

# The roles of MS2 RNA in MS2 capsid assembly

'Ottar Rolfsson

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

The University of Leeds  
Institute of Molecular and Cellular Biology

September 2009

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 overleaf. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

A PhD foundation year project served as a basis for the work presented in this thesis. The output from this project formed part of a jointly-authored publication. It is not discussed in detail in this thesis but is highly relevant for the work presented. Details of the publication are as follows:

*Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J., Homans, S. W., & Ashcroft, A. E. (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. J Mol Biol, 369(2), 541-552.*

**P.G.S.** provided advice on- and supervision of all experiments and wrote the manuscript. **O.R. (the candidate)** carried out all mass spectrometry analysis and contributed to writing/editing the manuscript (CWM). **G.S.T.** carried out NMR structural analysis and CWM. **G.B.** carried out light scattering analysis and CWM. **S.F.** carried out preliminary mass spectrometry analysis and CWM. **N.J.S.** CWM. **S.W.H.** provided supervision on NMR data analysis and CWM. **A.E.A.** provided advice, supervision and tuition on mass spectrometry data collection and analysis and CWM.

Part of the work detailed in chapter 3 has formed part of a jointly-authored publication.

Details of the publication are as follows:

*Rolfsson, O., Toropova, K., Morton, V., Francese, S., Basnak, G., Thompson, G.S., Homans, S. W., Ashcroft, A. E., Stonehouse, N. J., Ranson, N. A., & Stockley, P. G. (2008). RNA packing specificity and folding during assembly of the bacteriophage MS2. Computational and Mathematical Methods in Medicine: An Interdisciplinary Journal of Mathematical, Theoretical and Clinical Aspects of Medicine, 9(3), 339 – 349.*

**O.R. (the candidate)** carried out gel shift assays and mass spectrometry analysis and contributed to writing/editing the manuscript (CWM). **K.T.** carried out cryo-EM structural analysis and CWM. **V.M.** carried out mass spectrometry analysis and CWM. **S.F.** carried out preliminary mass spectrometry and gel shift assays with oligonucleotides. **G.B.** carried out light scattering analysis and CWM. **G.S.T.** carried out NMR structural analysis and CWM. **S.W.H.** provided supervision on NMR data analysis and CWM. **A.E.A.** provided advice, supervision and tuition on mass spectrometry data collection and analysis and CWM. **N.J.S.** CWM. **N.J.R.** provided advice, supervision and tuition on cryo-EM data collection and analysis and CWM. **P.G.S.** provided advice on- and supervision of all experiments and CWM.

The publications are supplied in the appendix.

## Acknowledgements

First and foremost I would like to thank my lovely, ever supporting and patient wife Sunna for sticking with me whilst undertaking this PhD project. Takk Sunna, eg elska þig meira. I would like to thank my supervisor Prof. Peter Stockley for his steady supervision, motivation and contagious enthusiasm throughout this project as well as for being a nice English chap. I would also like to acknowledge everybody in the Stockley lab. It's a fun mix of nerds from around the world of which I am one. A big thanks to Jamie for getting me sorted in the cloning and expression department, his expert knowledge of...everything and being a solid friend. Thank you to Dave for always being ready to answer questions and pretty much always knowing the answers. Thank you very much to everybody within the bionanogroup for their friendship and help, Zulfi and the gang...nanosilk will live forever. Thanks to Atul and Francis for trying to get me fit while in the lab. And to Francis for being the only true Englishman I met in England. A warm thank you very much to Gabby Basnak for being such a lovely and interesting compatriot. Thanks to Kat Toropova and Neil Ranson for all their helpful discussions on MS2 and providing the cryo-EM reconstructions which have complimented this project in so many ways. Thank you to all the people on the MS2 team especially Prof. Alison Ashcroft for helping me with my first project which turned out to be such a success. I would also like to thank Dr. Susan Schroder and Dr. Xiaobo Gu who performed the CE assays. I must also thank Dr. Stewart Warriner as without him I would never have come to sunny Leeds. Many more people have contributed to my PhD project, thank you to all. Finally I'd like to thank my mum and dad for always being there and my more recently acquired in-laws. I would like to acknowledge the Wellcome Trust for funding this project. It has been nice being encapsidated within the MS2 team and I will miss Leeds...no really I will.

## Abstract

Single strand (ss) RNA viruses are amongst the most prevalent viral pathogens in nature. A key event in the life cycle of many of these viruses is the packaging of their ssRNA genome into a capsid of defined size and shape. The mechanism by which genome packaging and capsid assembly proceeds is however poorly understood. Increased knowledge of this event is beneficial for novel anti-viral drug design, as well as contributing to our understanding of macromolecular assembly events. This project has explored the role(s) of the RNA genome in the capsid assembly process of the model ssRNA virus, bacteriophage MS2.

*In vitro* capsid reassembly reactions have been carried out using recombinant coat protein and ssRNA transcripts corresponding to different regions of the MS2 genome. These reactions have been assayed by size distribution analysis using native gel shift assays and sedimentation velocity analysis. This has allowed the effects of RNA size, sequence and structure on capsid assembly to be investigated. All the genomic RNAs transcripts, independent of sequence and size, promoted capsid assembly. The efficiency in which they each promote assembly was, however, different. This was shown to be due to the mechanism by which genomic RNA is packaged. It appears that coat proteins bind to RNA causing conformational changes that reduce its volume to that of the capsid interior. This was evident from the observed RNA length dependence on capsid assembly efficiency. Estimates of the hydrodynamic radii of assembly components and the inhibitory effect that ethidium bromide, a compound which stiffens RNA structure, has on capsid formation also supported this hypothesis.

The RNA structural transition was investigated using an RNA structure probing assay. The solution structures of the RNA transcripts were compared to the MS2 genome structure within the virion. Lead acetate was used to cause structure-specific cleavages within these RNAs which were then detected by reverse transcription using labelled primers. The results show that the RNA structure is partly conserved in solution and within the virion, implying that the conformational changes during encapsidation involve primarily tertiary structure rearrangement. The data suggest that the MS2 virion RNA has a defined structure within the virion. These results are consistent with cryo-electron microscopy of virions and capsids carried out by other members of the laboratory. One implication of this work is that compounds capable of inhibiting the conformational rearrangements required for virus assembly could serve as potent anti-viral therapeutics.

The work presented in this thesis has contributed to our understanding of how ssRNA is packaged into ssRNA virus capsids and, in particular, the roles it plays in capsid assembly.

# Table of Contents

ACKNOWLEDGEMENTS .....	3
ABSTRACT.....	4
TABLE OF CONTENTS .....	5
LIST OF FIGURES .....	9
LIST OF TABLES .....	12
ABBREVIATIONS.....	12
<b>1 INTRODUCTION.....</b>	<b>15</b>
1.1 CLASSIFICATION OF VIRUSES.....	15
1.2 SMALL ICOSAHEDRAL SSRNA VIRUSES AS RESEARCH MODELS.....	16
1.3 ARCHITECTURE AND SYMMETRY OF ICOSAHEDRAL VIRUSES.....	16
1.4 ICOSAHEDRAL CAPSID STRUCTURE .....	19
1.5 CAPSID ASSEMBLY .....	21
1.5.1 <i>Assembly is nucleated</i> .....	21
1.5.2 <i>Assembly and control of quasi equivalence</i> .....	23
1.6 THE ROLE OF GENOMIC RNA DURING CAPSID ASSEMBLY .....	25
1.7 SSRNA BACTERIOPHAGES .....	28
1.8 BACTERIOPHAGE MS2 .....	28
1.8.1 <i>The MS2 genome</i> .....	29
1.8.2 <i>The A-protein is important for phage infectivity</i> .....	31
1.8.3 <i>The replicase</i> .....	31
1.8.4 <i>The MS2 capsid and coat protein</i> .....	32
1.8.5 <i>The Lysis protein</i> .....	34
1.8.6 <i>Control of gene translation</i> .....	34
1.9 MS2 CAPSID ASSEMBLY .....	36
1.9.1 <i>Assembly induced by genomic length RNA</i> .....	36
1.9.2 <i>Assembly induced by TR</i> .....	37
1.9.3 <i>The MS2 virion structure and assembly</i> .....	40
1.10 THE AIMS OF THIS PROJECT.....	42
<b>2 MATERIALS AND METHODS .....</b>	<b>44</b>
2.1 MATERIALS.....	44
2.1.1 <i>General chemicals</i> .....	44
2.1.2 <i>Molecular biology products and reagents</i> .....	44

2.1.3	<i>Enzymes</i> .....	44
2.1.4	<i>Bacterial strains</i> .....	44
2.1.5	<i>Media and solutions</i> .....	45
2.1.6	<i>Vectors</i> .....	46
2.1.7	<i>Oligonucleotide primers</i> .....	46
2.2	<b>METHODS</b> .....	47
2.2.1	<i>Purification of MS2 recombinant capsids</i> .....	47
2.2.2	<i>Dissociation of MS2 capsids to produce MS2 CP</i> .....	48
2.2.3	<i>RT-PCR</i> .....	49
2.2.4	<i>PCR</i> .....	49
2.2.5	<i>In vitro transcription</i> .....	49
2.2.6	<i>Agarose gel electrophoreses</i> .....	50
2.2.7	<i>Isolation and purification of DNA and RNA</i> .....	50
2.2.8	<i>Cloning and sequencing of DNA</i> .....	51
2.2.9	<i>MS2 assembly reactions</i> .....	51
2.2.10	<i>Native agarose gel mobility shift assays (GEMSA assay)</i> .....	51
2.2.11	<i>Preparation of transmission electron microscopy grids</i> .....	52
2.2.12	<i>Sedimentation velocity assays</i> .....	52
2.2.13	<i>Lead acetate structure probing</i> .....	53
<b>3</b>	<b>PRODUCTION OF SUB-GENOMIC MS2 RNA FRAGMENTS AND THEIR EFFECTS ON MS2 CAPSID ASSEMBLY</b> .....	<b>54</b>
3.1	<b>INTRODUCTION</b> .....	<b>54</b>
3.1.1	<i>The selection of MS2 sub-genomic RNA fragments to study the effect on capsid assembly</i>	55
3.1.2	<i>Production strategy of the MS2 RNA sub-genomic fragments</i> .....	57
3.2	<b>RESULTS</b> .....	<b>60</b>
3.2.1	<i>Production of MS2 sub-genomic RNAs</i> .....	60
3.2.1.1	<i>Production of MS2 sub-genomic cDNAs by RT-PCR</i> .....	60
3.2.1.2	<i>Cloning and sequencing of the sub-genomic MS2 cDNAs</i> .....	64
3.2.1.3	<i>Introduction of T7 promoter sequence</i> .....	68
3.2.1.4	<i>Production of MS2 sub-genomic RNAs by in vitro transcription</i> .....	70
3.2.2	<i>The effect of MS2 sub-genomic RNA on MS2 capsid assembly</i> .....	72
3.2.2.1	<i>MS2 sub-genomic RNA promotes capsid like product formation</i> .....	72
3.2.2.2	<i>MS2 capsid assembly is cooperative</i> .....	79
3.2.2.3	<i>Smaller RNA is packaged more efficiently than larger RNA</i> .....	79
3.2.2.4	<i>The role of RNA structure and sequence on capsid assembly</i> .....	80
3.2.2.5	<i>MS2 capsid assembly assayed with light scattering</i> .....	82
3.2.2.6	<i>The 5' RNA induced capsid product shows RNase resistance</i> .....	83
3.2.2.7	<i>Assembled capsids contain the sub-genomic RNA</i> .....	85

3.3	DISCUSSION .....	87
3.3.1	<i>The production of MS2 sub-genomic RNA</i> .....	87
3.3.2	<i>Sub-genomic MS2 RNA nucleates MS2 capsid assembly</i> .....	89
<b>4</b>	<b>MS2 GENOMIC RNA ENCAPSIDATION STUDIED BY SEDIMENTATION VELOCITY ANALYSIS.....</b>	<b>93</b>
4.1	INTRODUCTION.....	93
4.1.1	<i>Sedimentation velocity analysis</i> .....	94
4.1.1.1	<i>Sedimentation boundary analysis</i> .....	96
4.2	RESULTS .....	100
4.2.1	<i>Characterisation of the components of MS2 capsid assembly by SV</i> .....	100
4.2.1.1	<i>The MS2 capsid and coat protein</i> .....	100
4.2.1.2	<i>Characterisation of genomic MS2 RNA</i> .....	104
4.2.2	<i>Analysis of MS2 capsid assembly by SV</i> .....	106
4.2.2.1	<i>Ethidium bromide inhibits folding of MS2 genomic RNA</i> .....	111
4.2.3	<i>Stem loops other than TR promote capsid assembly</i> .....	113
4.3	DISCUSSION .....	119
4.3.1	<i>MS2 RNA has an active role during MS2 capsid assembly</i> .....	119
4.3.2	<i>Assembly can be inhibited</i> .....	122
<b>5</b>	<b>COMPARISON OF MS2 RNA STRUCTURE <i>IN VIRION</i> VS. <i>IN VITRO</i>..</b>	<b>124</b>
5.1	INTRODUCTION.....	124
5.1.1	<i>RNA structure probing and lead ion hydrolysis</i> .....	126
5.1.2	<i>MS2 RNA structure probing strategy</i> .....	128
5.2	RESULTS .....	130
5.2.1	<i>Lead-induced hydrolysis of MS2 RNA assayed by PAGE</i> .....	130
5.2.1.1	<i>The solution structure of the MS2 sub-genomic RNAs is conserved</i> .....	134
5.2.1.2	<i>The MS2 RNA genome structure is partially conserved in virions</i> .....	136
5.2.1.3	<i>CP<sub>2</sub> binding can be mapped by lead ion induced hydrolysis</i> .....	137
5.2.2	<i>Lead induced hydrolysis of MS2 RNA assayed by CE</i> .....	141
5.2.2.1	<i>Assaying MS2 RNA lead induced hydrolysis positions with CE</i> .....	142
5.2.2.2	<i>Automated data analysis</i> .....	144
5.2.2.3	<i>CE assays confirm that MS2 RNA structure is partly conserved <i>in vitro</i> and <i>in virion</i></i> 145	
5.2.2.4	<i>Mapping CP<sub>2</sub>-RNA binding positions with CE</i> .....	147
5.2.2.5	<i>Comparison of CE and PAGE assay results</i> .....	151
5.3	DISCUSSION .....	153
5.3.1	<i>The MS2 genome has a defined structure within the MS2 virion</i> .....	153
5.3.2	<i>The structure of MS2 sub-genomic fragments is conserved in solution</i> .....	154

5.3.3	<i>Implications for MS2 capsid assembly</i> .....	155
5.3.4	<i>Genomic mapping of CP<sub>2</sub> binding sites</i> .....	156
5.3.5	<i>High throughput structural analysis of the MS2 genome structure</i> .....	157
<b>6</b>	<b>CONCLUSION</b> .....	<b>159</b>
	REFERENCES .....	166
	APPENDIX .....	179

## List of Figures

FIGURE 1-1. SPHERICAL CAPSIDS HAVE ICOSAHEDRAL SYMMETRY .....	18
FIGURE 1-2. THE T = 3 ICOSAHEDRAL CAPSID STRUCTURE OF TBSV .....	20
FIGURE 1-3. CAPSID ASSEMBLY IS A FINITE POLYMERISATION EVENT. ....	22
FIGURE 1-4. QUASI-EQUIVALENCE CAN BE INDUCED BY RNA BINDING. ....	24
FIGURE 1-5 THE RNA GENOME OF PAV IS ORGANISED AS A DODECAHEDRAL CAGE. ....	26
FIGURE 1-6. PROPOSED ASSEMBLY MECHANISM OF BPMV.....	26
FIGURE 1-7. GENETIC MAP OF THE MS2 GENOME.....	29
FIGURE 1-8. A PROPOSED SECONDARY STRUCTURE OF THE MS2 GENOME .....	30
FIGURE 1-9. THE STRUCTURE OF THE MS2 COAT PROTEIN. ....	33
FIGURE 1-10. THE STRUCTURE OF THE MS2 CAPSID.....	33
FIGURE 1-11. LEVIVIRUS RNA STRUCTURE CONTROLS GENE EXPRESSION.....	35
FIGURE 1-12. THE TR OPERATOR AND CP <sub>2</sub> -TR RIBONUCLEOPROTEIN COMPLEX .....	38
FIGURE 1-13. THE MS2 CAPSID ASSEMBLY MECHANISM.....	39
FIGURE 1-14. QUASI-EQUIVALENT CONTROL DURING CAPSID ASSEMBLY.....	40
FIGURE 1-15. THE ORGANISATION OF GENOMIC RNA WITHIN THE MS2 VIRION .....	41
FIGURE 3-1 SELECTION OF MS2 SUB-GENOMIC RNAS TO STUDY THE ROLE OF RNA IN MS2 CAPSID ASSEMBLY .....	56
FIGURE 3-2. THE IN VITRO STRATEGY FOR PRODUCTION OF MS2 CDNA STOCK. ....	58
FIGURE 3-3. THE PRODUCTION STRATEGY OF MS2 SUB-GENOMIC RNA FOR USE IN MS2 CAPSID ASSEMBLY REACTIONS. ....	59
FIGURE 3-4. ALIGNMENT OF THE THREE SUB-GENOMIC MS2 CDNA FRAGMENTS PRODUCED BY RT-PCR TO THE MS2 GENOME.....	60
FIGURE 3-5. RT-PCR OPTIMISATION WITH RESPECT TO MS2 RNA TEMPLATE CONCENTRATION.....	62
FIGURE 3-6. RT-PCR OPTIMISATION WITH RESPECT TO DNA POLYMERASE CONCENTRATION.....	63
FIGURE 3-7. RT-PCR OF THE 3'2578 BP AND 3528 BP SUB-GENOMIC MS2 CDNAs. ....	64
FIGURE 3-8. VECTOR CONSTRUCTS OF THE THREE CLONED SUB-GENOMIC MS2 CDNA FRAGMENTS.....	65
FIGURE 3-9. MULTIPLE SEQUENCE ALIGNMENT OF REPLICASE GENE PRODUCTS FROM SELECTED VIRUSES IN THE LEVIVIRIDAE FAMILY AND A HYPOTHETICAL REPLICASE GENE PRODUCT CARRYING ALL THE MUTATIONS IDENTIFIED IN THE REPLICASE	

CODING REGION FROM TABLE 3-1. ....	68
FIGURE 3-10. THE PRODUCTION OF SUB-GENOMIC MS2 cDNAs CONTAINING A T7 PROMOTER SEQUENCE .....	69
FIGURE 3-11. THE ALIGNMENT OF THE MS2 SUB-GENOMIC RNA TRANSCRIPTS TO THE MS2 GENOME.....	71
FIGURE 3-12. THE PURITY OF THE MS2 RNAs WAS ASSESSED ON DENATURING 1 % AGAROSE GELS. ....	71
FIGURE 3-13. IN VITRO CAPSID ASSEMBLY INDUCED BY MS2 VIRION RNA .....	75
FIGURE 3-14 IN VITRO CAPSID ASSEMBLY INDUCED BY MS2 5' SUB-GENOMIC RNA .....	76
FIGURE 3-15. IN VITRO CAPSID ASSEMBLY INDUCED BY THE 3' SUB-GENOMIC RNA.....	77
FIGURE 3-16. IN VITRO CAPSID ASSEMBLY INDUCED BY THE SUB-GENOMIC I RNA.....	78
FIGURE 3-17. IN VITRO CAPSID ASSEMBLY INDUCED WITH A 874 NT MS2 SUB-GENOMIC RNA LACKING THE TR OPERATOR.....	81
FIGURE 3-18. CAPSID ASSEMBLY ASSAYED WITH GEL FILTRATION COUPLED LIGHT SCATTERING .....	83
FIGURE 3-19. THE 5' SUB-GENOMIC RNA CAPSID PRODUCT SHOWS RNASE RESISTANCE. .....	84
FIGURE 3-20. MS2 CAPSID ASSEMBLY PRODUCTS CONTAIN SUB-GENOMIC RNA.....	86
FIGURE 4-1. SEDIMENTATION VELOCITY ANALYSIS .....	95
FIGURE 4-2. FORCES ACTING ON A SEDIMENTING PARTICLE IN SOLUTION UNDER AN APPLIED CENTRIFUGAL FIELD .....	96
FIGURE 4-3. SEDIMENTATION VELOCITY SIZE DISTRIBUTION ANALYSIS WITH THE C(S) MODEL .....	99
FIGURE 4-4. SEDIMENTATION COEFFICIENT DISTRIBUTIONS OF THE CP <sub>2</sub> , RECOMBINANT MS2 CAPSIDS AND MS2 VIRIONS. ....	101
FIGURE 4-5 SEDIMENTATION COEFFICIENT DISTRIBUTIONS OF THE MS2 GENOMIC RNAs .....	105
FIGURE 4-6. SEDIMENTATION COEFFICIENT DISTRIBUTIONS OF MS2 CAPSID ASSEMBLY REACTIONS INITIATED WITH THE GENOMIC MS2 RNAs .....	107
FIGURE 4-7. SV ASSAY OF ICRNA CONFIRMS THAT RNAs LACKING THE TR OPERATOR PROMOTE CAPSID ASSEMBLY .....	109
FIGURE 4-8. THE 5' RNA IS FOLDED INTO THE VOLUME OF THE MS2 CAPSID.....	112
FIGURE 4-9. GEMSA ASSAY OF CAPSID ASSEMBLY WITH THE 5' RNA FRAGMENT WITH AND WITHOUT 30 mM EtBr.....	112

FIGURE 4-10. THE RNA OLIGONUCLEOTIDES USED TO STUDY THE SEQUENCE SPECIFICITY OF QUASI-EQUIVALENT CONFORMER SWITCHING.....	114
FIGURE 4-11. NORMALISED SEDIMENTATION COEFFICIENT DISTRIBUTIONS OF MS2 CAPSID ASSEMBLY REACTIONS INITIATED WITH TR, F6, F7 AND QB. ....	116
FIGURE 4-12. SV QUANTIFICATION OF THE CAPSIDS PRODUCED WITH THE RNA STEM LOOPS.....	117
FIGURE 4-13. THE MS2 SUB-GENOMIC RNAs INTERACT WITH THE CAPSID INTERIOR IN A MANNER WHICH IS CONSISTENT WITH QUASI-EQUIVALENT CONFORMER SWITCHING .....	121
FIGURE 5-1. HOW DOES RNA FOLDING AFFECT RNA STRUCTURE?.....	125
FIGURE 5-2. THE PROPOSED MECHANISM OF LEAD ION HYDROLYSIS OF RNA. ....	127
FIGURE 5-3. A SCHEMATIC OF THE LEAD ION CLEAVAGE ASSAY .....	129
FIGURE 5-4. STRUCTURALLY PROBED REGIONS OF THE MS2 VIRION RNA AND THE SUB-GENOMIC RNA FRAGMENTS.....	131
FIGURE 5-5. LEAD ION PROBING OF MS2 VIRION RNA AND THE SUB-GENOMIC RNAs..	133
FIGURE 5-6. LEAD ION CLEAVAGE MAP OF THE 1450-2190 NUCLEOTIDE REGION OF THE MS2 GENOME. ....	135
FIGURE 5-7. MS2 RNA STRUCTURE IS CONSERVED IN VITRO AND IN VIRION. ....	136
FIGURE 5-8. LEAD ION CLEAVAGE MAP OF THE 1450-2200 NT REGION OF THE MS2 GENOME PRODUCED FROM PAGE RESULTS.....	139
FIGURE 5-9. LEAD ION CLEAVAGE POSITIONS MAPPED ONTO A PROPOSED SECONDARY STRUCTURE OF THE MS2 GENOME.....	140
FIGURE 5-10. CAPILLARY ELECTROPHORETIC ANALYSIS OF THE $Pb^{2+}$ INDUCED MS2 RNA CLEAVAGE POSITIONS.....	143
FIGURE 5-11. THE STRUCTURE OF THE MS2 SUB-GENOMIC RNAs IS PARTLY CONSERVED IN SOLUTION.....	146
FIGURE 5-12. A LEAD ION INDUCED CLEAVAGE MAP OF THE 515-2615 AND 2950-3486 NT REGIONS OF THE MS2 GENOME .....	148
FIGURE 5-13. COMPARISON OF THE NUMBER OF IDENTICAL $Pb^{2+}$ INDUCED CLEAVAGE POSITIONS OCCURRING WITHIN THE 1419-2160 NT REGION OF THE MS2 GENOME. ....	149
FIGURE 5-14. A LEAD ION INDUCED CLEAVAGE MAP OF THE 515-2615 AND 2950-3548 NT REGIONS OF THE MS2 GENOME .....	150
FIGURE 5-15. THE PAGE AND CE ASSAYS OF LEAD ION INDUCED CLEAVAGE CORRELATE WELL .....	152

## List of Tables

TABLE 1-1. THE BALTIMORE CLASSIFICATION OF VIRUSES.....	16
TABLE 2-1. MEDIA. ....	45
TABLE 2-2. SOLUTIONS. ....	45
TABLE 2-3. THE SEQUENCE SPECIFIC PRIMERS USED IN THE PRODUCTION OF MS2 cDNA CLONES AND IN REVERSE TRANSCRIPTION OF LEAD ION CLEAVAGE REACTION PRODUCTS .....	47
TABLE 4-1. THE HYDRODYNAMIC PROPERTIES OF THE MS2 CAPSID ASSEMBLY REACTION COMPONENTS .....	103
TABLE 4-2. THE OBSERVED QUANTITIES OF RNA AND CAPSIDS IN THE MS2 CAPSID ASSEMBLY REACTIONS ASSAYED BY SV .....	108
TABLE 4-3. HYDRODYNAMIC PROPERTIES OF THE OLIGONUCLEOTIDE SERIES .....	114
TABLE 4-4. SEDIMENTATION COEFFICIENTS OF IDENTIFIED COMPONENTS IN THE ASSEMBLY REACTIONS INDUCED WITH RNA STEM LOOPS .....	116

## Abbreviations

Aa	Amino acid
Å	$10^{-10}$ meters
ASU	Asymmetric subunit
AUC	Analytical Ultracentrifugation
BBV	Black Beetle Virus
b	Base
bp	Base pair
BPMV	Bean Pod Mottle Virus
BSA	Bovine serum albumin
BMV	Brome Mosaic Virus
cDNA	Complementary DNA
CE	Capillary electrophoresis
CP	Coat protein
CP <sub>2</sub>	Coat protein dimer

Cryo-EM	Cryo electron microscopy
Da	Dalton
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTPs	Deoxyribonucleotide triphosphates
ds	Double strand
DTT	Dithiothreitol
EDTA	Ethylen diamine tetraacetic acid
EtBr	Ethidium bromide
FHV	Flock House Virus
g	Grams
GEMSA	Gel mobility shift assay
h	Hour
hf	Host factor
HIV	Human Immunodeficiency Virus
IPTG	Isopropylthiogalactoside
iRNA	internal RNA
icRNA	internal control RNA
k	Kilo
l	Litres
M	Molar
min	Minutes
mRNA	MessengerRNA
nt	Nucleotide
PaV	Pariacato Virus
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RNase	Ribonuclease
rNTPs	Ribonucleotide triphosphates
rpm	Revolutions per minute
rt	Room temperature
S	Svedgberg or $10^{-13} \text{ sec}^{-1}$
SDS	Sodium dodecylsulphate

SDS-PAGE	Sodium dodecylsulphate polyacrylamide gel electrophoresis
sec	Seconds
ss	Single strand
SV	Sedimentation velocity analysis
T	Triangulation number
TCV	Turnip Crinkle Virus
TBSV	Tomato Bushy Stunt Virus
TEM	Transmission electron microscopy
TYMV	Turnip Yellow Mosaic Virus
T <sub>m</sub>	Melting temperature
UTR	Untranslated region
UV	Ultraviolet
U	Unit
vRNA	MS2 virion RNA
wt	Wild type
X-gal	5-bromo-4-chloro-3-indolyl- $\beta$ -galactoside

# 1 Introduction

---

Viruses are organisms that infect all forms of life at the cellular level. They are parasitic genes that have evolved to ensure their own replication which often results in deleterious effects on the host. In their simplest form they are composed of nucleic acid and a capsid comprised of multiple protein subunits. The nucleic acid encodes the information for construction of the molecular machinery necessary to infect and multiply within their host, while the protein capsid serves as a protective container during nucleic acid transport between host organisms (Wagner *et al.*, 2008).

Virus research has been at the forefront of discoveries into the properties of proteins and nucleic acids and the deciphering of biological mechanisms such as replication, translation and transcription (Summers, 2006). Underlying this is that viruses allowed the isolation of large quantities of homogenous protein or nucleic acid which represented an obstacle prior to the invention of heterologous expression systems (Schneemann, 2003). Viruses still represent model systems where molecular events are readily accessible to experiment. One of these events is the assembly of viral capsids, the understanding of which is beneficial towards novel anti-viral drug design as well as increasing understanding of macromolecular interactions and assembly events.

This thesis reports results where the MS2 virus has been used as a model system to investigate the formation of viruses containing a ssRNA genome enclosed in an icosahedral capsid. In particular it addresses how the viral genome interacts as a whole with multiple proteins to self assemble into a container of defined size and shape.

## 1.1 Classification of viruses

Nearly 5500 viruses are classified in a hierarchical manner by the International Committee on Taxonomy of Viruses (ICTV) based on their nucleic acid content, virion morphology and host type (Wagner *et al.*, 2008). In this way viruses are classified into orders, families, sub-families, genera and species. It is not known whether all viruses have a common ancestor and therefore a true linear classification is not possible. A similar but more manageable way of classifying viruses was proposed by David Baltimore (1971), in which viruses are classified solely on the way they produce

messenger RNA and their method of replication. In this manner viruses are categorised into six groups as outlined in **Table 1-1**.

**Table 1-1. The Baltimore classification of viruses.** *ssRNA viruses can be positive or negative strand based on whether the RNA within the virion can be directly translated into protein upon cell entry.*

Group	Genome	Example
I	dsDNA	Adenovirus
II	ssDNA	Parvovirus
III	dsRNA	Bluetongue virus
IV	<b>(+)ssRNA</b>	<b>MS2 phage</b>
V	(-)ssRNA	Rabies virus
VI	ssRNA with DNA intermediate	HIV
VIII	dsDNA with RNA intermediate	Hepatitis B

## 1.2 Small icosahedral ssRNA viruses as research models

Icosahedral ssRNA viruses represent ~ 15 % of all viruses classified by the ICTV. These viruses are responsible for diseases in animals and humans such as the common cold, gastroenteritis, foot and mouth disease, yellow fever and rubella. Because of the pathogenic nature of viruses, working with them presents a health hazard. This hazard can be circumvented by working with viruses with similar characteristics that are non-infectious to humans. A great deal of knowledge concerning virus life cycles, structure and assembly has been gathered on small (<50 nm diameter) icosahedral viruses infecting plants (Rossmann and Johnson, 1989) and bacteria (Calendar, 2006). As a result they are very well characterised both structurally and biochemically and therefore represent ideal model systems for virus research.

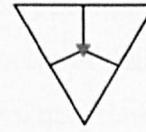
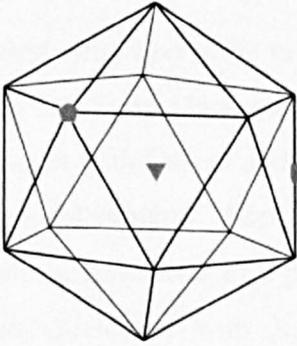
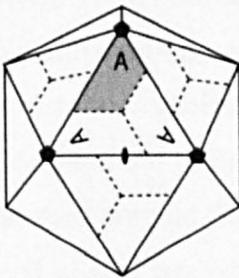
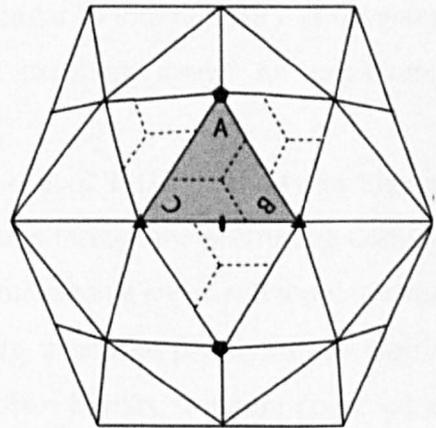
## 1.3 Architecture and symmetry of icosahedral viruses

In order to gain appreciation for virus structure and assembly an understanding in basic viral architecture is essential.

In 1956 Crick and Watson proposed that plant viruses, such as Turnip yellow mosaic virus (TYMV), which had been observed to be spherical by electron microscopy, would have the geometric form of a platonic solid (for example: tetrahedron, octahedron, icosahedron) and therefore be composed of asymmetric subunits (ASU) of a fixed number which could not exceed 60 (**Figure 1-1 A**). They

further suggested that using multiple copies of a single polypeptide would be the most efficient way of enclosing a genome with a limited coding capacity (Crick and Watson, 1956). X-ray diffraction studies of viruses confirmed that spherical viral capsids were highly ordered and adhered to icosahedral symmetry, however, they were not necessarily composed of just 60 subunits (Caspar, 1956). In a biological context this posed the question of how identical proteins, whose interactions with each other within the capsid were presumably identical, could generate a capsid of more than 60 subunits as this was not allowed based on the rules underlying icosahedral symmetry.

Realising the dynamic nature of biological macromolecules, Caspar and Klug (1962) devised the concept of quasi-equivalence. They predicted that capsids could be comprised of increasing numbers of identical protein subunits if these subunits could adapt very similar yet distinct conformations: quasi-equivalent conformations. Essentially they realised that the protein interactions within the capsid need not be the same between each and every protein subunit. In this way, capsids adhering to icosahedral symmetry of increasing size could be produced by increasing the number of identical protein subunits comprising the capsid. Caspar and Klug defined the arrangement of protein subunits necessary to form capsids of ever increasing sizes with the triangulation number  $T$  (Figure 1-1 B).  $T$  is defined by  $T = h^2 + hk + k^2$  where  $h$  and  $k$  are integers that define the positioning of the capsid five fold axes relative to one another (Caspar and Klug, 1962; Johnson and Speir, 1997). The theory of quasi-equivalence has proved highly successful at predicting virus capsid structure and furthermore allows the immediate appreciation of the size and protein organisation within a capsid through the concept of the  $T$  number (Harrison, 2001).

**A****3 ASU****B****T=1****T=3**

**Figure 1-1. Spherical capsids have icosahedral symmetry.** An icosahedron can only be constructed from a defined number of asymmetric building blocks. **A)** An icosahedron has 2-, 3- and 5-fold symmetry axes labelled as an ellipse, triangle and pentagon, respectively. These axes are used to define the positioning of protein subunits which comprise an icosahedral capsid. Each of the 20 triangular facets of the icosahedron is comprised of 3 asymmetric sub units (ASU). **B)** Icosahedral symmetry constraints and quasi-equivalence theory restrain the number of protein building blocks to  $60T$ , where  $T$  is the number of protein subunits comprising an ASU. A  $T = 1$  capsid is comprised of 60 identical protein subunits, where each protein corresponds to an ASU. A  $T = 3$  capsid is comprised of 180 protein subunits, where each ASU is comprised of three proteins. The areas highlighted in grey correspond to the ASU in the  $T = 1$  and  $T = 3$  capsids.

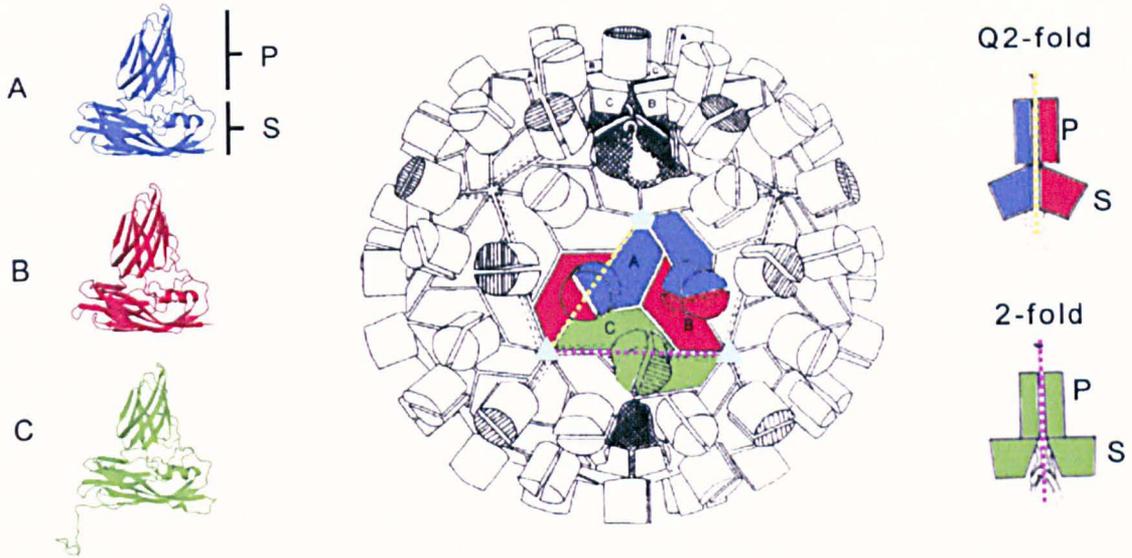
## 1.4 Icosahedral capsid structure

The first virus ever to be viewed at atomic resolution was Tomato bushy stunt virus (TBSV) by Dr. Steve Harrison and co-workers at Harvard. The structure of the capsid was consistent with Caspar and Klug's theory of virus structure in that it was comprised of 180 identical subunits observed in three distinct environments (Harrison *et al.*, 1978). Since then, the structures of a plethora of icosahedral viruses infecting all forms of life have been elucidated with X-ray crystallography and more recently cryo-electron microscopy (cryo-EM). Interestingly, many of these viruses, in particular viruses infecting eukaryotes, have structural characteristics similar to that of TBSV (Rossmann and Johnson, 1989; Harrison, 2001) making it an excellent model for explaining conserved features of viral structure.

The structural organisation of the capsid coat proteins of TBSV is shown in **Figure 1-2**. The coat protein (CP) is composed of two domains termed the protruding domain (P) and the shell domain (S). The S domain of the protein has a swiss roll topology and is composed of two sets of four anti-parallel  $\beta$ -strands, which lie parallel to the capsid surface. The swiss roll is the most common coat protein tertiary structure observed in ssRNA viruses infecting eukaryotes including Turnip crinkle virus (TCV) (Hogle *et al.*, 1986), Brome mosaic virus (BMV) (Lucas *et al.*, 2002), Norovirus (Prasad, *et al.*, 1999) and Rhinovirus (Rossmann *et al.*, 1985). This has been used as evidence for a common ancestry of these viruses (Rossmann and Johnson, 1989), however given the different amino acid (a.a.) sequence that underlie these folds, and that this protein topology is not exclusively found in viruses it is possible that the similarity is the result of convergent evolution to a highly advantageous protein structure (Williams and Westhead, 2002).

The main site of structural variation observed between the A, B and C quasi-equivalent conformations in TBSV is seen in the N-terminal region of the coat proteins. This region is ordered in the C-subunit but disordered in the A and B subunits. In the C subunit, the N-terminus is interlocked with two other N-termini of different C subunits related by 3-fold symmetry, forming a structure termed the  $\beta$ -annulus at the 3-fold axes. These interactions result in two types of dimer conformations, A/B and C/C which coalesce at different angles and thereby allow the successful closure of the capsid. The importance of the N-termini for quasi equivalent switching has since been observed for various virus particles where deletion of the N-termini results in defective virus particles or none at all (Erickson and Rossmann, 1982; Sorger *et al.*, 1986; Calhoun *et al.*, 2007).

This is however not applicable to all capsids composed of proteins with the jelly roll fold implying that quasi-equivalence is induced in different ways dependent on the virus type ( Prasad *et al.*, 1999; Satheshkumar *et al.*, 2005). The  $\beta$ -annulus 3-fold complex and its implications for capsid assembly are further discussed in the next section.



**Figure 1-2. The  $T = 3$  icosahedral capsid structure of TBSV.** From left, the capsid is comprised of 180 CP seen in three distinct environments termed A, B and C. The protein is comprised of two domains, a protruding domain (P) and a shell domain (S). The fold of the shell domain is conserved in various eukaryotic virus structure proteins. The N-terminus of the CP can be seen protruding downwards from the S domain of the C conformer. The organisation of the CP in the capsid is shown in a figure modified from Harrison *et al.* (1978). The CP are observed as two types of dimers within the capsid, A/B and C/C. The angle at which these dimers coalesce is different and serves to incorporate curvature into the capsid. This is highlighted to the right with a view perpendicular to the quasi 2-fold and 2-fold symmetry axes by which the subunits comprising the dimers are related. The quasi 2-fold and 2-fold axes are designated with dashed yellow and magenta lines, respectively. Understanding the manner in which these two dimer conformations are attained to produce an icosahedral capsid is key to understanding capsid assembly.

Not all icosahedral ssRNA capsid structures adhere to Caspar and Klug's theory of quasi-equivalence. The capsids of poliovirus and rhinovirus, for example are composed of four chemically distinct proteins: VP1-4. These proteins are however geometrically arranged in a similar manner as in a  $T = 3$  capsid, wherein VP1-3 all have the jelly roll topology and are interchanged for CP conformers 1-3 as observed in the TBSV structure. Capsids that are composed of chemically distinct coat proteins are said to be pseudo symmetric and designated with  $P$  numbers.

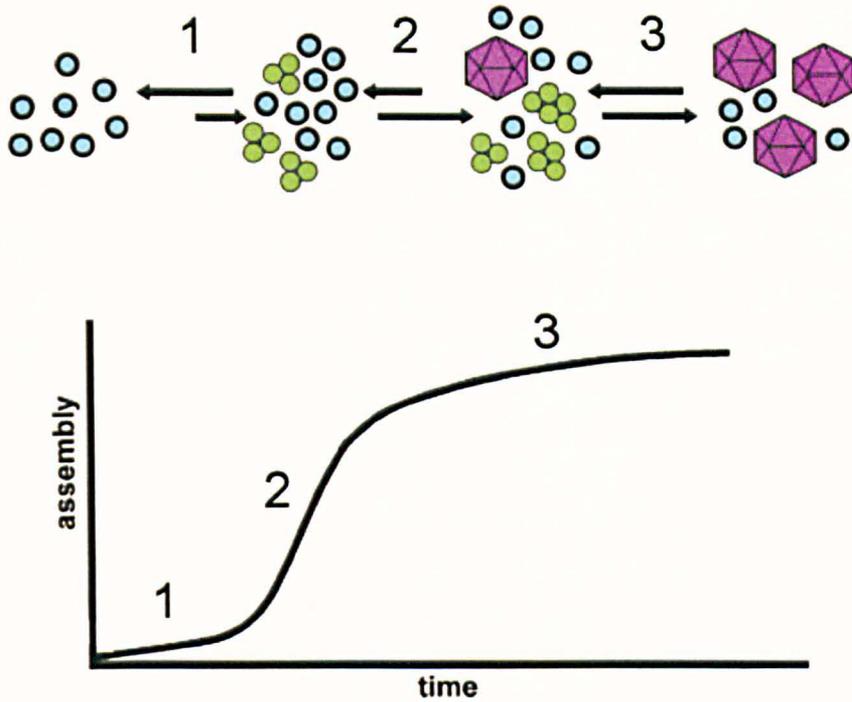
## 1.5 Capsid assembly

### 1.5.1 Assembly is nucleated

Prior to the availability of atomic resolution structures of viruses it had been shown that purified viral coat proteins of  $T = 3$  viruses infecting plants and bacteria could self assemble *in vitro* into structures resembling their respective capsids in the absence of viral RNA (Wagner *et al.*, 1971; Matthews and Cole, 1972). This implied that all the information necessary for constructing a capsid was intrinsic to the protein. Based on this, capsid assembly, from multiple subunits of identical proteins, is analogous to a finite crystallisation event. The only requirement for such a process is the formation of a nucleating species (Prevelige *et al.*, 1993).

The kinetics of capsid formation for small icosahedral viruses is consistent with a crystallisation event (**Figure 1-3**). The sigmoidal curve indicates that the rate of the reaction increases with time and that this is due to an increase in concentration of reactive species over time. This implies that assembly is initially slow but speeds up following an initial nucleation phase and as more binding sites become available in the growing capsid (Prevelige *et al.*, 1993; Zlotnick, 2005).

Because successful virus formation requires RNA to be packaged, a nucleation complex involving viral RNA is reasonable. Indeed, early *in vitro* capsid assembly reactions reported that assembly was induced by RNA at low concentrations under conditions where capsids otherwise would not form (Hung *et al.*, 1969; Hohn, 1969; Matthews and Cole, 1972). These results implied that specific coat protein binding to RNA might act to nucleate assembly. Capsid assembly has since been shown to be promoted by coat protein binding to specific RNA sequences within the virus genome thereby allowing selective RNA packaging. Such sequences are referred to as packaging sites and have been identified in viruses such as TCV (Sorger *et al.*, 1986), Flock house virus (FHV) (Zhong *et al.*, 1992), Human immunodeficiency virus (HIV-I), (Aldovini *et al.*, 1990) and bacteriophages as discussed in section 1.8.



**Figure 1-3. Capsid assembly is a finite polymerisation event.** Capsid assembly kinetics are consistent with a multistep reaction preceded by a nucleation event. Capsid subunits (blue) associate to form a nucleation complex (green) as evident by a lag phase during assembly. Once a critical concentration of nucleating species or assembly intermediates is reached, they react quickly and irreversibly with other subunits or intermediates to form capsids. The figure is modified from Zlotnick and Stray (2003).

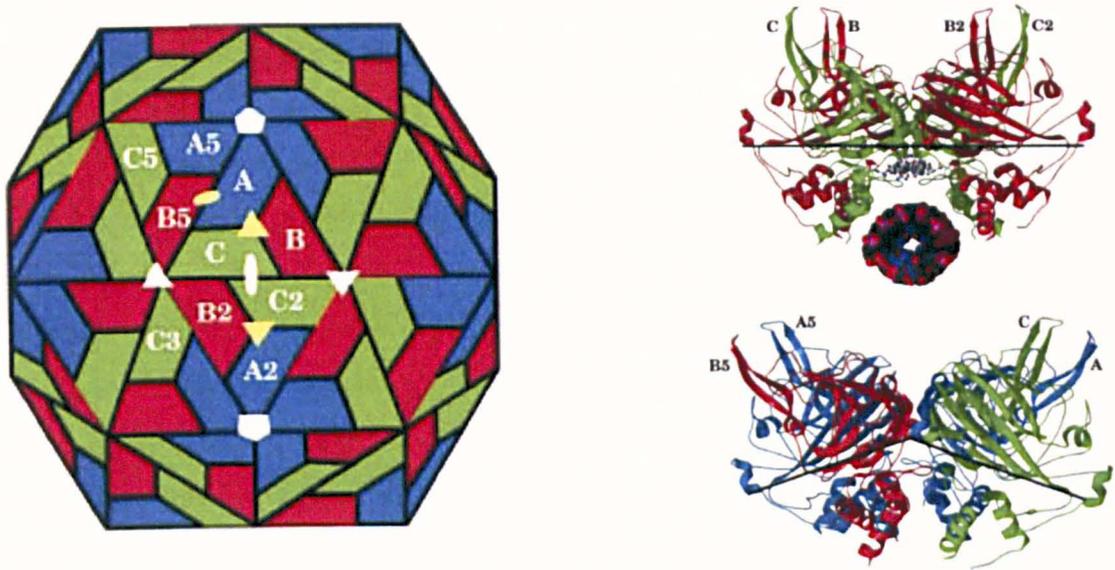
### 1.5.2 Assembly and control of quasi equivalence

The crystal structures of TBSV and other plant viruses that followed made it clear that the nucleating species for capsid assembly would most likely be associated with formation of the 3-fold or 5-fold symmetry axes. This was based on the observation that arranging capsid subunits in oligomeric pentamers, relating to the 5-fold axes or oligomeric hexamers, relating to the 3-fold axes, serves to define the overall curvature of the shell (Rossmann, 1983). This suggestion further implied that quasi conformer control would have to take place early during capsid assembly and would presumably be linked to the formation of the nucleating complex.

This was observed for Turnip crinkle virus (TCV), a close relative of TBSV, wherein capsid assembly was proposed to proceed from a nucleation complex of 3 C/C dimers forming a  $\beta$ -annulus complex at the capsid 3-fold axes following specific RNA binding. The RNA supposedly acts as a scaffold which facilitates assembly by aligning the C/C dimers in space in such a way that facilitates  $\beta$ -annulus formation. Following formation of the  $\beta$ -annulus, capsid assembly would then proceed by coat protein dimer addition as C/C or A/B conformers dependent upon whether they could bind to free N-termini already incorporated in the growing capsid shell. Proteolytic cleavage of the N-terminus resulted in  $T = 1$  capsid particles suggesting that the N-terminus is crucial for formation of the planar contacts at the 3-fold axes while the curvature necessary for formation of the 5-fold axes is intrinsic to the protein (Hogle *et al.*, 1986; Sorger *et al.*, 1986).

The mechanism proposed for TCV assembly demonstrated for the first time how quasi-equivalence is controlled through CP dimer interactions. The role of RNA in this process has further been highlighted in that deletion of the  $\beta$ -annulus does not necessarily result in particle polymorphism (Bertolotti-Ciarlett *et al.*, 2002; Satheshkumar *et al.*, 2005; Tang *et al.*, 2006). Conformational switching due to RNA binding was first observed for the nodaviruses FHV (Fisher and Johnson, 1993) and Black beetle virus (BBV) (Wery *et al.*, 1994) where it was shown that specific RNA binding acts as a conformational switch between flexed A/B like dimers and planar C/C like dimers (**Figure 1-4**). Similar flexed and planar dimers are observed for TBSV and TCV suggesting that RNA might contribute to quasi equivalence control in a similar manner for these plant viruses.

Quasi equivalent control of coat protein conformations is a mutually induced process where the contributions of the RNA-coat protein interactions and intra coat protein subunit interactions can vary. This is dependent upon virus type due to the manner in which protein subunits, such as the N-termini at the 3-fold axis, interact and stabilise quasi-equivalent conformers.

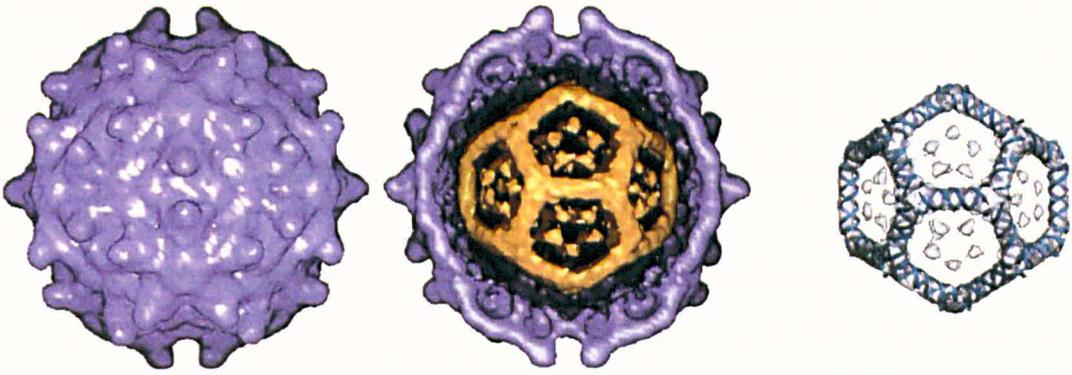


**Figure 1-4. Quasi-equivalence can be induced by RNA binding.** From left: BBV has a  $T = 3$  capsid composed of 180 subunits labelled A,B and C. The icosahedral symmetry axes are labelled with a white oval (2-fold), yellow oval (quasi 2-fold), white triangle (3-fold) yellow triangle (quasi 2-fold) and pentagon (5-fold). To the right, a view perpendicular to the 2-fold (top) and quasi 2-fold (bottom) symmetry axes shows how binding of an RNA A-type duplex wedges the protein dimer into a C/C conformation. The binding of RNA acts to rigidify the subunit contacts at the 2-fold axes thereby defining the planar 3-fold axes. Note that a  $T = 3$  capsid is composed of a mixture of bent and planar subunits, if only bent A/B subunits are used a  $T = 1$  capsid composed of 30 A/B dimers as 12 pentamers is produced. Similarly if only planar C/C dimers are used a flat sheet would result. The figure is adapted from Reddy et al. (1998).

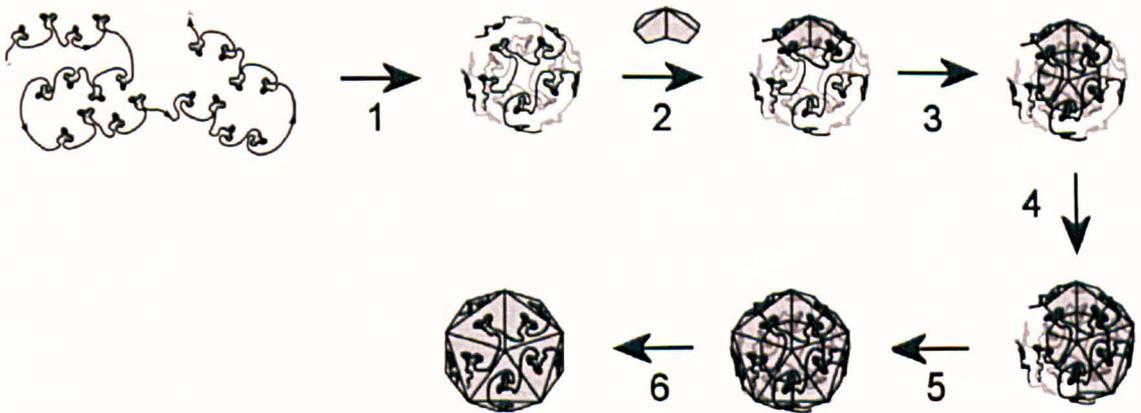
## 1.6 The role of genomic RNA during capsid assembly

RNA can act in at least two ways during capsid formation, namely it can initiate capsid assembly via a sequence recognition event and it can act as a conformational switch controlling protein conformations necessary to form a capsid of defined shape. The packaging of RNA into capsids and how RNA contributes to this process is poorly understood. This lack of basic knowledge arises from the observation that once assembly is initiated, the assembled capsid is the only stable species on the assembly pathway of transient species that are only present in limiting amount (Zlotnick *et al.*, 1994; Stockley *et al.*, 2007). This inevitably restricts research into characterising assembly reaction substrates and products. Structural analysis of large “substrate” RNA molecules in solution is experimentally difficult in part because of their size and asymmetry which does not allow their three dimensional structural characterisation with the modern analytical techniques commonly applicable to proteins. Characterisation of capsid “products” with X-ray crystallography and more recently cryo-EM has however been moderately successful and provided information on the structural organisation of RNA genomes within viral capsids, allowing hypotheses concerning RNA packaging to be made.

Using a combination of X-ray crystallography and cryo-EM, 23 % of the genomic RNA residing within the capsid of the nodavirus Pariacoto virus (PaV), closely related to FHV and BBV, was observed as a discontinuous dodecahedral cage directly beneath the capsid (Tang *et al.*, 2001) (**Figure 1-5**). Interestingly, the organisation of the RNA suggests a considerable deviation from the minimum free energy “tree like” structures associated with RNA. These results revitalised interest into understanding how RNA is reorganised into a compact and complex tertiary structure not normally associated with RNA (reviewed in Schneemann, 2006). A similar study into the plant virus Bean pod mottle virus (BPMV) revealed a non-random discontinuous nucleotide consensus sequence repeated more than 60 times within its ssRNA genome (Lin *et al.*, 2003) (**Figure 1-6**). These findings suggest a nucleotide template, beyond packaging signal recognition, which could help promote efficient capsid assembly as proposed earlier for TCV (Sorger *et al.*, 1986).



**Figure 1-5** The RNA genome of PaV is organised as a dodecahedral cage. From left, a cryo-EM reconstruction of PaV virus. Removal of half of the protein capsid shows the dodecahedral arrangement of RNA within the virion. X-ray crystallography of PaV shows that the dodecahedral cage is comprised of thirty 25 bp A-type RNA duplexes located underneath the C/C dimers at the 2-fold axes. RNA is also observed at the quasi 2-fold axes underneath the A/B dimers which is observed as islands of density surrounding the 5-fold axes. Note that the dodecahedron reflects structural ordering beneath the C/C dimers and is an icosahedrally averaged view of the genome. The figure is adopted from Tang et al. (2001).



**Figure 1-6.** Proposed assembly mechanism of BPMV. The non random nucleic acid sequence APuPyPyPyX is repeated in the BPMV genome suggesting that multiple packaging signals within the genome direct assembly. The viral genome folds to form a structural scaffold (1) which can direct capsid assembly via coat protein binding to secondary structural elements (2-6). The figure is modified from Lin et al. (2003)

A specific nucleotide sequence is however not always a requirement for capsid assembly. In pioneering work by Bancroft and Hohn into *in vitro* capsid assembly of viruses with  $T = 3$  capsids, viral coat proteins were observed to readily encapsidate not only their own RNA but also heterologous and non-viral RNA as well as polyanions such as polyvinyl sulfonate under conditions where viral particles do not otherwise form (Bancroft *et al.*, 1969; Hohn, 1969). Furthermore, follow-up research into the structure

of heterologous RNA within recombinant PaV particles suggests that the PaV genome is not a specific determinant of three dimensional RNA structure observed within the virion (Johnson *et al.*, 2004; Tihova *et al.*, 2004) (**Figure 1-5**). In these cases it seems that coat proteins are capable of imposing a structure upon RNA that is compatible with the capsid.

RNA has however also been shown to impose specific capsid morphology onto virus particles. For example, when recombinant BMV coat protein is expressed *in vivo* it forms  $T = 1$  virus particles composed of 60 coat protein dimers which exclusively contain the 1 kb non-replicating RNA encoding the coat protein. Interestingly the dimers are non-equivalent and interact within the capsid in distinctly different ways than what is seen in regular  $T = 3$  BMV capsids composed of 90 dimers enclosing approximately 3 kb of RNA. This suggests that RNA is capable of directing viral ribonucleoprotein complex formation (Krol *et al.*, 1999).

The structural plasticity of RNA and the hierarchical manner in which it folds, which is intrinsic to the nucleotide sequence (Tinoco Jr and Bustamante, 1999), suggests two scenarios for RNA packaging. In the first, viral RNA can fold on its own accord into a compact scaffold upon which capsid assembly can take place. With respect to RNA structure predictions based on the thermodynamics of nucleotide base pairing (Rudnick and Bruinsma, 2005; Yoffe *et al.*, 2008) along with presumed electrostatic self repulsions of the phosphate backbone (van der Schoot and Bruinsma, 2005) it is unlikely that an RNA genome can adopt a structure that is compatible with the capsid in this way. Investigations into the stability of the  $T = 1$  Satellite tobacco mosaic virus (STMV) capsid and genome have however indicated that an “RNA first” mechanism can not be ruled out (Kuznetsov *et al.*, 2005, Freddolino *et al.*, 2006). In the second packaging scenario, the coat protein subunits bind to the RNA and fold it into an assembly competent structure driven by non-modular interactions between the CP subunits and charge neutralisation of the phosphate backbone. It is likely that the structural co-evolution of RNA and capsid proteins has led to a highly synergistic virus assembly mechanism (Schneemann, 2006), however the details of this mechanism are in question.

So far, virus structure and assembly of simple icosahedral viruses has been discussed with respect to viruses with common structural features, accounted for by the common swiss roll capsid protein fold. Interactions between coat protein N-terminal domains and with viral RNA are of importance for controlling capsid assembly and

morphology of these virus types. How quasi equivalence is controlled within a virus capsid along with the details of genome packaging within the particle are old but yet still key questions at the forefront of structural virology (Harrison *et al.*, 1978). The bacteriophage MS2 has provided some of the most detailed answers to these questions. MS2 is distinct from the viruses already discussed in that the structure of the coat protein is unique to bacteriophages, however it still adheres to  $T = 3$  morphology and RNA acts to promote its assembly. This thesis concerns itself with the manner in which RNA promotes MS2 capsid assembly and the structural nature of the RNA genome during MS2 virus assembly. The MS2 virus is reviewed in the next sections.

## 1.7 ssRNA bacteriophages

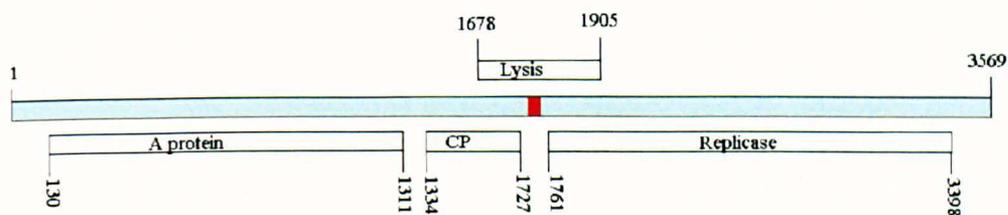
The ssRNA bacteriophages were discovered in 1961 by Loeb and Zinder through screening of raw New York sewage for viruses whose infectivity was confined to male strains of *Escherichia coli* (*E.coli*) and therefore dependent upon F-pili (Loeb and Zinder, 1961).

ssRNA bacteriophages form the virus family *Leviviridae* (or lightweight viruses) which is split into the genera *Levivirus* and *Allolevivirus* based on genome size (3.5 kb vs. 4.2 kb) and gene organisation (four vs. three genes). The *Levivirus* genus is further split into subgroups I and II based on sequence homology and serological properties within the genus. The *Allolevivirus* genus is similarly split into subgroups III and IV. MS2 (male specific, fraction 2) and Q $\beta$  of sub groups I and III respectively, represent the best characterised viruses of the *Leviviridae* family.

## 1.8 Bacteriophage MS2

The bacteriophage MS2 is composed of a 3569 nt RNA genome that encodes four gene products: an A-protein (or maturation protein), a coat protein, a lysis protein and a replicase apoprotein (**Figure 1-7**) (Fiers *et al.*, 1976; Beremand and Blumenthal, 1979). The MS2 phage is an excellent example of genetic economy where the limiting coding capacity of the genome has resulted in a read-through overlap of the lysis gene with respect to the CP and replicase genes (Beremand and Blumenthal, 1979). The A protein is preceded by a 129 nt 5' untranslated region (UTR) which is involved in control of gene translation (Groeneveld *et al.*, 1995) while the 3'UTR of 171 nt has unknown

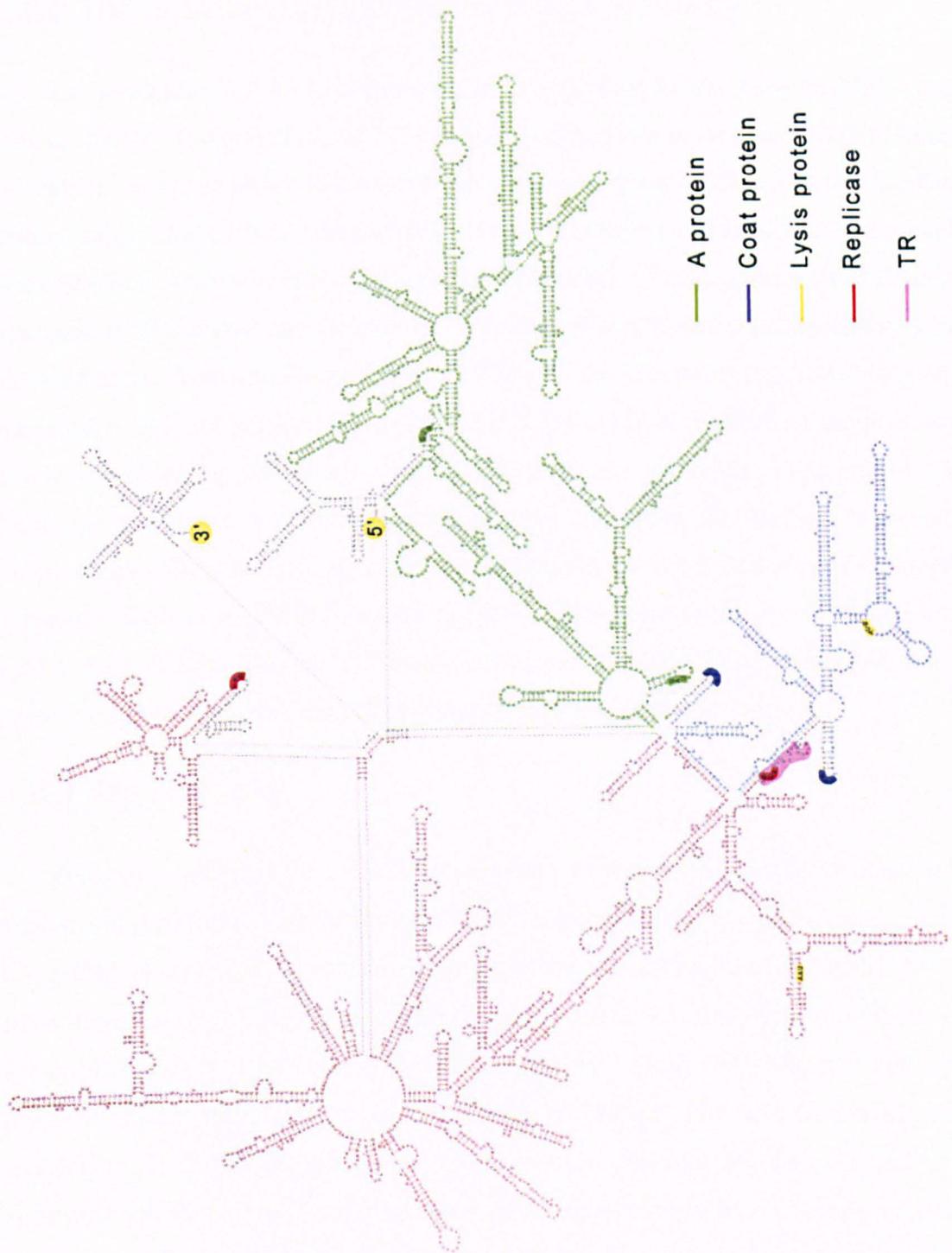
function. The genome is enclosed in a capsid composed of 180 copies of CP (see section 1.8.4) and a single maturation protein which together define the MS2 virion.



**Figure 1-7. Genetic map of the MS2 genome.** The nucleotide positions of the genes within the genome are obtained from the wild type MS2 sequence, nucleotide accession number NC\_001417.

### 1.8.1 The MS2 genome

MS2 was the first living organism to have its genome sequenced and structurally characterised by partial enzymatic digest. 75-85 % of the genome is proposed to be base paired, forming an intricate network of RNA secondary structural elements (Fiers *et al.*, 1976; Thomas *et al.*, 1976; Skripkin *et al.*, 1990) (**Figure 1-8**). The degree of secondary structure of the RNA is higher than can be expected on the basis of a random nucleotide sequence suggesting that the RNA has evolved towards compactness (Fiers, 1975; Yoffe *et al.*, 2008). The secondary structure of the genome is implicated in gene translation and RNA replication (reviewed in van Duin, 2006). Furthermore, the RNA hairpin structures, which are of abundance in the proposed RNA structure, are thought to act as potential binding sites for the CP during capsid assembly (see section 1.9.3). The structural role of the MS2 genome for capsid assembly is the main focus of chapters 4 and 5 with the goal of defining the structural transition of RNA during packaging and mapping discontinuous CP binding sites within the RNA genome.



**Figure 1-8.** A proposed secondary structure of the MS2 genome. The 5' and 3' termini are highlighted with yellow circles. The coding regions of the A-protein, coat protein and replicase are coloured in green, blue and red respectively. The start and stop codons of the genes are highlighted in the same colours. The lysis gene start and stop codons are highlighted in yellow. The TR operator, which is implicated in capsid assembly and replicase translation, is shaded in magenta. The figure is adapted from Fiers et al. (1975, 1976)

### 1.8.2 The A-protein is important for phage infectivity

The A-protein is a 44 kDa protein which is present in one copy in MS2 virions (Steitz, 1968). The infectivity of MS2 virions is dependent upon their ability to bind F-pili which in turn is dependent upon a functioning A-protein (Crawford and Gestland, 1965; Lodish *et al.*, 1965; Steitz, 1968). The protein forms a ribonucleoprotein complex with genomic RNA which is infectious in the absence of coat protein (Richelson and Nathans, 1967; Leipold and Hofschneider, 1976). Ribonuclease protection assays have shown that the A-protein interacts specifically with the genome at positions centered on nucleotides 393 and 3515 (Shiba and Suzuki, 1981). This, along with its importance for F-pilus binding, suggests that it protrudes both in and out of the MS2 capsid. The molecular details of A-protein association with the capsid are unclear, however its positioning on the capsid has recently been shown to be at the 5-fold axes of icosahedral symmetry (Hill *et al.*, 1997; Toropova, 2009). The mechanism by which the virion enters the cell after F-pilus attachment is not understood although the RNA and A-protein are known to enter the cell together (Kozak and Nathans, 1971).

### 1.8.3 The replicase

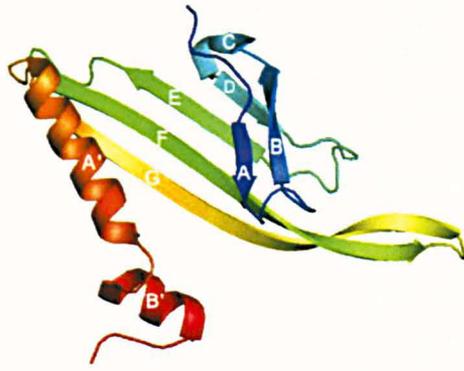
Following cell entry the (+) ssRNA genome serves as a template for translation and subsequent replication by a newly synthesised replicase. The (+) ssRNA and resulting (-) ssRNA is replicated in tandem. The replicase is a 60 kDa protein which forms a RNA dependent RNA polymerase with ribosomal subunit S1 and elongation factors Ef-Tu and Ef-Ts of the host cell. An additional host cell factor (Hf), which is not part of the holoenzyme, stimulates replication of the (+) ssRNA. The host factors S1, Ef-Tu and Hf are all implicated in template binding but to different degrees. S1 and Hf are required for replication of viral (+) ssRNA. In their absence RNA replication occurs but is non specific (reviewed in van Duin, 2006). S1 is further implicated in gene translation as discussed in 1.8.6. What is known of replicase function is based upon research into replicases from bacteriophages GA and Q $\beta$  of groups II and III of the *Leviviridae* family, respectively (Blumenthal *et al.*, 1976; Yonesaki and Haruna, 1981). These replicases are similar in terms of structure and replicate RNA within their genera. Replicase function is therefore assumed to be similar within the *Leviviridae* family (Yonesaki *et al.*, 1982).

### 1.8.4 The MS2 capsid and coat protein

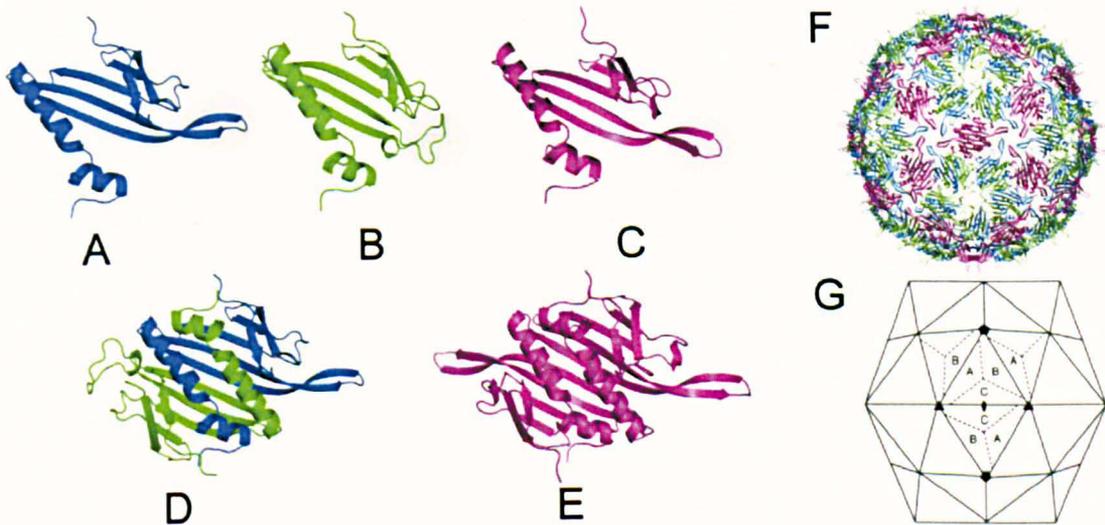
The MS2 capsid is a  $T = 3$  icosahedral protein container that is 27.5 nm in diameter. It is composed of 180 copies of the coat protein. The crystal structure of the MS2 virus was solved by Lars Liljas and co-workers which showed the details of the CP fold (**Figure 1-9**) and the arrangement of CP within the capsid (**Figure 1-10**) (Vålegard *et al.*, 1990; Golmohammadi *et al.*, 1993). The structure of the CP is conserved within the *Leviviridae* family (Plevka *et al.*, 2009). The MS2 CP fold is different from that observed for coat proteins of viruses infecting eukaryotes (**Figure 1-2**). This is most obvious from the lack of features corresponding to a P- domain and an N-terminal arm which are common in viral coat proteins with a jelly roll fold.

The capsid is composed of CP subunits observed in three distinct conformations as defined by a 16 aa polypeptide loop connecting the F and G  $\beta$ -strands (FG-loop) (**Figure 1-10, A-C**). The conformations of the A, B and C subunits define the quasi-equivalent capsid interactions. These interactions are different at the 3- fold and 5-fold symmetry axes. In the A and C subunits, the FG-loops are extended and interact at the capsid 3-fold axes forming a pore of approximately 1.6 nm. In the B subunits however the FG-loop is folded towards the capsid exterior forming a pore of 1.7 nm. It is through these pores that the A-protein is thought to make contact with viral RNA. The pores also render the capsid permeable to small molecules. This was exploited in experiments reported in chapter 5 where the resistance of the RNA genome within the capsid to lead ion induced hydrolysis was used as a tool to investigate genome structure.

The three CP subunits form two types of quasi-equivalent dimers ( $CP_2$ ) (**Figure 1-10, DE**). The capsid is composed of 60 A/B dimers and 30 C/C dimers. The CP is stable as a dimer in solution (Beckett and Uhlenbeck, 1988; Stockley *et al.*, 2007). CP subunits inter-digitate via their C-terminal  $\alpha$ -helices. The C-terminal  $\alpha$ -helices of one CP subunit sit in a hydrophobic groove of the neighbouring CP subunit. The groove is formed by the N-terminal A and B  $\beta$ -sheets and the C-terminal  $\alpha$ -helices (**Figure 1-10 DE**). The dimer is further stabilized by hydrogen bonds between G  $\beta$ -sheets of the interacting CP subunits. The  $CP_2$  presents a continuous, relatively flat surface of ten anti-parallel  $\beta$ -strands to the interior of the capsid which are implicated in RNA binding (Peabody, 1993).



**Figure 1-9. The structure of the MS2 coat protein.** The MS2 CP is a folded as a five-stranded anti-parallel  $\beta$ -sheet (labelled C-G) with two anti-parallel  $\beta$ -strands folding over it at the N terminus (labelled A & B) and a kinked  $\alpha$ -helix at the C terminus (labelled A' & B'). It is 13.7 kDa and composed of 129 aa. A cartoon model of the A coat protein subunit is shown. The figure is based on the pdb file 2MS2 (Golmohammadi et al., 1993).



**Figure 1-10. The structure of the MS2 capsid.** The MS2 capsid is composed of 180 CP subunits observed in three quasi-equivalent conformations termed A, B and C (their colour coding is constant throughout this thesis). The main site of structural variation between the subunits is in the conformation of the loop of polypeptide, highlighted in grey, connecting the F and G  $\beta$ -strands. The CP subunits form two types of non-covalent dimers termed A/B (D) and C/C (E). The dimers are shown as viewed from the capsid exterior. The positioning of the dimers within the capsid determines its size and symmetry. The capsid is shown in F as viewed down the 2-fold axes. The MS2 capsid adheres to  $T = 3$  icosahedral symmetry as shown in G. The A/B dimers lay on the quasi 2-fold axes and span the distance between the capsids icosahedral 5- and 3-fold axes. The C/C dimers lay on the 2-fold axes and span the distance between the capsids 3-fold axes. The figure is modified from Toropova et al., (2008).

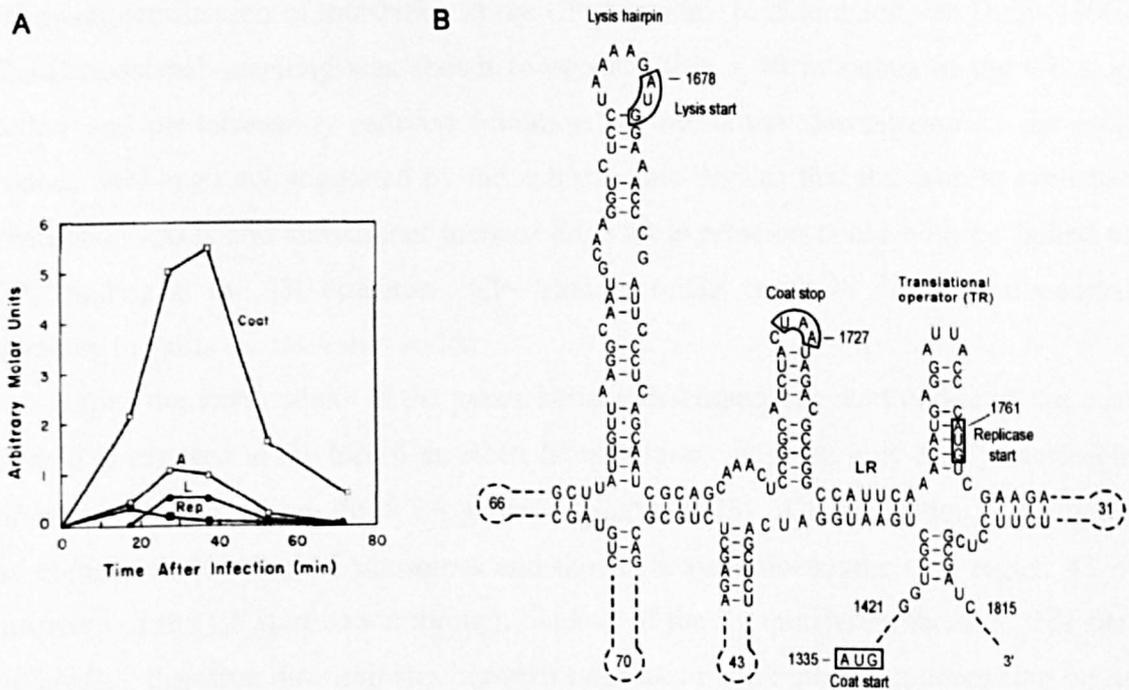
### 1.8.5 The Lysis protein

The lysis protein is a small 8.7 kDa membrane protein which is crucial for release of phage progeny (Beremand and Blumenthal, 1979). The cell lysis mechanism is not fully understood. It is thought to act by inhibiting peptidoglycan biosynthesis resulting in cell wall breakdown on the basis of its sequence homology to the lysis protein of the ssDNA phage  $\phi$  174. The A2 protein of *Allolevivirus* Q $\beta$  has been shown to function in a similar manner (reviewed in Bernhardt *et al.*, 2002). The control of translation of the lysis protein is crucial for effective phage release and indeed the translation of all the MS2 proteins is tightly regulated at the RNA level.

### 1.8.6 Control of gene translation

Radioisotope labelling of the f2 bacteriophage at multiple timepoints following infection have shown that viral proteins are expressed at different stages and to a different degree during infection (Beremand and Blumenthal, 1979) (**Figure 1-11A**). This effect is also observed *in vitro* (Lodish, 1968). This ensures correct protein stoichiometries for virion assembly and efficient use of host cell resources in order to maximise phage progeny at the time of cell lysis.

Regulation of translation has been shown to be dependent upon the structure of RNA flanking individual start codons. By involving nucleotides in stable secondary structures that otherwise recruit translational machinery to an RNA cistron, translation is effectively averted. Down regulation of the A-protein is achieved in this manner. The Shine-Delgarno sequence upstream of the A-protein gene is base paired in full length viral RNA which inhibits translation except off nascent RNA molecules (Steitz, 1973; Fiers *et al.*, 1975; Groeneveld *et al.*, 1995). Translation of the lysis gene and replicase genes is controlled in a similar manner. Here, however, the time frame for initiation of translation at the lysis and replicase cistrons is controlled by ribosomal readthrough of the CP gene rather than folding of the 5' RNA terminus as is the case for the A-protein (**Figure 1-11 B**) (Min Jou *et al.*, 1972; Kastelein *et al.*, 1982; Berkhout and van Duin, 1985). The ability of ribosomes to translate these proteins is therefore governed largely by RNA folding kinetics following either RNA replication or CP translation.



**Figure 1-11. Levivirus RNA structure controls gene expression.** *A*) *f2* protein expression levels following infection. *f2* is a group I Levivirus like MS2. The figure is adapted from (Beremand and Blumenthal, 1979). *B*) MS2 secondary structure in the region of the lysis and replicase start codons. The start codon of the CP, lysis and replicase genes are boxed. Lysis gene translation is dependent upon disruption of the lysis hairpin. Replicase gene expression is dependent upon disruption of the long range base paired stem (LR). These secondary structure elements obstruct ribosome binding to the start codons of the lysis and replicase genes but are disrupted upon ribosomal readthrough of the CP gene. CP<sub>2</sub> also inhibit replicase translation when bound to the TR operator. The figure is adapted form (van Duin, 2006).

An additional level of translational control is in place on the viral replicase. When the coat protein concentration becomes sufficiently high in the cell, it binds to a translational operator (TR) stem loop (**Figure 1-11B**) and acts as a translational repressor (Lodish and Zinder, 1966; Ward *et al.*, 1967; Bernardi and Spahr, 1972). This binding has dual functionality as it also initiates capsid assembly as detailed in 1.9.2. Replicase gene expression is therefore kept at low levels reflecting the catalytic nature of the molecule and is abolished when the viral life cycle enters its virion construction phase.

It is not known how lysis expression is kept low at early times of infection. However, in addition to ribosomal readthrough of the CP gene, translation of the lysis gene is dependent upon the positioning of the CP stop codon (Berkhout *et al.*, 1987). This has been explained by lateral diffusion or “scanning” of ribosomes on MS2 RNA

following termination of translation at the CP stop site. In Adhin and van Duin (1990) *E.coli* ribosomal scanning was shown to occur within a 40 nt radius of the CP stop codon and predominantly induced initiation of translation downstream to the stop codon. Although not suggested by the authors, this implies that the drop in replicase expression levels and subsequent increase in lysis expression could both be linked to CP<sub>2</sub> binding at the TR operator. CP<sub>2</sub> binding might result in directing ribosomal scanning towards the lysis start codon.

Unlike the start codons of the genes already discussed, the start codon of the coat protein is exposed at the top of an RNA hairpin loop. It is the only freely accessible ribosome binding site on the RNA genome (Steitz, 1973). CP translation is regulated by competitive binding of ribosomes and the replicase holoenzyme to a region 43 nt upstream of the CP start codon through binding of the S1 translation factor at this site. S1 binding therefore discriminates between translation and replication depending on its association with either the replicase holoenzyme or the ribosome (reviewed in van Duin 2006).

## 1.9 MS2 capsid assembly

MS2 virions are dissociated upon treatment with 60 % (v/v) glacial acetic acid which allows separation of CP from RNA and A-protein. The CP is easily purified by centrifugation followed by buffer exchange to provide CP<sub>2</sub> that can self assemble into capsids *in vitro* (Herrmann *et al.*, 1968). CP<sub>2</sub> can similarly be obtained by acid dissociation of recombinant MS2 capsids produced upon overexpression of the CP gene contained within a replicating plasmid in *E.coli* (Mastico *et al.*, 1993). Self assembly is dependent upon CP concentration with an optimum pH of 5.5-7.5. Capsid yield increases at lower temperatures and increasing salt concentration (Matthews and Cole, 1972).

### 1.9.1 Assembly induced by genomic length RNA

Assembly is promoted by RNA under conditions where self assembly of CP does not take place (Sugiyama *et al.*, 1967; Hohn, 1969; Matthews and Cole, 1972). Sugiyama and co-workers (1967) investigated MS2 capsid assembly *in vitro* by mixing various stoichiometries of radio labelled CP to viral RNA and analysed the components

by sucrose density centrifugation. They showed that virion RNA induces formation of a ribonucleoprotein complex (complex I) at low CP:RNA ratios which is comprised of 1-6 CP subunits. Complex I was specific to MS2 RNA as it could not be detected with rRNA, tRNA or poly-adenylic acid. At higher CP:RNA ratios, capsids were observed with morphology similar to MS2 phage as observed by transmission electron microscopy (TEM) but sedimenting slower than the virion, most likely due to the absence of the A-protein. As only these two components could be observed it was concluded that capsid assembly was co-operative and most likely took place through an initiation complex that might also be responsible for viral replicase gene suppression reported earlier (Lodish and Zinder, 1966).

Similar experiments were reported by Hohn (1969) in which a series of sub-genomic RNAs and non-native viral RNAs were assessed for their ability to form *Levivirus* fr capsids. Remarkably, all RNAs tested were capable of inducing capsid formation suggesting that a specific RNA sequence and structure were not a requirement for capsid assembly. Assembly in the presence of nucleotide phosphate alone however did not take place indicating that a polymer was required to string together CP subunits to form a capsid. As with MS2 (Sugiyama *et al.*, 1967) the assembly nucleating ability of non-native RNA was not associated with formation of an initiation complex such as could be seen with fr RNA, suggesting specific interactions between fr RNA and fr CP.

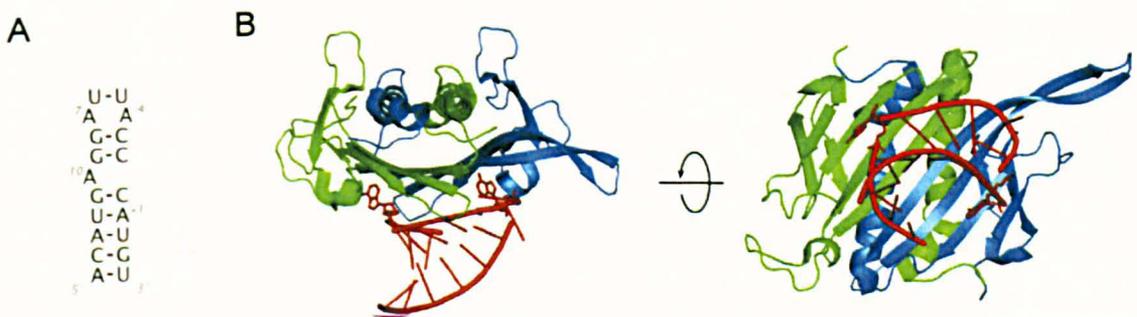
These results were among the first suggesting that selective viral RNA encapsidation and icosahedral capsid assembly would be dependent upon sequence specific recognition. For MS2, knowledge of sequence specific RNA binding and its implication for capsid assembly has provided one of the most detailed descriptions of capsid assembly at the molecular level.

### ***1.9.2 Assembly induced by TR***

Ribonuclease digestion of ribonucleoprotein complex I formed by viral RNA and coat protein established that coat protein binding covers a 59 nt region incorporating the replicase start codon (Bernardi and Spahr, 1972). Further analysis showed that the binding affinity was specific to a 19 nt stem loop operator (TR) incorporating the replicase start codon within the hairloop stem (Gralla *et al.*, 1974; Carey *et al.*, 1982a). The association constant for CP<sub>2</sub> binding to TR as compared to the RNA genome is

similar as observed by filter binding assays (Carey *et al.*, 1982b). As a consequence the TR operator is implicated in initiating capsid assembly *in vivo* (Ling *et al.*, 1970) as well as acting as a translational repressor of the replicase gene as already described in 1.8.6.

The CP<sub>2</sub>-TR interaction has been characterised both biochemically and structurally allowing the factors underlying RNA binding specificity to be defined. RNA binding discrimination is dependent upon the RNA adhering to a stem loop structure while binding affinity is dependent upon the nucleotide sequence (**Figure 1-12**) (Romaniuk *et al.*, 1987; Valegård *et al.*, 1994; Stockley *et al.*, 1995; Valegård *et al.*, 1997; Horn *et al.*, 2006).

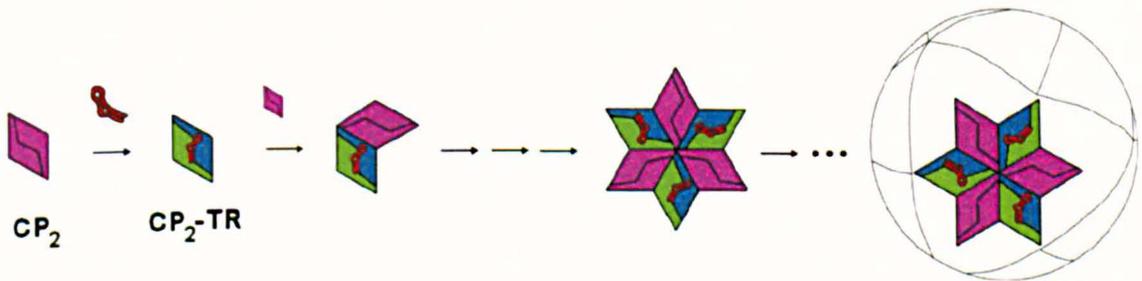


**Figure 1-12. The TR operator and CP<sub>2</sub>-TR ribonucleoprotein complex.** *A*) TR forms a 19 nt stem loop structure with a bulged adenine at position -10 and a hairpin of four nucleotides. The numbering is with respect to the replicase start codon (numbered +1). *B*) A cartoon model of the A/B dimer bound to TR (coloured red) is shown in side view and rotated 90 ° to show the dimer from the capsid interior (Pdb file: 2C4U, Grahn *et al.*, 2001). Crystal structures, obtained by soaking TR and oligonucleotide variants of TR into recombinant capsid crystals, along with site directed mutagenesis have shed light on the molecular details of CP<sub>2</sub>-RNA binding. TR forms a crescent like structure that is stabilised by a series of hydrogen bonds between CP<sub>2</sub> aa residues within the 10 strand β-sheet to seven phosphate groups at the 5' half of TR (Valegård *et al.*, 1997). For high affinity binding, the -4 and -10 nucleotide positions must be adenines (Romaniuk *et al.*, 1987) although exceptions do occur (RowSELL *et al.*, 1998). The two adenines (modelled as sticks in the figure) extrude from the stem loop and bind into shallow pockets on each of the two CP subunits through a series of intermolecular hydrogen bonds and hydrophobic contacts. Having a pyrimidine at the -5 position has also been shown to be beneficial for strong binding (Grahn *et al.*, 2001).

MS2 capsid assembly can be induced *in vitro* by addition of TR to purified CP<sub>2</sub> (Beckett and Uhlenbeck, 1988). Assembly initiation is not dependent upon the TR operator sequence as similar stem loops are capable of inducing assembly, however higher concentrations of oligonucleotide variants of TR and CP<sub>2</sub> are generally required.

This reflects the high affinity of the CP<sub>2</sub> for the TR stem loop and suggested that the CP<sub>2</sub>-TR complex initiates capsid formation (Beckett and Uhlenbeck, 1988; Beckett *et al.*, 1988). Recent investigations into MS2 capsid assembly in the presence of TR have provided molecular details of how this event occurs. Using a combination of gel chromatography, mass spectrometry and NMR the TR operator was shown to induce the switching between quasi-equivalent CP<sub>2</sub> conformations thereby allowing assembly to proceed efficiently (Stockley *et al.*, 2007).

In Stockley *et al.* (2007) it was observed that upon mixing CP<sub>2</sub> with TR at a stoichiometric ratio of 1:1, capsid assembly was inefficient with the dominating species in the assembly reaction corresponding to the CP<sub>2</sub>-TR complex as assayed by non-covalent mass spectrometry. Upon further addition of CP<sub>2</sub> to bring the stoichiometric ratio to 2:1, capsid assembly proceeded through a hexameric or pentameric intermediate by rapid CP<sub>2</sub> addition. These results implied that two types of CP<sub>2</sub> were required for efficient assembly, an RNA bound and RNA free coat protein dimer and that in the absence of free CP<sub>2</sub> capsid assembly is kinetically trapped as the CP<sub>2</sub>-TR complex. Subsequent NMR structural analysis of a coat protein mutant unable to assemble past a dimer, showed that RNA binding results in a conformational change within the FG-loop region from a symmetric C/C-like dimer to an asymmetric A/B-like dimer. These findings resulted in a proposed model for *in vitro* MS2 capsid assembly where TR acts as an allosteric switch to provide the two types of quasi equivalent conformers in solution both of which are necessary for efficient capsid assembly (**Figure 1-13**).

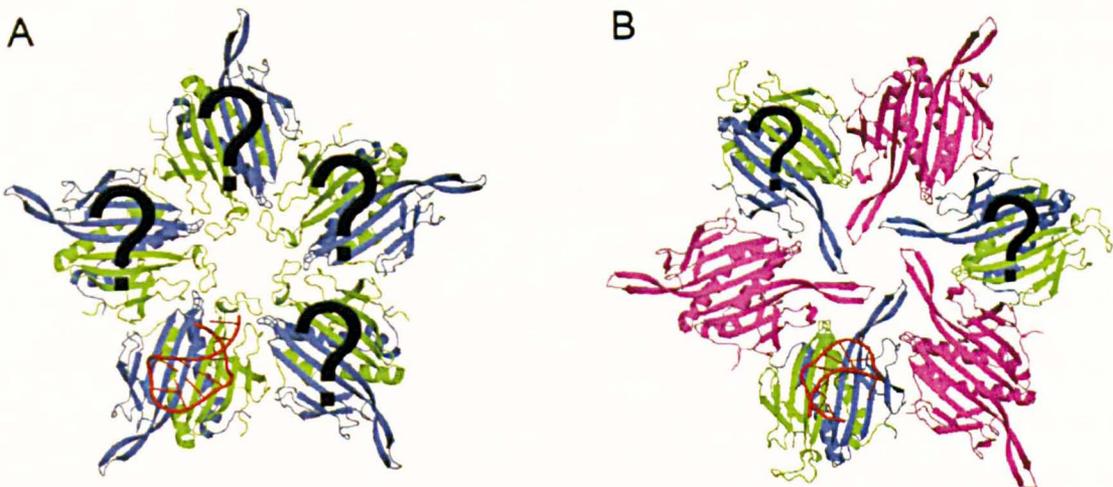


**Figure 1-13. The MS2 capsid assembly mechanism.** In the absence of TR, coat protein dimers are predominantly C/C like and capsid formation is negligible. The TR operator acts as a molecular switch providing the A/B-like CP<sub>2</sub> conformation in solution whose absence otherwise has a rate limiting effect on capsid assembly. Using non-covalent mass spectrometry, components of the assembly reaction were identified. Assembly takes place through a hexameric or pentameric intermediate by rapid dimer addition. These intermediates correspond to dimer arrangements around the capsid 3-fold or 5-fold symmetry axes. Formation of capsids through a hexameric intermediate excludes the pentameric intermediate during capsid assembly and vice versa. The hexameric intermediate is shown here. The figure is adapted from Stockley *et al* (2007).

The mass spectrometry side of the work reported in Stockley *et al.* (2007) represent results from a three month laboratory project that I undertook during a foundation year at the beginning my PhD and are part of my PhD project. It sparked further interest in capsid assembly and macromolecular assembly mechanisms which resulted in the project reported in this thesis. It was decided not to discuss these results in detail in this thesis. The publication is however supplied in the Appendix.

### 1.9.3 The MS2 virion structure and assembly

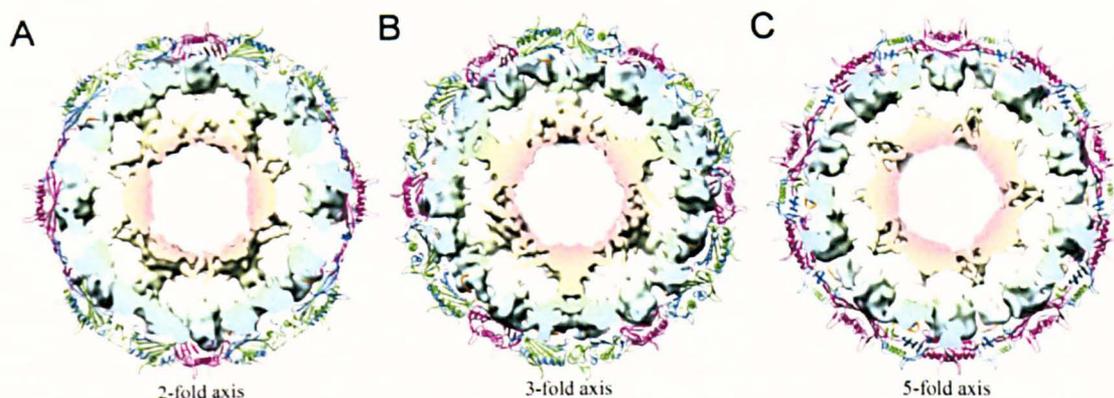
The *in vitro* assembly mechanism proposed for MS2 provides a detailed description of quasi equivalent control during capsid assembly. However, the situation *in vivo* must be different. The reactions reported in Stockley *et al.* (2007) were performed at close to equimolar ratios of CP<sub>2</sub> to TR raising the question of how quasi equivalence conformer switching is achieved during packaging of the full length viral genome ( **Figure 1-14**).



**Figure 1-14. Quasi-equivalent control during capsid assembly.** The two capsid intermediates identified by mass spectrometry corresponding to the capsid 5-fold (A) and 3-fold axis (B) are shown in cartoon model with TR bound at a single A/B dimer position. How are the conformations of A/B like dimers controlled *in vivo* when there is only one copy of the TR operator in the full length genome?

Recent cryo-EM reconstructions of viruses of the *Levivirus* genus have indicated that multiple sites within genomic RNA interact with the capsid interior (Konig *et al.* 2003; van den Worm *et al.*, 2006). This suggests that stem loop structures other than TR could act to control the conformation of an incoming CP<sub>2</sub> during capsid assembly. These results correspond with the idea that strong RNA binding is not necessary to

induce capsid assembly (Beckett *et al.*, 1988). A higher resolution cryo-EM reconstruction of the MS2 phage was recently reported by Toropova and co-workers.



**Figure 1-15. The organisation of genomic RNA within the MS2 virion.** A cryo-EM density map showing a 40 Å thick central region of the MS2 genome as observed down the 2-fold axes (A), the 3-fold axes (B) and the 5-fold axes (C). The crystal structure of the MS2 capsid is shown in cartoon representation enclosing the genomic RNA which is radially coloured from pale blue to pale red. The genome is folded as two concentric shells with roughly two thirds of the genome located in the outer shell and the remainder residing in the inner shell. Columns of density connect the two shells at the icosahedral 5-fold axes which can be observed in A and B. These columns are large enough to accommodate a single RNA duplex. Icosahedral averaging does not allow accurate assessment of how often RNA transcends the distance between the two RNA shells. The figure is adopted from Toropova *et al.* (2008).

In Toropova *et al.* (2008) an icosahedrally averaged cryo-EM structure of the MS2 virion at sub-nanometer resolution was reported. Interestingly, a high degree of ordered density was observed within the capsid shell which accounted for roughly 90 % of the RNA genome (**Figure 1-15**). The icosahedrally averaged model provided for the first time the details of RNA genome packaging within a  $T = 3$  viral capsid. The RNA genome is observed as a complex fold of two concentric shells connected at the 5-fold axes of icosahedral symmetry. The outermost RNA shell lies directly beneath the capsid forming a network that maps the binding sites for RNA on the inner surface of the capsid. The density at these positions is consistent with the positioning of the CP<sub>2</sub>-RNA binding site. Beneath the A/B dimers, density is observed that is extended towards the 5-fold axes forming a ring of continuous density around the 5-fold axes. Averaged density is also observed beneath every C/C dimer, however, it is distinct from the density associated with A/B dimer. This suggests that RNA interactions at these positions are different and is consistent with the idea of RNA induced conformer

control. The location of the outer shell of RNA is therefore suggestive of an assembly process assisted by repeated conformer switching to A/B dimers at correct positions to form the 5-fold axes (Toropova *et al.*, 2008).

The data reported in Toropova *et al.* (2008) imply an active role for the RNA genome in assembly of the MS2 capsid. The structure of the MS2 virion highlights some interesting puzzles concerning virus capsid assembly. How does the genome operate during capsid assembly? Does the RNA act throughout assembly or just at the beginning to promote formation of a seed complex? Is it a pre-folded template upon which capsid assembly takes place or is it a dynamic structural scaffold? How does the genome acquire the structure observed within the virion? Is this structure related to the solution structure of the genome? MS2 provides an ideal system to investigate the roles of genomic RNA during icosahedral capsid assembly as it is well characterised biochemically and structurally.

## 1.10 The aims of this project

The ability RNA to induce capsid formation and selective packaging of viral RNA *in vivo* suggests that icosahedral ssRNA virus assembly is a co-condensation process where coat protein assembles selectively around its cognate RNA. The roles of genomic RNA in this process are however poorly understood. The main focus of this project was to gain an understanding of the roles of RNA during ssRNA virus capsid assembly. The project formed part of a dedicated multidisciplinary group effort into understanding ssRNA virus assembly and structure.

In chapter 3 the production of MS2 genomic RNA corresponding to different segments of the MS2 genome is described. These RNAs allowed the importance of RNA size, sequence and structure for capsid assembly to be addressed. The chapter reports the production of three sub-genomic MS2 RNAs using standard molecular biology techniques. Additionally, their ability to promote capsid assembly as assayed by gel mobility shift assays is reported.

In chapter 4, the dependence of capsid assembly efficiency on RNA is further investigated using sedimentation velocity analysis. These assays allowed evaluation of the hydrodynamic properties of MS2 capsid assembly reaction components and were instrumental in gaining an understanding of how genomic RNA is packaged into

capsids. Capsid assembly induced with oligonucleotide variants of TR was also assayed in order to investigate how different stem loops induce capsid assembly.

In chapter 5, the solution structure of three sub-genomic RNA fragments is compared to the structure of the MS2 RNA genome *in virion* as observed by their susceptibility to lead ion hydrolysis. These experiments are discussed in light of an RNA structural transition during genome packaging as well as a proof of principle experiment that allows mapping CP<sub>2</sub> binding sites on the RNA genome.

The focus of this project was originally quite different, where the aim had been to generate recombinant MS2 capsids for use as platforms for random peptide display. These libraries would allow identification of capsids with cell specific adhesion properties using biopanning techniques which would allow production of cell specific RNA cargo delivery vehicles. An underlying requirement of this system was to gain further understanding of the MS2 RNA packaging process and RNA size restrictions. Although work was continuous throughout the project on producing a phage display system, progress was limited. Exciting results obtained with respect to how RNA affects capsid assembly however channelled work in that direction, the results of which are reported here.

## 2 Materials and Methods

---

### 2.1 Materials

#### 2.1.1 General chemicals

Inorganic chemicals were supplied by Fischer Scientific Inc., Sigma Aldrich Co. and Ambion.

#### 2.1.2 Molecular biology products and reagents

1 kb DNA ladder and dNTPs were supplied by Invitrogen. 2-Log DNA Ladder (0.1-10 kb) was supplied by New England Biolabs Inc. rNTPs were supplied by Roche Applied Sciences. RNA sample loading buffer was supplied by Sigma-Aldrich Inc. RNA size reference marker 0.28-6.6 kb was supplied by Sigma-Aldrich Inc. RNaseOUT ribonuclease inhibitor was supplied by Invitrogen. RNasequre was supplied by Ambion.

#### 2.1.3 Enzymes

All restriction enzymes were supplied by Fermentas Life Sciences. Taq DNA polymerase, Thermoscript reverse transcriptase, T4 DNA ligase and T7 RNA polymerase were supplied by Invitrogen. RQ1 RNase-free DNase and T4 polynucleotide Kinase were supplied by Promega Corporation. Transcriptor reverse transcriptase was supplied by Roche Applied Science. Reverse-iT One-Step RT-PCR Kit was supplied by ABgene. Immolase DNA polymerase was supplied by Bionline. Kod Hot Start polymerase was supplied by Novagen. PfuTurbo DNA polymerase was supplied by Stratagene. Phusion DNA polymerase was supplied by New England Biolabs Inc. Megascript *in vitro* transcription system was supplied by Ambion. All ribonucleases were supplied by Ambion.

#### 2.1.4 Bacterial strains

The cell strains XL1-Blue and XL10-gold were generally used for cloning and were supplied by Stratagene.

## 2.1.5 Media and solutions

All media was prepared with ddH<sub>2</sub>O and autoclaved at 121 °C for 40 minutes unless otherwise stated. The media supplements ampicillin or kanamycin were added to LB media at a final concentration of 50 µg/ml or 30 µg/ml respectively. For production of LB-agar plates for blue-white colour screening, filter sterilised X-gal and IPTG were added to the LB-agar medium at final concentrations of 80 µg/ml and 20 mM respectively.

**Table 2-1. Media.**

Media	Constituents l <sup>-1</sup>
Luria Broth (LB)	10 g tryptone 5 g yeast extract 5 g NaCl Adjusted to pH 7.0 with NaOH
LB agar	20 g bacto agar 1 L LB-medium
Soc Medium	20 g tryptone 5 g yeast extract 0.5 g NaCl Adjusted to pH 7.5 with NaOH 12.5 ml 1 M MgCl <sub>2</sub> (filter sterilised) 12.5 ml 1 M MgSO <sub>4</sub> (filter sterilised) 1 ml 1 M glucose (filter sterilised)

**Table 2-2. Solutions.**

Solutions	Constituents
Ω buffer	50 mM HEPES / NaOH / pH 7.2 100mM NaCl 10 mM DTT 5 mM EDTA
λ buffer	10 mM HEPES / NaOH /pH 7.2 100 mM NaCl 1 mM EDTA

Buffer A	10 mM Tris / HCl / pH 7.4 100 mM NaCl 1 mM MgSO <sub>4</sub> 0.1 mM EDTA 0.002 % NaN <sub>3</sub>
10 X MOPS buffer	400 mM MOPS / NaOH / pH 7.0 100 mM NaOAc 10 mM EDTA
10 X TBE	108 g Tris 55 g boric acid 0.5 M EDTA pH 8.0 with NaOH
Prt Loading dye	2 g SDS 20 ml glycerol 5 mg bromophenol blue 92 ml 100mM Tris ph 6.8 8 ml β-mercaptoethanol
DNA loading buffer	70 % Sucrose 0.005% bromophenol blue
5 x Assembly Buffer	200 mM CH <sub>3</sub> COONH <sub>4</sub> /pH 7.2 5 mM Mg(OAc) <sub>2</sub> DEPC treated H <sub>2</sub> O

### 2.1.6 Vectors

The pGEM-T Easy Vector System was supplied by Promega Corporation. The CloneSmart Blunt Cloning Kit was supplied by Lucigen Corporation. GeneJet Blunt Cloning Kit was supplied by Fermentas corporation.

### 2.1.7 Oligonucleotide primers

All oligonucleotide primers were supplied by MWG-Biotech. For reverse transcription of lead ion cleavage products, primers were either labelled with ATP  $\gamma$ -<sup>32</sup>P using T4 polynucleotide kinase (Promega) according to the manufactures instructions or purchased pre-modified with a 5' cy5-fluorophore. The sequences of the primers used in this study are given in **Table 2-3**.

**Table 2-3.** The sequence specific primers used in the production of MS2 cDNA clones and in reverse transcription of lead ion cleavage reaction products. The numbering of the primers refer to their positioning in the MS2 genome sequence (Genbank accession number: NC\_001417)

Primer	Sequence	Modification
1.F	GGGTGGGACCCCTTTCGG	-
1.F_T7	GATAATACGACTCACTATAGGGTGGGACCCCTTTCGG	-
19_F	GGTCCTGCTCAACTTCCTGTCG	-
992_F	GGGAAAAGGTGCCTTTCTCATT	-
992_F_T7	GATAATACGACTCACTATAGGGAAAAGGTGCCTTTCTCATT	-
1419_F_T7	GATAATACGACTCACTATAGGGGTCGCTGAATGGATCAGC	-
2166_F	GGCACAAGTTGCAGGATGC	-
2676_R	TTGTGGAAAATAGTTCCCATCG	-
3044_R	TTATAGAGGCGTGGATCTGACAT	-
3550_R	TAGTTACCAAATCGGGAGAATCC	-
3569_R	TGGGTGGTAACTAGCCAAGC	-
3550_R	TAGTTACCAAATCGGGAGAATCC	Cy5
3163_R	CCGTACGGAGTCTTGGTGTAT	Cy5
2676_R	TTGTGGAAAATAGTTCCCATCG	Cy5
2223_R	TAACGGTTGCTTGTTTCAGC	Cy5
1666_R	CCTTGCATTGCCTTAACA	Cy5
1115_R	CCCGTTATTACGTCAGTAACTGTTT	Cy5
566_R	GCCTCTGCTAAAGCAACACC	Cy5
1568_R	AAGCTCTACACCACCAACAGTCT	<sup>32</sup> P
1666_R	CCTTGCATTGCCTTAACA	<sup>32</sup> P
1812_R	CGCGAGGAAGATCAATACATA	<sup>32</sup> P
1951_R	TCATTACCAGAACCTAAGGTCGGA	<sup>32</sup> P
2061_R	GATCCCATGACAAGGATTTG	<sup>32</sup> P
2223_R	TAACGGTTGCTTGTTTCAGC	<sup>32</sup> P

## 2.2 Methods

### 2.2.1 Purification of MS2 recombinant capsids

Frozen *E.coli* cell paste containing overexpressed CP<sub>2</sub> as recombinant capsids was supplied by Dr. Gabriella Basnak. The paste was dissolved in  $\Omega$  buffer and sonicated with a pulse sequence of 20 on, 20 off for 30 cycles. The cell lysate was then centrifuged at 14750 RCF for 30 minutes at 4 °C in order to remove cell debris. The

supernatant was treated with DNase I for 30 minutes at room temperature followed by centrifugation at 14750 RCF for 30 minutes at 4°C in order to degrade and remove DNA. (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> was then added to the resulting supernatant to 40 % saturation in order to salt out the MS2 capsids. The resulting protein precipitate was isolated by centrifugation at 14750 RCF for 30 minutes at 4 °C and re-suspended in a minimal volume of buffer λ and then dialysed against the same buffer in order to remove (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. The dialysed MS2 protein sample was loaded onto 15 - 45 % (w/v) linear sucrose density gradient that was prepared using a Gradient Station apparatus from Biocomp Instruments. The gradient was centrifuged at 42000 RCF for 16 hours and the band corresponding to MS2 capsids was excised from the density gradient. Following dialysis against buffer λ the sample was then loaded onto a gel filtration column containing sepharose CL-4B and eluted off the column with buffer A at 1 ml/min. Purified MS2 capsids were precipitated in 80 % (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and stored at 4 °C.

### ***2.2.2 Dissociation of MS2 capsids to produce MS2 CP***

The purification of MS2 CP is a modified method originally devised by Osborne *et al.*, (1970) to purify bacteriophage R17 A-protein. 6-10 ml of the MS2 (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> capsid precipitate was spun down in a centrifuge at 13000 rpm for 20 minutes and the precipitate re-suspended in a minimal amount of 40 mM CH<sub>3</sub>COONH<sub>4</sub> and dialysed against 40 mM CH<sub>3</sub>COONH<sub>4</sub> at 4 °C to remove (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. MS2 CP was purified by dissociating recombinant phage capsids, suspended in 40 mM ammonium acetate (pH = 6.76), with two parts glacial acetic acid for 1 h at 4 °C. At the end of the time period the mixture was spun down and the supernatant filtered with 20 mM acetic acid on a NAP10 column (Sephadex G-25 DNA grade, Amersham Biosciences) which had been pre-washed with the eluting solvent. This yielded pure MS2 coat protein as judged by SDS-page and mass spectrometry. The concentration of the protein was calculated by measuring the absorbance at 280 nm and 260 nm and substituting into the formula  $[CP_2] = (A_{280} * 1.55 - A_{260} * 0.76) / l * \epsilon$ , where  $\epsilon = 33240$  and  $l =$  length of sample cell (Layne, 1957; Gill and von Hippel, 1989; Beckett and Uhlenbeck, 1988; Stockley *et al.*, 1995; Stockley *et al.*, 2007). The equation takes into account any nucleotide impurities which might not be removed during filtration of the CP following acid dissociation of the capsid (Osborn *et al.*, 1970). Recombinant MS2 capsids contain various sized RNAs most likely originating from the host from which they are purified or perhaps mRNA

encoding the recombinant protein as implied from both absorbance measurements of recombinant capsids and electrophoresis of RNA extracted from them.

### 2.2.3 RT-PCR

RT-PCR reactions were performed according to instructions with the reverse transcriptase enzyme used, using 1-2 µg of RNA template.

### 2.2.4 PCR

PCR reactions were mixed according to the instructions for the DNA polymerase used to amplify the DNA target unless otherwise stated. Reactions were then cycled on a programmable thermocycler (DNA Engine, BioRad) according to

1. 94 °C 2 min
2. 94 °C 40 sec
3. 58 °C 40 sec
4. 72 °C 2 min 50 sec (repeat 2-4 x29)
5. 72 °C 10 min
6. 4 °C ∞

with appropriate changes to denaturing, primer annealing and extension times and temperature when necessary.

### 2.2.5 *In vitro* transcription

Reactions were set up in 0.3 ml PCR tubes as follows:

- 5 µl 10x transcription buffer
- 2 µl 1 M DTT
- 2 µl RNasecure
- Nuclease free water to 40 µl including template DNA
- 300-500 ng template DNA

This mixture was incubated at 60 °C for 15 mins in order to activate the RNasecure prior to addition of

- 1.25 µl of each 100 mM rNTP
- 1 µl RNase OUT
- 1 µl Yeast Inorganic Pyrophosphatase
- 3 µl T7 RNA polymerase (150 U)

Reactions were incubated at 37 °C for 2-4 hours and then treated with DNase 1 according to:

- 6 µl 10 x DNase 1 buffer (supplied)
- 4 µl DNase 1
- Incubated at 37 °C for 20 min, then 65 °C for 10 min.

Reaction mixtures were loaded straight onto denaturing agarose gels and the RNA product excised and purified by electrophoresis and phenol extraction. *In vitro* transcription was also performed using a MEGAscript *in vitro* transcription kit (Ambion) according to the supplier's instructions, incubating at 37 °C for 2h and then precipitating the RNA immediately, following DNase I treatment, with ammonium acetate or lithium chloride. This procedure afforded RNA more homogenous in length and in much higher yield which consequently did not need to be gel purified prior to use in capsid assembly reactions.

### 2.2.6 Agarose gel electrophoreses

Native agarose gels were prepared by dissolving 1.0 g of agarose (Helena Biosciences) in 100 ml of TBE buffer by boiling the solution in a microwave for 1 min or as necessary. Upon cooling, EtBr was added to the solution to a final concentration of 500 ng/μl before casting the gel. Native agarose gels were run in TBE at 70 V for up to 2 1/2 h.

Denaturing agarose gels were prepared by dissolving 1.0 g of agarose in 72 ml H<sub>2</sub>O by boiling the solution in a microwave for 1 min or as necessary. Upon cooling, 10 ml of 10x MOPS buffer and 18 ml of 37 % formaldehyde (12.3 M) were added to the solution before casting the gel. Denaturing agarose gels were run in 1 x MOPS buffer at 90 V for 40 min.

### 2.2.7 Isolation and purification of DNA and RNA

DNA and RNA was purified by cutting it from agarose gels followed by electroelution at 20-100 V for 10-30 min or, in the case of DNA, by running the dissolved gel slice through a spin column (Qiaquick Gel extraction kit). DNA isolated by electroelution was extracted with phenol chloroform and precipitated with two equivalents of EtOH at a final NaOAc concentration of 100 mM. The DNA/RNA precipitate was washed with 70 % EtOH in H<sub>2</sub>O and resuspended in H<sub>2</sub>O and frozen at -20 °C until needed.

The concentrations of RNA/DNA was calculated by measuring absorbance at 260 nM. An OD of 1 = 50 ng/μl for DNA. An OD of 1 = 40 ng/μl for RNA. The molecular weights of the RNA fragments were calculated using the ambion molecular weight

calculator: vRNA = 1,148,304 g/mol, 5'RNA = 794,255 g/mol, 3' RNA = 829,571 g/mol, iRNA = 298,545 g/mol.

### 2.2.8 Cloning and sequencing of DNA

Purified DNA fragments were cloned into either pGEMT or pSMART according to the manufacturer's protocol. Following transformation of a competent *E.coli* cloning strain, cells were grown on LB agar O.N. with the appropriate antibiotic. Colonies were selected and, if possible, 8-10 minipreps were performed using a Wizard miniprep kit. Sequencing of purified plasmid DNA was performed by the Sequencing Service, University of Dundee using plasmid specific primers flanking the plasmid multiple cloning site distributed by the relevant vector manufacturer.

### 2.2.9 MS2 assembly reactions

Assembly reactions (standard procedure) with the genomic RNAs were carried out in 40 mM ammonium acetate, 1 mM Mg(OAc)<sub>2</sub> at pH = 7.2 at room temperature. RNA at a final concentration of 40 nM was titrated with various amounts of CP<sub>2</sub> in 20 µl or 320 µl final reaction volumes for GEMSA assays or sedimentation velocity assays respectively. For reactions induced with oligonucleotide stem loops, reactions were performed at a final concentration of 2 µM of oligonucleotide and carried out at 4° C. Reactions components were added in this order: 5x assembly buffer, H<sub>2</sub>O, RNA, CP<sub>2</sub> and incubated for a minimum of 3 h prior to analysis.

#### 2.2.10 Native agarose gel mobility shift assays (GEMSA assay)

MS2 capsid reassembly reactions were monitored using native agarose gel mobility shift assays. 20 µl reactions were performed at ratios ranging from 5:1 to 150:1 of CP<sub>2</sub>:RNA in a buffer containing 40 mM ammonium acetate and 1 mM Mg(OAc)<sub>2</sub>. The final concentration of RNA in these reactions was 30-50 nM dependent on the RNA length. Reactions were allowed to proceed for 3 h and then loaded and run in 1 % agarose gels at 90 V for 60- 80 min in TBE buffer. The gels were then stained with 5 µg/µl EtBr for 20 min before imaging under short wave UV.

Densitometry of the GEMSA were performed with the Gene Tools software package from SYNGENE.

In the instances where assembly products were treated with RNase A prior to electrophoresis, reactions were treated with 0.1 U of RNase A and tenfold dilutions thereof according to the manufacturers instructions for 15 min at room temperature.

### 2.2.11 *Preparation of transmission electron microscopy grids*

Copper TEM grids were prepared by Adrian Hick. The samples were prepared by placing the grid, carbon coated surface down, on the capsid sample for 30 sec. The grid was then washed with ddH<sub>2</sub>O prior to staining with 2 % uranyl acetate for 30 sec. Finally the grid was washed again with ddH<sub>2</sub>O to remove excess uranyl acetate and placed on filter paper and allowed to dry before viewing by TEM. All electron micrographs were taken at a 40000x magnification.

### 2.2.12 *Sedimentation velocity assays*

Capsid reassembly reactions were performed as described in 2.2.9. The 320  $\mu$ l assembly reaction along with 340  $\mu$ l of buffer in the reference sector was loaded into a two-sector meniscus matching epon centrepiece cell. Sedimentation velocity analysis was carried out using an Optima XL-I ultracentrifuge using a An-50 Ti rotor. Sedimentation velocity profiles were collected at 17,000 rpm by recording absorbance at 260 nm. Scans were recorded approximately every 12 minutes. 50 scans were normally recorded. Due to the large difference in sedimentation rate of the substrates and products of MS2 capsid assembly reactions initiated with the RNA oligonucleotides, these reactions were spun at both 17,000 rpm and 40,000 rpm collecting 50 and 100 scans were respectively. The resulting sedimentation profiles were analysed using the C(S) method in SEDFIT (Schuck, 2000). The buffer viscosity and density were calculated with the program SEDNTERP (<http://www.rasmb.bbri.org/>) and were calculated to be 0.001008 Pa s and 0.99835 g/ml respectively. A partial specific volume of 0.53 ml/g (Enger, *et al.*, 1963; Fiers *et al.*, 1967) was used in fitting all acquired data except that of the CP<sub>2</sub> and capsids alone where the partial specific volume of 0.68 ml/g was used (Strauss and Sinsheimer *et al.*, 1963). Due to diffusional broadening of sedimenting RNA species at these slow rotor speeds, the 25-50 first scans were normally used in data fitting. Size distribution profiles were used when a c(s) model fit with an rmsd of 0.02 or lower could be obtained.

### 2.2.13 *Lead acetate structure probing*

All reactions were performed in 10  $\mu\text{l}$  final volume. The 5' and 3' sub-genomic fragments were at a 0.1  $\mu\text{M}$  final RNA concentration. The iRNA was kept at 0.27  $\mu\text{M}$  final concentration. RNA concentrations were measured by absorbance at 260 nm assuming that 1 AU is equivalent to 40  $\mu\text{g}/\mu\text{l}$ . WT MS2 phage was used at final concentration of 0.285  $\mu\text{g}/\mu\text{l}$  or 0.15  $\mu\text{l}$  of 19 mg/ml phage stock solution per reaction. These concentrations were increased tenfold when the reactions were assayed by capillary electrophoresis in order to maximise sequence read length and signal intensity.

Prior to cleavage with lead ions, the MS2 sub-genomic RNAs were heated to 65  $^{\circ}\text{C}$  for 10 minutes and then cooled to r.t. at  $\pm 2$   $^{\circ}\text{C}/\text{min}$  on a thermal cycler in 6  $\mu\text{l}$  final volume. At this point 2  $\mu\text{l}$  of 5x assembly buffer was added to the sample resulting in a final concentration of 40 mM  $\text{NH}_4^+\text{CH}_3\text{COO}^-$ , 1 mM  $\text{Mg}(\text{OAc})_2$ , pH = 7.2. 2  $\mu\text{l}$  of 2 mM lead acetate was then added to each reaction to afford a final concentration of 0.4 mM of lead ions. RNA was incubated with lead acetate at room temperature for 5, 10, 30 and 60 minutes followed by addition of 5  $\mu\text{l}$  0.1 M EDTA, 1.5  $\mu\text{l}$  3 M NaOAc and 35  $\mu\text{l}$  EtOH and stored at -20  $^{\circ}\text{C}$  for  $\geq 2\text{h}$ . RNA was recovered as a pellet by centrifugation at 13000 rpm and washed once with 70 % EtOH prior to resuspension in 5  $\mu\text{l}$  DEPC treated H<sub>2</sub>O. WT MS2 capsids were treated similarly but were phenol/ chloroform extracted prior to precipitation with 3M NaOAc. Prior to reverse transcription, the appropriate labelled primer was annealed to the cleaved RNA by addition of 1  $\mu\text{l}$  of 1  $\mu\text{M}$  labelled primer stock solution and incubated at 65  $^{\circ}\text{C}$  for 10 min followed by snap cooling on ice. Reverse transcription was then performed with Transcriptor® reverse transcriptase (Roche Diagnostics) at 52  $^{\circ}\text{C}$  for 30 min according to the manufacturers description. Reverse transcription products were then precipitated by addition of 3  $\mu\text{l}$  of 3 M NaOAc and 60  $\mu\text{l}$  EtOH. ssDNA was recovered by centrifugation and washed once with 70 % EtOH, allowed to dry at r.t. for 15 minutes, packaged and sent for fragment analysis. When the reactions were assayed with gel electrophoresis, reverse transcription products were dissolved in 95 % formamide RNA loading buffer (Ambion) and run on 8 % denaturing polyacrylamide slab gels as described in Sambrook *et al.*, (2006). The gels were then exposed to film at -80  $^{\circ}\text{C}$  overnight.

## 3 Production of sub-genomic MS2 RNA fragments and their effects on MS2 capsid assembly

---

### 3.1 Introduction

The assembly of icosahedral capsids enclosing ssRNA is promoted by RNA. In the case of MS2, RNA stem loops (Beckett and Uhlenbeck, 1988; Stockley *et al.*, 2007) along with the full length RNA genome (Hung *et al.*, 1969; Sugiyama *et al.*, 1967) can induce capsid assembly. Selective encapsidation of genomic MS2 RNA indicates that RNA packaging and capsid assembly are a synergistic event (Ling *et al.*, 1971). The function of genomic RNA in this process is not understood. Recently, cryo-EM reconstructions of MS2 and other ssRNA viruses have suggested a possible structural role for RNA genomes during capsid assembly (Schneemann, 2006).

A prerequisite for investigating the role of the MS2 genome in MS2 capsid assembly was that the reaction components, the CP<sub>2</sub> and RNA, were accessible to experiment. The MS2 coat protein is easily isolated from acid dissociated recombinant capsids, as described in 2.2.2 using a well established in house method. A route towards generating MS2 RNA transcripts that would allow assessment of the effects of RNA size, sequence and structure was however required. MS2 RNA sub-genomic fragments of different sizes and corresponding to different regions of the MS2 genome would allow assessment of the catalytic properties of MS2 RNA during MS2 capsid assembly. The initial objective of this project was therefore to produce sub-genomic MS2 RNAs and assess their ability to promote MS2 capsid assembly. It was not clear whether sub-genomic RNA would promote capsid assembly as the scaffolding role of the genome suggested in Toropova *et al.* (2008) might be eradicated upon genome truncation.

This chapter reports the production of RNA transcripts that correspond to different regions of the MS2 genome and initial experiments into how these transcripts affect MS2 capsid assembly. The results section is split in two. In 3.2.1 the production of MS2 RNA sub-genomic fragments using molecular biology methods is described. In 3.2.2, investigations into how these sub-genomic RNA fragments affect MS2 capsid assembly is reported based on results from gel mobility shift assays (GEMSA) of *in vitro* capsid assembly reactions.

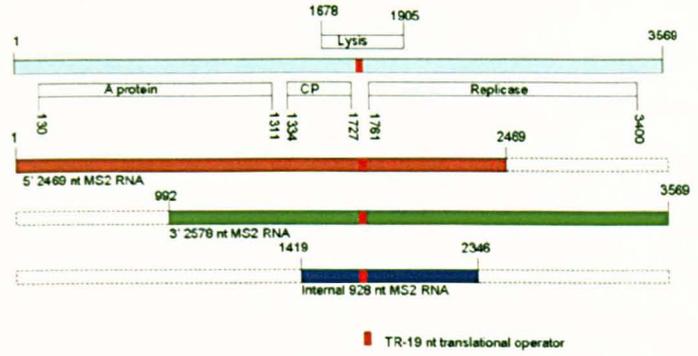
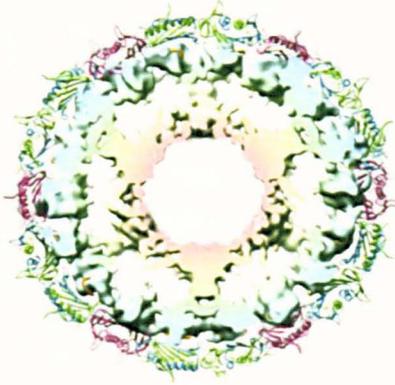
### 3.1.1 *The selection of MS2 sub-genomic RNA fragments to study the effect on capsid assembly*

The MS2 sub-genomic RNAs that were to be used in *in vitro* capsid assembly reactions, were chosen in light of the structural organisation of the MS2 genome within the MS2 virion.

The cryo-EM reconstruction reported by Toropova *et al.* (2008) showed that the MS2 genome is organised as two concentric shells of RNA connected at the capsid icosahedral 5-fold axes of symmetry. Roughly two thirds of the RNA genome comprise the outer shell of observed RNA density and is involved in contacts to the capsid interior while the remaining one third resides in the inner shell. The organisation of the MS2 genome based on the cryo-EM reconstruction was discussed in section 1.9.3.

While in practice, RNA transcripts representing any part of the MS2 genome could have been used, the cryo-EM reconstruction implies that the RNA within the capsid has a specific topology and/or particular fold. Further support for this idea has come from the way the A-protein is situated on the capsid 5-fold icosahedral symmetry axis (Toropova, 2009) and has been shown to bind to the MS2 genome, connecting to its 5' and 3' termini (Shiba and Suzuki, 1981). This idea, along with the estimates of the amount of RNA located in each of the RNA shells within the virion resulted in the strategy of producing MS2 sub-genomic fragments of sizes corresponding roughly to the size of RNA seen in each of the RNA shells (**Figure 3-1**).

It was decided to produce a minimum of three MS2 sub-genomic RNAs and investigate their effect on capsid assembly with respect to RNA size, structure and sequence. Two transcripts would be produced that represented two thirds of the MS2 genome from both its 5' and 3' termini along with an additional transcript that corresponded to roughly one third of the MS2 genome. In light of the genome organisation within the virion, it was anticipated that using these fragments to study capsid assembly would result in a high probability of observing different RNA induced effects on capsid assembly based on structure and sequence. Additionally these RNAs were to be used in parallel cryo-EM reconstructions of assembled capsids carried out by Dr. Katerina Toropova as part of a group effort to elucidate the role of MS2 genomic RNA in MS2 capsid assembly.



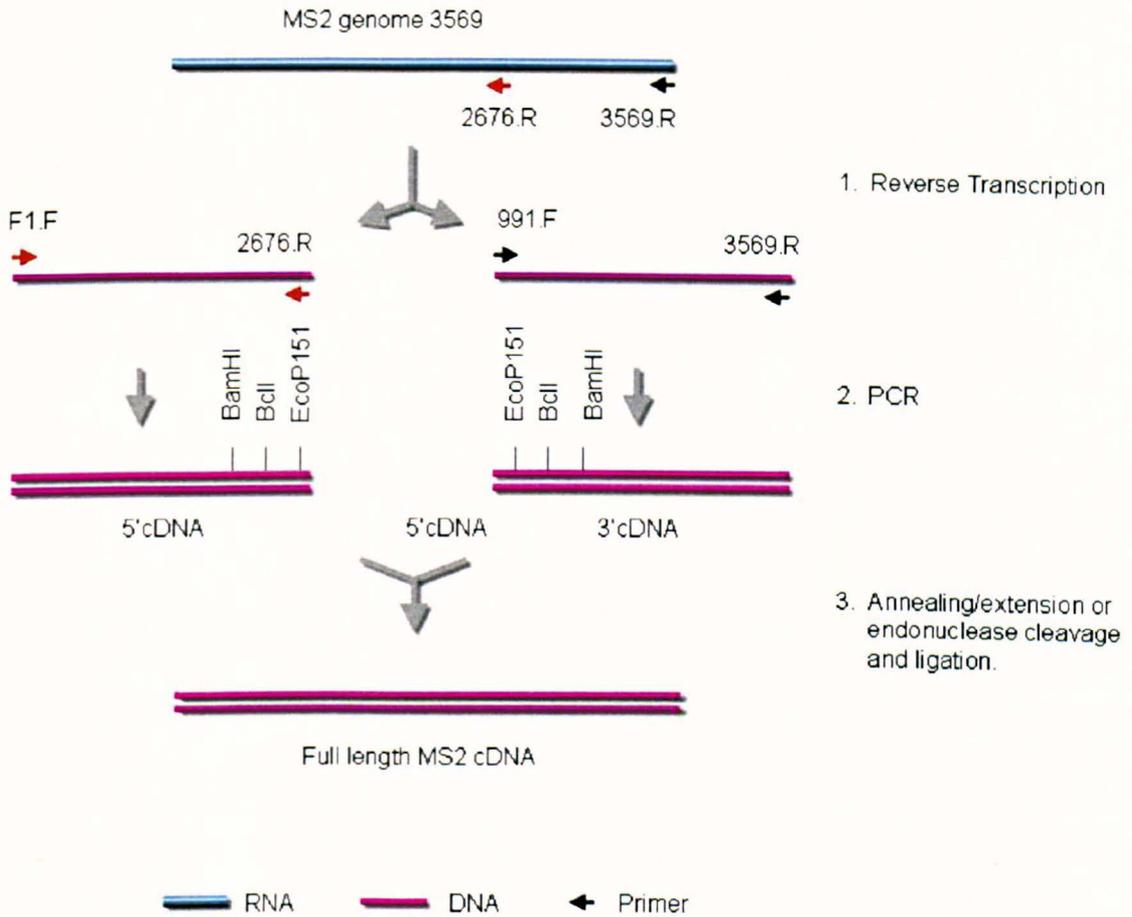
**Figure 3-1 Selection of MS2 sub-genomic RNAs to study the role of RNA in MS2 capsid assembly.** The sizes of the sub-genomic RNA fragments were to correspond roughly to the amount of RNA observed in each shell of RNA density observed within the MS2 virion in the cryo-EM reconstruction reported by Toropova et al. (2008) shown here in slab view down the icosahedral 3-fold axis. The RNA transcripts, whose production is described in 3.2, are shown aligned to the MS2 genome on the right.

### 3.1.2 Production strategy of the MS2 RNA sub-genomic fragments

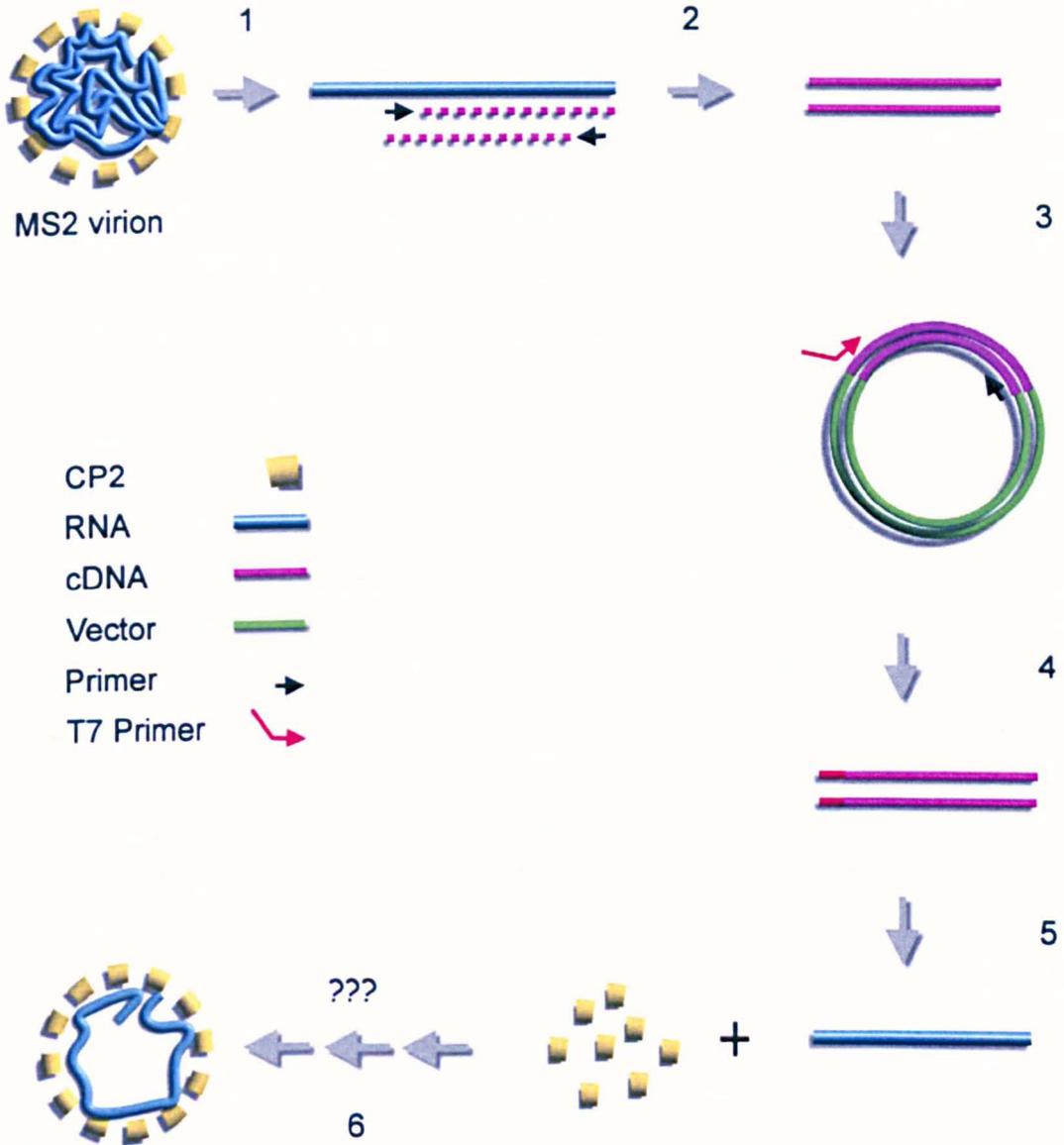
In order to produce the MS2 RNA sub-genomic fragments, it was decided to generate cDNA copies of the MS2 genome which could act as templates to generate RNA in *in vitro* RNA transcription reactions using T7 RNA polymerase. The three reactions underlying this route, namely reverse transcription, the polymerase chain reaction (RT-PCR) and *in vitro* transcription, represent central molecular biology techniques following their discovery by David Baltimore and Howard Temin in 1970 (for review see Verma, 1977), Kary Mullis (Saiki *et al.*, 1988) and Chamberlin *et al.* (1970), respectively. Using these methods the MS2 RNA sub-genomic fragments of choice could be produced. The access to a homogenous source of MS2 cDNA in good yield however, represented a key step for production of MS2 sub-genomic RNA.

Cloning of DNA into a replicating plasmid molecule that is maintained in bacteria represents the most straightforward procedure of generating large amounts of homogenous DNA. Previous investigations into MS2 genetics and protein translation had however indicated that maintaining MS2 cDNA within a replicating plasmid in *E.coli* was problematic as the MS2 cDNA was lethal to the cells resulting from continuous phage production even under controlled transcription conditions (Devos *et al.*, 1979; Hill, 1993; Shaklee, 1990). This type of “leaky” transcription has also been reported for bacteriophage Q $\beta$  (Taniguchi *et al.*, 1978)

To bypass this problem a purely *in vitro* approach towards generating the MS2 sub-genomic cDNA was devised and is shown in **Figure 3-2**. This strategy however proved non-practical as it resulted in various problems in downstream reactions when it came to the production of the sub-genomic RNAs. The strategy was therefore modified to include a cloning step which was successful and proved crucial in terms of obtaining good cDNA yields. **Figure 3-3** highlights the production process of the MS2 RNA sub-genomic fragments which is described in the following sections.



**Figure 3-2. The in vitro strategy for production of MS2 cDNA stock.** MS2 cDNA is lethal to *E.coli*. In order to bypass cloning of MS2 cDNA an in vitro approach towards generating MS2 cDNA was initially preferred. The figure shows the outline of the production of MS2 cDNA. 1: The MS2 RNA is reverse transcribed using two different primers (black and red arrows) to produce two different ssDNA strands representing MS2 RNA truncated at the 3' and 5' ends. 2: The single stranded DNA acts as a template in a PCR reaction generating a cDNA copy of the corresponding 3' and 5' truncated MS2 RNA. 3: The two cDNA fragments overlap by ~1700 nucleotides which allows production of full length MS2 cDNA through either ligation at a common restriction site or annealing the two fragments followed by DNA extension by a DNA polymerase. The full length cDNA could then act as a template for further PCR with T7 promoter containing primers to produce cDNA for use in in vitro transcription reactions with T7 RNA polymerase. The purely in vitro approach was abandoned because of low yields from RT-PCR reactions which resulted in various problems in downstream applications such as the annealing of the cDNAs and incorporation of the T7 promoter sequence by PCR. The 5' and 3' sub-genomic MS2 cDNA fragments were therefore cloned into plasmids which allowed production of MS2 RNA as outlined in Figure 3-3.



**Figure 3-3. The production strategy of MS2 sub-genomic RNA for use in MS2 capsid assembly reactions.** 1. Full length MS2 RNA is extracted from WT MS2 virions. 2. The full length RNA acts as a template in RT-PCR reactions to generate cDNA, the length of which is defined by the primer annealing positions. 3. The cDNA is cloned into a vector to provide a source of the MS2 cDNA fragment. 4. A T7 RNA polymerase promoter sequence is introduced into the cDNA sequence by amplification of the cDNA fragment with a T7 promoter containing primer. 5. The T7 containing cDNA is transcribed in vitro by T7 RNA polymerase to produce a MS2 sub-genomic RNA. 6. The MS2 sub-genomic RNA is used along with purified MS2 CP<sub>2</sub> to study the effect of MS2 genome sequence, length and orientation on capsid assembly.

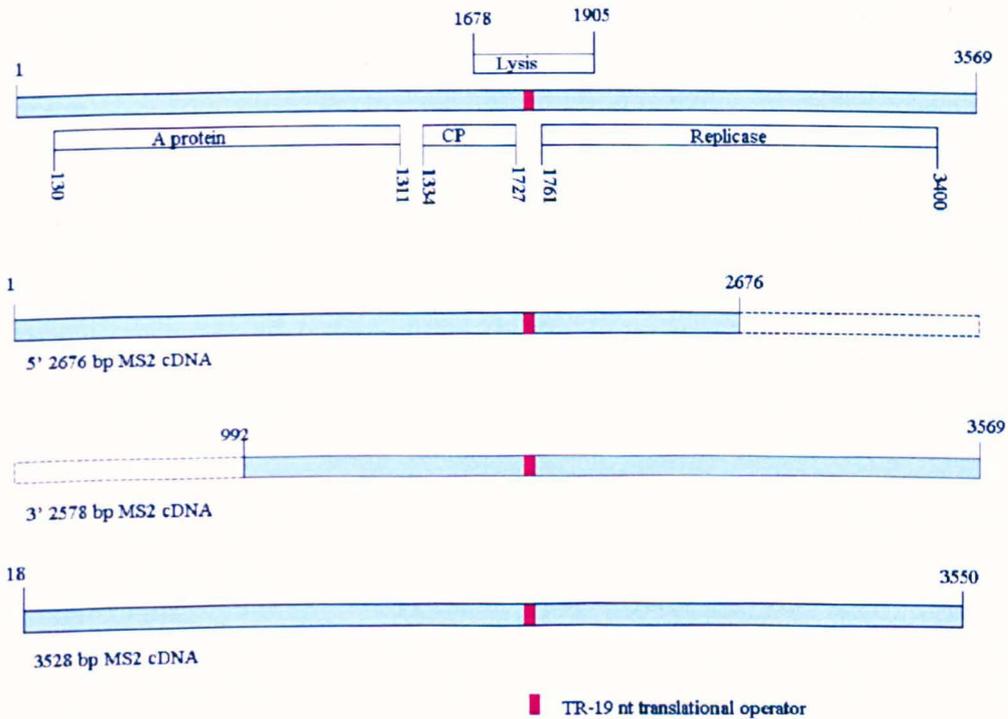
## 3.2 Results

### 3.2.1 Production of MS2 sub-genomic RNAs

#### 3.2.1.1 Production of MS2 sub-genomic cDNAs by RT-PCR

MS2 cDNA was produced by RT-PCR using MS2 virion RNA (vRNA) as template and sequence specific primers. The vRNA was purified by phenol extraction from MS2 virions which were a gift from Dr. David Peabody at the University of New Mexico.

Three MS2 cDNAs were produced which could be used to generate MS2 RNAs of sizes needed to study the effect of RNA on capsid assembly as described in 3.1.1.1. The alignment of these cDNAs to the MS2 genome is shown in **Figure 3-4**.



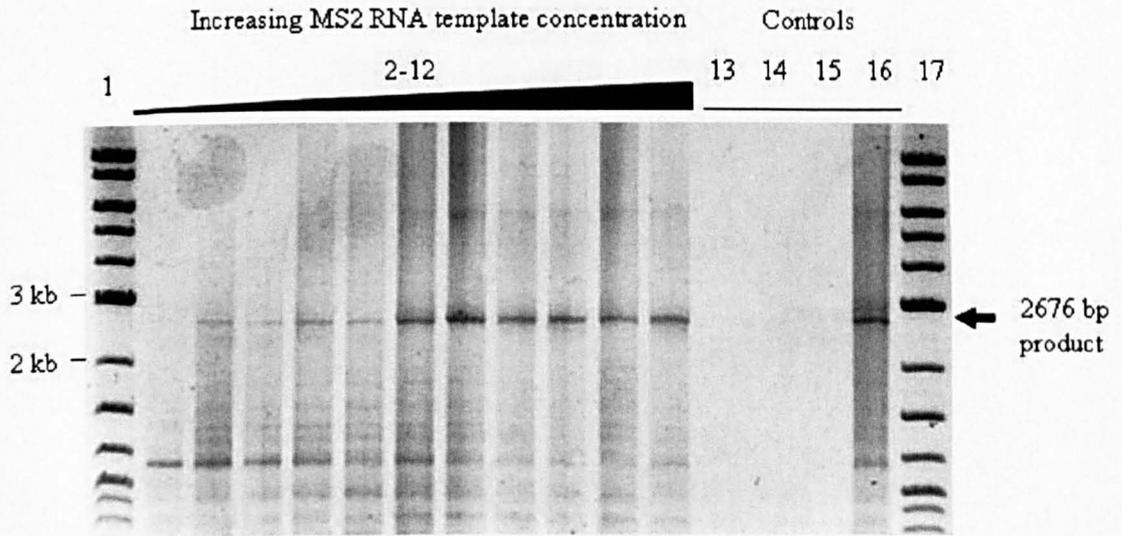
**Figure 3-4.** Alignment of the three sub-genomic MS2 cDNA fragments produced by RT-PCR to the MS2 genome. Two sub-genomic cDNAs corresponding to roughly two thirds of the MS2 genome from either the 5' or 3' terminus were produced along with a nearly-full length cDNA. The nearly-full length cDNA clone could be used to generate RNA corresponding to nearly any part of the MS2 genome. These fragments were cloned into vector constructs as described in section 3.1.2.2. to provide a stable source of homogenous cDNA that could be used to provide RNA transcripts for use in MS2 capsid assembly reactions.

Initial RT-PCR reactions were non-specific towards the desired cDNA product, produced the cDNA in low yield and required optimisation. RT-PCR reactions were optimised with respect to formation of cDNA corresponding to the 5' terminus of the MS2 genome using primers 1.F and 2676.R. This cDNA is referred to as the 5' sub-genomic MS2 cDNA and alignment to the MS2 genome is shown in **Figure 3-4**.

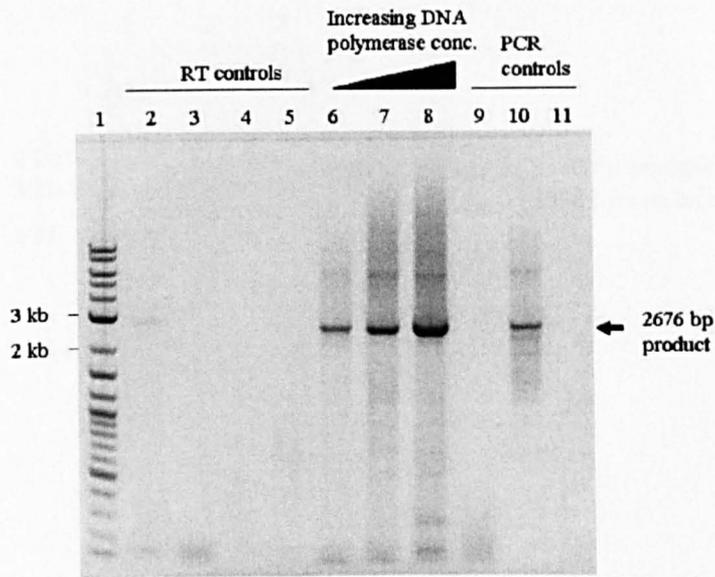
The 5' sub-genomic MS2 cDNA was successfully amplified by RT-PCR. The MS2 vRNA template was incubated in a buffered solution containing the primer 2676.R and Transcriptor reverse transcriptase and dNTPs for 40 min at 60 °C. The reverse transcription reaction was then used as a template to generate double stranded cDNA using PCR as described in section 2.2.4 with the primers 1.F and 2676.R. **Figure 3-5** shows the results of the reaction optimization with respect to MS2 RNA template concentration. A band corresponding to the 2676 bp product appeared when the RNA template amount used was 0.8 µg and became fairly constant when using in excess of 1.5 µg RNA template in the reverse transcription reaction. RT-PCR was also dependent on the amount of DNA polymerase used in the PCR step of the RT-PCR reactions. Increasing the DNA polymerase concentration four-fold resulted in a visible increase in the cDNA product of the expected size (**Figure 3-6**). Optimisation of MgCl<sub>2</sub> and dNTP concentrations did not affect product yield.

The 3' 2578 bp sub-genomic cDNA, produced using primers 992.F and 3569.R and the 3528 bp nearly-full length MS2 cDNA, produced using the primers 19.F and 3550.R were amplified in a similar fashion to the 5' sub-genomic cDNA albeit using the DNA polymerase Pfu in the PCR step. **Figure 3-7** shows an agarose gel of RT-PCR products corresponding to the 3' sub-genomic MS2 cDNA and the nearly-full length MS2 cDNA.

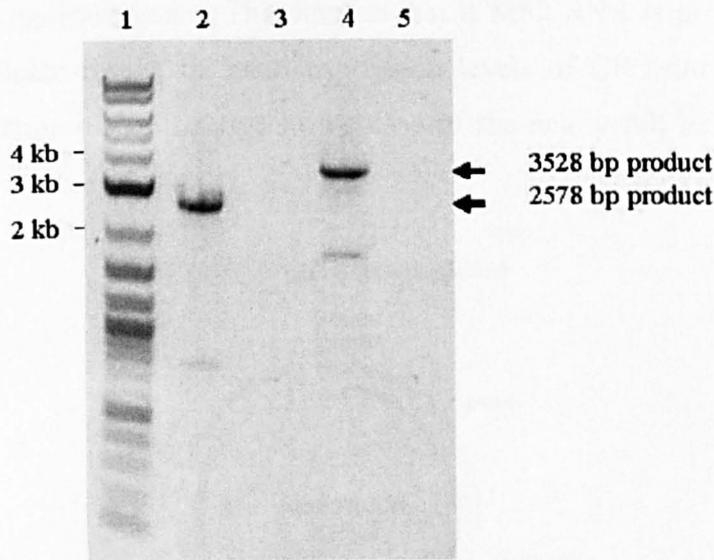
All the sub-genomic MS2 cDNAs were purified by electro-elution from agarose gels, precipitated and stored at -20 °C until needed in downstream reactions.



**Figure 3-5. RT-PCR optimisation with respect to MS2 RNA template concentration.** The figure shows a 1.2 % agarose gel of RT-PCR amplification products from reactions where an increasing amount of MS2 vRNA was used as template to generate the 5'2676 bp sub-genomic MS2 cDNA. The arrow marks the position of the expected 2676 bp product. **1:** 2- log DNA ladder. **2-12:** 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, and 8.0  $\mu\text{g}$  MS2 vRNA template. **13:** no MS2 vRNA control. **14:** no RT primer control. **15:** no reverse transcriptase control. **16:** no PCR primers control. **17:** 2-log DNA ladder. In all cases 10  $\mu\text{l}$  of reaction mixture were loaded in each lane from 50  $\mu\text{l}$  PCR reactions using 5 $\mu\text{l}$  of the first strand synthesis reaction as template. The expected yield of the 2676 nt cDNA product did not increase as the RNA substrate concentration was increased above 1.5  $\mu\text{g}$  suggesting that an optimum vRNA substrate concentration had been reached. The unexpected product observed in lane 16 was observed in repeats of this reaction control (see Figure 3-6) and most likely represent products produced by self priming of ssDNA, produced in the RT reaction, at it's 3' end.



**Figure 3-6. RT-PCR optimisation with respect to DNA polymerase concentration.** The figure shows a 1.2 % agarose gel of RT-PCR reaction products at increasing concentration of the DNA polymerase Immolase to produce the 5' 2676 bp sub-genomic MS2 cDNA. The arrow marks the position of the expected 2676 bp product. 1: 2-log DNA ladder. 2: RT reaction mix. 3: RT reaction mix, no RNA template control. 4: RT reaction mix, no 2676.R primer control. 5: RT reaction mix, no reverse transcriptase control. 6: 4 µl PCR, 1.0 Immolase unit. 7: 4 µl PCR, 2.5 Immolase units. 8: 4 µl PCR, 4 Immolase units. 9: 4 µl PCR, no RT, reaction template control. 10: 4 µl PCR, no primers control. 11: 4 µl PCR, no Immolase control. 10 µl of the RT reaction mix were loaded in lanes 2-5 corresponding to the 10 µl of RT-reaction mix used as a template in the subsequent PCR reactions. 4 µl of the PCR reaction mix were loaded in lanes 6-11. An obvious increase in the expected product yield followed increasing the DNA polymerase concentration. This reaction provided enough 5' 2676 bp sub-genomic MS2 cDNA for cloning experiments which would provide a stable source of this sub-genomic cDNA as described in the next section. Similar to what is seen in Figure 3-5, an unexpected product is observed in lane 10. Because no forward primer was available in this control reaction (only reverse primer from carry-over from RT reaction) it appears likely that this product originates from self priming of the ssDNA RT product, i.e. the 3' end of the ssDNA folds back upon itself and acts as a primer for DNA polymerisation.



**Figure 3-7. RT-PCR of the 3'2578 bp and 3528 bp sub-genomic MS2 cDNAs.** Expected products are indicated with black arrows. The figure shows a 1.0 % agarose gel of RT-PCR reaction products. 1: 0.5  $\mu$ g 1 kb DNA ladder. 2: 4  $\mu$ l of 2578 bp RT-PCR reaction mix. 3: 4  $\mu$ l of 2578 bp RT-PCR reaction mix-no template control. 4: 4  $\mu$ l of 3528 bp RT-PCR reaction mix. 5: 4  $\mu$ l of 3528 bp RT-PCR reaction mix-no template control. Both sub-genomic MS2 cDNAs were amplified in good yield and provided cDNA that was then cloned to provide a stable source of these sub-genomic cDNAs as described in section 3.2.1.2.

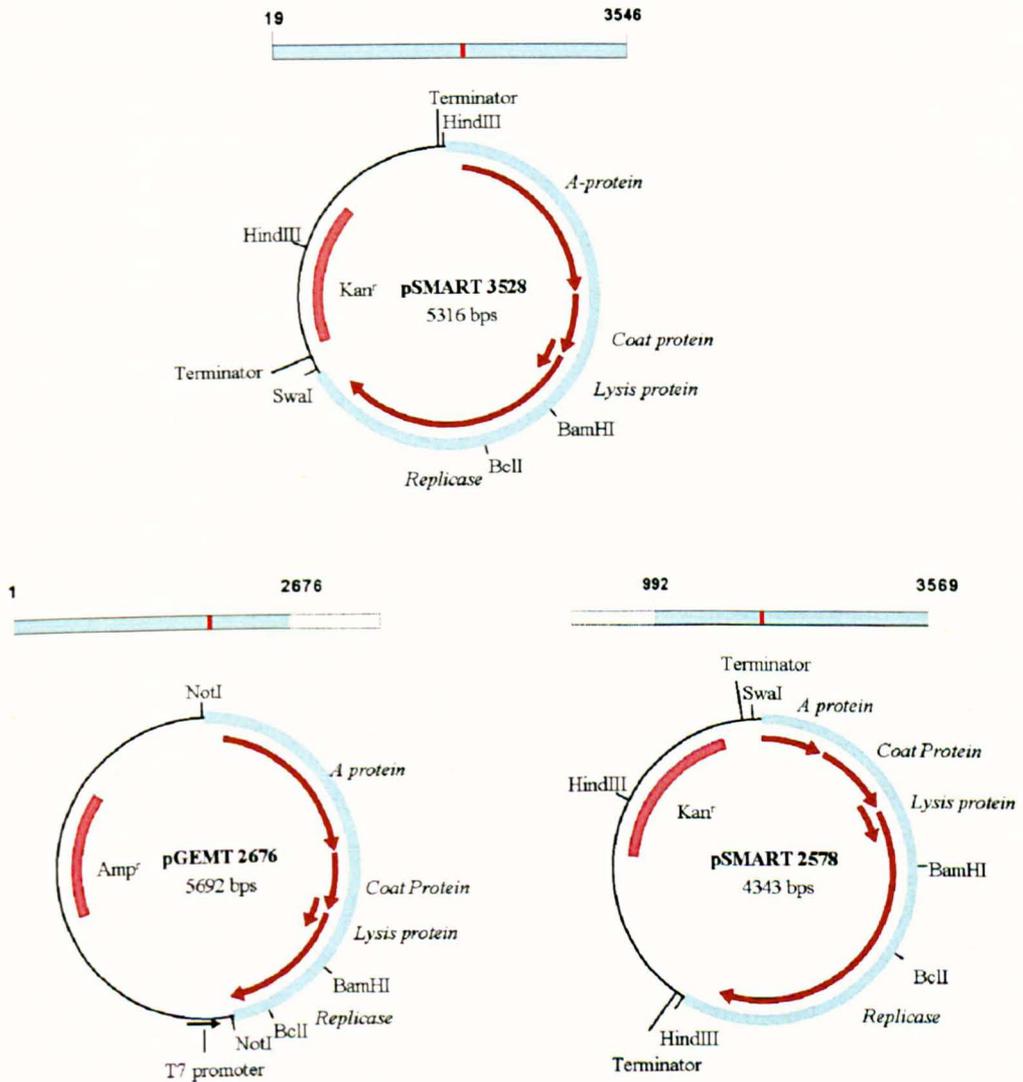
### 3.2.1.2 Cloning and sequencing of the sub-genomic MS2 cDNAs

The three sub-genomic MS2 cDNAs were cloned into vectors that would allow isolation of large quantities of the cDNA for use as template in PCR reactions to generate T7 promoter containing DNA as outlined in **Figure 3-3**. **Figure 3-8** shows the vector constructs containing the three sub-genomic MS2 cDNAs produced.

The 5' sub-genomic cDNA was successfully cloned into pGEM-T, a vector designed for direct cloning of PCR products. The 3' and 3528 bp MS2 sub-genomic cDNAs could however not be cloned into this vector despite attempts at optimising both vector:insert ligation conditions and *E.coli* cell strain and transformation conditions. The two fragments were ultimately cloned into a different vector, pSMART. pSMART is a cloning vector which, unlike pGEM-T, contains transcription terminators flanking the multiple cloning site thereby minimizing transcription of the cloned insert which could occur during transcription of other plasmid encoded genes.

The difficulties in cloning the 3' and nearly-full length MS2 cDNAs as compared to the 5' MS2 sub-genomic cDNA suggests that cloning of cDNA encoding the CP and replicase gene products is more toxic to *E.coli* cells than cloning of a cDNA lacking a

full copy of the replicase gene. This implies that if MS2 RNA is produced *in vivo*, a functioning replicase results in lethal expression levels of CP from numerous RNA copies or production of MS2 phage in the case of the nearly-full length MS2 cDNA clone.



**Figure 3-8. Vector constructs of the three cloned sub-genomic MS2 cDNA fragments.** Each cloned cDNA, produced by RT-PCR as described in section 3.2.1.1, is shown above the corresponding vector construct. The 2676 bp MS2 sub-genomic fragment was cloned into the vector pGEMT. The 3528 and 2578 bp MS2 sub-genomic fragments were cloned into the vector pSMART, a vector designed for the cloning of potentially lethal cDNA fragments. These vectors provided a stable source of cDNA that was used as template in PCR reactions to generate cDNA which in turn was used in *in vitro* transcription reactions as described in sections 3.2.1.3. All three vector inserts were sequenced. The DNA mutations identified in these vectors are given in **Table 3-1**.

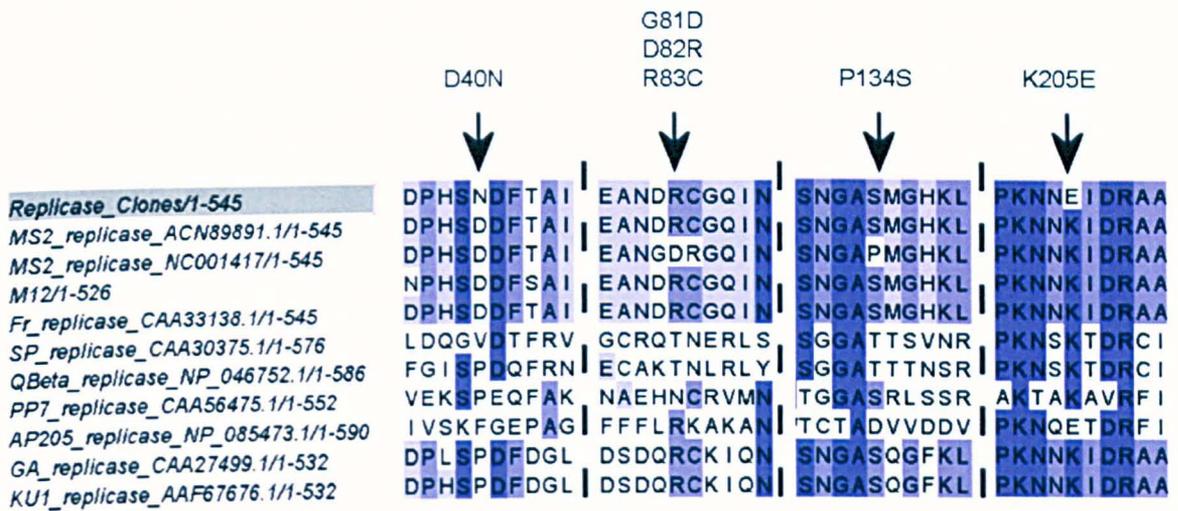
The clones carrying the MS2 sub-genomic cDNA inserts were sequenced and found to have a number of nucleotide mutations compared to the published sequence for MS2 phage (genbank accession ID: NC001417) (**Table 3-1**). There are however reasons to believe that some of the sequence variation might be due to the MS2 virion used in this study being of a different genotype than the MS2 virion originally sequenced or that the original sequencing was inaccurate. This is evident from comparison of NC001417 to the sequences of pLaACR62, an infectious MS2 cDNA clone created by Eric Remaut and co-workers (1982), and ACN89891 a more recent sequence for WT MS2 phage (Betancourt, 2009). A large proportion of the identified mutations in the MS2 sub-genomic cDNAs are also present in the infectious MS2 sequences pLaACR62 and ACN89891. Also, sequence alignment of a hypothetical replicase gene product carrying all the mutations identified in the replicase region of the MS2 sub-genomic cDNAs suggested that mutations occurring at aa residues GDR81-83DRC and P134S did not represent mutations but rather the WT MS2 phage sequence as these aa residues are highly conserved within the *Levivirus* genus (**Figure 3-9**).

Sequence variations that could not be accounted for in any WT MS2 sequence were however also identified. A total of six nucleotide mutations were found in the three MS2 sub-genomic cDNA clones that were not observed in the infectious MS2 sequences and were not shared between the MS2 sub-genomic cDNA clones (**Table 3-1**). Four of these result in missense aa mutations. D209G and G359R were identified in the A-protein genes of pGEMT-2676 and pSMART-2578 respectively. K205E and D40N were identified in the replicase gene of pSMART-2578 and pSMART-3528 respectively. Sequence alignments of the respective protein products from the three cDNA clones to the WT MS2 A-protein and replicase revealed that these mutated aa are highly conserved within the *Levivirus* genus.

The problems associated with the production of the three MS2 cDNA clones reflected previous results where isolation of full length MS2 cDNA was problematic as described in 3.1.2. It is possible that the mutations identified in the MS2 sub-genomic cDNA clones result in non functioning A-protein and replicase gene products and that these clones were isolated due to positive selection constraints imposed by the cloning procedure. It was decided to use these clones for the production of MS2 sub-genomic RNAs as the effects on RNA structure and function were anticipated to be minimal.

**Table 3-1. Nucleotide mutations identified from automated sequencing results in the MS2 sub-genomic cDNA clones pGEMT-2678, pSMART-2578 and pSMART-3528.** The mutation positions are relative to the MS2 cDNA GenBank reference sequence NC001417. Two additional WT MS2 cDNA clones, pLaCR62 and ACN89891.1, are also shown for comparison. These WT MS2 cDNA sequences are highlighted in yellow. Of the twenty mutations identified in all three sub-genomic MS2 cDNA clones, fourteen are shared in two out of the three WT MS2 cDNA reference sequences and between all MS2 sub-genomic clones. These positions are highlighted in grey. This suggests that nucleotide variation found in the MS2 sub-genomic clones is due to the MS2 virus used in this study being of a different genotype than the one originally sequenced. Mutations which could result in non functioning gene products were also identified. These are addressed in Figure 3-9.

DNA Mutation	Coding region	aa mutation	pGEMT-2676	pSMART-2578	pSMART-3528	MS2		
						pLaACR62	ACN89891.1	NC001417
G83A	No	No	No	-	Yes	-	No	No
A744G	A-protein	No	Yes	-	No	No	No	No
A755G	A-protein	D209G	Yes	-	No	No	No	No
G1204A	A-protein	G359R	No	Yes	No	No	No	No
G1878A	Replicase	D40N	No	No	Yes	No	No	No
G2002A	Replicase	G81D	Yes	Yes	Yes	Yes	Yes	No
G2004C	Replicase	D82R	Yes	Yes	Yes	Yes	Yes	No
A2005G	Replicase	D82R	Yes	Yes	Yes	Yes	Yes	No
T2006G	Replicase	R83C	Yes	Yes	Yes	Yes	Yes	No
C2007T	Replicase	R83C	Yes	Yes	Yes	Yes	Yes	No
T2159C	Replicase	No	Yes	Yes	Yes	Yes	Yes	No
C2160T	Replicase	P134S	Yes	Yes	Yes	Yes	Yes	No
A2361G	Replicase	K205E	No	Yes	No	No	No	No
T2426C	Replicase	No	Yes	Yes	Yes	Yes	Yes	No
C2429T	Replicase	No	Yes	Yes	Yes	Yes	Yes	No
A2591T	Replicase	No	-	Yes	Yes	Yes	Yes	No
C3451 insert	No	No	-	Yes	Yes	Yes	Yes	No
C3452 insert	No	No	-	Yes	Yes	Yes	Yes	No
C3462 del	No	No	-	Yes	Yes	Yes	Yes	No
C3463 del	No	No	-	Yes	Yes	Yes	Yes	No



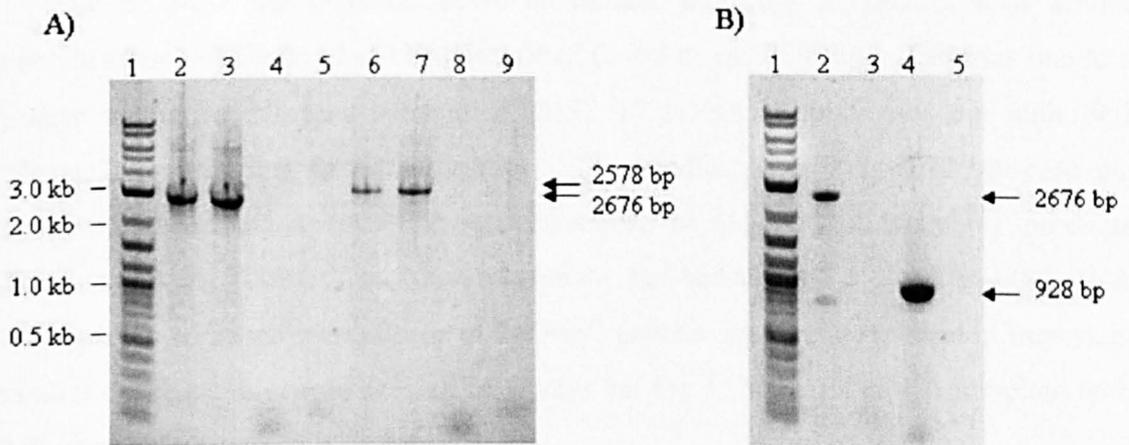
**Figure 3-9. Multiple sequence alignment of replicase gene products from selected viruses in the Leviviridae family and a hypothetical replicase gene product carrying all the mutations identified in the replicase coding region from Table 3-1.** A window of 10 aa residues is shown around each aa mutation. Alignment shows that the mutations in the replicase residues 81-83 and 134, which were present in all sequenced clones, are conserved to a high degree in the Levivirus Genus (all excluding Q $\beta$ , SP and AP205) suggesting that these residues represent an infectious MS2 sequence and that the original MS2 cDNA sequence of NC001417 is incorrect or originating from a different virus genotype. Both D40 and K205 are however highly conserved in the replicase's of the Levivirus Genus, indicating that these are most likely mutations of the MS2 sequence. A similar homology alignment for an A-protein product carrying the mutations D209G and G359R identified in pGEMT-2676 and pSMART-2578 showed that these aa positions are conserved within A-proteins of the Levivirus Genus.

### 3.2.1.3 Introduction of T7 promoter sequence

The next step in the production of MS2 sub-genomic RNA involved generating cDNA that could be used in *in vitro* transcription reactions using T7 RNA polymerase. The use of T7 RNA polymerase to generate RNA requires the cDNA template to be fused to a 21 bp T7 promoter sequence which is necessary for T7 RNA polymerase DNA recognition and initiation of transcription (for review of see Kochetkov *et al.*, 1998). The most straightforward way of producing cDNA fused to a T7 promoter sequence is by PCR using oligonucleotide primers to incorporate the T7 promoter sequence as outlined in **Figure 3-3**.

Three cDNAs containing the T7 promoter sequence were generated by PCR using the MS2 sub-genomic cDNA clones as template (**Figure 3-10**). These cDNAs corresponded to the RNAs required for capsid assembly reactions as described in 3.1.1. The T7 containing cDNA corresponding to the first 2676 nt of the MS2 genome from the 5' terminus was amplified from the MS2 sub-genomic clone pGEMT-2676.

Similarly, a 2578 bp cDNA representing the 3' terminus of the MS2 genome and an internal sub-genomic cDNA fragment of 928 bp were amplified from pSMART-2578 and pSMART-3528 respectively. Amplification of the T7 promoter containing cDNA was straightforward using Pfu polymerase as described in 2.2.8 and afforded cDNA that, following purification by phenol extraction and precipitation with sodium acetate, was used as template to generate MS2 sub-genomic RNAs.



**Figure 3-10. The production of sub-genomic MS2 cDNAs containing a T7 promoter sequence.** Three sub-genomic cDNAs were produced by PCR using T7 promoter containing primers using the MS2 sub-genomic cDNA clones as template. PCR reaction products were analysed on 1 % agarose gels. The expected 5'2676 bp, 3'2578bp and internal 928 bp MS2 sub-genomic cDNA products are indicated by arrows. **A)** 1: 0.5  $\mu$ g 1kb DNA ladder. 2-3: 2676 bp MS2 sub-genomic cDNA using primers 1.F\_T7 and 2676.R. 4: no primer control 5: no polymerase control. 6-7: 2578 bp MS2 sub-genomic cDNA using primers 992.F\_T7 and 3569.R. 8: no primer control. 9: no polymerase control. **B)** 1: 0.5  $\mu$ g 1 kb DNA ladder. 2: 2676 bp MS2 sub-genomic cDNA using primers 1.F\_T7 and 2676.R. 3: empty lane. 4: 928 bp MS2 sub-genomic cDNA using primers 1431.F\_T7 and 2346.R. 5: empty lane. These cDNAs were used to produce RNA as described in 3.2.1.4.

#### 3.2.1.4 Production of MS2 sub-genomic RNAs by *in vitro* transcription

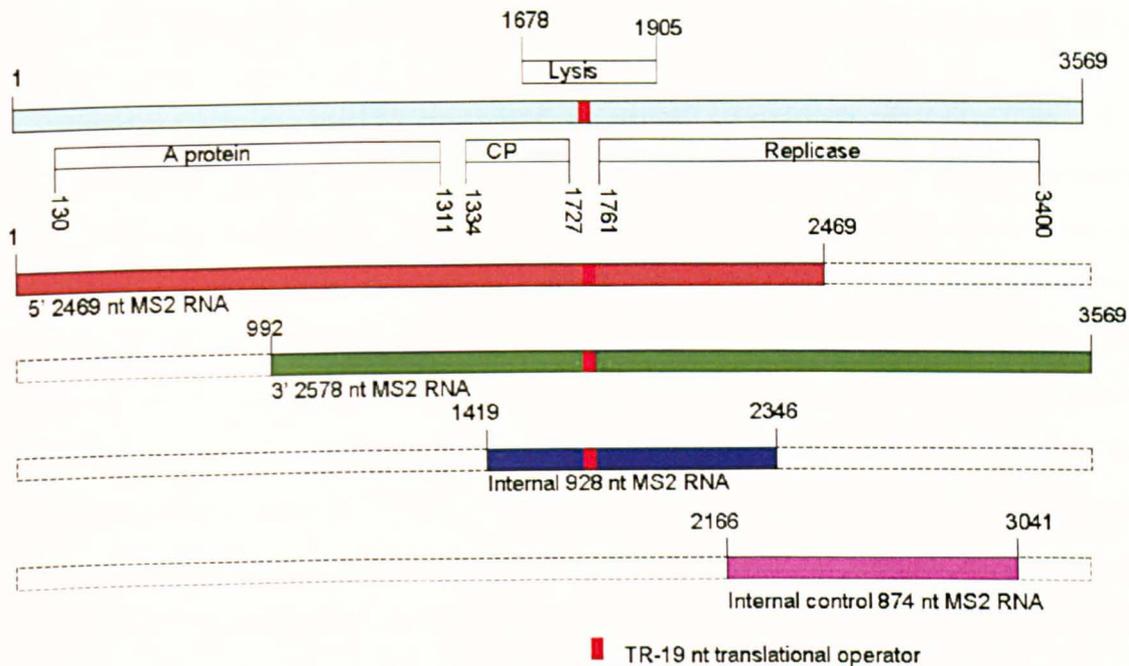
MS2 sub-genomic RNAs corresponding to two thirds of the MS2 genome from both the 5' and 3' genomic termini and an internal RNA (iRNA) corresponding to roughly one third of the MS2 genome were generated by *in vitro* transcription as outlined in **Figure 3-3** and described in 2.2.5. The RNA transcripts are shown aligned to the MS2 genome in **Figure 3-11**.

The 5' MS2 sub-genomic RNA is smaller than the 5' MS2 cDNA whose amplification and cloning has been described (2469 nt vs. 2676 bp). This was due to a mishap in the lab where a batch of 5' MS2 T7 cDNA template was cut with BclI (**Figure 3-2**) prior to *in vitro* transcription<sup>1</sup>. The resulting 5' 2469 nt MS2 sub-genomic RNA was then used in cryo-EM reconstructions of MS2 capsid assembly products (Toropova *et al.*, 2009). This was unfortunate, but because the 5' 2469 nt MS2 RNA still represented roughly two thirds of the MS2 genome and had been used in important parallel experiments it was decided to always cut the 5' MS2 T7 cDNA template with BclI prior to *in vitro* transcription.

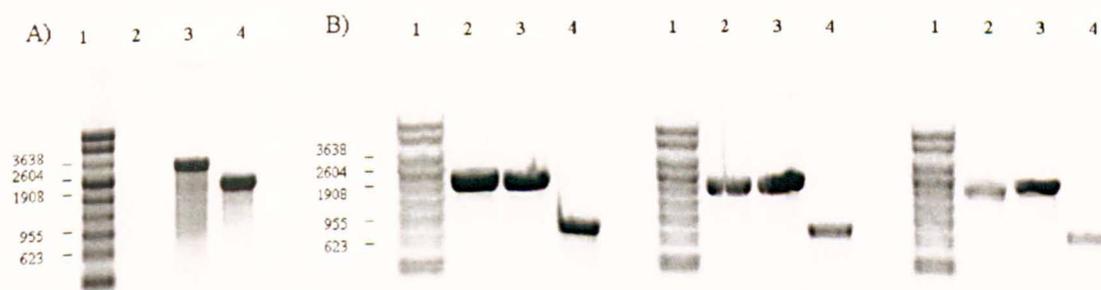
Characterisation of the MS2 sub-genomic RNAs is shown in **Figure 3-12**. Production of the RNAs in sufficient yield and purity for further reactions represented continuous work throughout this project as large quantities of RNA were required for analytical ultracentrifugation experiments (chapter 4). The best results were obtained using MEGAscript, an *in vitro* transcription kit supplied by Ambion for the production of large RNA transcripts. The yield and purity of sub-genomic RNAs generated with this kit were far superior to results obtained using the method outlined in 2.2.10 which is based on Sambrook *et al.* (2006), which showed little fidelity towards the correct size RNA product despite attempts at optimisation with respect to incubation time, template quantity and buffer composition. Using the Megascript kit, the sub-genomic MS2 RNAs were produced in sufficient yield and purity given that transcription was not allowed to proceed for more than two hours and that every precaution was taken to avoid RNase contamination. The sub-genomic RNAs were immediately precipitated following transcription with lithium chloride or ammonium acetate, kept in -80 °C and used in capsid assembly reactions within four to six weeks.

---

<sup>1</sup> The 5' RNA fragment was being used in parallel ligation reactions at the BclI restriction site (**Figure 3-2**) in order to generate full length MS2 cDNA.



**Figure 3-11.** *The alignment of the MS2 sub-genomic RNA transcripts to the MS2 genome.* The MS2 genome is shown in cyan with the positioning of the four gene products shown for reference (van Duin and Tsareva, 2006). The 5' sub-genomic RNA is shown in red. The 3' sub-genomic RNA is shown in green. The iRNA is shown in blue. An internal control RNA (icRNA) is shown in magenta which was produced similarly to the iRNA and is discussed in 3.2.2.4. The TR-operator is highlighted in bright-red. These four MS2 sub-genomic RNAs along with MS2 virion RNA were the RNA substrates that were used to study the effect of MS2 RNA on MS2 capsid assembly. The colour coding of the RNAs is constant throughout this thesis.



**Figure 3-12.** *The purity of the MS2 RNAs was assessed on denaturing 1 % agarose gels.* A) Denaturing agarose gel of MS2 vRNA extracted from MS2 virions. The 5' sub-genomic RNA is shown for comparison. 1: Sigma RNA standard. 2: empty lane. 3: MS2 vRNA. 4: 5' sub-genomic RNA. B) Denaturing agarose gel of the MS2 sub-genomic RNAs, transcribed in vitro from corresponding MS2 cDNA and purified by LiCl precipitation. Two repeats of the leftmost gel are shown. The sub-genomic RNA concentration is halved in identical lanes going from left to right. 1: Sigma RNA standard. 2: 5' 2469 nt sub-genomic RNA. 3: 3' 2578 nt sub-genomic RNA. 4: 928 nt iRNA. MS2 RNAs were used without gel purification in capsid re-assembly reactions.

### 3.2.2 The effect of MS2 sub-genomic RNA on MS2 capsid assembly

Investigations into ssRNA bacteriophage capsid assembly have shown that RNA stem loops (Beckett and Uhlenbeck, 1988) along with full length RNA genomes (Hung *et al.*, 1969; Sugiyama *et al.*, 1967) are capable of promoting the formation of capsid like structures *in vitro* as detailed in section 1.9. Although the packaging of sub-genomic RNA has also been reported (Beckett *et al.*, 1988) this RNA is considerably smaller (1 kb) than the MS2 sub-genomic RNAs (**Figure 3-11**) with an additional report suggesting a 1 kb size restraint on non-genomic RNA that could be packaged into the MS2 capsid (Pasloske *et al.*, 1998). Whether or not the MS2 sub-genomic RNAs would induce capsid assembly represented a principal question that this project sought to answer which would further allow addressing the roles of RNA in ssRNA virus capsid assembly.

The following sections describe the results from investigations into whether the sub-genomic RNAs could be used to study MS2 capsid assembly by characterisation of *in vitro* capsid assembly products by gel mobility shift assays (GEMSA) and transmission electron microscopy (TEM).

#### 3.2.2.1 MS2 sub-genomic RNA promotes capsid like product formation

MS2 capsid assembly reactions initialised with each of the three MS2 sub-genomic RNAs (**Figure 3-11**) were monitored at increasing CP<sub>2</sub> concentrations using native gel electrophoresis and TEM as described in sections 2.2.10 and 2.2.11 respectively. Capsid assembly with the vRNA was also monitored for comparison. The results are shown in **Figures 3-13** through **3-16** and are discussed below.

Interestingly, all the RNAs were observed to form capsid like products in the presence of purified CP<sub>2</sub> in reassembly buffer after 3 h at room temperature. This was evident from the GEMSA by the appearance of a band migrating at the same position as recombinant MS2 capsids which became visible at a CP<sub>2</sub>:RNA stoichiometry of 5-50:1 depending on the RNA. The TEM micrographs further confirmed the presence of MS2 capsid like products at CP<sub>2</sub>:RNA concentration ratios of 30:1. No capsid products were observed by TEM in the absence of RNA consistent with reports that CP<sub>2</sub> do not associate to form capsids at low ionic strength (Mattheaus and Cole, 1972). TEM micrographs at lower reaction ratios were not obtained.

These results correspond with a previous report by Sugiyama and coworkers (1967) where capsid like product formation with MS2 vRNA was observed at 23:1 of CP<sub>2</sub>:vRNA. The observed packaging of the MS2 sub-genomic RNAs however, represent the first results where truncated MS2 genomic segments of sizes approaching the full length viral genome are packaged into capsid like structures. They suggest that there is not a size restraint on the genomic RNA that can be encapsulated *in vitro* consistent with reports for the *Levivirus* bacteriophage fr (Hohn, 1969). The genomic RNAs were however observed to promote capsid assembly with different efficiency as addressed in 3.2.2.3.

The capsid assembly reactions did not exclusively form capsids. An assembly product migrating slower than both the capsid and the RNA was identified from the GEMSAs of capsid assembly at high CP<sub>2</sub>:RNA stoichiometries with all the genomic RNAs tested. These products appeared at CP<sub>2</sub>:RNA stoichiometric ratios of 50:1 and higher for the sub-genomic RNAs and at 90:1 for the vRNA. When an assembly reaction was performed at a 300:1 stoichiometry of CP<sub>2</sub>:5'RNA no material migrated into the gel but stayed in the well. When purified recombinant capsids were run on similar agarose gels no bands were seen that corresponded to these slower migrating species until at very high capsid concentrations (data not shown). These species were thought to correspond to aggregated capsids and/or incomplete capsids based on their slower migration in the gel. The TEM images confirmed this idea as an increased amount of incomplete capsid-like material was visible at high CP<sub>2</sub>:RNA molar ratios. The aggregated capsid like particles resembled in some ways the monster particles identified by Sorger and co-workers (1986) during *in vitro* capsid assembly of turnip crinkle virus at high coat protein concentrations.

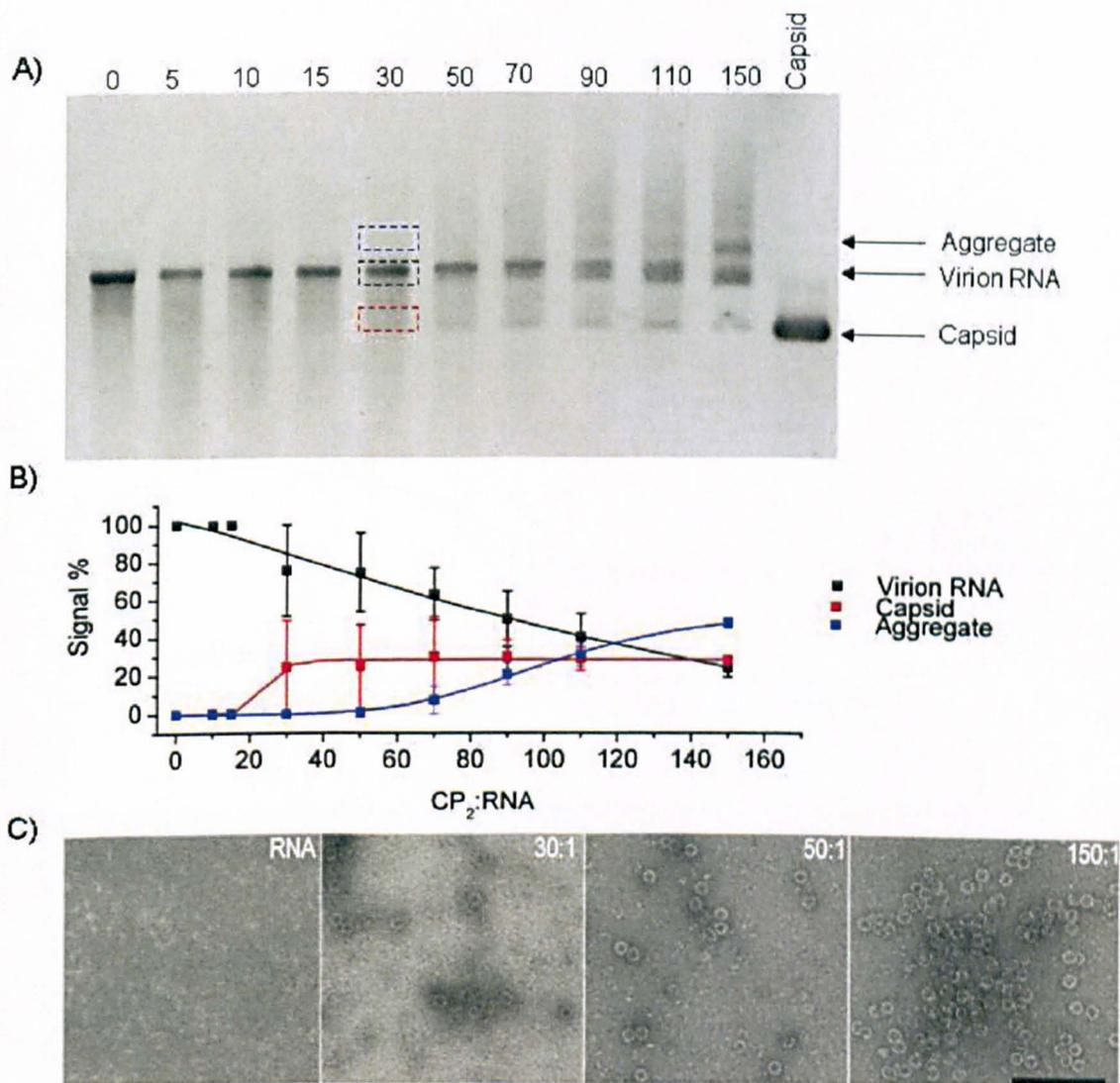
GEMSA of assembly reactions with the 5' and iRNA sub-genomic RNAs were performed under various conditions in order to maximise capsid formation and observe whether aggregation or the formation of incomplete capsids could be minimised. No difference in capsid formation was observed when capsid assembly reactions were allowed to proceed for 24 h as compared to 3 h. Lowering the temperature to 4 °C and incubating for 3 h or 24 h also resulted in no change. This suggested that the formation of the incomplete/aggregated capsid like products was non-reversible and that they most likely represented kinetically trapped material on the capsid assembly pathway. The formation of mis-assembled capsid products has been shown to be highly dependent upon coat protein subunit concentration and temperature (Matthews and Cole, 1972;

Nguyen *et al.*, 2007). As the aggregates were only seen at high CP<sub>2</sub>:RNA molar ratios it is tempting to conclude that their formation results from non-specific CP<sub>2</sub> binding which then hinders proper capsid assembly and is prone to aggregation.

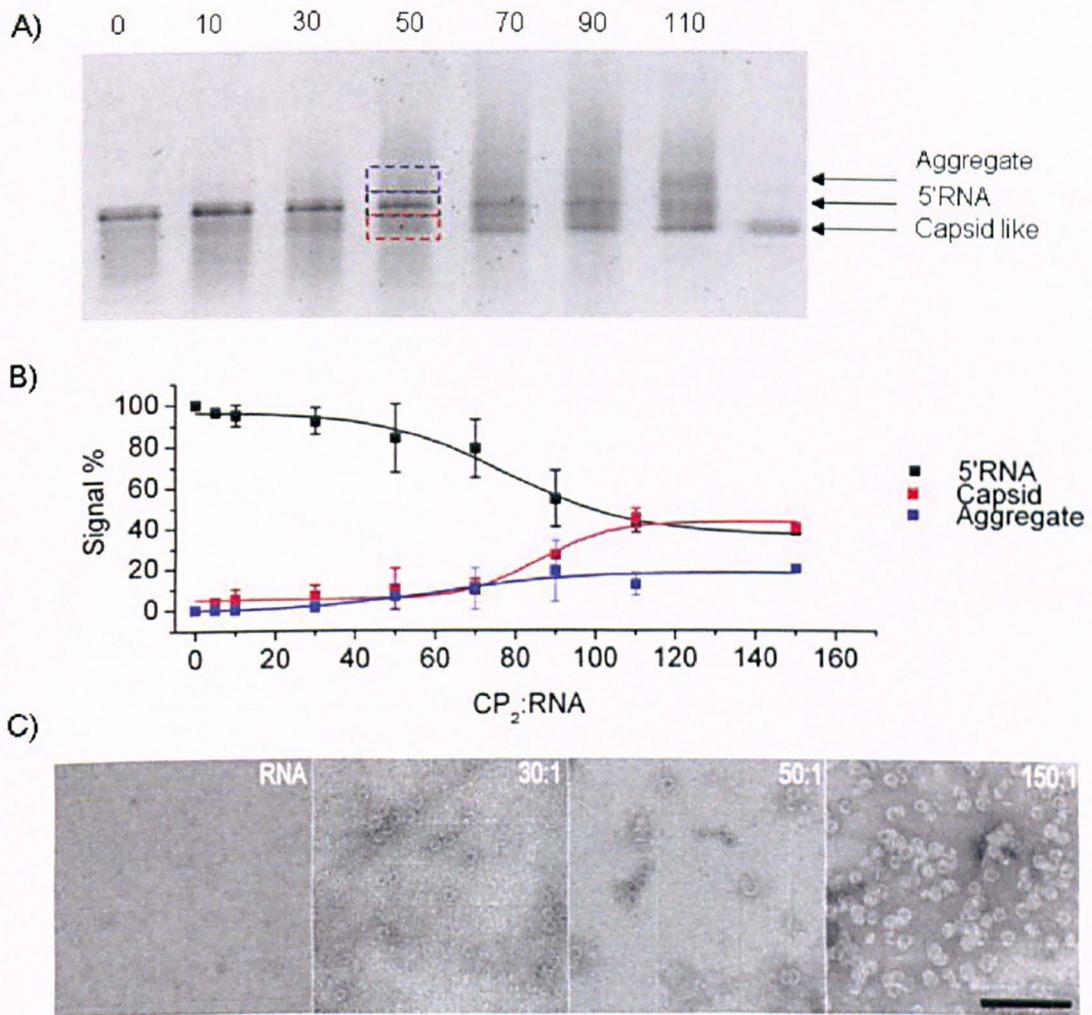
The ssRNA virus genome of bacteriophage R17, a close relative of MS2, has been observed to be accompanied with spermidine (Fukuma and Cohen, 1975), a positively charged polyamine that is believed to neutralise charge interactions of the RNA genome to allow more efficient packaging into to the volume of the viral capsid (Ames and Dubin, 1960). MS2 capsid assembly induced with the MS2 vRNA in the presence of 1 mM spermidine however did not have a marked effect on capsid formation. This was perhaps not surprising as Mg<sup>2+</sup>, present in the reassembly buffer, has been shown to compact MS2 vRNA in a similar manner to spermidine although higher concentrations of Mg<sup>2+</sup> are required (Leipold, 1977).

Further attempts at optimising capsid assembly reactions included addition of DTT to the assembly buffer and changing to a Tris/OAc buffer system at pH = 7.4. Again, these attempts did not result in hindering aggregate formation as viewed by agarose electrophoresis. It was decided not to experiment with the ionic strength of the assembly buffer as high salt concentrations have been shown to induce capsid assembly in the absence of RNA (Mathews and Cole, 1971) which would then compete with capsid formation induced by the protein-RNA interaction.

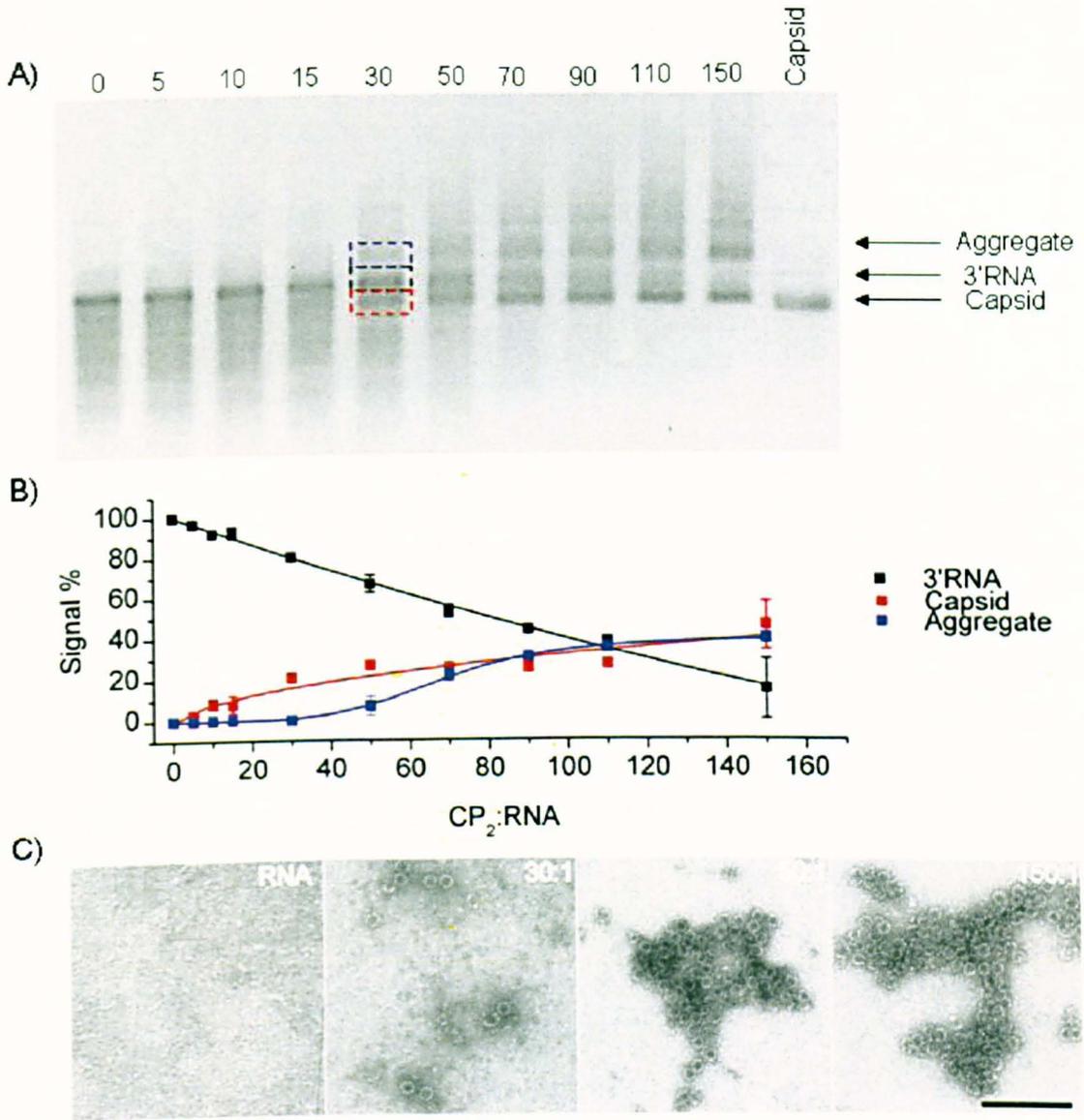
The GEMSA and TEM assays confirmed that the sub-genomic RNAs formed capsids *in vitro* and therefore could be used to study the role of MS2 RNA on capsid assembly. Further details on how the RNAs affected capsid assembly are addressed in 3.2.2.2-4.



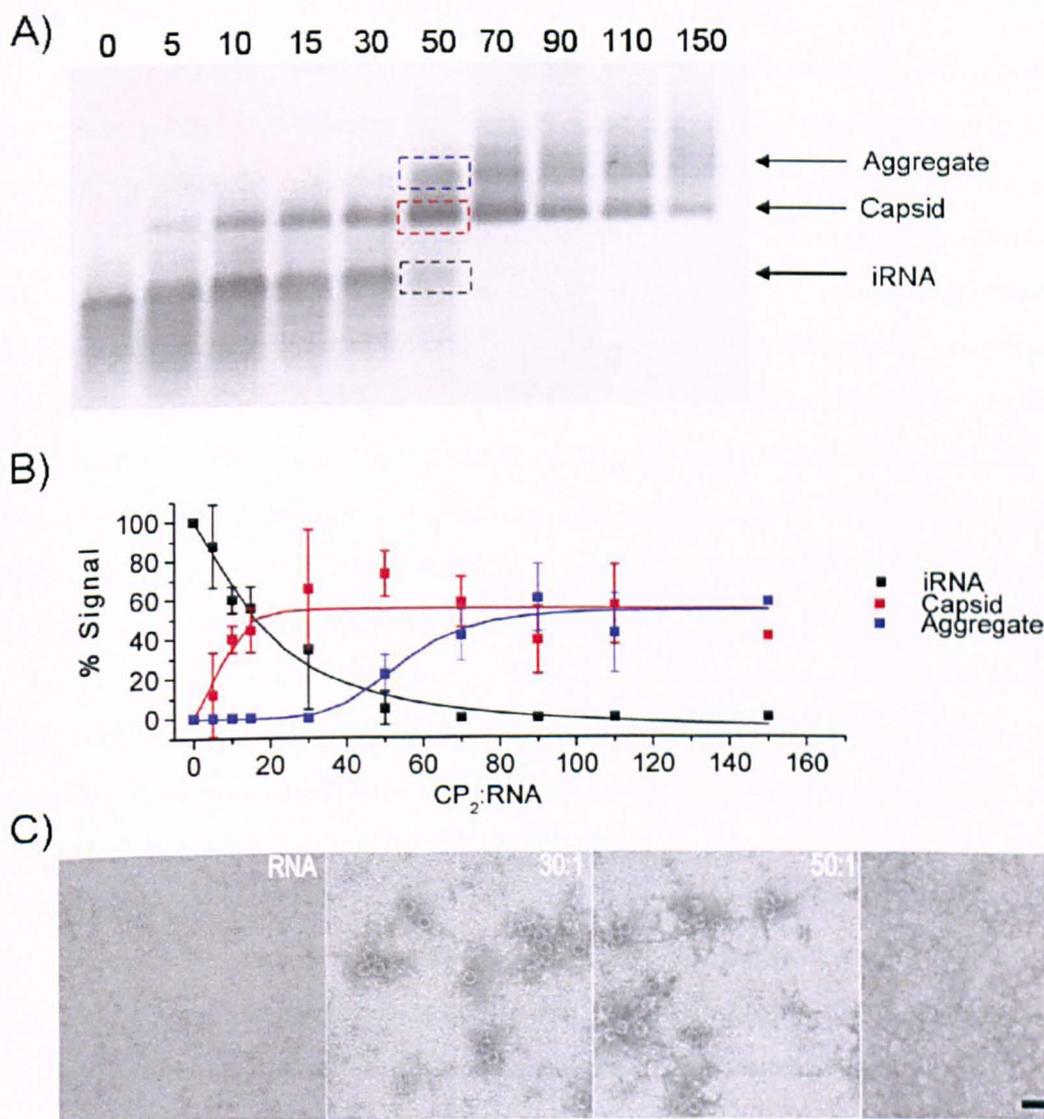
**Figure 3-13. In vitro capsid assembly induced by MS2 virion RNA.** Capsid assembly reactions were performed at increasing ratios of CP<sub>2</sub> to vRNA in a 20  $\mu$ l final volume of buffer consisting of 40 mM CH<sub>3</sub>COONH<sub>4</sub> and 1 mM Mg(OAc)<sub>2</sub>. The vRNA concentration was kept constant at 30 nM. After 3 h incubation at rt the reactions were analysed by GEMSA and TEM. **A** shows a 1 % native agarose gel of capsid assembly reactions. The number above each lane indicate the CP<sub>2</sub>:vRNA stoichiometric ratio of the assembly reaction in that lane. The migration position of vRNA, recombinant MS2 capsids and what is believed to be aggregated material is marked with an arrow. The three boxes indicate the area in each lane used to produce the densitometry plot shown in **B** and are colour coded according to **B**. The plot shows the percentage of vRNA, capsid and aggregate products at each CP<sub>2</sub>:RNA reaction ratio and is an average of GEMSAs as shown in **A**. The error bars represent standard deviation from average. **C** shows TEM micrographs of capsid assembly reactions at selected CP<sub>2</sub>:RNA stoichiometries which are indicated in the upper right corner of each micrograph. The black size bar corresponds to 200 nm. The results show that vRNA is packaged into virus like particles, consistent with previous reports, and that packaging is inefficient with only around a third of the RNA packaged. Furthermore the results imply that capsid assembly is cooperative with capsids observed at CP<sub>2</sub>:vRNA stoichiometries as low as 30:1 and that increasing the CP<sub>2</sub> concentration leads to aggregate formation.



**Figure 3-14 In vitro capsid assembly induced by MS2 5' sub-genomic RNA.** The packaging of the MS2 5' sub-genomic RNA was investigated in an identical manner as described for the vRNA in Figure 3-13. Capsid assembly reactions at increasing concentration ratios of CP<sub>2</sub>:5'RNA were analysed with GEMSA (A), densitometry of GEMSA (B) and TEM (C). The results were similar to the vRNA in that capsid like products were observed at low CP<sub>2</sub>:RNA ratios and increasing CP<sub>2</sub> concentration did not lead to the formation of more capsids but rather increased amounts of aggregated or incomplete capsids. In contrast to assembly induced with the vRNA (Figure 3-13), more capsids were seen to form with the 5' sub-genomic RNA at similar CP<sub>2</sub>:RNA ratios suggesting that smaller MS2 genomic RNA is packaged more efficiently than the full length RNA genome.



**Figure 3-15. *In vitro* capsid assembly induced by the 3' sub-genomic RNA.** The packaging of the MS2 3' sub-genomic RNA was investigated in an identical manner as described for the vRNA in Figure 3-13. Capsid assembly reactions at increasing concentration ratios of CP<sub>2</sub>:3'RNA were analysed with GEMSA (A), densitometry of GEMSA (B) and TEM (C). The results were similar to the vRNA in that capsids were observed at low CP<sub>2</sub>:RNA ratios and increasing CP<sub>2</sub> concentration did not lead to the formation of more capsids but rather increased amounts of aggregated capsid like material. In contrast to assembly induced with the vRNA (Figure 3-13), more capsids were seen to form with the 3' sub-genomic RNA at similar CP<sub>2</sub>:RNA stoichiometries. This was also observed for the 5' sub-genomic RNA (Figure 3-14) suggesting a size constraint on the RNA that is encapsulated. Comparison of these results with those of the 5' sub-genomic RNA did not imply an RNA structure or sequence constraint on RNA packaging.



**Figure 3-16. In vitro capsid assembly induced by the sub-genomic iRNA.** The packaging of the MS2 sub-genomic iRNA was investigated in an identical manner as described for the vRNA in Figure 3-13. Capsid assembly reactions at increasing stoichiometries of CP<sub>2</sub>:iRNA were analysed with GEMSA (A), densitometry of GEMSA (B) and TEM (C). Similar to the other genomic RNAs, the iRNA was observed to promote capsid assembly at low CP<sub>2</sub>:RNA ratios consistent with a co-operative assembly mechanism. At higher CP<sub>2</sub>:iRNA stoichiometries aggregate products were detected. Capsid assembly induced with the iRNA resulted in more capsid product formation at lower CP<sub>2</sub>:RNA stoichiometric ratios than had been observed for the 5' and 3' sub-genomic RNAs and the vRNA. This suggests that smaller RNAs are packaged more efficiently into MS2 capsids than larger RNAs.

### 3.2.2.2 MS2 capsid assembly is cooperative

Apart from the capsid aggregate already mentioned, only signals corresponding to the RNA and fully formed capsid product were detected in capsid assembly reactions assayed by GEMSA. This implies that CP<sub>2</sub> do not bind equally to all the genomic RNAs in solution as this would result in depletion of free CP<sub>2</sub> and inhibit formation of the final capsid product. Rather, the results suggest that CP<sub>2</sub> bind to genomic RNA which then promotes further growth of the capsid, perhaps through CP<sub>2</sub> dimer addition (Sorger *et al.*, 1986; Stockley *et al.*, 2007). When the capsid assembly reactions are assayed after 3 h at equilibrium, close to all CP<sub>2</sub> has gone into forming capsids, leaving only free RNA and the capsid in solution. This suggests that a CP<sub>2</sub> comprising a capsid is energetically favourable to a free CP<sub>2</sub> or a lone CP<sub>2</sub> bound to RNA due to stabilising interactions between the neighbouring CP<sub>2</sub> subunits and the genomic RNA within the capsid. The results observed for the vRNA and the sub-genomic MS2 RNAs correlate well with the findings of Sugiyama and coworkers (1967). Using sucrose gradient centrifugation, they characterised two components from *in vitro* MS2 capsid assembly reactions induced by vRNA which corresponded to RNA bound to 1-3 CP<sub>2</sub> and the capsid product.

The apparent lack of intermediates as well as the formation of capsid products at low CP<sub>2</sub>:RNA stoichiometries suggests that the packaging of the sub-genomic RNAs and vRNA into capsids is a co-operative process. This corresponds to accepted models of capsid formation where assembly occurs co-operatively and irreversibly following formation of a capsid intermediate seed complex (Sugiyama *et al.*, 1967; Hohn, 1969; Zlotnick, 1994; Zhang *et al.*, 2006; Stockley *et al.*, 2007).

### 3.2.2.3 Smaller RNA is packaged more efficiently than larger RNA

The GEMSA of capsid assembly suggest that the smaller sub-genomic RNAs promote capsid assembly more efficiently than the longer RNAs. This was evident from two observations. Firstly, capsids were observed at lower CP<sub>2</sub>:RNA stoichiometries for the smaller RNAs. Capsids were regularly detected with the iRNA at CP<sub>2</sub>:RNA reaction stoichiometries of 5:1 while they could not be observed with confidence at ratios lower than 30:1 for the 5' and 3' sub-genomic RNAs. Although capsids could also be detected in assembly reactions with the vRNA at 30:1, the yield of capsids compared to the sub-genomic fragments was notably lower. Secondly, at 110:1 CP<sub>2</sub>:RNA, capsid yield increased following a series of iRNA>5'RNA>3'RNA>vRNA.

Although the errors on the densitometry plots were quite high due to smearing of the products, the capsids represented roughly 55 % for the iRNA, 42 % for the 5' RNA, 35 % for the 3' RNA and 28 % for the vRNA of the total material in the capsid assembly reactions. Above 110:1 CP<sub>2</sub>:RNA, capsid yield was constant or decreased due to aggregate formation.

These results imply that there is a size, and therefore a structural constraint on RNA during packaging although there does not seem to be a constraint on the overall size of the genomic RNA as all the RNAs were indeed packaged.

#### 3.2.2.4 The role of RNA structure and sequence on capsid assembly

The size restraint during packaging of the MS2 sub-genomic RNAs implies that the different efficiencies observed are due to how easily the RNA structure can be accommodated within the viral capsid.

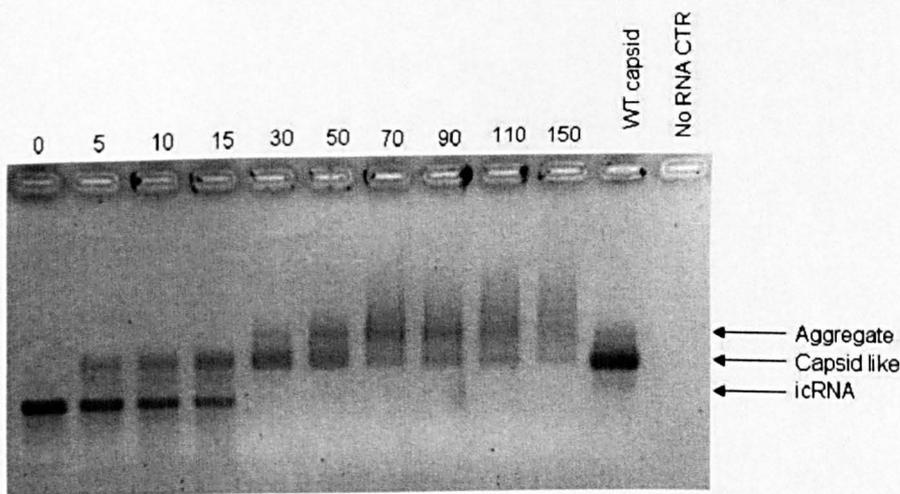
The iRNA was observed to be the most efficient at promoting capsid assembly. This is most likely due to a more compact structure of the iRNA compared to the 5', 3' and vRNA, which therefore allows more efficient packaging. Indeed the iRNA migrates faster than the MS2 capsid in the agarose gels as opposed to the 5', 3' and vRNA which migrate slower. Although migration in agarose gels depends upon both size and charge of the component, the migration rates suggest that the iRNA can be packaged directly from a pre-formed structure, while the 5' and 3' RNAs must be structurally reorganised or folded in order to fit inside the capsid. As the 5' and 3' sub-genomic RNAs promoted capsid formation better than the vRNA it follows that it is even harder to package the full length genome into the capsid shell. In chapter 4, the stokes radii of the MS2 sub-genomic RNAs was estimated and found to complement this idea.

The 5' and 3' sub-genomic RNA fragments encompass different regions of the MS2 genome and must have different structure, imposed by their different nucleotide sequence. Very few differences were observed between these two RNAs in their ability to promote capsid formation. If there are differences, it is likely they can be better observed by monitoring reaction kinetics rather than equilibrium states. The packaging of RNAs of different structures is sure to follow different reaction pathways although the end result, which is what was viewed by GEMSA, is the same.

The MS2 genomic RNAs used to study capsid assembly up to this point all contained the translational operator (TR) which has been shown to promote capsid

assembly *in vitro* and *in vivo* as discussed in section 1.9. Following reports suggesting that various genomic secondary structural elements other than TR could induce capsid assembly (Beckett and Uhlenbeck, 1988; Koning *et al*, 2003) and supported by the GEMSA results which suggested that CP<sub>2</sub> could bind to the genomic RNA at multiple positions as evident from aggregate formation, it was decided to monitor capsid assembly with a genomic RNA lacking the TR operator.

A MS2 genomic fragment of length corresponding to the iRNA but lacking the TR operator was produced in a similar manner as has been previously described for the iRNA. This internal control RNA (icRNA) corresponded to nucleotides 2166-3041 of the MS2 genome. Its alignment to the MS2 genome is shown in **Figure 3-11**. Its ability to promote capsid formation was assessed using GEMSA with the results shown in **Figure 3-17**. The icRNA promoted capsid assembly indicating that the TR operator is not necessary for successful *in vitro* capsid assembly. This supports ideas that CP<sub>2</sub> are able to bind a plethora RNA stem loop structures and that these sequences are repetitive in the MS2 RNA genome. The molecular properties that set TR aside from similar stem loop structures within the RNA genome, with respect to promoting capsid assembly, are explored further in chapter 4.



**Figure 3-17. *In vitro* capsid assembly induced with a 874 nt MS2 sub-genomic RNA lacking the TR operator.** A 874 nt sub-genomic RNA lacking the TR operator (icRNA) was produced and its ability toward promoting capsid formation analysed with GEMSA in a similar manner as described in Figure 3-13. The icRNA was able to promote capsid formation confirming reports that secondary structural elements within the MS2 genome other than TR are capable of inducing capsid assembly (Beckett *et al.*, 1988; Toropova *et al.*, 2009). Surprisingly the icRNA seemed the most efficient at promoting capsid assembly of all the MS2 genomic RNAs under observation despite the absence of the TR operator. This corresponded well with the idea that smaller sub-genomic RNAs are packaged more efficiently than larger RNAs, discussed in 3.2.2.3.

Having shown that all MS2 RNAs with or without the TR operator could promote capsid assembly it was decided to investigate whether this also applied to non-MS2 RNA. Capsid assembly was performed with a 1.89 kb control mRNA corresponding to the elongation factor 1 $\alpha$  gene of *Xenopus Laevis*<sup>2</sup>. The results were not definitive as the migration rate of the RNA was very close to the migration rate of the capsid. This made it difficult to distinguish between unbound RNA and capsid however they did indicate that the foreign RNA was more prone to aggregation with CP<sub>2</sub> rather than capsid formation perhaps an indication of non-specific RNA binding.

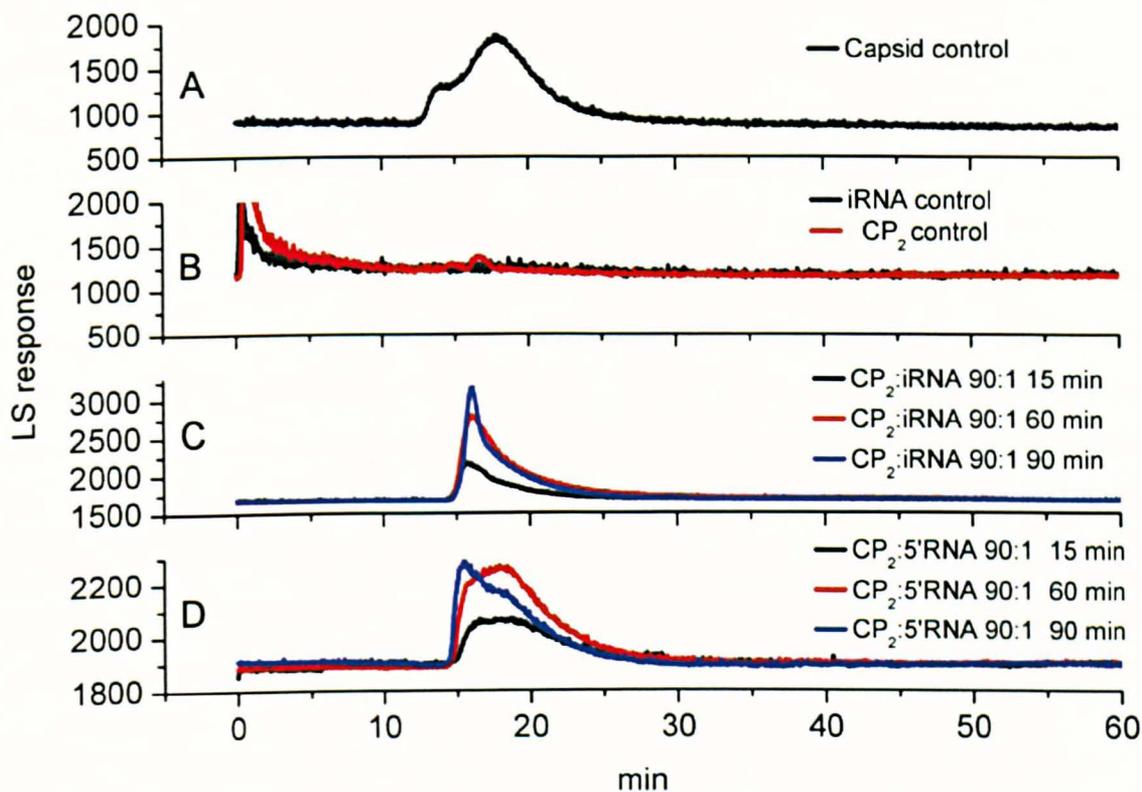
### 3.2.2.5 MS2 capsid assembly assayed with light scattering

In order to verify the observed effects of the sub-genomic RNAs on capsid assembly and get information on assembly kinetics, reassembly reactions were monitored by measuring changes in laser light scattering. All molecules scatter light with intensities which can be related to their molecular weight (Tinoco *et al.*, 1995). Light scattering has been used successfully to monitor capsid assembly kinetics as it allows the change in molecular weight to be monitored (Zlotnick *et al.*, 2000; Casini *et al.*, 2004). Furthermore, a light scattering detector which monitors the output from a gel filtration column (GFLS) allows the production of size distribution profiles of biological samples. GFLS has been used successfully in this laboratory to assay MS2 capsid assembly reactions initiated with oligonucleotide RNAs (Stockley *et al.*, 2007; Basnak *et al.*, 2009).

MS2 capsid assembly was promoted with the 5' and iRNA sub-genomic RNAs in assembly buffer and assayed at timepoints up to 90 min at 37 °C (**Figure 3-20**). Unfortunately the resolution of assembly products was not good. Capsid products were not sufficiently well separated from capsid-like aggregates. The results do however imply that the reactions have reached equilibrium as changes in light scattering were not observed between the 60 and 90 min timepoints. Identical reactions were assayed with the iRNA at 4° and at room temperature with no apparent changes in elution profiles after 60 min. It was decided not to optimise gel filtration conditions as advised by the GFLS facility manager Dr Gabriella Basnak.

---

<sup>2</sup> This fragment was transcribed from pTRI-Xef supplied as a template control for *in vitro* transcription with the Megascript transcription kit from Ambion.



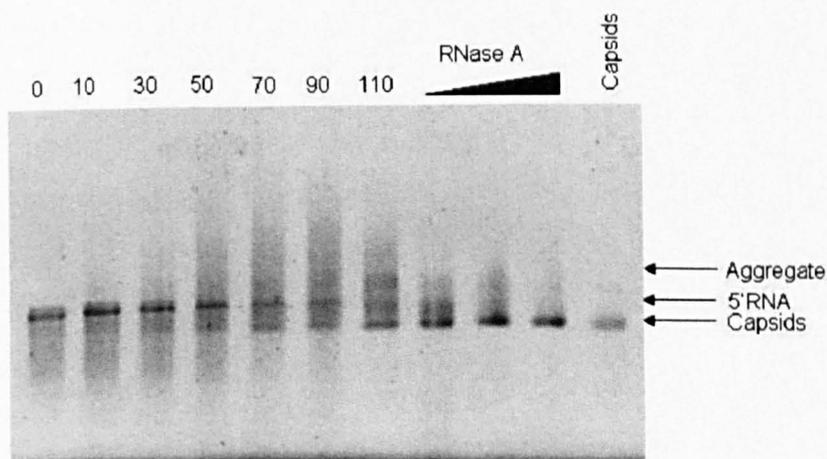
**Figure 3-18. Capsid assembly assayed with gel filtration coupled light scattering.** Assembly reactions induced with the 5' RNA and iRNA were assayed at a molar ratio of 90:1 at three time points. The traces show light scattering response vs. elution time off the gel filtration column. **A** shows recombinant capsids which elute off the column after 18 min. **B** shows the elution profiles of CP<sub>2</sub> and iRNA. A small signal can be seen at 15 and 17 min in the CP<sub>2</sub> sample after 1 h incubation in assembly buffer. The iRNA was not detected with LS however an absorbance trace that was collected simultaneously along with the LS trace indicated that iRNA is eluted off the column at 22-38 min. The 5' RNA was not assayed on its own. **C** and **D** show elution profiles of capsid assembly reactions with the iRNA and 5' RNA at increasing time points respectively. The dominant LS peak observed at 15-25 min is composed of at least two components. The sharper peak at ~16 min has been shown to correspond to aggregated capsid like products by TEM (Dr. Gabriella Basnak, personal communication). Taking into account the GEMSA of capsid assembly it is likely that these products represent capsids and aggregate which are not separated well during gel filtration.

### 3.2.2.6 The 5' RNA induced capsid product shows RNase resistance

The TEM micrographs of the symmetric capsid products assembled with the sub-genomic RNAs (**Figures 3-13-16**) indicated that they were similar in size and shape to MS2 virions (Katerina Toropova, personal communication). It has however been shown that *Levivirus* phage particles reconstructed *in vitro* with vRNA are different from infectious virions as observed by their sedimentation properties (addressed in section 4.2.1.1) and their susceptibility to RNase (Hung *et al.*, 1969; Sugiyama, 1967;

Hohn, 1969). These differences are thought to arise from both the absence of the A-protein and protrusion of RNA from particles which are then susceptible to RNase. RNase susceptibility is therefore an indication of fully formed MS2 capsid particles as the capsids confer ribonuclease resistance (Argetsinger and Gussin, 1966).

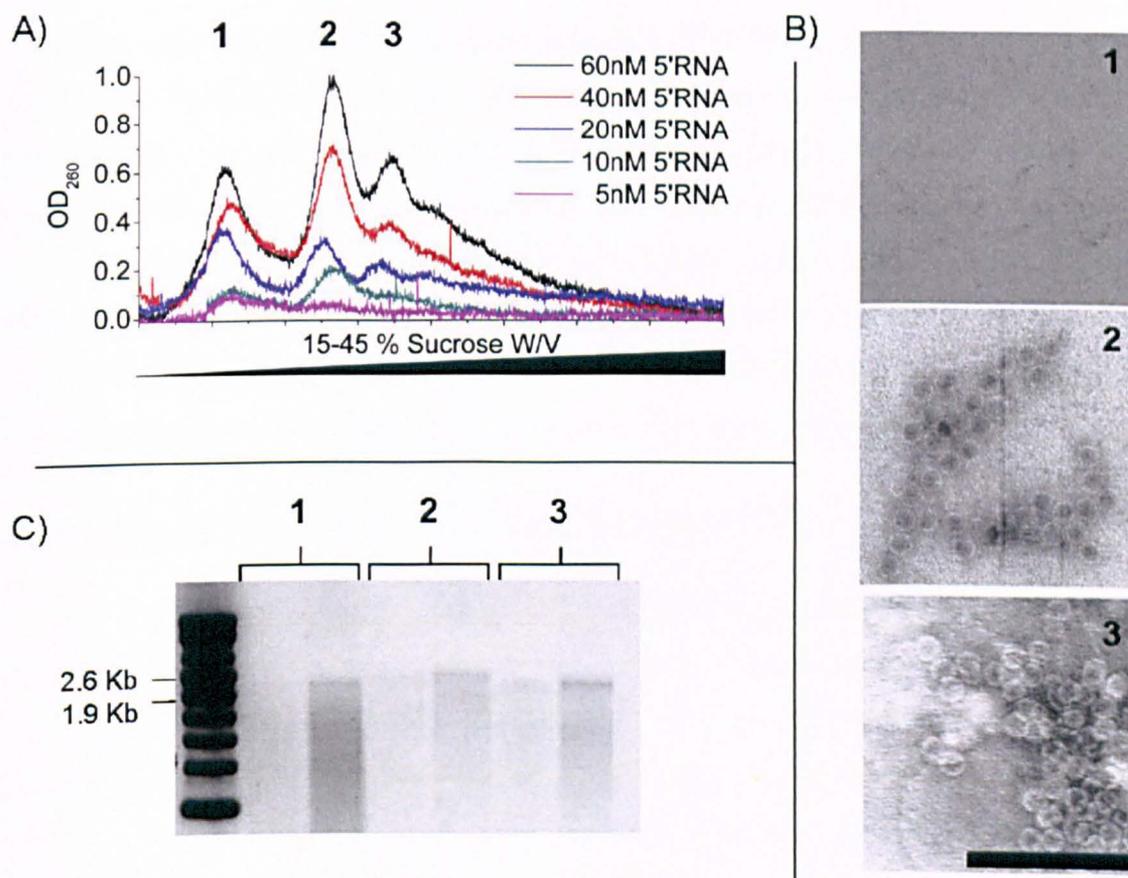
In order to investigate whether the capsids assembled with the sub-genomic RNAs represented fully formed capsids, a capsid assembly reaction initiated with the 5' RNA at a molar ratio of 110:1 of CP<sub>2</sub>:RNA was treated with RNase A (**Figure 3-18**) for 15 min at increasing concentrations of RNase A. The 5' RNA within the capsid appeared resistant to nuclease cleavage while aggregate products were degraded suggesting that the assembled capsids protect packaged RNA against ribonuclease cleavage while aggregates are more prone to degradation. Similar reactions were performed for up to 1 h which similarly showed that ~50 % of RNA was degraded upon treatment with RNase A. This corresponds with the yields of capsids obtained from *in vitro* assembly with the 5' RNA as shown in **Figure 3-14** and was discussed in 3.2.2.3.



**Figure 3-19.** *The 5' sub-genomic RNA capsid product shows RNase resistance. Capsid assembly reactions were performed at increasing molar ratios of CP<sub>2</sub>:5'RNA and analysed on a 1 % native agarose gel. The CP<sub>2</sub>:5'RNA molar ratio is indicated above each lane. Identical reactions at 110:1 CP<sub>2</sub>:5'RNA were treated with increasing amounts of RNase A. A decrease in signal intensity corresponding to capsid aggregates was observed while no decrease in intensity of the signal corresponding to MS2 capsids could be seen.*

### 3.2.2.7 Assembled capsids contain the sub-genomic RNA

In order to verify whether the capsids that were reconstructed with the sub-genomic RNAs contained the correct sized RNA, capsid assembly reactions induced with the 5' RNA were subjected to sucrose density centrifugation (**Figure 3-19**). Capsid assembly was performed at a molar ratio of 60:1 CP<sub>2</sub>:5'RNA rather than 110:1 to allow identification of all reaction components. Three samples were isolated off the sucrose gradients which corresponded to the 5' RNA, capsid and aggregate components as suggested by comparison to the 5' RNA GEMSA (**Figure 3-14**). Denaturing agarose electrophoresis of the samples showed that full length RNA was present in the capsid samples indicating that the correct length RNA was indeed being incorporated into capsids.



**Figure 3-20. MS2 capsid assembly products contain sub-genomic RNA.** Capsid assembly products were purified on sucrose density gradients and analysed by denaturing agarose electrophoresis and TEM. **A)** Absorbance traces of sucrose density gradients of MS2 capsid assembly with the 5' sub-genomic RNA following centrifugation at 17000 rpm for 16 h. Five reactions were analysed at increasing RNA concentrations, all at a molar ratio of 60:1 CP<sub>2</sub>:5'RNA, in order to optimise the detection of assembly products using a UV-detector that was coupled to a fraction collector for sample isolation. The absorbance of all samples is shown normalised to the highest peak in the 60 nM RNA sample (black trace). The samples corresponding to the peaks labelled 1-3 from each reaction were pooled to afford samples 1-3. **B)** TEM micrographs of these samples indicated that capsids were only present in samples 2 and 3. The diameter of 10 capsid particles was measured directly off the TEM micrographs and found to be the same. The black size bar represents 200 nm. **C)** The RNA from samples 1-3 was purified by phenol/chloroform extraction and analysed on a 1 % denaturing MOPS gel. Full length 5' RNA can be seen in all samples but is less degraded in 2 and 3. The slight difference in migration rates is due to different amounts of the RNA carrier Glycoblue in each sample.

### 3.3 Discussion

RNA nucleates ssRNA virus capsid assembly. Employing the bacteriophage MS2 as a model system, the properties of genomic RNA during capsid assembly can be investigated. Using MS2 sub-genomic RNA of defined lengths, it was anticipated that the importance of RNA size and structure on MS2 capsid assembly could be explored.

The access to sub-genomic MS2 RNA of different lengths and the question whether MS2 sub-genomic RNA would promote assembly of capsids *in vitro* represented initial questions in this project. Below, the successful production of MS2 sub-genomic RNA as well as reports on how these RNAs promote capsid assembly is discussed.

#### 3.3.1 The production of MS2 sub-genomic RNA

The structure of the MS2 genome within the MS2 virion provided a foundation for further research into the structural role of the MS2 genome during capsid assembly. The organisation of the genome into two concentric shells within the virion which represent, in principle, roughly 1/3 and 2/3 of the genome provided an incentive for producing RNA fragments whose sizes corresponded to the amount of RNA in these shells. The sub-genomic RNA fragments were produced in four steps which involved RT-PCR to produce cDNA, cloning of the cDNA, amplification from the cDNA clones with a T7 promoter containing primer and finally *in vitro* transcription.

It proved difficult to generate cDNA that corresponded exactly to the desired cDNA product lengths due to RNA sequence constraints affecting DNA primer design. Reverse transcription reactions were initially non-specific towards the desired cDNAs but were optimised successfully by manipulating reaction conditions. It proved pivotal to use polymerase enzymes with high thermal stability in RT-PCR reactions which allowed efficient denaturation of the RNA and DNA templates during amplification. This was perhaps not surprising given the high degree of base pairing and compactness reported for the MS2 RNA genome (Fiers *et al.*, 1976; Skripkin *et al.*, 1990). Three MS2 cDNA products were produced corresponding to two thirds of the MS2 genome from the 5' and 3' termini along with a nearly-full length cDNA sub-genomic fragment. These cDNAs allowed the generation of nearly any sized RNA using the strategy already outlined.

The MS2 cDNAs were cloned into replicating plasmids and maintained in *E.coli* in order to obtain them in relatively homogenous and copious amounts for downstream reactions. Initially, cloning of the cDNAs was avoided as production of MS2 cDNA clones had been reported to be problematic do to the lysogenic nature of the MS2 phage (Remaut *et al.*, 1982; Hill, 1993). However the cDNAs obtained from RT-PCR could not be used to template further production of T7 containing cDNA and efforts at producing a full length cDNA by ligation or annealing of the 5' and 3' cDNAs were hampered by low yields from RT-PCR reactions. Cloning of the MS2 cDNAs proved essential in that it provided the three cDNAs in good yield and they were used successfully to template the production of T7 promoter containing cDNA by PCR.

The cDNA clones deviated slightly in sequence from the reported wild type sequence of MS2 (Fiers *et al.*, 1976). This deviation was shown to be most likely due to different infectious phage genotypes, supported by comparison to other infectious MS2 RNA sequences. The MS2 phage genome has recently been shown to rapidly incorporate adaptive mutations induced by environmental growth conditions (Betancourt, 2009). True mutations of conserved amino acids within the maturation gene of the 5'- and 3' cDNA clones and replicase gene of the 3' and nearly-full length cDNA clones were however also identified. Given the difficulties in cloning these cDNAs, it is likely that the cloning procedure resulted in positive selection of non-lethal cDNA clones. This was further suggested in experiments, not described in detail here, where expression from the 3528 nt nearly-full length cDNA, which had been cloned into a pET expression vector, resulted in CP<sub>2</sub> production with no detectable formation of phage as judged by SDS-PAGE and plaque assays of *E.coli* cell extract and growth medium. The specific effects of these mutations on the virus life cycle were not pursued.

Despite the MS2 cDNA fragments containing possible mutations it was decided to use these cDNAs for the production of MS2 RNA. It is generally accepted that ssRNA viral genomes exist as ensembles of non-identical but closely related sequences which in part accounts for their rapid adaptability to environmental changes (Domingo *et al.*, 1997). This sequence variety is generated by the error prone replicase and does not seem to affect the ability of RNA to be encapsidated as suggested by the large number of defective phage (Rhode *et al.*, 1995; Brakmann *et al.*, 2001). This implies that the structural plasticity of the RNA genome is not likely to be affected by the introduction of a small number of nucleotide mutations.

The sub-genomic RNA fragments were transcribed *in vitro* from their corresponding cDNA templates affording three MS2 RNA sub-genomic transcripts that were used to assay their ability to promote capsid assembly.

### 3.3.2 Sub-genomic MS2 RNA nucleates MS2 capsid assembly

The MS2 sub-genomic RNAs were shown to promote capsid assembly as observed by their ability to generate products that migrated at similar rates as recombinant and virion capsids within native agarose gels. This had been partly anticipated in light of the pioneering work of Tomas Hohn into the role of fr RNA on fr capsid assembly (Hohn, 1969). The conservation of coat protein aa residues associated with RNA binding (Peabody, 1993) and similar capsid crystal structure of MS2 and fr (Liljas *et al.*, 1994) had also suggested that the RNA packaging mechanism for these two phages was likely to be similar. With the exception of GA, this similarity with respect to RNA genome packaging can not be extended to other phages within the *Leviviridae* as the structure of the B-type coat protein FG-loop is dissimilar for these viruses (Plevka *et al.*, 2009). This suggests a different mechanism of quasi-equivalent control of coat protein dimers for these viruses and therefore perhaps a different mechanism of RNA packaging.

The MS2 capsid assembly products were, as observed by TEM, shown to resemble *in vitro* capsid assembly products reported earlier for icosahedral bacteriophages (Sugiyama *et al.*, 1969; Hohn, 1969; Matthews and Cole, 1972). TEM was mainly used as an analytical tool to verify the presence of MS2 capsids in assembly reactions. No particular emphasis was put into further characterisation of the capsid products using TEM. As part of a group effort to investigate the MS2 genome packaging mechanism, all detailed structural analysis using EM and cryo-EM with the sub-genomic RNAs was in the hands of Dr. Katerina Toropova (Toropova, 2009) and are discussed in section 4.3.

The assembled capsid products were shown to contain the correct size RNA when capsid assembly was induced with the 5' sub-genomic RNA and the encapsulated RNA shown to be partly resistant to RNase cleavage.

By monitoring capsid reassembly reactions with gel mobility shift assays at different molar ratios of CP<sub>2</sub> to RNA, the concentration dependence of capsid assembly with the sub-genomic RNAs was investigated. These reactions showed that capsid

formation is dependent upon CP<sub>2</sub> concentration. Capsid formation proceeds at low CP<sub>2</sub>:RNA molar ratios or down to 5:1 CP<sub>2</sub>:RNA depending on the RNA. This confirms the ability of RNA on promoting capsid assembly. This suggests that genomic RNAs are able to recruit CP<sub>2</sub> which are able to form capsids at concentrations where the CP<sub>2</sub> might otherwise be expected to be evenly bound to the entire RNA population. This corresponds to accepted models of capsid formation where assembly occurs co-operatively and irreversibly following formation of a capsid intermediate nucleation complex (Stockley *et al.*, 2007; Zlotnick, 1994). The MS2 nucleation complex was first identified during *in vitro* assembly at equal molar ratios of CP<sub>2</sub>:vRNA (Capecchi and Gussin, 1965; Sugiyama *et al.*, 1969). Non-covalent mass spectrometry of assembly reactions with TR indicate that this complex might correspond to either a hexamer of CP<sub>2</sub> or a pentamer of CP<sub>2</sub> (Stockley *et al.*, 2007). The ability of RNA to form this complex is possibly a dominating factor in the ability of RNA to promote capsid assembly. A co-operative assembly pathway takes advantage of the limited concentration of CP<sub>2</sub> in solution. The virus perhaps takes advantage of this in order to minimise encapsidation of host RNA.

At higher ratios of CP<sub>2</sub>:RNA, approximately 50:1 and above, the formation of aggregates in the assembly reactions became apparent. TEM micrographs indicated that these aggregates correspond to incomplete capsid-like material. It is likely that these aggregates form due to non-specific CP<sub>2</sub> binding to genomic RNA and initiation of assembly at multiple sites on the same RNA strand. This would inevitably produce capsid intermediates incapable of combining to produce a  $T=3$  capsid. Using a sub-genomic MS2 RNA lacking TR it was shown that capsid assembly can take place in the absence of TR so this scenario is not unlikely at high CP<sub>2</sub> concentrations. Taking into account the results from chapter 4, where evidence is provided that the RNA genome is folded into the volume of the capsid, it is likely that non-specific binding could also hinder folding of the RNA into an assembly competent structure producing kinetic traps that are then prone to aggregation.

The CP<sub>2</sub> concentration dependence of capsid assembly efficiency was not the same for the genomic RNAs. This implied that RNA size is an important factor for efficient capsid assembly. This was observed from the genomic RNAs efficiency to promote capsid formation being inversely proportional to RNA size. With respect to the co-operative manner in which assembly is observed to proceed, the length dependence could be associated with the number of CP<sub>2</sub> binding sites available on the sub-genomic

RNAs. At identical reaction ratios, the CP<sub>2</sub> will have to “sample” fewer binding sites on short RNAs as compared to long RNAs in order to associate and form an assembly competent seed complex i.e. at identical reaction ratios, the smaller RNAs will on average have CP<sub>2</sub> bound in closer proximity to on another as compared to the larger RNAs. This would result in a higher probability of formation of an assembly nucleation complex at lower reaction ratios for smaller RNAs. In addition, if the co-operative mechanism of assembly is due to the RNA directing the “migration” of CP<sub>2</sub> towards the growing capsid then CP<sub>2</sub> will on average be bound to RNA closer to the growing capsid on small RNAs as compared to large. The distance the CP<sub>2</sub> would have to “migrate” to be incorporated into a capsid enclosing a small RNA would therefore be shorter and the coat protein dimer less likely to diffuse into the bulk solution than if a larger RNA were being encapsidated. Another factor which could influence the RNA length dependence on capsid formation efficiency is negative charge repulsion of the RNA phosphate backbone. It is perhaps easier to encapsidate smaller RNAs as charge neutralisation upon CP<sub>2</sub> binding will be greater for smaller RNAs resulting in lower charge repulsion of the RNA backbone during packaging. Additionally the cryo-EM of the MS2 virion (Toropova *et al.*, 2008) suggests that it could be harder to package larger RNAs due to a more complex conformation that these RNAs must adopt in order to fit into the capsid.

The results confirmed reports suggesting that various secondary structure elements within the MS2 genome are capable of promoting capsid assembly (Beckett and Uhlenbeck, 1988; Koning *et al.*, 2003). A sub-genomic RNA lacking the TR operator was packaged efficiently. What property of the RNA does then allow for efficient packaging? Assembly reactions with a foreign mRNA sequence indicated that this sequence was not as efficient at inducing assembly as genomic MS2 RNA. It is unlikely that this RNA does not contain stem loop structures capable of inducing capsid assembly. Reports that indicate that large foreign albeit icosahedral viruses RNAs can be packaged while ribosomal RNAs of similar size cannot seem to indicate that the structural plasticity of the RNA may be of importance (Sugiyama *et al.*, 1967; Hohn, 1969). Furthermore, in Yoffe *et al* (2008) it was argued that evolutionary constraints imposed by icosahedral capsids on ssRNA have resulted in viral RNAs being spatially more compact than RNAs of similar nucleotide composition. Finally, the cryo-EM reconstruction of the MS2 phage suggests that the arrangement of stem-loop structures within the RNA genome could be of importance for assembly. This is discussed further in 4.3.2.

The results reported here suggest two possible reaction mechanisms in which genomic length RNA might promote assembly. In the first, RNA is only necessary to induce formation of the nucleation complex after which capsid assembly can proceed most likely through dimer addition with no further help from the RNA. If this is the case, the key to the catalytic properties of RNA lie in its ability to promote the formation of the pentameric/hexameric nucleation complex. In the second mechanism, after formation of the nucleation species, RNA could assist in capsid formation throughout the assembly process. The two identified attributes of genomic RNA during capsid assembly, namely co-operativity and RNA size dependence, imply that the latter mechanism is more likely to be true. Although capsid assembly is co-operative in the absence of RNA (Matheus and Cole, 1972) the size dependence implies that the RNA is an active participant in the assembly process. Taking into account the manner in which the genome interacts with the capsid shell (Toropova *et al.*, 2008) and the control of quasi-equivalent CP<sub>2</sub> conformations which must occur at defined positions within the capsid (Stockley *et al.*, 2007), this catalytic assembly mechanism for genomic RNA seems plausible.

Initial efforts into monitoring the capsid assembly reactions with GEMSA were not successful as the migration of the 5'-, 3'- and vRNA and capsid products in the gels was very similar resulting in poor resolution of reaction products. This was eventually solved by running the agarose gels at low voltages for longer periods of time ( 2 ½ h) the results of which have already been discussed. This problem however fuelled ideas that the size/hydrodynamic radius of the RNA was perhaps an underlying requirement for packaging and that the diameter of the RNA to be packaged had to conform to the inner diameter of the MS2 capsid. This seemed plausible as a recent report had suggested that some viral RNA tertiary structures are highly stable, unfolding only at high temperature (Kuznetsov *et al.*, 2005). This suggested that the RNA perhaps acted as a prefolded structural template upon which capsid assembly could take place.

In order to verify the observed RNA length dependence on capsid assembly and investigate further the manner in which genomic RNA is packaged into the MS2 capsid it was decided to assay the assembly reactions with analytical ultracentrifugation. The results of these experiments are described in the next chapter.

## 4 MS2 genomic RNA encapsidation studied by sedimentation velocity analysis

---

### 4.1 Introduction

In the previous chapter, the production of sub-genomic MS2 RNAs of different lengths and corresponding to different regions of the MS2 genome was reported. The RNAs were analysed with respect to their ability to promote MS2 capsid formation *in vitro*. Interestingly, these investigations suggested that the efficiency of capsid formation is dependent upon RNA length. This suggests that the role of the MS2 genome during capsid assembly is not confined to a single CP<sub>2</sub>-RNA recognition event. The results indicate a complex assembly process where multiple CP<sub>2</sub>-RNA interactions are of importance. This correlates well with the complex RNA fold of the virion RNA observed in the cryo-EM model of the MS2 phage. Furthermore, the results confirm suggestions that stem loop structures other than TR must be capable of quasi-equivalent conformer switching between A/B and C/C type coat protein dimers both of which are required for capsid formation (Stockley *et al.*, 2007; Toropova *et al.*, 2008; Beckett *et al.*, 1988). The manner in which genomic ssRNA is packaged into virus particles is of enormous interest. ssRNA viruses are amongst the most prevalent viral pathogens. These viruses all share the common task of packaging their genomes into a confined protein container. The mechanism by which this occurs is not understood (Rudnick and Bruinsma, 2005; Schneemann, 2006; Yoffe *et al.*, 2008). The observed RNA length-dependence on packaging efficiency implies a capsid assembly mechanism where RNA is compacted into the MS2 capsid.

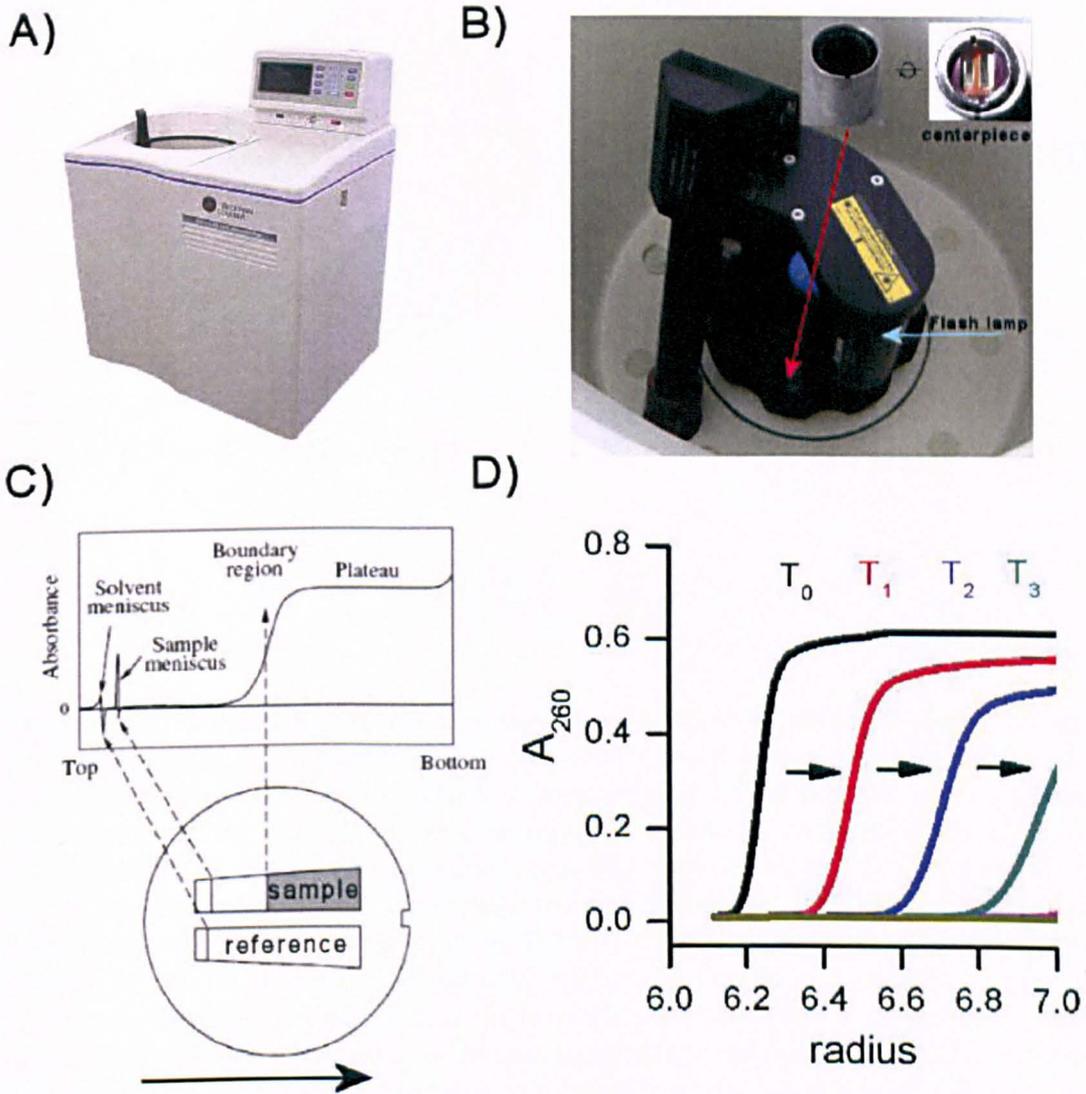
In order to gain further insight into the mechanism in which RNA is packaged into the MS2 capsid, a different analytical technique was sought which could complement or confirm the findings reported in chapter 3. This chapter reports results wherein MS2 capsid assembly was assayed with analytical ultracentrifugation. In section 4.2.2, sedimentation velocity analysis of MS2 capsid assembly, induced with MS2 genomic RNA and the MS2 sub-genomic RNAs is reported. In section 4.2.3, similar analysis of MS2 capsid assembly reactions, albeit induced with RNA oligonucleotides is reported. These results are then discussed with respect to the packaging mechanism of the MS2 genome and the sequence specificity underlying coat protein dimer quasi-equivalent

conformer switching. The results presented in this chapter have contributed to two manuscripts which have been submitted to peer-reviewed journals at the time this is written: Rolfsson and Toropova *et al.* (2009) and Basnak *et al.* (2009).

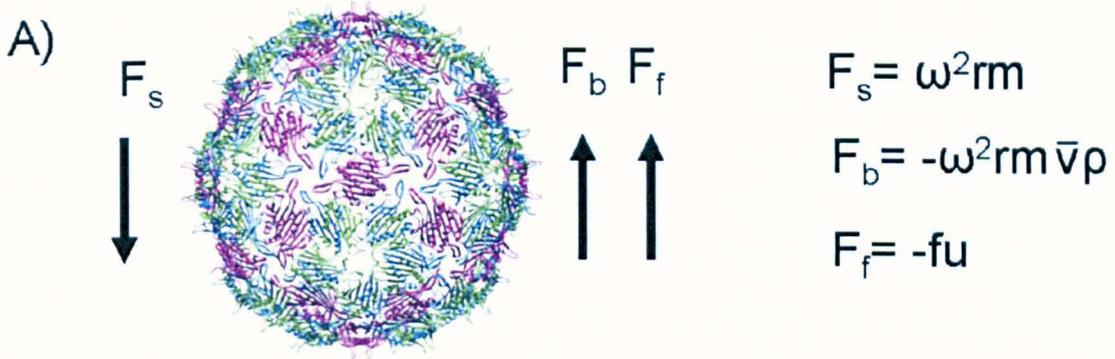
#### 4.1.1 Sedimentation velocity analysis

Molecules can be fractionated and characterised based on their sedimentation rate under an applied centrifugal force. This type of analysis is referred to as sedimentation velocity analysis (SV) and can be performed using an analytical ultracentrifuge. The analytical ultracentrifuge is comparable to a preparative centrifuge, however it is equipped with a laser light detection system that allows the direct measurement of the sedimentation rate of macromolecules in solution. This is achieved by measuring the change in the concentration distribution of a sedimenting particle over time (**Figure 4-1**). When particles such as virus capsids sediment under a centrifugal force, a boundary is formed which represents the concentration gradient formed during sedimentation. The displacement of the boundary over time is dependent upon the sedimentation rate of the molecule in solution and to a lesser extent its diffusion during sedimentation (reviewed in Stafford, 1997; Lebowitz *et al.*, 2002).

The method was largely developed by Svedberg in 1940. Although it has since been surpassed in terms of speed and sensitivity, sedimentation velocity analysis is still an important analytical method due to its versatility. Biological macromolecules of nearly any size can be investigated within a broad range of buffer conditions and the results are directly related to the molecular properties of the macromolecule based on the physical postulates comprising hydrodynamic theory (Schuck, 2000). The sedimentation rate of molecules is dependent upon molecular mass, shape and buoyancy as implemented in the Svedberg equation (**Figure 4-2**). The frictional coefficient implemented in the Svedberg equation connects the sedimentation rate of molecules to their size and shape as defined in the Stokes-Einstein equation (Tinoco *et al.*, 1995). Sedimentation velocity analysis therefore allows characterisation of the hydrodynamic properties of molecules by measuring their sedimentation rate (reviewed in Lebowitz *et al.*, 2002).



**Figure 4-1. Sedimentation velocity analysis.** A) A Beckman XI-1 analytical ultracentrifuge. The sliding top door on the ultracentrifuge allows access to a vacuum chamber shown in B which houses a fixed angle rotor and a xenon flash lamp. The rotor houses eight AUC centerpieces. Each centerpiece contains two sector shaped compartments which hold the sample and reference buffer in dialysis equilibrium. A centerpiece is shown in inset from the side and a top down view. C) A schematic of a centerpiece is shown. The arrow indicates the direction of sedimentation. The concentration distribution of a sample is obtained by measuring the absorbance at a fixed wavelength by scanning the sector cells while the rotor spins. This produces a sedimentation profile as an absorbance vs. radius plot where the radius represents the distance from the axis of rotation. The absorbance profile describes the concentration distribution of the sample within the sector shaped sample volume. D) Collection of sedimentation profiles at different time points ( $T_0$ - $T_3$ ) during sedimentation allows the rate of sedimentation of a molecule to be calculated from the movement of the sedimentation boundary over time.



B)

$$F_s + F_b + F_f = 0 \quad (1) \quad f_0 = 6\pi\eta r \quad (3)$$

$$S = \frac{u}{\omega^2 r} = \frac{m(1-\bar{v}\rho)}{f} \quad (2)$$

**Figure 4-2. Forces acting on a sedimenting particle in solution under an applied centrifugal field.** *A*) A particle, such as a virus capsid, in a centrifugal field is affected by a sedimentation force ( $F_s$ ) which is proportional to the particle mass ( $m$ ) and the acceleration of the particle in the centrifugal field defined as angular velocity squared times the radius ( $\omega^2 r$ ). The buoyant force ( $F_b$ ) opposes  $F_s$  and is proportional to the weight of displaced solvent.  $F_b$  is proportional to angular acceleration, the mass and partial specific volume of the particle ( $\bar{v}$ , which is the inverse of density) and the density ( $\rho$ ) of the solvent. The driving force that causes the particle to sink is the difference between  $F_s$  and  $F_b$ . As the particle sinks however it experiences frictional drag ( $F_f$ ) which is proportional to the particles frictional coefficient and sedimentation velocity ( $u$ ).  $F_f$  increases with increased velocity of the particle and balances out the driving force of sedimentation which results in the particle sedimenting at terminal velocity. *B*) At terminal velocity the sum of the forces acting on the sedimenting particle are zero (1). Rearrangement of the forces acting upon the particle produces the quantity ( $u/\omega^2 r$ ) which is referred to as the sedimentation coefficient ( $S$ ).  $S$  has the dimension of seconds but is expressed in Svedberg units, defined as  $10^{-13}$  sec. The Svedberg equation (2) shows that sedimentation rate is proportional to the mass and buoyancy but inversely proportional to the size/friction of a particle. The Stokes equation (3) can be combined with the Svedberg equation to allow estimation of the degree of hydration and shape of a sedimenting particle.

#### 4.1.1.1 Sedimentation boundary analysis

The appeal of SV analysis has been boosted by the development of methods for whole sedimentation boundary analysis. Traditionally, the sedimentation rate of a particle has been identified by measuring the displacement of the boundary at its midpoint (Stafford, 1997; Tinoco *et al.*, 1995). Whole boundary analysis however involves inspection of the displacement and shape of the whole concentration gradient

during sedimentation. This is beneficial as it takes into account the transport properties of molecules during sedimentation which ultimately affect sedimentation and allows identification of all sample components simultaneously (Stafford, 1997; Schuck, 2000; Howlett *et al*, 2006).

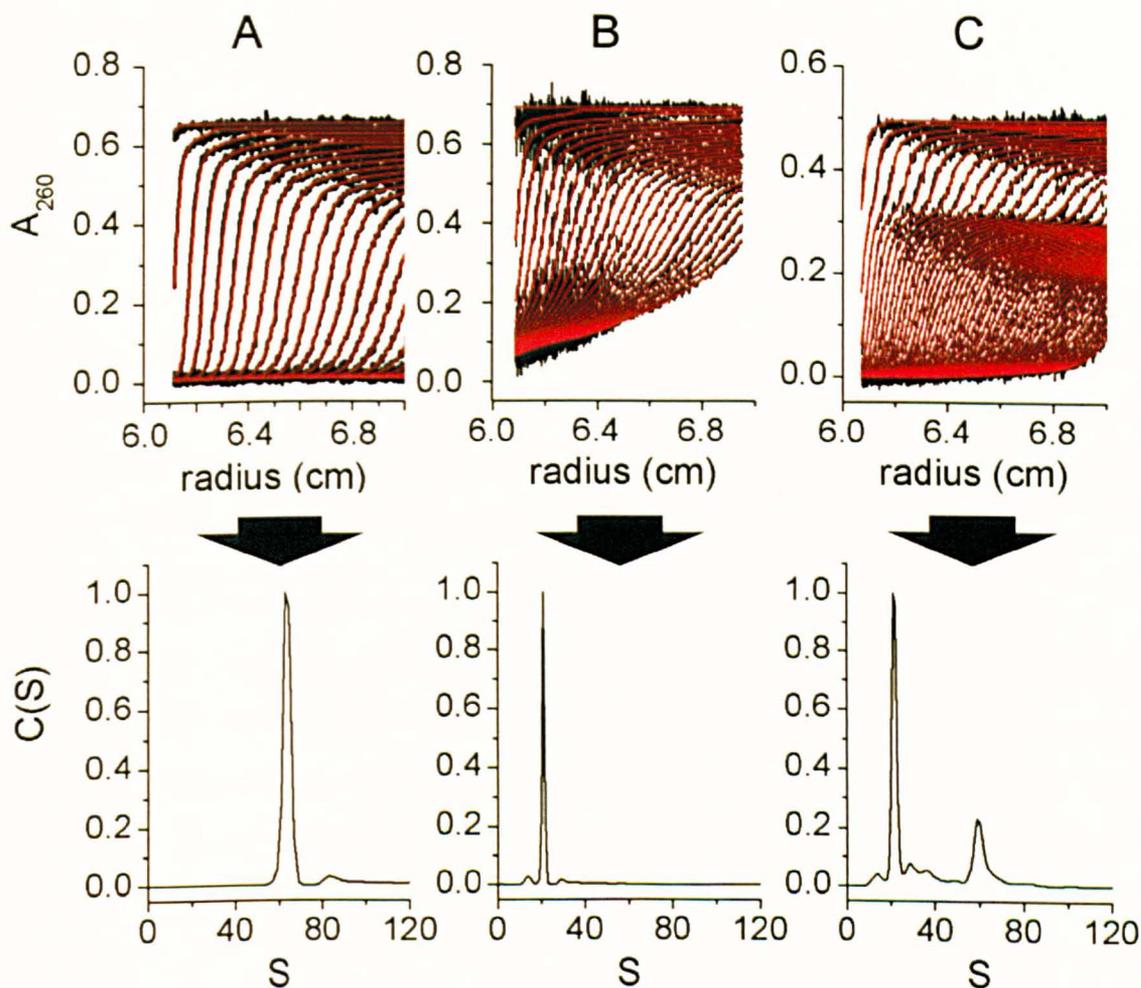
An example of boundary analysis is shown in **Figure 4-3**. For a fast sedimenting homogenous sample which conforms to ideal behaviour during sedimentation, analysis is straightforward. This is observed for a sample such as purified MS2 capsids as shown in the top half of **Figure 4-3 A** which has concentration distribution profiles which are well resolved and of similar shape over time. While the displacement of the concentration boundary due to sedimentation is most obvious, the change in boundary displacement is also caused by molecular diffusion which opposes the concentration gradient formed during sedimentation. This results in spreading of the boundary during sedimentation which can complicate data analysis. Boundary spreading is most prominent at later stages during sedimentation when a large concentration gradient has formed towards the bottom of the sample cell due to sample accumulation. This can be observed in the top half of **Figure 4-3 B**. The spreading of the concentration boundary is also attributed to non-ideal behaviour. For a highly negatively charged RNA molecule this would correspond to increased electrostatic repulsion at later times during sedimentation (Ralston, 1995).

Perhaps most crucially, the shape of the boundary during sedimentation is informative of the sample composition. When particles of different size and molecular weight sediment, the concentration gradient formed during sedimentation will reflect the stoichiometry of the particles in solution producing a boundary shape that is characteristic of the composition of the sample. This effect can be observed in the top half of **Figure 4-3 C** where components which differ largely in molecular weight sediment with different rates. SV analysis thereby allows the size distribution analysis of sample components.

Sedimentation velocity analysis uses the strongly size dependent time course of boundary migration to build a size distribution profile. As shown in **Figure 4-3 B** the sedimentation profiles are however convoluted by the hydrodynamic properties of the particles which complicates data analysis because it becomes more difficult to resolve subpopulations of the size distribution. A whole boundary concentration distribution model has been described by Peter Schuck (2000) which takes into account the effect of boundary broadening due to diffusion. This model, referred to as the  $c(s)$  model, is

fairly powerful in reporting the qualitative features of sample size distribution. The model fits the sedimentation profiles to Lamm equation solutions which describe the concentration distribution of a dilute sample sedimenting over time in a sector shaped cell (Schuck, 2000). For a tutorial review see Lebowitz *et al.* (2002). The fitting model is implemented in the computer program Sedfit (Lebowitz *et al.*, 2002) which allows automatic computational analysis of sedimentation profiles. **Figure 4-3** shows the size distribution profiles obtained by fitting the sedimentation distribution profiles to the  $c(s)$  model. The output of the fitting process is a  $c(s)$  plot which shows a continuous concentration dependent sedimentation coefficient distribution over a region of selected  $S$ -values. The model is dependent upon prior knowledge of the density of the solvent, the partial specific volume of the sedimenting component and the shape of the molecule. However errors in estimates of these properties do not affect the resolution of sample components drastically if the components sediment with fairly different rates (Schuck, 2000).

The hydrodynamic properties of ssRNA viruses within the Levivirus genus are well characterised (Boedtke and Gesteland, 1975). The sedimentation coefficients and partial specific volume for both the MS2 phage and MS2 RNA have been determined by Strauss and Sinsheimer (1963) and others (Overby *et al.*, 1966; Fiers *et al.*, 1967). Density centrifugation, where sample components are analysed after centrifugation through a density gradient, has also been used successfully to characterise MS2 capsid assembly reactions with full length viral RNA (Sugiyama *et al.*, 1967) as outlined in section 1.91. The gel mobility shift assays of MS2 capsid assembly reported in chapter 3 did initially not provide good resolution of assembly reaction components. Furthermore a gel filtration light scattering assay did not provide resolution between reassembled capsids and aggregated material. It was decided to assay capsid assembly reactions using SV applying the  $c(s)$  method to provide size distribution profiles of the MS2 capsid assembly reactions with the sub-genomic MS2 RNAs described in chapter 3.



**Figure 4-3. Sedimentation velocity size distribution analysis with the  $c(s)$  model.** The top three plots show three examples of sedimentation profiles (black lines) for MS2 capsids (A), the 5' sub-genomic MS2 RNA (B) and a capsid assembly reaction (C). They are shown here as an example in order to highlight macromolecular behaviour during sedimentation and the ability of sedimentation velocity to distinguish between sample components. Large particles such as MS2 capsids sediment fast which is evident from a wide boundary distribution (A). The 5' RNA fragment sediments slower which is evident from the more tightly spaced concentration profiles (B). The individual boundaries are however also spread out over a larger region within the sample cell as compared to the MS2 capsid. This is due to the concentration gradient build up and back diffusion during sedimentation. An assembly reaction produces sedimentation profiles with a double S-like shape which correspond to a mixture of components (C). The sedimentation profiles can be fit to a continuous size distribution  $c(s)$  Lamm equation model (red traces) affording the sedimentation coefficient distribution plots shown below. The capsid assembly distribution profiles clearly show two sedimenting components which are well resolved on the  $c(s)$  plot.

## 4.2 Results

### 4.2.1 Characterisation of the components of MS2 capsid assembly by SV

In order to be able to assay MS2 capsid assembly reactions using SV it was necessary to obtain the sedimentation coefficients of all the MS2 capsid assembly components. This entailed monitoring the size distribution of both the MS2 capsid, CP<sub>2</sub> and all the genomic MS2 RNAs individually. These could then be used as standards in the SV assays of MS2 capsid assembly reactions.

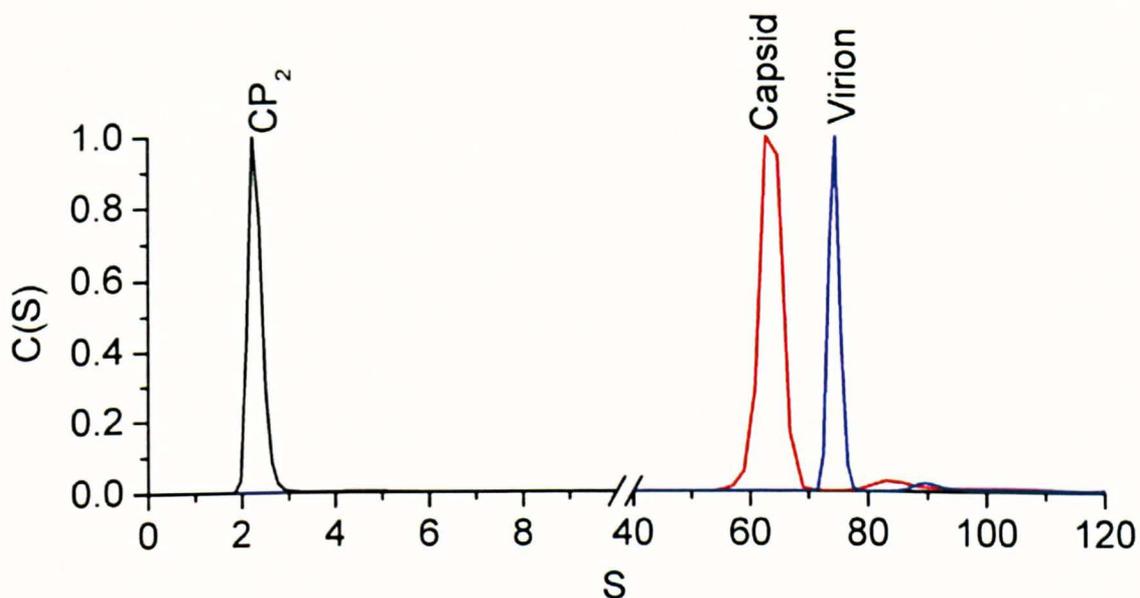
#### 4.2.1.1 The MS2 capsid and coat protein

The sedimentation distribution profiles of the MS2 virion, the MS2 recombinant capsid and the MS2 CP<sub>2</sub> were obtained by monitoring their sedimentation over time and fitting the concentration distributions to the  $c(s)$  model as outlined in 2.2.12. This allowed the characterisation of these components based on their sedimentation rate. **Figure 4-4** shows the sedimentation coefficient distribution plot obtained for these three components. The identified sedimentation coefficients are given in **Table 4-1**. Examples of sedimentation distribution profiles of the components are shown in Supplementary Figure 1 in the appendix.

The CP<sub>2</sub> was observed to sediment as a homogenous population with an  $S_{20w}$  of 2.3. The CP<sub>2</sub> has previously been observed to sediment at 1.9 S in 20 mM acetic acid but at 2.3 S in TMK buffer at pH = 7.4 (Mr. Andy Baron, personal communication). The sedimentation coefficient of the coat protein of the closely related Levivirus f2 has similarly been reported as 1.9 S under acidic conditions (Matthews and Cole, 1972). This difference in sedimentation rate suggests self association of the coat protein at physiological pH and is consistent with the idea that the stable form of the CP is that of a dimer (Beckett and Uhlenbeck, 1988; Stockley *et al.*, 2007).

The observed  $S_{20w}$  of the recombinant MS2 capsid was 65.8. This value corresponds well with a previously determined in-house  $S_{20w}$  of 66.1 for recombinant capsids (Mr. Andy Baron, personal communication). The MS2 virion sedimented faster with an  $S_{20w}$  of 75.1 which is most likely due to its higher molecular mass. It is unlikely that the capsid sediments slower due to increased friction. Cryo-EM reconstructions of recombinant MS2 capsids assembled with the TR operator and the MS2 virion indicate

an identical diameter for these particles (Toropova *et al.*, 2008). The MS2 virion has previously been reported to have an  $S_{20w}$  of 78.5-81 (Strauss and Sinsheimer, 1963; Overby *et al.*, 1966 Fiers, 1967). The lower sedimentation rate observed for the MS2 virion compared to these values is most likely due to different experimental conditions.



**Figure 4-4. Sedimentation coefficient distributions of the CP<sub>2</sub>, recombinant MS2 capsids and MS2 virions.** The three components were analysed individually in assembly buffer. The CP<sub>2</sub> was allowed to equilibrate for 3 h prior to analysis. The sedimentation distribution plots were overlaid to produce the *c(s)* plot shown. The purified CP<sub>2</sub> (black), capsid (red) and virion (blue) all sediment as fairly homogenous populations at 2.3  $S_{20w}$ , 65.8  $S_{20w}$  and 75.1  $S_{20w}$  respectively. A small component sedimenting slightly faster than the main peak can be seen in the capsid and virion samples indicating a degree of self association. The virion is different from the recombinant capsid in that it contains the A-protein and the full length MS2 genome. The recombinant capsid does not have the A-protein and contains random cellular RNA. Due to these differences in shape and mass, the MS2 virion sediments faster than the recombinant capsid. Capsids devoid of RNA have been shown to sediment with an  $S_{20w}$  of 45 (Mathews and Cole, 1972).

If a molecule of known mass and density is spherical and unsolvated its radius can be calculated and inserted into the Stokes equation (**Figure 4-2**) to determine the particle frictional coefficient (Lebowitz *et al.*, 2002). This frictional coefficient ( $f_0$ ) will be the smallest frictional coefficient that a particle of that molecular mass can have.  $F_0$  can be inserted into the Svedberg equation to calculate the maximum S value ( $S_{max}$ ) of the particle. The ratio of  $S_{max}/S_{observed}$  is equal to the frictional ratio of the particle ( $f/f_0$ ) which is an indicator of particle hydration and shape. A  $f/f_0 = 1$  will indicate a completely spherical particle, a completely unsolvated molecule or both while a

deviation from unity will indicate a degree of hydration or a non-spherical shape (Tinoco *et al.*, 1995). For the MS2 virion the  $S_{\max}$  is 86. Along with the  $S_{20w}$  of 75.1 obtained for the MS2 virion the calculated  $f/f_0$  is 1.14. This correlates well with the spherical shape of the MS2 virion. The known molecular mass and frictional ratio can be used to obtain the Stokes radius of the MS2 virion which was calculated at 11.6 nm. **Table 4-1** lists the frictional ratios of the MS2 recombinant capsid and CP<sub>2</sub>. Similar frictional ratios and Stokes radii were obtained for the MS2 recombinant capsid and the virion, while the values are lower for the CP<sub>2</sub> as expected. The Stokes radius of the MS2 virion is smaller than the radius observed by X-ray crystallography (13.75 nm) (Golmohammadi *et al.*, 1993). It is possible that this error is due to the estimates of the density of the buffer and the estimation of the diffusion coefficient which is implemented in the  $c(s)$  fitting of the data. The obtained frictional ratios and Stokes radii are however intercomparable and indicate a similar size for the MS2 capsid and virion which is consistent with previous reports.

**Table 4-1. The hydrodynamic properties of the MS2 capsid assembly reaction components.** Sedimentation was monitored in 40 mM ammonium acetate, 1 mM Mg(OAc)<sub>2</sub>, pH= 7.2 at 20°C. The table shows the average experimentally observed sedimentation coefficients for the MS2 capsid assembly components ( $S_{exp}$ ) and the standard sedimentation coefficients corrected for buffer conditions, in water at 20 °C ( $S_{20w}$ ). The estimated frictional ratio ( $f/f_0$ ), observed molecular weight ( $M_{exp}$ ) and calculated molecular weight ( $M_{calc}$ ) are also shown. The observed and calculated molecular weights are similar as indicated by  $M_{exp/calc}$ .  $R_s$  denotes the Stokes radius which is obtained from the  $M_{exp}$  and the  $f/f_0$ . All the RNAs were found to have sedimentation coefficients and  $R_s$  proportional to their size. Including EDTA in the assembly buffer results in a slower sedimentation rate of the larger RNAs.

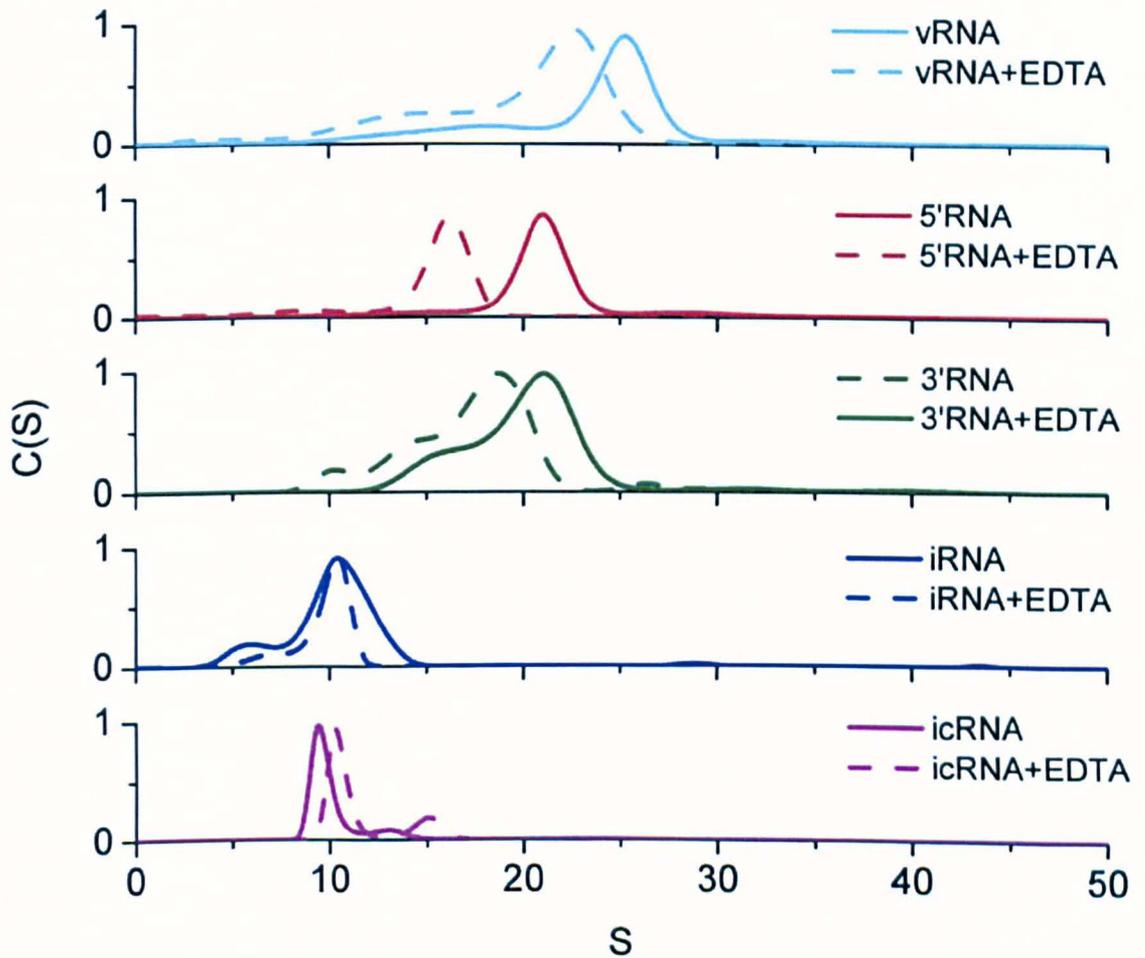
	$S_{exp}$	$S_{20w}$	$f/f_0$	$M_{exp}$ (kDa)	$M_{calc}$ (kDa)	$M_{exp}/M_{calc}$	$R_s$ (nm)
MS2 Virion	74.6 ± 0.52	75.1	1.1	3663.2	3662.2	0.03 %	11.6
MS2 capsids	65.4 ± 2.0	65.8	1.2	3237.8	-	-	11.7
CP <sub>2</sub>	2.3 ± 0.17	2.3	1.4	27.6	27.5	0.52 %	2.8
Virion RNA	25.7 ± 1.48	25.8	3.0	1140.7	1148.3	0.66 %	18.4
Virion RNA+ 25 mM EDTA	22.1 ± 1.76	22.3	3.4	1140.7	1148.3	0.66 %	21.3
5'RNA	20.4 ± 1.33	20.5	2.9	783.8	794.3	1.33 %	15.9
5'RNA + 25 mM EDTA	17.3 ± 1.21	17.4	3.4	773.1	794.3	2.73 %	18.4
3'RNA	21.2 ± 1.30	21.3	2.9	825.9	829.6	0.45 %	16.1
3'RNA + 25 mM EDTA	18.5 ± 1.59	18.6	3.5	823.2	829.6	0.77 %	19.5
iRNA	9.7 ± 0.85	9.8	3.2	301.1	298.5	0.84 %	12.8
iRNA + 25 mM EDTA	10.6 ± 1.76	10.7	3.0	302.9	298.5	1.45 %	11.8
icRNA	9.7 ± 0.47	9.7	3.1	285.2	281.1	1.45 %	12.1
icRNA + 25 mM EDTA	10.5 ± 1.43	10.5	2.9	290.5	282.1	3.26 %	11.4

#### 4.2.1.2 Characterisation of genomic MS2 RNA

The sedimentation rate of the MS2 virion RNA extracted from MS2 phage and the MS2 sub-genomic fragments were assayed with SV in a similar manner as described in the previous section. The obtained sedimentation coefficients and calculated hydrodynamic properties of these RNAs are given in **Table 4-1**. The sedimentation coefficient distributions are shown in **Figure 4-5**. The alignment of the sub-genomic RNAs to the MS2 genome is shown in **Figure 3-11**.

The MS2 virion RNA sedimented with a dominating component at  $25.7 \pm 1.48$  S. The sedimentation rate of the MS2 genome has been reported previously at 25.5-28 S dependent upon solvent, which is in good agreement with these results (Strauss and Sinsheimer, 1963; Fiers, 1967). The sedimentation distribution plot indicated that the RNA was not structurally homogenous. Sedimenting material observed as a shoulder centered at 16.6 S was detected. This material corresponded to 20 % of the total material. This peak is due to either conformational heterogeneity of the RNA or RNase contamination prior to or during analysis. Fiers (1967) reported that the MS2 genome sediments as three distinct populations at 16 S, 21 S and 27 S which he concluded was due to contaminating enzymatic cleavage of the RNA. Strauss and Sinsheimer (1963) reported the sedimentation coefficient of 14 S for the RNA denatured in formaldehyde indicating that the slower sedimenting material could correspond to full length RNA.

The sub-genomic MS2 RNAs were observed to sediment slower than the full length genome as expected based on their lower molecular mass. These RNA samples were each dominated by a major species with an S-value proportional to its length and corresponding to sedimentation coefficients of RNAs of similar size (Fiers, 1967; Brosius *et al.*, 1978; Brosius *et al.*, 1980). The 3' sub-genomic RNA had a wider sedimentation distribution on average than the 5' RNA perhaps indicating a broader distribution of conformational states. A fraction of the 3'- and iRNA samples were observed to sediment slower than the dominating species, similar to what was observed for the full length MS2 genome indicating a degree of conformational heterogeneity of these RNAs. Every precaution was taken in order to avoid RNase contamination during sample handling and RNAs were homogenous as observed by denaturing agarose electrophoresis prior to analysis.



**Figure 4-5 Sedimentation coefficient distributions of the MS2 genomic RNAs.** The RNAs were analysed in assembly buffer in the presence and absence of 25 mM EDTA. The RNAs all sediment with *S*-values proportional to their length. The sedimentation coefficients are given in **Table 4-1**. Removal of magnesium ions by addition of EDTA to the assembly buffer allowed assessment of the extent of tertiary structure of the RNAs. Addition of EDTA results in the longer RNAs sedimenting more slowly.

In order to investigate the extent of tertiary structure in the genomic RNAs in assembly buffer, the change in RNA population size distributions upon removal of magnesium ions with EDTA was assessed. RNA structure has been shown to be sensitive to  $Mg^{2+}$  which stabilises tertiary structure through charge neutralisation of the phosphate backbone (Shcherbakova *et al.*, 2008). The capsid assembly buffer contains magnesium ions at 1 mM concentration. EDTA chelates  $Mg^{2+}$  and it was therefore anticipated that addition of EDTA to the assembly buffer would interrupt RNA tertiary structure due to  $Mg^{2+}$  stabilisation. Treatment with EDTA resulted in a 10–15 % reduction in the observed  $S_{20,w}$  values for the 5'-, 3'- and vRNA with a corresponding increases in the frictional coefficient and stokes radii. The slower sedimentation rate is

suggestive of a more expanded RNA structure indicating that these RNAs have a tertiary fold in assembly buffer that is partly dependent upon magnesium ion concentration. The vRNA has previously been shown to be more compact at higher  $Mg^{2+}$  concentrations by TEM under denaturing conditions (Jacobson, 1976). The iRNA and icRNA were not observed to expand upon addition of EDTA.

The drop in sedimentation rates of the larger RNAs is not as significant as observed for 23 S rRNA (2903 nt), which is similar in size to the 5'- and 3'RNA. Upon removal of magnesium ions a 60 % decrease in sedimentation rate was observed with a corresponding increase in frictional ratio (Yi and Wong, 1982). This difference confirms the compactness of the MS2 genome due to the high degree of complementary base pairing as compared with RNAs of similar length (Fiers *et al.*, 1976; Yoffe *et al.*, 2008)

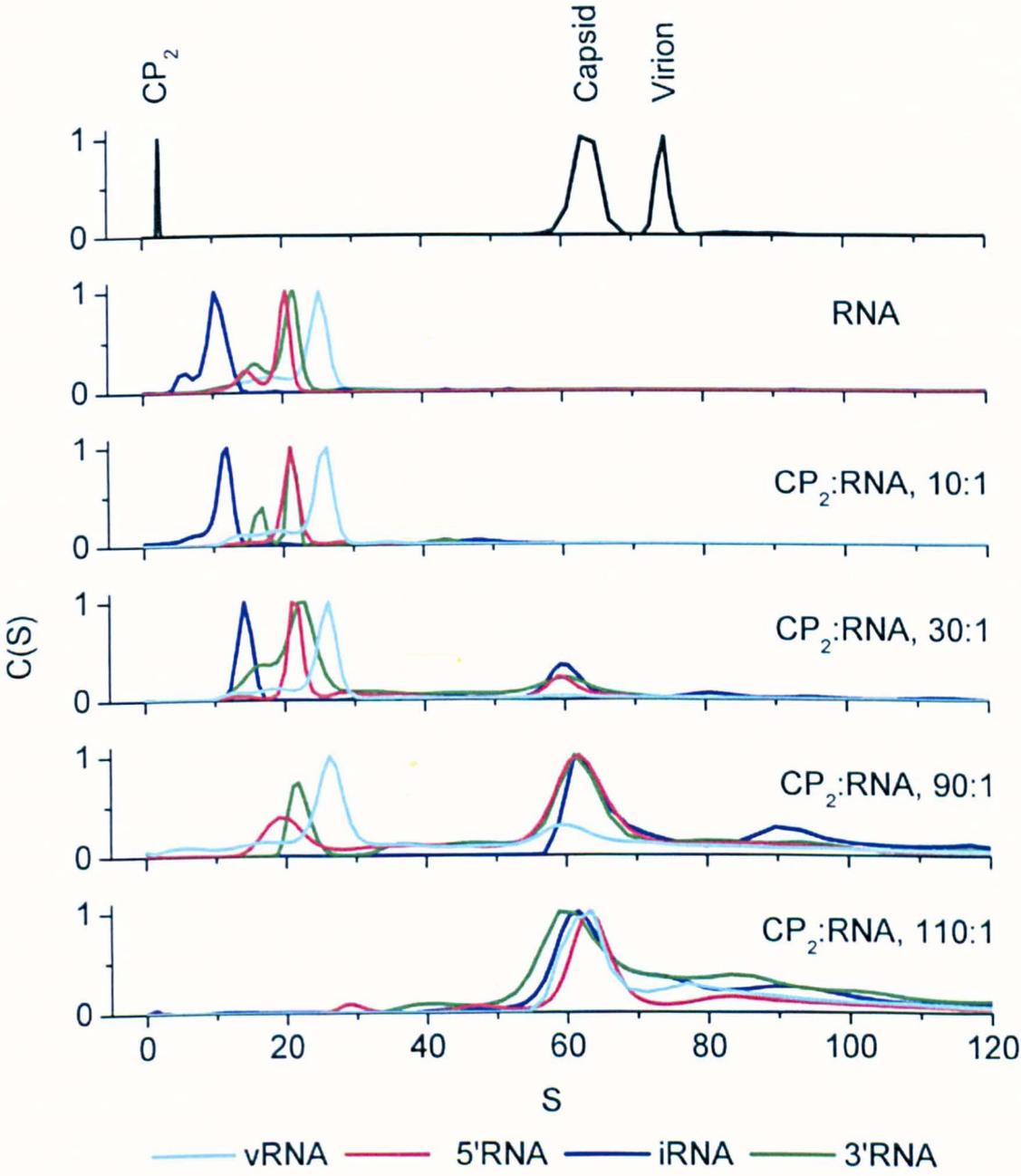
The frictional coefficient ratios of all the genomic RNAs analysed indicate a hydrated or highly asymmetric RNA structure in assembly buffer. Interestingly, the hydrodynamic radii of all the RNAs is larger than that observed for the virion and capsid. These results indicate that encapsidation of these RNAs must involve conformational rearrangement of their solution structures prior to or during capsid assembly.

#### ***4.2.2 Analysis of MS2 capsid assembly by SV***

Having successfully characterised the MS2 capsid assembly reactants and capsid products based on their sedimentation rates, MS2 capsid assembly reactions could now be investigated with SV. Capsid assembly was performed at increasing reaction stoichiometries of CP<sub>2</sub>:RNA in assembly buffer and analysed after 4 h as described in 2.2.12.

**Figure 4-6** shows the sedimentation coefficient distributions of MS2 capsid assembly with each RNA fragment at reaction stoichiometries from 10:1 to 110:1 of CP<sub>2</sub>:RNA. The amounts of sedimenting components corresponding to both RNA and capsid in the assembly reactions are given in **Table 4-2**. As the protein concentration is increased an increasing amount of RNA sedimenting with S values comparable to the recombinant capsid is observed. These species had sedimentation coefficients of 62 - 63 S independent of the genomic RNA and correspond to the sedimentation rate of

capsids reconstituted with vRNA which has been previously reported (Sugiyama *et al.*, 1967; Hohn, 1969; Mathews and Cole, 1972; Beckett and Uhlenbeck, 1987).



**Figure 4-6. Sedimentation coefficient distributions of MS2 capsid assembly reactions initiated with the genomic MS2 RNAs.** Normalised *c(s)* vs. *S* plots of capsid assembly reaction components (two top-most plots) and of individual capsid assembly reactions with each of the RNAs at increasing CP<sub>2</sub> concentrations after 4 h at 20 °C in 40 mM ammonium acetate, 1 mM Mg(OAc)<sub>2</sub>, pH = 7.2. The shorter RNA fragments promote capsid formation at lower CP<sub>2</sub> concentrations than the longer RNA fragments suggesting that it is harder to package larger RNAs into the volume of the capsid. The apparent lack of intermediates supports a co-operative capsid assembly process. The amount of sedimenting species corresponding to RNA and capsid are given in Table 4-2.

**Table 4-2. The observed quantities of RNA and capsids in the MS2 capsid assembly reactions assayed by SV. The assembly reaction stoichiometry is shown in the leftmost column. The numbers in each row represent sedimenting material corresponding to either RNA or capsid reported as a percentage of the total sedimenting material at the indicated reaction stoichiometry.**

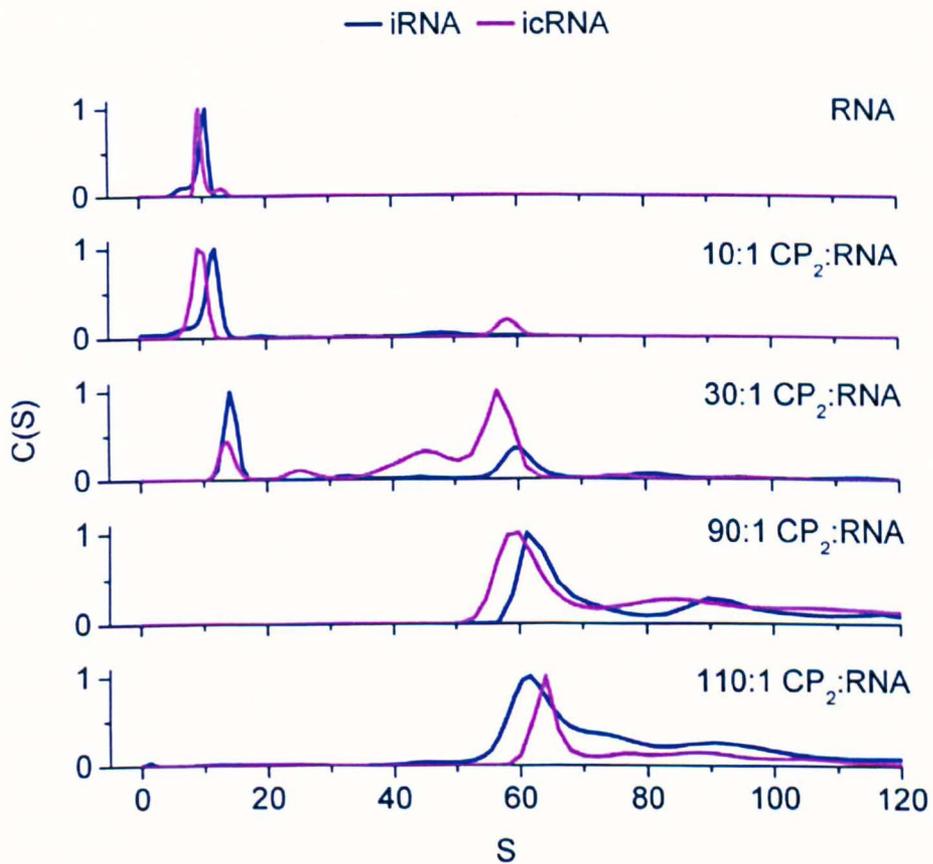
CP <sub>2</sub> :RNA	vRNA		5'RNA		3'RNA		iRNA	
	RNA (%)	Capsid (%)						
0	100.0	-	100.0	-	100.0	-	100.0	-
10	89.8	-	84.0	-	89.3	10.6	79.5	12.1
30	81.2	6.3	46.6	26.5	74.4	25.1	40.2	35.9
90	41.8	21.0	18.5	49.5	14.6	50.7	0.0	44.3
110	0.0	47.7	0.0	56.6	0.0	50.4	0.0	55.4

Few species were observed that sedimented with intermediate S-values between the genomic RNA and the capsid product. Those that did could not be detected in a reproducible manner. An exception to this was the sedimenting component corresponding to the iRNA which sedimented faster at increasing CP<sub>2</sub> concentrations perhaps indicating an RNA-CP<sub>2</sub> assembly intermediate complex. A similar change in sedimentation rate of the peaks corresponding to other genomic RNAs was not observed. At higher concentrations of CP<sub>2</sub>, assembly products sedimenting faster than the MS2 capsid became visible, most likely corresponding to capsid-like aggregates as observed by TEM and described in the previous chapter. The results are consistent with the conclusions from the GEMSA assays and are suggestive of a co-operative capsid assembly process.

The RNAs are packaged with different efficiencies as judged by their ability to promote formation of species sedimenting at the same rate as the  $T = 3$  capsid. The iRNA is packaged most efficiently. Capsids were detected at low CP<sub>2</sub>:iRNA reaction ratios and all of the iRNA was protein bound at a reaction ratio of 90:1. The 5'- and 3'RNAs also promoted capsid assembly at low CP<sub>2</sub>:RNA ratios, however at 90:1, more species corresponding to free RNA could be observed as compared to the iRNA. The vRNA was least efficient at promoting capsid assembly. Fewer capsids were detected sedimenting at 30:1 and 90:1 as compared to the sub-genomic RNAs. These results suggest that the efficiency of encapsidation is inversely proportional to the length of the RNA. This is consistent with the GEMSA assays of capsid assembly. The implication of these results together with the obtained hydrodynamic radii of the assembly components suggest that the MS2 genomic RNA is folded into the MS2 capsid during

capsid assembly and is therefore not a prefolded structural scaffold upon which capsid assembly takes place.

Similar to the GEMSA results, few differences were observed between the 5' and 3' RNAs which suggested a difference in their ability to promote capsids dependent upon their sequence and/or structure. The icRNA and the iRNA, which are of similar size, also appeared to promote capsid formation with similar efficiencies (**Figure 4-7**). The lack of the TR operator does not appear to affect the ability of genomic RNA to produce capsids consistent with the idea that various stem loops are capable of inducing capsid assembly (Beckett and Uhlenbeck, 1988; Koning *et al.*, 2003; Toropova *et al.*, 2008). The observed lack of sequence specificity indicates that CP<sub>2</sub> folds MS2 genomic RNA in a manner which is independent of the genomic sequence. The results do not suggest that assembly is independent of RNA sequence in general.



**Figure 4-7. SV assay of icRNA confirms that RNAs lacking the TR operator promote capsid assembly.** Normalised sedimentation distribution plots of the iRNA and the icRNA show that the absence of the TR operator does not result in a reduced ability to promote MS2 capsid formation. The icRNA confirms that smaller RNAs are packaged more efficiently than larger RNAs.

The sedimenting species corresponding to the  $T = 3$  capsids all sedimented with similar rates independent of the RNA packaged and slower than the MS2 virion. The difference in sedimentation rate of capsids reconstituted with the vRNA and MS2 virions is thought to be due to the absence of the A-protein (Sugiyama *et al.*, 1967) and the possibility that RNA protrudes from the particle which results in frictional drag during sedimentation (Hohn, 1969). The RNase cleavage assay reported in section 3.2.2.6 does not indicate whether this is the case. A difference in gel mobility shift would be expected if a large portion of the RNA protruding from the capsid would be excised. This is not clear from **Figure 3.19**. In order to investigate this, assembly reactions with the 5' RNA at 110:1 CP<sub>2</sub>:RNA were treated with RNase A in a similar manner as before and assayed with SV. The results from these reactions were unfortunately not conclusive. Reaction controls of 5' RNA in the absence of both RNase A and CP<sub>2</sub> was slightly degraded, however capsid products sedimented slightly slower in the presence of RNase compared to in its absence (data not shown). This suggests that the reconstituted MS2 RNA capsid products are not completely ribonuclease resistant and that RNA could protrude from the capsids. The RNA could however also be susceptible to ribonuclease due to temporary expansion of the capsid or loss of CP<sub>2</sub> subunits from the capsid. The RNase susceptibility of the reconstituted capsids correspond with results which have indicated that capsids, containing genomic length RNA are not ribonuclease resistant in the absence of the A-protein (Lodish *et al* 1965; Argetsinger and Gussin, 1966).

The  $c(s)$  sedimentation distribution model which was used to fit the data is strictly applicable to non-interacting species showing ideal behaviour. During fitting of the sedimentation velocity profiles an identical frictional ratio was used in fitting of all the data. If the RNAs protrude from the reconstituted capsid the frictional coefficients will be different for the different RNAs. Further investigations will have to be performed in order to understand how the reconstituted capsids are associated with different RNAs. An interesting experiment which would show whether the RNA is indeed fully enclosed in the particle would be to obtain electron micrographs of capsids reconstituted with an RNA that is end labelled with a heavy metal such as gold. The electron micrographs would then give ideas about the location of RNA termini. MS2 capsids have been successfully reconstituted with the TR operator which has been end labelled with a gold nanoparticle (Atul Mohan personal communication).

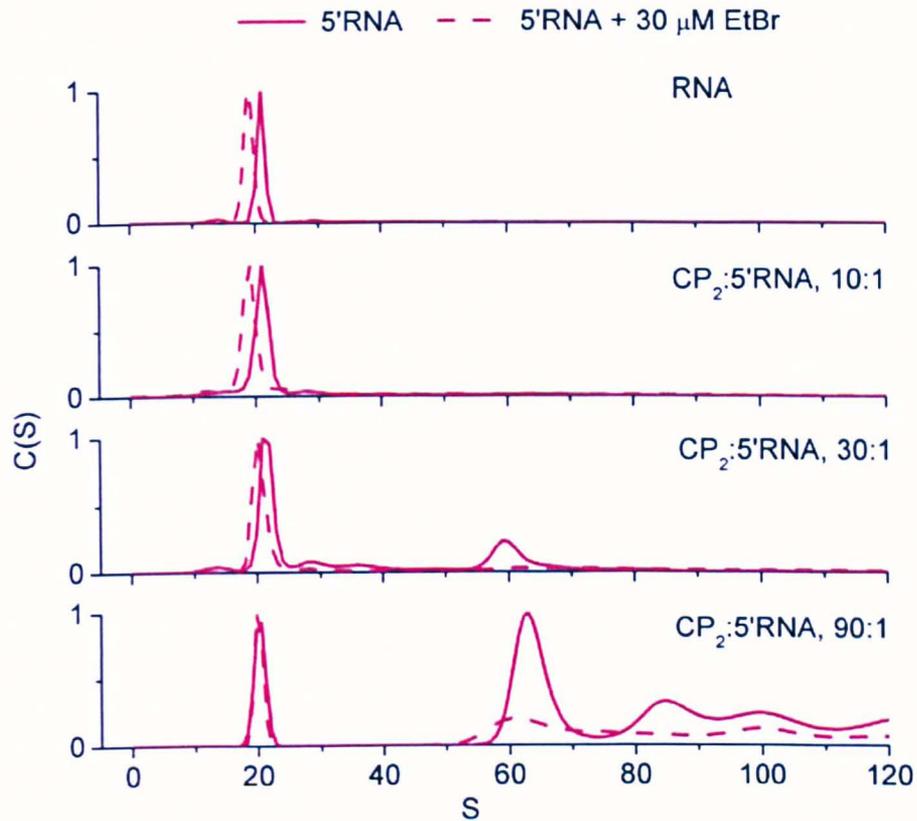
#### 4.2.2.1 Ethidium bromide inhibits folding of MS2 genomic RNA

An obvious consequence of the structural rearrangement of the RNA that must occur during encapsidation is that reagents preventing or inhibiting RNA folding should reduce packaging efficiency. To test this idea, capsid assembly reactions were performed in the presence of ethidium bromide (EtBr). EtBr intercalates between nucleotide base pairs (Waring, 1965; LePecq and Paolotetti, 1967) which results in distortion and stiffening of DNA and RNA structure by inducing a change in sugar pucker, glycosidic torsional angle and phosphodiester torsional angles (Sobell *et al.*, 1977; White *et al.*, 1987). The final result is a distorted RNA structure which should form MS2 capsids less efficiently than in the absence of EtBr. This is indeed what was observed (**Figure 4-8 and 4-9**).

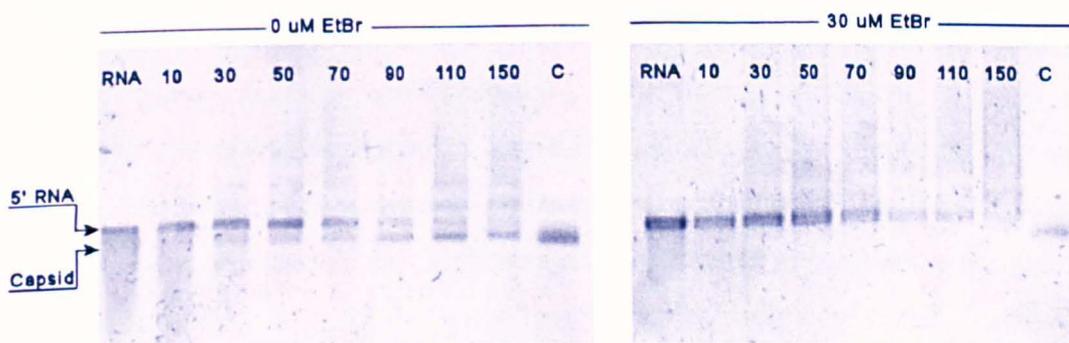
MS2 capsid assembly with the 5' RNA fragment was monitored with SV analysis and GEMSA in the presence of EtBr at a 30  $\mu$ M concentration. The concentration is close to the saturating concentration of EtBr bound to RNA at the genomic RNA concentration used in the assembly experiments and corresponds to one EtBr molecule bound every 5-10 nucleotides as observed by fluorescence binding measurements of total rat liver ribosomal RNA and tRNA (Waring, 1965; LePecq and Paolotetti, 1967; Bittmann, 1969).

In the presence of EtBr the 5' sub-genomic RNA sedimented at  $19.1 \pm 0.9$  S compared to  $20.4 \pm 1.3$  S in the absence of EtBr which correspond to average  $S_{20w}$  of 19.2 and 20.5, respectively. An increase in frictional ratio from 2.9 to 3.1 and an increase in hydrodynamic radius by 2 nm was estimated based on the sedimentation rates. These values perhaps indicate a slightly expanded RNA conformation in the presence of EtBr, however the  $c(s)$  model does not take into account the change in buffer density and viscosity upon addition of EtBr.

In the presence of EtBr assembly becomes notably less efficient. Many fewer capsids are observed to sediment at a 90:1 CP<sub>2</sub>:5'RNA reaction stoichiometry in the presence of EtBr. Similarly, few if any capsids are detected with the GEMSA in the presence of EtBr. These results suggest that RNA packaging becomes more difficult due to inhibitory effects induced by binding of EtBr to RNA and is consistent with the idea that an RNA conformational change accompanies genome packaging.



**Figure 4-8.** The 5' RNA is folded into the volume of the MS2 capsid. Normalised  $c(s)$  vs.  $S$  plots of capsid reassembly reactions at increasing  $CP_2:5'$ RNA ratios in the presence and absence of  $30 \mu\text{M}$  EtBr which corresponds to one EtBr molecule bound every 5-10 nucleotides. The RNA sediments slower in the presence of EtBr. Capsid assembly is notably less efficient in the presence of EtBr. This suggests that EtBr restricts the RNAs capability to adopt a structural conformation that can be packaged into the capsid.



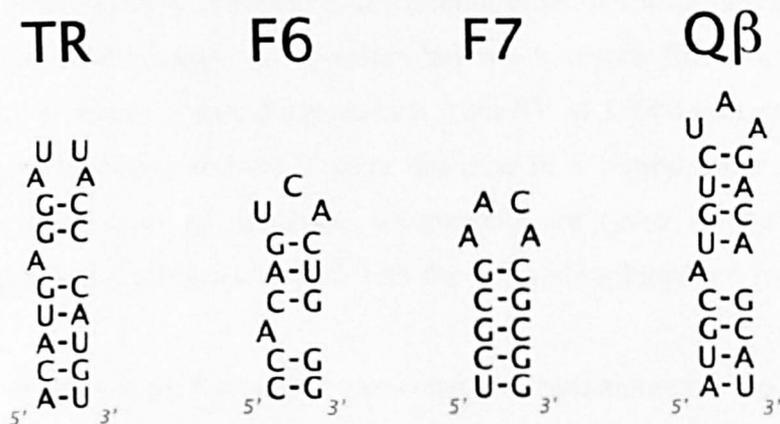
**Figure 4-9.** GEMSA assay of capsid assembly with the 5' RNA fragment with and without  $30 \mu\text{M}$  EtBr. The numbers above each lane indicate the molar ratio of  $CP_2$  to 5'RNA. Recombinant capsids are indicated with a C. In the presence of EtBr no material migrating at the same position as the capsid is observed. The GEMSA assay confirms the SV analysis and indicates that EtBr inhibits capsid assembly.

### 4.2.3 Stem loops other than TR promote capsid assembly

A molecular mechanism in which quasi-equivalent control of the A/B and C/C CP<sub>2</sub> conformers is achieved during capsid assembly has been proposed (Stockley *et al.*, 2007). The 19 nt stem loop operator TR has been shown to act as a molecular switch which upon binding to CP<sub>2</sub> induces a conformational change from a symmetric C/C like dimer to an asymmetric A/B like dimer (see section 1.9.2). The TR operator is however only present in a single copy in the MS2 genome and therefore, in principle, multiple switching events are required to generate the MS2 phage capsid. Since coat protein dimers bind only to RNA with secondary structure (Romaniuk *et al.*, 1987) the assembly experiments with the icRNA assayed by AUC and GEMSA suggest that stem loops are present within the MS2 genome that are capable of inducing a CP<sub>2</sub> conformational change in a similar manner as TR. In order to investigate the sequence dependency of quasi-equivalent conformer switching, MS2 capsid assembly was investigated by SV using RNA stem loops different from TR. These experiments were part of a dedicated group effort wherein the efficiency of different stem loops to induce capsid assembly was monitored with gel filtration light scattering (GFLS), TEM, and mass spectrometry.

MS2 capsid assembly was monitored in the presence of three RNA stem loops in addition to TR (**Figure 4-10**). These RNAs correspond to two RNA aptamers (F6 and F7) that are known to bind CP<sub>2</sub> at the same site as TR but with less affinity (Rowse *et al.*, 1998; Lago *et al.*, 1998) and the translational operator of the Allevovirus Q $\beta$ . The Q $\beta$  stem loop is functionally equivalent to TR but is discriminated against by MS2 CP<sub>2</sub> *in vivo* and is a poor binder *in vitro* (Ling *et al.*, 1970; Horn *et al.*, 2006).

The RNAs were analysed using sedimentation velocity analysis in a similar manner as before. The obtained S values and estimated hydrodynamic properties are given in **Table 4-3**. The sedimentation coefficient distributions of the RNAs are shown in **Figure 4-11** (second from top). The oligonucleotides sedimented with S<sub>20w</sub> values between 2.0 and 2.28 indicating that these RNAs are homogenous under these conditions. The F6 aptamer sedimented faster than the other oligonucleotides despite having the lowest mass, perhaps indicating a more compact structure as compared to TR, F7 and Q $\beta$ .



**Figure 4-10.** The RNA oligonucleotides used to study the sequence specificity of quasi-equivalent conformer switching. The stem loops are different in nucleotide sequence, length, hairpin size and with respect to the presence of a bulged adenine. TR F6 and Qβ are also different in that a different number of base pairs separates the bulged adenine from the loop.

**Table 4-3.** Hydrodynamic properties of the oligonucleotide series. The oligonucleotides were analysed in assembly buffer at 4 °C by SV at 40000 rpm. The table is similar to Table 4-1 and shows the observed sedimentation coefficients and the estimated hydrodynamic properties of the capsid assembly components.

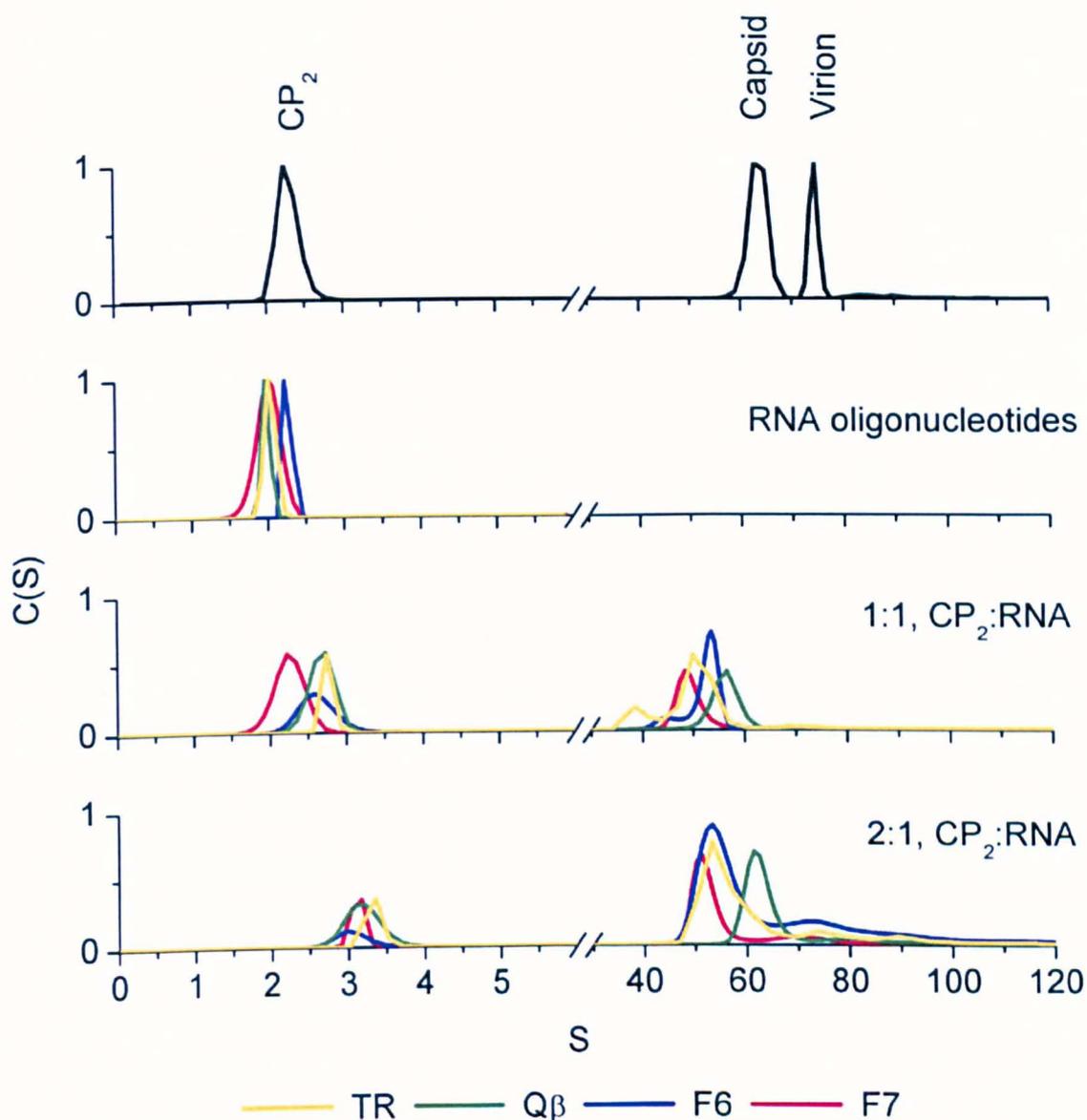
	$S_{exp}$	$S_{20w}$	$f/f_0$	$M_{exp}$ (kDa)	$M_{calc}$ (kDa)	$M_{exp}/M_{calc}$	$R_s$ (nm)
TR	$2.04 \pm 0.09$	2.05	1.1	5.90	6.05	0.8	1.21
F6	$2.26 \pm 0.17$	2.28	1.0	5.90	4.45	32.7	1.08
F7	$2.03 \pm 0.16$	2.07	1.0	4.96	4.48	10.6	1.01
Qβ	$1.99 \pm 0.07$	2.00	1.2	6.30	6.37	1.1	1.32
Capsid	$65.4 \pm 2.0$	65.8	1.2	3237	-	-	11.74
CP <sub>2</sub>	$2.3 \pm 0.17$	2.32	1.4	27.6	27.45	0.52	2.84

MS2 capsid assembly was performed with the RNA stem loops in a similar manner as described for the genomic RNAs. A notable difference however was that the samples were incubated at 4° C in order to emulate similar reactions assayed with mass spectrometry and GFLS. In addition, samples were spun subsequently at 17000 rpm and 40000 rpm in order to characterise the slow sedimenting RNA oligonucleotide species which sedimented too slowly at 17000 rpm for analysis. An intermediate rotor velocity could not be used because the assembled capsids sedimented too fast at rotor speeds above 17000 resulting in too few sedimentation boundaries for SV analysis. The sedimentation profiles from the two runs were fitted independently to the continuous sedimentation coefficient distribution model and then normalised with respect the amount of sedimenting material at the two different rotor speeds. **Figure 4-11** shows

the combined sedimentation coefficient distribution plots of capsid assembly reactions with the four RNA stem loops. For S-values below six, results from SV at 40000 rpm are shown. For S-values above thirty, results from SV at 17000 rpm are shown. No species sedimenting between 5-30 S were detected in a reproducible manner. The sedimentation coefficients of identified components are given in **Table 4.4**. The amounts of capsid-like products formed with the RNA stem loops are given in **Figure 4-12**.

The SV assays of capsid assembly show that at a stoichiometric equivalent of CP<sub>2</sub> to RNA, two dominating species are detected in all assembly reactions which sediment faster than both the corresponding RNA stem loop and the free CP<sub>2</sub>. The slower sedimenting species observed at 2.3-2.9 S, most likely correspond to ribonucleoprotein complexes similar to the CP<sub>2</sub>-TR complex which has been detected by filter binding assays (Beckett and Uhlenbeck, 1987) and non-covalent mass spectrometry (MS) (Stockley *et al.*, 2007). CP<sub>2</sub>-RNA complexes have similarly been detected by MS in assembly reactions initiated with the non-cognate stem loops (Basnak *et al.*, 2009). Upon addition of a second stoichiometric equivalent of CP<sub>2</sub> these species sediment faster at 3.0-2.6 S suggesting that the composition of the sedimenting ensemble component has changed. This is consistent with the idea that even when the amount of RNA that is encapsidated is taken into account (**Figure 4-12**), CP<sub>2</sub> is saturated with RNA at a 1:1 reaction stoichiometry while the reverse is true at a 2:1 reaction stoichiometry. The different sedimentation rates of these species at the two reaction ratios reflect this.

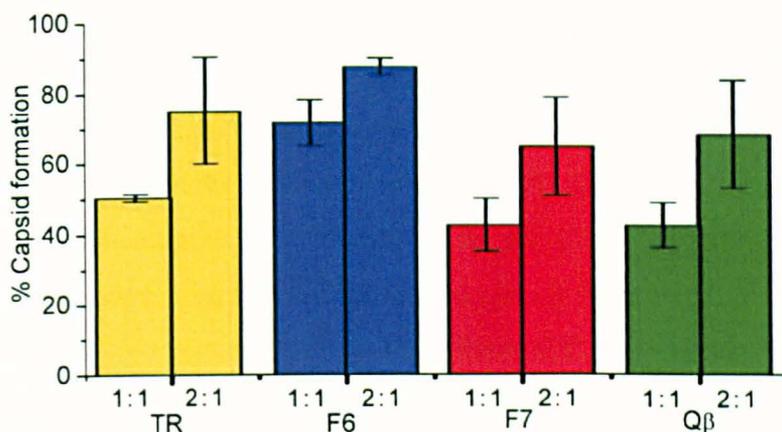
The similar sedimentation rates of the slower sedimenting complex formed during assembly with the stem loops suggests they promote assembly in a similar manner to TR. This was confirmed by mass spectrometry. The non-cognate stem loops were observed to promote formation of assembly competent capsid assembly intermediates in a similar manner as reported for TR (Basnak *et al.*, 2009).



**Figure 4-11.** Normalised sedimentation coefficient distributions of MS2 capsid assembly reactions initiated with TR, F6, F7 and Q $\beta$ . Assembly reactions were performed in assembly buffer and incubated for 4h at 4°C prior to analysis. The sedimentation coefficient distributions of CP<sub>2</sub>, capsid and virion and the RNA stem loops are shown for comparison on the two upper c(s) vs. S plots. Upon addition of CP<sub>2</sub>, the species corresponding to RNA sediment faster and have sedimentation coefficients of 3.0-3.6 at a 2:1 reaction stoichiometry consistent with the formation of a CP<sub>2</sub>-RNA complex. Products with sedimentation coefficients similar to recombinant capsids are detected at 50-63 S. More capsids are formed at the reaction ratio of 2:1.

**Table 4-4.** Sedimentation coefficients of identified components in the assembly reactions induced with RNA stem loops. The table shows the S-values of the slow sedimenting species (CP<sub>2</sub>-RNA) and the capsid at 1:1 and 2:1 stoichiometric ratios of CP<sub>2</sub>:RNA.

Component	TR		Q $\beta$		F6		F7	
	1:1	2:1	1:1	2:1	1:1	2:1	1:1	2:1
CP <sub>2</sub> -RNA	2.8±0.2	3.6±0.1	2.7±0.3	3.2±0.5	2.6±0.3	3.0±0.4	2.3±0.1	3.1±0.1
Capsid	53.0±4.0	55.8±4.6	56.5±1.7	62.9±2.8	53.3±4.9	55.4±3.9	49.9±1.8	52.4±3.8



**Figure 4-12. SV quantification of the capsids produced with the RNA stem loops.** The bar chart shows that all the oligonucleotides promote capsid assembly to a similar degree over the 4 h incubation period at the same reaction ratios. The error bars represent standard deviation from average of at least two assembly reactions.

A second, faster sedimenting species was also detected in all the assembly reactions. These species sedimented at rates between 49.9-62.9 S dependent upon the RNA oligonucleotide and the CP<sub>2</sub>:RNA reaction ratio. These S-values are similar to that of the recombinant capsid (65.4 S) and correspond well with the sedimentation coefficient of 53 S obtained by Beckett and Uhlenbeck for capsids, reconstituted with TR (Beckett and Uhlenbeck, 1987). TEM micrographs of the larger assembly component isolated by GFLS further suggest that these faster sedimenting species correspond to capsids (Basnak *et al.*, 2009).

**Figure 4-12** shows the amount of capsids formed in the assembly reactions. Strikingly, these data show that the RNA stem loops all promote capsid assembly with similar efficiencies and that there are few, if any, differences between the four stem loops, even at a 1:1 reaction stoichiometry, after incubation for 4 h. For TR this was unexpected because the results suggest that the kinetic trapping effect which occurs upon binding of CP<sub>2</sub> to TR and inhibits assembly of capsids at a 1:1 stoichiometry due to the lack of C/C like dimers in solution, is not observed after an incubation period of 4 h. After 4h it appears that the kinetic trapping effect of the A/B dimer like CP<sub>2</sub>-TR complex has been relieved.

The implications of these finding for capsid assembly become clear when the kinetic effects of these RNAs are investigated. Interestingly, gel filtration coupled light scattering assays of assembly reactions under similar conditions showed that the three

non-cognate stem loops are all better initiators of capsid assembly than TR. Many more capsids are observed to form at early time points during assembly with F6, F7 and Q $\beta$  as compared to TR (Basnak *et al.*, 2009). This suggests that the kinetic trapping effect, observed at early time points during assembly with TR, is not emulated with the non-cognate stem loops and that the efficiency in which RNA stem loops in general promote capsid formation is modulated by their affinity for the CP<sub>2</sub>. If binding of all the non-cognate RNA stem loops causes a conformational change from a C/C like to an A/B like dimer then lower affinity binding will result in more C/C type dimers being present in solution due to faster dissociation of the CP<sub>2</sub>-RNA complex. Because capsid assembly must proceed in the presence of both types of quasi equivalent dimer, this explains why the non-cognate stem loops are better initiators of assembly than TR. The higher affinity of TR as compared to the stem loops results in a lower C/C dimer concentration because more dimers will be saturated with TR resulting in the A/B dimer conformation dominating in solution (Basnak *et al.*, 2009). With respect to genomic length RNA, these results are consistent with the idea that multiple stem loops can trigger the assembly reaction, even RNAs that have very low affinity for the CP<sub>2</sub>.

## 4.3 Discussion

The manner in which ssRNA viruses, enclose their genome within icosahedral capsids is not understood in detail. The results reported in this chapter give valuable insight into the dynamic behaviour of the RNA genome during ssRNA virus capsid assembly. Monitoring MS2 capsid assembly with analytical ultracentrifugation provided good resolution of assembly reaction components which allowed to assess the effect of RNA size and sequence on capsid assembly. These experiments highlighted important aspects of the MS2 capsid assembly mechanism. Firstly, MS2 genomic RNA must be folded into the volume of the capsid during assembly. Secondly, various stem loop structures are capable of controlling quasi-equivalent CP<sub>2</sub> conformations. Together these results suggest that quasi-equivalent conformer switching occurs throughout/during assembly. The results highlight the already multifunctional roles of viral RNA. Its function is not restricted to encoding protein products, serving as a template for its own replication and to “isolated” coat protein recognition events. The genome functions as a whole, as a dynamic structural scaffold upon which capsid assembly takes place. Beyond MS2, these results offer new insight into the roles of RNA in ssRNA virus assembly and suggest a possible route to novel therapeutics capable of binding but misfolding RNA.

### 4.3.1 MS2 RNA has an active role during MS2 capsid assembly

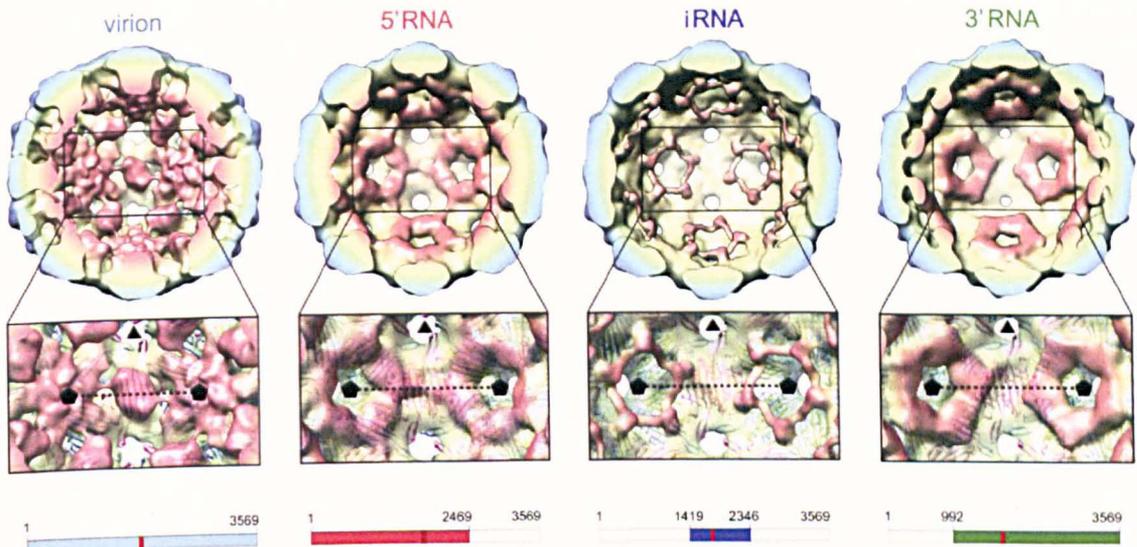
The GEMSA assays of capsid assembly reported in chapter 3 had suggested that assembly induced with genomic MS2 RNA was co-operative and that there was a size constraint on the packaging efficiency of genomic RNA. These effects were confirmed by analysing capsid assembly by analytical ultracentrifugation. The size constraint implies that it is more difficult to package larger RNAs into the volume of the capsid than it is smaller RNAs. The different hydrodynamic radii of the MS2 genomic RNAs and the observation that they are all larger than the MS2 capsid offer clues as to why this is the case. The hydrodynamic radii of the sub-genomic RNAs in assembly buffer indicate that a conformational change is associated with packaging the genomic RNAs. The larger hydrodynamic radii of the MS2 genome and the MS2 sub-genomic RNAs as compared to the MS2 capsid suggest that the CP<sub>2</sub> induce this conformational change. Coat protein dimers therefore have RNA chaperone activity (Weeks *et al.*, 1997) and

fold the RNA into the volume of the capsid. The RNA packaging process is distinct from RNA folding which is predominantly dependent upon nucleotide base-pairing to stabilize RNA tertiary structure (Tinoco and Bustamante, 1999; Scherbakova *et al.*, 2008). Folding of the genome is further suggested by the findings that capsid assembly becomes less efficient in the presence of EtBr. This indicates that EtBr intercalation inhibits the RNAs ability to adopt a conformation compatible with the volume of the capsid interior. This idea implies that the length dependence of RNA packaging is related to the higher energy barrier associated with packaging larger RNAs due to the need to bring more negatively charged phosphodiester groups into close proximity and/or the greater variety of conformational space which needs to be sampled by larger RNAs in order to fit inside the capsid.

Interpretation of the length dependence of packaging efficiency can however also be dependent upon whether the RNAs are all packaged with the same or with different folds. Cryo-EM reconstructions of the capsid assembly products induced with the sub-genomic MS2 RNAs suggest that their tertiary structure within reconstituted capsids is similar. This implies that the sub-genomic RNAs adhere to a tertiary structural template within the capsids which, interestingly, is observed to be defined by interactions with the CP<sub>2</sub> at specific positions in the capsid shell (**Figure 4-13**). The ordered electron density which is not due to the protein capsid is observed surrounding the capsids icosahedral 5-fold axes in all the assembled capsids (Toropova, 2009; Rolfsson and Toropova, 2009). This is consistent with RNA induced quasi-equivalent conformer switching occurring at the A/B dimer positions in the capsid (Stockley *et al.*, 2007; Toropova *et al.*, 2008) and the observations reported in this chapter that variations of stem loops can trigger this effect.

The sizes of the sub-genomic MS2 RNAs were initially selected as they, in principle, corresponded to the size of genomic RNA seen in two shells of RNA density in the structure model of the MS2 virion (**Figure 1-15**). All the sub-genomic structure models showed little density attributed to RNA which was not in close contact with the capsid interior i.e. no inner RNA shell was visible. Possible reasons for this are that the sub-genomic RNAs are packaged with different RNA structures from that of the virion. A cryo-EM reconstruction of capsids reconstituted with vRNA however showed that the inner shell is reformed suggesting that the tertiary structural arrangement of RNA within *in vitro* assembled capsids and MS2 virions is similar. Perhaps the lack of an inner shell with the sub-genomic RNAs therefore represents a greater disorder of the RNA

segments which take part in forming the inner shell (Toropova, 2009; Rolfsson and Toropova, 2009). Given a requirement for quasi-equivalent conformer switching, a large proportion of the sub-genomic RNAs would however be expected to be in contact with the capsid as roughly 2/3 of the RNA in the MS2 virion are located in the outer shell (Toropova *et al.*, 2008).



**Figure 4-13.** *The MS2 sub-genomic RNAs interact with the capsid interior in a manner which is consistent with quasi-equivalent conformer switching. Cryo-EM reconstructions of MS2 capsid assembly products at 16-18 Å resolution as viewed down the 2-fold axes of icosahedral symmetry. The rear half of each structure model is shown. The MS2 virion is shown for comparison on the left, with the inner shell of RNA density removed. All the capsids are similar in the protein capsid region. Electron density which can be associated with the sub-genomic RNAs is located pre-dominantly beneath A/B type coat protein dimers. This is evident as a ring of density surrounding the icosahedral 5-fold axes. There is less electron density observed beneath the C/C dimer positions which reflects little RNA occupancy at these positions. This suggests that genomic RNA controls CP<sub>2</sub> conformation throughout the capsid assembly process. Cryo-EM reconstructions of the capsid assembly products were obtained by a collaborating PhD student Katerina Toropova. The figure is modified from Rolfsson and Toropova, 2009.*

Together with the SV results, the cryo-EM reconstructions suggest that both the RNA and the CP<sub>2</sub> act as chaperones of each others conformation! During assembly the RNA is therefore folded by the CP<sub>2</sub> while at the same time the genomic RNA controls the quasi-equivalent conformation at A/B dimer positions. These results strongly suggest a nucleotide template, beyond isolated packaging signal recognition, that could help promote efficient capsid assembly. It therefore appears that global RNA structure has an active role in capsid assembly. The precise positioning of RNA stem loops

within the genome could act to facilitate capsid formation by bringing CP<sub>2</sub> with the correct conformations into close proximity with one another throughout the assembly process.

Support for this idea has been obtained by monitoring MS2 capsid assembly induced with oligonucleotide variants of TR which contain nucleotide extensions in either the 5' or 3' direction. GFLS and MS assays show that increasing the length of TR in either the 5' or 3' direction to incorporate an adjacent stem loop, positively affects the rate of capsid formation (see **Figure 1-11 B** for nucleotide structure adjacent to TR operator). If however the nucleotide extensions are non-cognate, capsid assembly becomes inefficient and aggregate formation dominates (Basnak *et al.*, 2009). Furthermore, in Basnak *et al.* (2009) an oligonucleotide variant of TR with an extension in the 5' direction was shown to differentiate between capsid assembly mechanisms, capsid assembly occurring preferentially through the 3-fold hexamer of dimers rather than the 5-fold pentamer of dimers (**Figure 1-14**). These results confirm the scaffolding nature of genomic RNA and suggest that genomic RNA can direct capsid assembly.

#### 4.3.2 Assembly can be inhibited

The experiment with EtBr demonstrates that reagents that alter the stiffness or structure of viral genomes reduce RNA packaging efficiency. This is an important result. The implication is that small molecular inhibitors, such as complementary oligonucleotides or aminoglycoside derivatives which are capable of binding and inhibiting correct folding of genomic RNA, could act as novel anti viral therapeutics (Turner *et al.*, 2006; Ironmonger *et al.*, 2007; Rolfsson *et al.*, 2008). These could represent potential clinical tools since incompletely assembled virions could be immunogenic but non-infectious. Similarly, RNA mimics that encompass viral coat protein binding sites could be effective anti-viral agents because they would either trap coat proteins in unproductive aggregates or redirect the normal assembly pathway. It appears that viral RNA and its protective coat proteins have to make very many precise interactions with each other, and if such contacts occur improperly anywhere during the assembly process the formation of infectious progeny virus could be impaired. Few current therapeutics target viral capsid self-assembly pathways while none target the viral RNA directly. Reagents with such potential could prove successful at inhibiting viral assembly pathways.

The results discussed here suggest that there is a network of RNA stem loop structures capable of quasi equivalent conformer switching at defined positions within the MS2 genome. The requirement for quasi-equivalent conformer switching at precise positions within the RNA genome must constrain the assembly process in such a way that only RNAs with solution structures which have a defined spacing of stem loops which allow correct conformer switching are packaged efficiently. The observed length dependence on RNA packaging efficiency could therefore be dependent on the structure of the MS2 RNAs prior to assembly. Knowledge of the exact binding positions of the CP<sub>2</sub> to the RNA genome could provide further understanding of how the genome contributes to capsid assembly and would be beneficial for understanding the topology of the RNA fold within the virion. In the next chapter, experiments are reported where the solution structures of the sub-genomic RNAs were compared to the structure of the MS2 genome inside the virion in order to address these ideas.

## 5 Comparison of MS2 RNA structure *in virion* vs. *in vitro*

---

### 5.1 Introduction

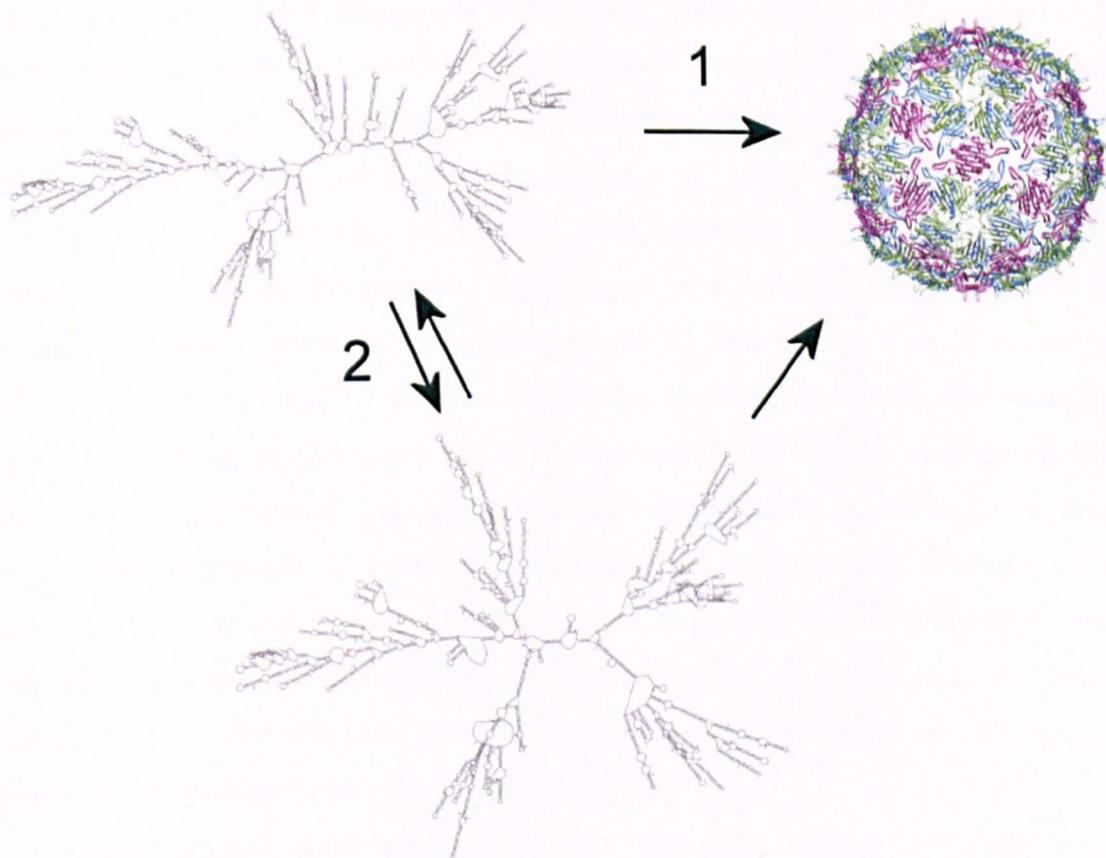
The results reported in chapter 4 along with the cryo-EM reconstruction of assembly products (Toropova, 2009) suggest that genomic RNA folds into a complex structure within the capsid which is defined by RNA-coat protein interactions at the A/B dimer positions within the capsid. This indicates that capsid assembly is a mutually induced process where RNA acts to control coat protein conformers while the coat protein dimers act as chaperones that fold the RNA into a structure capable of fitting inside the  $T = 3$  capsid.

This type of assembly model raises two interesting series of questions. Firstly, does the observed RNA length-dependence on packaging efficiency depend upon how easily the sub-genomic RNAs adopt an assembly competent tertiary structure? Their ability to attain such a structure would in turn be dependent upon the type of structural rearrangement required for efficient packaging (**Figure 5-1**). In other words, do the sub-genomic RNAs share a similar secondary structure in solution and is this structure conserved within the virion?

Secondly, the proposed assembly mechanism implies that there are multiple stem loop structures within the MS2 genome capable of inducing a conformational change from a C/C to an A/B type dimer both of which comprise the MS2 capsid. **Figure 5-1** gives an example of two minimum energy structure predictions of the MS2 genome. Although mfold is not accurate at predicting long RNA structures (Zuker *et al.*, 2003; Deigan *et al.*, 2009), it is obvious from these predictions as well as from the secondary structure proposed in Fiers *et al.* (1976) (**Figure 1-8**) that there are multiple secondary structure elements within the genome that could induce allosteric conformer switching. The ability to map coat protein dimer binding positions within the RNA genome would be beneficial for understanding the RNA fold within the capsid.

This chapter describes results wherein the susceptibility of RNA to lead ion induced hydrolysis was used to compare the structures of the three MS2 sub-genomic RNAs *in vitro* to that of the virion RNA residing within the MS2 phage (*in virion*). In section 5.2.1, a 740 nt genomic region, shared within all the RNAs, is compared based on lead ion induced cleavage positions within each RNA, as assayed with polyacrylamide

sequencing gels. In section 5.2.2, results from similar lead ion structure probing experiments are reported as assayed using high-throughput capillary electrophoresis. These results are then discussed with respect to the solution and virion RNA structure and its implications for capsid assembly. Additionally the ability of these assays to map coat protein dimer binding sites within the RNA genome are explored.



**Figure 5-1. How does RNA folding affect RNA structure?** Folding of the MS2 genome implies a change in RNA tertiary structure during RNA packaging. Depending on whether or not the RNA structure is conserved in the virions, the packaging could also induce a change in RNA secondary structure. This would depend on whether the genome can be directly folded from an RNA secondary structure in solution that is conserved within the capsid (1) or if structural reorganisation of the RNA genome is a prerequisite for RNA encapsidation (2). A structural comparison of genomic RNA in vitro and genomic RNA in virion would indicate whether RNA packaging is accompanied by RNA secondary structure reorganisation and thereby provide insight into whether the genome structure is conserved in virion. The RNA structures shown here are minimum energy structure predictions of the MS2 genome from the program *mfold* and have a free energy difference of 7 kcal/mol. They are depicted to highlight the type of structural rearrangement that might occur during RNA packaging. *Mfold* predicts a further 9 structures which differ in minimum free energy of 7 kcal/mol or less.

### 5.1.1 RNA structure probing and lead ion hydrolysis

RNA structure can be investigated by using a plethora of enzymatic and chemical probes, each capable of cleavage or modification of RNA in a manner which is often sequence and/or structure specific. The reactivity of nucleotides within an RNA molecule towards these reagents reflects their immediate structural environment. This is because nucleotide reactivity is governed by steric factors conferred by RNA structure or a bound RNA ligand, which ultimately affects reaction rates. The mapping of nucleotide positions within an RNA molecule with respect to their reactivity towards probes with distinct RNA cleavage or modification specificities therefore provides information on RNA structure (reviewed in Moine *et al.*, 1998).

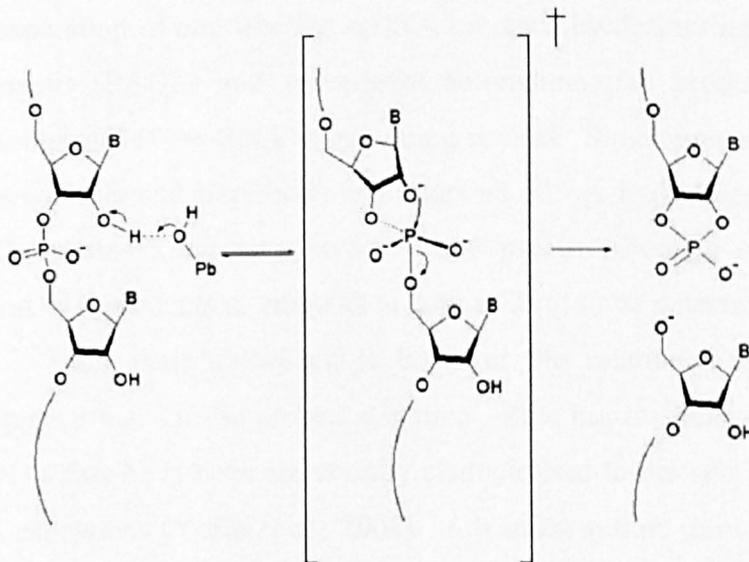
The best known method of probing RNA structure is by sequence specific ribonuclease digestion. For example, ribonuclease T1 cleaves RNA specifically at the 3' side of unpaired guanines while ribonuclease V1 cleavage is specific for helical RNA. The parallel use of these ribonucleases thereby allows the discrimination between nucleotide environments and allows the elaboration of RNA structure models. The MS2 genome structure was originally characterised *in vitro* by chemical analysis of oligonucleotide products obtained by partial digest of the RNA with ribonuclease T1 (**Figure 1-8**) (Fiers *et al.*, 1976). Obviously structure probing of viral genomes *in virion* with ribonucleases is not possible as capsids render the genome nuclease resistant. Capsids such as MS2 with large pores at the 5- and 3-fold icosahedral axes are however permeable to small molecules (Valegård *et al.*, 1994).

A range of small organic compounds have been identified that selectively modify nucleotides. Their mechanism of action often relies on selective chemical modification of purines and pyrimidines which inhibit nucleotide base pairing and allow their position within an RNA molecule to be determined using reverse transcription (see below). The central region of the MS2 genome has been structurally probed *in vitro* using a combination of chemical and ribonuclease probes (Skripkin *et al.*, 1990) which afforded results similar to those obtained by Fiers (1976). In addition to small organic compounds, RNA structure can also be investigated using divalent metal ions.

RNA is hydrolysed upon treatment with divalent metal ions. The best studied examples are lead ions ( $\text{Pb}^{2+}$ ) which were first observed to depolymerise RNA sequence specifically and in a concentration dependent manner by Farkas (1968). Lead ions were later shown to bind and cleave tRNA<sup>Phe</sup> in a structure dependent manner (Werner *et al.*,

1976; Brown *et al.*, 1983). Since then they have been utilised in structure probing of various RNAs and found use in mapping divalent metal ion binding sites (reviewed in Kirsebom and Ciesiolka, 2005).

Lead ions cleave RNA in two ways that are distinguished by their efficiency of RNA hydrolysis, although the mechanism of cleavage of the phosphodiester bond is essentially thought to be the same (**Figure 5-2**). In the first case, phosphodiester bonds are cleaved efficiently and with high specificity at low lead ion concentrations (<100  $\mu\text{M}$ ). These cleavages are associated with tight metal ion binding sites that co-ordinate the lead hydrate allowing optimal alignment for 2'OH deprotonation. This has been observed in the site specific cleavage of tRNA<sup>Phe</sup> (Brown *et al.*, 1983) and 16 S and 23 S ribosomal RNAs (Winter *et al.*, 1997). High efficiency cleavage is however quite rare with most cleavages occurring at higher concentrations of lead (> 0.5 mM) and usually occurs at several consecutive phosphodiester bonds (Kirsebom and Ciesiolka, 2005).



**Figure 5-2.** *The proposed mechanism of lead ion hydrolysis of RNA. From left, lead ions form lead ion hydrates in aqueous solution. Lead ion hydrate acts as a base and deprotonates the 2'OH group of a ribose. Subsequent nucleophilic attack by the deprotonated oxygen on the phosphorus results in cleavage of the phosphodiester bond through formation of a penta-coordinated transition state, shown here in brackets. Cleavage efficiency is thought to depend on the proper orientation of the lead ion hydrate for deprotonation of the 2'OH moiety and the conformational flexibility of the phosphate backbone to allow formation of the penta-coordinated transition state. The mechanism was proposed by Brown *et al.* (1983).*

Because the lead hydrate reacts equally well with all accessible 2'OH groups on RNA, a difference in cleavage specificity of the RNA molecule is an indicator of the immediate structural environment of the cleaved bond. Lead ion cleavage at higher

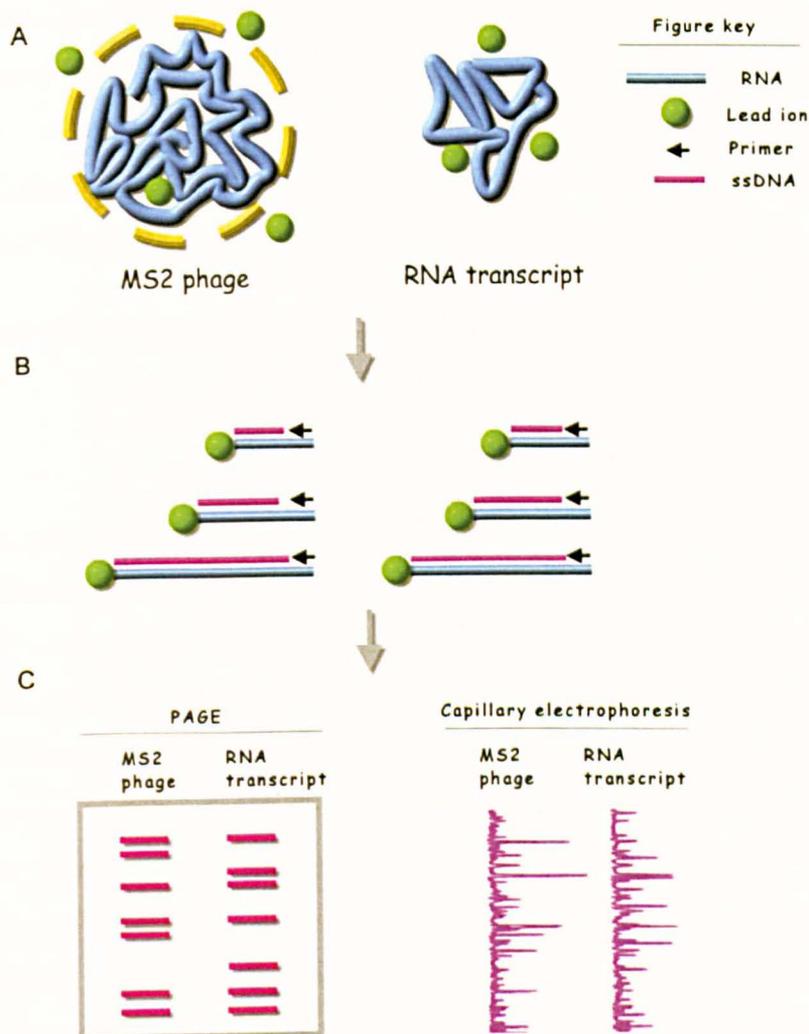
concentrations is dependent on the conformational flexibility of the RNA backbone. Cleavages occur mainly in single stranded regions excluding regions involved in stacking or other higher order interactions (Ciesiolka *et al.*, 1998). In comparison to small molecules that specifically modify RNA, lead hydrolysis is less sequence specific. Its main advantage is that it allows the detection of subtle conformational changes which occur upon RNA structure rearrangement or ligand binding (Huntzinger *et al.*, 2005). It follows that lead ions are well suited for structural comparison of the MS2 sub-genomic RNAs, their small size enabling them to penetrate MS2 capsids allowing further comparison to the RNA genome *in virion*.

RNA structure probing reactions must be assayed in a manner which allows identification of nucleotide cleavage positions within the RNA. This can be achieved by primer extension of cleaved or modified RNA products. Primer extension analysis relies on reverse transcription of modified unlabelled RNA using a radio-labelled DNA primer. The separation of end-labelled ssDNA products by denaturing polyacrylamide gel electrophoresis (PAGE) and subsequent autoradiography produces a structural snapshot or “footprint” of the RNA region being probed. Simultaneous electrophoresis of appropriate controls and nucleotide size markers allows hydrolysis positions to be identified. These assays are very sensitive and precise allowing single nucleotide resolution of ssDNA products at amounts as low as 2 pg to be detected (Boorstein and Craig, 1989). Their main drawback is however the relatively small window of nucleotide sequence that can be probed at a time. This has in some ways constrained the size of RNAs that have been structurally characterised to the sub-kilo base region, although with exceptions (Yoffe *et al.*, 2008). Advances in high throughput nucleotide sequencing methods have recently been realised with respect to RNA structure probing (Mitra *et al.*, 2008; Wilkinson *et al.*, 2008) which could bypass the necessity to analyse RNA structure by PAGE. The use of capillary electrophoresis (CE) to analyse RNA structure probing reactions has promised to allow RNA structure to be explored in a wider structural context than has been possible to date (Watts *et al.*, 2009).

### ***5.1.2 MS2 RNA structure probing strategy***

In order to investigate how the conformation of RNA is altered during capsid assembly, it was decided to assay the sub-genomic RNAs, whose production was described in chapter 3 and the RNA genome within the MS2 virion with respect to their

susceptibility to lead ion induced hydrolysis. Because lead ions also allow probing of RNA-ligand interactions it was anticipated that CP<sub>2</sub> binding sites could be mapped using this strategy. The outline of the experimental assay is shown in **Figure 5-3**.



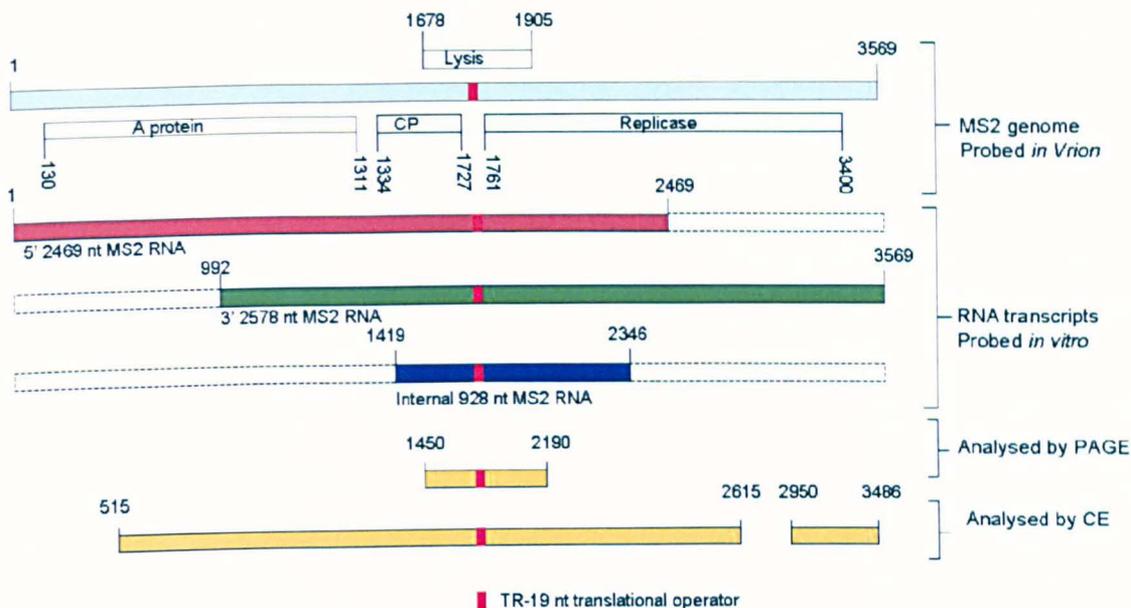
**Figure 5-3. A schematic of the lead ion cleavage assay.** The MS2 capsid is permeable to small molecules such as lead ions allowing structural comparison of RNA in vitro and in virion. **A)** MS2 phage and an MS2 sub-genomic RNA are treated individually with lead ions produced by the solvation of lead acetate. This hydrolyses the RNAs at predominantly single stranded positions producing a pool of various sized RNA fragments. **B)** The pool of RNA is proportionately post-labelled by reverse transcription using a genome specific radio-labelled primer. This generates ssDNA of sizes that are defined by the primer annealing position within the RNA genome and the lead hydrolysis position. **C)** Size fractionation of the ssDNA by PAGE provides a banding pattern or a structural “footprint” of each RNA. Comparison of banding patterns allows regions of similar structure or regions involved in coat protein dimer interactions to be identified. Size fractionation of ssDNA can also be performed by capillary electrophoreses which produces a electropherogram or trace. Comparison of traces of different RNAs similarly allows identification of identical hydrolysis positions.

## 5.2 Results

The structural comparison of the MS2 sub-genomic RNAs and the virion RNA inside the MS2 capsid was performed by assaying their susceptibility to lead ion induced hydrolysis. The lead ion cleavage reactions were assayed using PAGE and CE. Although these assaying methods produced similar results, it proved difficult to combine the two data sets due to problems relating to nucleotide resolution of the assays coupled with the consecutive manner in which lead ions cleave RNA. These assays are therefore reported separately and finally compared in section 5.2.2.5.

### 5.2.1 Lead-induced hydrolysis of MS2 RNA assayed by PAGE

The 5'-, 3'-, and iRNA sub-genomic fragments were treated independently with lead acetate in a time dependent manner at room temperature. Following incubation with lead acetate, reactions were quenched with EDTA and the RNAs precipitated at high salt concentration. The MS2 phage was treated similarly with the addition of a phenol chloroform extraction step, following treatment with EDTA, to remove CP<sub>2</sub>. The RNA samples were then used as templates in reverse transcription reactions (a detailed description of the experiments is given in section 2.2.13). Primer extension was performed using six primers that had been radiolabelled at their 5' termini with <sup>32</sup>P. The primers were spaced at roughly 130 nt intervals within the MS2 genome which resulted in partial overlap of the primer extension products when resolved on 6-8 % polyacrylamide sequencing gels. Subsequent autoradiography of the gels allowed the structure of the MS2 RNAs to be compared over an approximately 740 nt region of the MS2 genome (**Figure 5-4**). This region of the genome was specifically chosen as it represents a region which is present in all the MS2 sub-genomic RNAs and incorporates the TR operator. In addition, this region of the genome has been exhaustively probed by van Duin and co-workers (Skripkin *et al.*, 1990) which along with the structure model proposed by Fiers (1976) allowed the results to be compared to previous structure proposals.



**Figure 5-4. Structurally probed regions of the MS2 virion RNA and the sub-genomic RNA fragments.** The regions analysed by PAGE and CE, are shown in yellow aligned to the MS2 genome and the MS2 sub-genomic RNA fragments. The structural data from PAGE and CE assays covers 74 % of the MS2 genome.

All the MS2 sub-genomic RNAs and the MS2 phage were probed simultaneously and assayed with a single primer in one go in order to produce results that could be directly compared in terms of lead induced cleavage patterns. **Figure 5-5** shows the results of a lead induced cleavage assay with the primer 1812.R. Lead induced hydrolysis positions can be identified as black signals, corresponding to ssDNA products of defined lengths, whose intensity increases with increased lead ion incubation time. The absence of a signal in the control lane is a further indicator of lead ion induced hydrolysis. In **Figure 5-5 B**, two signals are readily observed for the 5' sub-genomic RNA corresponding to  $Pb^{2+}$  cleavage sites. One occurs within the 3' region of the TR operator and another adjacent to the 5' end of the TR operator and are indicated with black arrows. Various non-randomly distributed  $Pb^{2+}$  cleavages of lesser intensities are also observed in the TR region. Signals are also observed where the intensity is similar in all the lanes including the control as indicated by a red arrow. These sites most likely represent reverse transcriptase stop positions where the reverse transcriptase has fallen off the RNA transcript due to underlying secondary structure. The MS2 genome is highly structured which was reflected in the PAGE assays by numerous transcription stop sites.

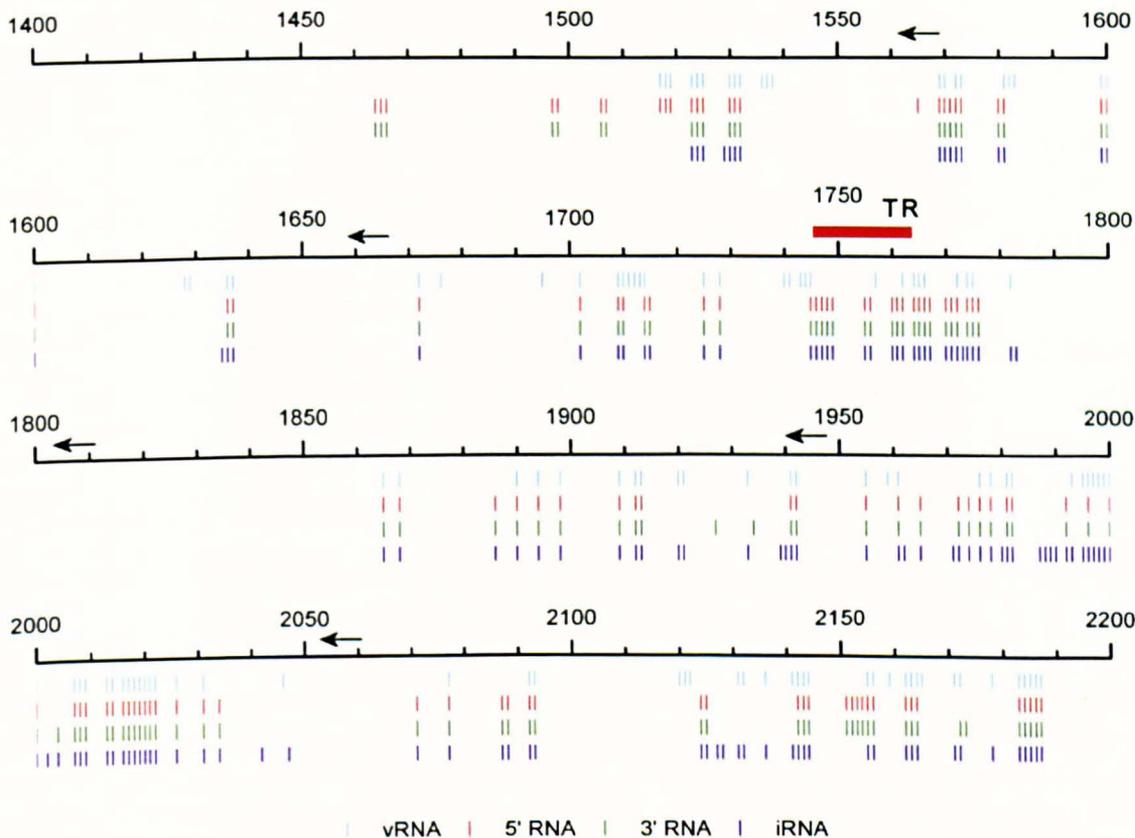


**Figure 5-5. Lead ion probing of MS2 virion RNA and the sub-genomic RNAs.** *A* shows an autoradiograph of a 6 % PAGE assay of primer extension products obtained by reverse transcription of virion RNA, 5' RNA, 3' RNA and iRNA treated with  $Pb^{2+}$ . A 30 nt region covering the TR operator (red) is emphasized in **B**. Samples were treated with 0.4 mM  $Pb^{2+}$  in assembly buffer for 5, 10, 30 and 60 minutes as indicated by a black triangle. A no  $Pb^{2+}$  control reaction is indicated with a  $\div$ . The red and black arrows indicate examples of transcription termination and  $Pb^{2+}$  induced cleavage, respectively. Dideoxy sequencing ladders (G and C), a hydroxyl ladder (OH) and size standard (L) allowed the estimation of lead induced cleavage positions. The autoradiograph represents a structural snapshot of a region that is intrinsic to all the RNAs probed, this is highlighted in **C** which shows the approximate position of the genomic region being probed. Interestingly, the structural footprints of the MS2 sub-genomic RNAs and the MS2 genome in virion are similar.

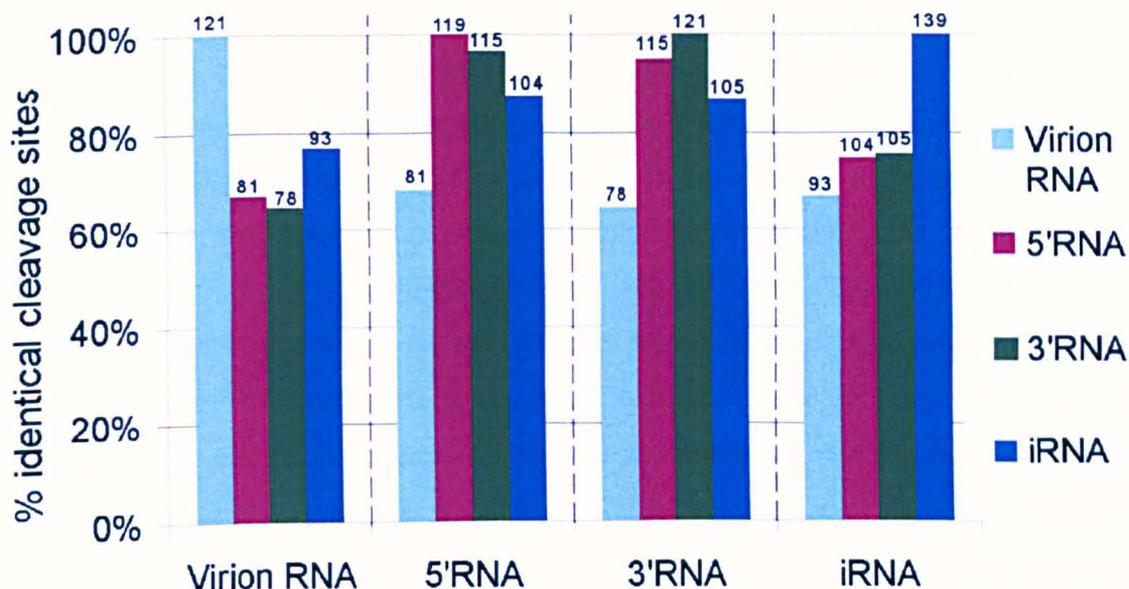
### 5.2.1.1 The solution structure of the MS2 sub-genomic RNAs is conserved

Interestingly, the structural footprints of the sub-genomic MS2 RNAs were similar. The two highlighted  $\text{Pb}^{2+}$  cleavage sites in **Figure 5-5 B** are also clearly detected in the 3' RNA and iRNA. Furthermore, the non-randomly distributed pattern of cleavages observed within the TR operator in the 5' RNA is mirrored in the 3' and iRNA sub-genomic MS2 RNAs. The intensities of the  $\text{Pb}^{2+}$  cleavage positions were however considerably higher in the iRNA. This was because approximately three times the amount of iRNA was probed with respect to the 5' and 3' sub-genomic RNAs in order to keep the stoichiometry of phosphodiester bonds to  $\text{Pb}^{2+}$  the same in all samples. The higher intensities observed for the iRNA therefore do not necessarily indicate a more dynamic structure but rather that there is more material detected in the iRNA primer extension reactions. Examination of **Figure 5-5 A** shows that the structural similarity with respect to lead ion cleavage is not isolated to the TR region. Careful examination of this gel and repeats of this experiment with the 1812.R primer indicated that the structure of the sub-genomic RNAs was fairly similar over the roughly 160 nt region probed. This suggested that the conformation of these RNAs over the probed region in the assembly buffer was similar.

In order to investigate whether the conformation of the sub-genomic RNAs was conserved outside this region, primer extension reactions were performed that incorporated flanking regions that allowed structural comparison of nucleotides 1450-2190 of the MS2 genome (**Figure 5-4**). The identified lead ion cleavage positions within this region are shown in **Figure 5-6**. Perhaps surprisingly, the lead ion cleavage footprint over this region is strikingly similar for the three sub-genomic RNAs. A total of 119, 121 and 139 lead ion induced hydrolysis sites were identified within the 5', 3' and iRNA's, respectively. **Figure 5-7** shows the number  $\text{Pb}^{2+}$  cleavage sites occurring at identical nucleotide positions within the three sub-genomic RNAs. 104  $\text{Pb}^{2+}$  cleavage sites were identified at identical positions within the three sub-genomic fragments. This represents a large percentage of the identified cleavage sites for each RNA transcript. 87 % of the cleavage sites identified in the 5' sub-genomic fragment were also identified in the 3' and iRNA, while this percentage was 86 % and 75 % for the 3' RNA and iRNA respectively. These results suggest that the solution structures of the three sub-genomic RNAs are similar *in vitro*.



**Figure 5-6. Lead ion cleavage map of the 1450-2190 nucleotide region of the MS2 genome.** The map highlights all identified lead ion cleavage positions within the sub-genomic RNAs *in vitro* and the genomic RNA within the MS2 virion. Cleavage positions are indicated with vertical bars that are colour coded according to the RNA and are accurate to  $\pm 1$  nt. The map can be analysed like a PAGE gel if rotated  $90^\circ$ . All cleavage sites are shown. No discrimination was made between strong and weak cleavage positions when the cleavage sites were identified from autoradiographs. The annealing positions of the primers used in primer extension reactions are shown as black arrows, with the exception of the primer 2223.R. Lead ions cleave the sub-genomic RNAs at many identical positions producing very similar cleavage patterns. This suggests that they each exist as an ensemble of structurally similar conformations. A large proportion of the identified cleavage sites are also detected in the vRNA suggesting that MS2 genome structure is partly conserved within the MS2 capsid. The numbers of identical cleavage positions identified with respect to each RNA are shown in Figure 5-7. Figure 5-8 further highlights regions of similar conformational flexibility between all the RNAs probed.



**Figure 5-7.** *MS2 RNA structure is conserved in vitro and in virion.* A large proportion of identified  $Pb^{2+}$  hydrolysis sites within the probed region of the MS2 genome occur at identical nucleotide positions within the RNA transcripts and virion RNA. The bar chart shows the number of identical  $Pb^{2+}$  cleavages as a percentage of the total number of  $Pb^{2+}$  cleavages identified for a particular RNA. For example: 121  $Pb^{2+}$  cleavages were identified in the virion RNA, of those 121 sites, 81, 78 and 93 were also identified in the 5', 3- and iRNAs, respectively. 64-77 % of the  $Pb^{2+}$  cleavages identified in the virion RNA therefore also occur in the RNA transcripts in vitro suggesting that in vitro MS2 RNA structure is partly conserved within the MS2 virion. Shared cleavage sites are even more numerous between the sub-genomic RNAs suggesting a similar secondary structure for these RNAs within the probed region.

#### 5.2.1.2 The MS2 RNA genome structure is partially conserved in virions

Comparison of the lead ion induced hydrolysis positions of the MS2 sub-genomic RNAs and the RNA genome within the virion reveals that these RNAs are structurally similar (Figure 5-6 and 5-7). A total of 121 lead ion induced cleavage sites were identified in the MS2 genome within the virion over the probed region. 78 of these cleavage sites or approximately 64 % were detected in all three sub-genomic MS2 RNAs *in vitro*. This suggests that the conformational flexibility of a large part of the RNA region probed is similar *in vitro* and *in virion* and that RNA structure is partly conserved in the MS2 virion.

The fact that there is a defined cleavage pattern obtained for the MS2 virion RNA suggests that the RNA genome does not exist as an ensemble of conformationally distinct structures within the MS2 virion, rather it has a defined structure. This is in

excellent agreement with the cryo-EM reconstruction of the MS2 phage which indicates that the genome adopts a defined tertiary fold in the capsid (Toropova *et al.*, 2008).

There are also regions where the susceptibility of phosphodiester bonds to hydrolysis is different between the virion RNA and the RNAs in solution. These are observed where lead ions induce cleavage of the phosphate backbone within the sub-genomic RNAs but not within the MS2 virion and vice versa. There are a multitude of cleavage positions detected in the sub-genomic RNAs that are masked within the MS2 virion. It is hard to speculate as to the exact reason for these structural differences as they could be due to differences in underlying secondary structure or due tertiary structure interactions associated with a more compact structure of the RNA within the virion. The high degree of structural similarity of all the assayed RNAs with respect to lead ion induced cleavage however suggests that a large contribution of the differences observed must be due to CP<sub>2</sub> binding.

### 5.2.1.3 CP<sub>2</sub> binding can be mapped by lead ion induced hydrolysis

RNA-protein complexes can be detected based on the change observed in lead ion induced cleavage reactivity over a particular region of RNA in the presence and absence of protein. The change in reactivity is attributed to the shielding effect of RNA bound protein (Kirsebom and Ciesiolka, 2005). The cryo-EM reconstruction of the MS2 virion suggested that the CP<sub>2</sub> interacts at multiple positions with the MS2 genome (Toropova *et al.*, 2008). It was decided to investigate whether the lead ion induced structural footprints of the sub-genomic RNAs *in vitro* and RNA genome in virion could be used to identify CP<sub>2</sub> binding sites within the RNA genome.

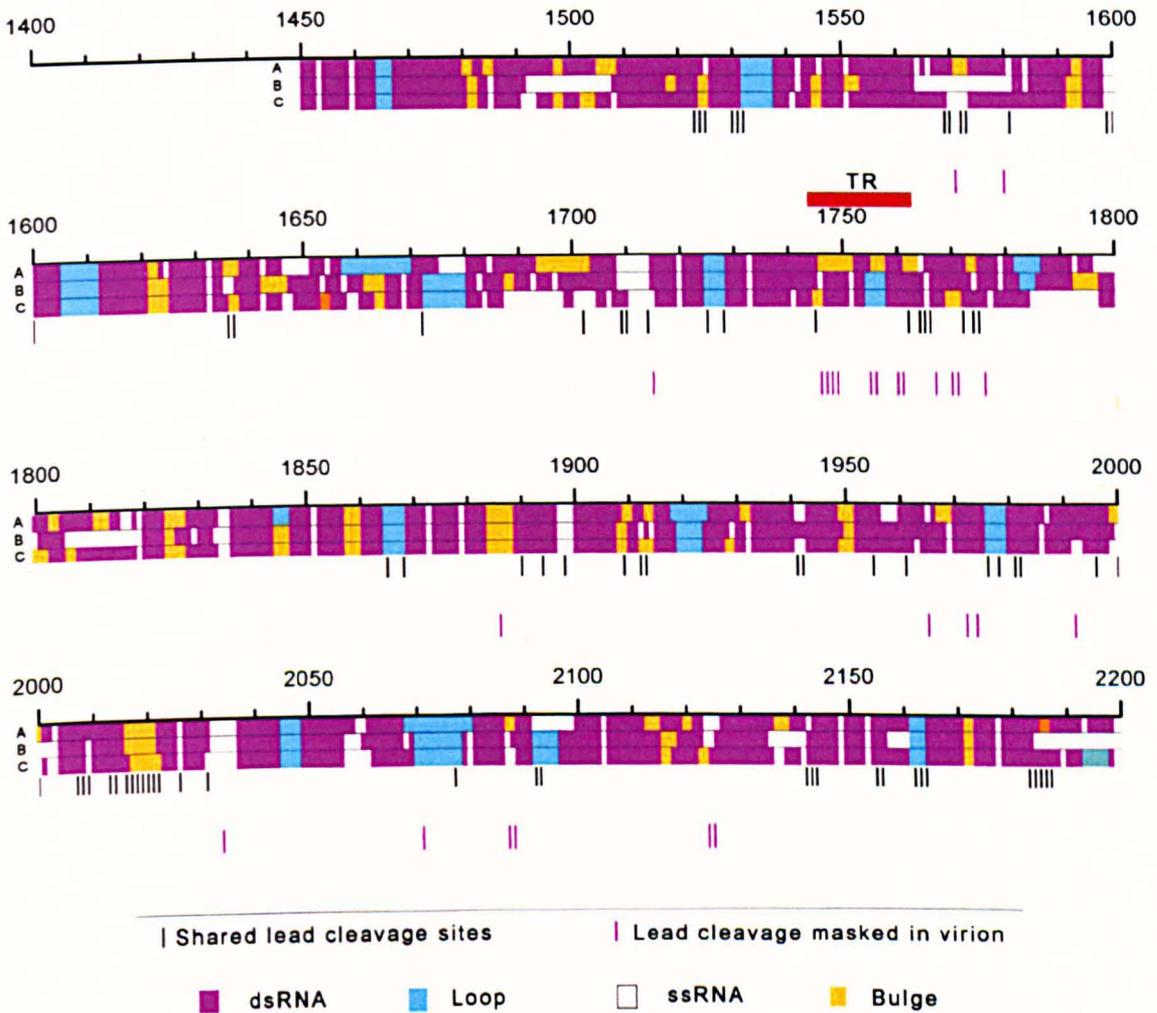
**Figure 5-8** highlights the positions of the 78 identical lead ion induced cleavage positions detected in the RNA genome *in virion* and all the sub-genomic RNAs *in vitro*. 104 identical cleavage positions were observed for the sub-genomic MS2 RNAs alone which indicates that 26 of the cleavage sites that are shared between the sub-genomic RNAs are masked within the virion. These cleavage positions represent RNA regions at which CP<sub>2</sub> binding is most likely to occur. Because coat protein dimer binding has been mapped to the TR operator (Gralla *et al.*, 1974; Carey *et al.*, 1982a) it represented an RNA region where lead cleavage patterns were expected to be different for the RNAs *in vitro* vs. *in virion*. As can be seen in **Figure 5-5** and **Figure 5-8** the cleavage pattern of the TR operator is different *in vitro* and *in virion*. Most notably, the cleavages observed at 1755U and 1756U, which are intrinsic to the TR loop, are not observed in the

genomic RNA within the virion. This suggests that lead ion induced hydrolysis can be used to map CP<sub>2</sub> binding sites within the RNA inside the MS2 virion.

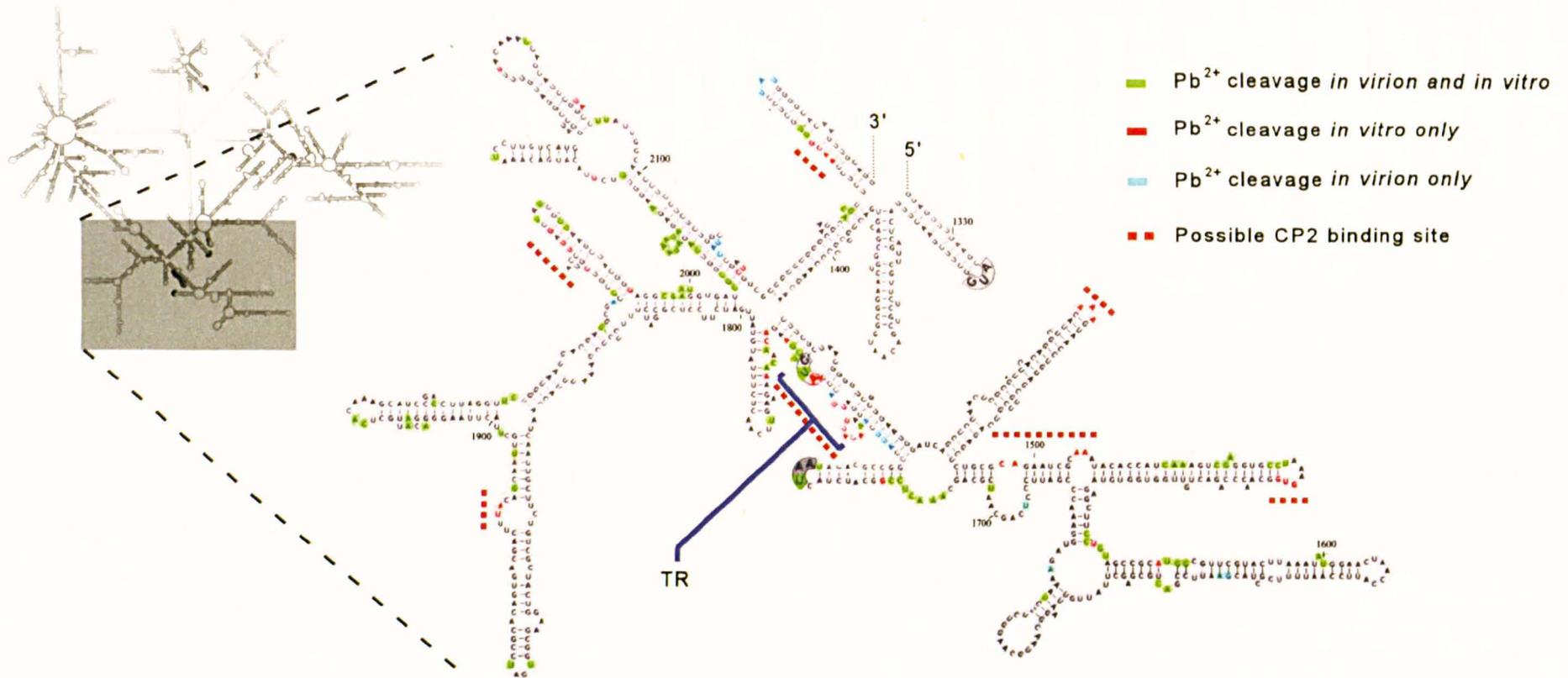
Coat protein dimers have been shown to bind RNA stem loop structures (Beckett and Uhlenbeck, 1988). Lead ions induce cleavage predominantly in single stranded regions such as RNA loops and bulges although hydrolysis is ultimately thought to be dependent upon the flexibility of the RNA backbone. In order to investigate whether the masked Pb<sup>2+</sup> cleavage sites fell within predicted stem-loop structures, which could then be correlated with CP<sub>2</sub> binding, their positions were aligned to three structure predictions of the MS2 genome (Figure 5-8). Figure 5-9 further maps the lead ion cleavage positions identified specifically onto the MS2 secondary structure proposed by Fiers *et al* (1976). The observed results show that the masked Pb<sup>2+</sup> cleavage sites do not necessarily occur within proposed stem-loop regions. This corresponds to results which have shown that the susceptibility of a hairpin to Pb<sup>2+</sup> induced cleavage is dependent upon the loop size and nucleotide composition (Ciesiolka *et al.*, 1998). The intensity of the two observed cleavages detected within the TR hairpin *in vitro* are for example quite weak which perhaps reflects the rigidity of the loop due to base stacking (Valegård *et al.*, 1994). Many of the masked Pb<sup>2+</sup> induced cleavages do however occur within or very close to predicted hairpins or single stranded regions. These cleavage sites could be due to CP<sub>2</sub> binding as it is plausible that an RNA-CP<sub>2</sub> interaction could effect the conformation of the RNA in the vicinity of the binding site. This is observed for binding at the TR operator where the pattern of lead ion cleavage at the 5' side of the TR region *in virion* is clearly different from the cleavage pattern *in vitro* (Figure 5-5, black arrow number 1). It is difficult to conclude that the Pb<sup>2+</sup> induced cleavage sites masked within the virion at other positions than TR correspond to CP<sub>2</sub> binding sites. These sites appear to be non-randomly distributed within the genome as expected, but they are not observed to cluster in the same manner as observed within the TR region. It is highly likely that the masked ion cleavages within the virion also represent regions which are involved in interactions with other regions of the folded genome.

RNA structure proposals of the MS2 genome (Fiers *et al.*, 1976) and mfold minimum energy structure predictions indicate that there are multiple stem loops within the MS2 genome that could act as binding targets for CP<sub>2</sub>, as implied by the cryo-EM model of MS2. Surprisingly, although these predictions differ in structure, a simple count of the number of stem loops predicted for the genome reveals that the number of stem loops corresponds well with the number of A/B dimers within the MS2 capsid (n=

60). The genome structure proposed in Fiers *et al.*, (1976) has exactly 60 stem loops while the ten lowest free energy structure predictions obtained from mfold have an average of  $60.6 \pm 2.7$  stem loops (data not shown).



**Figure 5-8. Lead ion cleavage map of the 1450-2200 nt region of the MS2 genome produced from PAGE results.** The 78 lead ion induced cleavage sites that occur in all RNAs are marked with a black band. These are regions where RNA structure can be expected to be similar. The 26 lead ion induced cleavage positions that occur in the sub-genomic RNAs *in vitro* but not in the virion RNA are marked with a magenta band. These are cleavage sites that are masked in the virion which could be associated with CP<sub>2</sub> binding. The TR operator (nucleotides 1746-64) is aligned to the map in red. Cleavage positions within the TR operator *in vitro* are masked in virion. The map is aligned to two structural proposals (A: Fiers *et al.*, 1976 and B: Skripkin *et al.*, 1990) and one structure prediction (C: mFold minimum free energy structure) of this region of the genome. Most of the lead ion cleavage sites occur within or very close to regions that are predicted to be single stranded.



**Figure 5-9.** Lead ion cleavage positions mapped onto a proposed secondary structure of the MS2 genome. All PAGE identified  $Pb^{2+}$  cleavage sites are highlighted on the secondary structure proposed by Fiers et al. (1976). The emphasized region corresponds to nucleotides 1323-2180. Nucleotides which are cleaved are colour coded dependent on whether the  $Pb^{2+}$  cleavage was detected in vitro (red), in virion (cyan) or both (red). Regions where CP<sub>2</sub> binding is likely to occur are marked with a red dotted line. Many cleavages fall within or close to proposed single stranded regions, however, they are also detected in proposed double stranded regions. The differences are likely due to experimental conditions.  $Pb^{2+}$  probing was carried out at room temperature while enzymatic probing performed by Fiers et al was carried out at 4°.

The masked lead induced cleavage sites are an indicator of the structural rearrangement which occurs upon CP<sub>2</sub> binding. It seems that a more detailed structural analysis of the whole RNA genome allowing quantification of ssDNA products and perhaps incorporating complementary chemical probes would be required in order to accurately map CP<sub>2</sub> binding.

### 5.2.2 Lead induced hydrolysis of MS2 RNA assayed by CE

The structural footprinting data of the sub-genomic RNAs and the virion RNA reported in section 5.2.1 strongly suggests that the MS2 RNA secondary structure is partially conserved in virions. It was decided to assay the full length genome *in virion* and all the sub-genomic RNAs *in vitro* with respect to lead induced hydrolysis in order to get a complete structural footprint of the RNA genome. This would provide further confirmation of conserved RNA structure *in virion*. The mapping of CP<sub>2</sub> binding sites would also perhaps become more straightforward, taking into account the discontinuous manner in which CP<sub>2</sub> must interact with RNA within the virion. In addition, information on the topology of the MS2 genome double shell fold within the virion could be obtained (Toropova *et al.*, 2008).

Assaying lead acetate reactions using PAGE is a time consuming process. Furthermore, the sequencing gels used to comprise the marker maps in **Figures 5-6 and 5-8** did not allow more than a ~200 nt region to be probed at a time. This is close to the upper limit of the nucleotide “window” that can be observed in a single PAGE assay without compromising excessively on nucleotide resolution (Sambrook *et al.*, 2006). Recently, capillary electrophoresis (CE) has emerged as a method to assay RNA structure probing reactions (Mitra *et al.*, 2008; Vasa *et al.*, 2008). Like the PAGE assay, CE relies on indirect labelling of modified or hydrolysed RNA products by reverse transcription with a labelled primer. The primer is however fluorescently labelled which results in fluorescently labelled ssDNA reverse transcription products. The DNA products are fractionated using a DNA sequencer and detected as peak profiles which correspond to labelled ssDNA (**Figures 5-3 and 5-10**). Addition of a fluorescently labelled size standard to the primer extension reactions allows the positions of RNA modification or cleavage to be identified with relation to the primer binding site. Additionally, dideoxy sequencing reactions can be assayed in parallel. Although the analysis of fluorescently labelled ssDNA by CE has been around for quite

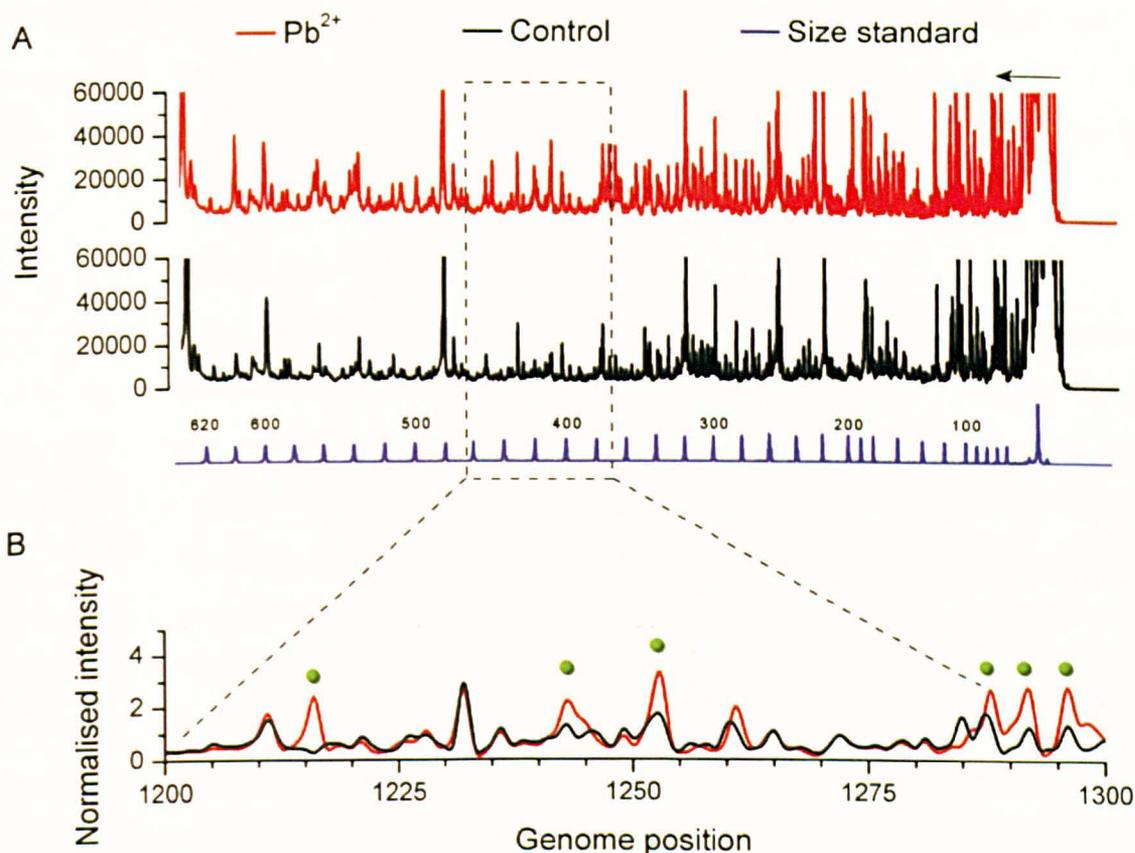
some time, the limitation in terms of usage for RNA structure probing has been due to a lack of software capable of analysing and quantifying the data output (Mitra *et al.*, 2008). Two such software analysis packages have recently been described, both of which analyse the peak profiles by fitting them to Lorentzian peaks whose number and position is constrained by the size standard using a non-linear least squares fitting model (Mitra *et al.*, 2008; Vasa *et al.*, 2008). The intensities of the peaks can then be used to quantify nucleotide reactivities with respect to an RNA probe and compile an RNA structure model. The use of CE to assay RNA structure probing reactions is beneficial over PAGE assays in that larger regions of genomic RNA can be probed at nucleotide resolution in a high throughput manner (Wilkinson *et al.*, 2008; Watts *et al.*, 2009).

It was decided to attempt to assay lead ion induced cleavage reactions of the MS2 sub-genomic RNAs and the full length MS2 genome *in virion* with capillary electrophoresis. The following sections report preliminary and pioneering results into genome wide structure probing of MS2 RNA.

#### 5.2.2.1 Assaying MS2 RNA lead induced hydrolysis positions with CE

The CE assays were performed in collaboration with Dr Susan J Schroder at the University of Oklahoma. The MS2 sub-genomic RNAs and the virion RNA within the MS2 capsid were treated with  $Pb^{2+}$  in a similar manner as described in section 5.2.1. Following primer extension of hydrolysed MS2 RNA, the ssDNA products were desiccated and shipped to the USA where they were resuspended and fractionated using a Beckmann CEQ 8000 Genetic Analysis System operated by Dr Xiabo Gu.

The  $Pb^{2+}$  induced hydrolysis reactions required optimisation in order to maximise nucleotide sequence read length and signal intensity. This was successfully achieved by increasing the amount of MS2 sub-genomic RNA and MS2 phage that was being probed by a factor of ten as compared to the cleavage reactions assayed by PAGE. Additionally, the sub-genomic RNAs and MS2 virion were incubated with  $Pb^{2+}$  ions for 30 min which resulted in optimal detection of  $Pb^{2+}$  induced cleavage. This provided nucleotide sequence read lengths in excess of 600 nt. **Figure 5-10** shows a typical sequencing trace of a  $Pb^{2+}$  probed MS2 RNA.



**Figure 5-10. Capillary electrophoretic analysis of the  $Pb^{2+}$  induced MS2 RNA cleavage positions.** Treatment of the sub-genomic RNAs and the virion RNA inside the MS2 phage with  $Pb^{2+}$  was similar to that outlined in Figure 5-3. Primer extension was carried out with primers labelled at the 5' termini with a Cy5 fluorophore and the reactions were analysed in the presence of a Beckman WellRed size standard. In A the sequence traces of primer extension reactions of the 5' sub-genomic MS2 RNA in the presence and absence of  $Pb^{2+}$  with the primer 1666.R are shown. The size standard which is co-electrophoresed with the primer extension products is also shown. The arrow indicates the positioning of the primer within the sequence trace. The traces are fed into the software CAFA which fits the signals to the appropriate number of nucleotides as defined by the size standard and aligns the two traces. Normalisation of the peak intensities allows comparison of trace files and the identification of cleavage positions indicated with green spheres (B).

The trace data were fed into the software algorithm CAFA (capillary automated footprinting analysis) which allowed automated alignment of the trace files of all four MS2 RNA samples obtained with identical primers. CAFA also fits the trace files and estimates the intensity at each nucleotide positions based on Lorentzian peak areas (Mitra et al., 2008). The data output from CAFA therefore promises to allow quantification of the hydrolysed RNA products, given that the data is of good quality with respect to nucleotide resolution and signal intensity. Based on the nucleotide sequence read length of roughly 600 nt, the MS2 sub-genomic RNAs and MS2 virion

RNA were probed with lead ions and analysed using eight primers which were spaced at roughly 500 nt within the MS2 genome in one go. Two of the primers did not work which resulted in a drop of coverage at the 5' end of the MS2 genome and a roughly 330 nt region close to the 3' end of the genome (**Figure 5-4**).

#### 5.2.2.2 Automated data analysis

Unfortunately the CE data were not of such quality that allowed automated sequence comparison and quantification of hydrolysis positions. Results from automated data analysis are however supplied in Supplementary Figure 8 in the appendix. The alignment of the nucleotide traces was often off by ~2 nt which inhibited data analysis by simple deduction of the control trace from the  $\text{Pb}^{2+}$  hydrolysis reaction trace to identify regions of lead ion cleavage. Trace mis-alignment can be observed at positions 1260-1262 in **Figure 5-10 B** which results in false positives lead ion cleavages being identified. This effect was predominantly observed at distances over ~300 nt from the primer annealing positions. The alignment problem had been anticipated and therefore, ddGTP and ddCPT sequencing reactions were carried out with all the primers used, however the read length from these reactions was unsatisfactory. Proper alignment of the sequence traces is crucial for accurate assignment of lead ion induced cleavage positions, particularly in light of the subtle conformational changes that  $\text{Pb}^{2+}$  ion probing is capable of detecting. For example, the cleavage pattern in **Figure 5-5 B** (black arrow 1) is slightly different over a region of four nt within the virion RNA as compared to the sub-genomic RNAs which could only be detected by CE given proper trace alignment.

Another problem that presented itself was slight bleedthrough of the ladder signal into the reaction traces. This was observed as signals with ~20 nt spacing within some CE traces dependent upon the primer used during reverse transcription. The intensity of these signals was dependent upon the signal intensity ratio of the ladder compared to the reaction trace. The dyes used to detect the primer extension products (Cy5) and the size standard (Wellred) have well resolved fluorescence emission maxima of 688 nm and 820 nm, respectively. It seems a simple solution to this problem could involve optimisation of the amount of ladder added to each primer extension reaction with respect to the primer being used as yields from primer extension reactions will be dependent upon the primer used.

The amount of noise present as RT-stops or cleavages in the control reactions lacking  $Pb^{2+}$  was rather high in all the samples analysed presenting an additional problem. This had the effect of masking lead ion cleavage sites by producing false negatives. As opposed to the PAGE assays, these false negatives could not be identified with as much accuracy as the reactions assayed with CE were only assayed at a single  $Pb^{2+}$  incubation time point.

Despite the shortcomings of the CE data set, comparison of the traces by eye revealed that they complemented the results from the PAGE assays quite well.

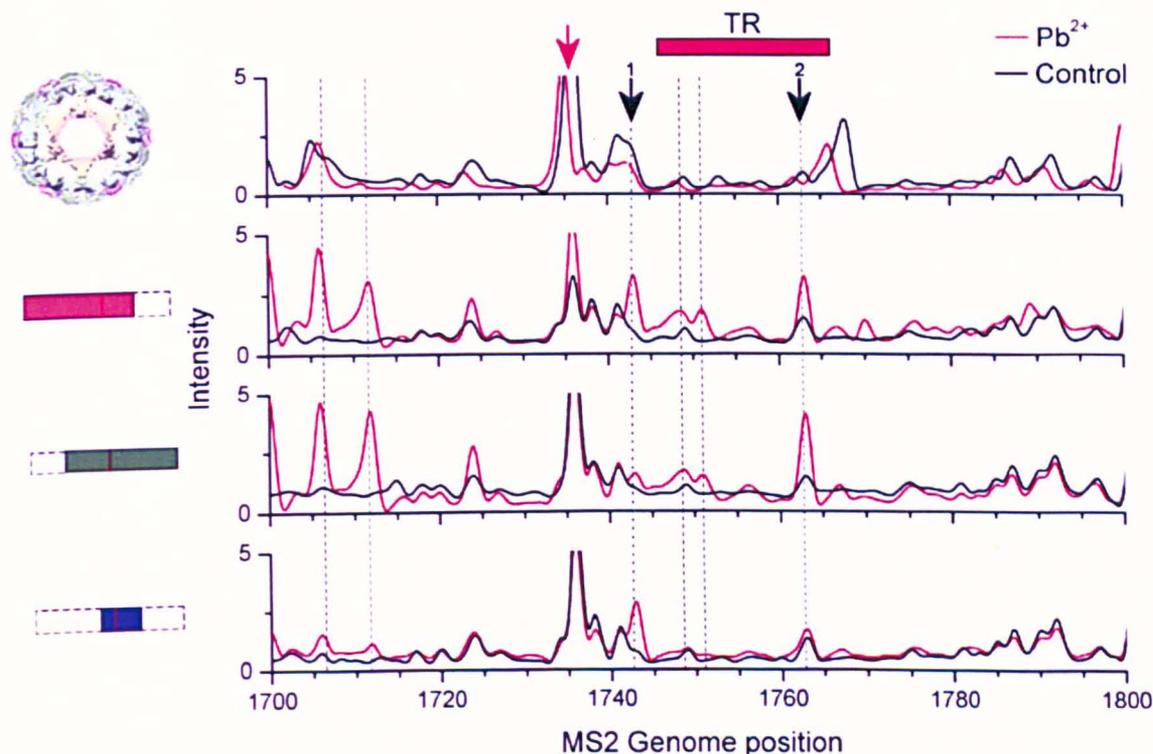
### 5.2.2.3 CE assays confirm that MS2 RNA structure is partly conserved *in vitro* and *in virion*

Examination of the CE traces confirmed the results obtained from the PAGE assays. **Figure 5-11** shows an alignment over a 100 nt region incorporating the TR operator of the three sub-genomic RNA fragments. Lead ion induced cleavage positions are conserved over this region between the sub-genomic RNAs *in vitro* suggesting a similar conformation for these RNAs in solution. **Figure 5-12** shows all identified lead ion cleavage positions within the 5'-, 3'- and iRNA sub-genomic fragments. Lead ions induce hydrolysis of these RNAs at many identical nucleotide positions within regions where these RNAs overlap.

The overlapping region of all the sub-genomic RNAs which was assayed with CE corresponds to nucleotides 1419-2160 of the MS2 genome. This corresponds to roughly the same region which was assayed by gel electrophoresis. **Figure 5-13** shows the number of  $Pb^{2+}$  induced cleavage sites occurring at identical nucleotide positions within the MS2 sub-genomic fragments over this region. 66 % of the lead ion cleavage sites identified in the 5' sub-genomic RNA were also identified in the 3' RNA and iRNA while this percentage was 51 % and 82 % for the 3' RNA and iRNA respectively.

The MS2 genomic region which was assayed in which only the 5'- and 3' MS2 sub-genomic fragments overlap encompasses nucleotides 992-2165 of the MS2 genome. This region corresponds to 48 % and 46 % of the 5'- and 3' RNA sequences, respectively. 146 identical  $Pb^{2+}$  cleavage positions out of a total of 183 for the 5' RNA and 199 for the 3' RNA were identified within this region which corresponds to 80 % and 73 % of the total number of cleavages identified for each sub-genomic fragment. The observation that lead ions induce cleavage of the sub-genomic RNAs at many similar nucleotide positions strongly implies that their solution structure is partly

conserved in assembly buffer. This correlates with the results obtained from the PAGE assays and implies that the ensemble of structures that these RNAs exist in prior to capsid assembly is analogous.



**Figure 5-11.** *The structure of the MS2 sub-genomic RNAs is partly conserved in solution.* The figure shows nucleotide traces of primer extension products obtained by reverse transcription of MS2 virion RNA, 5' RNA, 3' RNA and iRNA lead ion induced hydrolysis products. A 100 nt window incorporating the TR operator is shown. The traces shown in each row for the  $Pb^{2+}$  probed (red) and control sample (black) correspond to the RNAs indicated on the left. The intensities of the peaks within a trace correspond to the RNAs indicated on the left. The positions of some lead ion induced cleavages are conserved between the sub-genomic RNAs (dashed lines) suggestive of a similar RNA structure. The conserved lead ion induced cleavage sites all appear to be masked within the MS2 virion. The arrows correspond to the arrows shown in Figure 5-5 and highlight an identical lead ion induced cleavage pattern obtained for the MS2 RNAs independent of the nucleotide size fractionation assay.

Complementary to the PAGE assays, the CE assays suggested that the susceptibility of the MS2 genome within the MS2 virion to lead ion induced cleavage is somewhat similar to that observed *in vitro*. 217 lead ion cleavage sites were identified in an area that covers 74 % of the MS2 genome. These cleavage positions are shown in **Figure 5-12**. 118 of these sites were conserved within the sub-genomic RNAs. These sites are

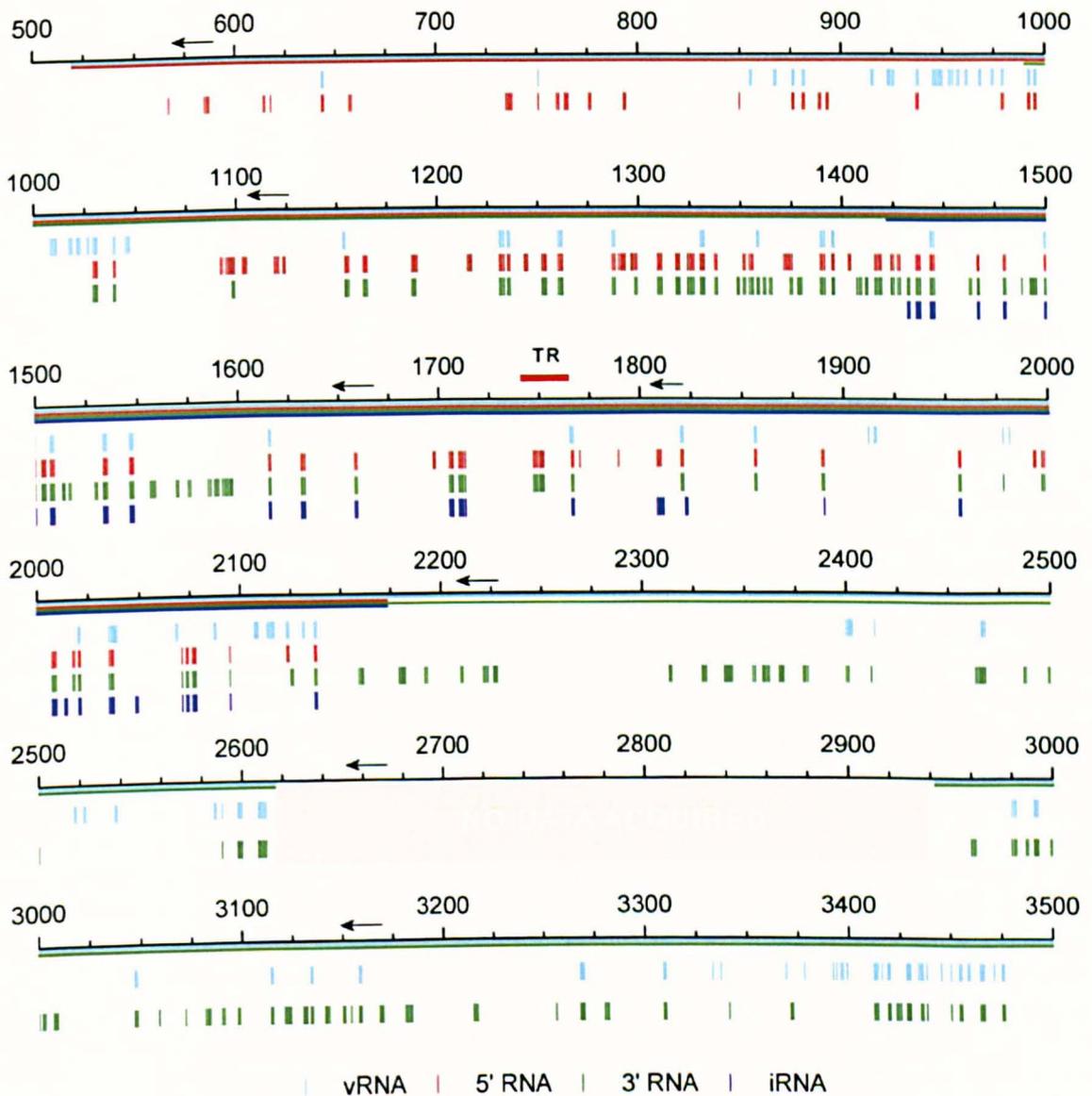
mapped in **Figure 5-14**. With respect to the 5' sub-genomic RNA, 139 lead induced cleavage positions were identified in the virion RNA over the ~1600 nt assayed region in which the 5' RNA and virion RNA overlap. 72 of these sites are conserved in the 5' RNA *in vitro*. Similarly, 175 cleavage sites were mapped within the ~2159 nt region assayed in which the 3' sub-genomic RNA and the MS2 virion RNA overlap 110 of which were conserved in the 3' RNA *in vitro*. Finally, 53 cleavage sites were identified in the phage RNA of which 25 were also identified within the iRNA over the ~750 nt region in which these RNAs overlap.

These data show that 50-60 % of the identified lead ion cleavage positions detected in the virion RNA within the MS2 phage are also detected in the sub-genomic RNAs *in vitro*. This correlates well with the PAGE results which showed that 64 % of the lead ion induced cleavage sites occurring in the phage also occur within the MS2 sub-genomic RNA fragments *in vitro*. These numbers suggest that the solution conformation of the MS2 genomic RNA is partly conserved in the MS2 virion following capsid assembly. The results also further confirm the idea that the MS2 genome adheres to a specific structure within the virion as mentioned in 5.2.1.2.

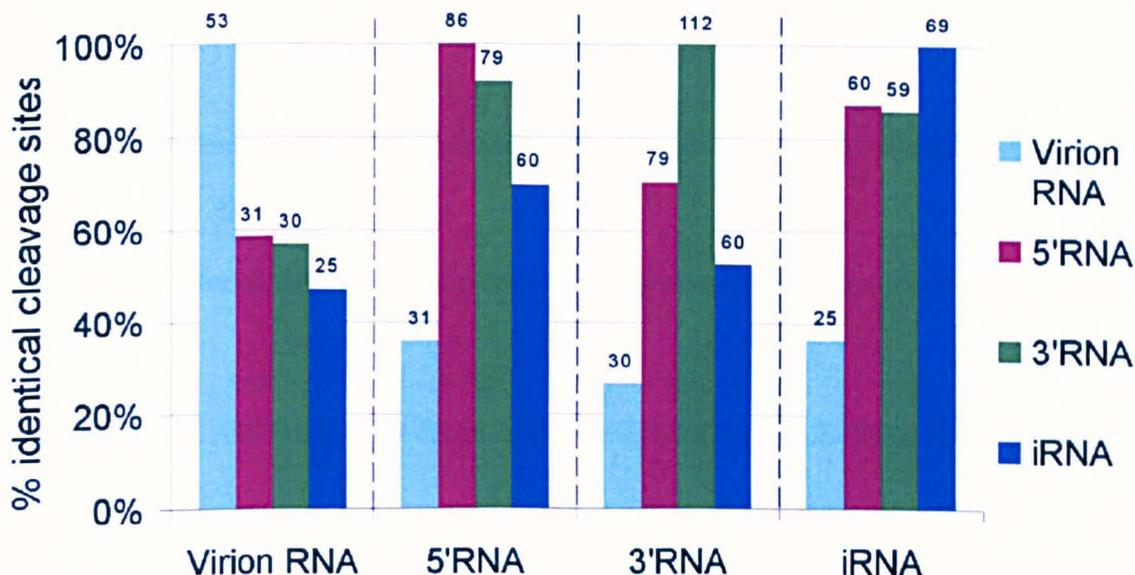
#### 5.2.2.4 Mapping CP<sub>2</sub>-RNA binding positions with CE

**Figure 5-14** shows the nucleotide positions at which lead ion induced cleavage is masked in the MS2 phage RNA as compared to the sub-genomic RNAs *in vitro*. These are positions where the conformational flexibility of the RNAs is different due to a difference in structural environment. There are cleavage positions that occur in the TR operator *in vitro* which are masked within the MS2 virion. The difference in lead ion cleavage efficiency over the TR operator is shown in **Figure 5-11**. The different lead ion induced hydrolysis pattern observed within the TR operator is suggestive of a different structural environment *in virion* as compared to *in vitro*. These results complement the PAGE assays in that they indicate that CP<sub>2</sub> binding can be assayed by monitoring the susceptibility of virion RNA to lead ion induced cleavage.

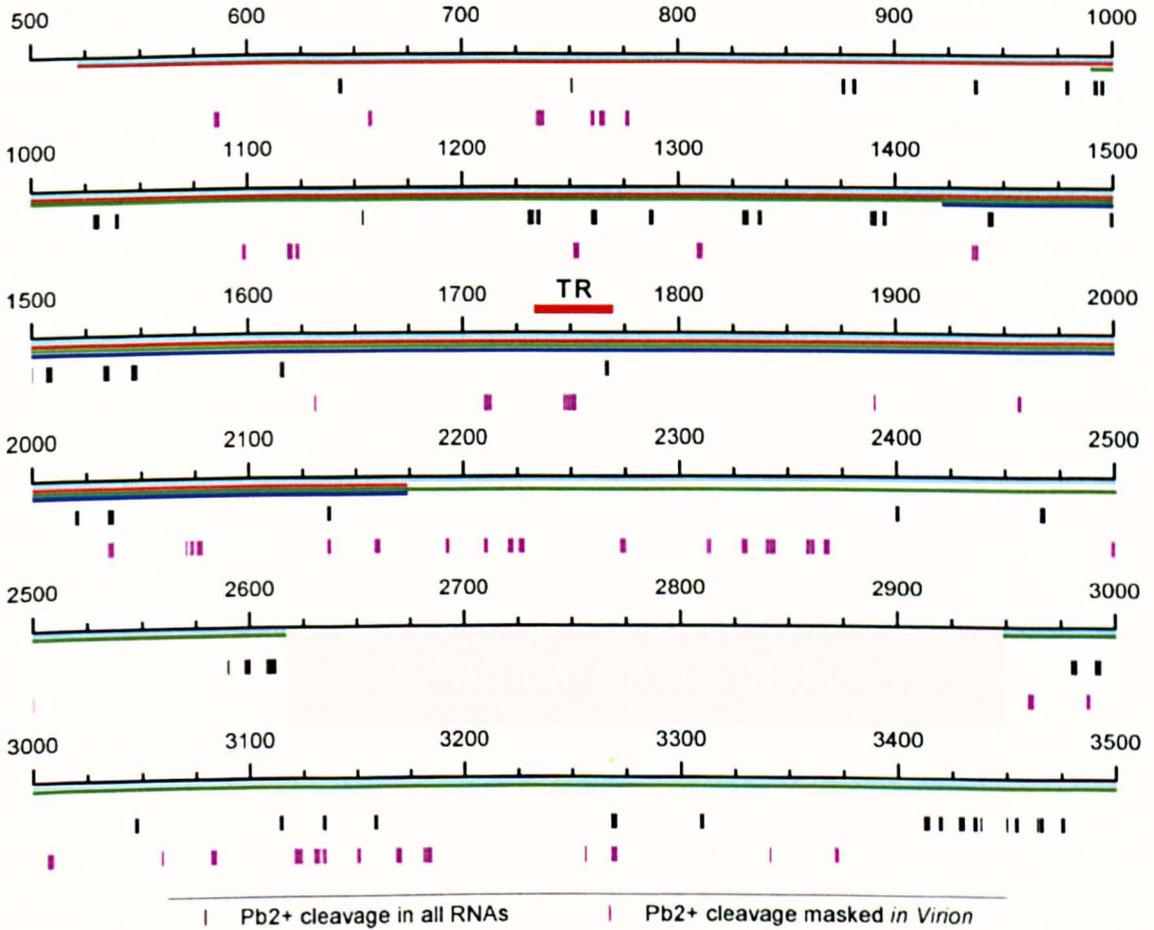
As with the PAGE results, whether or not CP<sub>2</sub> binding takes place around other masked cleavage positions can not be determined based on these results. It seems that a more complete data set which in addition allows quantification of ssDNA products will be key in order to map CP<sub>2</sub> binding sites and gain information on the topology of the MS2 RNA genome *in virion*.



**Figure 5-12. A lead ion induced cleavage map of the 515-2615 and 2950-3486 nt regions of the MS2 genome.** The map highlights all identified lead ion cleavage positions within the sub-genomic RNAs *in vitro* and the genomic RNA within the MS2 virion. Cleavage positions are indicated with vertical bars that are colour coded according to the RNA and are accurate to  $\pm 2$  nt. 217, 225, 355 and 75 lead ion cleavage sites were identified in the MS2 virion RNA, 5'-, 3'- and iRNA, respectively. The map can be analysed like a PAGE gel if rotated  $90^\circ$ . All cleavage sites are shown. No discrimination was made between strong and weak cleavage positions when the cleavage sites were identified from electropherograms. The annealing positions of the primers used in primer extension reactions are shown as black arrows, with the exception of the primer 3554.R. Horizontal lines, colour coded with respect to each MS2 RNA and positioned underneath the nucleotide bar, indicate which RNAs contribute to the cleavage map at any particular position. There are many lead ion induced cleavage positions that are shared between both the sub-genomic RNAs and the MS2 genome within the MS2 virion suggestive of similar RNA structure at these positions.



**Figure 5-13.** Comparison of the number of identical  $Pb^{2+}$  induced cleavage positions occurring within the 1419-2160 nt region of the MS2 genome. The bar chart shows the number of identical  $Pb^{2+}$  cleavages as a percentage of the total number of  $Pb^{2+}$  cleavages identified for a particular RNA using the CE assay. For example: 53 lead ion cleavage positions were identified in virion RNA within the 1419-2160 nt region of the MS2 genome. 31 of these 53 sites were also observed in the 5' RNA, 30 in the 3' RNA and 25 in the iRNA. 47-58 % of all identified cleavage sites within the virion are therefore also observed in the sub-genomic transcripts. The percentage of identical cleavage sites between individual RNA sub-genomic transcripts in vitro is higher still. The Figure is similar to **Figure 5-7**. Although fewer cleavage positions were identified with the CE assay as compared to the PAGE assay the results are similar in that a large proportion of cleavage sites are conserved within the MS2 virion while an even larger proportion is conserved between the sub-genomic RNAs in vitro. The results strengthen the suggestion of conserved MS2 RNA structure in vitro and in virion.



**Figure 5-14.** A lead ion induced cleavage map of the 515-2615 and 2950-3548 nt regions of the MS2 genome. Lead cleavage sites that are conserved between the sub-genomic RNAs and the MS2 genome within the virion are shown with a vertical black bar. These sites occur where the conformational flexibility of the sub-genomic RNAs *in vitro* and the MS2 virion RNA inside the phage capsid are similar. The lead ion induced cleavage data coverage over the genome is different. This is indicated with colour coded bars as in Figure 5-10. For example, a lead ion cleavage site that is conserved within the genome at positions 515-992 is only conserved between the MS2 virion RNA and the 5' sub-genomic RNA, however, a hydrolysis site at genome positions 992-1419 is conserved within the MS2 virion RNA, the 5'- and 3' sub-genomic RNAs. Lead cleavage positions that are masked in the MS2 phage virion are positioned with a vertical magenta bar. These are sites of possible CP<sub>2</sub> interaction as the susceptibility to lead ion induced hydrolysis at these positions is different. Lead ion induced cleavage positions are masked within the TR operator sequence which is shown as a red bar aligned to the map.

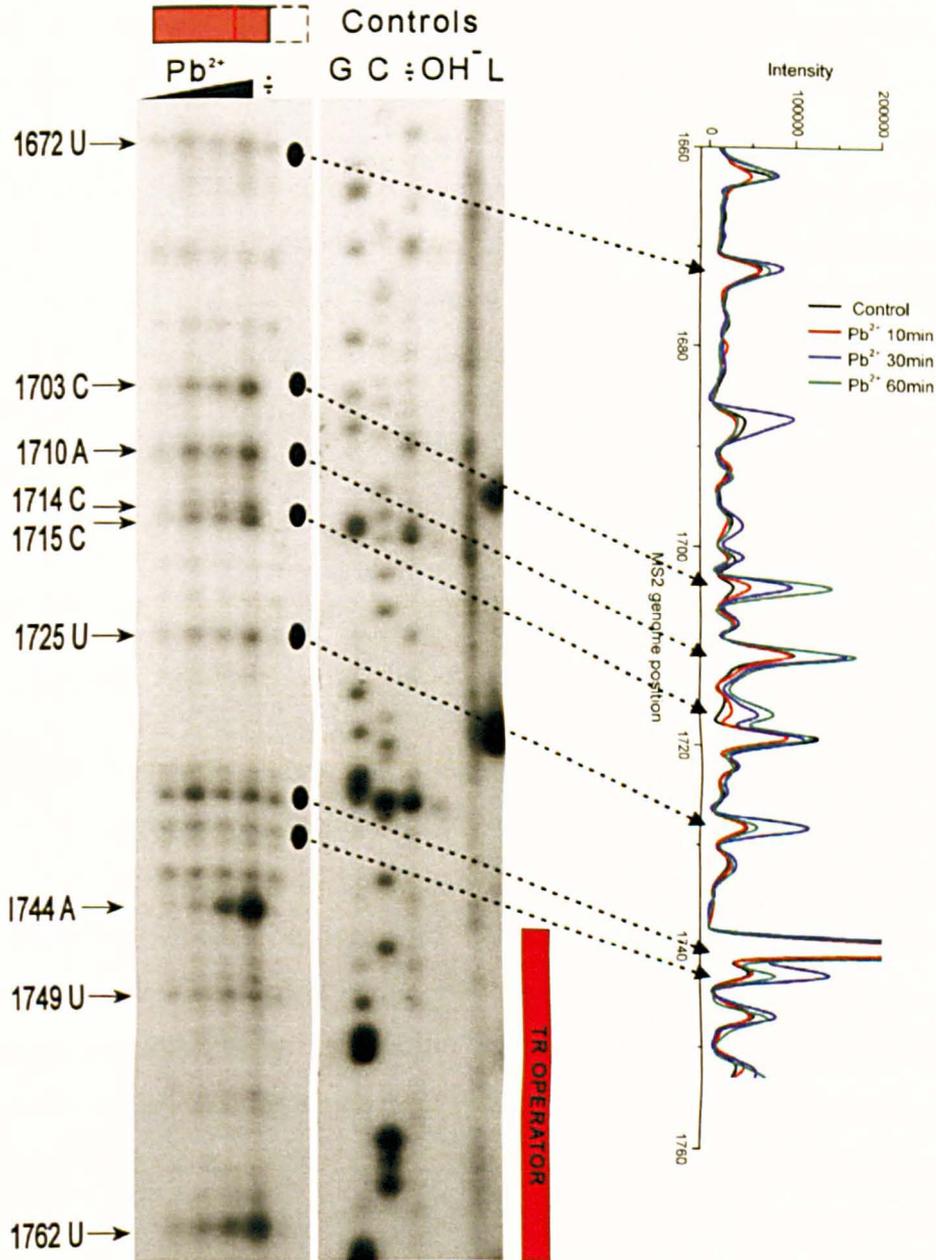
### 5.2.2.5 Comparison of CE and PAGE assay results

Comparison of the CE and PAGE data over the region of shared nucleotide sequence between (nucleotides 1450-2160) reveals that the cleavage patterns are not entirely identical. Many more lead cleavage sites were for example identified off the PAGE gels (**Figure 5-6** vs. **5-12**). This seems to indicate that these assays do not correlate well. Comparison of the cleavage over the TR operator (**Figure 5-5** vs. **Figure 5-11**) however shows that the lead ion induced hydrolysis patterns of the sub-genomic RNAs assayed by CE and PAGE are nearly identical. The TR operator is in both cases flanked upstream by a strong termination of transcription site indicated by a red arrow and a strong hydrolysis site (black arrow 1) and has a strong cleavage site close to its 3' side (black arrow 2). The differences observed therefore seem to be due to the quality of the data sets and the manner in which they were analysed.

The genome wide CE data is in many ways preliminary and represents results from mainly a single experiment. The reactions analysed by CE were also performed at tenfold higher RNA concentration than those assayed by PAGE. Furthermore, the lead ion cleavage reactions were assayed at multiple time points by PAGE but only at a single time point by CE, perhaps resulting in the loss of identified cleavage sites. This along with other problems with the CE data, mentioned in section 5.2.2.2, seems to result in discrepancies between the two data sets.

As mentioned in 5.2.2.1 the CE were optimised with respect to  $Pb^{2+}$  incubation time prior to assaying the full length genome with  $Pb^{2+}$  at a fixed 30 min timepoint. The results from these  $Pb^{2+}$  induced cleavage time point CE assays allow a more detailed comparison with the PAGE assays. **Figure 5-15** shows that the  $Pb^{2+}$  induced timepoint reactions assayed with PAGE and CE are in excellent agreement.

For future work, in order to obtain good CE traces which will allow quantification of cleavage data with minimum background noise, it will be important to optimise the lead ion cleavage assay for CE. A good way to do this would be to shorten reaction times from 30 min to 2 min and increase the  $Pb^{2+}$  concentration. This would minimise non-specific cleavage which perhaps takes place in assembly buffer over time and thereby reduce signal noise. Increasing the  $Pb^{2+}$  concentration and performing cleavage over a range of  $Pb^{2+}$  concentration would optimise the cleavage assay with respect to the larger amount of RNA required for the CE assays and provide a more informative data set.



**Figure 5-15.** The PAGE and CE assays of lead ion induced cleavage correlate well. A 100 nt region of the 5' sub-genomic RNA is shown as assayed by primer extension using the primer 1812R. In both cases the 5' sub-genomic RNA was probed with lead ions at multiple timepoints. The PAGE assay is identical to that shown in Figure 5-5. The black triangle indicates  $Pb^{2+}$  probing for 5, 10, 30 and 60 min with  $\div$  indicating a no  $Pb^{2+}$  control reaction. The CE electropherogram shows aligned traces of primer extension reactions of 5' RNA treated with  $Pb^{2+}$  for 10, 30 and 60 min. The two assays are in good agreement with one another.

## 5.3 Discussion

The biochemical analysis of the MS2 genome residing within the MS2 virion and of the MS2 sub-genomic RNA fragments reported in this chapter provides valuable information on the structural organisation of large viral ssRNA genomes both in solution and in their most compact state within the virion. By taking advantage of the permeability of the MS2 capsid to small molecules, the structure of the MS2 genome within the MS2 virion could be compared to ssRNAs of similar sequence *in vitro* on the basis of their susceptibility to lead ion induced hydrolysis. These experiments allowed comparison of MS2 RNA backbone structure with near single nucleotide resolution. The results are the first where the structure of a large ssRNA genome has been analysed biochemically inside a capsid.

### 5.3.1 *The MS2 genome has a defined structure within the MS2 virion*

Cryo-EM reconstructions of various ssRNA viruses have revealed glimpses into genome structure within virions. Although the degree of RNA that is accounted for in these structural models varies, they have shown that genomic RNA is at least partially ordered within virions (Schneemann, 2006). In the Nodavirus PaV foreign RNA exhibits similar icosahedral order within virions implying that the structural organisation is non-specific with respect to RNA sequence (Johnson *et al.*, 2004; Tihova *et al.*, 2004). For MS2, for which there exists the most complete structural model of any ssRNA within a virion, the icosahedral ordering of the RNA genome does not propagate to atomic resolution (Toropova *et al.*, 2008; Valegård *et al.*, 1990; Golmohammadi *et al.*, 1993). The loss of electron density that can be associated with viral RNA at atomic resolution has also been observed for the Picorna virus Seneca Valley Virus (Venkataraman *et al.*, 2008). This raises the question whether RNA genome structure is defined at the level of RNA 2° structure within virions. The lead ion cleavage assays of the MS2 genome within the MS2 virion suggest that this is indeed the case. The structural footprints of the RNA genome imply that there is a defined RNA structure or ensemble of structures that the genome adheres to within the virion. The MS2 RNA genome is therefore not randomly packaged into the double shell tertiary structure observed in the MS2 cryo-EM model (Toropova *et al.*, 2008). This suggests that the

structure underlying the double shell conformation of the MS2 genome is conserved and that the absence of RNA density at atomic resolution is due to icosahedral averaging rather than structural heterogeneity of the RNA genome within the virion. These results correlate well with the X-ray structure of BPMV wherein a non random nucleotide sequence was observed to contact the BPMV capsid suggesting a defined RNA structure within the BPMV virion (Lin *et al.*, 2003).

### 5.3.2 *The structure of MS2 sub-genomic fragments is conserved in solution*

Three MS2 sub-genomic RNAs which incorporate different regions of the MS2 genome were analysed with respect to their susceptibility towards lead ion induced hydrolysis. The probed region corresponds to nucleotides 1419-2190 of the MS2 genome in which the sub-genomic MS2 RNAs overlap. A large proportion of identified lead ion hydrolysis sites within this region were observed at identical nucleotide positions within the sub-genomic MS2 RNAs. Further comparison of the 5' and 3' sub-genomic RNAs showed that lead ion cleavage positions were also conserved to a large extent outside this region. The high proportion of lead ion cleavage sites that were shared between these RNAs suggests that their structural conformation in solution is similar.

These results reflect the idea that RNA structure is dominated by secondary structure elements formed by nucleotide residues closely spaced within the RNA molecule (Fresco *et al.*, 1960; Tinoco and Bustamante, 1999). This is evident from the similar footprints of the sub-genomic RNAs despite their different nucleotide extensions in either the 5' or 3' direction with respect to the probed genomic region. This suggests that the MS2 sub-genomic RNAs exist as an ensemble of solution structures which are structurally similar within regions of identical nucleotide sequence.

In light of the results reported in chapter 4, which indicated that the MS2 RNA structure is larger with respect to the size of the MS2 capsid, it seems plausible that the sub-genomic RNAs exist as linear arrays of secondary structural elements in solution. This type of structure would correspond to recent atomic force microscopy analysis of ssRNAs extracted from  $T = 3$  viruses (Kuznetsov *et al.*, 2005). In Kuznetsov *et al.* (2005) viral RNA extracted from polio virus and TYMV amongst others was shown to conform to a linear arrangement of local stem loop structures at room temperature.

### 5.3.3 Implications for MS2 capsid assembly

In light of the results presented in chapter 4 which indicate that CP<sub>2</sub> act as chaperones that fold the RNA genome during capsid assembly, the structure probing data provide insight into the RNA structural transition that occurs during genome packaging. The observation that a large proportion of identified lead ion induced cleavage positions are conserved within the virion implies that the MS2 genome is folded from secondary structure elements that are ultimately defined by RNA sequence. It appears that the tertiary fold within the virion therefore reflects a structure that is a compacted version of the RNA solution structure. Similar results have been obtained by McPherson and colleagues who showed that removal of coat proteins from various ssRNA viruses results in the slow unfolding of compact RNA structures into linear secondary structure elements as observed by AFM. This reflects an apparent reversal of the encapsidation process. Remarkably, the unfolded viral RNAs could refold into compact structures upon lowering the temperature suggesting that genomic RNA alone programs its encapsidated conformation (Kuznetsov *et al.*, 2005; Kuznetsov *et al.*, 2006). It appears that this could also be the case for MS2.

This is plausible in light of the cryo-EM reconstruction models of the MS2 phage and sub-genomic RNAs which makes clear that a discontinuous network of CP<sub>2</sub>-RNA binding sites exists within the MS2 capsid (Toropova *et al.*, 2008; Toropova, 2009). It is unlikely that this network is not at least partly conserved in solution taking into account that the free energy changes involved in complete disruption of an RNA stem loop such as TR as compared to CP<sub>2</sub>-TR binding are fairly similar (Gell *et al.*, 2008; Carey *et al.*, 1983) and that the positioning of RNA stem loops is defined by RNA sequence. It is unlikely that a CP<sub>2</sub>, the binding of which to RNA is structure dependent (Romaniuk *et al.*, 1987), can force a stem loop structure upon the RNA genome at positions where stem loop formation is unfavourable due to underlying RNA sequence. Reshuffling of CP<sub>2</sub> binding sites during folding of the RNA genome therefore seems unlikely although the lead ion cleavage data do not rule out such structural rearrangement. In chapters 3 and 4, the efficiency of capsid formation with the sub-genomic RNAs was shown to be dependent upon the size of the RNA being encapsidated. The structural similarity of sub-genomic RNAs *in vitro* and to the genome inside the MS2 virion imply that this difference is therefore perhaps not due to a different rearrangement of RNA secondary structure prior to or during packaging.

Rather, the higher electrostatic repulsions between the phosphate backbone of larger RNAs that must be overcome in order for them to be packaged (Belyi and Muthukumar, 2006) and increased conformational space that the larger RNAs must sample in order to conform to a structure compatible with the MS2 capsid seem to be of key importance.

#### 5.3.4 Genomic mapping of CP<sub>2</sub> binding sites

A large proportion of all identified lead ion cleavages were conserved between the MS2 genome and sub-genomic fragments, however, many were also identified that suggested that the structural environment of the MS2 genome is different inside the virion vs. *in vitro*. This is in accordance with the estimated hydrodynamic radii of the MS2 assembly components in chapter 4. The tertiary structure of the MS2 RNA in solution must be different of that in the virion on account of a more expanded structure.

The cryo-EM reconstruction of the MS2 phage suggests that two thirds of the RNA genome will make a minimum of 60 contacts to the capsid interior. A contribution to the difference in the RNA genome structure and/or structural environment *in vitro* vs. *in virion* that can be detected with lead ion hydrolysis will therefore be due to CP<sub>2</sub> binding. This was indeed detected as the structural footprint within and surrounding the TR operator sequence was different. This was an important finding for two reasons. Firstly, although the high affinity of the CP<sub>2</sub> for the TR stem loop indicates that the CP<sub>2</sub>-TR operator complex is conserved in the full length genome within the virion (Carey *et al.*, 1983b), this has never been proven (Beckett and Uhlenbeck, 1988). The different lead ion cleavage pattern over the TR operator *in virion* suggests that this is indeed the case. This supports the idea that TR is located in the outer shell of RNA density in the cryo-EM structure model (Toropova *et al.*, 2008). It also supports the idea that there is a defined structure for the RNA within the virion. The MS2 genome is believed to be circularised through contacts with the A-protein at its 5' and 3' termini (Shiba and Suzuki, 1981). If this is the case then the CP<sub>2</sub>-TR complex would provide a structural anchor that together with the A-protein would result in defined spatial order of the RNA genome within the virion. Secondly it suggests that the discontinuous network of CP<sub>2</sub> binding sites within the MS2 genome could be identified using the lead ion cleavage assay.

Multiple potential CP<sub>2</sub> binding sites were identified over the 740 nt central region of the MS2 genome that was probed using PAGE. Considering an average RNA stem

loop size of 20 nucleotides, the lead ion cleavage positions masked in the virion were 10-12 and discontinuously spaced over the probed region (**Figure 5-7**). Supposing that 60 CP<sub>2</sub> must bind to the RNA genome to conform to an A/B dimer conformation and that their spacing is even within the genome, approximately 12 CP<sub>2</sub> binding sites would be expected to be found within the probed region. These numbers are quite similar, however the comparison is quite naïve as there will be contributions towards RNA structure variation from both the A-protein and different tertiary RNA interactions within the genome which must accompany a more compact RNA structure. The PAGE data were also not optimised to allow detailed comparison as the goal of these experiments had initially been to compare footprints to get an idea of structural similarity. In order to define all CP<sub>2</sub> binding sites a more detailed structural analysis was attempted by assaying the whole genome using CE.

### ***5.3.5 High throughput structural analysis of the MS2 genome structure***

The lead ion cleavage reactions that were assayed by PAGE strongly suggest an underlying structural template that the MS2 genome adheres to within the virion. This strengthens the idea of the multifunctional role of the MS2 genome in regards to CP<sub>2</sub> conformational control during capsid assembly suggested in chapter 4. In order to gain full understanding of this role, knowledge of global genome structure and topology within the MS2 virion is required.

By assaying the lead ion cleavage reactions with capillary electrophoresis it was anticipated that the fine details of RNA structure could be analysed in a rapid manner. This proved to be the case as the time it took to generate the CE results could be measured in weeks as compared to months for the PAGE data. The CE data confirmed the findings that MS2 genomic structure is conserved in solution and to an extent inside the virion. The data are however preliminary and require repeating in order to lower sequence background levels, increase data coverage and data information. This can be achieved in a relatively simple manner and will be crucial in order to obtain quantifiable results which are necessary to elucidate CP<sub>2</sub> binding positions within the genome. Such data would also provide structural constraints which could be used to build an accurate RNA structure model of the MS2 RNA genome within the virion (Mathews *et al.*, 2004).

RNA structure probing using CE is a novel technique. At this moment in time only two independent laboratories have published results where nucleotide traces have been used to elucidate RNA structure. Both of these labs were involved in producing the bioinformatics software capable of analysing sequencing data (Mitra *et al.*, 2008; Vasa *et al.*, 2008). The most exciting appeal of CE for RNA structure probing is the rapid manner in which a manageable and informative dataset can be produced. This promises to push RNA structure probing forward as larger RNAs are now amenable to experiment. Knowledge of large RNA structure is very limited (Yoffe *et al.*, 2008). Perhaps not surprisingly, the MS2 genome is one of the few large RNAs (1 > kb) which has been structurally characterised. That research provided information on how RNA structure regulates MS2 gene translation (van Duin, 2006). The manner in which RNA controls folding of protein domains has recently been suggested based on the complete structural characterisation of a viral genome (Watts *et al.*, 2009). These results emphasise that RNA structure constitutes an important part of the genetic code. The results presented in this thesis have highlighted a further global structural role for viral RNAs beyond single secondary structure elements. The CE results presented in this chapter show that elucidation of the structural motifs involved in capsid formation and their organisation within the MS2 capsid are within reach.

## 6 Conclusion

---

The main focus of this thesis has been to investigate the manner in which MS2 capsids assemble and in particular how genomic RNA influences the capsid assembly process. The ssRNA virus MS2 is of interest because it represents a model system in which the various aspects of both the structure and life cycle of ssRNA viruses can be studied. ssRNA viruses represent a substantial part of all known viral pathogens. A key event in the life cycle of these viruses is the enclosure of their genomic material within a protein capsid. Understanding the mechanism in which these capsids form and the attributes of underlying nucleic acid-protein interactions which must accompany their formation could be beneficial for the production of novel anti-viral therapeutics as well as increasing the understanding of macromolecular assembly mechanisms. MS2 is a good model system because it is safe to work with and is well characterised both structurally and biochemically. Importantly, the MS2 capsid can be disassembled and reassembled *in vitro* in a process which is induced by RNA and has allowed the role of RNA in capsid assembly to be explored.

In chapter 3 of this thesis the production of RNA transcripts of different sizes and corresponding to discrete regions of the MS2 genome was described. In chapters 3 and 4, these RNA transcripts were shown to promote capsid assembly with different efficiencies as assayed by gel retardation assays and analytical ultracentrifugation. In chapter 5, the structures of the RNA transcripts were investigated in solution and shown to be similar to the structure of the MS2 genome within MS2 virions as assayed by biochemical structure probing.

MS2 capsids are composed of 180 coat protein subunits. The capsid has icosahedral symmetry with  $T = 3$  subunit arrangement i.e. the coat proteins are observed in three quasi-equivalent conformations in the capsid. The coat proteins form two types of dimers (A/B and C/C) which serve as the building blocks of the capsid. A/B dimers are arranged around the capsid 5-fold axes while the C/C dimers span the distance between the capsids 3-fold axes. A loop of polypeptide connecting the F and G  $\beta$ -strands of the coat protein has a different conformation in the B subunit but an identical conformation in the A and C subunits. This results in a symmetric C/C dimer structure while the A/B dimer is asymmetric (Valegård *et al.*, 1990). For correct capsid assembly

to take place the dimers must be arranged with the appropriate conformations throughout the capsid.

How is quasi equivalent conformer packaging achieved during capsid assembly? A current proposed mechanism indicates that discrimination between coat protein conformers occurs upon binding to a 19 nt RNA stem loop translational operator of the MS2 replicase gene (Stockley *et al.*, 2007). The translational operator (TR) has high affinity for coat protein and nucleates capsid assembly (Bernardi and Spahr, 1972; Carey *et al.*, 1982; Beckett and Uhlenbeck, 1988; Beckett *et al.*, 1988). Using mass spectrometry it was shown that upon mixing CP<sub>2</sub> with TR at equimolar ratios, a CP<sub>2</sub>-TR complex was formed and capsid assembly was negligible. Upon further addition of CP<sub>2</sub> however, capsid formation proceeded rapidly. This suggested that two types of dimer were required for assembly, an RNA bound and an RNA free CP<sub>2</sub>. NMR structure studies of a non-assembly competent coat protein mutant, showed that in the absence of RNA the coat protein dimers are predominantly symmetric while binding of TR results in a conformational change in the FG-loop region which results in an asymmetric dimer. Together these results suggest that TR acts as a molecular switch providing A/B dimers in solution whose conformation is otherwise largely C/C like (Stockley *et al.*, 2007).

But how might capsid assembly proceed in the presence of the 3569 nt MS2 genome where only one copy of the TR operator is present? A cryo-electron microscopy 3D structure model of the MS2 phage in which ~90 % of the RNA genome was accounted for indicated that the genome is well ordered within the MS2 virion and that it exists as an ensemble of similar tertiary structures. The RNA is packaged as two concentric shells with roughly 2/3 of the RNA density located in an outer shell while the remainder is located in an inner shell. The two shells are connected at the icosahedral 5-fold axes where the observed electron density is large enough to accommodate an RNA duplex. The outer shell is in close proximity to the capsid interior at positions which correspond to known TR binding sites on the CP<sub>2</sub>. The observed density of RNA beneath the A/B dimers is on average different from the RNA beneath the C/C dimers. This is consistent with the model of RNA induced conformational control of CP<sub>2</sub> conformers during capsid assembly. The cryo-EM reconstruction therefore implies that multiple sites within the MS2 genome are capable of quasi-equivalent control, possibly in a similar manner as observed for TR (Toropova *et al.*, 2008).

The data reported in Stockley *et al.* (2007) and Toropova *et al.* (2008) suggested a function of MS2 RNA in capsid assembly and served as an incentive for further research into the roles of genomic RNA during MS2 capsid assembly.

In chapters 3 and 4 results from experiments were reported wherein the role of the MS2 genome during capsid assembly was investigated by monitoring how sub-genomic RNAs affect capsid assembly efficiency. In order to produce MS2 genomic RNA, three cDNA transcripts corresponding to different regions of the MS2 genome were produced which, in principle, allowed the generation of any sized RNA corresponding to the MS2 genome using the technique of reverse transcription. Taking into account the relative amounts of RNA observed in each RNA shell within the MS2 virion (Toropova *et al.*, 2008), two RNA transcripts corresponding to roughly two thirds of the RNA genome from both the 5' and 3' termini and one corresponding to roughly 1/3 of the MS2 genome were produced. In addition, a fourth transcript corresponding to roughly 1/3 of the genome but lacking the TR operator was produced. These sub-genomic RNAs along with the full length MS2 genome allowed the assessment of the effect of RNA length, sequence and structure on capsid assembly.

In chapter 3, the sub-genomic RNAs were shown to induce capsid formation as observed by their ability to promote formation of material which migrated with similar rates as recombinant capsids in native agarose gels (GEMSA). The capsids appeared similar to MS2 and fr capsids, which have been reconstituted with genomic length RNA previously, in that they had similar morphology as observed by transmission electron microscopy (TEM) (Sugiyama *et al.* 1967,; Hohn *et al.*, 1969; Mathews and Cole, 1972). In chapter 4, the assembled capsids were further shown to sediment at rates which corresponded to recombinant MS2 capsids. These experiments showed that MS2 sub-genomic transcripts are capable of promoting formation of MS2 capsids *in vitro* and that truncation of the MS2 genome does not reduce its ability to be packaged. This suggests that the double shell tertiary structure of the genome is a consequence of packaging a large RNA into the capsid volume but not a requirement for capsid formation.

How is genomic RNA packaged? The GEMSA and sedimentation velocity assays of capsid assembly induced with the genomic RNAs of different sizes suggest that it is folded into the volume of the capsid. Smaller sub-genomic RNAs were observed to promote capsid formation at lower CP<sub>2</sub>:RNA reaction ratios than larger RNAs. This suggests that there is an energy barrier which must be overcome in order for these

RNAs to be packaged which is dependent upon the length of the genomic RNA. This is consistent with reducing their hydrodynamic radii. Estimates of the genomic RNAs hydrodynamic radii implied that a conformational change, wherein the RNAs are compacted, accompanies RNA packaging. The presence of Ethidium bromide (EtBr) in the assembly buffer resulted in a lower packaging efficiency of the 5' sub-genomic RNA, suggesting that reagents which rigidify RNA structure reduce packaging efficiency which is consistent with folding of RNA during packaging. In light of the cryo-EM reconstruction of the MS2 virion (Toropova *et al.*, 2008), these suggestions are in agreement with the idea that it will be harder to package a larger RNA due to a more complex fold which it must adopt within the capsid. The MS2 genome does not act as a static or prefolded scaffold upon which capsid assembly takes place. It appears that MS2 CP<sub>2</sub> subunits act as chaperones which bind and fold RNA into the volume of the capsid. This corresponds to proposed ideas of genomic packaging of various ssRNA plant and animal viruses (Lin *et al.*, 2003; Johnson *et al.*, 2004; Kuznetsov *et al.*, 2005). Another idea concerning the packaging of ssRNA viruses is that only nascent transcripts from the RNA polymerase are packaged into viral capsids. Reports on sub-genomic RNA encapsidation *in vivo* for the ssRNA viruses Brome mosaic virus (BMV) (Annamalai *et al.*, 2006) and Flock house virus (FHV) (Venter *et al.*, 2007) have suggested that only, replicating RNA is packaged. This corresponds to proposals that viral RNA is folded into an assembly competent structure immediately following replication (Hung *et al.*, 1969; Larson and McPherson, 2001). If capsid assembly follows this path *in vivo*, how or when does the RNA then fulfill its physiological roles in RNA replication and protein translation? The results reported in this thesis suggest that non-nascent sub-genomic RNA can be packaged and emphasize a chaperone activity of coat protein dimers. The packaging of viral ssRNA *in vivo* is however bound to be different from the scenario *in vitro*. The reports on FHV and BMV genome packaging *in vivo* suggest that capsid assembly could be affected by factors beyond coat protein-RNA interactions. The polymerase and other cellular machinery perhaps contribute to the assembly process. For MS2, an *in vivo* assembly component which has not been studied with respect to its affect on assembly is the A-protein. The A-protein binds to the MS2 genome near to its 5' and 3' termini effectively circularising the genome (Shiba and Suzuki, 1982). This is sure to restrict the conformational space available to the genome and increase capsid formation efficiency. For future work it

would be interesting to assess how the A-protein contributes to RNA packaging and capsid assembly.

The experiment with EtBr suggests that compounds capable of hindering correct RNA folding could be potent inhibitors of ssRNA virus assembly. It is however possible that EtBr perturbs binding of the CP<sub>2</sub> to RNA or association of CP<sub>2</sub>. For future work, an interesting experiment would be to monitor the formation of the CP<sub>2</sub>-TR complex with the CP mutant W82R which does not assemble beyond a dimer but binds TR. This would give ideas about how EtBr affects CP<sub>2</sub> binding to RNA. An assembly experiment with wild type CP<sub>2</sub> and TR in the presence and absence of EtBr would then indicate whether capsid assembly was less efficient due to perturbed CP<sub>2</sub> interactions. Additionally, different nucleotide intercalators could be assayed towards their ability to inhibit capsid formation which would indicate whether this effect is restricted to EtBr.

Capsid assembly induced with the sub-genomic RNAs was observed to take place in a co-operative manner. This was evident from the bimodal distribution of RNA and by the formation of capsids at low CP<sub>2</sub>:RNA reaction ratios as observed in GEMSA and sedimentation velocity assays of capsid assembly reactions. The encapsidation process of the sub-genomic RNAs is in this way similar to what has been described for the *in vitro* encapsidation of the full length MS2 genome (Sugiyama *et al.*, 1967). At high reaction ratios of CP<sub>2</sub>:RNA, approximately 50:1 and above, the formation of products larger than capsids became apparent. TEM micrographs indicated that these products most likely corresponded to incomplete capsid-like material. The data suggest that as the CP<sub>2</sub> concentration is increased, nonspecific binding of CP<sub>2</sub> to RNA increases, perhaps hindering folding of the RNA into an assembly competent structure. These structural kinetic traps on the assembly pathway are then prone to aggregation.

Specific regions of the MS2 genome were not observed to be of importance for successful RNA packaging. Sub-genomic RNAs of similar size but having a different sequence were packaged with similar efficiencies. Even in the absence of the TR operator, sub-genomic RNA was packaged. These results confirm reports that the TR operator is not required for inducing capsid assembly (Beckett and Uhlenbeck, 1988; Beckett *et al.*, 1988) and that there are secondary structure elements within the MS2 genome capable of quasi-equivalent conformer switching similar to TR (Toropova *et al.*, 2008). In chapter 4, results which formed part of a broader study to investigate the manner in which RNA stem-loops promote capsid assembly were reported. They showed that stem loops with low affinity to the CP<sub>2</sub> are capable of promoting capsid

assembly in a similar manner to TR. These results suggest that quasi-equivalent conformer control is not dependent on sequence specific binding and support the results obtained with the sub-genomic RNAs.

The results do however not suggest that RNA structure is not of importance for capsid assembly. Cryo-EM reconstructions obtained by a fellow PhD student, Katerina Toropova, of the capsid assembly products containing the sub-genomic RNAs indicate that they interact with the capsid at A/B dimer positions. This is consistent with CP<sub>2</sub> conformer switching occurring throughout the assembly process and is suggestive of a network of stem-loop structures within the MS2 genome. These results, along with those reported in chapters 3 and 4 of this thesis, are suggestive of a highly synergistic assembly process where CP<sub>2</sub> bind and fold the RNA while at the same time quasi-equivalent conformer switching is induced by the RNA at defined positions within the genome allowing  $T = 3$  shell formation (Rolfsson and Toropova *et al.*, 2009).

RNA structure is sure to have an effect on the capsid assembly process. Knowledge of the manner in which this might take place is likely to be gained by monitoring reaction kinetics. Gel filtration coupled light scattering and mass spectrometry have shown that capsid assembly kinetics can be different and that assembly can follow different molecular pathways dependent upon the type of RNA stem loop used to promote capsid formation (Basnak *et al.*, 2009). Sugiyama *et al.*, (1967) noticed that non-cognate MS2 RNAs do not produce capsid nucleation complexes such as those detected when assembly is initiated *in vitro* with the MS2 genome suggesting that RNA directs capsid assembly. For future work, it would be interesting to monitor the kinetics of cognate and non-cognate RNA folding during assembly. An informative experiment would be to “shuffle” the stem loop distribution of the MS2 genome which would disrupt the MS2 RNA stem loop template. If the current ideas on MS2 capsid assembly are correct then an RNA with such stem loop distribution would result in aggregate formation as A/B dimers would be formed at positions which would perhaps not be compatible with the  $T = 3$  shell. In effect the CP<sub>2</sub> binding site network, intrinsic to the MS2 genome, as suggested by the cryo-EM reconstructions would be disrupted.

In chapter 5, the structures of the sub-genomic RNAs in solution along with the MS2 genome within the virion were investigated on the basis of their susceptibility to lead ion induced hydrolysis. These experiments afforded insight into MS2 virion RNA structure and the RNA structural transition which must accompany MS2 genome

packaging. The MS2 virion RNA was shown to have a defined structure within the virion. This correlates with the cryo-EM reconstruction of the MS2 virion (Toropova *et al.*, 2008). The implication is that RNA is not packaged randomly into viral capsids and suggests a specific packaging mechanism for virion RNA.

The virion RNA structure was observed to be partly conserved in solution. The sub-genomic RNAs were shown to be structurally similar to both each other and to the virion RNA with respect to the nucleotide positions at which lead ion induced hydrolysis occurred. In light of the results presented in chapters 3 and 4, these findings suggest that viral RNA is folded into a tertiary structure from secondary structure motifs which are largely conserved in the virion. This is consistent with reports which have shown that upon removal of coat proteins from  $T = 3$  viruses such as Polio and BMV, their ssRNA slowly unwinds into a linear array of secondary structure elements (Kuznetsov *et al.*, 2005).

The lead ion structure probing assay showed promise in that it could represent a method in which all the CP<sub>2</sub> dimer binding sites within the MS2 genome could be identified. A difference in lead ion induced cleavage positions over the TR operator region was observed in the sub-genomic RNA transcripts *in vitro* as compared to the RNA within the MS2 virion. This suggested that the change in the structural environment of RNA associated with CP<sub>2</sub> binding can be detected with the lead ion cleavage assay. Mapping the CP<sub>2</sub> binding sites within the MS2 genome would be beneficial for understanding both the topology of the MS2 genome within the virion and would provide information on the scaffolding nature of the RNA during capsid assembly. A novel technique was pursued in order to structurally characterise the whole MS2 genome and all the sub-genomic RNAs with respect to lead ion induced cleavage in a high throughput manner using capillary electrophoresis. These results, although preliminary, were very promising and serve as a foundation for further high throughput structural characterisation of the MS2 genome. Work on this is currently underway.

## References

- Adhin, M. R., & van Duin, J. (1990). Scanning model for translational reinitiation in eubacteria. *Journal of Molecular Biology*, 213(4), 811-818.
- Aldovini, A., & Young, R. A. (1990). Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J Virol*, 64(5), 1920-1926.
- Annamalai, P., & Rao, A. L. (2007). In vivo packaging of brome mosaic virus RNA3, but not RNAs 1 and 2, is dependent on a cis-acting 3' tRNA-like structure. *J Virol*, 81(1), 173-181.
- Argetsinger, J. E., & Gussin, G. N. (1966). Intact ribonucleic acid from defective particles of bacteriophage R17. *Journal of Molecular Biology*, 21(3), 421-434.
- Baltimore, D. (1971). Expression of animal virus genomes. *Microbiol. Mol. Biol. Rev.*, 35(3), 235-241.
- Bancroft, J. B., Hiebert, E., & Bracker, C. E. (1969). The effects of various polyanions on shell formation of some spherical viruses. *Virology*, 39(4), 924-930.
- Basnak, G., Morton, V. L., Rolfsson, O., Stonehouse, N. J., Ashcroft, A. E., & Stockley, P. G. (2009). Viral genomic ssRNA directs the pathway towards a T=3 capsid. submitted.
- Beckett, D., & Uhlenbeck, O. C. (1988). Ribonucleoprotein complexes of R17 coat protein and a translational operator analog. *J Mol Biol*, 204(4), 927-938.
- Beckett, D., Wu, H. N., & Uhlenbeck, O. C. (1988). Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *J Mol Biol*, 204(4), 939-947.
- Belyi, V. A., & Muthukumar, M. (2006). Electrostatic origin of the genome packing in viruses. *Proc Natl Acad Sci U S A*, 103(46), 17174-17178.
- Beremand, M. N., & Blumenthal, T. (1979). Overlapping genes in rna phage: a new protein implicated in lysis. *Cell*, 18(2), 257-266.
- Berkhout, B., Schmidt, B. F., van Strien, A., van Boom, J., van Westrenen, J., & van Duin, J. (1987). Lysis gene of bacteriophage MS2 is activated by translation termination at the overlapping coat gene. *Journal of Molecular Biology*, 195(3), 517-524.
- Berkhout, B., & van Duin, J. (1985). Mechanism of translational coupling between coat protein and replicase genes of RNA bacteriophage MS2. *Nucleic Acids Res*, 13(19), 6955-6967.
- Bernardi, A., & Spahr, P. F. (1972). Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17. *Proc Natl Acad Sci U S A*, 69(10), 3033-3037.
- Bernhardt, T. G., Wang, I.-N., Struck, D. K., & Young, R. (2002). Breaking free: "Protein antibiotics" and phage lysis. *Research in Microbiology*, 153(8), 493-501.

- Bertolotti-Ciarlet, A., White, L. J., Chen, R., Prasad, B. V. V., & Estes, M. K. (2002). Structural Requirements for the Assembly of Norwalk Virus-Like Particles. *J. Virol.*, 76(8), 4044-4055.
- Betancourt, A. J. (2009). Genomewide patterns of substitution in adaptively evolving populations of the RNA bacteriophage MS2. *Genetics*, 181(4), 1535-1544.
- Bittman, R. (1969). Studies of the binding of ethidium bromide to transfer ribonucleic acid: absorption, fluorescence, ultracentrifugation and kinetic investigations. *J Mol Biol*, 46(2), 251-268.
- Blumenthal, T., Young, R. A., & Brown, S. (1976). Function and structure in phage Qbeta RNA replicase. Association of EF-Tu-Ts with the other enzyme subunits. *J Biol Chem*, 251(9), 2740-2743.
- Boedtker, H., & Gesteland, R. F. (1975). Physical properties of RNA bacteriophages and their RNA. In N. Zinder (Ed.), *RNA Phages: Cold Spring Harbour*.
- Boorstein, W. R., & Craig, E. A. (1989). Primer extension analysis of RNA. *Methods Enzymol*, 180, 347-369.
- Brakmann, S., & Grzeszik, S. (2001). An error-prone T7 RNA polymerase mutant generated by directed evolution. *Chembiochem*, 2(3), 212-219.
- Brosius, J., Dull, T. J., & Noller, H. F. (1980). Complete nucleotide sequence of a 23S ribosomal RNA gene from *Escherichia coli*. *Proc Natl Acad Sci U S A*, 77(1), 201-204.
- Brosius, J., Palmer, M. L., Kennedy, P. J., & Noller, H. F. (1978). Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. *Proc Natl Acad Sci U S A*, 75(10), 4801-4805.
- Brown, R. S., Hingerty, B. E., Dewan, J. C., & Klug, A. (1983). Pb(II)-catalysed cleavage of the sugar-phosphate backbone of yeast tRNA<sup>Phe</sup>--implications for lead toxicity and self-splicing RNA. *Nature*, 303(5917), 543-546.
- Calendar, R. (2006). *The Bacteriophages (Vol. second)*. Oxford: Oxford University Press.
- Calhoun, S. L., Speir, J. A., & Rao, A. L. (2007). In vivo particle polymorphism results from deletion of a N-terminal peptide molecular switch in brome mosaic virus capsid protein. *Virology*, 364(2), 407-421.
- Capecci, M. R., & Gussin, G. N. (1965). Suppression in vitro: Identification of a Serine-sRNA as a "Nonsense" Suppressor. *Science*, 149(3682), 417-422.
- Carey, J., Cameron, V., de Haseth, P. L., & Uhlenbeck, O. C. (1983). Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. *Biochemistry*, 22(11), 2601-2610.

- Carey, J., & Uhlenbeck, O. C. (1983). Kinetic and thermodynamic characterization of the R17 coat protein-ribonucleic acid interaction. *Biochemistry*, 22(11), 2610-2615.
- Casini, G. L., Graham, D., Heine, D., Garcea, R. L., & Wu, D. T. (2004). In vitro papillomavirus capsid assembly analyzed by light scattering. *Virology*, 325(2), 320-327.
- Caspar, D. L. (1956). Structure of bushy stunt virus. *Nature*, 177(4506), 475-476.
- Caspar, D. L., & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol*, 27, 1-24.
- Chamberlin, M., McGrath, J., & Waskell, L. (1970). New RNA Polymerase from *Escherichia coli* infected with Bacteriophage T7. *Nature*, 228(5268), 227-231.
- Ciesiolka, J., Michalowski, D., Wrzesinski, J., Krajewski, J., & Krzyzosiak, W. J. (1998). Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs. *J Mol Biol*, 275(2), 211-220.
- Crawford, E. M., & Gesteland, R. F. (1964). The adsorption of bacteriophage R-17. *Virology*, 22(1), 165-167.
- Crick, F. H., & Watson, J. D. (1956). Structure of small viruses. *Nature*, 177(4506), 473-475.
- Deigan, K. E., Li, T. W., Mathews, D. H., & Weeks, K. M. (2009). Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A*, 106(1), 97-102.
- Devos, R., van Emmelo, J., Contreras, R., & Fiers, W. (1979). Construction and characterization of a plasmid containing a nearly full-size DNA copy of bacteriophage MS2 RNA. *J Mol Biol*, 128(4), 595-619.
- Domingo, E., & Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annu Rev Microbiol*, 51, 151-178.
- Erickson, J. W., & Rossmann, M. G. (1982). Assembly and crystallization of a T = 1 icosahedral particle from trypsinized southern bean mosaic virus coat protein. *Virology*, 116(1), 128-136.
- Farkas, W. R. (1968). Depolymerization of ribonucleic acid by plumbous ion. *Biochim Biophys Acta*, 155(2), 401-409.
- Fiers, W. (1967). Studies on the bacteriophage MS2. 3. Sedimentation heterogeneity of viral RNA preparations. *Virology*, 33(3), 413-424.
- Fiers, W. (1975). Chemical Structure and Biological Activity of Bacteriophage MS2 RNA. In N. Zinder (Ed.), *RNA Phages*. New York: Cold Spring Harbour Laboratories.
- Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G., & Ysebaert, M. (1976). Complete nucleotide sequence of bacteriophage MS2 RNA:

primary and secondary structure of the replicase gene. *Nature*, 260(5551), 500-507.

Fisher, A. J., & Johnson, J. E. (1993). Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. *Nature*, 361(6408), 176-179.

Freddolino, P. L., Arkhipov, A. S., Larson, S. B., McPherson, A., & Schulten, K. (2006). Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14(3), 437-449.

Fresco, J. R., Alberts, B. M., & Doty, P. (1960). Some molecular details of the secondary structure of ribonucleic acid. *Nature*, 188, 98-101.

Fukuma, I., & Cohen, S. S. (1975). Polyamines in bacteriophage R17 and its RNA. *J Virol*, 16(2), 222-227.

Gell, C., Sabir, T., Westwood, J., Rashid, A., Smith, D. A. M., Harris, S.-A., & Stockley, P. G. (2008). Single-Molecule Fluorescence Resonance Energy Transfer Assays Reveal Heterogeneous Folding Ensembles in a Simple RNA Stem-Loop. *Journal of Molecular Biology*, 384(1), 264-278.

Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, 182(2), 319-326.

Golmohammadi, R., Valegard, K., Fridborg, K., & Liljas, L. (1993). The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J Mol Biol*, 234(3), 620-639.

Grahn, E., Moss, T., Helgstrand, C., Fridborg, K., Sundaram, M., Tars, K., Lago, H., Stonehouse, N. J., Davis, D. R., Stockley, P. G., & Liljas, L. (2001). Structural basis of pyrimidine specificity in the MS2 RNA hairpin-coat-protein complex. *RNA*, 7(11), 1616-1627.

Gralla, J., Steitz, J. A., & Crothers, D. M. (1974). Direct physical evidence for secondary structure in an isolated fragment of R17 bacteriophage mRNA. *Nature*, 248(445), 204-208.

Groeneveld, H., Thimon, K., & van Duin, J. (1995). Translational control of maturation-protein synthesis in phage MS2: a role for the kinetics of RNA folding? *RNA*, 1(1), 79-88.

Harrison, S. C. (2001). The familiar and the unexpected in structures of icosahedral viruses. *Current Opinion in Structural Biology*, 11(2), 195-199.

Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K., & Bricogne, G. (1978). Tomato bushy stunt virus at 2.9 Å resolution. *Nature*, 276(5686), 368-373.

Herrmann, R., Schubert, D., & Rudolph, U. (1968). Self-assessment of protein subunits from bacteriophage fr. *Biochem Biophys Res Commun*, 30(5), 576-581.

Hill, H. R. (1993). Molecular analysis and applications of bacteriophage MS2 structural proteins. University of Leeds, Leeds.

- Hill, H. R., Stonehouse, N. J., Fonseca, S. A., & Stockley, P. G. (1997). Analysis of phage MS2 coat protein mutants expressed from a reconstituted phagemid reveals that proline 78 is essential for viral infectivity. *J Mol Biol*, 266(1), 1-7.
- Hogle, J. M., Maeda, A., & Harrison, S. C. (1986). Structure and assembly of turnip crinkle virus. I. X-ray crystallographic structure analysis at 3.2 Å resolution. *J Mol Biol*, 191(4), 625-638.
- Hohn, T. (1969). Role of RNA in the assembly process of bacteriophage fr. *J Mol Biol*, 43(1), 191-200.
- Horn, W. T., Tars, K., Grahn, E., Helgstrand, C., Baron, A. J., Lago, H., Adams, C. J., Peabody, D. S., Phillips, S. E., Stonehouse, N. J., Liljas, L., & Stockley, P. G. (2006). Structural basis of RNA binding discrimination between bacteriophages Qbeta and MS2. *Structure*, 14(3), 487-495.
- Howlett, G. J., Minton, A. P., & Rivas, G. (2006). Analytical ultracentrifugation for the study of protein association and assembly. *Curr Opin Chem Biol*, 10(5), 430-436.
- Hung, P. P., Ling, C. M., & Overby, L. R. (1969). Self-assembly of Q-beta and MS2 phage particles: possible function of initiation complexes. *Science*, 166(913), 1638-1640.
- Huntzinger, E., Possedko, M., Winter, F., Moine, H., Ehresmann, C., & Romby, P. (2005). Probing RNA Structures with Enzymes and Chemicals in vitro and in vivo. In K. A. Hartman & A. Bindereif & A. Schon & E. Westhof (Eds.), *Handbook of RNA Biochemistry* (Vol. 1). Weinheim: Wiley-VCH.
- Ironmonger, A., Whittaker, B., Baron, A. J., Clique, B., Adams, C. J., Ashcroft, A. E., Stockley, P. G., & Nelson, A. (2007). Scanning conformational space with a library of stereo- and regiochemically diverse aminoglycoside derivatives: the discovery of new ligands for RNA hairpin sequences. *Org Biomol Chem*, 5(7), 1081-1086.
- Jacobson, A. B. (1976). Studies on secondary structure of single-stranded RNA from bacteriophage MS2 by electron microscopy. *Proc Natl Acad Sci U S A*, 73(2), 307-311.
- Johnson, J. E., & Speir, J. A. (1997). Quasi-equivalent viruses: a paradigm for protein assemblies. *J Mol Biol*, 269(5), 665-675.
- Johnson, J. M., Willits, D. A., Young, M. J., & Zlotnick, A. (2004). Interaction with capsid protein alters RNA structure and the pathway for in vitro assembly of Cowpea chlorotic mottle virus. *Journal of Molecular Biology*, 335(2), 455-464.
- Kastelein, R. A., Remaut, E., Fiers, W., & van Duin, J. (1982). Lysis gene expression of RNA phage MS2 depends on a frameshift during translation of the overlapping coat protein gene. *Nature*, 295(5844), 35-41.
- Kirsebom, L. A., Ciesiolka, J. (2005). Pb<sup>2+</sup> induced Cleavage of RNA. In R. K. Hartmann, Bindereif, A., Schön, A., Westhof, E. (Ed.), *Handbook of RNA*

Biochemistry (Vol. 1). Weinheim: Wiley-VCH.

Kochetkov, S. N., Rusakova, E. E., & Tunitskaya, V. L. (1998). Recent studies of T7 RNA polymerase mechanism. *FEBS Lett*, 440(3), 264-267.

Koning, R., van den Worm, S., Plaisier, J. R., van Duin, J., Pieter Abrahams, J., & Koerten, H. (2003). Visualization by cryo-electron microscopy of genomic RNA that binds to the protein capsid inside bacteriophage MS2. *J Mol Biol*, 332(2), 415-422.

Kozak, M., & Nathans, D. (1971). Fate of maturation protein during infection by coliphage MS2. *Nat New Biol*, 234(50), 209-211.

Krol, M. A., Olson, N. H., Tate, J., Johnson, J. E., Baker, T. S., & Ahlquist, P. (1999). RNA-controlled polymorphism in the *in vivo* assembly of 180-subunit and 120-subunit virions from a single capsid protein. *Proceedings of the National Academy of Sciences of the United States of America*, 96(24), 13650-13655.

Kuznetsov, Y. G., Daijogo, S., Zhou, J., Semler, B. L., & McPherson, A. (2005). Atomic force microscopy analysis of icosahedral virus RNA. *Journal of Molecular Biology*, 347(1), 41-52.

Kuznetsov, Y. G., & McPherson, A. (2006). Atomic force microscopy investigation of Turnip Yellow Mosaic Virus capsid disruption and RNA extrusion. *Virology*, 352(2), 329-337.

Lago, H., Fonseca, S. A., Murray, J. B., Stonehouse, N. J., & Stockley, P. G. (1998). Dissecting the key recognition features of the MS2 bacteriophage translational repression complex. *Nucleic Acids Res*, 26(5), 1337-1344.

Larson, S. B., & McPherson, A. (2001). Satellite tobacco mosaic virus RNA: structure and implications for assembly. *Curr Opin Struct Biol*, 11(1), 59-65.

Layne, E. (1957). Spectrophotometric and turbidimetric methods for measuring proteins. *Methods in Enzymology*, 3, 447-454.

Lebowitz, J., Lewis, M. S., & Schuck, P. (2002). Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci*, 11(9), 2067-2079.

Leipold, B. (1977). Effect of spermidine on the RNA-A protein complex isolated from the RNA bacteriophage MS2. *J Virol*, 21(2), 445-450.

Leipold, B., & Hofschneider, P. H. (1976). Characterization of an RNA-a protein complex isolated from the RNA bacteriophage M12. *J Virol*, 19(3), 792-800.

LePecq, J. B., & Paoletti, C. (1967). A fluorescent complex between ethidium bromide and nucleic acids. Physical-chemical characterization. *J Mol Biol*, 27(1), 87-106.

Liljas, L., Fridborg, K., Valegard, K., Bundule, M., & Pumpens, P. (1994). Crystal Structure of Bacteriophage fr Capsids at 3.5 Å Resolution. *Journal of Molecular Biology*, 244(3), 279-290.

- Lin, T. W., Cavarelli, J., & Johnson, J. E. (2003). Evidence for assembly-dependent folding of protein and RNA in an icosahedral virus. *Virology*, 314(1), 26-33.
- Ling, C. M., Hung, P. P., & Overby, L. R. (1970). Independent assembly of Q[beta] and MS2 phages in doubly infected *Escherichia coli*. *Virology*, 40(4), 920-929.
- Lodish, H. F. (1968). Bacteriophage f2 RNA: control of translation and gene order. *Nature*, 220(5165), 345-350.
- Lodish, H. F., Horiuchi, K., & Zinder, N. D. (1965). Mutants of the bacteriophage f2 : V. On the production of noninfectious phage particles. *Virology*, 27(2), 139-155.
- Lodish, H. F., & Zinder, N. D. (1966). Mutants of the bacteriophage f2: VIII. Control mechanisms for phage-specific syntheses. *Journal of Molecular Biology*, 19(2), 333-348.
- Loeb, T., & Zinder, N. D. (1961). A bacteriophage containing RNA. *Proc Natl Acad Sci U S A*, 47, 282-289.
- Lucas, R. W., Larson, S. B., & McPherson, A. (2002). The crystallographic structure of brome mosaic virus. *J Mol Biol*, 317(1), 95-108.
- Matthews, K. S., & Cole, R. D. (1972). Shell formation by capsid protein of f2 bacteriophage. *Journal of Molecular Biology*, 65(1), 1-4.
- Min Jou, W., Haegeman, G., Ysebaert, M., & Fiers, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature*, 237(5350), 82-88.
- Mitra, S., Shcherbakova, I. V., Altman, R. B., Brenowitz, M., & Laederach, A. (2008). High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res*, 36(11), e63.
- Moine, S., Ehresmann, B., Ehresmann, C., & Romby, P. (1998). Probind RNA structure and Function in Solution. In R. W. Simons & M. Grundberg-Manago (Eds.), *RNA Structure and Function* (pp. 77): Cold Spring Harbour Laboratory Press.
- Nguyen, H. D., Reddy, V. S., & Brooks, C. L. (2007). Deciphering the Kinetic Mechanism of Spontaneous Self-Assembly of Icosahedral Capsids. *Nano Lett.*, 7(2), 338-344.
- Osborn, M., Weiner, A. M., & Weber, K. (1970). Large scale purification of A-protein from bacteriophage R17. *Eur J Biochem*, 17(1), 63-67.
- Overby, L. R., Barlow, G. H., Doi, R. H., Jacob, M., & Spiegelman, S. (1966). Comparison of two serologically distinct ribonucleic acid bacteriophages. I. Properties of the viral particles. *J Bacteriol*, 91(1), 442-448.
- Pasloske, B. L., Walkerpeach, C. R., Obermoeller, R. D., Winkler, M., & DuBois, D. B. (1998). Armored RNA technology for production of ribonuclease-resistant viral RNA controls and standards. *Journal Of Clinical Microbiology*, 36(12), 3590-3594.

- Peabody, D. S. (1993). The RNA binding site of bacteriophage MS2 coat protein. *Embo J*, 12(2), 595-600.
- Plevka, P., Kazaks, A., Voronkova, T., Kotelovica, S., Dishlers, A., Liljas, L., & Tars, K. (2009). The structure of bacteriophage phiCb5 reveals a role of the RNA genome and metal ions in particle stability and assembly. *J Mol Biol*, 391(3), 635-647.
- Prasad, B. V. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G., & Estes, M. K. (1999). X-ray crystallographic structure of the Norwalk virus capsid. *Science*, 286(5438), 287-290.
- Prevelige, P. E., Jr., Thomas, D., & King, J. (1993). Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys J*, 64(3), 824-835.
- Ralston, G. An introduction to analytical ultracentrifugation: Beckman, [www.Beckmancoulter.com](http://www.Beckmancoulter.com)
- Reddy, V. S., Giesing, H. A., Morton, R. T., Kumar, A., Post, C. B., Brooks, C. L., & Johnson, J. E. (1998). Energetics of quasiequivalence: Computational analysis of protein-protein interactions in icosahedral viruses. *Biophysical Journal*, 74(1), 546-558.
- Remaut, E., Stanssens, P., & Fiers, W. (1981). Plasmid vectors for high-efficiency expression controlled by the PL promoter of coliphage lambda. *Gene*, 15(1), 81-93.
- Richelson, E., & Nathans, D. (1967). Association of bacteriophage proteins and RNA in , infected with MS2. *Biochemical and Biophysical Research Communications*, 29(6), 842-849.
- Rohde, N., Daum, H., & Biebricher, C. K. (1995). The Mutant Distribution of an RNA Species Replicated by Q[beta] Replicase. *Journal of Molecular Biology*, 249(4), 754-762.
- Rolfsson, O., Toropova, K., Morton, V., Francese, S., Basnak, G., S. Thompson, G., Homans, S. W., Ashcroft, A. E., Stonehouse, N. J., Ranson, N. A., & Stockley, P. G. (2008). RNA packing specificity and folding during assembly of the bacteriophage MS2. *Computational and Mathematical Methods in Medicine: An Interdisciplinary Journal of Mathematical, Theoretical and Clinical Aspects of Medicine*, 9(3), 339 - 349.
- Rolfsson, O., Toropova, K., Ranson, N. A., & Stockley, P. G. (2009). Mutually-induced conformational switching of RNA and coat protein during assembly of a viral capsid. submitted.
- Romaniuk, P. J., Lowary, P., Wu, H. N., Stormo, G., & Uhlenbeck, O. C. (1987). RNA binding site of R17 coat protein. *Biochemistry*, 26(6), 1563-1568.
- Rossmann, M. G., Abad-Zapatero, C., Murthy, M. R., Liljas, L., Jones, T. A., & Strandberg, B. (1983). Structural comparisons of some small spherical plant viruses. *J Mol Biol*, 165(4), 711-736.

- Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., & et al. (1985). Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature*, 317(6033), 145-153.
- Rossmann, M. G., & Johnson, J. E. (1989). Icosahedral RNA Virus Structure. *Annual Review of Biochemistry*, 58(1), 533-569.
- Rowse, S., Stonehouse, N. J., Convery, M. A., Adams, C. J., Ellington, A. D., Hirao, I., Peabody, D. S., Stockley, P. G., & Phillips, S. E. (1998). Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nat Struct Biol*, 5(11), 970-975.
- Rudnick, J., & Bruinsma, R. (2005). Icosahedral packing of RNA viral genomes. *Phys Rev Lett*, 94(3), 038101.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(4839), 487-491.
- Sambrook, J. M., P. Russel, D. (2006). *The Condensed Protocols From Molecular Cloning: A Laboratory Manual*: Cold Spring Harbor.
- Satheshkumar, P. S., Lokesh, G. L., Murthy, M. R. N., & Savithri, H. S. (2005). The role of arginine-rich motif and beta-annulus in the assembly and stability of Sesbania mosaic virus capsids. *Journal of Molecular Biology*, 353(2), 447-458.
- Schneemann, A. (2006). The structural and functional role of RNA in icosahedral virus assembly. *Annu Rev Microbiol*, 60, 51-67.
- Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J*, 78(3), 1606-1619.
- Shaklee, P. N. (1990). Negative-strand RNA replication by Q beta and MS2 positive-strand RNA phage replicases. *Virology*, 178(1), 340-343.
- Shcherbakova, I., Mitra, S., Laederach, A., & Brenowitz, M. (2008). Energy barriers, pathways, and dynamics during folding of large, multidomain RNAs. *Curr Opin Chem Biol*, 12(6), 655-666.
- Shiba, T., & Suzuki, Y. (1981). Localization of A protein in the RNA-A protein complex of RNA phage MS2. *Biochim Biophys Acta*, 654(2), 249-255.
- Skripkin, E. A., Adhin, M. R., de Smit, M. H., & van Duin, J. (1990). Secondary structure of the central region of bacteriophage MS2 RNA. Conservation and biological significance. *J Mol Biol*, 211(2), 447-463.
- Sobell, H. M., Tsai, C.-C., Jain, S. C., & Gilbert, S. G. (1977). Visualization of drug-nucleic acid interactions at atomic resolution : III. Unifying structural concepts in understanding drug-DNA interactions and their broader implications in understanding protein-DNA interactions. *Journal of Molecular Biology*, 114(3), 333-365.

Sorger, P. K., Stockley, P. G., & Harrison, S. C. (1986). Structure and assembly of turnip crinkle virus. II. Mechanism of reassembly in vitro. *J Mol Biol*, 191(4), 639-658.

Stafford, W. F. (1997). Sedimentation velocity spins a new weave for an old fabric. *Curr Opin Biotechnol*, 8(1), 14-24.

Steitz, J. A. (1968a). Identification of the A protein as a structural component of bacteriophage R17. *J Mol Biol*, 33(3), 923-936.

Steitz, J. A. (1968b). Isolation of the A protein from bacteriophage R17. *J Mol Biol*, 33(3), 937-945.

Steitz, J. A. (1973). Discriminatory ribosome rebinding of isolated regions of protein synthesis initiation from the ribonucleic acid of bacteriophage R17. *Proc Natl Acad Sci U S A*, 70(9), 2605-2609.

Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J., Homans, S. W., & Ashcroft, A. E. (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. *J Mol Biol*, 369(2), 541-552.

Stockley, P. G., Stonehouse, N. J., Murray, J. B., Goodman, S. T., Talbot, S. J., Adams, C. J., Liljas, L., & Valegard, K. (1995). Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. *Nucleic Acids Res*, 23(13), 2512-2518.

Strauss, J. H., Jr., & Sinsheimer, R. L. (1963). Purification and properties of bacteriophage MS2 and of its ribonucleic acid. *J Mol Biol*, 7, 43-54.

Sugiyama, T., Hebert, R. R., & Hartman, K. A. (1967). Ribonucleoprotein complexes formed between bacteriophage MS2 RNA and MS2 Protein in vitro. *Journal of Molecular Biology*, 25(3), 455-460.

Summers, W. C. (2006). Phage and the early development of molecular biology In R. Calendar (Ed.), *The Bacteriophages* (Vol. second). Oxford: Oxford University Press.

Tang, J. H., Johnson, J. M., Dryden, K. A., Young, M. J., Zlotnick, A., & Johnson, J. E. (2006). The role of subunit hinges and molecular "switches" in the control of viral capsid polymorphism. *Journal of Structural Biology*, 154(1), 59-67.

Tang, L., Johnson, K. N., Ball, L. A., Lin, T., Yeager, M., & Johnson, J. E. (2001). The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nat Struct Biol*, 8(1), 77-83.

Taniguchi, T., Palmieri, M., & Weissmann, C. (1978). A Qbeta DNA-containing hybrid plasmid giving rise to Qbeta phage formation in the bacterial host [proceedings]. *Ann Microbiol (Paris)*, 129 B(4), 535-536.

Thomas, G. J., Jr., Prescott, B., McDonald-Ordzie, P. E., & Hartman, K. A. (1976). Studies of virus structure by laser-Raman spectroscopy. II. MS2 phage, MS2 capsids

and MS2 RNA in aqueous solutions. *J Mol Biol*, 102(1), 103-124.

Tihova, M., Dryden, K. A., Le, T. V. L., Harvey, S. C., Johnson, J. E., Yeager, M., & Schneemann, A. (2004). Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. *Journal of Virology*, 78(6), 2897-2905.

Tinoco, I., Jr., & Bustamante, C. (1999). How RNA folds. *J Mol Biol*, 293(2), 271-281.

Tinoco, I., Jr., Sauer, K., & Wang, J. C. (1995). *Physical Chemistry Principles and Applications in Biological Sciences*. New Jersey: Prentice Hall.

Toropova, K. (2009). *Cryo-electron microscopy of bacteriophage MS2*. University of Leeds, Leeds.

Toropova, K., Basnak, G., Twarock, R., Stockley, P. G., & Ranson, N. A. (2008). The Three-dimensional Structure of Genomic RNA in Bacteriophage MS2: Implications for Assembly. *Journal of Molecular Biology*, 375(3), 824-836.

Turner, K. B., Hagan, N. A., & Fabris, D. (2006). Inhibitory effects of archetypical nucleic acid ligands on the interactions of HIV-1 nucleocapsid protein with elements of Psi-RNA. *Nucleic Acids Res*, 34(5), 1305-1316.

Valegard, K., Liljas, L., Fridborg, K., & Unge, T. (1990). The three-dimensional structure of the bacterial virus MS2. *Nature*, 345(6270), 36-41.

Valegard, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J., & Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature*, 371(6498), 623-626.

Valegard, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G., & Liljas, L. (1997). The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J Mol Biol*, 270(5), 724-738.

van den Worm, S. H., Koning, R. I., Warmenhoven, H. J., Koerten, H. K., & van Duin, J. (2006). Cryo electron microscopy reconstructions of the Leviviridae unveil the densest icosahedral RNA packing possible. *J Mol Biol*, 363(4), 858-865.

van der Schoot, P., & Bruinsma, R. (2005). Electrostatics and the assembly of an RNA virus. *Phys Rev E Stat Nonlin Soft Matter Phys*, 71(6 Pt 1), 061928.

Van Duin, J., & Tsareva, N. (2006). Single-stranded RNA phages. In R. Calendar (Ed.), *The Bacteriophages* (second ed.). Oxford: Oxford University Press.

Vasa, S. M., Guex, N., Wilkinson, K. A., Weeks, K. M., & Giddings, M. C. (2008). ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*, 14(10), 1979-1990.

Venkataraman, S., Reddy, S. P., Loo, J., Idamakanti, N., Hallenbeck, P. L., & Reddy, V.

- S. (2008). Structure of Seneca Valley Virus-001: an oncolytic picornavirus representing a new genus. *Structure*, 16(10), 1555-1561.
- Venter, P. A., & Schneemann, A. (2007). Assembly of two independent populations of flock house virus particles with distinct RNA packaging characteristics in the same cell. *Journal of Virology*, 81(2), 613-619.
- Verma, I. M. (1977). The reverse transcriptase. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 473(1), 1-38.
- Wagner, E. K., Hewlett, M. J., Bloom, D. C., & Camerini, D. (2008). *Basic Virology*: Blackwell Publishing Ltd.
- Wagner, G. W., & Bancroft, J. B. (1971). The assembly of RNA-free mixed-coat capsids of small spherical viruses. *Virology*, 45(1), 321-323.
- Ward, R., Shive, K., & Valentine, R. (1967). Capsid protein of f2 as translational repressor. *Biochem Biophys Res Commun*, 29(1), 8-13.
- Waring, M. J. (1965). Complex formation between ethidium bromide and nucleic acids. *J Mol Biol*, 13(1), 269-282.
- Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr., Swanstrom, R., Burch, C. L., & Weeks, K. M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, 460(7256), 711-716.
- Weeks, K. M. (1997). Protein-facilitated RNA folding. *Curr Opin Struct Biol*, 7(3), 336-342.
- Werner, C., Krebs, B., Keith, G., & Dirheimer, G. (1976). Specific cleavages of pure tRNAs by plumbous ions. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 432(2), 161-175.
- Wery, J.-P., Reddy, V. S., Hosur, M. V., & Johnson, J. E. (1994). The Refined Three-Dimensional Structure of an Insect Virus at 2.8 Å Resolution. *Journal of Molecular Biology*, 235(2), 565-586.
- White, S. A., & Draper, D. E. (1987). Single base bulges in small RNA hairpins enhance ethidium binding and promote an allosteric transition. *Nucleic Acids Res*, 15(10), 4049-4064.
- Wilkinson, K. A., Gorelick, R. J., Vasa, S. M., Guex, N., Rein, A., Mathews, D. H., Giddings, M. C., & Weeks, K. M. (2008). High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol*, 6(4), e96.
- Williams, A., & Westhead, D. R. (2002). Sequence relationships in the legume lectin fold and other jelly rolls. *Protein Eng.*, 15(10), 771-774.
- Winter, D., Polacek, N., Halama, I., Streicher, B., & Barta, A. (1997). Lead-catalysed

specific cleavage of ribosomal RNAs. *Nucleic Acids Res*, 25(9), 1817-1824.

Yi, Q.-M., & Wong, K.-P. (1982). The effects of magnesium ions on the hydrodynamic shape, conformation, and stability of the ribosomal 23S RNA from. *Biochemical and Biophysical Research Communications*, 104(2), 733-739.

Yoffe, A. M., Prinsen, P., Gopal, A., Knobler, C. M., Gelbart, W. M., & Ben-Shaul, A. (2008). Predicting the sizes of large RNA molecules. *Proc Natl Acad Sci U S A*, 105(42), 16153-16158.

Yonesaki, T., Furuse, K., Haruna, I., & Watanabe, I. (1982). Relationships among four groups of RNA coliphages based on the template specificity of GA replicase. *Virology*, 116(1), 379-381.

Yonesaki, T., & Haruna, I. (1981). In Vitro Replication of Bacteriophage GA RNA. Subunit Structure and Catalytic Properties of GA Replicase. *J Biochem*, 89(3), 741-750.

Zhang, T., & Schwartz, R. (2006). Simulation study of the contribution of oligomer/oligomer binding to capsid assembly kinetics. *Biophys J*, 90(1), 57-64.

Zhong, W., Dasgupta, R., & Rueckert, R. (1992). Evidence that the packaging signal for nodaviral RNA2 is a bulged stem-loop. *Proc Natl Acad Sci U S A*, 89(23), 11146-11150.

Zlotnick, A. (1994). To Build a Virus Capsid - an Equilibrium-Model of the Self-Assembly of Polyhedral Protein Complexes. *Journal of Molecular Biology*, 241(1), 59-67.

Zlotnick, A. (2005). Theoretical aspects of virus capsid assembly. *J Mol Recognit*, 18(6), 479-490.

Zlotnick, A., Aldrich, R., Johnson, J. M., Ceres, P., & Young, M. J. (2000). Mechanism of Capsid Assembly for an Icosahedral Plant Virus. *Virology*, 277(2), 450-456.

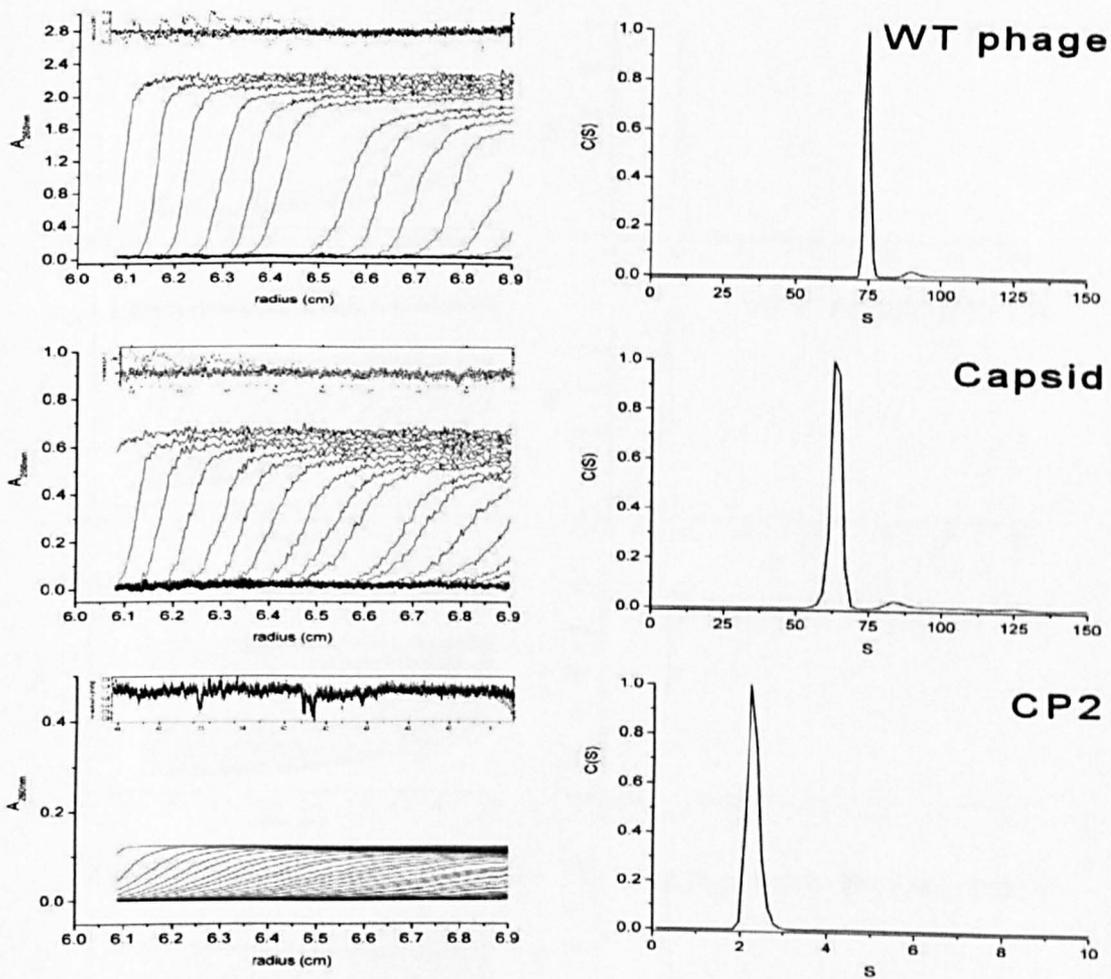
Zlotnick, A., & Stray, S. J. (2003). How does your virus grow? Understanding and interfering with virus assembly. *Trends in Biotechnology*, 21(12), 536-542.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, 31(13), 3406-3415.

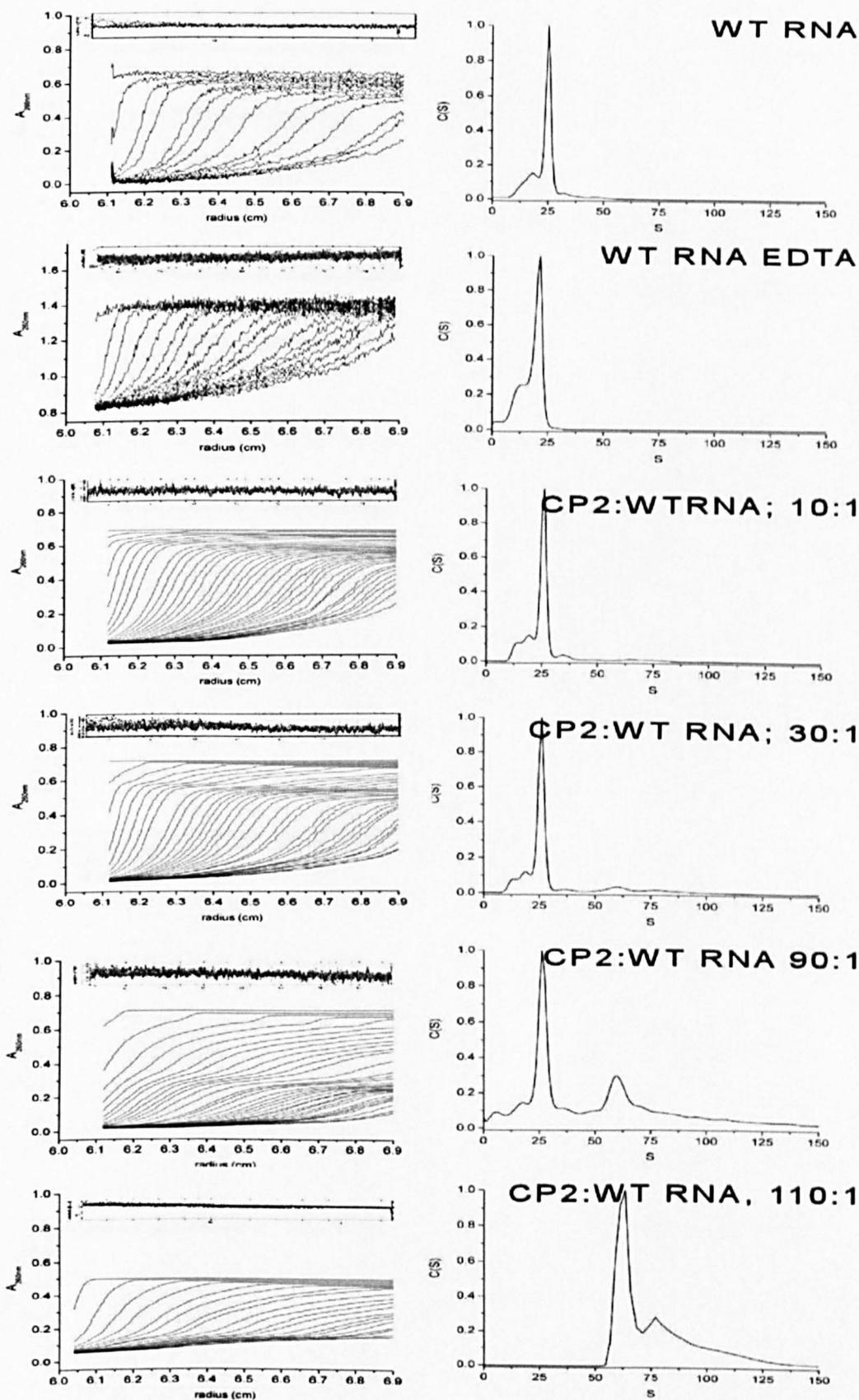
## Appendix

### 1. Analytical ultracentrifugation data supplementary

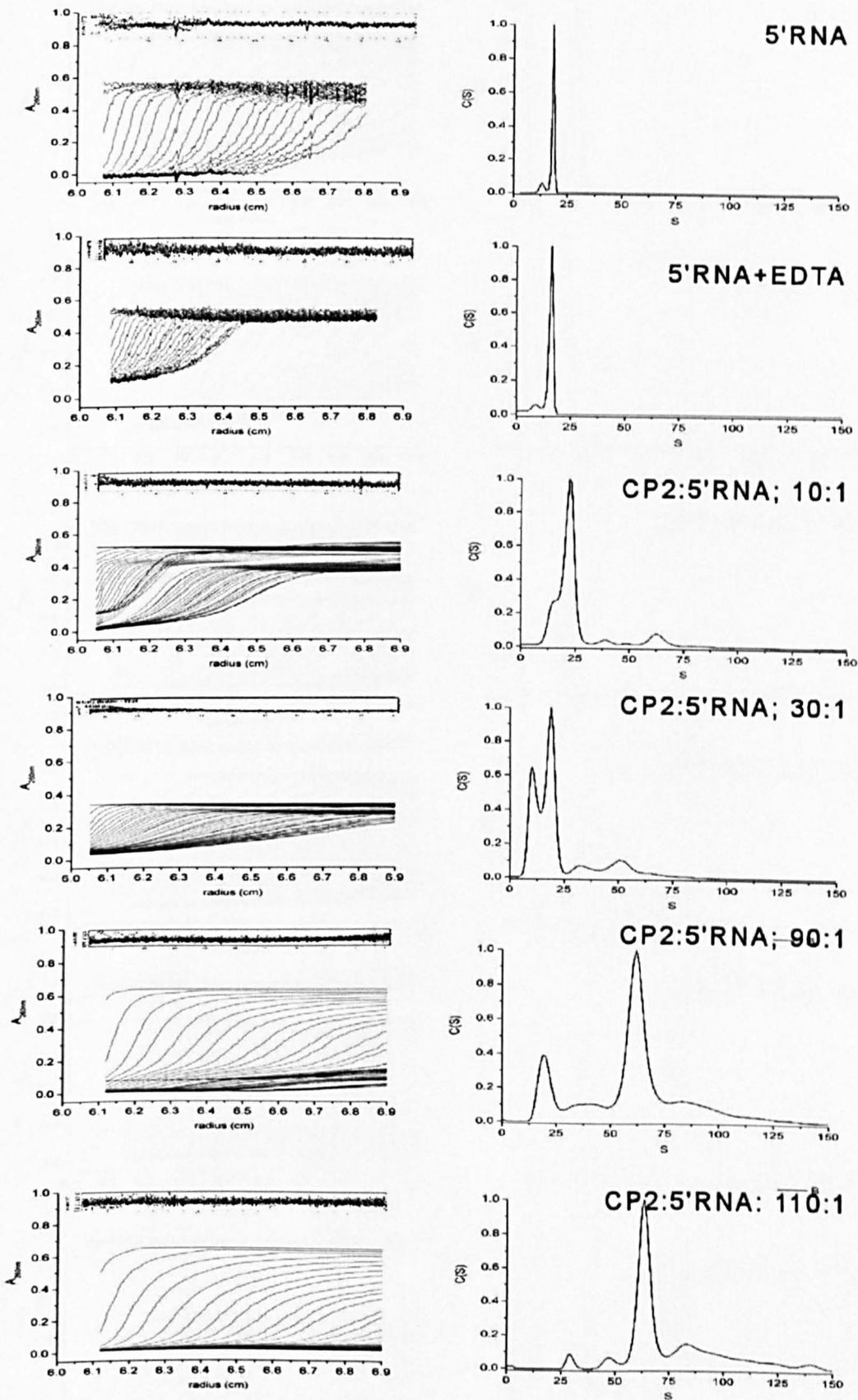
The following six figures, supplementary figures 1-6, show examples of the sedimentation velocity profiles obtained for all MS2 capsid assembly components and assembly reactions. In the figures the sedimentation velocity profiles are shown on the left with the C(S) plot, obtained by fitting the sedimentation profiles using the C(S) model in the program SEDFIT as described in section 2.2.12, shown on the right. In some cases the sedimentation velocity profiles have been smoothed for clarity. Note that in some cases specific sedimentation velocity profiles are missing such as for the WT phage in AUC supplementary figure-1 (no scan visible at radius ~6.4 cm). This is because of a malfunction in the connection between the X1-1 Beckmann analytical ultracentrifuge and the data acquisition software. This malfunction does not affect data analysis. Above the sedimentation profiles are shown the residuals of fitting to the C(S) model. They are shown as an absorbance vs. radius plot. The scale on the y-axis of the residual plots span  $\pm 0.05$  AU from zero as most residuals were within that range.



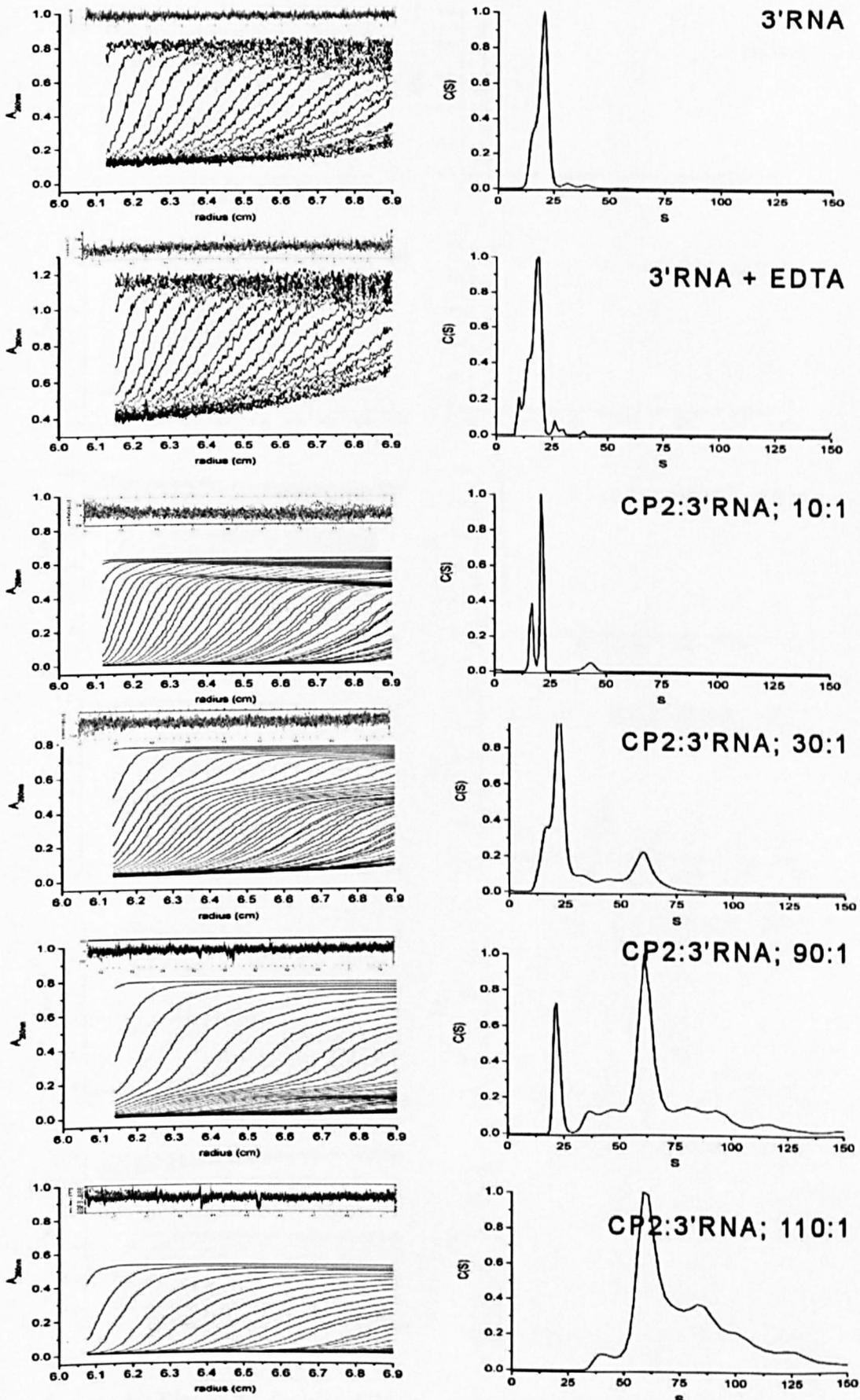
*Supplementary Figure-1. WT phage, recombinant capsid and coat protein.*



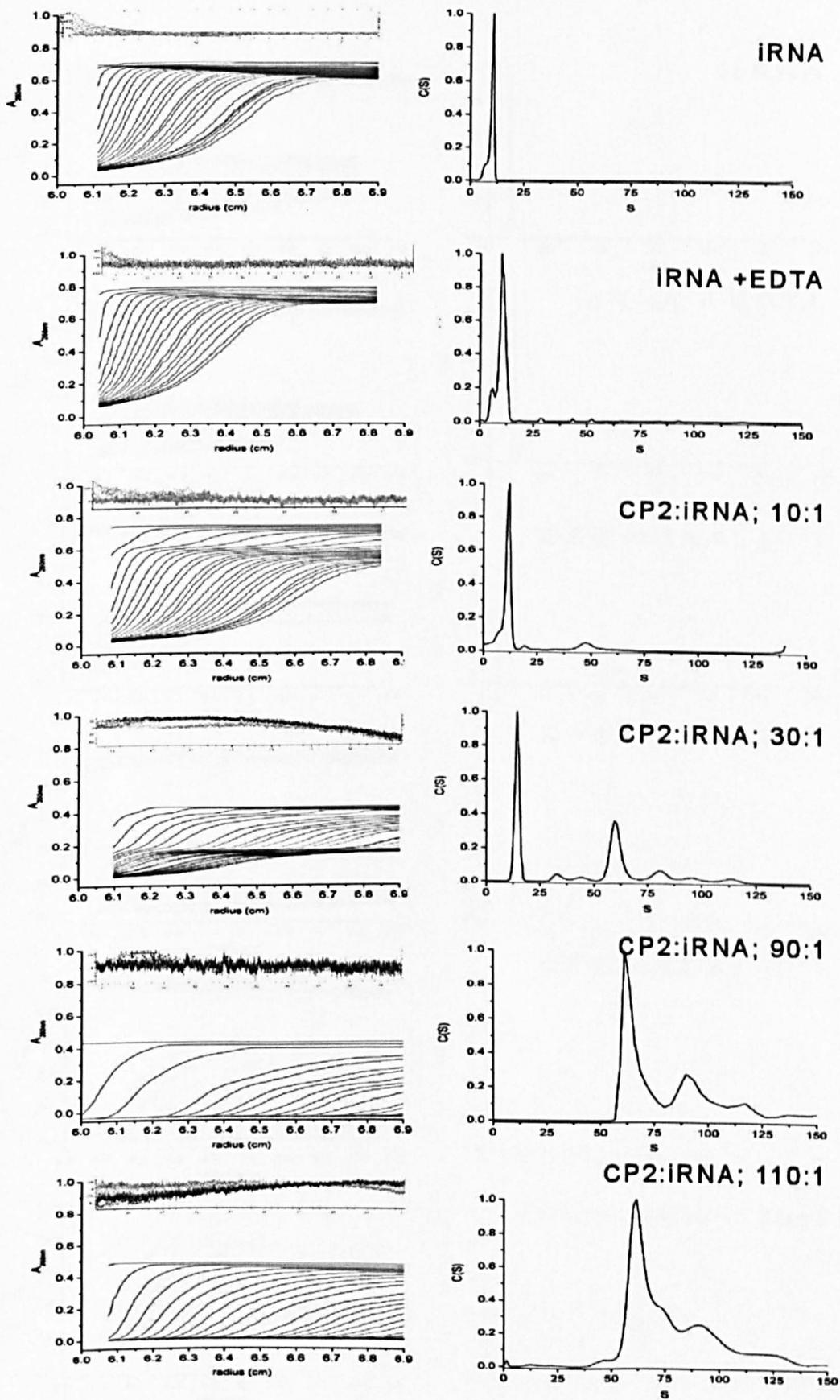
**Supplementary Figure-2.** WT MS2 RNA and capsid assembly reactions initiated with it.



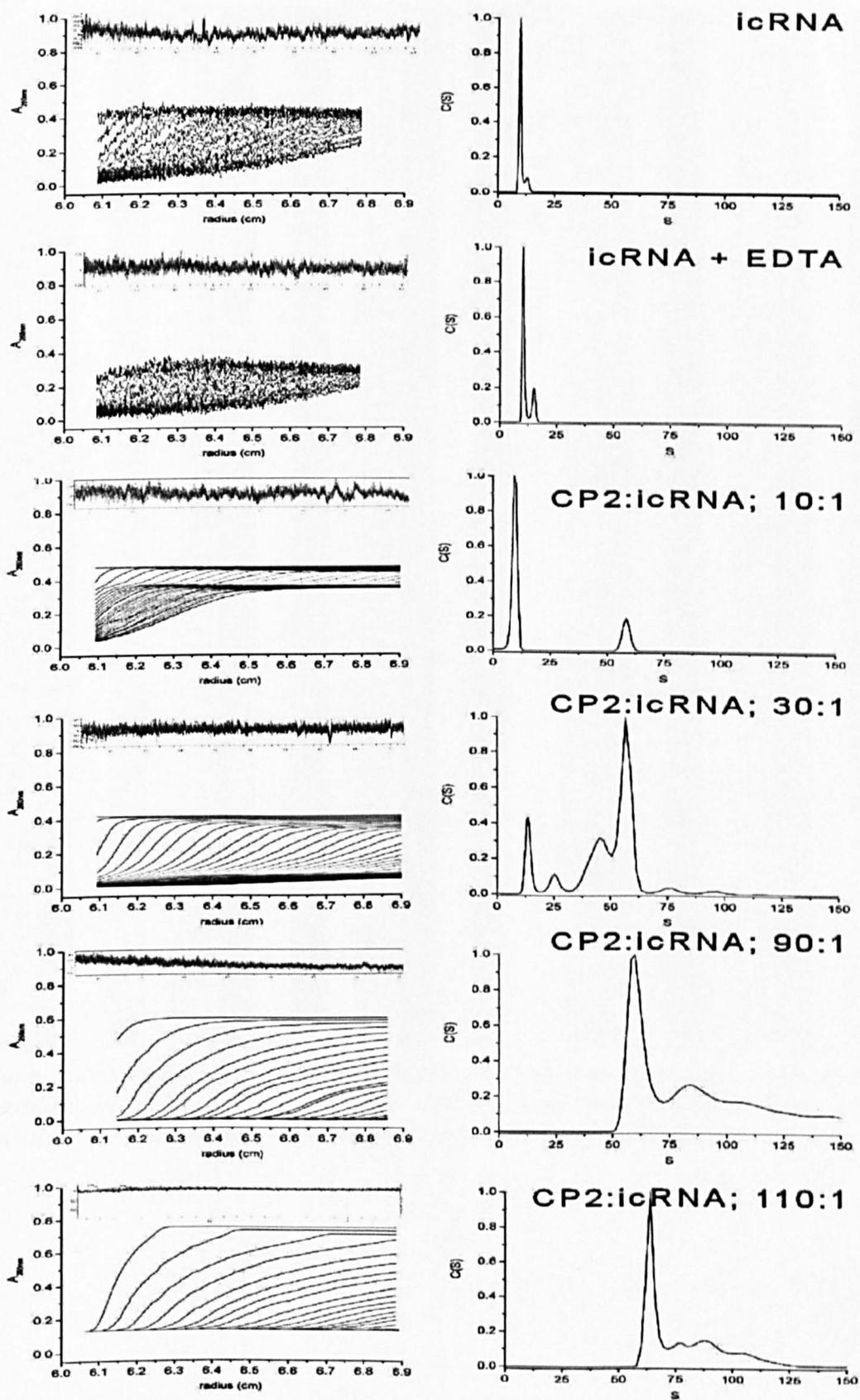
**Supplementary Figure-3.** 5' sub-genomic MS2 RNA and MS2 capsid assembly reactions initiated with it.



*Supplementary Figure-4. 3' sub-genomic MS2 RNA and MS2 capsid assembly reactions initiated with it.*

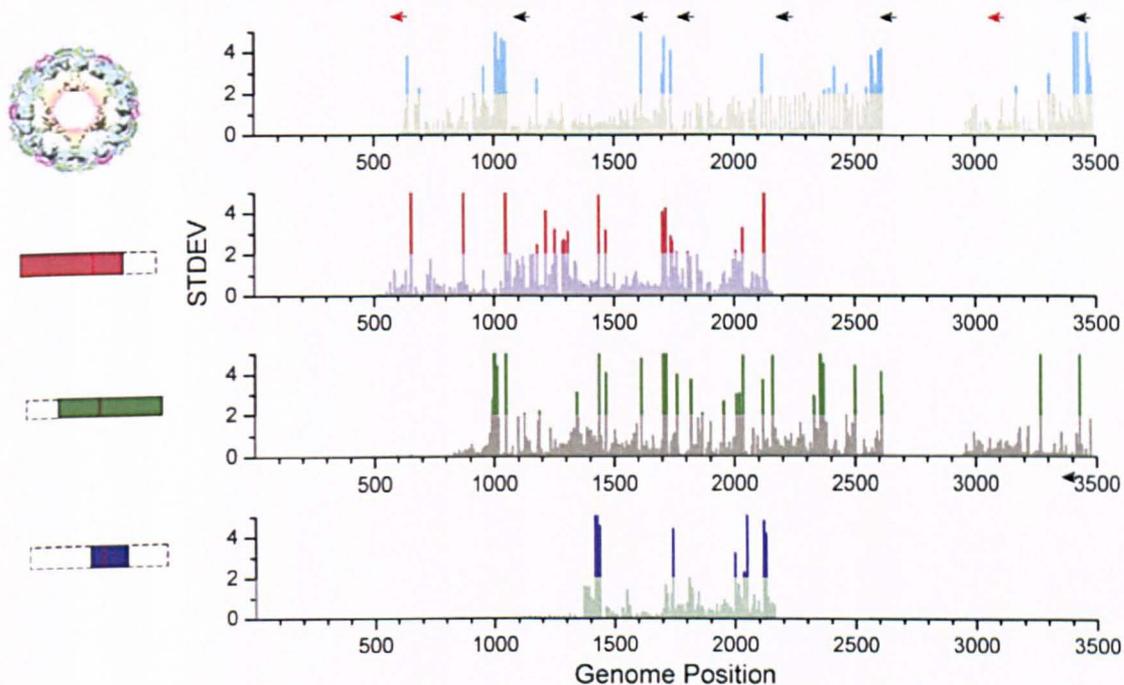


**Supplementary Figure-5.** Internal 928 nt sub-genomic MS2 RNA and MS2 capsid assembly reactions initiated with it.



**Supplementary Figure-6.** Internal control 874 nt sub-genomic MS2 RNA and MS2 capsid assembly reactions initiated with it.





**Supplementary Figure 8. Automated DATA analysis of CE data set.** Intensities are expressed as standard deviation from the normalised average intensity following subtraction of a control electropherogram. More specifically, for each primer, a  $Pb^{2+}$  cleavage electropherogram and a control electropherogram were obtained. The intensities of each electropherogram were normalised to their average intensities. The control dataset was then subtracted from the  $Pb^{2+}$  dataset in order to remove positions which might have occurred due to RT stop sites or RNA degradation. The above figure shows the signal intensity at each nucleotide position of this dataset expressed as standard deviation. Cleavage intensities can be categorised as strong or weak which might be useful for generating constraint based RNA structure predictions of the MS2 genome. STDEV below 2 represent weak  $Pb^{2+}$  cleavage positions while a STDEV above 2 corresponds to a strong  $Pb^{2+}$  hit. The arrows mark the primer annealing positions used for reverse transcription. The red arrows correspond to primers 566.R and 3163.R which did not work. Data coverage upstream of these primers is therefore lacking.



phage	5RNA	3RNA	RNA																				
1823			1	2014			1	2143				2389			1	3082			1	3414			1
1865				2016				2144				2378			1	3083			1	3416			1
1866	1	1	1	2017			1	2151				2380			1	3084			1	3419			1
1867	1	1	1	2018			1	2152				2381			1	3090			1	3420			1
1868				2019				2153				2400	1		1	3091			1	3423			1
1886				2020	1	1	1	2154				2401	1		1	3098			1	3424			1
1889		1	1	2021	1	1	1	2155				2402	1		1	3099			1	3425			1
1890		1	1	2022	1	1	1	2156				2403	1		1	3114	1		1	3428	1		1
1894				2026				2159				2412			1	3115	1		1	3429	1		1
1898				2031				2160				2414	1			3121			1	3430	1		1
1909				2034				2161				2483			1	3122			1	3434	1		1
1912	1			2035	1	1	1	2162				2454			1	3123			1	3435	1		1
1913				2036	1	1	1	2163				2465			1	3124			1	3436	1		1
1915	1			2037	1	1	1	2164				2466	1		1	3130			1	3438	1		1
1916	1			2038	1			2165				2467	1		1	3131			1	3445	1		1
1920				2039	1			2171				2468	1		1	3132			1	3450	1		1
1921				2042				2172				2487			1	3134	1		1	3454	1		1
1927				2046				2173				2490			1	3135	1		1	3455	1		1
1933				2047				2178				2499			1	3141			1	3458	1		1
1934				2048			1	2179				2500			1	3142			1	3459	1		1
1939				2049			1	2180				2517	1		1	3143			1	3464	1		1
1940				2068	1			2181				2518	1		1	3150			1	3465	1		1
1941				2069	1			2182				2522	1		1	3151			1	3466	1		1
1942				2070				2183				2523	1		1	3154			1	3471	1		1
1955				2071		1	1	2184				2537	1		1	3158	1		1	3475	1		1
1956		1	1	2073	1	1	1	2185				2538	1		1	3159	1		1	3476	1		1
1957		1	1	2074	1	1	1	2186				2596	1		1	3168			1				
1958				2076	1	1	1	2187				2597	1		1	3169			1				
1959				2077	1	1	1	2192				2590	1		1	3170			1				
1961				2078	1	1	1	2193				2598	1		1	3181			1				
1962				2087	1			2210				2599	1		1	3182			1				
1965				2088	1			2211				2600	1		1	3183			1				
1971				2092				2221				2609	1		1	3184			1				
1972				2093				2222				2609	1		1	3215			1				
1974				2095		1	1	2223				2610	1		1	3216			1				
1976				2107	1			2226				2611	1		1	3217			1				
1978	1		1	2108	1			2227				2612	1		1	3258			1				
1980				2109	1			2228				2960			1	3268	1		1				
1981	1			2113	1			2273				2961			1	3269			1				
1982				2114	1			2274				2962			1	3270	1		1				
1987				2115	1			2275				2980	1		1	3280			1				
1988				2116	1			2313				2981	1		1	3281			1				
1989				2117	1			2314				2982	1		1	3282			1				
1990				2120				2329				2997			1	3309	1		1				
1992				2121				2330				2998			1	3310	1		1				
1993		1		2122				2331				2991	1		1	3333	1		1				
1994		1		2123	1	1		2340				2992	1		1	3337	1		1				
1995				2124	1	1		2341				2993	1		1	3341			1				
1996				2125			1	2342				2999			1	3369	1		1				
1997	1	1		2126			1	2343				3000			1	3371			1				
1998	1	1		2127				2344				3002			1	3372			1				
1999				2128				2354				3003			1	3378	1		1				
2000				2131				2355				3007			1	3392	1		1				
2002				2132	1			2359				3008			1	3394	1		1				
2004				2136				2360				3009			1	3396	1		1				
2007	1	1	1	2137	1	1	1	2361				3047	1		1	3397	1		1				
2008	1	1	1	2138	1	1	1	2362				3048	1		1	3399	1		1				
2009	1	1	1	2141				2367				3059			1	3412	1		1				
2013			1	2142				2368				3072			1	3413	1		1				

**Supplementary Figure 9b. MS2 nucleotide position of all lead ion induced hydrolysis positions in all probed RNA samples identified from capillary electrophoresis electropherograms. Part II, MS2 nucelotide region 1823-3467. Lead ion cleavage positions are marked with a "1" at the relevant nucleotide positions.**

# A Simple, RNA-Mediated Allosteric Switch Controls the Pathway to Formation of a $T=3$ Viral Capsid

Peter G. Stockley\*, Ottar Rolfsson, Gary S. Thompson, Gabriella Basnak  
Simona Francese, Nicola J. Stonehouse, Steven W. Homans  
and Alison E. Ashcroft

*Astbury Centre for Structural  
Molecular Biology, University of  
Leeds, Leeds LS2 9JT, UK*

Using mass spectrometry we have detected both assembly intermediates and the final product, the  $T=3$  viral capsid, during reassembly of the RNA bacteriophage MS2. Assembly is only efficient when both types of quasiequivalent coat protein dimer seen in the final capsid are present in solution. NMR experiments confirm that interconversion of these conformers is allosterically regulated by sequence-specific binding of a short RNA stem-loop. Isotope pulse-chase experiments confirm that all intermediates observed are competent for further coat protein addition, i.e., they are all on the pathway to capsid formation, and that the unit of capsid growth is a coat protein dimer. The major intermediate species are dominated by stoichiometries derived from formation of the particle threefold axis, implying that there is a defined pathway toward the  $T=3$  shell. These results provide the first experimental evidence for a detailed mechanistic explanation of the regulation of quasiequivalent capsid assembly. They suggest a direct role for the encapsidated RNA in assembly *in vivo*, which is consistent with the structure of the genomic RNA within wild-type phage.

© 2007 Elsevier Ltd. All rights reserved.

\*Corresponding author

**Keywords:** virus assembly; RNA–protein interaction; mass spectrometry

## Introduction

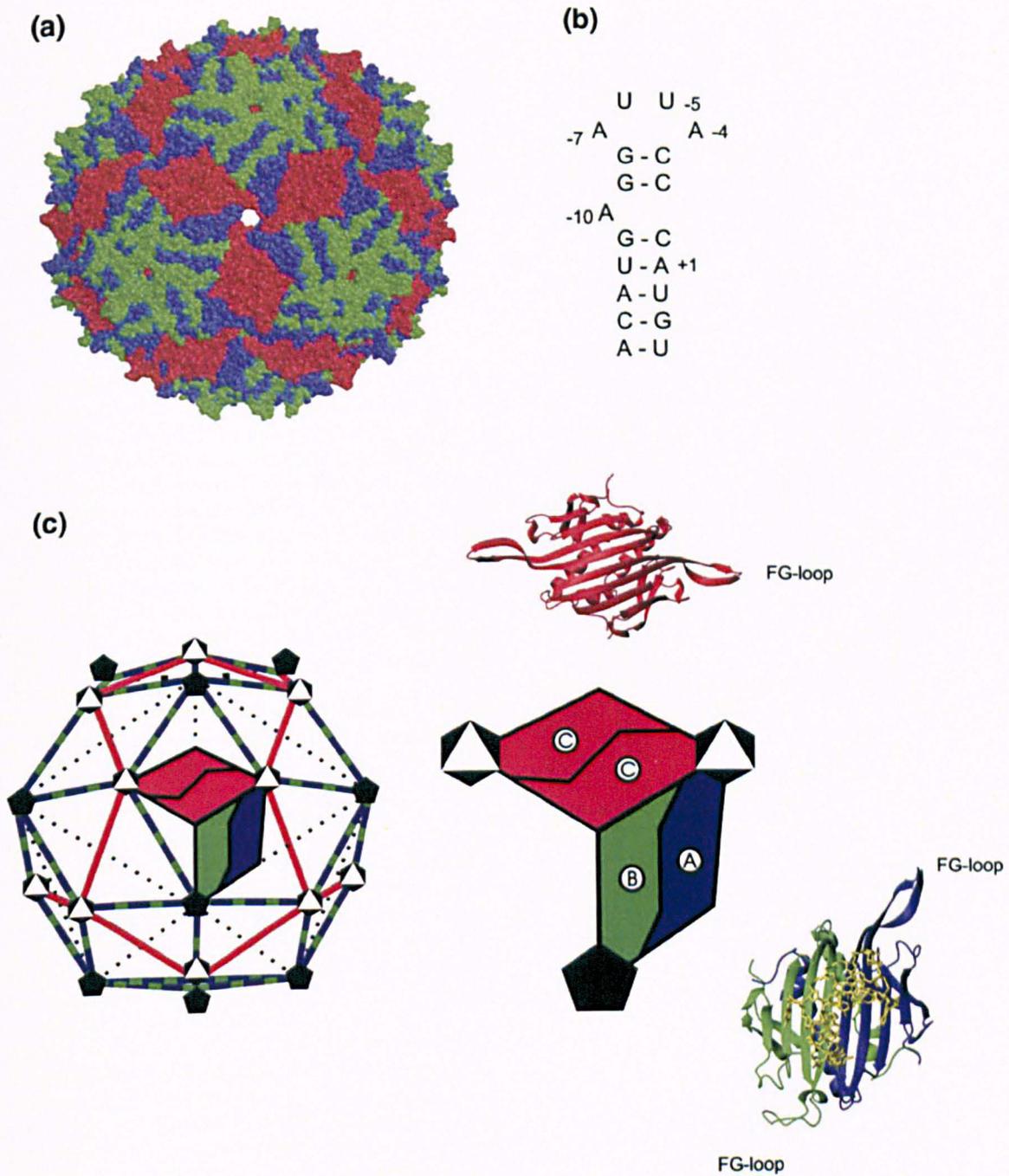
Simple spherical viruses consisting of a nucleic acid genome enclosed in a protein shell composed of multiple copies of one, or very few, coat protein (CP) subunits generally have isometric capsids. The positioning of the CP subunits is predictable from the allowed tessellations of a sphere.<sup>1,2</sup> We studied the RNA bacteriophage MS2, which conforms to a  $T=3$  shell in the Caspar–Klug nomenclature. The triangulation number ( $T=3$ ) implies that the phage CP subunit can exist in three quasiequivalent conformations (A, B, and C), allowing 180 copies of the subunit to create an enclosed capsid. This is indeed the case (Figure 1); the CP subunits form interdigitated noncovalent dimers (CP<sub>2</sub>) that act as the basic capsid building blocks composed of A/B

or C/C conformers.<sup>3,4</sup> Within each monomer, these conformers are defined by the orientation of the loop of the polypeptide connecting the F and G  $\beta$ -strands, with A and C subunits having extended loops while B subunit loops fold back toward the globular core of the protein (Figure 1(c)). In the complete capsid, A/B dimers surround the particle fivefold axes with a ring of B-type loops, while A- and C-type loops alternate around the threefold axes.

A major unresolved problem in structural virology is understanding the detailed molecular mechanism(s) that gives rise to the assembly of capsids of the correct size and symmetry. Various proposals for the formation of  $T=3$  shells based on the initial formation of three<sup>5</sup> or fivefold<sup>6</sup> assembly initiation complexes have been made. However, capsid assembly is spontaneous and usually very rapid, with the result that it has been technically difficult to isolate and characterize intermediates beyond these “initiation complexes” on the pathway to the final products.<sup>7,8</sup> The MS2 system is an ideal model for investigating such phenomena due to the extensive biochemical and structural information available.<sup>3,9–13</sup> Capsid reassembly can be

Abbreviations used: CP, coat protein; ESI-MS, electrospray ionization-mass spectrometry.

E-mail address of the corresponding author:  
[stockley@bmb.leeds.ac.uk](mailto:stockley@bmb.leeds.ac.uk)



**Figure 1.** The structures and solution components associated with MS2 bacteriophage assembly. (a) A space-filling representation of the X-ray structure of the  $T=3$  capsid of wild-type MS2<sup>3</sup> (Protein Data Bank 2MS2). (b) The sequence and secondary structure of the TR RNA. (c) The distribution of symmetry axes (black and white symbols) within the icosahedral surface lattice of the MS2 capsid and the arrangement of the quasiequivalent dimers, AB (blue/green) and CC (red), within it. Alongside are ribbon models for the structures of each, with the A/B dimer bound to TR RNA, and an enlarged cartoon of their relationship within the capsid.

triggered *in vitro* by a sequence-specific RNA-protein interaction between a CP<sub>2</sub> and an RNA stem-loop (TR) of just 19 nt<sup>14,15</sup> that encompasses the start codon of the viral replicase on the RNA genome (Figure 1). RNA-protein binding thus achieves two functions *via* a single molecular recognition event: translational repression of replicase and creation of an assembly competent complex on viral RNA.

We used electrospray ionization-mass spectrometry (ESI-MS) to investigate the mechanism of capsid reassembly and here show for the first time the time-dependent formation of both intermediates and the final product in a single assay. ESI-MS has proven to be a powerful technique for the analysis of noncovalently bound protein complexes because the mild ionization conditions are able to preserve often

fragile macromolecular structures, thus permitting mass analysis of the intact complex<sup>16,17</sup> Our analysis of the yield and kinetics of the assembly of both intermediates and the final capsid leads to one of the first mechanistic descriptions of the control of quasiequivalent interactions during virus assembly, in this case *via* an allosteric switch controlled by interaction with page RNA.

## Results and Discussion

### An RNA-mediated conformational change

The reassembly experiments were analyzed using a customized nanoelectrospray ionization–time-of-flight instrument with an extended  $m/z$  range (up to  $m/z$  60,000), collisional cooling capabilities,<sup>18–21</sup> and a temperature-controlled, automated injection and ionization system (see Materials and Methods). Capsid reassembly *in vitro* was achieved by raising the pH of a solution of acid-disassembled coat protein in the presence of TR RNAs.<sup>22,23</sup> Working in solutions of final pH 5.2–5.7 and in buffers compatible with mass spectrometry slows the reassembly process, allowing us to observe early and intermediate stages.

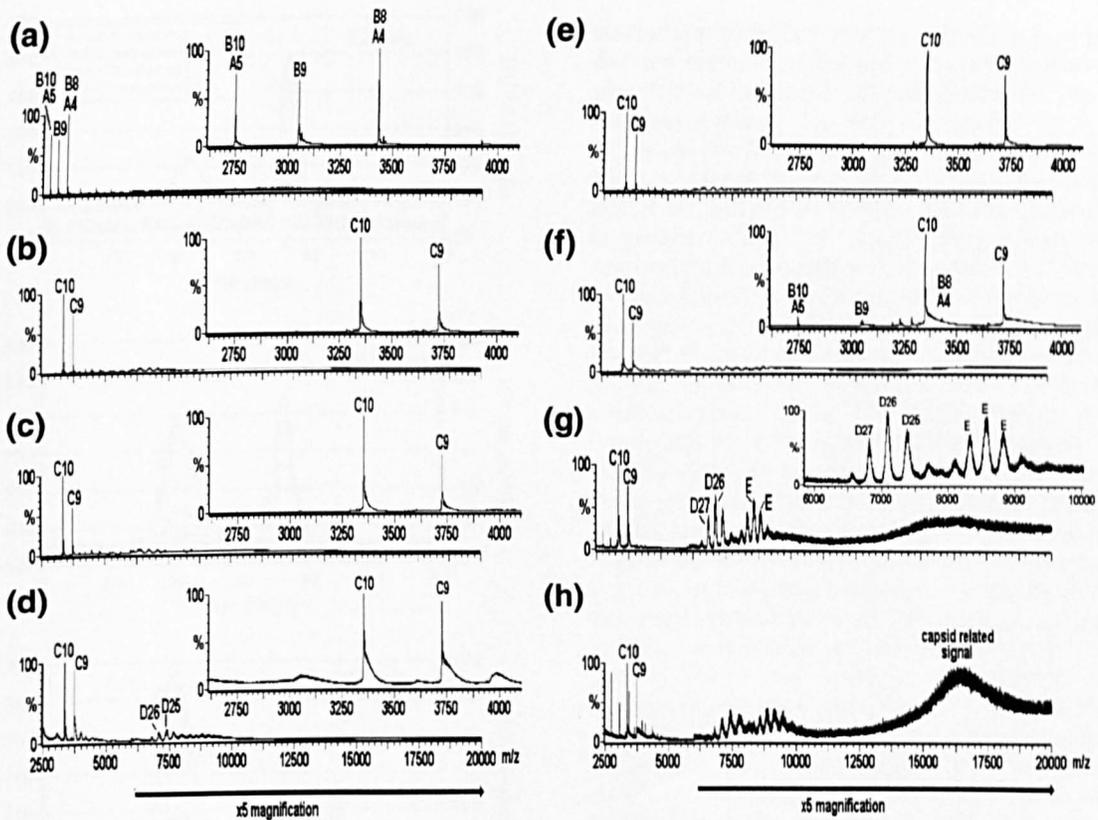
In our ESI-MS experiments the possible molecular components in the reassembly reaction consist of the following: the CP; the CP dimer (CP<sub>2</sub>); the TR RNA (TR), the RNA–protein dimer complex (CP<sub>2</sub>:TR), and various higher-order species. Systematic analysis of reassembly conditions by gel filtration-light-scattering prior to the MS assay allowed us to choose conditions for the reaction that slow the production of the final product, the  $T=3$  capsid. These conditions favor very high overall yields of the desired end product and essentially eliminate the production of mis-structured aggregates that could otherwise complicate the analysis.

At pH  $\sim 3.0$  the CP obtained by acid disassembly of recombinant  $T=3$  capsids<sup>24</sup> shows a spectrum dominated by a mixture of CP (13.7 kDa) and CP<sub>2</sub> (Figure 2(a); components labeled A and B, respectively), even at concentrations well above the dissociation constant for dimer formation determined at neutral pH, suggesting that the CP<sub>2</sub> is not completely stable at this pH, at least in the gas phase of the mass spectrometer. These major species are accompanied by minor species with masses consistent with the CP trimer, tetramer, and pentamer (Figure 2(a); unlabeled, minor peaks (<1%) in the range  $m/z$  4000–7000). The addition of TR to a solution of coat protein at a final pH  $\sim 5.2$ – $5.7$  and at a stoichiometry of 1:1 (CP<sub>2</sub> to TR) changes the spectrum radically, resulting in the complete disappearance of CP, CP<sub>2</sub>, and the other minor species and the emergence of the complex CP<sub>2</sub>:TR of mass 33.5 kDa (Figure 2(b); component C). The presence of this complex as the major species implies that all of the CP<sub>2</sub> binds TR RNA and is thus stabilized. The narrow ESI charge state distribution ( $9^+$ – $10^+$  ions) of the complex reflects its well-defined structure. This

species is accompanied by masses consistent with trace amounts of higher-order oligomers in the region  $m/z$  4000–10,000, which do not appear to alter significantly in concentration over time (Figure 2(c)) until after some 5 h, when certain higher-order species increase slightly in intensity (e.g., component D in Figure 2(d); see below for an explanation).

The spectra are significantly different, however, when reactions are performed at a final stoichiometry of 2:1 CP<sub>2</sub> to TR (Figure 2(f)–(h)), by allowing pre-equilibration of the complex (1:1, CP<sub>2</sub>:TR) for 10 min (Figure 2(e)) followed by the addition of an equal aliquot of CP<sub>2</sub> lacking RNA (Figure 2(f)). The proportion of the higher-order species increases rapidly and dramatically and is not consistent with a simple mass action effect (Figure 2, cf. (c) and (g)). This effect was not seen when equal aliquots of CP<sub>2</sub> and TR were added to 1:1 reactions. Several higher-order species are clearly visible in the 2:1 reaction, some of which we are able to assign to unambiguous stoichiometries, as well as weaker species whose signals are too broad to measure accurately. For the mass assignment, the measured masses, based on a minimum of three adjacent charge states, were compared with the calculated masses of all possible multiples of CP and TR. A significant higher mass species of 182.7 kDa is consistent with the complex  $[3(\text{CP}_2:\text{TR})+3\text{CP}_2]$  (Figure 2(g) and (h); component D). In addition to CP, CP<sub>2</sub>, CP<sub>2</sub>:TR, and  $[3(\text{CP}_2:\text{TR})+3\text{CP}_2]$  (components A–D, respectively), other, minor, higher-order species consistent with  $[\text{CP}_2:\text{TR}+\text{CP}]$  (mass 47 kDa),  $[2(\text{CP}_2:\text{TR})]$  (mass 67 kDa), and  $[3(\text{CP}_2:\text{TR})+\text{CP}_2+\text{CP}]$  (mass 142 kDa) were unambiguously identified (Figure 2; not labeled). Other species of masses >200 kDa (e.g., component(s) E, mass  $\sim 288$ – $300$  kDa) were more difficult to identify due to low-intensity broad peaks, which made unambiguous charge state assignment difficult and hence gave rise to the possibility of several theoretically compatible stoichiometries. The presence of complexes with CP monomer components may arise because of the presence of the monomer under the starting conditions. Over time the peaks corresponding to all of these species decreased in intensity. In contrast to the time-course for the 1:1 CP<sub>2</sub> to TR reaction (Figure 2(b)–(d)), the higher-order aggregates ( $m/z$  4000–10,000) are accompanied by broad unresolved higher mass to charge signals ( $> m/z$  14000) visible in the spectra for the 2:1 reaction from the earliest time-point sampled ( $\sim 1$  min; Figure 2(f)–(h)). Such broad peaks have been mass assigned to the MS2  $T=3$  capsid by others<sup>25</sup> and are also seen when we analyze the intact recombinant capsid.<sup>24</sup>

To confirm that the high mass peaks corresponded to the formation of  $T=3$  shells, identical reassembly reactions were analyzed by gel filtration-light-scattering (Figure 3). The 1:1 reaction forms capsids only very slowly, whereas the 2:1 equivalent has capsids present from the earliest time-points. These data imply that the small amounts of reassembly intermediates seen in Figure 2(b)–(d) are caused by a limiting concentration of an RNA-free CP compo-

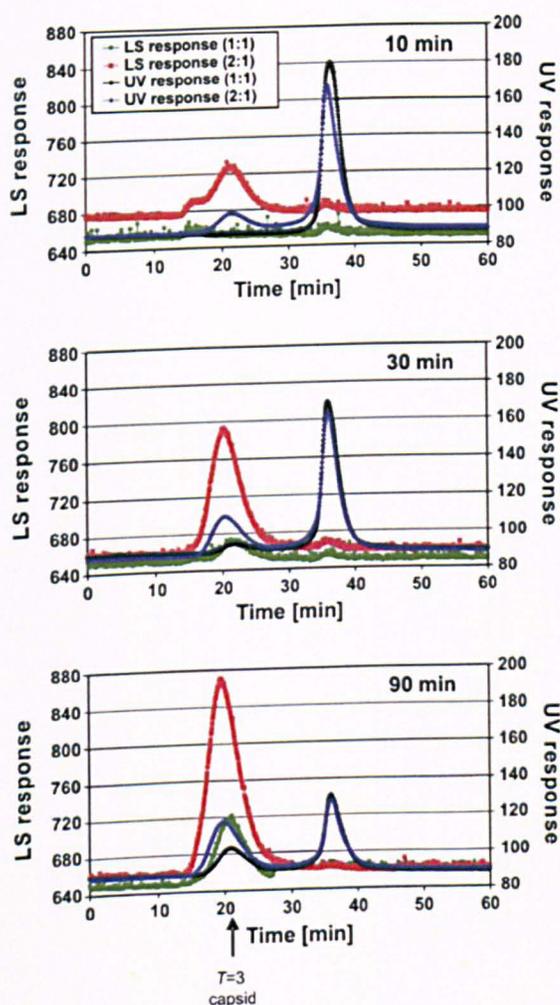


**Figure 2.** Virus capsid reassembly monitored over time by real-time nanoESI-MS. The spectra were acquired over the range  $m/z$  500–30,000 for samples in ammonium acetate (40 mM) at pH 5.2–5.7. The components are labeled as follows: A = CP; B = CP<sub>2</sub>; C = CP<sub>2</sub>:TR; D = 182.7 kDa, assigned as [3(CP<sub>2</sub>:TR)+3CP<sub>2</sub>], and E = ~288–300 kDa, unassigned. The number immediately following each letter is the charge state of those particular ions, i.e., [M + nH]<sup>n+</sup>. The spectra are as follows: (a) CP alone (8 μM); pH 3.2; (b)–(d) are for reassembly at pH 5.2–5.7 at a stoichiometry of CP<sub>2</sub> to TR 1:1 (8 μM:8 μM), at 1, 90, 300 min, respectively. Spectrum (e) is the 1:1 pre-equilibrated starting point for reassembly at a stoichiometry of CP<sub>2</sub> to TR 2:1 (16 μM:8 μM), spectra (f)–(h), at 1, 120, and 180 min, respectively. Note that spectra (b) and (e) are essentially equivalent. The bars below the spectra indicate a magnification factor of 5 for all ions above  $m/z$  6000 to enhance clarity. The insets in (a)–(f) highlight the range  $m/z$  2600–4100 to enhance clarity in the region containing ions arising from components A, B, and C. The inset in (g) highlights the range  $m/z$  6000–10,000 to enhance clarity of the ions arising from components D and E and is normalized to the D26 ions, which are ~10% of the intensity of the full spectrum.

ment in the assembly pathway. Increasing the concentration of CP<sub>2</sub> triggers a dramatic increase in the efficiency of assembly of the higher-order species observed in Figure 2(f)–(h). Further increasing the concentration of RNA-free CP subunits in the MS assay (to a 4:1 CP<sub>2</sub> to TR ratio) results in more rapid loss of these higher-order species, including the CP<sub>2</sub>:TR complex, from the  $m/z$  spectrum, suggesting that they do participate in capsid formation rather than being simply MS artifacts. The spectrum is then dominated by peaks originating from CP monomer and dimer, together with a trace of trimer (data not shown). A reasonable inference from these effects is that the RNA-free coat protein added to the preformed CP<sub>2</sub>:TR complex is in a different quasiequivalent conformation from the one in the complex. Because both A/B and C/C dimers are required to build the  $T=3$  capsid, mixing these conformers at high concentration allows the assembly reaction to proceed efficiently. The 1:1 CP<sub>2</sub>:TR mixture also makes  $T=3$  shells, although under these conditions the RNA-free sample does not. Because the life-time of the CP<sub>2</sub>:TR is 42 s,<sup>26,27</sup> both

types of dimer are available in the 1:1 reaction but in these reactions the concentration of the RNA-free form would be limiting.

To probe whether TR binding leads to conformational change within the CP subunit, using NMR we carried out measurements with a mutant CP subunit (W82R) that does not assemble beyond dimer but does bind TR RNA with the same affinity as the wild-type (unpublished results). Residue 82 lies outside both the FG-loop and the TR binding site. The backbone resonances from the <sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N triple resonance NMR spectrum of the apo-CP<sub>2</sub>(W82R) were analyzed with 87% of the backbone NH resonances being assigned, including virtually all the nonproline resonances in the FG loop region. The full details of these experiments will be published elsewhere. The degeneracy of all pairs of backbone resonances present in apo-CP<sub>2</sub>(W82R) resulting from pairs of symmetry-related atoms in the two subunits of the dimer, encompassing the FG loop, show that in the absence of RNA the protein subunits are equivalent on the NMR chemical shift timescale (Figure 4). Furthermore, <sup>1</sup>H–<sup>15</sup>N heteronuclear nu-



**Figure 3.** Gel filtration-light-scattering assays of capsid reassembly. MS2 CP<sub>2</sub> and TR were mixed in 40 mM ammonium acetate, pH 5.2–5.7, to form a 1:1 reaction, incubated at 4 °C for 10, 30, or 90 min, and then loaded onto a Sepharose 6 gel filtration column equilibrated in 50 mM Tris-acetate, pH 7.4, and eluted at 0.36 ml/min. The outflow from the column was analyzed simultaneously *via* UV absorbance (black trace) and light-scattering (green trace). Similar reactions were then set up, preincubated at a molar ratio of 1:1 for 10 min, and then an additional aliquot of protein was added to create the 2:1 reaction. The traces are shown as UV absorbance (blue trace) and light-scattering (red trace). The position at which authentic T=3 capsids elute is marked with an arrow; the peak at ~35 min corresponds to the starting materials and smaller complexes.

clear Overhauser effect spectra (data not shown) show considerable motion within the FG loop on picosecond to nanosecond timescales, indicating that the chemical shift equivalence seen in these regions results from an average of a large number of conformers.

The addition of TR to the labeled sample leads to a change in the spectrum of CP<sub>2</sub>(W82R), with resonances for the *apo*-CP<sub>2</sub> state disappearing and pairs of peaks appearing at perturbed positions, which are associated with the [CP<sub>2</sub>(W82R):TR] complex. The

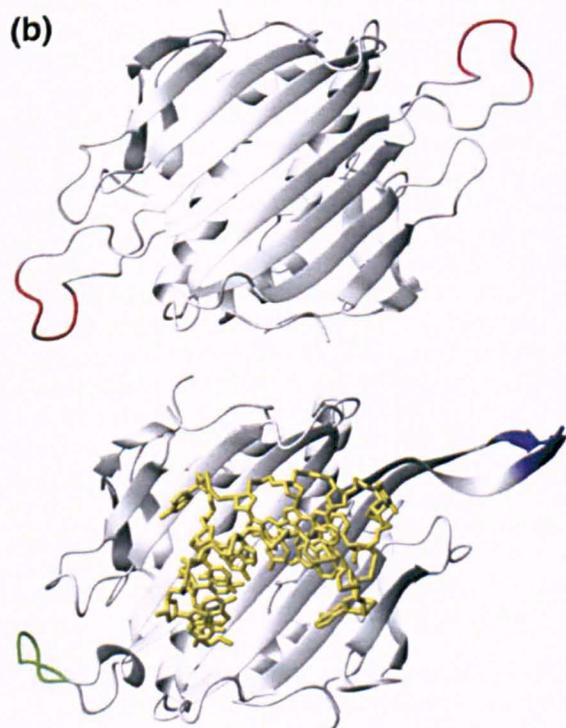
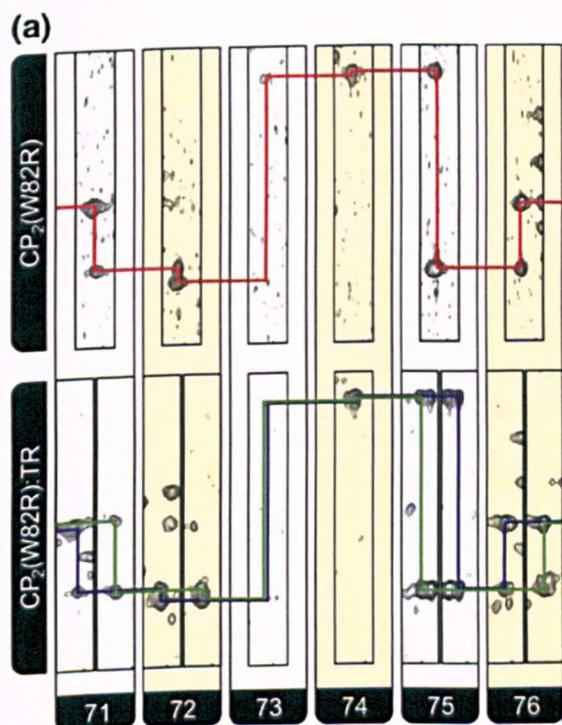
data show that the complex is in slow exchange with the *apo* state, as expected for a RNA–protein complex with a nanomolar binding affinity. The sequential assignment for the backbone HN, N, and C<sup>α</sup> resonances of residues 71–76 were made from HNCA, HncoCA, and <sup>1</sup>H–<sup>15</sup>N heteronuclear single-quantum coherence spectra and the position of the fragment within the CP<sub>2</sub>(W82R) primary sequence assured by the presence of distinctive C<sup>α</sup> resonances for residues 73 and 74, which are a GG pair, one of only two such pairs within the primary sequence of the CP. Pairs of resonances can be seen for residues 71, 72, 75, and 76, which are clearly delineated by shift differences in the <sup>1</sup>H+<sup>15</sup>N dimensions (Figure 4). Comparison of the observed differences in shifts for the FG loop <sup>1</sup>H amide resonances were made with those predicted by the Heigh Mallion ring current model using the program MOLMOL and ring current intensities suitable for RNA.<sup>28,29</sup> These calculations show that the observed changes in shifts in the presence and absence of TR are approximately an order of magnitude greater than that expected from theory (data not shown), confirming that the observed shifts are not caused by RNA binding *per se* but arise either from RNA-induced changes in the conformations of the loops or from fixing of a defined conformational state from within an equilibrium of states present in the *apo* protein. In summary, binding of TR results in conformational changes involving the FG loops and creates asymmetry in the complex. We propose that the conformations of the loops in the RNA-bound dimer are likely to be closer to those in an A/B coat protein dimer, compared to C/C-like or ill-defined conformers in the *apo* protein.

Why then does the W82R mutant fail to assemble beyond the TR:CP<sub>2</sub> complex? To address this question we carried out an ESI-MS chasing experiment with combinations of wild-type and W82R CP. The W82R mutant dimer forms a CP<sub>2</sub>(W82R):TR complex when mixed 1:1 with TR with very little evidence of any higher-order species. Interestingly, the charge state distribution of the CP<sub>2</sub>(W82R):TR complex was slightly but noticeably wider ( $n=7^+–10^+$ ) than that of the wild-type complex ( $n=9^+–10^+$ ) analyzed under the same buffer and instrumental conditions, indicating a more dynamic structure (data not shown). Upon the addition of wild-type CP<sub>2</sub> to the preformed CP<sub>2</sub>(W82R):TR complex, both the CP<sub>2</sub>:TR and the CP<sub>2</sub>(W82R):TR complexes were detected with no higher-order species visible. The reverse experiment, *i.e.*, performing the wild-type CP<sub>2</sub>:TR complex (with its associated dimer, trimer, and traces of other higher-order species) followed by the addition of CP<sub>2</sub>(W82R), resulted again in the observation of only the CP<sub>2</sub>:TR and CP<sub>2</sub>(W82R):TR complexes, in this case with the concomitant disappearance of the higher-order species (data not shown). There were no signals corresponding to a mixed CP-CP(W82R):TR complex. These results are consistent with our expectation that W82R can bind TR normally but is unable to support further interactions with other coat protein dimers, thus acting as an inhibitor of reassembly.

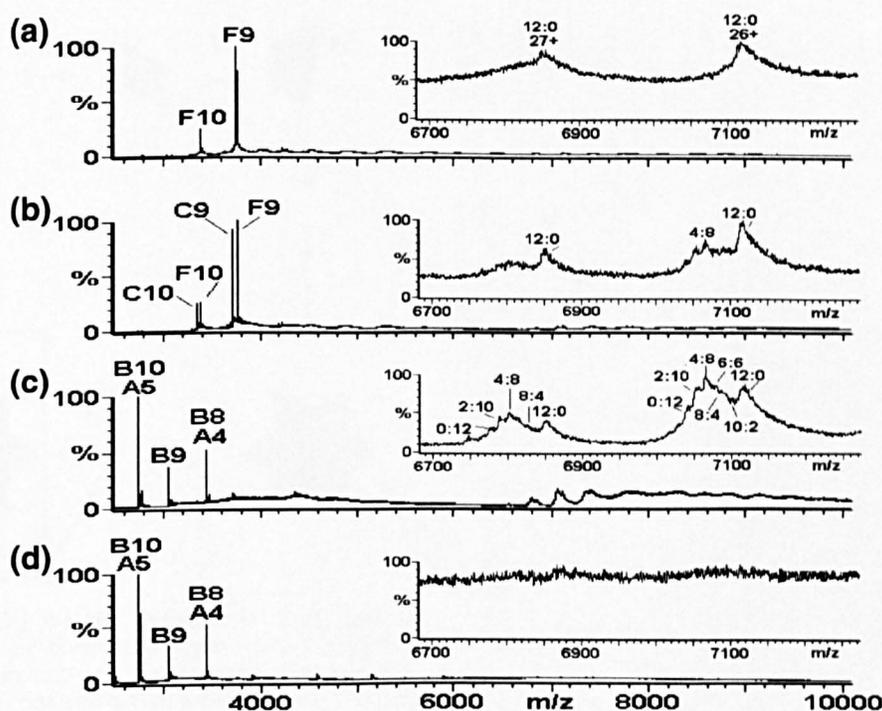
It is a reasonable assumption that the RNA-induced conformational change in the FG loops observed with W82R protein also occurs with the wild-type coat protein. This RNA-mediated allosteric effect at the FG loops explains the dependency of the kinetics and yield of assembly products on the presence of RNA-free and RNA-bound forms of the coat protein dimer.

### The pathway to $T=3$ shell formation

To confirm that the intermediates seen in the MS spectra were on-pathway to capsid formation and to determine the mass step size of capsid growth, we repeated the reassembly assay using isotopically labeled wild-type CP. The CP<sub>2</sub>:TR complex was performed as above with a 1:1 ratio of reactants but using <sup>15</sup>N-labeled CP<sub>2</sub> (Figure 5(a), component F). The intermediate complex [3(CP<sub>2</sub>:TR)+3CP<sub>2</sub>] discussed above was also visible in low yield in the region  $m/z$  6700–7200, as expected. The measured molecular mass of this component indicated a totally <sup>15</sup>N-labeled [3(CP<sub>2</sub>:TR)+3CP<sub>2</sub>] complex, i.e., with all 12 CP subunits <sup>15</sup>N-labeled, and hence is labeled 12:0 (<sup>15</sup>N to <sup>14</sup>N). RNA-free <sup>14</sup>N-CP<sub>2</sub> was then added to make the 2:1 reaction (Figure 5(b)). There were then two peaks of approximately equivalent intensity corresponding to <sup>14</sup>N and <sup>15</sup>N-labeled CP<sub>2</sub>:TR complexes (components C and F), but there is no evidence for a heterodimer complex, indicating that while the TR molecule can dissociate from the CP<sub>2</sub>:TR complex and reassociate with a different CP<sub>2</sub>, the dimer does not dissociate to monomer under these conditions. Inspection of the region  $m/z$  6700–7200 containing the multiply charged peaks for the complex [3(CP<sub>2</sub>:TR)+3CP<sub>2</sub>] in the 2:1 reaction reveals that the <sup>15</sup>N-labeled complex peaks are still present (Figure 5(b), inset, labeled 12:0) with additional peaks of the same charge state but with lower  $m/z$  values, i.e., of lower molecular mass. Further addition of <sup>14</sup>N-CP<sub>2</sub> (<sup>15</sup>N to <sup>14</sup>N ratio now 1:3) causes these peaks to intensify (Figure 5(c)) and finally (<sup>15</sup>N to <sup>14</sup>N ratio now 1:4) disappear (Figure 5(d)), as expected from the previous experiments. The masses of these components indicated the presence of complexes consisting of <sup>15</sup>N to <sup>14</sup>N CP in the ratios of 10:2, 8:4, 6:6, 4:8, 2:10, and 0:12 subunits, respectively, i.e., equivalent to all stoichiometries differing by dimer addition. The mass differences between these species consolidate the assignment of this complex and confirm the presence of 12 CP molecules. Again, no evidence was found for complexes of this stoichiometry with mixed <sup>15</sup>N:<sup>14</sup>N CP<sub>2</sub> units. The predominance of the <sup>15</sup>N to <sup>14</sup>N 4:8 ratio species in Figure 5(c) is consistent with the ratio of



**Figure 4.** Observed changes in the chemical shifts for the FG loop residues 71–76 in both *apo* and RNA-bound complexes. (a) Strips, alternately colored white and yellow for each residue, from HNCA spectra of CP<sub>2</sub>(W82R) (top) and CP<sub>2</sub>(W82R):TR (bottom). The assignment pathway for CP<sub>2</sub>(W82R) is shown in red while the split pathways for the complex are in blue and green. Note that no assignment of resonances to specific FG loop conformations is made or assumed. (b) Cartoon representation of (top) the backbone of CP<sub>2</sub>(W82R) showing the position of residues 71–76 in red for the two CP subunits derived from the crystal structure of this mutant<sup>38</sup> (Protein Data Bank 1MSC) and (bottom) the equivalent view for a wild-type A/B dimer (green/blue) bound to TR RNA (yellow sticks), which is the conformation we assume is closest to that present in solution.



**Figure 5.** Mapping the capsid reassembly pathway using isotopic chase experiments. Virus capsid reassembly from  $^{15}\text{N}$ -labeled  $\text{CP}_2$  chased by adding increasing quantities of  $^{14}\text{N}$ - $\text{CP}_2$ , with each reaction monitored by real-time nanoESI-MS at  $t=1$  min. The spectra were acquired over the range  $m/z$  500–10,000 for samples in 40 mM ammonium acetate, pH 5.2–5.7. The larger panels show the range  $m/z$  2500–10,000 and the components are labeled as follows: A = CP; B =  $\text{CP}_2$ ; C =  $\text{CP}_2$ :TR; F =  $^{15}\text{N}$ - $\text{CP}_2$ :TR. The number immediately following each letter is the charge state of those particular ions. Spectra are (a)  $^{15}\text{N}$ - $\text{CP}_2$  to TR 1:1 (8  $\mu\text{M}$ :8  $\mu\text{M}$ ),  $t=1$  min; (b)  $^{15}\text{N}$ - $\text{CP}_2$  to TR 1:1 (8  $\mu\text{M}$ :8  $\mu\text{M}$ ) with  $^{14}\text{N}$ - $\text{CP}_2$  (8  $\mu\text{M}$ ) added,  $t=1$  min; (c)  $^{15}\text{N}$ - $\text{CP}_2$  to TR 1:1 (8  $\mu\text{M}$ :8  $\mu\text{M}$ ) with  $^{14}\text{N}$ - $\text{CP}_2$  (24  $\mu\text{M}$ ) added,  $t=1$  min; (d)  $^{15}\text{N}$ - $\text{CP}_2$  to TR 1:1 (8  $\mu\text{M}$ :8  $\mu\text{M}$ ) with  $^{14}\text{N}$ - $\text{CP}_2$  (32  $\mu\text{M}$ ) added,  $t=1$  min. Inset spectra highlight the range  $m/z$  6700–7200, showing the 26+ and 27+ charge states of the major reassembly intermediate [ $3(\text{CP}_2$ :TR)+ $3\text{CP}_2$ ]. The peaks are labeled with the ratio of  $^{15}\text{N}$ : $^{14}\text{N}$  CP contained in the species, i.e., 12:0, 10:2, 8:4, 6:6, 4:8, 2:10, and 0:12.

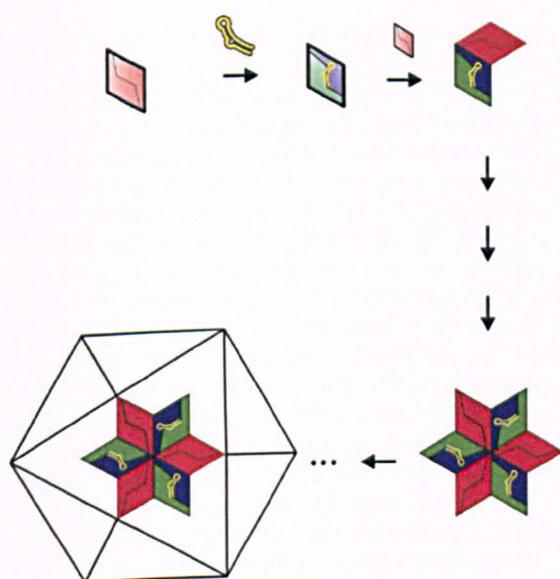
$^{15}\text{N}$  to  $^{14}\text{N}$  CP present in this particular reassembly reaction (i.e., 1:3).

Thus, it appears that the TR in the reassembly reaction rapidly equilibrates between the  $^{14}\text{N}$  and  $^{15}\text{N}$ -labeled coat protein dimer subunits, presumably because of the short half-life of the  $\text{CP}_2$ :TR complex.<sup>26,27</sup> As the initially formed [ $3(\text{CP}_2$ :TR)+ $3\text{CP}_2$ ]  $^{15}\text{N}$ -labeled complexes continue on the assembly pathway to form capsids, new [ $3(\text{CP}_2$ :TR)+ $3\text{CP}_2$ ] complexes are formed from a mixture of the remaining  $^{15}\text{N}$ -labeled  $\text{CP}_2$  species and the added  $^{14}\text{N}$  material. Gradually, the percentage of all  $^{15}\text{N}$ -labeled [ $3(\text{CP}_2$ :TR)+ $3\text{CP}_2$ ] decreases with increasing  $\text{CP}_2$  concentration, while the mixed  $^{14}\text{N}$ : $^{15}\text{N}$  (Figure 5(b) and (c)), and eventually all  $^{14}\text{N}$  complexes, become predominant. Eventually, the addition of further  $\text{CP}_2$  dimers results in the disappearance of this complex, as well as other higher-order species, as capsid formation progresses (Figure 5(d)). One prediction of this interpretation is that performing the experiment in the opposite direction, i.e., performing the complex with  $^{14}\text{N}$   $\text{CP}_2$  and chasing with  $^{15}\text{N}$ -labeled material, should produce an equivalent, but mirror image, spectrum. This is indeed the case (see Supplementary Information).

The data confirm that the intermediate species detected in the MS spectra are competent to bind additional coat proteins and are therefore likely to

be on the capsid assembly pathway. The unit of capsid growth is clearly a coat protein dimer under these conditions. Although some intermediate species are visible for extended periods during the assembly process, there does not appear to be a significant build-up of partially formed structures that later coalesce. This last observation is consistent with cryo-EM reconstructions of species formed in reassembly reactions that show particles of increasing completeness over time.<sup>23</sup>

The results lead directly to a molecular mechanism for  $T=3$  capsid formation (Figure 6). RNA-free  $\text{CP}_2$  in solution has a preponderant, quasiequivalent conformation distinct from that created when TR binds, both of which are needed to allow the capsid to be built. This is consistent with the previous observation that some  $\text{CP}_2$  subunits will not bind TR until they have undergone a conformational change.<sup>12</sup> The TR binding site lies entirely within a  $\text{CP}_2$  and it makes distinct contacts with each monomer, suggesting that it might bind preferentially to an asymmetric dimer, e.g., an A/B unit. However, the NMR data from the apo- $\text{CP}_2$  suggest that the dynamic motions of the FG loop regions also propagate along the attached F and G  $\beta$ -strands and overlap the RNA binding site, providing a potentially direct mechanism to couple RNA binding and FG loop conformational preference. A 1:1 reassembly reaction would thus be



**Figure 6.** Model for the initiation of  $T=3$  capsid assembly. A pair of chemically equivalent (C/C-like) subunits in solution (top left, shaded pink) binds a TR RNA, resulting in a conformational switch into an A/B-like asymmetric dimer (shaded green and blue). This complex binds an additional C/C-like dimer, creating the first higher-order intermediate on the pathway to capsid formation. A variety of subpathways involving the addition of coat protein dimers (arrows) can then generate the major assigned intermediate encompassing the first threefold axis of the virus particle. We have mass assignments for all the cartoon species shown in the diagram, as well as a number of components that may be in the subassembly pathway, such as  $2[\text{CP}_2:\text{TR}]+2\text{CP}_2$  &  $3[\text{CP}_2:\text{TR}]+2\text{CP}_2+\text{CP}$ . These latter species are not always present in the reactions and when present are in trace amounts.

dominated by A/B-like species, whereas the 2:1 reactions described above would, in addition, contain more symmetrical dimers but with very dynamic FG loop regions. These could be C/C-like. Formation of A/B dimers would in principle permit formation of a capsid fivefold axis *via* interacting B-type loops. Such dimers could also contribute to the formation of a threefold axis by the A conformers interacting with C/C dimers. The stimulatory effect of adding RNA-free  $\text{CP}_2$  to the preformed  $\text{CP}_2:\text{TR}$  suggests that formation of particle threefold axes is the dominant pathway toward  $T=3$  shell formation under these conditions. This is consistent with the fact that none of the assigned mass peaks in the spectra corresponds to a fivefold intermediate,  $5(\text{CP}_2:\text{TR})$ ; rather, they are essentially all consistent with the threefold stoichiometry,  $[3(\text{CP}_2:\text{TR})+3\text{CP}_2]$ , implying that there is a unique assembly pathway. Formation of the threefold complex determines that assembly yields a  $T=3$  rather than a  $T=1$  capsid; the latter would require only complexes with fivefold stoichiometries.

### Biological implications

The mechanism described above immediately begs an obvious question: what triggers the C/C-

like dimer in solution to adopt the A/B conformer at appropriate positions in the growing capsid? *In vivo* there is only one copy of the TR stem-loop per genome and there are no obvious candidates for closely related sequences/structures in the remainder of the phage RNA. However, we have shown that very many RNA stem-loop variants can bind at the TR binding site. Indeed, we have characterized over 30 such species by X-ray crystallography.<sup>13,30</sup> In such experiments preformed capsids are soaked with RNAs that penetrate into the interior of the particles *via* the pores at the symmetry axes, from where they can bind to all possible binding sites within the interior, including both A/B and C/C conformers. The orientation of the RNA at C/C dimers is degenerate but it is unique at A/B, allowing the details of the RNA-protein complex to be determined. These results suggest that the difference in RNA affinity between A/B and C/C dimers in a capsid is probably modest, although for many of the RNA variants we have studied the measured *in vitro* affinity for isolated dimers is very low. Note that the recombinant  $T=3$  shells used to generate the  $\text{CP}_2$  starting material here were described previously as “empty,” because they appear to lack any MS2 RNA or other long RNA sequences.<sup>24</sup> They are not, however, free of cellular RNA fragments, which may serve as the triggers of quasiequivalent switching when the recombinant protein is expressed at high levels.

It appears, therefore, that if there were suitably oriented genomic sequences/structures outside the TR region, coat protein dimers adding to a growing shell could bind them, thus fixing the appropriate quasiequivalent conformation. Cryo-EM reconstruction of the intact phage reveals the existence of such ordered RNA segments located underneath the protein shell,<sup>31</sup> supporting this idea. We have also shown that RNAs that encompass the TR sequence but have 5' or 3' genomic extensions appear to alter the efficiency of capsid formation<sup>23</sup>. This mechanism may well play a direct role in the life-cycle of the phage. The TR sequence is a transient signal that is only generated when ribosomes read through the coat protein gene, disrupting a long-range base-pairing interaction, the Min Jou sequence.<sup>32</sup> Binding of a  $\text{CP}_2$  to the TR stem-loop would lead to translational repression of the replicase cistron and simultaneously mark the MS2 genome for packaging into capsids. Many of the genomes marked in this way, however, might only be partially replicated and it might be inappropriate to begin assembly immediately after the repression complex forms. If protein-free genomic RNAs present higher affinity binding sites for  $\text{CP}_2$  at low concentrations, this would allow all progeny phage genomes to complete replication and carry an assembly initiation species. As the concentration of  $\text{CP}_2$  rises to levels in excess of the genomic RNA copy number, these C/C-like species can begin the assembly process by interacting with the A/B dimer bound at TR. Presumably, the additional A/B dimers required to complete assembly would then be

triggered to form at appropriate points by interactions with genomic sequences outside the TR region, where both RNA-protein and protein-protein binding energy can contribute to the affinity of the interaction.

The pathway to capsid formation described above corresponds to the mechanism of autostery first described by Caspar, in which free subunits are held in an assembly-incompetent state until they make contact with the growing viral shell or another viral ligand.<sup>33</sup> In many other viral systems quasiequivalent conformer formation is dominated by alteration of the orientation of distinct protein domains or extended "arm" regions.<sup>5</sup> Many such proteins share structurally conserved protein domains and these have been co-opted by viruses from diverse sources from bacteria to humans.<sup>34</sup> The RNA phage coat proteins, in contrast, contain only a single domain and lack such extended arms. They form a unique protein fold that has to date no other viral homologues. This isolation may be because in these cases assembly is so dependent on interactions with the viral genome that it places such strict evolutionary constraints on the genomic sequence that the fold has not been used elsewhere. A related issue is the conformation of the phage genome during the assembly process. Does CP<sub>2</sub> binding to the growing shell provide the driving force for genome folding and condensation or does the protein bind to a prefolded RNA? AFM measurements on a number of other single-stranded RNA viruses suggest that the genomic RNAs can fold stably into a compact structure with dimensions similar to that of the interior of the final protein shell, suggesting that the latter proposal is more likely.<sup>35</sup>

The data presented here demonstrate the power of mass spectrometry to dissect the molecular details of a viral assembly pathway and gain insight into MS2 phage capsid construction by the identification of on-pathway intermediates. The ability to monitor mass and intensity changes in co-existing, noncovalently bound, high mass oligomeric species in real time is a vital asset for these studies. Such information is potentially very important for designing novel inhibitors of viral life-cycles, e.g., in the RNA phage, reagents that block the conformational changes required to create the A/B dimer would be potent inhibitors of assembly. Additionally, there are growing applications of virus-like particles in therapeutic and bionanotechnologic applications in which fundamental information on assembly mechanisms would be extremely useful.<sup>36,37</sup> ESI-MS can clearly contribute significantly to interdisciplinary studies of such systems.

## Materials and Methods

### Preparation of coat proteins and TR RNA

#### Proteins

Wild-type recombinant coat protein in the form of *T*=3 shells lacking specifically bound RNA were prepared by overexpression in *Escherichia coli* as described previously.<sup>24</sup>

These capsids were purified and then disassembled coat protein dimers isolated by treatment with glacial acetic acid, followed by exchange into 20 mM acetic acid.<sup>12,22</sup>

NMR experiments were carried out on samples of the MS2 mutant coat protein, W82R, that does not assemble beyond the dimer.<sup>38</sup> W82R protein was purified by successive chromatography over diethylaminoethanol and sulphopropyl exchange media following the methodology described previously<sup>38</sup> with slight modifications. Following elution from the sulphopropyl column in 20 mM phosphate, pH 5.8, all samples were treated with 10  $\mu$ L/ml of 2 M Tris, pH 8.5, to prevent crystallization and were then dialyzed into 50 mM Tris, pH 8.5, 100 mM KCl, 10 mM DTT, and 20 mg/l complete protease inhibitor for long-term storage at 4 °C. The purity and identity of the protein samples were assessed by SDS-PAGE, ESI-MS, and N-terminal sequencing of the first eight residues in the protein.

Isotopically labeled reagents were purchased from (Spectra Stable Isotopes Inc., Columbia, MD, USA), Goss Scientific Instruments (Great Baddow, Essex, UK), or Cambridge Isotope Laboratories (Andover, MA, USA). W82R samples were prepared by growing cells in M9 medium prepared with 50% D<sub>2</sub>O/50% H<sub>2</sub>O (v/v) and supplemented with 40 mg/l FeCl<sub>3</sub> and 1% (v/v) 100 $\times$  BME vitamin solution. Isotopes were introduced into the growth medium by the addition of 4 g/l of uniformly <sup>13</sup>C-labeled 50% <sup>2</sup>H-random-labeled glucose and 1 g/l of <sup>15</sup>NH<sub>4</sub>Cl. Plasmid selection during expression was maintained using 100  $\mu$ g/l carbenicillin. Cells were grown to an *A*<sub>600</sub> of 0.8 starting from a labeled starter culture grown overnight and were induced with 0.1 mM isopropyl- $\beta$ -D-thiogalactopyranoside and harvested after a further 8 h of growth.

<sup>15</sup>N-labeled wild-type MS2 coat protein was prepared identically with wild-type protein, with the exception that the cells were grown in <sup>15</sup>NH<sub>4</sub>Cl-containing minimal medium.

#### RNAs

Synthetic oligonucleotides encompassing the TR sequence (Figure 1) or the -5C variant were prepared as described previously.<sup>39</sup> The -5C variant binds to the coat protein dimer significantly more tightly than the wild-type sequence, with half-lives of >300 and 42 s, respectively.<sup>27</sup> The crystal structure of the -5C variant bound to the coat protein in a preformed capsid<sup>10</sup> shows that this effect is likely the result of an additional intramolecular hydrogen bond within the RNA, from the exocyclic amino group of C to the phosphodiester bridging the -6 and -7 positions. This presumably stabilizes the conformation of the RNA in solution closer to that resulting from protein binding reducing the free energy cost of distortion. There are no differences in contacts to the protein. Therefore, the -5C variant was used in the NMR experiments.

#### Reassembly reactions

##### Reaction conditions

Assembly reactions were carried out in 40 mM ammonium acetate, pH 6.8, at 8 °C. Initially, CP<sub>2</sub> was mixed with TR to a final concentration of 8  $\mu$ M, and further aliquots of CP<sub>2</sub> were added as indicated. The concentration of the CP<sub>2</sub> solution was calculated from UV absorbance, and the sample was passed through a 0.2- $\mu$ m syringe filter prior to measurement.

### Electrospray ionization-mass spectrometry

Samples were analyzed by positive ionization nanoelectrospray using an LCT Premier Mass Spectrometer (Waters Corp., Manchester, UK) with collisional cooling capabilities equipped with a NanoMate (Advion, Inc., Ithaca, NY, USA) temperature-controlled automated sample handling and ionization interface, which was thermostated at 8 °C. A capillary voltage of 1.9 kV was set with a nitrogen gas flow of 0.5 psi for sample introduction and ionization. The sampling cone voltage was optimized at 80 V, ion guide 1 at 130 V, and aperture 1 at 60 V. The analyzer pressure was  $7.2 \times 10^{-7}$  mBar and the pressure in the transfer optics region was  $4.9 \times 10^{-3}$  mBar. Data were acquired over the range  $m/z$  500–30,000 and data processing was performed using the MassLynx software supplied with the mass spectrometer. An external calibration using CsI clusters was applied to the data.

### Gel filtration-light-scattering assays

For assays of reassembly by gel filtration, reactions at differing stoichiometries were set up with slight variations on the protocol for mass spectrometry. For convenience, the concentration of CP<sub>2</sub> in the 1:1 reaction was 10 and 20 μM for the 1:2 sample, with the final TR concentration being 10 μM for both. The final 100-μl reaction volume was made up with assembly buffer (40 mM ammonium acetate, pH 6.8), and the samples were incubated on ice for various times until injection onto a Tricon high-performance column of Sepharose 6 10/300 GL (GE Healthcare Bio-Sciences AB, Sweden) thermostated at 12 °C. The eluant was 50 mM Tris acetate, pH 7.4 (chosen to avoid degassing problems with ammonium acetate), at a flow rate 0.36 ml/min.

The outflow of the column was monitored for UV absorption and for light-scattering using a static light-scattering detector PD 2000 (Precision Detectors, Inc., Bellingham, Mass., USA).

### NMR data collection

Conditions providing optimum solubility of CP<sub>2</sub> were determined from a set of hanging drop trials<sup>40</sup> in a range of buffers at a pH between 6.0 and 7.0 at 25 °C. Suitable solution conditions were thus determined to be 1% v/v glycerol, 100 mM Bis-Tris, pH 7.0, 100 mM KCl. In addition, the buffer also contained 10% D<sub>2</sub>O, 2 mg/ml complete protease inhibitor, 10 mM NaN<sub>3</sub>, 2 mM EDTA, 1 mM DTT, and 0.05 mM sodium 2,2-dimethyl-2-silapentane-5-sulfonate. Samples were introduced into the required solution conditions and concentrated using Vivaspin centrifugal concentrators (Sartorius, Ltd., Epsom, UK). All experiments were carried out in Shigemi microcells (Shigemi, Inc. Allison Park, PA, USA) containing 250–300 μl of solution with concentrations of CP<sub>2</sub> (W82R) between 0.5 and 1.0 mM. NMR experiments were measured on a Unity Inova 750-MHz NMR spectrometer (Varian, Inc., Palo Alto, CA USA) equipped with four radio-frequency channels and a triple-resonance (<sup>1</sup>H-<sup>15</sup>N-<sup>13</sup>C) room temperature probe with Z axis magnetic field gradients. The set of triple-resonance NMR experiments applied to the CP<sub>2</sub> complex consisted of <sup>1</sup>H-<sup>15</sup>N heteronuclear single-quantum coherence, HNCA, HNcoCA, HNCO, HNcaCO, HNcaCB, and HncocaCB spectra, which were acquired using Varian Biopack pulse sequences. In the case of the HNCA, HNcoCA, HncaCB, and HncocaCB experiments the pulse sequences were

modified<sup>41</sup> so as to maintain optimal transverse relaxation optimized spectroscopy effects in partially deuterated proteins during the C<sup>α</sup>/<sup>1</sup>H chemical shift evolution period. The HNcoCA sequence was modified to have a C<sup>α</sup> selective inversion pulse. <sup>1</sup>H-<sup>15</sup>N heteronuclear nuclear Overhauser effect spectra were acquired on the CP<sub>2</sub> dimer using a Varian Unity Inova 600-MHz NMR spectrometer equipped with four radiofrequency channels and a triple-resonance (<sup>1</sup>H-<sup>15</sup>N-<sup>13</sup>C) probe with Z axis magnetic field gradients.

In contrast to the experiments measured on CP<sub>2</sub> the, HNCA and HNcoCA spectra of the CP<sub>2</sub>:TR complex were measured using a Varian cryogenic probe with receiver coils and preamplifiers cooled to -253 °C. The limited assignment of the CP<sub>2</sub>:TR complex used spectra from <sup>1</sup>H-<sup>15</sup>N heteronuclear single-quantum coherence, HNCA, and HNcoCA experiments. The buffer for the CP<sub>2</sub>:TR samples used for NMR spectroscopy was supplemented with 800 U RNasin (Promega, Southampton, UK), which was brought to the final solution conditions using a Vivaspin centrifugal concentrator. Solutions of the CP<sub>2</sub>:TR complex were prepared by addition of a slight excess of TR to CP<sub>2</sub> as determined by the loss of signals from the *apo* CP<sub>2</sub> dimer.

All experiments were measured with the use of transverse relaxation optimized spectroscopy to reduce T<sub>2</sub> relaxation effects and improve dispersion. Data processing and analysis were carried out using NMRPipe<sup>42</sup> and Nmrview<sup>43</sup> on computers running under Linux. Segments of assigned residues from CP<sub>2</sub> were mapped to the primary sequence of the protein using the program Mapper2.<sup>44</sup> <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N chemical shifts were referenced directly (<sup>1</sup>H) and indirectly (<sup>13</sup>C, <sup>15</sup>N) to sodium 2,2-dimethyl-2-silapentane-5-sulfonate.

## Acknowledgements

We thank Drs Hugo Lago and Andrew Smith for preliminary mass spectrometry and help with data collection, respectively, and Dr Abdul Rashid for production of some of the materials used in these experiments. We also thank Professors Sheena Radford (Leeds), Bentley Fane (Arizona), and Jon King (Cambridge, Mass.) and Dr Reidun Twarock (York) for their helpful discussions of the results and the manuscript. This work was supported in part by the Leverhulme Trust, the Wellcome Trust, and the UK BBSRC. OR is a Wellcome Trust-funded PhD student.

## References

1. Caspar, D. & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24.
2. Twarock, R. (2004). A tiling approach to virus capsid assembly explaining a structural puzzle in virology. *J. Theoret. Biol.* **226**, 477–482.
3. Valegård, K., Liljas, L., Fridborg, K. & Unge, T. (1990). The three-dimensional structure of the bacterial virus MS2. *Nature*, **345**, 36–41.
4. Golmohammadi, R., Valegård, K., Fridborg, K. &

- Liljas, L. (1993). The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J. Mol. Biol.* **234**, 620–639.
5. Sorger, P. K., Stockley, P. & Harrison, S. (1986). Structure and assembly of turnip crinkle virus. 2. Mechanism of Reassembly in vitro. *J. Mol. Biol.* **191**, 639–658.
6. Rossmann, M. G., Abad-Zapatero, C., Erickson, J. W. & Savithri, H. S. (1983). RNA-protein interactions in some small plant viruses. *J. Biomol. Struct. Dyn.* **1**, 565–579.
7. Rossmann, M. G. & Johnson, J. E. (1989). Icosahedral RNA virus structure. *Annu. Rev. Biochem.* **58**, 533–573.
8. Zlotnick, A., Aldrich, R., Johnson, J. M., Ceres, P. & Young, M. J. (2000). Mechanism of capsid assembly for an icosahedral plant virus. *Virology*, **277**, 450–456.
9. Witherell, G. W., Gott, J. M. & Uhlenbeck, O. C. (1991). Specific interaction between RNA phage coat proteins and RNA. *Prog. Nucl. Acid Res. Mol. Biol.* **40**, 185–220.
10. Valegård, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. (1994). Crystal structure of a bacteriophage RNA coat protein operator system. *Nature*, **371**, 623–626.
11. Stockley, P. G., Stonehouse, N. J. & Valegård, K. (1994). Molecular mechanism of RNA phage morphogenesis. *Int. J. Biochem.* **26**, 1249–1260.
12. Lago, H., Parrott, A. M., Moss, T., Stonehouse, N. J. & Stockley, P. G. (2001). Probing the kinetics of formation of the bacteriophage MS2 translational operator complex: identification of a protein conformer unable to bind RNA. *J. Mol. Biol.* **305**, 1131–1144.
13. Horn, W. T., Tars, K., Grah, E., Helgstrand, C. & Baron, A. J. (2006). Lago, et al. Structural basis of RNA binding discrimination between bacteriophages Q $\beta$  and MS2. *Structure*, **14**, 487–495.
14. Beckett, D. & Uhlenbeck, O. C. (1988). Ribonucleoprotein complexes of R17 coat protein and a translational operator analog. *J. Mol. Biol.* **204**, 927–938.
15. Beckett, D., Wu, H.-N. & Uhlenbeck, O. C. (1988). Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *J. Mol. Biol.* **204**, 939–947.
16. Loo, J. A. (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16**, 1–23.
17. Ashcroft, A. E. (2005). Recent developments in electrospray ionisation mass spectrometry: noncovalently bound protein complexes. *Nat. Prod. Rep.* **22**, 452–464.
18. Krutchinsky, A. N., Chernushevich, I. V., Spicer, V. L., Ens, W. & Standing, K. G. (1998). Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer. *J. Am. Soc. Mass Spectrom.* **9**, 569–579.
19. Chernushevich, I. V. & Thomson, B. A. (2004). Collisional cooling of large ions in electrospray mass spectrometry. *Anal. Chem.* **76**, 1754–1760.
20. Van Berkel, W. J. H., Van den Heuvel, R. H. H., Versluis, C. & Heck, A. J. R. (2000). Detection of intact megaDalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry. *Protein Sci.* **9**, 435–439.
21. Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402–1407.
22. Sugiyama, T. & Nakada, D. (1967). Control of translation of MS2 RNA cistrons by MS2 coat protein. *Biochemistry*, **57**, 1750–1774.
23. Stockley, P. G., Ashcroft, A. E., Francese, S., Thompson, G. S., Ranson, N. A., Smith, A. M. et al. (2005). Dissecting the fine details of assembly of a T=3 phage capsid. *J. Theoret. Med.* **6**, 119–125.
24. Mastico, R. A., Talbot, S. J. & Stockley, P. G. (1993). Multiple presentation of foreign peptides on the surface of an RNA-free spherical bacteriophage capsid. *J. Gen. Virol.* **74**, 541–548.
25. Tito, M. A., Tars, K., Valegård, J., Hajdu, J. & Robinson, C. V. (2000). Electrospray time-of-flight mass spectrometry of the intact MS2 virus capsid. *J. Am. Chem. Soc.* **122**, 3550–3551.
26. Carey, J. & Uhlenbeck, O. C. (1983). Kinetic and thermodynamic characterization of the R17 coat protein-ribonucleic acid interaction. *Biochemistry*, **22**, 2610–2615.
27. Talbot, S. J., Goodman, S., Bates, S. R. E., Fishwick, C. W. G. & Stockley, P. G. (1990). Use of synthetic oligoribonucleotides to probe RNA-protein interactions in the MS2 translational operator complex. *Nucl. Acids Res.* **18**, 3521–3528.
28. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51–55.
29. Case, D. A. (1995). Calibration of ring-current effects in proteins and nucleic acids. *J. Biomol. NMR*, **6**, 341–346.
30. Valegård, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G. & Liljas, L. (1997). The three dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence specific protein-RNA interactions. *J. Mol. Biol.* **270**, 724–738.
31. Koning, R., van den Worm, S., Plaisier, J. R., van Duin, J., Pieter Abrahams, J. & Koertgen, H. (2003). Visualization by cryo-electron microscopy of genomic RNA that binds to the protein capsid inside bacteriophage MS2. *J. Mol. Biol.* **332**, 415–422.
32. Berkhout, B. & van Duin, J. (1985). Mechanism of translational coupling between coat protein and replicase genes of RNA bacteriophage MS2. *Nucl. Acids Res.* **13**, 6955–6967.
33. Caspar, D. (1980). Movement and self-control in protein assemblies. Quasi-equivalence revisited. *Biophys. J.* **32**, 103–138.
34. Bamford, D. H., Grimes, J. M. & Stuart, D. I. (2005). What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* **15**, 655–663.
35. Kuznetsov, Y. G., Daijogo, S., Zhou, J., Semler, B. L. & McPherson, A. (2005). Atomic force microscopy analysis of icosahedral virus RNA. *J. Mol. Biol.* **347**, 41–52.
36. Wu, M., Brown, W. L. & Stockley, P. G. (1995). Cell-specific delivery of bacteriophage-encapsidated ricin-A chain. *Bioconjug. Chem.* **6**, 587–595.
37. Wang, Q., Lin, T. W., Tang, L., Johnson, J. E. & Finn, M. G. (2002). Icosahedral virus particles as addressable nanoscale building blocks. *Angew Chem. Int. Ed. Engl.* **41**, 459–462.
38. Ni, C. Z., Kodandapani, R., Wickersham, J., Peabody, D. S., Ely, K. R. & Syed, R. (1995). Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. *Structure*, **3**, 255–263.
39. Murray, J. B., Collier, A. K. & Arnold, J. R. P. (1994). A general purification procedure for chemically synthesised oligoribonucleotides. *Anal. Biochem.* **218**, 177–184.

40. Lepre, C. A. & Moore, J. M. (1998). Microdrop screening: a rapid method to optimize solvent conditions for NMR spectroscopy of proteins. *J. Biomol. NMR*, **12**, 493–499.
41. Eletsky, A., Kienhofer, A. & Pervushin, K. (2001). TROSY NMR with partially deuterated proteins. *J. Biomol. NMR*, **2**, 177–180.
42. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe—a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR*, **6**, 277–293.
43. Johnson, B. A. & Blevins, R. A. (1994). NMR View—a computer-program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
44. Güntert, P., Salzmann, M., Braun, D. & Wuthrich, K. (2000). Sequence-specific NMR assignment of proteins by global fragment mapping with the program MAPPER. *J. Biomol. NMR*, **18**, 129–137.

*Edited by M. F. Summers*

(Received 21 January 2007; received in revised form 4 March 2007; accepted 5 March 2007)  
Available online 15 March 2007

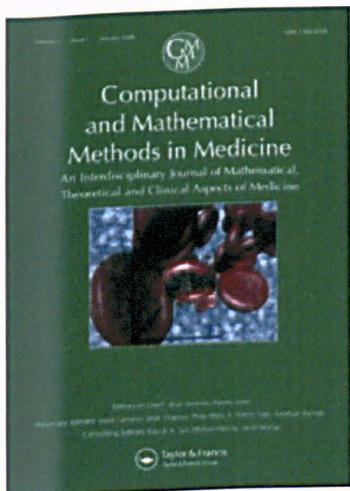
This article was downloaded by: [University of Leeds]

On: 14 January 2009

Access details: Access Details: [subscription number 773557621]

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Computational and Mathematical Methods in Medicine

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713653639>

### RNA packing specificity and folding during assembly of the bacteriophage MS2

Ottar Rolfsson <sup>a</sup>; Katerina Toropova <sup>a</sup>; Victoria Morton <sup>a</sup>; Simona Francese <sup>a</sup>; Gabriella Basnak <sup>a</sup>; Gary S. Thompson <sup>a</sup>; Stephen W. Homans <sup>a</sup>; Alison E. Ashcroft <sup>a</sup>; Nicola J. Stonehouse <sup>a</sup>; Neil A. Ranson <sup>a</sup>; Peter G. Stockley <sup>a</sup>

<sup>a</sup> Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK

Online Publication Date: 01 September 2008

To cite this Article Rolfsson, Ottar, Toropova, Katerina, Morton, Victoria, Francese, Simona, Basnak, Gabriella, S. Thompson, Gary, Homans, Stephen W., Ashcroft, Alison E., Stonehouse, Nicola J., Ranson, Neil A. and Stockley, Peter G.(2008)'RNA packing specificity and folding during assembly of the bacteriophage MS2',Computational and Mathematical Methods in Medicine,9:3,339 — 349

To link to this Article: DOI: 10.1080/17486700802168445

URL: <http://dx.doi.org/10.1080/17486700802168445>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## RNA packing specificity and folding during assembly of the bacteriophage MS2

Ottar Rolfsson, Katerina Toropova, Victoria Morton, Simona Francese<sup>1</sup>, Gabriella Basnak,  
Gary S. Thompson, Stephen W. Homans, Alison E. Ashcroft, Nicola J. Stonehouse, Neil  
A. Ranson and Peter G. Stockley\*

*Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK*

*(Received 18 February 2008; final version received 1 March 2008)*

Using a combination of biochemistry, mass spectrometry, NMR spectroscopy and cryo-electron microscopy (cryo-EM), we have been able to show that quasi-equivalent conformer switching in the coat protein (CP) of an RNA bacteriophage (MS2) is controlled by a sequence-specific RNA–protein interaction. The RNA component of this complex is an RNA stem-loop encompassing just 19 nts from the phage genomic RNA, which is 3569 nts in length. This binding results in the conversion of a CP dimer from a symmetrical conformation to an asymmetric one. Only when both symmetrical and asymmetrical dimers are present in solution is assembly of the  $T = 3$  phage capsid efficient. This implies that the conformers, we have characterized by NMR correspond to the two distinct quasi-equivalent conformers seen in the 3D structure of the virion. An icosahedrally-averaged single particle cryo-EM reconstruction of the wild-type phage (to  $\sim 9 \text{ \AA}$  resolution) has revealed icosahedrally ordered density encompassing up to 90% of the single-stranded RNA genome. The RNA is seen with a novel arrangement of two concentric shells, with connections between them along the 5-fold symmetry axes. RNA in the outer shell interacts with each of the 90 CP dimers in the  $T = 3$  capsid and although the density is icosahedrally averaged, there appears to be a different average contact at the different quasi-equivalent protein dimers: precisely the result that would be expected if protein conformer switching is RNA-mediated throughout the assembly pathway. This unprecedented RNA structure provides new constraints for models of viral assembly and we describe experiments aimed at probing these. Together, these results suggest that viral genomic RNA folding is an important factor in efficient assembly, and further suggest that RNAs that could sequester viral CPs but not fold appropriately could act as potent inhibitors of viral assembly.

**Keywords:** RNA–protein interactions; MS2; ssRNA virus; genomic RNA structure and folding; cryo-electron microscopy; viral assembly

### Introduction

The ssRNA viruses are one of the largest groups of viruses, and include human immunodeficiency virus, hepatitis C virus, Norwalk viruses and the *Picornaviridae* such as rhinovirus, poliovirus, hepatitis A and foot and mouth disease virus, as well as plant viruses such as tomato bushy stunt virus [25]. These viral pathogens infect human, animal and plant hosts and for many of these novel routes towards anti-viral therapy are urgently required. All these viruses face the common challenge of packaging their nucleic acids into protective shells composed of a defined number of one, or very few, type(s) of coat protein (CP) subunits, reflecting constraints of genetic

---

\*Corresponding author. Email: stockley@bmb.leeds.ac.uk

economy and quasi-equivalence [6,8]. Many ssRNA viruses have evolved lifecycles during which capsid assembly is triggered by sequence/structure-specific interaction between CPs and their RNA genomes [26,38]. A major unresolved problem in structural virology, however, has been to understand the detailed molecular mechanism(s) that gives rise to the highly efficient assembly of capsids of the correct size and symmetry in such systems. This is especially true for viruses whose capsids exhibit quasi-equivalent symmetry in their CP lattices [6]. Various proposals for formation of protein shells based on the initial formation of 3-fold [26], or 5-fold [23] assembly initiation complexes have been made previously. Capsid assembly is, however, spontaneous and usually very rapid, with the result that it has been technically difficult to isolate and characterize intermediates on the pathway to the final products beyond these 'initiation complexes' [22,41]. Understanding these processes in mechanistic detail may well lead to novel routes for anti-viral therapy.

The RNA bacteriophage MS2 is an ideal model for investigating such phenomena owing to the extensive biochemical and structural information that is available [12,14,27,34,35,39]. We have therefore undertaken detailed mechanistic and structural studies of this phage in order to address the specific questions of how the genomic RNA is encapsidated specifically, how the formation of the  $T = 3$  quasi-equivalent shell is controlled during the assembly reaction, and what role(s), if any, the ssRNA genome plays during these processes. MS2 is a member of the *Leviviridae* family of viruses that infect male *Escherichia coli* cells via an initial attachment to the bacterial F-pilus [21,27]. It has a single-stranded, positive-sense RNA genome of 3569 nucleotides that encodes just four gene products: CP, replicase, lysis protein and maturation protein (Figure 1). CP is the most highly expressed of the gene products and 180 copies assemble to form an icosahedral protein shell that encapsidates the genome in the mature virion. A single copy of the maturation protein is also incorporated into the virion and functions during infection by binding the F-pilus of target cells. A complex between the maturation protein and the genomic RNA is the only viral component to enter host cells during infection [40].

The crystal structure of the wild-type MS2 bacteriophage has been determined to 2.8 Å resolution [9,34]. The CP fold was unique amongst known icosahedral viruses at that time, although it has since been shown to be shared by the other *Leviviridae* CPs [32]. The main chain folds into a five-stranded anti-parallel  $\beta$ -sheet with two anti-parallel  $\beta$ -strands folding over it at the N-terminus and with a kinked  $\alpha$ -helix at the C-terminus (Figure 2). The C-terminal  $\alpha$ -helices of two CP monomers interdigitate to form non-covalent dimers ( $CP_2$ ) in the capsid and these can be isolated as dimers in solution by acid dissociation [31]. In the  $T = 3$  capsid, the CP is found in three distinct conformations, termed A, B and C, consistent with the quasi-equivalent symmetry required to construct a  $T = 3$  structure. The main site of variation between conformers is in the loop between the F and G,  $\beta$ -strands (FG-loop), which is extended in A and C conformers, but bent back toward the main body of the subunit in the B conformer (Figure 2). The capsid thus contains two types of dimer; an asymmetric A/B dimer, and a symmetric C/C dimer. The positioning of these different building blocks within the protein shell controls the size and symmetry of the viral particle. Unusually, the capsid has large ( $\sim 14$  Å dia.) pores at the 5-fold symmetry axes and it is not known whether these play a role in the viral life cycle, although, mutation of a conserved proline residue (Pro78) lining the pore at 5-fold axes results in production of assembled  $T = 3$  capsids but fails to produce phage from an infectious recombinant clone [11].

Capsid reassembly can be triggered *in vitro* by a sequence-specific RNA-protein interaction between acid-disassembled CP subunits, which form dimers at more neutral pHs and an RNA stem-loop (TR) of just 19 nt derived from the genomic sequence that encompasses the start codon of the viral replicase ([2,3]; Figure 2). RNA-protein binding thus achieves two functions via a single molecular recognition event: translational repression of the replicase and creation of an assembly competent complex on viral RNA.

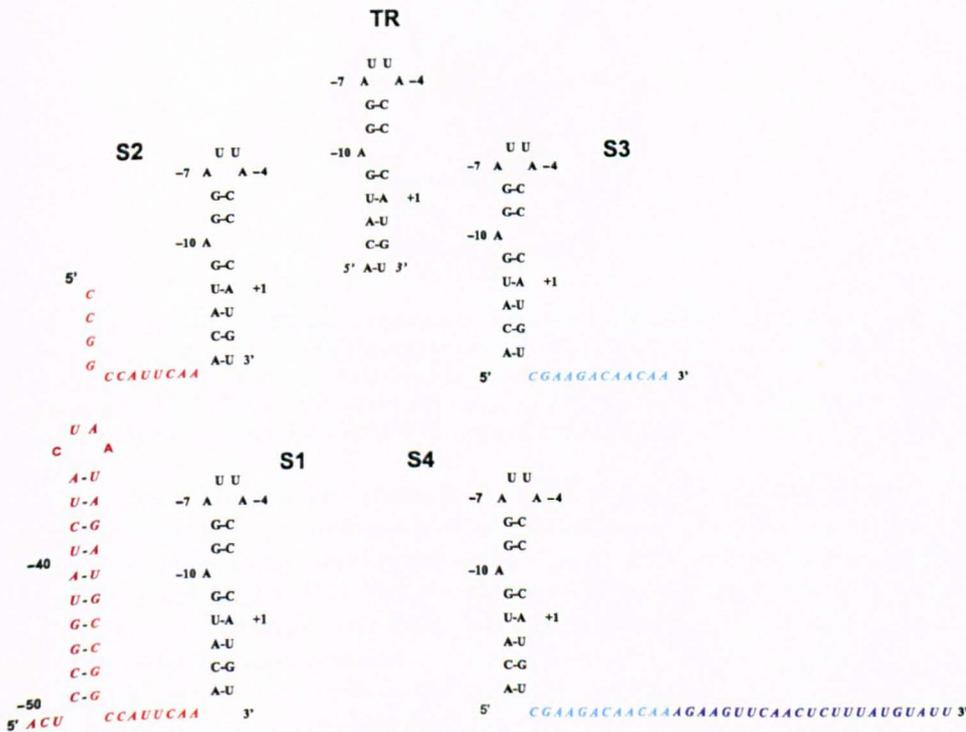
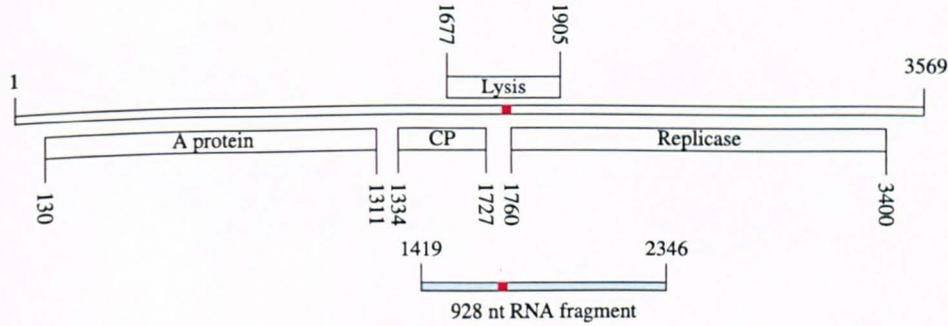


Figure 1. Genomic map of bacteriophage MS2 and sequences of the RNAs used in reassembly experiments. The upper panel shows a diagrammatic representation of the single-stranded 3569 nt long MS2 genome and the positions of the genes it encodes. The red bar shows the location of the TR assembly stem-loop. The sub-genomic fragment is shown highlighted in light blue. The lower panel shows the sequences and putative secondary structures of the short RNA fragments used for reassembly experiments. Fragments S1–S4 represent 12 and 35 nt extensions at the 5' (increasing red) or 3' (increasing blue) ends of the TR stem-loop, respectively. The TR fragment is numbered with respect to the start codon of the replicase cistron (A<sub>-1</sub>UG).

## Results and discussion

### Specific encapsidation of the short genomic RNA

Results from co-infections by two closely related RNA phages, MS2 and Q $\beta$ , in the same bacterial cell suggest that there is sequence selective encapsidation by the cognate CP subunits of their respective genomic RNAs [16]. The molecular basis of this selectivity has been established by the pioneering *in vitro* experiments of Uhlenbeck and colleagues, who showed

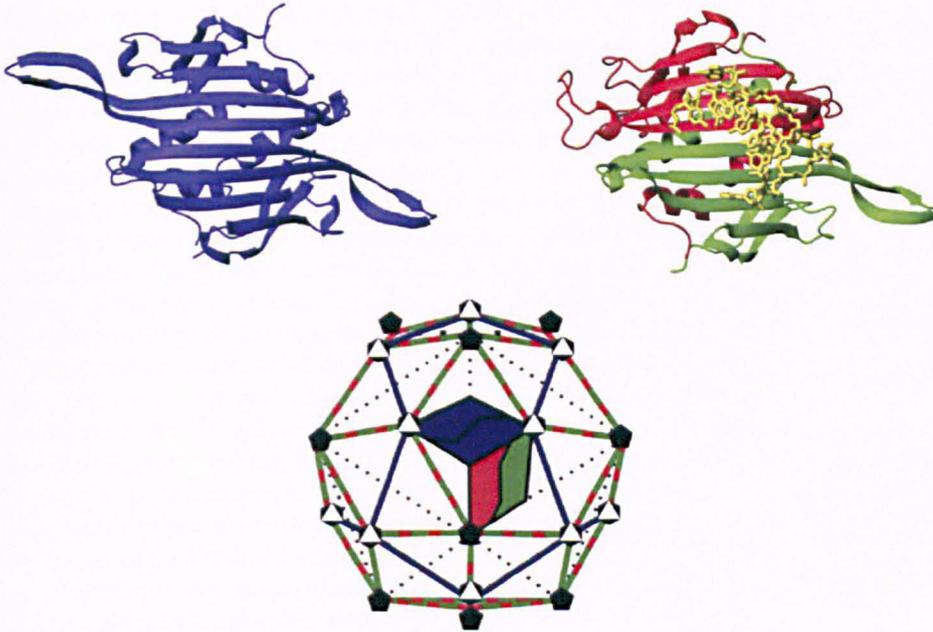


Figure 2. Structures of the MS2 quasi-equivalent coat protein dimers and their relationship within a  $T = 3$  shell. The C/C (red/red) and A/B (blue/green) quasi-equivalent dimers are shown as ribbon representations viewed along the 2-fold axes from the inside of the phage particle. The TR RNA (yellow stick model) is shown bound to the A/B dimer. Structures are taken from pdb1ZDH. The relationship of A/B and C/C dimers to the final  $T = 3$  capsid is shown below. Adapted from Ref. [29]. Available in colour online.

that sequence-specific interaction between R17 (an MS2 homologue) and its genomic RNA was controlled by a single 19 nt stem-loop within the 3569 nt genomic RNA [4,5]. We have recently extended these studies using X-ray crystallography to provide the first atomic level mechanism for packaging discrimination [12]. The *in vitro* affinity of the MS2 stem-loop for its CP is very similar to that for the intact genomic RNA. This recognition sequence encompasses the 5' start codon of the phage replicase gene, and therefore also provides a molecular explanation for the translational repression of the replicase cistron by CP expression [17]. The 'appearance' of the 19 nt translational operator stem-loop (TR) is controlled by phage gene expression. Initially, following phage RNA entry into the bacterium, the TR sequence is sequestered in a long-range secondary structure, base pairing with a portion of the CP gene – the 'Min Jou' sequence [13]. CP gene translation is required to free the TR sequence and initiate folding into the stem-loop operator. Beckett and Ulhenbeck showed that in isolation this TR fragment triggers reassembly of  $T = 3$  shells by the CP more efficiently than polyA [2,3].

These observations raise a number of important questions. How does the TR stem-loop fold into the operator conformation? How is the TR sequence recognized specifically by the CP dimer? How does the CP dimer ( $CP_2$ ): TR complex lead to assembly of a shell with the correct size and symmetry? Single-molecule fluorescence data, that will be described in detail elsewhere (Gell *et al.* in prep), suggest that TR folds as a two-state system, consistent with its role as a genetic switch. Previous stopped-flow fluorescence measurements suggest that isolated 19 nt stem-loops bind to CP dimers in a diffusion-controlled reaction, although, these data showed some evidence that protein conformational changes were occurring during the binding [14].

Expression of a recombinant MS2 CP gene in *E. coli* leads to formation of  $T = 3$  shells containing small amounts of cellular RNA [18]. These are easily purified and crystallize

isomorphously with the wild-type phage, even though they lack the genomic RNA and the single copy of the maturation protein. Soaking RNA oligonucleotides encompassing the TR sequence [35], sequence or chemical variants [10,36], or even aptamer consensus sequences [7,24], into such crystals allows the RNA to access the inner surface of the protein shell, presumably via the pores through the capsid, where they bind to every CP dimer. This technique has allowed the 3D structures of a large number of different RNA–protein complexes to be determined by X-ray diffraction difference maps. TR binds to A/B quasi-equivalent dimers in a unique orientation, allowing the interaction to be resolved at atomic resolution (Figure 2). It is clear from such experiments that the TR binds across the inner face of the CP<sub>2</sub>, a sheet comprising ten β-strands. The RNA in this site is at least 12 Å away from the nearest amino acid residues of the loops whose conformations define the quasi-symmetry of the capsid. Building a  $T = 3$  capsid requires these loop conformations to be set in a local symmetry context and their separation from TR binding site does not immediately reveal the mechanism underlying this switch. The conformation of the TR bound to protein differs from that seen in solution because the  $-10$  A residue (see Figure 1 legend for numbering scheme) becomes extruded from the stem, which takes up an A duplex structure. The  $-10$  A base and the loop residue  $-4$  A bind into identical pockets, but in different orientations, on the surface of each protein monomer. Thus, the bulk conformation of the CP dimer is unchanged by binding TR RNA, consistent with the very rapid on rate observed in the stopped-flow experiments [14]. These studies therefore did not clarify how quasi-equivalent conformer switching is achieved during assembly.

### *The mechanism of $T = 3$ shell formation*

The major insight into the molecular mechanism controlling  $T = 3$  shell formation was made by studying TR-mediated reassembly of CP dimers by non-covalent mass spectrometry [29]. Reaction mixtures containing equimolar amounts of CP<sub>2</sub> and TR rapidly formed the ‘initiating’ RNA–protein complex but these were only very slowly converted into species with higher stoichiometries. Addition of a further aliquot of CP<sub>2</sub> lacking RNA, however, resulted in rapid formation of higher order species including broad peaks in the mass spectrum that correspond to the peaks observed when the recombinant  $T = 3$  shell is analysed. It appears that both RNA-bound and RNA-free forms of the CP<sub>2</sub> are required for efficient and rapid capsid assembly.

Since there are two types of CP<sub>2</sub> dimer conformer in the final capsid this result is suggestive of RNA-binding resulting in conformational change in the protein dimer. This possibility was investigated using NMR spectroscopy to determine the chemical shift values of the CP subunits with and without bound TR using a protein mutant, W82R that does not assemble beyond dimer [20]. The data showed very clearly that in the absence of RNA the chemical shifts of the residues in the FG-loop are the same in both subunits, whereas they become distinct when the TR stem loop is bound. The RNA binding site is sufficiently far from the FG-loops that these effects are not due to direct alteration in the local chemical environment. In other words, the NMR data show unambiguously that TR acts allosterically to promote formation of an asymmetric (presumably A/B-like) CP<sub>2</sub> from an RNA-free species that is symmetric. Analysis of the intermediate species in the mass spectra of the reassembly reaction (Figure 3) revealed the presence of complexes with the following stoichiometries: CP<sub>2</sub>:TR of mass 33.5 kDa (Figure 3; component A), [3(CP<sub>2</sub>:TR) + 3CP<sub>2</sub>] of mass 182.7 kDa (Figure 3; component B), [3(CP<sub>2</sub>:TR) + 2CP<sub>2</sub> + CP] of mass 169.2 kDa (Figure 3; component C) and [5(CP<sub>2</sub>:TR) + 5CP<sub>2</sub>] of mass 304.8 kDa (Figure 3; component D). There was no sign of a species corresponding to a fully formed 5-fold axis 5(CP<sub>2</sub>:TR), suggesting that these species result from formation of a 3-fold axis assembly intermediate. The [5(CP<sub>2</sub>:TR) + 5CP<sub>2</sub>] species has only recently been tentatively assigned and could correspond to a 5-fold intermediate with additional subunits

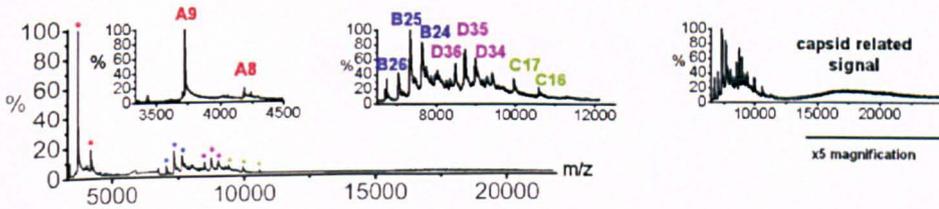


Figure 3. Mass spectrum of a TR-mediated assembly reaction. Virus reassembly reaction in the ratio 2:1 CP<sub>2</sub>:TR after 30 min. The spectra were acquired over the range *m/z* 500–30,000, for sample in ammonium acetate (40 mM) at pH 5.2–5.7. The observed intermediates are labelled as follows; A = CP<sub>2</sub>:TR; B = 182.7 kDa, previously assigned as [3(CP<sub>2</sub>:TR) + 3CP<sub>2</sub>], C = 169.2 kDa species assigned as [3(CP<sub>2</sub>:TR) + 2CP<sub>2</sub> + CP], and D = 304.8 kDa species assigned for the first time as [5(CP<sub>2</sub>:TR) + 5CP<sub>2</sub>]. Available in colour online.

bound. A minimal 5(CP<sub>2</sub>:TR) possible intermediate could be unstable, explaining why it has not been seen before. Alternatively, combinatorial analysis (Stansifer and Twarock, pers. comm.) suggests that the 3-fold species corresponds to the more probable of these two possibilities, implying that despite the suggestive 5-fold stoichiometry this intermediate is more likely to form as an extension of the 3-fold stoichiometry, [3(CP<sub>2</sub>:TR) + 3CP<sub>2</sub>]. Once such a structure forms the phage has ensured that the protein will assemble a *T* = 3 structure and not a *T* = 1 capsid. Additional experiments, including ones with isotopically-labelled protein, showed that all the intermediates were competent for further assembly, *i.e.* were on the assembly pathway and that the unit of capsid growth is a CP<sub>2</sub>.

The data described above suggest a detailed molecular mechanism for *T* = 3 shell formation (Figure 4) in which TR binding to the CP dimer in solution induces a conformational change in the FG-loops of the protein monomers yielding an A/B-like dimer from a C/C-like precursor. In other words, the TR stem-loop acts as an allosteric ligand. Whilst this is a satisfactory explanation of the data for *in vitro* reassembly reactions with multiple copies of the TR:CP<sub>2</sub> complex present it does not necessarily apply to assembly *in vivo* where the genomic RNA contains only one copy of the TR stem-loop. There are a number of options in this case. Capsid assembly might need only to be initiated via a single TR stem-loop induced conformer switch. Alternatively, RNA sequences outside the TR region may mimic TR action during the assembly process and this would be consistent with a recent cryo-electron microscopy (cryo-EM) reconstruction of the wild-type phage particle at intermediate resolution [33]. This shows an array of RNA directly beneath the protein shell but with the average density below an A/B dimer differing from that below the C/Cs (Figure 5(A)).

In order to examine the putative role(s) of TR flanking sequences during assembly, we carried out reassembly reactions with RNA oligonucleotides carrying genomic 5' or 3' flanking sequences attached to the TR site (Figure 1, lower panel; [28]). All of the RNAs promote capsid formation, as judged by the positions of migration on gels of the RNAs at higher protein concentrations (Figure 5(B), upper panels). Perhaps, surprisingly the longer RNAs seem to be less efficient than TR in this reaction, cf. lanes 4–6 in the various gels. The two longest RNAs also produce smears that migrate slower than the capsid species (lanes 7–9). Electron micrographs suggest that this is the result of the formation of partially formed capsids. These results can be rationalized by reference to the model of the assembly reaction (Figure 4). In the presence of TR there are only two significantly populated types of protein conformer, A/B and C/C, and these can each be added to a growing *T* = 3 capsid. For the longer RNAs, however, there is the possibility that more than one CP dimer could bind to a single RNA, resulting in a complex mixture of unknown conformations of proteins. As a consequence, the concentrations

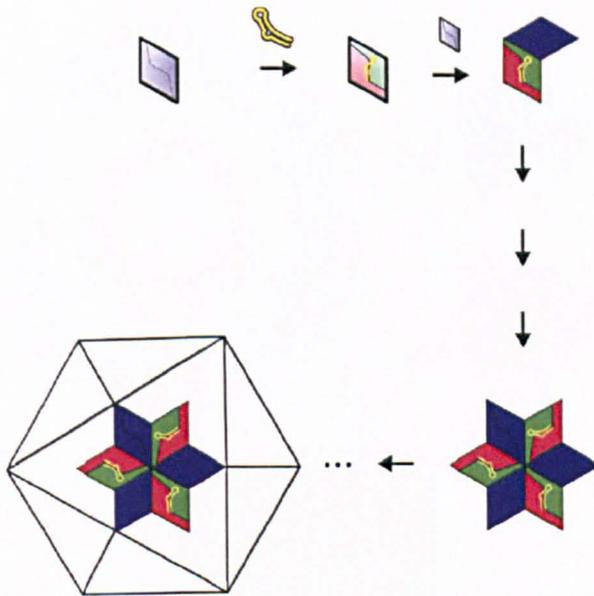


Figure 4. Model for the initiation of  $T = 3$  capsid assembly. A pair of chemically equivalent (C/C-like) subunits in solution (top left, shaded pink) binds a TR RNA resulting in a conformational switch into an A/B-like asymmetric dimer (shaded green and blue). This complex binds an additional C/C-like dimer creating the first higher order intermediate on the pathway to capsid formation. A variety of sub-pathways involving addition of CP dimers (arrows) can then generate the major assigned intermediate encompassing the first 3-fold axis of the virus particle. Adapted from Ref. [29]. Available in colour online.

of species able to participate in capsid formation are potentially lowered. It appears that the longer RNAs are acting as 'traps' on the assembly pathway.

*In vivo* of course the assembly reaction has to package a much longer genomic RNA and ensure that it acquires the complex 3D-fold seen in cryo-EM reconstructions that includes a second inner shell of RNA as well as the network seen directly below the CP (Figure 5(A)). The trapping of proteins seen in the reassembly reaction with the shorter fragments cannot simply scale up with the length of the RNA or assembly *in vivo* would never take place. A number of proposals have been made about the assembly of positive-strand RNA viruses, including the idea that only nascent transcripts from the polymerase are able to be packaged. There is some evidence of such requirements in pariacoto virus [37], a member of the *Nodaviridae* and brome mosaic virus (BMV) [1]. It has been argued that this type of packaging mechanism is driven by the fact that nascent transcripts fold using short range base pairing to create an RNA conformation easily adapted to the small volumes of the final capsid [15]. The apparent 'trapping' effect of the TR extension RNAs could be the result of such a mechanism. In order to test this idea, we have cloned an internal MS2 genomic fragment 928 nt long that encompasses the TR stem-loop signal. This has also been used successfully in *in vitro* reassembly experiments (Figure 5(B), lower panel). Once again a smear migrating more slowly than the *bona fide*  $T = 3$  capsid is created and electron micrographs suggest that this may be due to aggregation of partially formed capsids. Clearly there is no absolute requirement to have nascent transcripts, even with a large sub-genomic fragment, but at high protein:RNA molar ratios, RNAs may end up being shared between assembling capsids creating aggregates. Since there is little time during a phage infection to resolve such structures, the ratio of RNA:CP must be tightly controlled *in vivo* to avoid this problem.

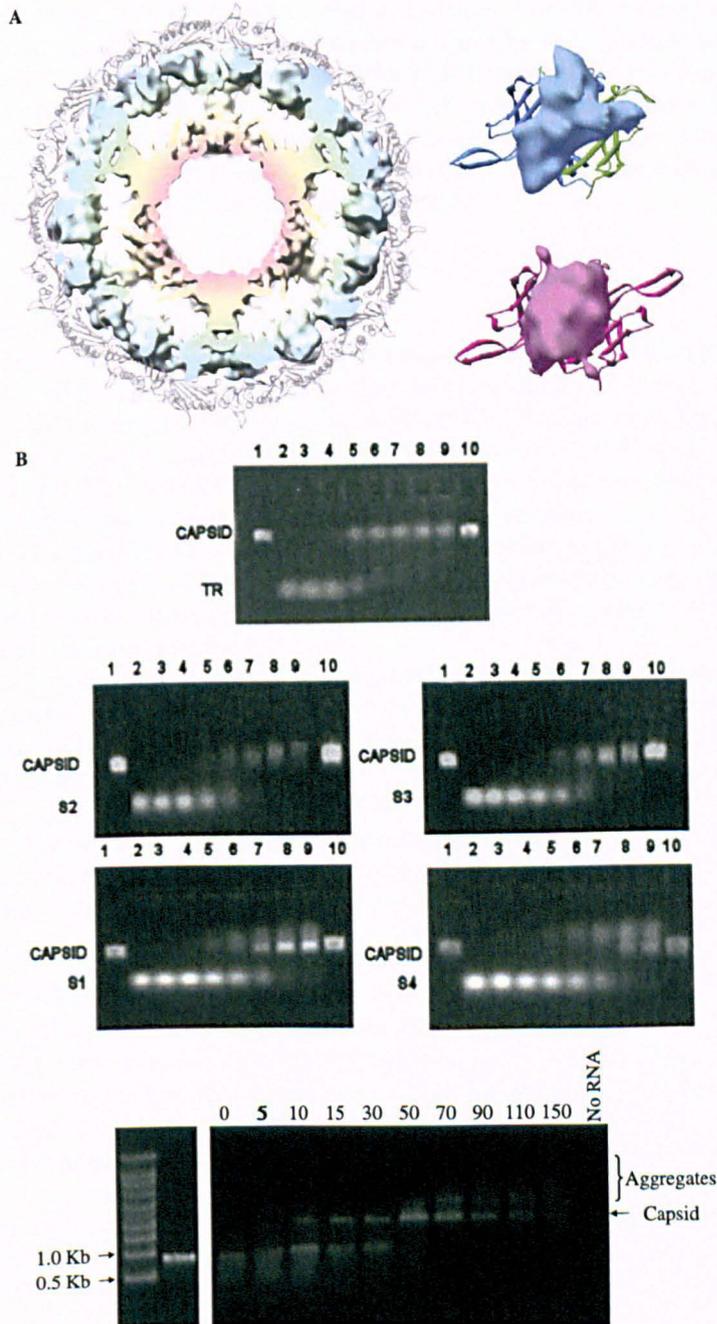


Figure 5. *In vitro* assembly of capsids using RNA fragments. (A) Cartoon (left) showing a cross-section view of the cryo-EM reconstruction of the MS2 wild-type phage virion [33]. Regions of the cryo-EM density corresponding to the MS2 CP shell (shown in cartoon representation in pale grey) have been masked away, leaving density for encapsidated genomic RNA. Genomic RNA is observed in two shells. The first lies directly beneath the protein capsid and maps precisely to the known binding sites for RNA on the icosahedral CP lattice. The second is at lower radius, and the two shells are linked by density on each icosahedral 5-fold axis. On the right the averaged electron density beneath A/B (blue/green) and C/C (red/red) dimers in this structure

These data suggest that RNA mimics that encompass viral CP binding sites could be effective anti-viral agents because they would either trap CPs in unproductive aggregates or redirect the normal assembly pathway. Viral RNA and its protective CPs have to make very many precise interactions with each other, and if such contacts occur improperly anywhere during the process the formation of infectious progeny virus should be impaired. Current therapeutics do not target viral self-assembly pathways but reagents with such potential could prove to be potent inhibitors of viral assembly pathways.

## Materials and methods

### Proteins, RNAs and reagents

Wild-type bacteriophage MS2, maintained in 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.1 mM MgCl<sub>2</sub> and 0.01 mM EDTA (buffer A), was a gift from Professor David S. Peabody, University of New Mexico (Albuquerque). Wild-type, recombinant MS2 CP was over-expressed in *E. coli* and purified using previously described methods [18,30]. Purified  $T = 3$  capsids were stored in 10 mM HEPES (pH 7.2), 100 mM NaCl and 1 mM EDTA. CP<sub>2</sub> for reassembly reactions was obtained by treating the recombinant capsids with glacial acetic acid to induce capsid dissociation to disassembled CP monomers, followed by exchange into 20 mM acetic acid to promote dimerization [31]. To ensure that complete disassembly had occurred before reassembly, the CP<sub>2</sub> were examined using negative stain EM [30]. Synthetic oligonucleotides were prepared as described previously [19].

The 928 nt sub-genomic fragment was prepared by *in vitro* transcription from a cDNA template corresponding to nucleotides 1419–2346 of the MS2 genome using a MEGAscript T7 kit (Ambion) according to the manufacturer's instructions. The cDNA template was generated by PCR using the forward primer GATAATACGACTCACTATAGGGGTCGCTGAATGGATCAGC and reverse primer CTGTAAACACTCCGTTCCCTACA from the plasmid pSmart\_3528. The plasmid pSmart-3528 is a pSMART (Lucigen Corporation, Middleton, WI, USA) plasmid encompassing cDNA corresponding to nucleotides 19-3550 of the MS2 genomic sequence (AN: V00642) as verified by automated DNA sequencing.

### MS2 capsid reassembly reactions

Reassembly reactions were performed with the short oligonucleotides (S1–S4 and TR) by titrating the RNA with increasing amounts of CP<sub>2</sub> in reassembly buffer A (20 mM Tris-acetate, pH 7.0–7.5), allowing the reactions to proceed for 1 h on ice and then analysing the products on agarose-acrylamide mixed gels (Figure 5). The recombinant  $T = 3$  capsid was used as a marker (lanes 1 and 10). The RNAs (1  $\mu$ M; lanes 2) were titrated with 0.1, 0.2, 0.5, 1, 2, 4 and 5  $\mu$ M CP<sub>2</sub> (lanes 3–9).

Equivalent experiments with the large RNA fragment were carried out by titrating the RNA with increasing amounts of CP<sub>2</sub> in reassembly buffer B (40 mM NH<sub>4</sub>OAc, 1 mM Mg(OAc)<sub>2</sub>, pH = 5.8–6.5). The reassembly reactions were performed at 50 nM RNA with 5–150 molar excess of CP<sub>2</sub> in a 20  $\mu$ l final reaction volume and allowed to proceed for 3 h at 4°C. 5  $\mu$ L of loading buffer (40% sucrose, 0.25% bromophenol blue) was then added to each reaction prior to

---

are shown. Adapted from Ref. [33]. (B) Upper panel: agarose-acrylamide gels of reassembly reactions with short fragments. Samples were – lanes 1 and 10 = recombinant  $T = 3$  shells as a control; lanes 2–9 = RNAs (TR, S1–S4, as indicated) at 1  $\mu$ M with 0, 0.1, 0.2, 0.5, 1, 2, 4 and 5  $\mu$ M CP<sub>2</sub> added. Lower panel: shows an agarose gel of a reassembly reaction with the 928 nt fragment (Figure 1) at differing molar excesses of CP dimers (indicated above each lane). Available in colour online.

agarose electrophoreses in a 0.8% agarose gel at 80 V for 60 min. Agarose gels were stained post electrophoreses in TBE running buffer containing 5 µg/ml EtBr and visualized at 254 nm.

### Acknowledgements

We thank The Wellcome Trust, The Leverhulme Trust and the UK BBSRC for financial support, Professor David Peabody, University of New Mexico for his gift of wild-type MS2 bacteriophage, and Dr Abdul Rashid for oligonucleotide synthesis. OR, KT and SF thank The Wellcome Trust, the University of Leeds Interdisciplinary Institute in Bionanosciences, and the EU Marie Curie scheme for PhD studentship support, respectively. NAR holds a University of Leeds Research Fellowship.

### Note

1. Present address: BMRC, Sheffield Hallam University, Howard Street S1 1WB, Sheffield, UK.

### References

- [1] P. Annamalai and A.L.N. Rao, *Packaging of brome mosaic virus subgenomic RNA is functionally coupled to replication-dependent transcription and translation of coat protein*, *J. Virol.* 80(20) (2006), pp. 10096–10108.
- [2] D. Beckett and O.C. Uhlenbeck, *Ribonucleoprotein complexes of R17 coat protein and a translational operator analog*, *J. Mol. Biol.* 204 (1988), pp. 927–938.
- [3] D. Beckett, H.N. Wu, and O.C. Uhlenbeck, *Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly*, *J. Mol. Biol.* 204(4) (1988), pp. 939–947.
- [4] J. Carey and O.C. Uhlenbeck, *Kinetic and thermodynamic characterization of the R17-coat protein ribonucleic-acid interaction*, *Biochemistry* 22(11) (1983), pp. 2610–2615.
- [5] J. Carey, V. Cameron, P.L. Dehaseth, and O.C. Uhlenbeck, *Sequence-specific interaction of R17-coat protein with its ribonucleic-acid binding-site*, *Biochemistry* 22(11) (1983), pp. 2601–2610.
- [6] D. Caspar and A. Klug, *Physical principles in the construction of regular viruses*, *Cold Spring Harb. Symp. Quant. Biol.* 27 (1962), pp. 1–24.
- [7] M.A. Convery, S. Rowsell, N.J. Stonehouse, A.D. Ellington, I. Hirao, J.B. Murray, D.S. Peabody, S.E.V. Phillips, and P.G. Stockley, *Crystal structure of an RNA aptamer protein complex at 2.8 angstrom resolution*, *Nat. Struct. Biol.* 5(2) (1998), pp. 133–139.
- [8] F.H.C. Crick and J.D. Watson, *Structure of small viruses*, *Nature* 177(4506) (1956), pp. 473–475.
- [9] R. Golmohammadi, K. Valegård, K. Fridborg, and L. Liljas, *The refined structure of bacteriophage MS2 at 2.8 Å resolution*, *J. Mol. Biol.* 234 (1993), pp. 620–639.
- [10] E. Grahm, N.J. Stonehouse, J.B. Murray, S. Van den Worm, K. Valegård, K. Fridborg, P.G. Stockley, and L. Liljas, *Crystallographic studies of RNA hairpins in complexes with recombinant MS2 capsids: Implications for binding requirements*, *RNA-A Publ. RNA Soc.* 5(1) (1999), pp. 131–138.
- [11] H.R. Hill, N.J. Stonehouse, S.A. Fonseca, and P.G. Stockley, *Analysis of phage MS2 coat protein mutants expressed from a reconstituted phagemid reveals that proline 78 is essential for viral infectivity*, *J. Mol. Biol.* 266(1) (1997), pp. 1–7.
- [12] W.T. Horn, K. Tars, E. Grahm, C. Heigstrand, A.J. Baron, H. Lago, C.J. Adams, D.S. Peabody, S.E.V. Phillips, N.J. Stonehouse et al., *Structural basis of RNA binding discrimination between bacteriophages Q beta and MS2*, *Structure* 14(3) (2006), pp. 487–495.
- [13] W.M. Jou, G. Haegeman, M. Ysebaert, and W. Fiers, *Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein*, *Nature* 237 (1972), pp. 82–88.
- [14] H. Lago, A.M. Parrott, T. Moss, N.J. Stonehouse, and P.G. Stockley, *Probing the kinetics of formation of the bacteriophage MS2 translational operator complex: Identification of a protein conformer unable to bind RNA*, *J. Mol. Biol.* 305(5) (2001), pp. 1131–1144.
- [15] S.B. Larson and A. McPherson, *Satellite tobacco mosaic virus RNA: Structure and implications for assembly*, *Curr. Opin. Struct. Biol.* 11 (2001), pp. 59–65.
- [16] C.M. Ling, P.P. Hung, and L.R. Overby, *Independent assembly of Q β and MS2 phages in doubly infected E. coli*, *Virology* 40 (1970), pp. 920–929.
- [17] H.F. Lodish, *Regulation of in vitro protein synthesis by bacteriophage RNA by RNA tertiary structure*, *RNA Phages* (1975), p. 301.

- [18] R.A. Mastico, S.J. Talbot, and P.G. Stockley, *Multiple presentation of foreign peptides on the surface of an RNA-free spherical bacteriophage capsid*, J. Gen. Virol. 74 (1993), pp. 541–548.
- [19] J.B. Murray, A.K. Collier, and J.R. Arnold, *A general purification procedure for chemically synthesized oligoribonucleotides*, Anal. Biochem. 218(1) (1994), pp. 177–184.
- [20] C.Z. Ni, R. Syed, R. Kodandapani, J. Wickersham, D.S. Peabody, and K.R. Ely, *Crystal-structure of the MS2 coat protein dimer – implications for RNA-binding and virus assembly*, Structure 3(3) (1995), pp. 255–263.
- [21] W. Paranchych, *Attachment, ejection and penetration stages of the RNA phage infectious process*, RNA phages (1975), pp. 85–112.
- [22] M.G. Rossmann and J.E. Johnson, *Icosahedral RNA virus structure*, Ann. Rev. Biochem. 58 (1989), pp. 533–573.
- [23] M.G. Rossmann, C. Abadzapatero, J.W. Erickson, and H.S. Savithri, *RNA–protein interactions in some small plant-viruses*, J. Biomol. Struct. Dyn. 1(2) (1983), pp. 565–579.
- [24] S. Rowsell, N.J. Stonehouse, M.A. Convery, C.J. Adams, A.D. Ellington, I. Hirao, D.S. Peabody, P.G. Stockley, and S.E.V. Phillips, *Crystal structures of a series of RNA aptamers complexed to the same protein target*, Nat. Struct. Biol. 5(11) (1998), pp. 970–975.
- [25] A. Schneemann, *The structural and functional role of RNA in icosahedral virus assembly*, Ann. Rev. Microbiol. 60 (2006), pp. 51–67.
- [26] P.K. Sorger, P.G. Stockley, and S.C. Harrison, *Structure and assembly of turnip crinkle virus II: Mechanism of reassembly in vitro*, J. Mol. Biol. 191(4) (1986), pp. 639–658.
- [27] P.G. Stockley, N.J. Stonehouse, and K. Valegard, *Molecular mechanism of Rna phage morphogenesis*, Int. J. Biochem. 26(10–11) (1994), pp. 1249–1260.
- [28] P.G. Stockley, A.E. Ashcroft, S. Francese, G.S. Thompson, N.A. Ranson, A.M. Smith, S.W. Homans, and N.J. Stonehouse, *Dissecting the fine details of assembly of a T = 3 phage capsid*, Comput. Math. Methods Med. 6 (2005), pp. 119–125.
- [29] P.G. Stockley, O. Rolfsson, G.S. Thompson, G. Basnak, S. Francese, N.J. Stonehouse, S.W. Homans, and A.E. Ashcroft, *A simple, RNA-mediated allosteric switch controls the pathway to formation of a T = 3 viral capsid*, J. Mol. Biol. 369(2) (2007), pp. 541–552.
- [30] N.J. Stonehouse and P.G. Stockley, *Effects of amino acid substitution on the thermal stability of MS2 capsids lacking genomic RNA*, FEBS Lett. 334(3) (1993), pp. 355–389.
- [31] T. Sugiyama and D. Nakada, *Control of translation of MS2 RNA cistrons by MS2 coat protein*, Biochemistry 57 (1967), pp. 1744–1750.
- [32] K. Tars, M. Bundule, K. Fridborg, and L. Liljas, *The crystal structure of bacteriophage GA and a comparison of bacteriophages belonging to the major groups of E. coli leviviruses*, J. Mol. Biol. 271(5) (1997), pp. 759–773.
- [33] K. Toropova, G. Basnak, R. Twarock, P.G. Stockley, and N.A. Ranson, *The three-dimensional structure of genomic RNA in bacteriophage MS2: Implications for assembly*, J. Mol. Biol. 375(3) (2008), pp. 824–836.
- [34] K. Valegard, L. Liljas, K. Fridborg, and T. Unge, *The three-dimensional structure of the bacterial virus MS2*, Nature 345 (1990), pp. 36–41.
- [35] K. Valegard, J.B. Murray, P.G. Stockley, N.J. Stonehouse, and L. Liljas, *Crystal structure of a bacteriophage–RNA coat protein–operator complex*, Nature 371(6498) (1994), pp. 623–626.
- [36] K. Valegard, J.B. Murray, N.J. Stonehouse, S. vandenWorm, P.G. Stockley, and L. Liljas, *The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein–RNA interactions*, J. Mol. Biol. 270(5) (1997), pp. 724–738.
- [37] P.A. Venter and A. Schneemann, *Assembly of two independent populations of flock house virus particles with distinct RNA packaging characteristics in the same cell*, J. Virol. 81(2) (2007), pp. 613–619.
- [38] N. Wei and T.J. Morris, *Interactions between viral coat protein and a specific binding region on turnip crinkle virus-RNA*, J. Mol. Biol. 222(3) (1991), pp. 437–443.
- [39] G.W. Witherell, J.M. Gott, and O.C. Uhlenbeck, *Specific interaction between RNA phage coat proteins and RNA*, Prog. Nucleic Acid Res. Mol. Biol. 40 (1991), pp. 185–220.
- [40] K. Wong and W. Paranchych, *Effect of ribonuclease on penetration of R17 phage A-protein and RNA*, Can. J. Microbiol. 22(6) (1976), pp. 826–831.
- [41] A. Zlotnick, R. Aldrich, J.M. Johnson, P. Ceres, and M.J. Young, *Mechanism of capsid assembly for an icosahedral plant virus*, Virology 277 (2000), pp. 450–456.