Evaluating a Single-Stranded RNA Genome as an Anti-Viral Drug Target.

Amy Magoolagan Barker

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

The University of Leeds
School of Molecular and Cellular Biology

January 2018

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

Firstly, I would like to thank Professor Peter Stockley for all his help and support whilst working under his supervision. I cannot emphasise enough how grateful I am to you for turning my incoherent babble into something that I can be proud of! Thank you Peter, and Dr Iain Manfield, for plucking me out of the hyperdispersant industry and into the world of RNA bacteriophages. I have had the opportunity to travel to some amazing places, use some seriously neat kit and work with some fantastic scientists, thank you.

To Amanda, Nikesh and Alex; thank you for keeping me focussed and reminding me why we're so lucky to work in science #postdocgoals. Thank you to Becca for kick-starting the XFP project, I'm really excited to have the opportunity to work with you on this. Simon and Emma, I am thankful for your unfaltering daily encouragement and for putting up with me. Some weeks we spend more time together than I do with my own family! I appreciate your patience and banter.

To my siblings Cara, Dr Lucy, Holly, Sara and Rachel I hope this document is evidence enough to show that I can think about things other than guinea pigs, nail varnish and glühwein. Chaz and Camilla, thank you for mentally supporting me through motherhood, I'm blessed to have you by my side. Thank you to my aunties Tricia, Andrea and Rachael for reminding me that family is everything, I love you guys. To my parents and in-laws, I have to acknowledge you for your stability, reliability and most importantly, childcare. Without you I could not have got this thesis finished. Mum, I appreciate everything you've sacrificed for me, you've made me the woman I am today. Finally, a thank you from the bottom of my heart goes to Matt. For not questioning when I just need to "pop into the lab", giving me an understanding hug when my experiments don't work and helping me with my Luddite PC skills. Your emotional support and unconditional love is what keeps me going. To my daughter, Maddy, I hope you have the opportunity in life to work on something you thoroughly enjoy. Be curious, kind and let the world inspire you.

It sounds like I'm saying goodbye writing these acknowledgements, when actually it feels like just the beginning.

Abstract

In ssRNA viruses, generating progeny can create antigenic drift as a result of protein mutations. This enables viruses to evade vaccine control meaning novel anti-viral targets are always in demand. The discovery of the evolutionarily conserved packaging signal (PS)-mediated assembly mechanism has been proposed as a novel anti-viral target and the work described here is the first to direct drugs towards the RNA stem-loops (SLs) involved. Using *T*=3 bacteriophage MS2 as a model for studying (+)ssRNA viruses, the secondary structure of RNAs derived from genomic RNA (gRNA) were explored as anti-viral targets. Using *in silico* modelling, ligands were identified and tested *in vitro* to verify binding to MS2's highest affinity PS, translational repressor (TR). One ligand, mitoxantrone (MTX), was shown to bind RNA stem-loop "TR" with K_D = 302 ± 22 nM. MTX is known to bind RNA at unpaired adenosine bases, removing the bulged adenosine base from TR increased the K_D to 1.11 ± 0.17 x10^{-6} M. TR-mediated MS2 reassembles in the presence of MTX formed species ≤67 kDa. MTX did not bind to MS2 coat protein dimer, inferring that it exerts its anti-assembly effect(s) via RNA binding. gRNA-mediated reassemblies in the presence of MTX produced mis-assembled capsids. *In vivo* MTX reduced the infectivity of wild-type MS2 in *E.coli*. After three passages, the presence of MTX generated escape mutants. Next Generation Sequencing showed that MS2 generated a unique nucleotide mutation, G393U in the 5' region, encoding for the Maturation Protein (MP), within a SL that also binds the MP. Comparing *in virio* to *in vitro* reassembled particles by X-ray footprinting showed an increase in hydroxyl radical reactivity, implying this region of RNA binds the MP *in virio*. ssRNA bacteriophage use their MP to recognise the host cell and to protect the ends of its viral RNA, an interaction that is necessary for successful infection. Targeting important RNA:protein complexes that underpin infectious virion assembly is a novel route for anti-viral therapy.

Table of Contents

	ACKNOWLEDGEMENTS	iii
	ABSTRACT	iv
	TABLE OF CONTENTS	v
	LIST OF FIGURES	x
	LIST OF TABLES	xvi
	ABBREVIATIONS	xvii
1	**Introduction.**	**1**
	1.1 Structural Classification of Viruses	3
	1.2 Fighting Viral Infections with Vaccines	5
	1.3 Introduction to Anti-Viral Drugs	7
	1.3.1 Targeting Virus Attachment and Cell Entry	7
	1.3.2 Inhibition of Genome Synthesis	9
	1.3.3 Virion Assembly as an Anti-Viral Target	11
	1.3.4 Preventing the Release of Infectious Virions	14
	1.4 Introduction to Virus Structure	15
	1.5 RNA Bacteriophage as a Model Virus	17
	1.5.1 (+)ssRNA Bacteriophage MS2	18
	1.5.2 Evidence for PS Mediated Assembly	23
	1.6 Nucleic Acid Drug Targets	24
	1.7 Methods for Studying RNA:Ligand Interactions	32
	1.7.1 Hydrodynamic Techniques	32
	1.7.2 Spectroscopic Techniques	33
	1.8 Previous Work on MS2 RNA Binding Ligands	35

	1.9 Thesis Objectives	38
2	**Materials and Methods.**	**41**
	2.1 Synthesis and Purification of RNA	43
	2.1.1 Synthesis of MS2 Genome	43
	2.1.2 Producing sub-genomic cDNA fragments of MS2 genome	43
	2.1.3 In Vitro Transcription Reactions	43
	2.1.4 Synthesis of RNA Stem-Loops	43
	2.1.5 RNA Purification by High Performance Liquid Chromatography (HPLC)	46
	2.1.6 Extraction of gRNA from Wild-Type MS2	50
	2.1.7 Reverse Transcription-PCR of 620 Base Region of MS2	50
	2.1.8 Generating Full Length MS2 cDNA for Next-Generation Sequencing	51
	2.2 Gel Electrophoresis Techniques	51
	2.2.1 Denaturing 1% (w/v) Agarose for Gel Purifying RNA	51
	2.2.2 Native 1% (w/v) Agarose Gel for Large Plasmids/DNA	51
	2.2.3 Denaturing 10% (w/v) Polyacrylamide Gel for RNA	51
	2.2.4 Native 15% (w/v) polyacrylamide Gel for DNA/RNA	51
	2.2.5 Denaturing SDS 15 % (w/v) Polyacrylamide Protein Gel	51
	2.3 *In vivo* Methods: Preparing *E. coli* cell cultures and Propagating Bacteriophage	52
	2.3.1 E. Coli. strain C3000 Preparation	52
	2.3.2 Propagation and Purification of Bacteriophages MS2 and GA	52
	2.3.3 Propagation and Purification of Bacteriophage Qβ	53
	2.3.4 Caesium Chloride Purification of Bacteriophages	53
	2.3.5 Bacteriophage Infection Assays	54
	2.4 Plaque Assays: Quantitating Infectivity	54
	2.5 Bacterial Cell Growth Kinetics	56

	2.6 Western Blot to Identify Wild-Type MS2 Coat Protein	56
	2.7 Sedimentation Velocity Analytical Ultracentrifugation of RNAs	57
	2.8 Circular Dichroism Spectroscopy of RNAs	58
	2.9 Absorbance Spectroscopy of RNAs in the presence and absence of MTX	59
	2.10 Mass Spectrometry Techniques	59
	2.10.1 Native, Positive Ion Mode MS for Studying VLP formation	59
	2.10.2 Denaturing, Negative Ion Mode MS for RNA Mass Determination	59
	2.10.3 Ion Mobility MS to study RNA:Ligand Interactions	60
	2.11 cDNA Preparation for NextGen Sequencing	60
	2.12 MiSeq Data Analysis and Instrumentation	61
	2.13 X-ray Footprinting (XRF)	61
3	**Assaying Small Ligands Binding to TR.**	**61**
	3.1 Work Flow for Identifying RNA Binding Ligands *In Silico*	61
	3.1.1 ROCS Ligands Generated from In Silico Search	63
	3.2 Method Optimisation for Studying RNA:Ligand Interactions	67
	3.2.1 RNA:Ligand Binding by svAUC	67
	3.2.2 Ligand Binding Using CD	70
	3.2.3 Fluorescence Spectroscopy of Mitoxantrone in the presence and absence of TR	73
	3.2.4 UV-Visible Absorbance Spectroscopy of Mitoxantrone in the presence and absence of RNAs	74
	3.2.5 Mass Spectrometry of RNAs in the presence and absence of Mitoxantrone and Dimethyl Amorilide	80
	3.3 Discussion	92

4	**Effects of TR binding ligands on MS2 Assembly.**	**93**
	4.1 Ligand Inhibition of TR-mediated Reassembly of MS2 *in vitro*	95
	4.2 Using Native MS to Study the Reassembly of MS2 with TR in the presence and absence of Mitoxantrone	98
	4.3 gRNA-Mediated MS2 Reassembly *in vitro*	102
	4.3.1 Effects of Mitoxantrone on Reassembly of MS2 VLPs with gRNA	104
	4.3.2 Quantitating gRNA-Mediated MS2 Reassembly in the presence and absence of Mitoxantrone	105
	4.4 Discussion	108
5	**Effects of Mitoxantrone on MS2 Infectivity.**	**111**
	5.1 Establishing a Quantitative Assay of MS2 Infectivity	115
	5.2 Quantitating the Effect of Mitoxantrone on MS2 Infectivity	117
	5.3 Testing Anti-Viral Effect of Mitoxantrone on Other Bacteriophages	118
	5.4 Characterising the Effect of Mitoxantrone on MS2 Infectivity	119
	5.5 MS2 Infectivity after pre-incubation with MTX	121
	5.6 Characterisation of Mitoxantrone-Resistant MS2	122
	5.7 Probing gRNA in Virions using XRF	131
	5.8 Discussion	132
6	**Conclusions and Further Work.**	**137**
7	**References.**	**145**
8	**Appendix.**	**165**
	Appendix 8.1 Data analysis of svAUC of TR and Ligands.	168
	Appendix 8.2 Thermal Unfolding of TR Monitored by CD.	169
	Appendix 8.3 Thermal Unfolding of TR Monitored by Absorbance in Buffer A.	169
	Appendix 8.4 *E. coli* C3000 cell growth Monitored by absorbance at 600 and 650 nm.	170

	Appendix 8.5 Quantitating the Effect of Mitoxantrone on MS2 CP Production and gRNA Replication.	171
	Appendix 8.6 TR Ion Mobility MS of (4+) ion A) Raw m/z spectra. B) Arrival time Distribution.	172
	Appendix 8.7 ESI MS of Ligands in Cocktail #1 A) 2-CL-1B-MECA. B) MTX. C) A3. D) Quercin. E) Terazosin.	173
	Appendix 8.8 Details of RNA Binding Ligands.	175
	Appendix 8.9 NextGen Sequencing Results Showing Identified, Aligned Genomes and % Total RNA.	176
	Appendix 8.10 cDNA Sequences of MS2 Samples.	177
	Appendix 8.11 Nucleic Acid Insertion and Deletions Identified in NGS Data.	181
	Appendix 8.12 FASTA Protein Sequences of MS2.	182
	Appendix 8.13 FASTA Protein Sequence of Qβ Replicase. Reading N to C terminal.	184
	Appendix 8.14 Percent Identity Matrix Comparing the Replicase protein sequence in Qβ versus MS2.	184
	Appendix 8.15 Multiple FASTA Replicase Protein Sequence Alignment of Qβ and MS2.	185
	Appendix 8.16 A) Analysing 5'wt MS peaks from Figure 5.17. B) Analysing 5'mutant MS peaks from Figure 5.18.	187

LIST OF FIGURES

1.1	Baltimore's classification of viruses.	4
1.2	Schematic of Group IV (+)ssRNA replication and protein translation.	4
1.3	The electrostatic attachment between virus:cell can be blocked by acting as either a ligand or receptor mimic.	9
1.4	Molecular comparison of Guanosine and Acyclovir.	10
1.5	Anti-viral drugs target different aspects of the virus lifecycle.	11
1.6	Changes in hydrodynamic radius of fluorescently labelled MS2 viral RNA upon addition of CP_2.	12
1.7	Examples of anti-viral drugs that target virus assembly.	14
1.8	Ligands can be used to prevent the release of Influenza A Virus.	15
1.9	Icosahedral axes of symmetry.	16
1.10	Examples of three "complex" virus structures.	17
1.11	MS2 genome schematic detailing the four proteins coded in the RNA genome.	18
1.12	Timing and relative amounts of phage protein synthesised in Group I bacteriophage f2.	19
1.13	Quasi-equivalence in the capsid of bacteriophage MS2.	21
1.14	MS2 RNA stem loops involved in CP_2 binding.	21
1.15	RNA stem loops identified by SELEX that bind to MS2 CP_2.	22
1.16	Cut-away view along 2-fold axis of STMV, BPMV and Pariacotovirus showing distribution of icosahedrally ordered internal RNA.	24
1.17	Watson and Crick base-paired DNA A-T and G-C.	26
1.18	Space-filling models comparing the Structure B-DNA and A-RNA ds helices.	26
1.19	Λ and Δ enantiomers of $[Ru(o\text{-}phen)_3)]^{2+}$.	27

1.20	Amine structures used as molecular building blocks for customised DNA base pair binding.	28
1.21	An example of a customised polyamine DNA-binding molecule.	28
1.22	Four examples of RNA secondary structure.	29
1.23	Crystal structure of a complex between Neomycin B and 16S rRNA.	30
1.24	Ethidium intercalated between two adenosine-uridine base pairs.	31
1.25	Sequence and secondary structure of HIV-1 TAR RNA and its customised binding peptide L-22.	32
1.26	The effect of EtBr on the efficiency of MS2 capsid assembly.	37
1.27	Effects of multivalent cations on the hydrodynamic radius of MS2 5'-RNA sub-genomic fragment.	37
2.1	Plasmid map of MS2-containing vector.	43
2.2	Schematic of solid phase synthesis of RNA oligonucleotides.	46
2.3	Ion exchange HPLC chromatograms of RNA stem loops TR, TR-10A, B3 and TCV.	47
2.4	Ion exchange HPLC chromatograms of RNA stem loops 5'wild-type, 5'mutant, Qβ and GA.	51
2.5	Absorbance trace of CsCl fractionation of wild-type MS2.	53
2.6	Photos of plaque assays.	55
2.7	Schematic representation of boundary formation during an svAUC experiment.	61
3.1	Schematic showing proposed workflow for identification and testing of anti-viral molecules.	62
3.2	Examples of ligands shortlisted by literature review on RNA-binding molecules.	65
3.3	Chemical structures of the top ligands Identified by ROCS *in silico* screening.	68
3.4	svAUC analysis of MS2 RNAs.	71

3.5	svAUC of MS2 TR RNA incubated with shortlisted ligands from ROCS screening.	74
3.6	Raw CD spectra showing thermal unfolding of TR stem loop.	75
3.7	CD Spectra showing raw data of TR with MTX and correcting data for ligand CD.	76
3.8	MTX fluorescent spectra after Ex_{633nm} shows how λ_{max} shifts with MTX concentration.	77
3.9	Absorbance spectra of MTX dilutions.	78
3.10	Absorbance spectra titrating MTX with TR.	79
3.11	Determining the affinity of MTX for TR.	80
3.12	Predicted secondary structure of stem-loops TR and TR-10A.	81
3.13	Comparing the affinity of MTX for TR and TR-10A.	83
3.14	Negative ion mode MS of TR RNA.	83
3.15	Negative ion mode MS of MTX.	84
3.16	Negative ion mode MS of MTX binding to TR, RNA:MTX.	86
3.17	Predicted secondary structure of TR monomer and dimer.	86
3.18	Negative ion mode MS of TR-10A RNA.	87
3.19	Negative ion mode MS of MTX binding to TR-10A.	89
3.20	Assessing RNA:MTX binding interaction. Comparing MS data at 1:1 ratio RNA:MTX for TR and TR-10A.	90
3.21	Quantitatively comparing MS data at 1:1 ratio RNA:MTX for TR and TR-10A.	92
3.22	Negative ion mode MS of DMA binding to TR.	92
3.23	Negative ion mode MS of TR MTX and DMA.	93
3.24	Solution structure by NMR of the tau SRE RNA–MTX complex.	97
4.1	SDS-Page of dissociated recombinant MS2 CP fractions.	98

4.2	Native agarose gel of MS2 resembled with TR in the presence of low molecular weight ligands.	99
4.3	Binding of MTX to TR analysed by native gel electrophoresis and visualised by ethidium bromide staining and fluorescent FAM scanning.	100
4.4	nsTEM of TR-mediated MS2 reassembled VLPs in the presence of low molecular weight ligands.	100
4.5	MS2 reassembly intermediates identified by native MS.	102
4.6	Native MS of TR and CP_2 reassembly reaction.	103
4.7	Native MS of TR and CP_2 reassembly reaction in the presence of MTX.	104
4.8	TR-mediated MS2 reassembly in the presence of MTX analysed by native agarose gel.	105
4.9	54 PS identified by CLIP-Seq (Rolfsson *et al.* 2015).	106
4.10	Agarose gels of MS2 gRNA preparations.	107
4.11	Effect of MTX on the electrophoretic properties of gRNA.	108
4.12	The effect of MTX on gRNA reassembled MS2.	109
4.13	Demonstrating VLP classification by nsTEM.	110
4.14	Effect of MTX on gRNA MS2 VLP formation by nsTEM.	111
4.15	Quantification of VLPs formed from gRNA reassembled MS2 in the presence of MTX.	112
4.16	Schematic demonstrating the effect of MTX on the affinity of gRNA towards MS2 CP_2.	115
5.1	Effects of ligands on *E. coli* cell growth by monitoring absorbance at 650 nm.	116
5.2	Quantitation of *E. coli* C3000 cells grown in increasing concentrations of MTX.	117

5.3	Photos comparing the effects of MTX on MS2 infections by using the plaque assay.	122
5.4	Comparing the effects of MTX on the viral titre of MS2, GA and Qβ.	123
5.5	Western blot of cell lysates showing MS2 CP.	123
5.6	Schematic of MS2 genome detailing 620-bp region selected for RT-PCR.	124
5.7	Denaturing agarose gel of RT-PCR products from extracted RNA from MS2 infection in the presence and absence of MTX.	125
5.8	Comparing titres of phage infected with MS2 pre-incubated with MTX against infections inoculated with MTX 15 minutes post infection.	126
5.9	Comparing MS2 viral titer after passaging samples in the presence and absence MTX.	126
5.10	Purified bands of cDNA generated from extracted RNA from MS2 infections ion the presence and absence of MTX.	130
5.11	MS2 protein amino acid mutations identified by NGS analysis.	131
5.12	Secondary structure change of nucleotides 386-418 after mutation G393U.	133
5.13	Negative ion mode MS of 5'wt RNA binding to MTX.	134
5.14	Negative ion mode MS of 5'mutant RNA binding to MTX.	134
5.15	Hydroxyl radical footprinting of TR.	135
5.16	Hydroxyl radical footprinting reactivities and secondary structures around the putative MP binding site.	137
5.17	Structure of MS2 MP *in situ* as determined by cryo-EM (Dai *et al.* 2016).	138
5.18	Asymmetric structure of MS2 showing MP and interacting RNA stem loops (Koning 2016).	143
6.1	Stem-loop RNA sequences of bacteriophages.	145

6.2	Secondary Structures of identified packaging signals.	146
6.3	Negative ion mode MS of TR in the presence and absence Cocktail #1.	167
8.1	Data analysis of svAUC showing the effect of ligands on TR RNA.	169
8.2	Thermal unfolding of TR RNA monitored by CD.	169
8.3	Thermal unfolding of TR monitored by absorbance.	170
8.4	*E. coli* C3000 cell growth monitored by absorbance at 600 and 650 nm.	171
8.5	Quantitating the effect of MTX on MS2 CP expression and gRNA replication.	172
8.6	TR Ion Mobility MS of (4+) ion.	173
8.7	ESI-MS of Ligands in Cocktail #1.	174
8.8	Details of RNA binding ligands.	175
8.9	NextGen sequencing results showing identified, aligned genomes and % total RNA.	176
8.10	cDNA Sequences of MS2 Samples.	177
8.11	Nucleic acid insertion and deletions identified in NGS data.	181
8.12	FASTA Protein Sequences of MS2.	182
8.13	FASTA Protein Sequence of Qβ Replicase.	184
8.14	Percent Identity Matrix Comparing the Replicase protein sequence in Qβ versus MS2.	184
8.15	Multiple FASTA Replicase Protein Sequence Alignment of Qβ and MS2.	185
8.16	Analysing 5'wt and 5'mutant mass spectras.	187
8.17	MS2 PSs identified by CLIP-Seq (Rolfsson *et al.* 2015).	188
8.18	Publication Record During PT-PhD	189

LIST OF TABLES

1.1	Examples of different cellular molecules that viruses use to attach to and enter a cell.	8
1.2	Secondary structure properties of dsDNA and RNA duplexes.	25
2.1	Primer Sequences used for Generating MS2 DNA.	42
2.2	RNA oligonucleotide sequences and information.	45
2.3	Primer names and sequences used for RT-PCR.	48
3.1	ROCS definitions of chemical functional groups used to facilitate the identification of compounds that are similar to the start library.	63
3.2	Hydrodynamic properties of MS2 RNAs in the presence of cations.	69
3.3	Effects of small ligands on the sedimentation properties of TR.	72
3.6	Comparing (4-) ion intensities of monomeric TR with MTX at different ratios of RNA:MTX.	85
3.7	Comparing (4-) Ion intensities of monomeric TR-10A with MTX at different ratios of RNA:MTX.	88
3.8	Comparing (4-) ion intensities of monomeric TR with DMA at RNA:MTX at different ratios.	91
4.1	Number of full and partial VLPs of gRNA reassembled MS2 in the presence and absence of MTX.	115
5.1	Results of NGS of MS2 inoculum stock, MS2 passaged three times and MS2 passaged three times in the presence of MTX.	129
6.1	Comparing stem-loops identified by Dai *et al.* (2016) Dykeman *et al.* (2013) and Rolfsson *et al.* (2015)	142

ABBREVIATIONS

ACV Acyclovir
AO Acridine Orange
ASP Aspartic Acid
AUC Analytical Ultracentrifugation
CA Capsid Protein
CD Circular Dichroism
CLIP Cross-linking Immunoprecipitation
CMV Cytomeglavirus
CP Coat Protein
CP_2 Coat Protein Dimer
CPK Corey Pauling Koltun Chemical Colouring
CTD C-terminal Domain
dATP Deoxy Ribose Adenosine Triphosphate
ddATP Di-Deoxy Ribose Adenosine Triphosphate
DEPC Diethyl Pyrocarbonate
dGTP Deoxy Ribose Guanosine Triphosphate
DMA Dimethyl Amorilide
DMSO Dimethyl Sulfoxide
DNA Deoxyribonucleic Acid
EDTA Ethylenediaminetetraacetic Acid
EM Electron Microscopy
ESI Electrospray Ionisation
ESI-MS Electrospray Ionisation Mass Spectrometry
FHV Flock House Virus
FMDV Foot and Mouth Disease Virus
FRET Förster Resonance Energy Transfer
GAG Group-specific Antigen
GLU Glutamic Acid
gRNA Genomic RNA
GTP Guanosine Triphophate
HAP Heteroaryldihydropyrimidine
HBV Hepatitis B Virus
HCV Hepatitis C Virus
HIV Human Immunodeficiency Virus
HPV Human Papilloma Virus
HRP Horseradish Peroxidase
HSV Herpes Simplex Virus
ICTV International Committee on Taxonomy of Viruses
IFN Interferon
IPV Inactivated Poliovirus Vaccine
K_D Dissociation Constant
LB Luria-Bertani Media

MALDI Matrix Associated Laser Desorption Ionisation
MCCB Medicinal Chemistry and Chemical Biology
MOI Multiplicity of Infection
MOPS 3-(N-morpholino)propanesulfonic Acid
MP Maturation Protein
MS Mass Spectrometry
MST Microscale Thermophoresis
MTX Mitoxantrone
NCBI National Center for Biotechnology Information
NGS Next Generation Sequencing
NHS National Health Service
NMR Nuclear magnetic Resonance
nsTEM Negative Stain Transmission Electron Microscopy
NTD N-terminal Domain
OH Hydroxyl-group
OPV Oral Poliovirus Vaccine
PaV Pariacotovirus
ParV Parechovirus
PCR Polymerase Chain Reaction
PDB Protein Data Bank
PFU Plaque Forming Units
pI Isoelectric Point
PS Packaging Signal
RdRp RNA Dependent RNA Polymerase
R_h Hydrodynamic Radius
RNA Ribonucleic acid
RT Reverse Transcriptase
SDS Sodium Dodecyl Sulfate
SELEX Systematic Evolution of Ligands by Exponential Enrichment
SL Stem-loop
SMM Small Molecule Microarrays
STNV Satellite Tobacco Necrosis Virus
svAUC Sedimentation Velocity Analytical Ultracentrifugation
TAR Trans-Activation Response
TAT Trans-Activator of Transcription
TBE Tris-Borate EDTA
TMV Tobacco Mosaic Virus
TR Translational Repressor
WHO World Health Organisation
UV-Vis Ultraviolet Visible
VARNA Visualization Applet for RNA secondary structure
VLP Virus-Like Particle
VMD Visual Molecular Dynamics

Chapter 1

1 Introduction

This page has been left intentionally blank.

1.1 Structural Classification of Viruses

Viruses are infectious agents that affect all kingdoms of life. Their negative impacts on global health, food security and their cost to world economy make them an important field of interest in the scientific community. This becomes very apparent when you consider the statistics. An estimated 240 million people are chronically infected with Hepatitis B Virus (HBV) and more than 686,000 people die every year as a consequence due to cirrhosis and liver cancer (WHO 2016). HBV infection can be prevented by vaccination but there is no effective treatment once a chronic infection develops (WHO 2009). It is not only human health that is affected by viruses. Foot and Mouth Disease Virus (FMDV) is a highly contagious disease, infectious to cloven-hooved animals worldwide. The cost of FMDV control alone is estimated at an annual global impact of US$ 6.5 - 21 billion (Knight-Jones and Rushton, 2013). To protect humans, animals and plants from the deleterious effects of viruses we need to understand how they work at a molecular level, starting with their biological structure. Viruses are dynamic assemblies, the simplest exist as a single-stranded ribonucleic acid (ssRNA) genome encapsulated inside a protective protein shell (Caspar and Klug, 1962). Currently the International Committee on Taxonomy of Viruses (ICTV) classifies members within virus families using genomics; the interpretation of evolutionary relationships by analysis of nucleic acid and protein sequence similarities (Adams *et al.* 2015). Previous to this, the first ever attempt at the classification of viruses included information on the virus genome, be it single (ss) or double stranded (ds), positive (+) or negative (-) sense, RNA or DNA, to group viruses into families. This was called "Baltimore Classification" and describes seven groups, shown in Figure 1.1 (Baltimore, 1971). Group I dsDNA viruses typically transcribe early and late messenger RNA (mRNA) at different time points in the virus lifecycle. Early mRNA are transcribed from the dsDNA genome prior to synthesis of viral DNA, late mRNA are transcribed from progeny DNA using DNA-dependent DNA polymerases. These viruses often replicate in the cell nucleus. ssDNA viruses belong in Group II, these usually have circular genomes and replicate via a “rolling

circle" mechanism forming a dsDNA intermediate. Group III viruses have a dsRNA genome, they typically replicate in the nucleus and have segmented genomes encapsulated within a double layer protein shell. Viral RNA-dependent RNA polymerase (RdRp) is carried by the virion and synthesises mRNA from each of the dsRNAs, each encoding for one protein.

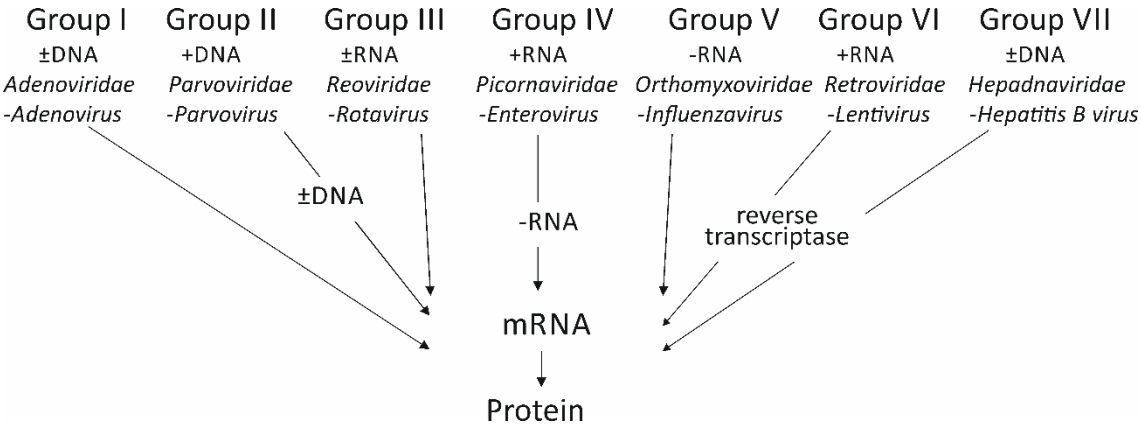


Figure 1.1 Baltimore Classification of viruses is governed by the type of genetic material contained within its protective protein shell. Groups I to VII are shown above with family and genus examples.

Group IV (+)ssRNA viruses can be directly translated, demonstrated by Figure 1.2. Their genomes usually encode a replicase which can produce both (+) and (-) ssRNA.

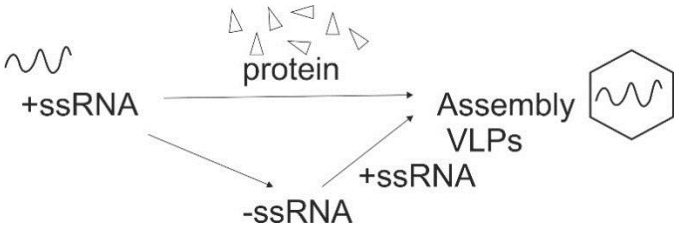


Figure 1.2 Group IV (+)ssRNA can be directly translated into protein. (-)ssRNA is made purely as a template to generate more (+)ssRNA. Examples of (+)ssRNA families include picorna, flavi and caliciviridae.

In Group V, (-)ssRNA requires viral RdRp to bind the encapsidated genome and then sequentially transcribe each gene; this can be carried out within the cytoplasm or nucleus. Group VI (+)ssRNA uses reverse transcriptase (RT) to make (+)ssRNA into a ssDNA intermediate, which is then spliced into the host genome using an enzyme called Integrase. Once integrated, the virus can persist in the host cell and serve as a template for the transcription and replication of the viral

genome, leading to the production of new viruses. Finally, Group VII dsDNA viruses replicate via a ssRNA intermediate (Oxford *et al.* 2016). Viral mRNA and sub-genomic RNA serve as templates for RT to synthesise more DNA. One of the main differences between DNA and RNA viruses is the accuracy of the enzymes involved in producing base-paired complementary strands. Polymerases involved in DNA synthesis can "proof-read". When an incorrect base is added, i.e. it does not conform to Watson-Crick base pairing, steric hindrance stalls the synthesis which allows for the 3' end of the nascent strand to shift towards the active site for the enzyme exonuclease (Bebenek and Kunkel, 2004). Here, 1-3 mismatched bases are removed and synthesis is resumed back in the polymerase active site. RNA replicases have a lower fidelity and make approximately 1 base error per 10^3-10^6 bases compared to DNA error rate of 1 base per 10^7-10^9 bases (Steinhauer *et al.* 1992). This mutation rate explains how RNA viruses evolve so readily (Holland et al. 1982). Within a virus population there are many sub-species, known as quasi-species, these changes in amino acid sequences are the driving force behind viral antigenic drift (Sanjuán and Domingo-Calap, 2016).

1.2 Fighting Viral Infections with Vaccines

There are two different approaches to fighting viral infections. The first to be discovered was vaccination. Smallpox virus has infected humans for thousands of years, Pharaoh Ramses V was thought to have died from an acute infection 1157 B.C., as characterised by small skin lesions (Behbehani, 1983). Chinese medicine dating back to the Sung Dynasty (960 to 1280 A.D.) used a technique described as "variolation" to protect against the virus. This involved inhaling dried pus obtained from lesions from Smallpox virus sufferers (Dixon, 1962). The first well-documented vaccine developed was against Smallpox virus (Jenner, 1801). Jenner noted that dairy maids who had suffered Cowpox virus infections appeared protected against subsequent Smallpox virus outbreaks. He noted this protection could be transmitted from person to person by injecting fluid from the pustules of sufferers directly into the skin of patients. Immunity could be certain

and it was much safer than variolation (Jenner, 1798). Today, vaccination is the most common prevention against viral infection using a non-infectious virus strain. These “disabled” viruses enable the human body to recognise foreign biomolecules resulting in the production of neutralising antibodies that motivate the immune system to remove virions and infected cells. The first commercial vaccine developed was against Yellow Fever virus using an attenuated virus strain. Prior to administration the virus was weakened via multiple passages through other hosts (e.g. chicken embryos) (Theiler and Smith, 1937). A second disabling technique was established using formaldehyde, chemically denaturing the RNA genome renders the virion uninfectious. In Poliovirus, Inactivated Poliovirus Vaccine (IPV), generates the same innate immune response as wild-type virus (Salk, 1954). Today, an attenuated Poliovirus vaccine is administered orally worldwide (Oral Poliovirus Vaccine (OPV)) (Sabin, 1957), and both IPV and OPV have been pivotal in the Global Poliovirus Eradication Initiative (WHO 1988). During January-September 2017 there was only 10 recorded cases of wild-type Poliovirus infection worldwide (http://polioeradication.org). Currently, in the U.K. the NHS offers an inoculation program as a first line of defence against viral infections. From the age of 2 months, children are immunized against diphtheria, tetanus, whooping cough (pertussis), Poliovirus and Haemophilus Influenzae type B, Rotavirus and Meningitis B. Vaccines against measles, mumps, rubella and Human papilloma virus (HPV) (girls only) are all given at a later age. Protection against Influenza A is administered to young children and immune-compromised adults annually (NHS 2015). There are disadvantages to vaccination, antigenic drift can cause the protective effect of antibodies induced by one strain to be reduced over a function of time, resulting in individuals being relatively unprotected against the newest strains in circulation. The mass production of live virions prior to denaturation is hazardous and the thermal and genetic instability of attenuated strains are unavoidable (Brandau *et al.* 2003, Agol, 2006). Recently, a novel VLP-based Poliovirus has been successfully cultured in plants and shown promising anti-viral effects (Marsian *et al.* 2017). Using a synthetic production method means the hazards associated with industrial-scale

vaccine production are minimised, a technique already successful with vaccines derived against HPV and HBV (Ma *et al.* 2007, Biemelt *et al.* 2003, Streatfield, 2005).

1.3 Introduction to Anti-Viral Drugs

The second approach to fighting viral infections is using clinically approved therapeutic drug treatments, targeting different aspects of the virus life cycle (Howe-Grant, 1993).

1.3.1 Targeting Virus Attachment and Cell Entry

Viruses need to infect a host to generate their own progeny. Firstly, the virus must attach (adsorb) to its preferred cellular surface receptor. This attachment is non-covalent and driven by charged surface molecules presented by both the virus and host. This can be a broad interaction – as demonstrated by Rabies Virus which can infect the nervous system and salivary glands of a wide range of mammal species (WHO 2013). On the other hand, it can be a highly specific interaction as demonstrated by HIV-1 which has a much narrower receptor range (Clapham and McKnight, 2001). Table 1.1 shows some well-characterised examples of the cellular molecules involved in the entry pathway of human viruses.

An obvious way to prevent viral infection is to target the molecules involved in the attachment of a virus to its receptor. Sterically hindering protein:protein interfaces is an active area of research (Brito and Pinney, 2017). Figure 1.3 A demonstrates how this can occur via ligand or receptor mimic. Ligand mimics are small molecules which bind to a cells exterior. These molecules can be easily administer although “smothering” cell surfaces may cause adverse toxic deffects to the host (Golan *et al.* 2011). Receptor mimics are used to decorate virus surfaces and render them unable to find their cell binding pocket, suppressing or delaying infection. An example of a receptor mimic has been demonstrated targeting HIV-1 envelope glycoprotein (env). Approximately 72 copies of HIV-1 env protein (trimer) decorate each virion, the env monomer consisting of a gp120 "knob" and gp41 transmembrane domain. In order for a

successful infection gp120 must bind to a cell surface glycoprotein, CD4. A synthetic receptor mimic "eCD4-Ig" has been developed which binds to a conserved amino acid region within gp120, efficiently neutralising 100% of a diverse panel of HIV-1, HIV-2 and Simian Immunodeficiency Virus isolates from trials with mice and rhesus macaques, Figure 1.3 B (Gardner *et al.* 2015).

Table 1.1 Examples of different cellular molecules that viruses use to attach to and enter a cell.

Virus	**Cell Surface Receptors**	**Reference**
Influenza A	Sialic acid.	Matrosovich *et al.* 2013.
HIV-1	CD4 and chemokine co-receptor CC-CKR5.	Dragic *et al.* 1996, Ruibal-Ares *et al.* 2004.
Hepatitis C	Low density lipoprotein receptors, EGFR, CD81, SR-B1.	Agnello *et al.* 1999, Miao *et al.* 2017.
Rabies	Acetylcholine receptor, neural cell adhesion molecule, nerve growth factor, gangliosides, phospholipids.	Lafon, 2005.
Rhinovirus	Intracellular adhesion molecule 1.	Greve *et al.* 1989, Bella and Rossmann, 2000.
Hepatitis B	Immunoglobulin A receptor.	Neurath and Strick, 1990, Glebe and Urban, 2007.
Adenovirus Type 2	Integrins $\alpha_v\beta_{3/5}$	Bergelson *et al.* 1997.
Poliovirus	Transmembrane glycoprotein CD155.	Mendelsohn *et al.* 1989.

Enveloped viruses enter cells by membrane fusion and most non-enveloped viruses enter by receptor-mediated endocytosis (Kielian and Jungerwirth, 1990, Fuchs and Blaas, 2012). Once internalised into a cell, the presence of virions causes an innate immune response, releasing Type 1 interferons (IFNs). These secretory proteins raise awareness to nearby cells of an impending viral attack, recruiting natural killer cells and macrophages to clear the infection.

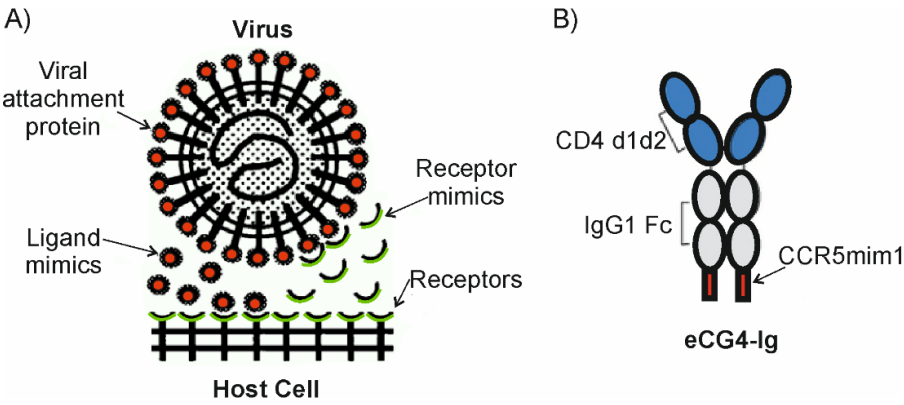


Figure 1.3 The electrostatic attachment between virus:cell can be blocked by acting as either a ligand or receptor mimic. A) Schematic demonstrating virus:cell interaction adapted from Lentz (1990). B) An example of a HIV-1 receptor mimic, sulfopeptide CCR5mim1 (red) binds to the HIV-1 glycoprotein. Adapted from Gardner et al. (2015).

There are over 20 different IFNs, however the presence of just one can prevent a virus from producing and replicating its genomic material. IFN-α can be administered as an anti-viral treatment against HBV and HCV infections (Davis, 1998; Jacobs, 2000; Lui *et al.* 2013).

1.3.2 Inhibition of Genome Synthesis

Cell entry mechanisms involve changes in the cytoplasmic environment close to the cell membrane. These changes, such as pH, cause the viral coat protein shell to degrade – releasing its genomic contents revealing a target for anti-viral drugs. In the 1960s, the first anti-viral drug experiments were against Herpes Simplex Virus (HSV). Growing cultures of cells and infecting them with HSV in the presence of drugs was a slow process. In the 1980s genomic information became available, this greater level of understanding gave lead to better drugs, with fewer off-target side-effects. An example of this is “Acyclovir”, a synthetic guanosine analogue. It acts as a competitive substrate for DNA polymerase and is active against HSV-1 and HSV-2, see Figure 1.4 (Corey *et al.* 1983; Balfour Jr. 1999; Deval, 2009).

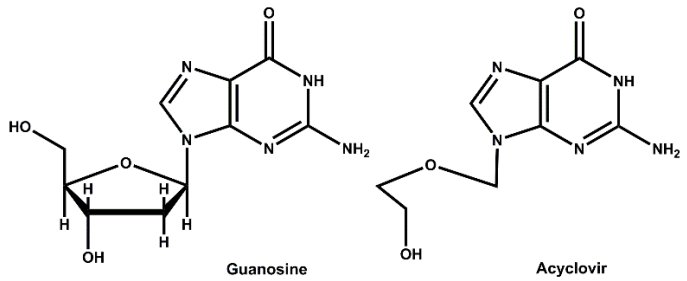


Figure 1.4 A molecular comparison of Guanosine and Acyclovir (ACV). Enzymes in the cell convert ACV to the active form ACV-triphosphate which competes with dGTP as a substrate for viral DNA polymerase, acting as a chain terminator (Deval, 2009).

There are many ways of inhibiting nucleic acid synthesis, mRNA can be targeted directly by using short interfering RNAs (siRNAs) which work by anti-sense inhibition of a selected gene. The 21 nt phosphorothionate "Formivirisen" is complementary to the immediate-early region of mRNA of Cytomegalovirus (CMV), Figure 1.5 A (Highleyman, 1998, Grillone and Lanz, 2001). A common anti-viral agent targeting DNA synthesis is against RNase H, this subunit of reverse transcriptase is used to cleave RNA/DNA substrates generated in the viral lifecycle. Inhibiting the function of RNase H stops degradation of dsRNA:DNA, preventing (-)ssRNA synthesis and protein translation (Champoux and Schultz, 2009). Other drugs can target DNA replication, for example protease inhibitors can bind reversibly to the non-structural proteins involved rendering them unavailable for use. "Telaprevir" is a serine protease inhibitor against chronic HCV, Figure 1.5 B (Perni *et al.* 2006, Wilby *et al.* 2012). Also used against HIV-1, Telaprevir stops the cleavage of capsid polyprotein *gag* into multiple structural proteins that are involved in mature HIV-1 virus assembly (Lv *et al.* 2015).

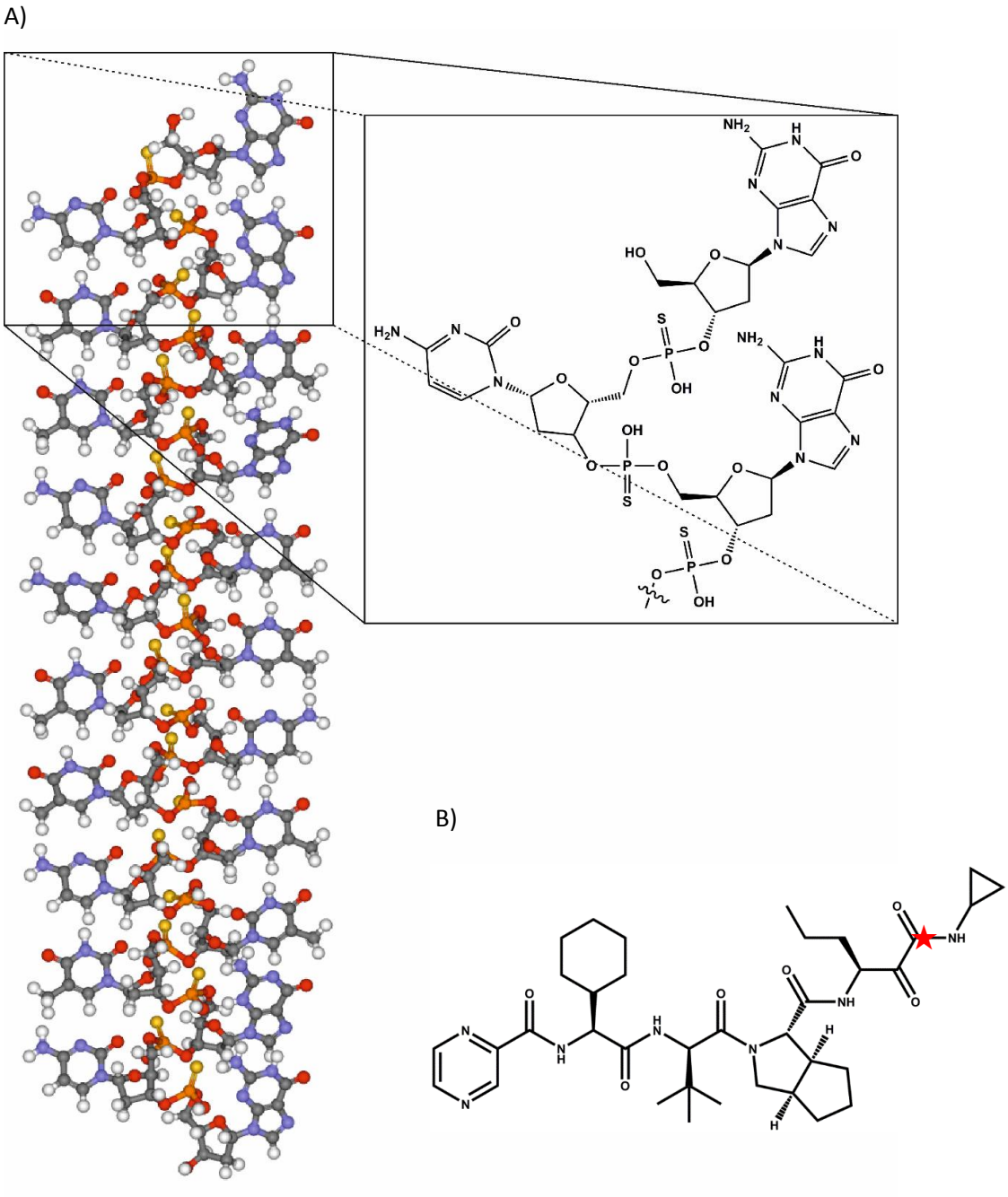


Figure 1.5 Examples of anti-viral drugs that target different aspects of the virus lifecycle. A) Formivirsen is a 21-base synthetic phosphorothionate oligonucleotide complement to a region of mRNA in CMV. B) Telaprevir inhibits chronic HCV replication by binding non-structural proteins involved in the virus life cycle. A covalent but reversible bond is formed between the hydroxyl group of Ser139 HCV protease and the keto-carbonyl group of Telaprevir (★) which prevents the proteolytic cleavage and the formation of mature HCV virions.

1.3.3 Virus Assembly as an Anti-Viral Target

There are many different opinions on how viruses assemble from their constituent components; encouraged by the presence of scaffold proteins, changes in solution conditions (pH, ionic

strength or presence of metal ions) or condensation of CP around a nucleic acid template (Casjens and King, 1974, Adolph and Butler, 1975, Roberts and Steitz, 1967, Schneemann, 2006). Borodavka *et al.* (2013) showed by single-molecule fluorescence that the volume of free viral genomic RNA (gRNA) in solution is larger than the internal dimensions of its protein capsid shell; bacteriophage MS2 gRNA alone measures hydrodynamic radius (R_h) 10-16 nm, whereas the interior of its capsid only measures R_h 10 nm. For virus-like particle (VLP) formation and for a decrease in entropy to occur, Borodavka observed hydrodynamic collapse of gRNA upon addition of coat protein dimer (CP_2) supporting the hypothesis of RNA condensation around CP_2. Figure 1.6 demonstrates the pre-collapse state of gRNA (#) and collapsed gRNA intermediate (★). The increase in R_h for capsid assembly is highlighted in green. The formation of one ordered capsid from multiple free CP_2 subunits is driven by many favourable interactions. These interactions can be strong hydrogen bonds or weak Van der Waals interactions (Perlmutter and Hagan, 2015, van der Schoot and Bruinsma, 2005, Kegel and van der Schoot, 2006). The accumulation of many weak interactions may lead to strong, cooperative bonding (Rupp, 2009). Most viruses can form empty VLPs using multiple weak interactions alone and they can also assemble around any polyanion polymer present (Hu *et al.* 2008, Bancroft *et al.* 1969, Sikkema *et al.* 2007).

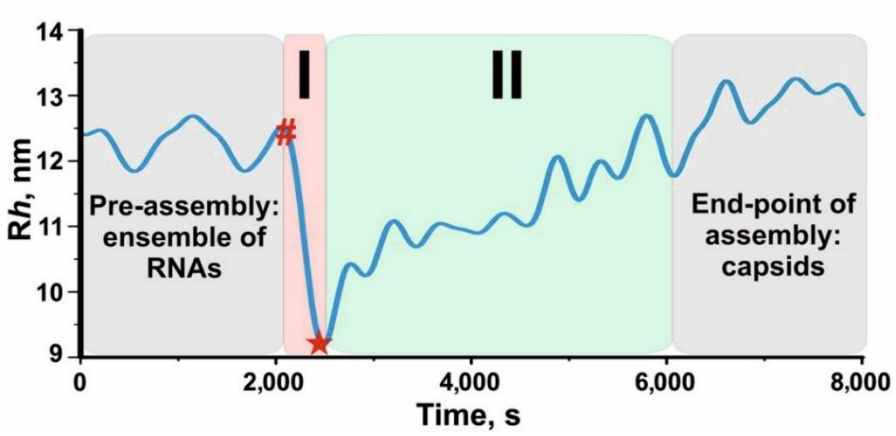


Figure 1.6 Changes in R_h of fluorescently labelled viral RNA upon addition of CP_2. (#) Indicates the pre-collapse state of gRNA and (★) formation of RNA-protein complex (I). The assembly of VLPs is highlighted in green (II). Adapted from Borodavka et al. 2013.

In the *Picornaviridae* (+)ssRNA family, CP subunits exist in solution as multimers known as protein capsomers. From the Enterovirus genus, Poliovirus can be seen existing as capsomers during an infection cycle (Su and Taylor, 1976). In most Enteroviruses, pentamers of the CP (trimeric VP0-VP3-VP1) join to create a 14 S subunit (Rombaut *et al.* 1991), and these 14 S subunits assemble into empty ~80 S particles. The roles of RNA-free VLPs is debated. Are they just failed attempts at assembly or are they a store for capsomers until they are required for virion formation? Work done on Enterovirus-E (formerly Bovine Enterovirus) has shown that empty VLPs are not "dead-end" failure assembly products (unpublished work by Dr. Rebecca Chandler-Bostock, University of Leeds). RNA-free capsids were labelled with ^{35}S-Met/Cys and in the presence of gRNA they were able to assemble into full virus particles. VLP assembly has been explored as an anti-viral target, physically blocking amino acids on the interacting surfaces of protein molecules is a highly-specific way to prevent the formation of higher-order species (Cheng and Brooks, 2015). In HIV-1, *in silico* screening identified a compound that binds electrostatically to an apical site on the N-terminal domain of the HIV-1 capsid protein (CA) (Tang *et al.* 2003). CA protein is a sub-domain of HIV-1 *gag* protein and is present in the formation of immature and mature particles (Sundquist and Kräusslich, 2012). CAP-1 inhibits the CA:CA interactions involved in HIV-1 assembly, Figure 1.7 panel A (Kelly *et al.* 2007). An alternative to blocking protein:protein interactions is to strengthen their association creating unfunctional aggregates. Stray *et al.* (2005) showed in HBV that the presence of heteroaryldihydropyrimidine (HAP) compounds altered the kinetics of assembly, creating structures incompatible with a functional capsid, Figure 1.7 panel B (Bourne et al. 2008, Venkatakrishnan *et al.* 2016).

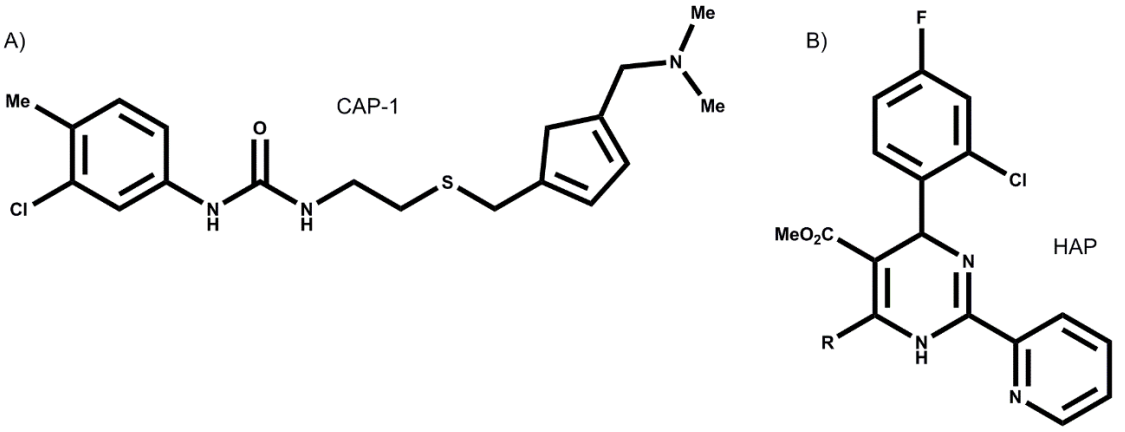


Figure 1.7 Examples of anti-viral drugs that target virus assembly. A) Molecular structure of CAP-1. CAP-1 produces malformed HIV-1 particles of lower particle size. B) Molecular structure of HAP compounds. HAP compounds inhibit allosteric changes between core protein subunits in HBV yielding protein aggregates.

1.3.4 Preventing the Release of Infectious Virions

The last phase of a virus life cycle involves the release of progeny virus from the host cell. The lytic pathway describes how a virus enters a cell, takes over the host's replication mechanism, makes viral DNA, viral proteins and then lyses the cell using enzymes that degrade the inside of the cell wall. In some animal viruses, approximately 100-200 new virions leave the disintegrated host cell to infect other cells (Oxford *et al.* 2016). Some enveloped viruses escape the host cell without total disruption of the membrane, they "bud" out, taking a portion of the membrane with them. A line of defence against virion release in Influenza A and B is to physically block the ion channel membrane protein (M_2) associated with virion release using drugs (Wang *et al.* 1993). Amantadine can block the M_2 ion channel and can treat infections for 2-4 days, however resistance to the drug occurs (Krumbholz *et al.* 2009). Its close relative Rimantadine, also works with a similar mode of action and is preferred as it has less severe side effects (Pielak *et al.* 2011), Figure 1.8. Both are also active in blocking the p7 ion channel involved in the release of HCV (Foster *et al.* 2014). A second line of defence against the release of Influenza A particles is targeting neuraminidase, an enzyme that decorates the outside of Influenza particles and is crucial for releasing viruses from infected cells (Edingert *et al.* 2014). "Oseltamivir" or "Tamiflu"

as it is commercially known, is a small molecule that is well absorbed when administered orally and is 99.6% effective against the 2009 seasonal H1N1 viral strain (Nicholson *et al.* 2000, Ward *et al.* 2005, Center for Disease Control and Prevention H1N1 Recommendations (2010)).

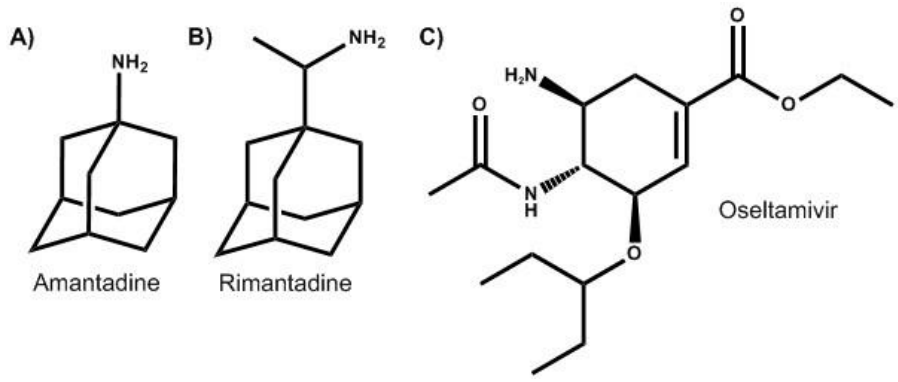


Figure 1.8 Ligands can be used to prevent the release of Influenza A Virus. Molecular structures of A) Amantadine and B) Rimantadine. Both physically block the membrane ion channel used for the release of infectious virus. C) Oseltamivir is used to prevent neuraminidase from releasing Influenza particles from the cell membrane.

Some viruses are lysogenic, and this pathway is common to Group VII *Retroviridae.* They incorporate their genetic material into the genome of the host, where viral nucleic acid becomes part of the host cell chromosome and is replicated with it and transmitted to daughter cells at each subsequent cell division (Oxford *et al.* 2016). By remaining inert (latent), they cause no harm to the host cells and can lie dormant until the lytic stage in which the virus reproduces. Identification of viruses during the lysogenic phase is difficult as they can remain at low, undetectable levels for many years (Arts and Hazuda, 2012). Latent HIV-1 infections can be treated using expensive anti-retroviral therapies, suppressing virus production and increasing patient survival rates. However, interruption in drug delivery schedules can lead to viral rebound (Scheid *et al.* 2016).

1.4 Introduction to Virus Structure

Crick and Watson (1956) proposed that all small viruses, regardless of their physical appearance, were built up of identical protein subunits packed together in a regular manner. Originally

thought of as spherical in shape, it was clear that the rounded structures were faceted – and could be best described as equilateral triangles built up into an icosahedron. This was shown first by X-ray diffraction of Tomato Bushy Stunt Virus (Caspar, 1956). Figure 1.9 demonstrates the symmetry axes of an icosahedron.

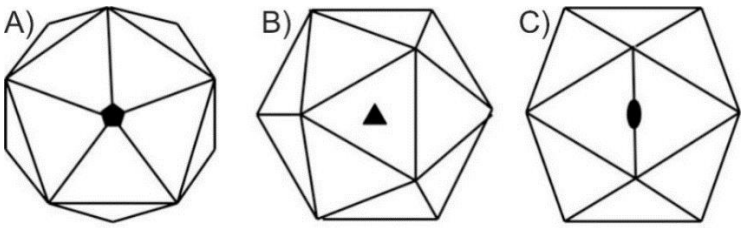


Figure 1.9 Icosahedral axes of symmetry. A) An icosahedron displayed along the 5-fold rotation axis passes through the vertices of the icosahedron. B) 3-fold axis passes through the middle of each triangular face. C) 2-fold axis through the centre of each edge. Figure adapted from Prasad et al. (2012).

It was noticed that some spherical viruses consist of more than 60 subunits (Brenner and Horne, 1959); however, the size of a virus capsid protein is proportional to the size of its genome. Chemically identical protein subunits occupying different spatial arrangements are known as being "quasi-equivalent" (Caspar and Klug, 1962). This allows for larger icosahedral virus structures to form. The relationship between quasi-equivalent conformations within an icosahedral shell is categorised by their Triangulation Number (*T*), which describes the repeating structural subunits. *T* number is specified by a vector (h, k) from an origin point, where f is a positive integer and h and k are values with no common factor, Equation 1.

$$T = (h^2 + hk + k^2)f^2 \qquad \textit{Equation 1}$$

Other virus structures include "helical" and "complex", where "complex" describes any virus that does not fall into either "icosahedral" or "helical" as it may fall into both. See Figure 1.10 for examples. The generation of high-resolution virus structures is crucial in the fight against viral infection (Earl and Subramanian, 2016). For example, there has been limited success in generating neutralising antibodies against HIV-1, which was thought to be due to the structural

flexibility of the glycosylated envelope proteins (Wei *et al.* 2003). Further understanding of the virus' structure, specifically glycoprotein structures, has uncovered a new generation of potent, broadly neutralizing antibodies against HIV-1 (Caskey *et al.* 2015, Ward and Wilson, 2017).

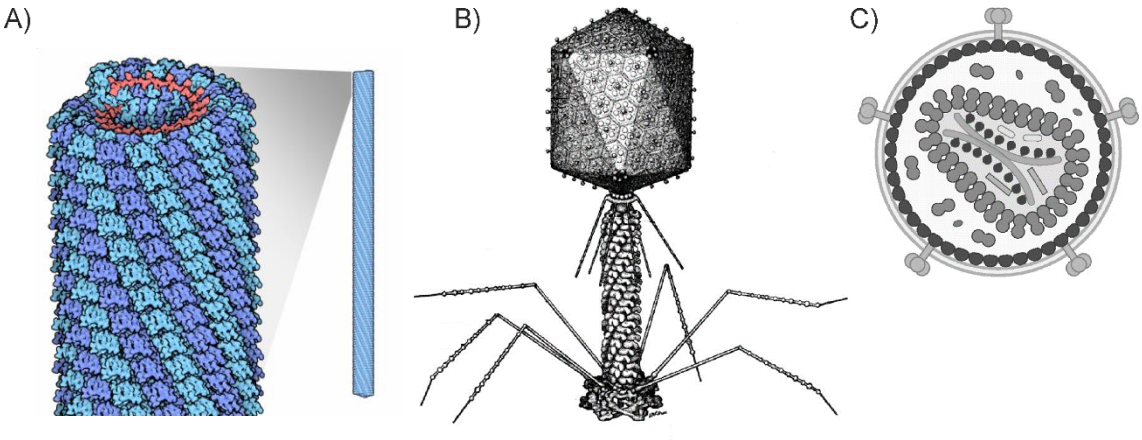


Figure 1.10 Examples of three "complex" virus structures. A) TMV ssRNA (red) coiled inside a protein coat stacking together in a repeating, helical chimney (blue) (Namba et al. 1989). B) Schematic of dsDNA bacteriophage T4. Showing its icosahedral prolate head, helical contractile tail, baseplate and tail fibres (Eiserling and Black, 1994). C) Schematic of the mature HIV-1 particle showing multi-layered complexity. The cone-shaped capsid protein encases 2 copies of the viral RNA genome as a nucleoprotein complex. (Ganser-Pornillos et al. 2008).

1.5 RNA Bacteriophage as a Model Virus

Group IV (+)ssRNA viruses are the largest group in both number of families and number of individual viruses (ICTV, 2016). The largest family within this group is *Picornaviridae*, responsible for infecting humans with *Enteroviruses* (e.g. Poliovirus), *Hepatoviruses* (e.g. Hepatitis A Virus) and *Rhinoviruses* (the common cold), to name a few. As mentioned in Section 1.2, the low fidelity of the RNA polymerase involved in generating progeny virus creates antigenic drift. This enables viruses to evade vaccine control and means new neutralising antibodies, vaccines and anti-viral therapeutics are always in demand. RNA bacteriophages (phage), are commonly used as models for studying (+)ssRNA viruses (Havelaar *et al.* 1993). Phages are viruses that infect prokaryotes, they are extremely common in the environment and it is estimated that there are 10^{10} bacteriophage per litre of sea water (Suttle, 2005). Icosahedral phages are classified into Groups

I – IV based on a number of characteristics including serological cross-reactivity (RNA phages only), molecular weight and sedimentation properties of their genomes. MS2 is the best known representative of Group I and GA, Qβ and SP of groups II, III, and IV respectively. The phages can also be sub-categorised by genome length, the longer length (approx. 600 nucleotides) encoding for an extra "read-through" protein. Collectively Groups I and II are known as Group A, and III and IV as Group B, Group B having the larger genome (Calendar, 1988).

1.5.1 (+)ssRNA Bacteriophage MS2

Bacteriophage MS2 is from the *Leviviridae* family and is composed of a 3569 nucleotide (+)ssRNA genome that encodes four gene products: a maturation protein (MP, also known as the A-protein), CP, lysis protein and a replicase protein, Figure 1.11 (Fiers *et al.* 1976). MS2 infects *E. coli* by adsorption to extracellular F-pili, via a mechanism that is not wholly understood.

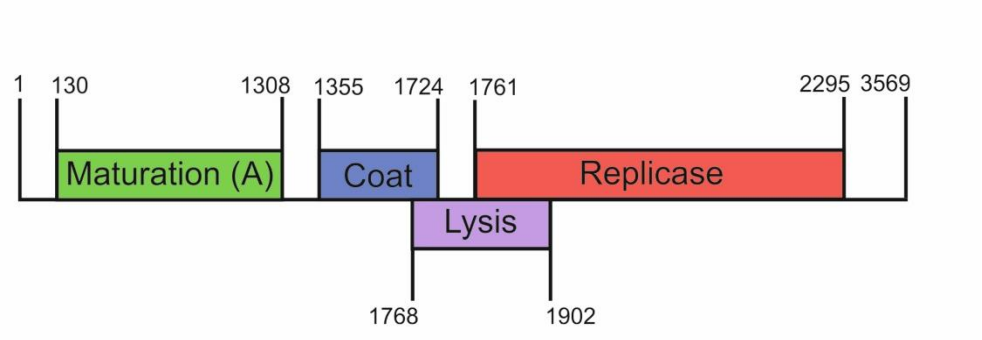


Figure 1.11 MS2 genome schematic detailing the four proteins coded in the RNA genome. The genome also contains non-coding regions at the 5' and 3' ends and a 26 nt intercistronic region between the MP and CP genes.

The first gene product encoded at the 5' end is the MP. There is only one copy per virion and binding to the *E. coli* F-pilus requires that the MP is accessible (Steitz *et al.* 1968). At least part of the MP is localised to the surface of the virion as verified by chemical labelling methods (Curtiss and Krueger, 1974) and also by cryo-EM tomography (Dent *et al.* 2013). Dent *et al.* reported an asymmetric feature common to all MS2 virions showing both the genome and MP interact with the F-pili receptor. Using RNase experiments the MP has been shown to attach

specifically to the 5’ and 3’ end of MS2 gRNA (Fiers *et al.* 1975). These sections of RNA:MP bind at a 1:1 ratio and bind with a higher affinity than other gRNA fragments (Shiba and Suzuki, 1981). The next protein to be encoded in the MS2 genome is CP, which is produced in the highest quantity of all the MS2 proteins, as demonstrated by Figure 1.12 (Beremand and Blumenthal, 1979). Replicase protein translation is dependent on CP translation. This is determined by a nucleotide sequence in the CP gene (1420-1432) which forms a long-distance base-pair interaction with the replicase start site, named the "Min Jou" sequence (Min Jou *et al.* 1972). When ribosomes bind the start of the CP gene to start translation, the “Min Jou” interaction falls apart releasing the replicase start codon for ribosomal binding (Berkhout *et al.* 1985).

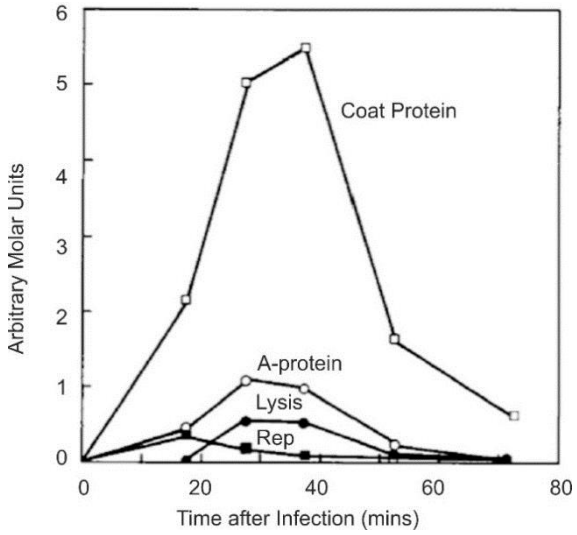


Figure 1.12 Timing and relative amounts of phage protein synthesised in Group I bacteriophage f2. Adapted from (Beremand and Blumenthal, 1979).

Most of MS2's *in vitro* characterisation is inferred from other Group A bacteriophage such as fr and GA, or from experiments with bacteriophage Qβ because their replicase are easily purified (van Duin, 1988). The last protein encoded in the MS2 genome is the lysis protein (L). It can be detected ~20 minutes into the infection, and again this gene is only accessible to ribosomes during translation of CP. Regulation of L protein production is important for ensuring virus is released from infected cells only when maximum progeny virions have been produced. The exact mechanism for L to cause cell lysis is unknown (Goessens *et al.* 1988), recent advances

have suggested the E. *coli* chaperone protein DnaJ (also known as Heat shock protein 40) plays an important role (Chamakura *et al.* 2017). Other related phages from the *Microviridae* and *Alloleviviridae* families insert their lysis proteins into the cell membrane, deactivating the host cell enzymes responsible for cell wall synthesis (Bernhardt *et al.* 2002). Using X-ray crystallography, Valegård *et al.* (1990) showed MS2 has a 28 nm, *T*=3 shell with 180 CP arranging as 90 non-covalent dimers (CP_2) in an icosahedral surface lattice, Figure 1.13 C (Golmohammadi *et al.* 1993). Structural studies by Uhlenbeck and colleagues demonstrated that a sequence-specific interaction between bacteriophage R17 CP (an MS2 variant) and its gRNA is controlled by a single 19-nt stem-loop (SL) within the 3569-nt genome, Figure 1.14 panel A (Carey *et al.* 1983). This SL "TR" encompasses the start codon of the replicase gene and it has nanomolar affinity towards MS2 CP_2 (Lago *et al.* 2001). Using native mass spectrometry, Stockley *et al.* (2007) provided evidence that TR binding to CP_2 generates a quasi-equivalent conformer, switching symmetric CP_2 C/C to A/B, both of which are necessary for building a *T*=3 capsid (Basnak *et al.* 2010), Figure 1.13 panels A and B. This RNA:protein interaction is highly specific and it has been shown that RNA variants of TR can selectively discriminate between CP of MS2 and closely related protein mutants (Fouts *et al.* 1997, Rowsell *et al.* 1998, Hirao *et al.* 1999). Seven phosphate groups on the 5' end of the TR stem interact with the β-sheet surface of a CP_2 (Witherall *et al.* 1991). For binding, a pyrimidine base must occupy position -5 (numbering from replicase start codon) and there must be an unpaired purine base in the stem at position -10 (Wu and Uhlenback, 1987, Helgstrand *et al.* 2002, Stockley *et al.* 2005), Figure 1.14 panel B.

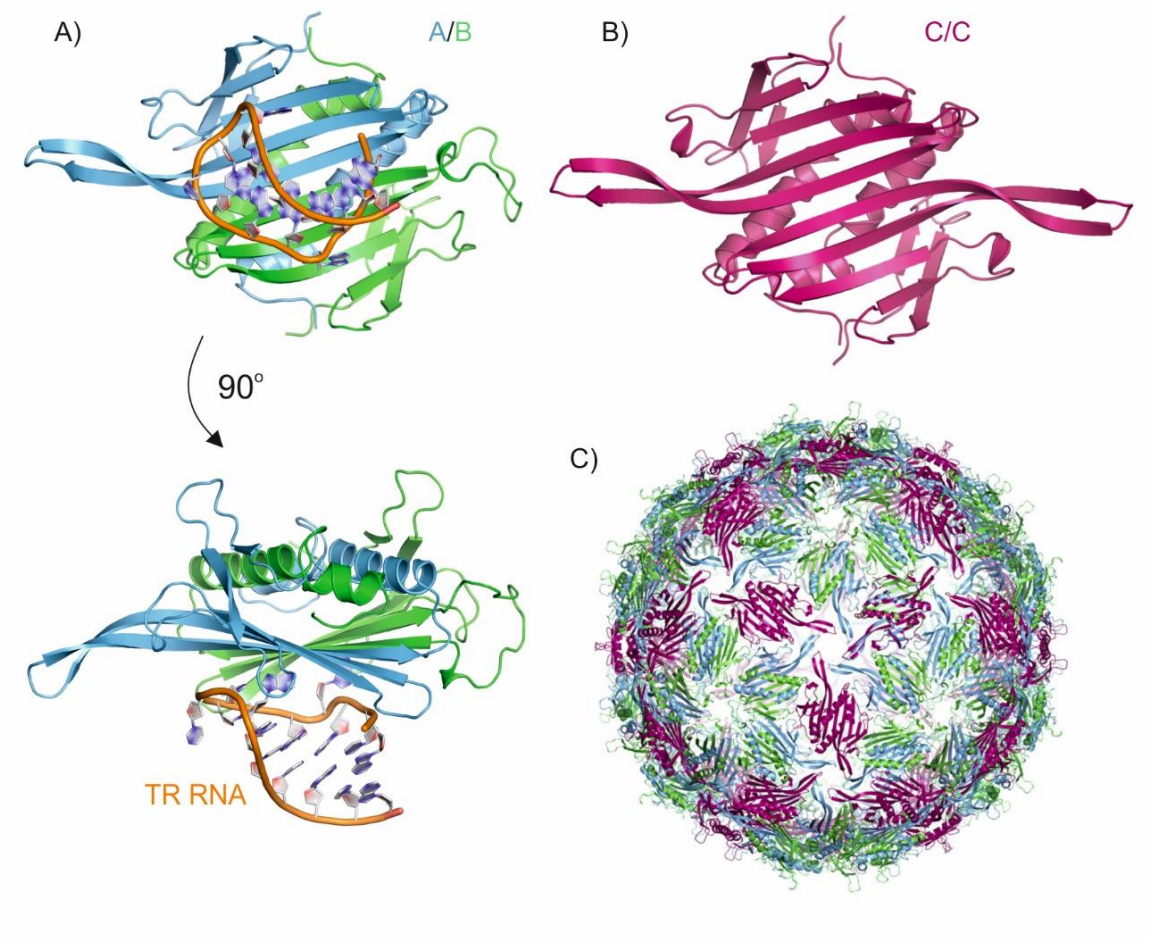


Figure 1.13 Quasi-equivalence in the capsid of bacteriophage MS2. A) Crystallographic structure of the TR oligonucleotide bound to an asymmetric A/B CP dimer, below rotated to 90° to show nucleotides involved in binding, RNA is shown as an orange ribbon (Grahn et al. 1998, PDB 1ZDK). (B) Crystallographic structure of symmetrical C/C CP dimer. C) The X-ray crystal structure of MS2 viewed down the 3-fold axis (Van Den Worm et al. 1997, PDB 1AQ3). Figure adapted from Professor Neil Ranson, University of Leeds and Toropova et al. 2008.

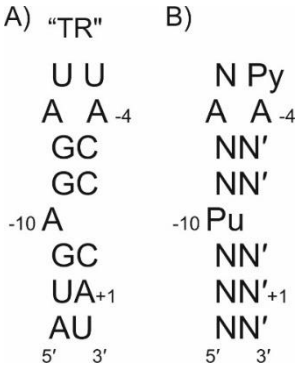


Figure 1.14 MS2 RNA Stem Loops involved in CP_2 binding.
A) Predicted secondary structure of MS2 packaging signal, translational operator "TR", nts 1747-1763.
B) Consensus sequence of RNA oligonucleotides necessary for CP_2 binding (Stockley et al. 1995). Py = pyrimidine, Pu = purine base.

MS2 CP can be over-produced in *E. coli*, existing in solution as *T*=3 VLPs (Valegård *et al.* 1994). When treated with acetic acid these shells dissociate into CP_2 and can be re-assembled *in vitro* with gRNA. Reassembled capsids incorporating one copy of the MP have not been explored due to MP's insoluble nature (Beckett and Uhlenbeck, 1988). Reassembly assays have been characterised with fragmented and full-length gRNA (Rolfsson, PhD Thesis, 2009). MS2 CP_2 can bind and package many different non-cognate RNAs as explored by Stockley (1995), however only cognate RNAs form stable interactions that are capable of assembling into VLPs (Basnak *et al.* 2010). Cooperativity between SLs in MS2 gRNA are important for successful encapsulation and subsequent VLP formation (Rolfsson *et al. 2010,* Dykeman *et al.* 2013). Dykeman *et al.* (2013) determined the sequences and locations of SLs using Hamiltonian paths, a concept from graph theory, in combination with bioinformatics and structural studies. Figure 1.15 A shows "F6" and "F7" RNA sequences determined by systematic evolution of ligands by exponential enrichment (SELEX) (Shtatland *et al.* 2000) and the consensus sequence motifs used in Dykeman's paper, Figure 1.15 B. CP_2-binding RNAs can be described in terms of their relative affinities and distributions within the genome which defines conformations for the packaged gRNA. gRNA SLs involved in assembly have been termed packaging signals (PSs) and have been discovered in other viruses (Stockley *et al.* 2013).

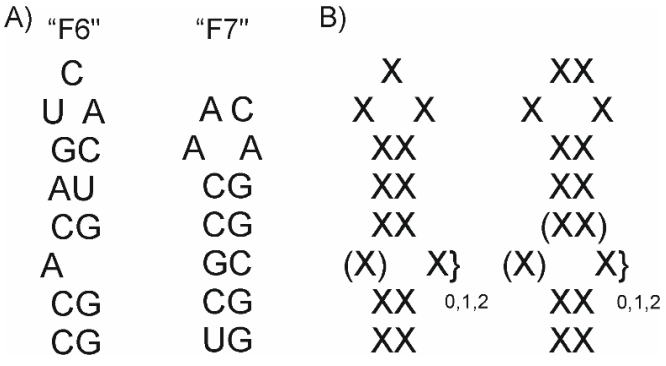


Figure 1.15 RNA SLs that bind to MS2 CP_2. A) SLs F6 and F7 determined by SELEX (Shtatland et al. 2000). B) The search motifs used for determining consensus sequences and positions within the MS2 genome. (X) and (X-X) denote a nucleotide or base pair, respectively, that can be omitted, while $X\}_{0,1,2}$ indicates a location for 0, 1 or 2 additional nucleotides. Adapted from Dykeman et al. (2013).

1.5.2 Evidence for PS-Mediated Assembly

Satellite Tobacco Necrosis Virus (STNV) is an example of a *T*=1 (+)ssRNA virus and with a diameter of only 17 nm it is one of the smallest known viruses. STNV was one of the first virus structures to be solved at atomic resolution (Bancroft et al. 1967). The recombinant expression, reassembly and disassembly conditions of STNV CP were characterised *in vitro* and the high-resolution 1.4 Å structure of the recombinant STNV CP Virus-Like Particle (VLP) was determined by X-ray crystallography (Phillips *et al.* 2012). SELEX of STNV CP isolated RNA fragments that bound with high affinity (Bunka *et al.* 2011). This publication led to further work identifying that assembly initiation is by specific RNA-induced stabilisation of a trimeric CP capsomere (Ford *et al.* 2013). At low CP concentrations, as per *in vivo*, the effects of PS recognition and the cooperativity resulting from their appropriate placement within gRNA is important (Patel *et al.* 2015).

The role of gRNA is to carry genetic information through to the next generation of progeny virus. The viral genome has many roles in a virus' lifecycle that rely on its ability to form multiple secondary structures. (+)ssRNA can exist in a highly extended conformation acting as a template for replication by polymerases and it can function as mRNA for translation by ribosomes. On the other hand, gRNA is highly ordered and compact when visualised in assembled virions (Schneemann, 2006). Nodaviridae, such as Flock House Virus (FHV) and Pariacotovirus (PaV), show dodecahedron cages of density relating to RNA, demonstrated by Figure 1.16 (Fisher *et al.* 1993).

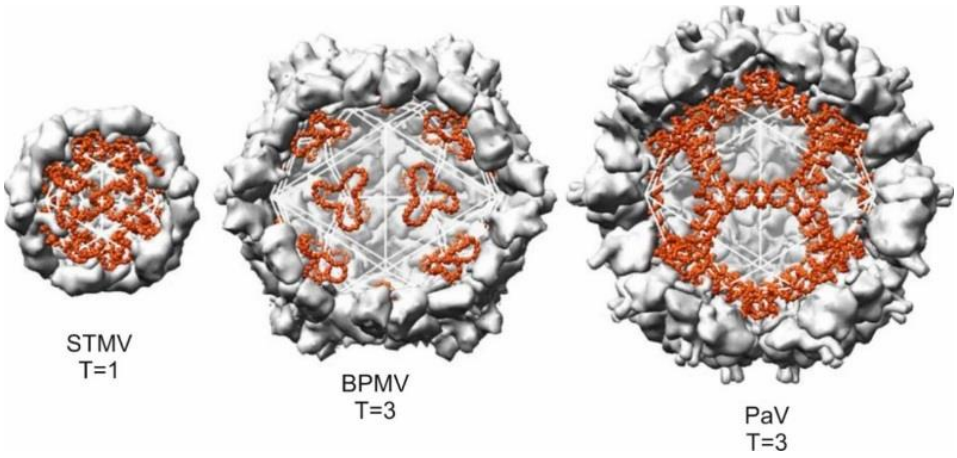


Figure 1.16 Cut-away view along 2 fold axis of Satellite Tobacco Mosaic Virus, Bean Pod Mottle Virus, and Pariacotovirus showing distribution of icosahedrally ordered, internal RNA. The protein capsid is represented in grey at low resolution, whereas RNA is represented in orange as a CPK model of the sugar-phosphate backbone adapted from Schneemann, 2006.

Cross-linking studies of Tobacco Yellow Mosaic Virus, an alphavirus-like (+)ssRNA virus, show extensive contacts between RNA and CP, suggesting that substantial parts of the RNA must be icosahedrally ordered (Böttcher and Crowther, 1996). Seitsonen *et al.* (2010) identified high order CP:RNA densities below the 5-fold vertices in human Parechovirus (ParV), a member of the *Picornaviridae* family. Using a combination of RNA SELEX, bioinformatic analysis and reverse genetics, it has been shown in ParV that RNA segments bind to CP in a sequence-specific manner (Shakeel *et al.* 2015). Disruption of either the RNA CP recognition motif or its contact amino acid residues is deleterious for viral assembly. RNA-binding sites on the CP are evolutionarily conserved across the ParV genus, suggesting that they represent a potential broad-spectrum anti-viral target (Shakeel *et al.* 2016).

1.6 Nucleic Acid Drug Targets

For many years research has been carried out on ligands that bind to DNA and RNA with the notion of finding molecules that can inhibit or control specific biological processes (Eguchi *et al.* 2014, Boutros and Ahringer, 2008). The secondary structures of base-paired DNA and RNA have different spatial conformations due to the 2'OH on the RNA ribose.

DNA nucleotides can form base pair interactions with other DNA nucleotides to form a duplex. Figure 1.17 shows the hydrogen bonding involved in A-T and G-C nucleotide base pairs. dsDNA can exist in three forms A-DNA, B-DNA and Z-DNA (Wilkins *et al.* 1953, Franklin and Gosling, 1953, Pohl and Jovin, 1972). The structure of B-DNA was determined by Wilkins et al. (1953), to be superceded by Franklin's A-form discovery later that year (Franklin and Gosling, 1953). A and B dsDNA are right-handed helices, and the main difference between them is the size of the major and minor grooves. Whilst A-DNA shows equal spacing between strands, B-DNA exhibits larger major and smaller minor grooves, as shown in Figure 1.18. Unlike A and B, Z-DNA is a left-handed helix, which can be induced to form under different physiological conditions, e.g. after addition of polyvalent salts (Behe and Felsenfeld, 1981, Thomas et al. 1991).

Table 1.2 Secondary structure properties of ds DNA and RNA duplexes. A-RNA is closest in structure to B-DNA. Z-DNA helical twist is sequence dependent. Py = pyrimidine Pu = purine. Data from Voet et al. 2013.

	A-DNA	**B-DNA**	**Z-DNA**	**A-RNA**
Base Pair (bp) per turn	11.6	10	12	11
Helical twist (° per bp)	34	36	9 (Py-Pu), 51 (Pu-Py)	30.9
° bp tilt normal to helix axis	20	6	7	16.7

In contrast to ssDNA, ssRNAs typically fold back on themselves to create short Watson–Crick base-paired (bp) SLs. Whilst dsRNA (A-form) closely resembles A-DNA there are still notable differences in the size of the major and minor grooves which explains why there are molecules which uniquely bind to DNA or RNA (Figure 1.18 and Table 1.2).

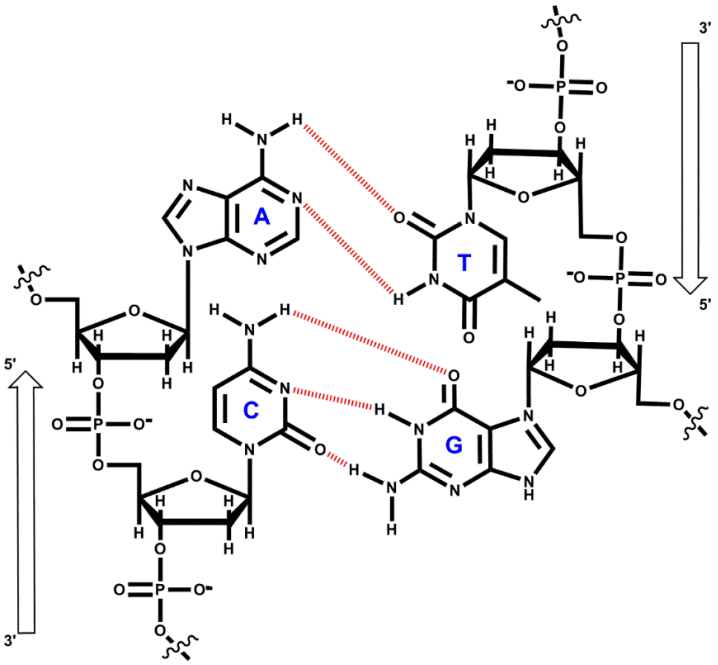


Figure 1.17 Watson and Crick base paired DNA A-T and G-C. Two anti-parallel DNA strands use hydrogen bonding (dashed red) to form Watson and Crick base pairs leading to the formation of a dsDNA right-handed helix (Watson and Crick, 1953).

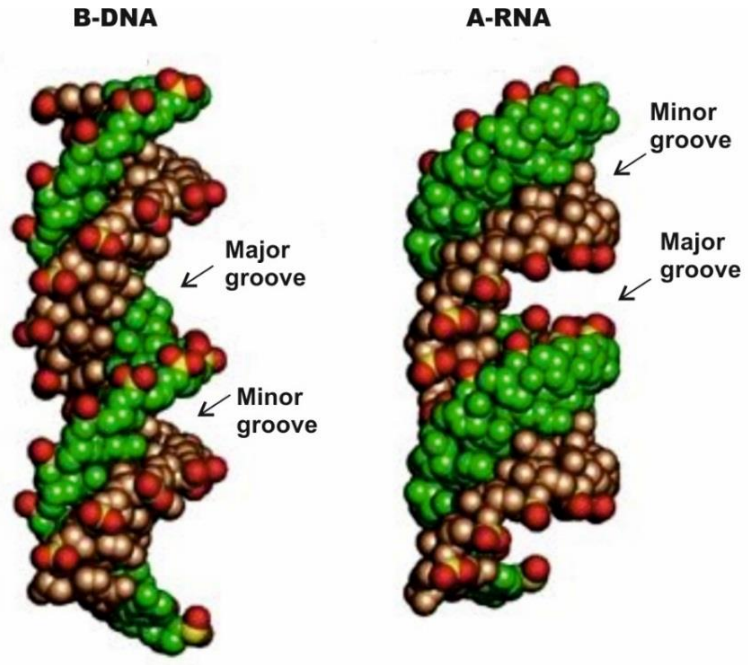


Figure 1.18 Space-filling models comparing the structure of B-DNA and A-RNA ds helices. Colours show each single strand in green and brown, red denotes phosphate-oxygen backbone. The A-form RNA helix is more tightly packed than B-form DNA. The difference in the size of the major and minor grooves allows specific targeting of DNA versus RNA. Figure adapted from Rana, 2007.

dsDNA can be targeted *in vivo* with the anti-cancer compound cis-diammine dichloroplatinum (II) (cisplatin). By forming a coordination bond to DNA (specifically guanosine), cisplatin can

inhibit DNA replication resulting in selective toxicity for rapidly dividing cancer cells (Jamieson and Lippard, 1999). Some ligands can bind DNA through both intercalation and electrostatics. First shown binding to calf thymus dsDNA, tris(phenanthroline)ruthenium (II) ($[Ru(o\text{-}phen)_3)]^{2+}$) exists as two enantiomers in solution, Figure 1.19 (Nordén *et al.* 1976). The enantiomer Δ-Ru binds to DNA with one of its phenanthroline ligands intercalated between base pairs while Λ-Ru does not intercalate but binds electrostatically (Yamagishi, 1984). Each enantiomer exhibits a degree of binding specificity, with Λ preferring A-T and Δ G-C base pairs (Satyanarayana *et al.* 1993).

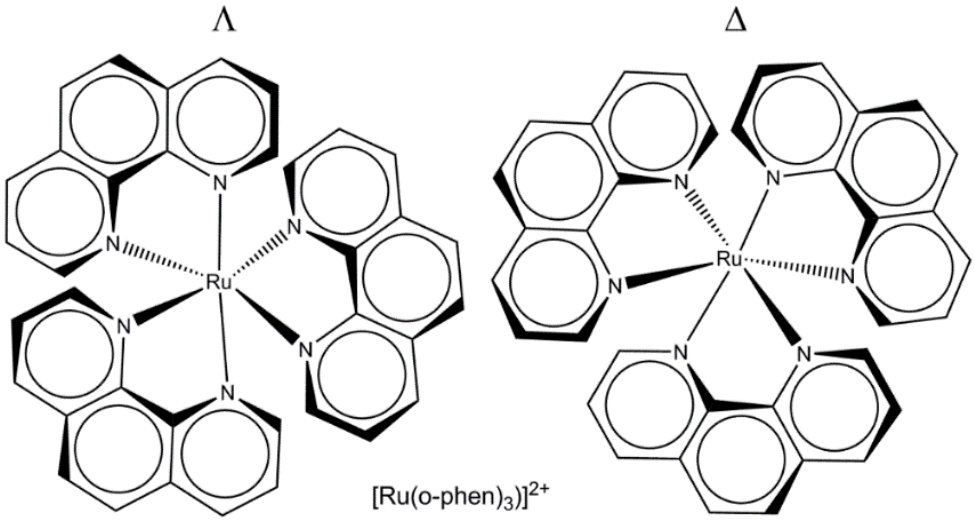


Figure 1.19 Λ and Δ enantiomers of $[Ru(o\text{-}phen)_3)]^{2+}$. Each binds with differing affinity towards dsDNA base pairs (Nordén 1976). Λ enantiomer prefers A-T base pairs and Δ enantiomer prefers G-C.

Dervan *et al.* (2005) demonstrated the idea of customised DNA-binding molecules. Using polyamide building blocks they synthesised short polymers to fit into the minor groove of dsDNA (Figure 1.20). These bind to specific 6–8 bp sequences with high affinity (low nanomolar) and can inhibit transcription by sterically blocking the binding of a broad range of transcription factors, Figure 1.22 (Wade *et al.* 1992, Gottesfeld *et al.* 1997).

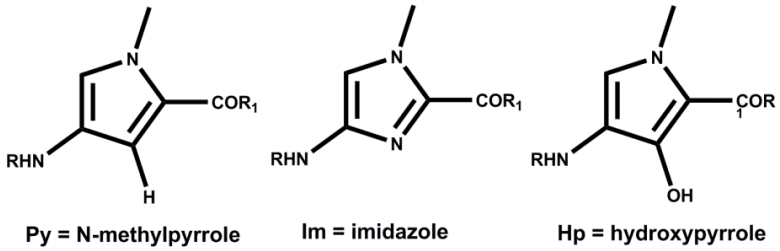


Figure 1.20 Amine structures used as molecular building blocks for customised DNA binding. Five-membered heterocyclic nature of Py, Im and Hp residues. Polymers containing 2 or more residues are synthesised from these starting materials by stepwise addition.

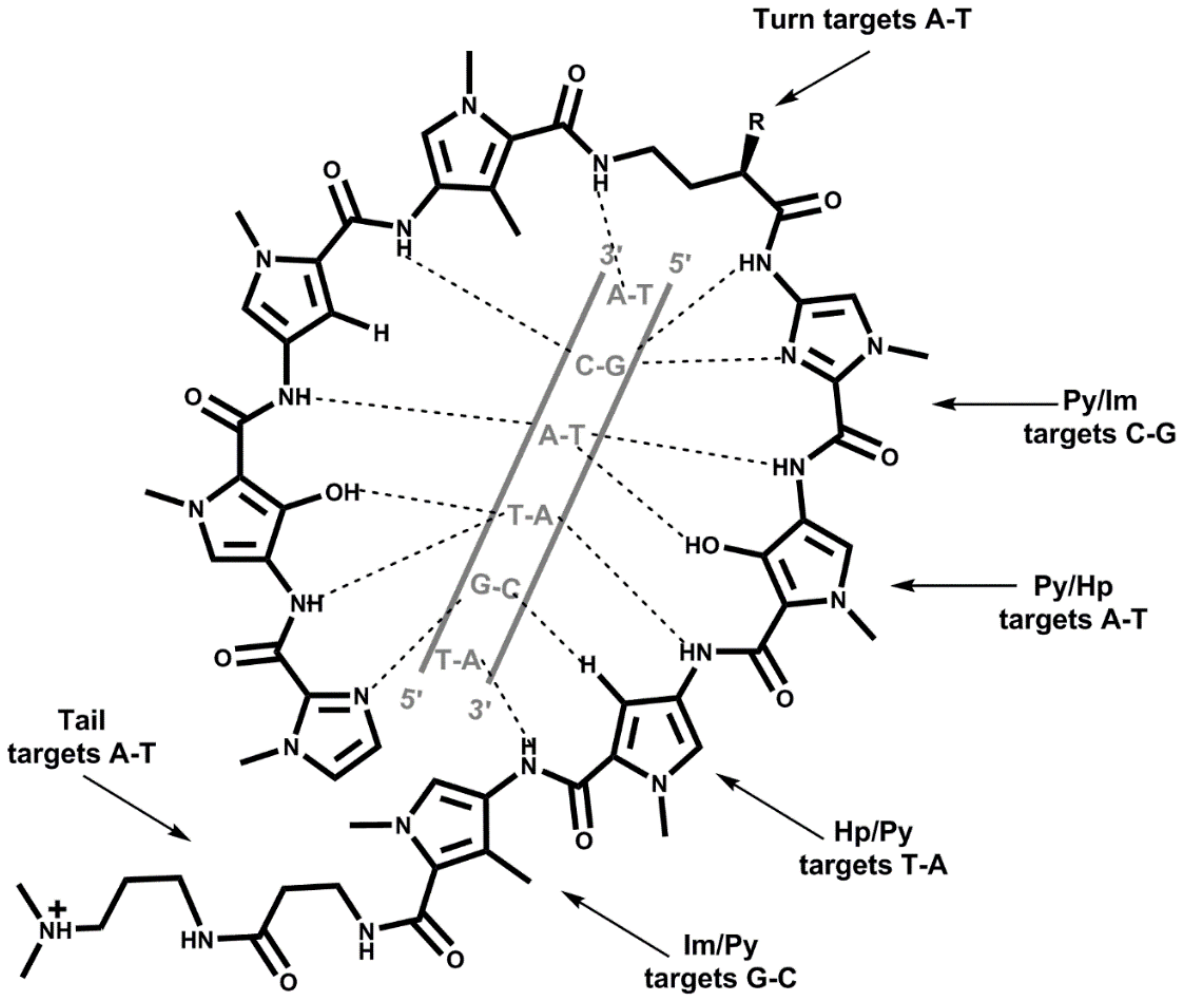


Figure 1.21 An example of a customised polyamine DNA binding molecule. Schematic detailing how a polyamide binds to a short dsDNA duplex (grey). Dashed lines represent hydrogen bonds between polymer subunits and DNA base pairs. Each polyamide strand is usually oriented N→C with respect to the 5'→3' direction of the DNA helix (White 1997, Swalley 1999) Hydrogen bonds shown are not to scale.

As determined by X-ray crystallography and NMR there are four main types of RNA secondary structure: helix (Schaffer *et al.* 2016), bulge (Zhao *et al.* 2005), internal loop (Aboul-ela *et al.* 1996) and hairpin (Codutti *et al.* 2015), Figure 1.22. These contain different local chemical environments suitable for binding of specific or non-specific ligands.

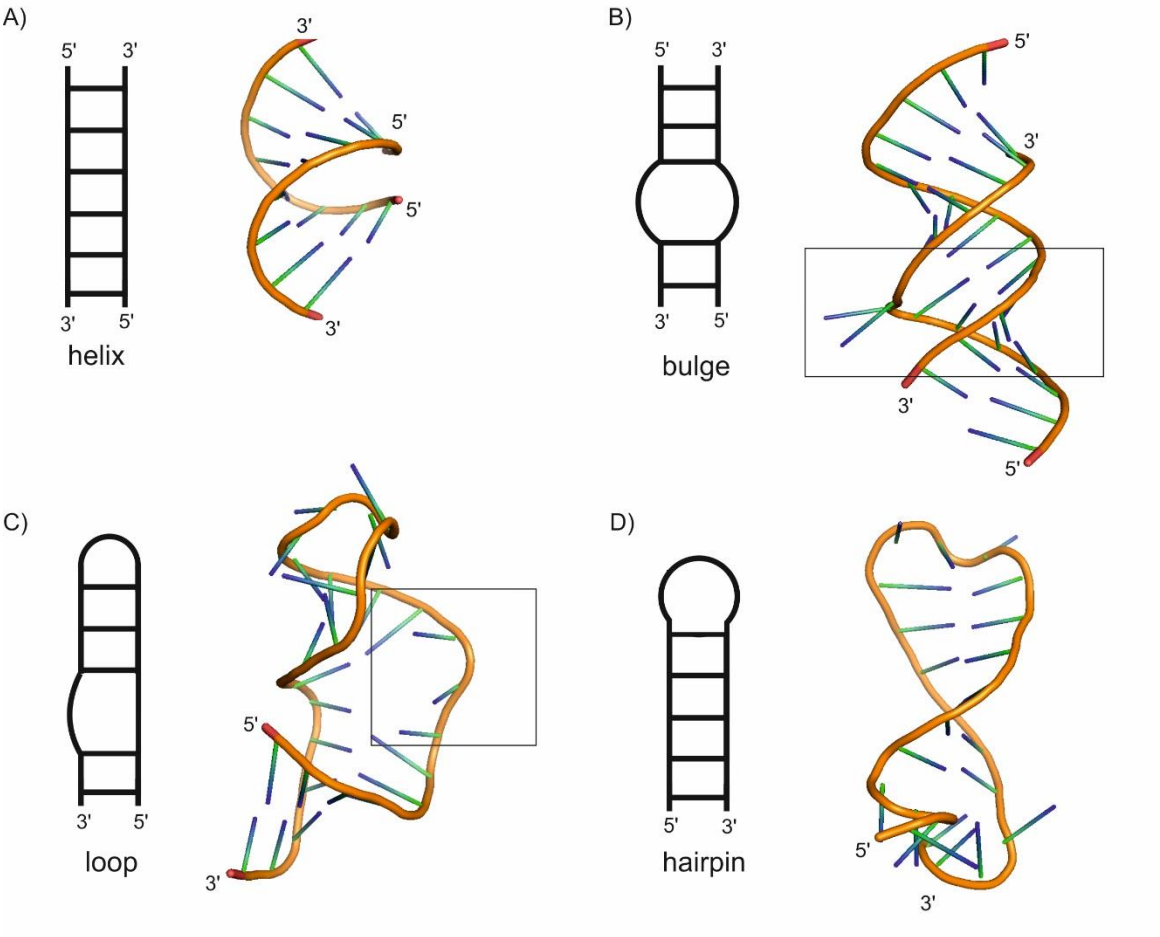


Figure 1.22 Four examples of RNA secondary structure. Orange ribbon represents phosphate backbone and blue/green denotes RNA bases. A) Crystallised RNA helix of 8 bp (Schaffer et al. 2016, PDB 4U3P). B) Bulged SL of rRNA known to bind to aminoglycosides (Zhao et al. 2005, PDB 1ZX7). C) HIV-1 TAR RNA with looped nucleotide in stem (Aboul-ela et al. 1996, PDB 1ANR). D) mRNA hairpin murine tumour necrosis factor α (Codutti et al. 2015, PDB 2N2O). RNA structures drawn in PyMOL (DeLano, 2002).

dsRNA stacked bases can bind intercalating ligands, and the negatively charged phosphate backbone of the duplex can bind molecules electrostatically. Spermidine is one such molecule, present in living cells as a degradation product of the amino acids L-arginine and L-methionine (Tabor and Tabor, 1984). Spermidine has been shown to bind electrostatically to the major groove of dsRNA, using X-ray crystallography (Quigley *et al.* 1978, Frydman *et al.* 1992).

Aminoglycosides are molecules that have antibiotic properties that bind bacterial rRNAs altering the fidelity of translation. Their ability to dock via electrostatics to flexible RNA has been extensively explored (Moitessier *et al.* 2006). Figure 1.23 shows how the dsRNA stem of the bacterial ribosome 16S rRNA (A-site) binds to two Neomycin B molecules. Binding of Neomycin B between the bulged adenosine base and the minor groove of two helices disturbs the fidelity of the tRNA selection step during protein synthesis (Davies *et al.* 1965, Vicens and Westhof, 2003).

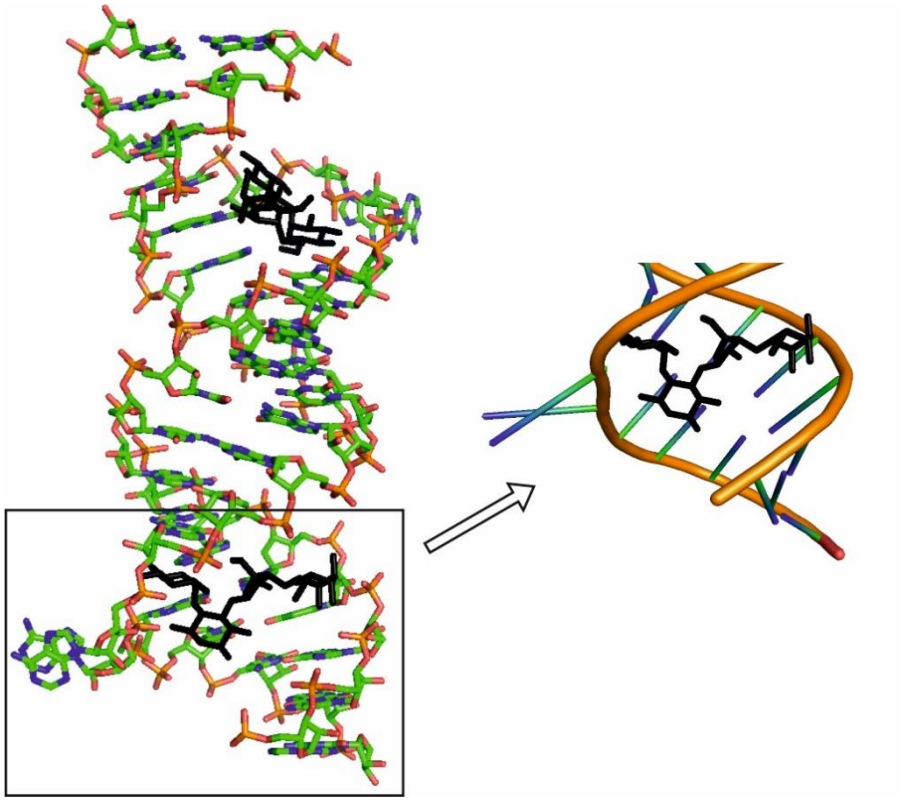


Figure 1.23 Crystal structure of a complex between Neomycin B and 16S rRNA. Two neomycin B molecules binding electrostatically to the A-site on the bacterial ribosome 16S rRNA. Inset shows simplification of the interaction; neomycin B binding between the bulged adenosine base and the minor groove of two helices (François et al. 2005, PDB 2ET4). Drawn in PyMOL (DeLano, 2002).

Both RNA and DNA secondary structures are stabilised by π-stacking interactions. These are exploited during purification or in experimental assays via dye intercalation (Reha *et al.* 2002). Most dye molecules have electron-rich aromatic regions and will readily form π-stacking interactions between base pairs (Reha *et al.* 2002). For example, the partial positive charge on

cyclic ethidium molecules is important for mediating electrostatic attraction and hydrogen bonding interactions with RNA phosphate groups, Figure 1.24 (Waring, 1965, Giacomoni and Le Bret, 1975, Olmsted and Kearns, 1977). Acridine orange (AO), is an organic compound used as a nucleic acid-selective fluorescent cationic dye, similar to EtBr but useful for cell cycle determination as it interacts differently with DNA and RNA. AO intercalates DNA (λ_{max} = 525 nm), whereas AO binds electrostatically to RNA (λ_{max} = >630 nm) making it suitable for detecting different stages of the cell life cycle, including proliferation and apoptosis (Darzynkiewicz, 1990).

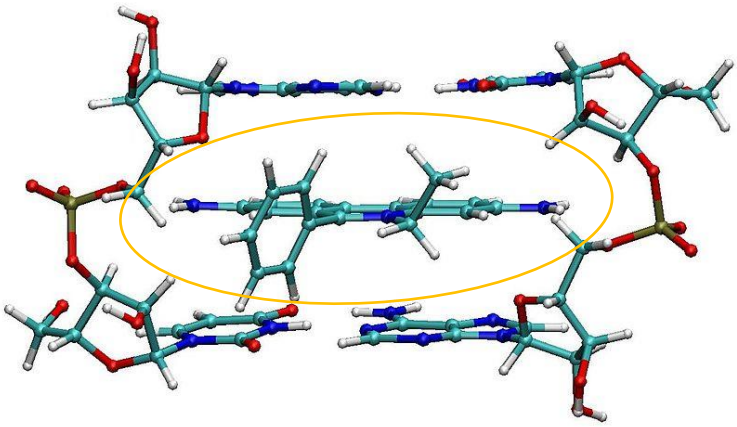


Figure 1.24 Ethidium (circled in yellow) intercalated between two A-U base pairs. Figure adapted from Langner, 2006, using VMD software (Humphrey et al. 1996).

Another intercalating ligand is tetracycline, a widely used therapeutic antibiotic of low toxicity. In the presence of Mg^{2+} it is known to intercalate at the A-site of 30 S rRNA and inhibits protein synthesis by blocking the binding of aminoacylated tRNA (White and Cantor, 1971, Brodersen *et al.* 2000, Ross *et al.* 1998). Viral RNA has been targeted electrostatically using peptide mimetics (Aboul-ela *et al.* 1998). Varani *et al.* (2009) used a customized *de novo* approach against the regulatory HIV-1 "TAR" RNA and its complex with the HIV protein "Tat" and transcription factor cyclin T1/cdk9 kinase (cyclin), Figure 1.25 A (Davidson *et al.* 2011, Bardaro *et al.* 2009). Using specific targeting of the RNA sequence, the peptide ligand blocks the conformational change of TAR that is required to form its complex with Tat, rendering it unavailable to participate in transcription elongation (Faber *et al.* 2000, Murchie *et al.* 2004).

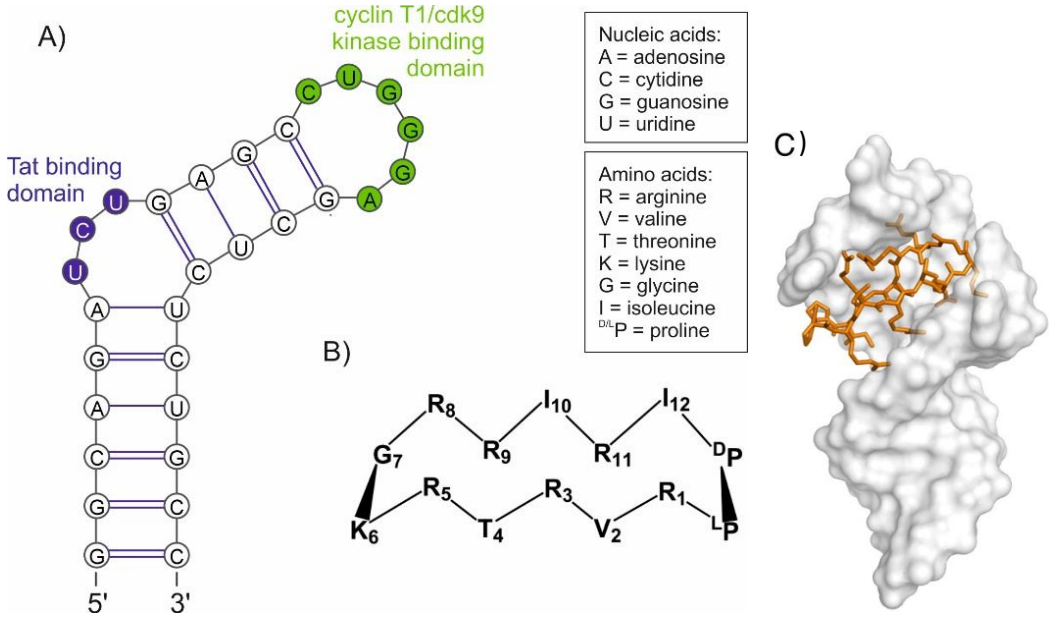


Figure 1.25 Sequence and secondary structure of HIV-1 TAR RNA and its customised binding peptide L-22. A) TAR RNA residues are coloured according to their binding targets, blue = Tat binding domain and green = cyclin T1/cdk9 kinase (Aboul-ela et al. 1998). Image drawn in VARNA (Darty et al. 2009). B) Cyclic peptide "L-22" inhibitor of the HIV-1 "Tat-TAR" transcription complex and C) L-22 (orange) buried deep within the major groove of the loop and bulge residues as resolved by NMR (Davisdon et al. 2009).

1.7 Methods for Studying RNA:Ligand Interactions

1.7.1 Hydrodynamic Techniques

There are many techniques available for studying RNA-ligand interactions *in vitro*. The sedimentation properties of RNA molecules in solution can provide information on their shape and secondary structure. Boedtker (1960) showed with gRNA from Tobacco Mosaic Virus (TMV) that increasing the concentration of Na^+ ions increased the sedimentation coefficient (S value) of RNA by compacting it, compaction as verified by an increase in absorbance. In hydrodynamic properties, reductions in S values suggest a change in conformation. Later work by Boedtker (1968) reported a reduction in S value for RNA upon treatment with formaldehyde, which is known to denature dsRNA helices. The reduction is a consequence of the conversion of dsRNA

to ssRNA in solution. Increases in S-values can indicate an increased compaction, increased symmetry or an increase in RNA mass due to ligand binding and/or RNA multimerisation (Schuck *et al.* 2016).

1.7.2 Spectroscopic Techniques

Circular dichroism (CD) is the difference between left and right-handed optically polarised light and can be used to look at changes in the optical properties of an RNA molecule (Sosnick *et al.* 2000). RNA is composed of D-ribose sugars and dsRNA forms a right-handed duplex, and its chiral CD spectrum is highly informative. Areas of interest in the CD spectrum are assigned to specific RNA secondary structures (Jaeger *et al.* 1990). If ligand binding causes RNA denaturation or unfolding then a change in CD signal is expected. When heat denaturing RNA, a decrease in λ_{max} at 260 and 280nm is observed, relating to destabilisation of A-U and G-C bp respectively. CD has been used to evaluate an mRNA G-quadruplex and its complexes with three alkaloid molecules: nitidine, palmatine and jatrorrizine (Tan *et al.* 2013). G-quadruplexes are secondary RNA structures that arise from the association of four guanines into a cyclic arrangement stabilized by hydrogen bonding (Rhodes and Lipps, 2015). The CD signal decreased with thermal melting of the G-quadruplex, and differences in the change in signal allows calculation of the relative affinities of the different ligands.

RNA molecules can be chemically modified and this is exploited in many of the techniques used to investigate RNA structure and function. The spectroscopic properties of fluorescently labelled oligonucleotides changes when ligand binding changes the RNA conformation and/or molecular dynamics (Perez-Gonzalez *et al.* 2016). Parrott *et al.* (2000) incorporated a fluorescent 2-aminopurine (AP) base into different positions within the MS2 SL TR. Emission spectra show that positioning of the AP base within the oligonucleotide sequence reflects local conformations, as expected. By measuring fluorescence intensity, they showed upon addition of CP_2 fluorescence was enhanced, unaffected or decreased depending on the site of substitution. This technique

can also determine the degree of RNA unfolding as shown with HIV-1 internal ribosome entry site (Carnivali *et al.* 2010). By dual-labelling oligos with a pair of Förster resonance energy transfer (FRET) dyes, the distance between donor and acceptor can be measured and it provides a measure of ligand-induced conformational changes in an RNA target. Qualitative information on the interaction of heteroaromatic compounds with nucleic acids can be obtained from UV−vis or fluorescence spectroscopy (Blakeley *et al.* 2012). Yang *et al.* (2016) recently investigated RNA thermal stability by monitoring FRET efficiency. Artigas *et al.* (2015) measured ligand fluorescence emission spectra in the absence and presence of increasing amounts of HIV-1 transactivation response element RNA (Tau). Hypochromic and bathochromic effects were observed in the fluorescence spectra of Ametantrone-containing ligands during the titration with RNA, indicative of binding. Microscale thermophoresis (MST) can be used to follow the movement of dye-labelled molecules across temperature gradients (Weinken *et al.* 2010). Their movement within the temperature gradient is dependent on size, charge, and hydration shell of the RNA molecule. When a binding ligand is introduced one or more of those properties changes, allowing for calculation of reaction affinities. Entzian *et al.* (2016) showed ATP binding to its RNA aptamer DH25.42. By generating sequence mutants they were able to map ATP's exact binding site.

Non-covalent interactions can be detected by native mass spectrometry (MS), removing the need for RNA modification (Thomas *et al.* 2006, Ganisl *et al.* 2012). Drug-DNA complexes have been studied using native MS, the technique is able to differentiate between intercalation and electrostatic associations (Rosu *et al.* 2006, Rosu *et al.* 2008). There are two main types of ionisation associated with MS analysis of RNA molecules; matrix associated laser desorption ionisation (MALDI) and electrospray ionisation MS (ESI-MS, also known as native MS) (Karas and Hillenkamp, 1988, Fenn *et al.* 1989). By dissolving RNA in organic solvents and applying an electric current, deprotonated RNA ions are formed and enter the mass spectrometer (Thomas *et al.* 2006). This generates $(M - nH)^{n-}$, where M is the oligonucleotide analyte and n is the charge

state formed by n protons being lost (Banerjee *et al.* 2012). Deprotonation of RNAs in ESI MS, occurs at the phosphate backbone. MS has been used to establish the composition and stoichiometry of RNA:ligand interactions from their observed molecular masses (Fabris *et al.* 2010, Thomas *et al.* 2006). Griffey *et al.* (1999) developed novel MS techniques to characterise the binding of aminoglycosides to prokaryotic rRNA A-site subdomain. They further developed MS as a tool to create site-specific cleavage of RNA molecules, allowing the specific nucleotides involved in aminoglycoside binding to be identified.

NMR spectroscopy can be an attractive technique to detect RNA–ligand interactions, even if the K_D is in the millimolar range (Latham *et al.* 2009, Palmer, 2014). NMR experiments developed for screening RNA binders rely on the detection of changes in RNA observable by NMR such as chemical shift, nuclear Overhauser effect and relaxation rates (Furtig *et al.* 2003). Buck *et al.* (2007) used NMR to observe a 3-stage RNA conformational folding. Using the abundance of natural isotopes in RNA and exhaustively exchanging into heavy water (D_2O), allows determination of hydrogen bonding and adjacent base pair interactions. High resolution RNA structures were determined for the first time using solid state NMR by Marchanka *et al.* (2015). The technique is usually dominated by broad signals and resonance overlaps but by using a series of experiments with segmentally labelled RNA the structure was determined for a ribonucleoprotein complex; a 23mer RNA with ribosomal protein "L7ae" from *Pyrococcus furiosus*. ^{15}N-labelling of RNA can be used to define the binding of ligands on the RNA signals by analysing the chemical shift variation of ^{15}N-labelled imino groups.

1.8 Previous Work on MS2 RNA Binding Ligands

Our research group has studied the interaction of MS2 gRNA with the intercalating ligand EtBr. Using sedimentation velocity AUC (svAUC), Rolfsson *et al.* (2010) monitored unlabelled RNA *in vitro*. Using this technique the hydrodynamic properties of a 2469 nt 5'-gRNA fragment of MS2

and its reassembled virus-like particles (VLPs) were compared in the presence and absence of EtBr. In the presence of EtBr the RNA sedimentation coefficient was reduced from $S_{20,w}$ 20.5 to 19.2, suggesting unfolding of the RNA (Rolfsson, PhD Thesis 2009). $S_{20,w}$ is calculated from the experimental sedimentation coefficient, extrapolating values to report at 20°C, in water. This allows for comparison of molecules in different buffers. This reduction in RNA flexibility has an effect on the ability of gRNA to bind MS2 CP dimers, which reduces the amount of *T*=3 VLPs, Figure 1.26. As previously mentioned, the hydrodynamic collapse of gRNA is essential for successful encapsidation during VLP formation (Borodavka *et al.* 2013). Using single molecule fluorescence correlation spectroscopy and RNA uniquely labelled with a fluorescent dye it is possible to monitor its apparent hydrodynamic radius (Maiti *et al.* 1997, Eigen, 1994). Monitoring fluorescent fluctuations in a set confocal volume (~fL) and measuring how long a molecule takes to transverse the set volume, gives characteristic diffusion times (τ_d) obtained from fitting the correlation function to a model (Gell *et al.* 2006, Rigler and Elson, 2001). The average τ_d is used to calculate the molecule's hydrodynamic radius (Zander *et al.* 2003). Smaller molecules will spend a shorter time in the confocal volume, e.g. τ_d ~80 µs for free Alexa Fluor® 488 dye versus e.g. τ_d ~0.4–1 ms for gRNA (Borodavka, PhD Thesis 2013). Using a fragment of bacteriophage MS2 gRNA, Borodavka showed that addition of cationic ligands decreased the hydrodynamic radius of gRNA in aqueous solution, Figure 1.27. Tailoring ligands to bind to RNA may inhibit hydrodynamic collapse and the formation of VLPs.

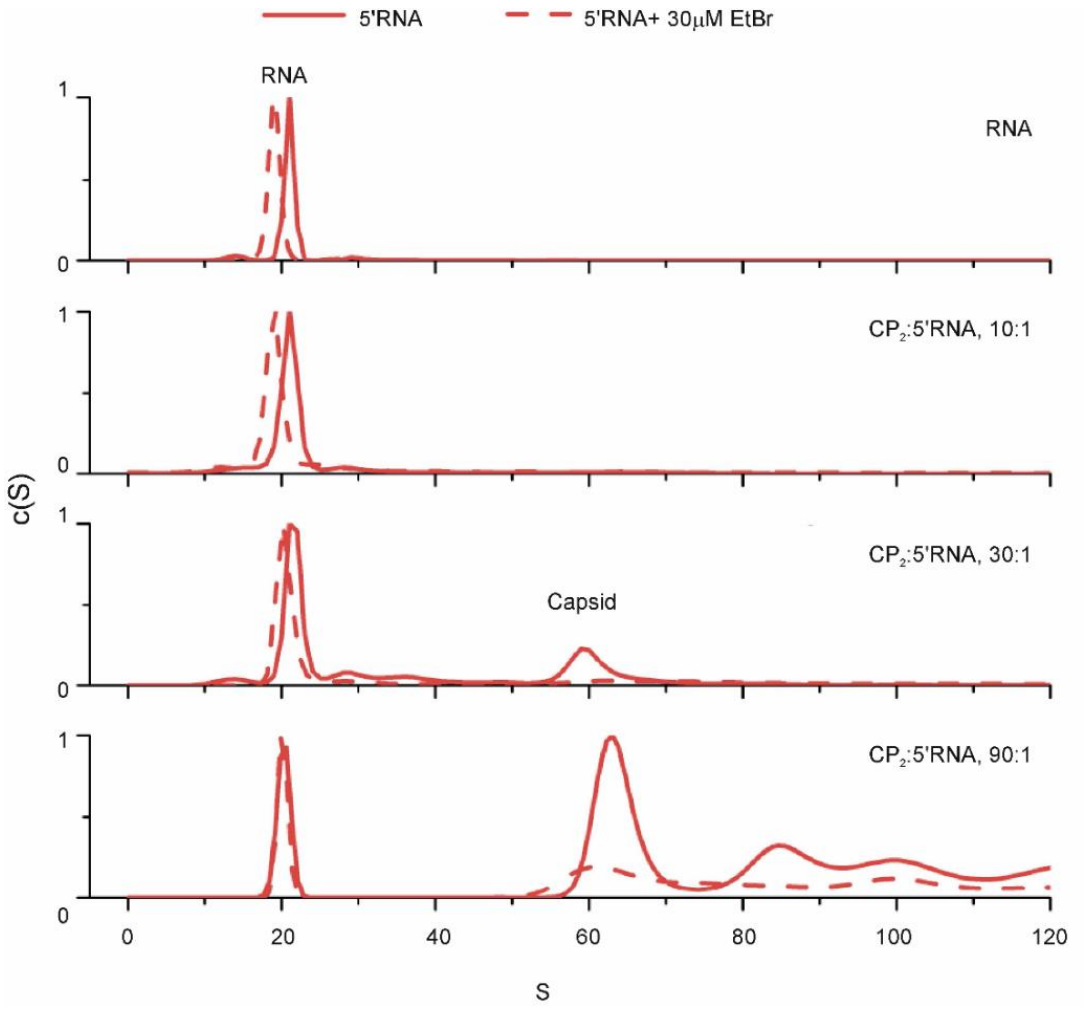


Figure 1.26 The effect of EtBr on the efficiency of MS2 capsid assembly. Capsid assembly was monitored at increasing stoichiometric ratios of CP_2:5'RNA ±30 μM EtBr. svAUC analysis confirms that assembly is restricted in the presence of EtBr. Adapted from Rolfsson et al. (2010).

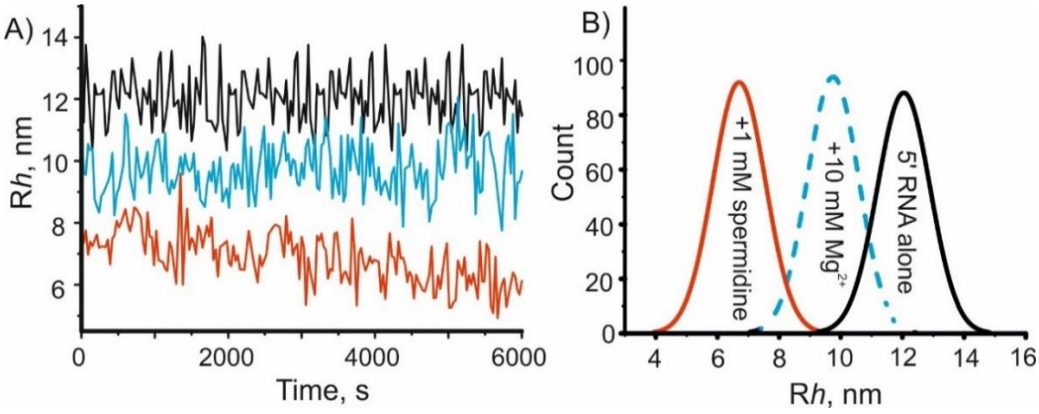


Figure 1.27 Effects of multivalent cations on the hydrodynamic radius of MS2 5′ RNA sub-genomic fragment. MS2 5′ RNA (2577 nt) in monovalent buffer is shown in black, followed by separate reactions in which divalent (blue) or trivalent (red) ions were added. A) smFCS raw data showing the hydrodynamic radius (Rh) of RNA over time. B) The distribution plot shows how these cationic ligands decrease the R_h of 5′ RNA from ~12 nm to ~7 nm. Adapted from Borodavka et al. (2013).

1.9 Thesis Objectives

The aim was to investigate the role of RNA SLs in virus assembly and to utilise their secondary structure as a potential anti-viral target. By exploring the structures of small molecular weight ligands described in Section 1.6, RNAs derived from MS2 gRNA were screened. Reaction affinities and stoichiometries were reported using the techniques described in Section 1.7 and the mechanism of assembly inhibition was investigated. Further work verified if ligands inhibited phage infections *in vivo* and whether resistant mutations were produced. By examining resistant mutants at a nucleic acid level we showed a better understanding of how RNA PSs are involved in virus infection and how to exploit them.

Chapter 2

2. Materials and Methods

This page has been left intentionally blank.

2.1. Synthesis and Purification of RNA

2.1.1. Synthesis of MS2 Genome

A full-length 3569 nt MS2 genomic cDNA template was produced by reverse transcribing wild type phage using Superscript III reverse transcriptase (200 units/reaction, ThermoFisher) and random hexamer oligonucleotide primers (Invitrogen). Primers MS2_F1 and MS2_R1 (Table 2.1) were utilized to amplify the resulting cDNA using high-fidelity Phusion DNA polymerase (1 units/reaction, New England Biolabs). The resulting PCR product was used as a template in a second amplification with 5'-phosphorylated primers MS2_F2 and MS2_R2 (Integrated DNA Technologies). The obtained PCR product was agarose gel purified using QIAquick gel extraction kit (Qiagen) and ligated into a pSMART HCAmp vector (Lucigen) following the manufacturer's protocol. XL1 Blue competent cells (Novagen) were used for transformation with the ligated products and the resulting colonies were PCR-screened and plasmid DNA was isolated from the positive clones, sequenced using primers SL1 and SR2 (Lucigen Corporation). *Hpa*I digestion (20 units/reaction, New England Biolabs) of the resulting plasmid yielded a linearized DNA template that was used for *in vitro* transcriptions.

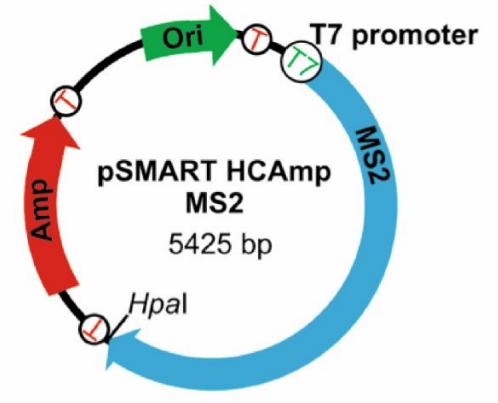


Figure 2.1 Plasmid map of MS2 containing vector.

Table 2.1 Primer names and sequences used for generation of MS2 DNA.

Primer	Primer sequence and restriction enzyme site used
MS2_F1	5'-GGG TGG GAC CCC TTT CGG GGT CCT GCT CAA CTT CCT GTC GAG CTA ATG CCA TT-3'
MS2_R1	5'-GTA ACT AGC CAA GCA GCT AGT TAC CAA ATC GGG AGA ATC CCG GGT CCT CTC TTT-3'
MS2_R2	5'-*GTTAAC*TGGGTGGTAACTAGCCAAGCAGCTAG (*Hpa*I)-3'
MS2_F2	5'-TAATACGACTCACTATAGGGTGGGACCCCTTTCGG-3'

2.1.2. Producing sub-genomic cDNA fragments of MS2 genome

Sub-genomic clones were obtained from pSMART 2676 and pSMART 2578, as previously described in Rolfsson *et al.* (2010). Plasmids were linearized with *Hind*III (10 units/reaction, New England Biolabs) 37°C, 1 h. Sub-genomic fragments 5' (nucleotides 1-2676) and 3' (nucleotides 992-3569) ends were used. For a 50 µL reaction: 10X buffer (5 µL), plasmid (1 µg), nuclease free H_2O (38 µL), linearized with *Hind*III (10 units/reaction, New England Biolabs) 37°C, 1 h. Ran to check for cleavage in native 1% agarose (w/v) TBE gel, see Section 2.2.2.

2.1.3. *In Vitro* Transcription Reactions

For 25 µL reaction: Use Ambion Megascript Kit T7 Transcription Kit Cat No. AM1334.

Reagents were mixed in order: 10X buffer, RNaseOUT (final concentration 1 units/µL), ribonucleotide triphosphates (A, C, G, T) and template (1-2 µg), with enzyme T7 (2 µL) dilute with DEPC-treated water to 20 µL. Incubate 37°C for 2 h. Quench reaction using Turbo DNase (2 units/reaction, Fisher Scientific) and 10X buffer, incubate 37°C for 30 min. Run on denaturing 1% (w/v) agarose gel, see Section 2.2.1.

2.1.4. Synthesis of RNA stem loops

Using solid-phase synthesis RNA oligonucleotides were made with *N*-benzoyl-protected (NBz) Adenosine, *N*-$_i$methylformamidinyl-protected (NMe) Guanosine and *N*-acetyl-protected (NAc) Cytosine; no protective group is required for Uracil. During synthesis the 2' hydroxyl group was

protected by O-tert-butyldimethylsilyl (TBDMS) and the 3'OH was protected by 2-cyanoethyl-*N,N*'-diisopropyl (DMTr). All synthesis reagents were from Link Technologies Ltd. See Figure 2.2 for reaction schematic.

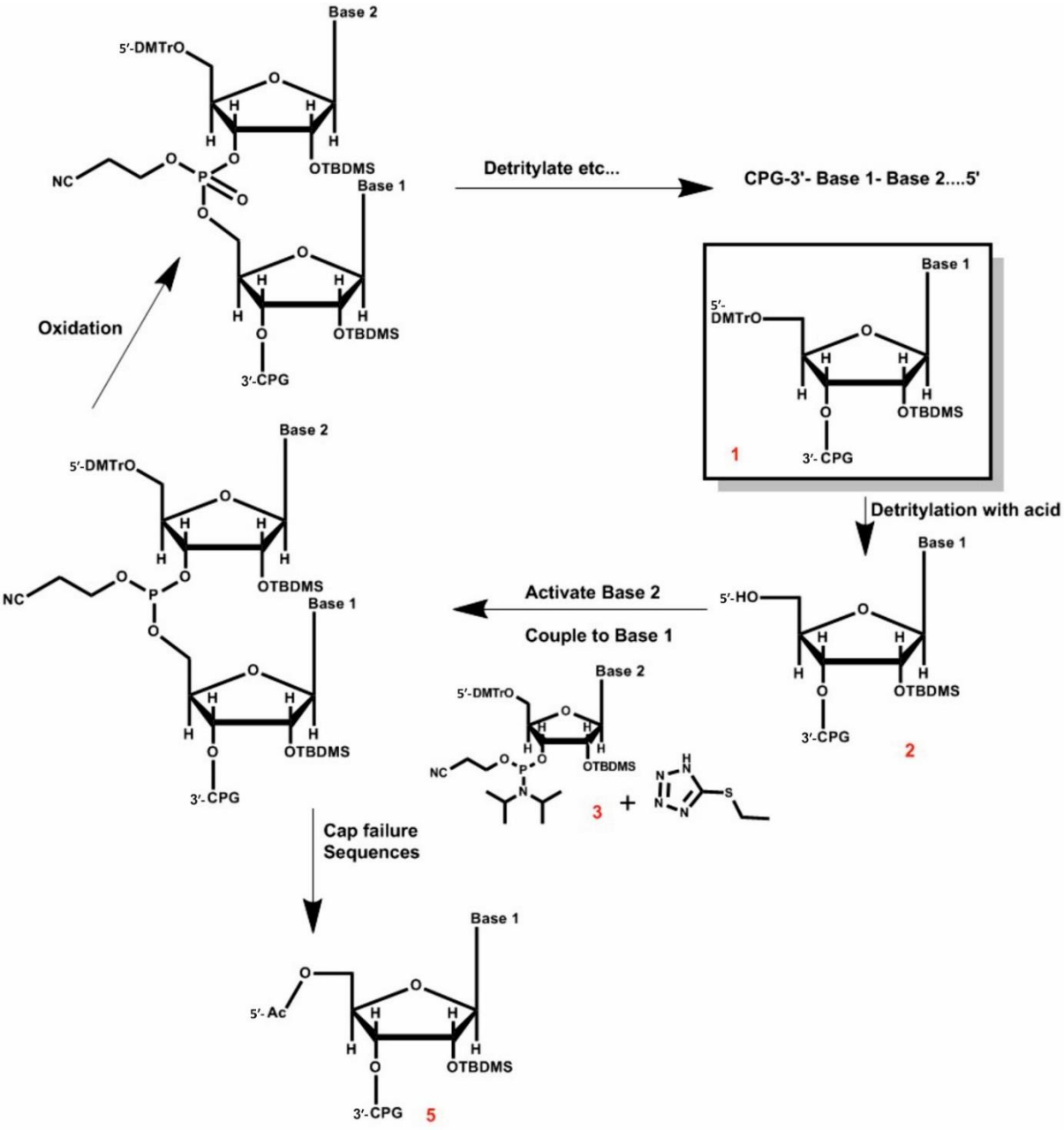


Figure 2.2 Schematic of solid phase RNA synthesis. Starting with 3'-end protected nucleotide coupled to CPG (1), after detritylation (DMTr) of 4'OH (2) it is ready for addition of the next base (3), in the presence of coupling reagent. All failure sequences (5) are capped to prevent growth of failure sequences. Successful base additions (4) are pushed forward for oxidation of the phosphate group (6) and then ready for next base addition or for subsequent cleavage from CPG, removal of protecting groups and purification as outlined in text.

Ammonia-saturated methanol at room temperature for 16 h was used to remove protecting groups and to cleave RNA from CPG resin. Methanol was removed under vacuum and the pellet was re-suspended in anhydrous dimethyl sulfoxide (DMSO). One volume of triethylamine trihydrofluoride (TEA.3HF) was added and incubated at room temperature to remove TBDMS from the 2′ hydroxyl groups. RNAs were butan-1-ol precipitated and re-suspended into DEPC-treated water (Severn Biotech).

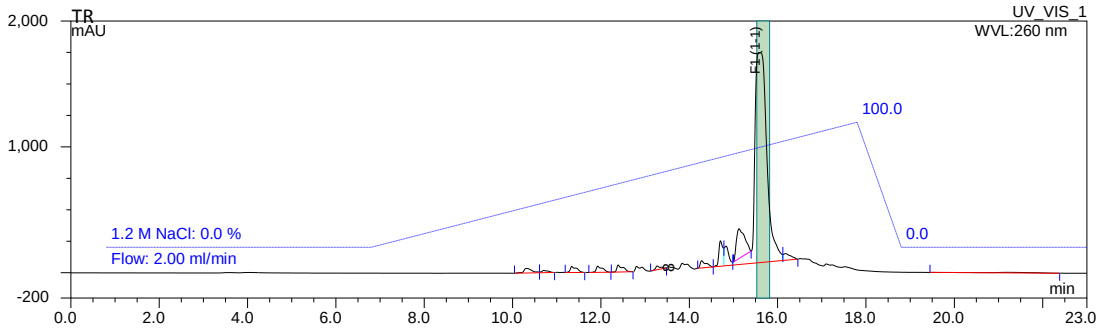
2.1.5. RNA Purification by High Performance Liquid Chromatography (HPLC)

Each un-modified RNA (<37 nucleotides) was purified by ion-exchange HPLC using a Dionex BioLC and a DNAPac PA-100 anion exchange column (Dionex). The purification was at 60°C with a gradient from 100% 18.2 mΩ autoclaved water to 1.2 M NaCl. Dye-labelled oligonucleotides were purified by reverse phase HPLC using a Dionex BioLC and a Luna “C18” column (Phenomenex). The purification was run at 55°C with a gradient from triethylammonium acetate (50 mM) to 100% triethylammonium acetate (50 mM in 50% acetonitrile). See Figures 2.2 and 2.3 for purification chromatograms. Each full-length RNA were eluted as a major species, after elution of minor failure sequences. RNA fractions were collected, as highlighted in green, lyophilized and desalted into 18.2 mΩ H_2O. Length of synthesised RNAs were checked on 10% (w/v) 29:1 acrylamide:bis-acrylamide urea (7.5 M) gels, see Section 2.2.3. For RNA synthesis yields and sequence information see Table 2.2.

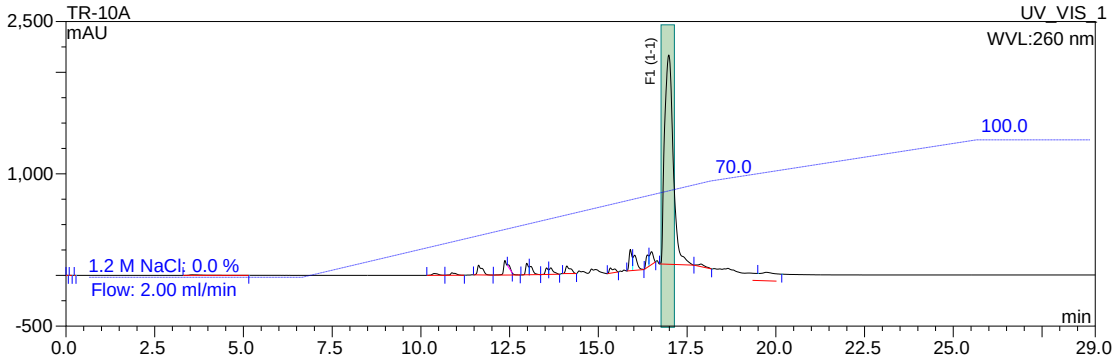
Table 2.2 RNA Oligonucleotides Sequences and Information (1 µmol synthesis). RNAs 1-6 Generated in house by solid-phase synthesis, 7-9 purchased from IDT and purified in house. Estimated melting temperatures (Tm) determined by IDT OligoAnalyzer 3.1.

RNA	Name	Comment	Sequence 5'→3'	MWt. g.mol^{-1}	Extinction Coefficient (ε_{260nm}) M^{-1}.cm^{-1}	Tm (°C)	Yield (nmoles)
1	TR	MS2 aptamer	ACA UGA GGA UUA CCC AUG U	6,045.7	185,000	48.1	71.9
2	TR-10A	Mutant MS2 aptamer	ACA UG GGA UUA CCC AUG U	5,716.0	171,800	51.4	72.8
3	B3	STNV aptamer	CCU UUU CAA GAC AUG GCA ACA AUG CAC ACA G	9,523.8	282,700	54.4	132.5
4	PS6	HParV aptamer	UCU UCC ACA UGU UUU GAU GAA UUU GGC UGA	9,490.6	264,400	55.7	96.2
5	PS1	HepB aptamer	UUU GUU UAA AGA CUG GGA GGA GUU GGG GGA GGA G	11,155.7	336300	60.0	117.5
6	TCV	TCV aptamer	CUG AGG AGC AGC CAA AGG GUA AAU UGC AAG CAC UCA	11,863.0	355,400	60.0	34.0
7	5'wt	Wt of Nt 386-418 MS2 genome	CGA GGA GAA AGC CCG GUU UCG GCU UCU CCC UCG	10,231.2	292,100	66.7	71.5
8	5'mutant	Mutant of Nt 386-418 MS2 genome G393U	CGA GGA UAA AGC CCG GUU UCG GCU UCU CCC UCG	10,192.1	289,600	65.4	53.0
9	Qβ	Qbeta aptamer	AUG CAU GUC UAA GAC AGC AU	6,374.9	195,800	47.3	73.1
10	GA	GA aptamer	AAC AUA AGG AAA ACC UAU GUU	7,346.5	215,100	36.3	71.2

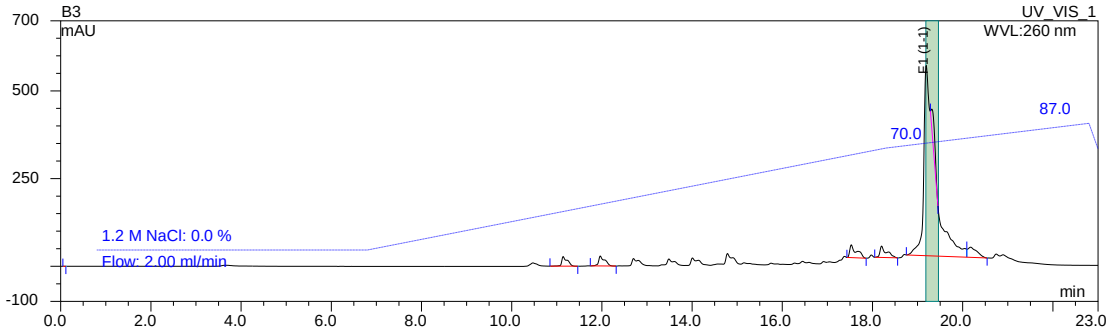
1)



2)



3)



4)

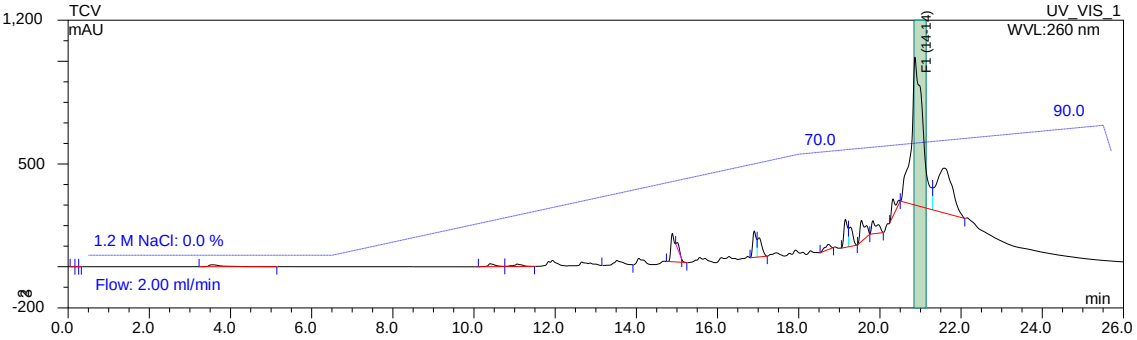
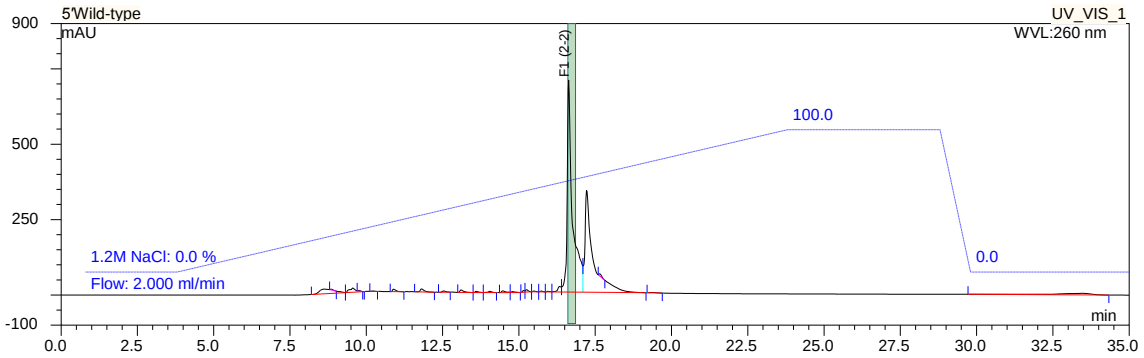
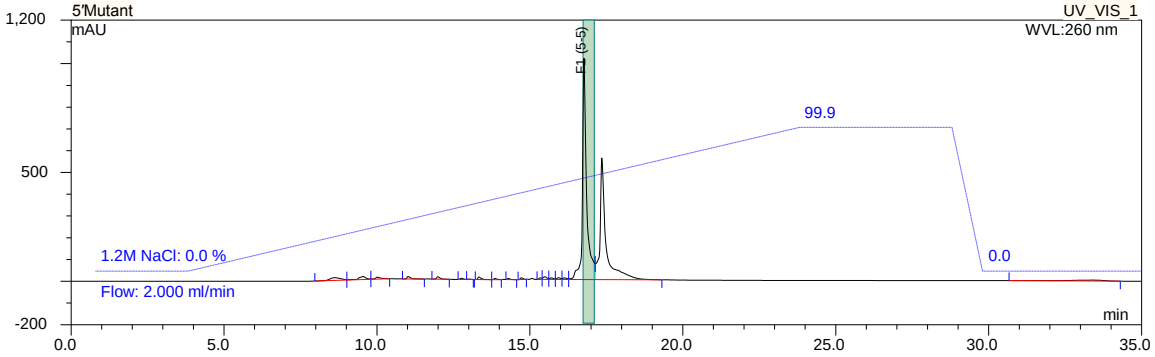


Figure 2.3 Ion-Exchange Chromatograms of RNA Oligonucleotide Purification. 1) TR. 2) TR-10A. 3) B3. 4) TCV. Major eluting product was collected, highlighted in green. See Table 2.1 for further information.

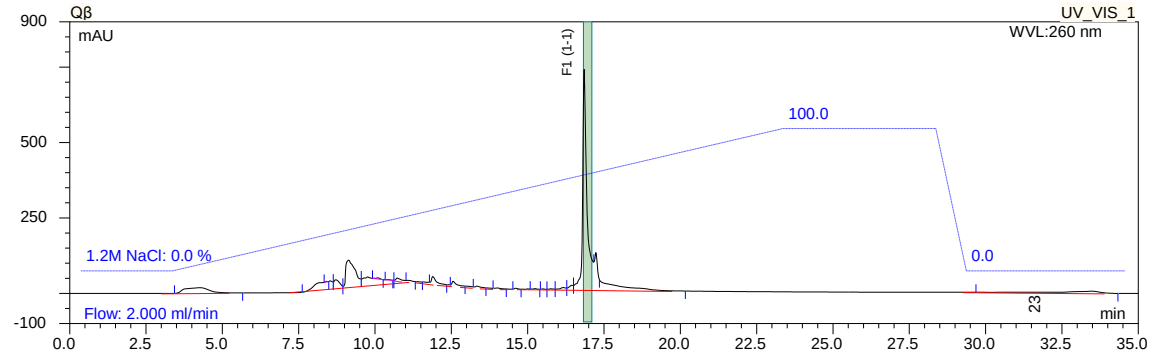
5)



6)



7)



8)

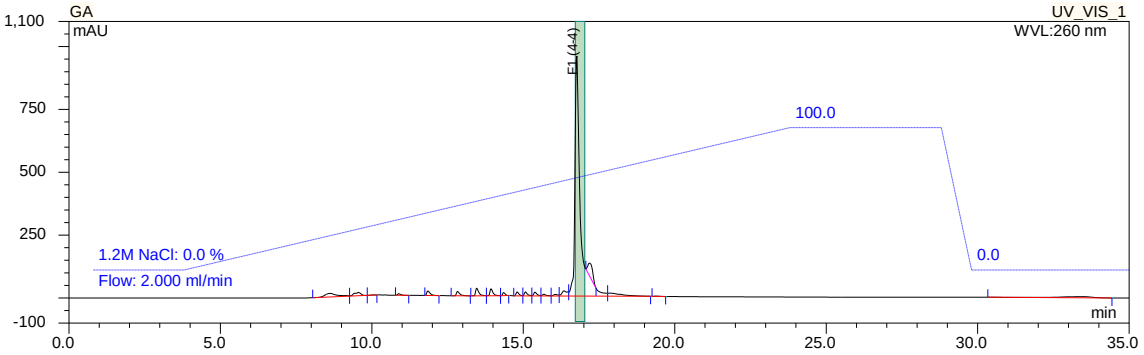


Figure 2.4 Ion-Exchange Chromatograms of RNA Oligonucleotide Purification. 5) 5'Wild-type. 6) 5'Mutant. 7) Qβ. 8) GA. Major eluting product was collected, highlighted in green. See Table 2.1 for further information.

2.1.6. Extraction of gRNA from wild-type MS2

After an MS2 infection assay, described in Section 2.3.5, samples (100 µL) were purified by ion-exchange using an RNAeasy Kit (Qiagen). RNA was isolated by using guanidine-isothiocyanate to denature bound proteins and a silica-membrane allowed for collection and purification. The Beer-Lambert law was used to quantitate RNA using ε_{260nm}= 46241.3 $M^{-1}.cm^{-1}$ as calculated from the nucleic acid sequence (https://eu.idtdna.com/calc/analyzer).

2.1.7. Reverse Transcription-PCR of 620 base region of MS2

Samples of purified MS2 RNA were heated at 95° for 5 min then incubated on ice. Reverse transcription was carried out using SuperScript III (200 units/reaction, Invitrogen) as per the manufacturer's instructions. Incubating at 52°C for 2 h, generated a 620 base DNA copy of the region of interest (Shin and Sobsey, 2003). PCR was performed for 30 rounds using KAPA2G Robust (Kapa Biosystems) as per the manufacturer's instructions. PCR products were purified using QIAquick PCR Purification Kit (Qiagen) as per the manufacturer's instructions. It uses a silica membrane assembly to bind DNA in high-salt buffer, followed by elution of DNA in low-salt buffer or water. The purification procedure removes primers, nucleotides, enzymes, and salt impurities from DNA samples.

Table 2.3 Primer names and sequences used for RT-PCR.

Primer	Comment	Sequence
MS2_620_F	Nucleotides 1748-1759	5'-ATG AGG ATT ACC CAT GTC GAA G-3'
MS2_620_R	Reverse complement of nucleotides 2310-2920	5'-TCC CTA CAA CGA GCC TAA ATT C-3'

Reverse transcription reaction mix (Invitrogen) 15 µL: 5X buffer (7 µL), 10 mM dNTP (1 µL), and reverse primer MS2_620_R (100 pmol), reverse transcriptase (200 U), DEPC-treated H_2O (5 µL).

PCR reaction mix (Kapa Biosystems) 15 µL: 5X buffer (10 µL), 10 mM dNTP (1 µL), primer MS2_620_F (25 pmol), KAPA2G enzyme (0.25 µL), DEPC-treated H_2O (3.5 µL).

To verify that the correct DNA was synthesised, samples (10 µL) were mixed with glycerol (75% (v/v), 5 µL) and run on 2% (w/v) agarose TBE gel for 90 V for 1 h at room temperature.

2.1.8. Generating Full Length MS2 cDNA for Next-Generation Sequencing

The protocol was repeated as detailed in Section 2.1.7, but using the primers (MS2_F1 and MS2_R1) as listed in Section 2.1.3.

2.2. Gel Electrophoresis Techniques

2.2.1. Denaturing 1% (w/v) Agarose for Gel Purifying RNA

Agarose (0.5 g) was dissolved in 18.2mΩ H_2O (36 mL), with gentle heating. In a fume cupboard 10X MOPS (5 mL), formaldehyde (35%, 9 mL) was added. EtBr (0.5 µL, 10 $mg.mL^{-1}$ stock) was added to samples before loading. Gel run in 1x MOPS at 80 V for 1 h at 4°C.

2.2.2. Native 1% (w/v) Agarose Gel for Large Plasmids/DNA

Agarose (0.7 g) was dissolved in 1X TBE (70 mL), with gentle heating. EtBr (0.5 µL, 10 $mgmL^{-1}$ stock) was added to samples before loading. Gel run in 1X TBE at 80 V for 1 h at room temperature.

2.2.3. Denaturing 10% (v/v) Polyacrylamide Gel for RNA

Made as per manufacturer's instructions (SequaGel, National Diagnositics). Gel was pre-heated in 1X TBE at 14 W for 60 min before loading samples. Amended recipe to 20% (v/v) gel which was run for 3 h at 200 V for RNA <40mer, as directed by manufacturer.

2.2.4. Native 15% (v/v)Polyacrylamide Gel for DNA/RNA

Acrylamide/Bis-acrylamide (29:1 v/v 30% solution) (18 mL) was mixed with with 5X TBE (7.2 mL), 18. 2mΩ H_2O (10.8 mL), and add 10% w/v APS (360 µL) and TEMED (36 µL) to polymerise. Gels were run at 200 V for 2.5 h at 4°C. Visualised post-run with 10^5 (v/v) dilution of EtBr or SybrGold stain (both stains were 1/10,000 (v/v) dilutions).

2.2.5. Denaturing SDS 15 % (v/v) Polyacrylamide Protein Gel

To make 2 (10 x 8 x 0.75 cm) gels: 15% Separating gel mix H_2O (3.75 mL), 30% acrylamide (7.5 mL), Tris-HCl 1.5 M pH 8.8 (3.75 mL) and 10% SDS (0.5 mL) were added to 10% APS (150 µL) and

TEMED (15 µL) to polymerise. For Stacking gel H_2O (4.2 mL), 30% stock acrylamide (Protogel, 0.65 mL), Tris-HCl 0.5 M pH 6.8 (1.6 mL), 10% SDS (0.25 mL) were mixed and added to 10% APS (67 µL) and TEMED (6.7 µL) to set. Samples were boiled in denaturing SDS gel loading Laemmli buffer (5 min at 105°C). Gels were run in 1X SDS PAGE running buffer at 200 V for 1 h at room temperature and visualised using InstantBlue™.

2.3. *In vivo* Methods: Preparing *E. coli.* Cell Cultures and Propagating Bacteriophage

2.3.1. *E. coli.* Strain C3000 Preparation

Working aseptically, *E. coli* strain expressing F-Pili, C3000 (ATCC® 15597), was spread onto agar and grown overnight at 37°C. A single colony was picked and cultured in minimal LB (10 mL) with (100 µL) MS2 supplement (D-glucose (0.5 M), $CaCl_2.2H_2O$ (165 mM), thiamine (0.3 mM) all 0.2 µm filtered) and incubated at 37°C, 200 rpm, overnight and stored at 4°C until needed. Fresh overnight cultures were made from single colonies every week to keep fresh cell stocks.

Minimal LB: tryptone (10 g), yeast (1 g), NaCl (8 g), pH 7.5, 18.2 mΩ water (1 L) autoclaved.

2.3.2. Propagation and Purification of Stock Bacteriophages MS2 and GA

C3000 cells were grown to OD_{600} = 0.2-0.4 (37°C, 220 rpm), then infected with sterile phage stock at a multiplicity of infection (MOI) = 5-10. Bacteriophages were wild-type MS2 (ATCC® 15597-B1™) and purified GA was a gift from Professor David Peabody, University of New Mexico. Flasks were held stationery at 37°C for 15 min to allow phage to adsorb onto bacterial pili. Infections were incubated for another 5-6 h at 37°C, 100 rpm. To harvest bacteriophage, ammonium sulfate (280g/L of lysate) and EDTA (final concentration 5 mM) were added and left stirring slowly overnight at 4°C. Bacteriophage were precipitated by centrifugation (10,000 *g* for 30 min at 4°C) and re-suspended in 50 ml of TNE buffer (100 mM NaCl, 50 mM Tris-HCl, 5 mM EDTA, pH = 7.2). To remove cell debris, an equal volume of $CHCl_3$ was added. The mix was shaken

vigorously to form a suspension (5-10 min), then separated by centrifugation (1 min, 3,000 *g*). The aqueous phase containing bacteriophage was collected for further purification.

2.3.3. Propagation and Purification of Stock Bacteriophage Qβ

E. coli bacteriophage Qβ (ATCC® 23631-B1™) was propagated as per the manufacturer's instructions, using *E.coli.* C3000 cells and TYG medium.

TYG: tryptone (10 g), glucose (1g), yeast (1 g), CaCl2 (0.3g), NaCl (8 g), pH 7.5, 18.2 mΩ water (1 L) autoclaved.

To isolate and purify Qβ, the protocol detailed in 2.3.2 was used, initiating from the ammonium sulfate precipitation step.

2.3.4. Caesium Chloride (CsCl) Purification of Bacteriophages

Bacteriophage were diluted in TNE buffer and CsCl added to a final density of $\rho = 1.38 \pm 0.01 g.mL^{-1}$ (0.55 g of CsCl per 1 g of sample in TNE buffer). The solution was left overnight on a roller mixer at 4°C and density double checked by weighing 1 mL. Using a SW32Ti Beckman rotor, CsCl mixtures were centrifuged at 25,000 rpm, 4°C for 56 h in 38.6 mL tubes. There are several visible bands by measuring for absorbance at 280 nm. Using automated fractionation (BioComp Systems Inc.) the major band at 280 nm corresponded to the wild-type MS2, see Figure 2.5.

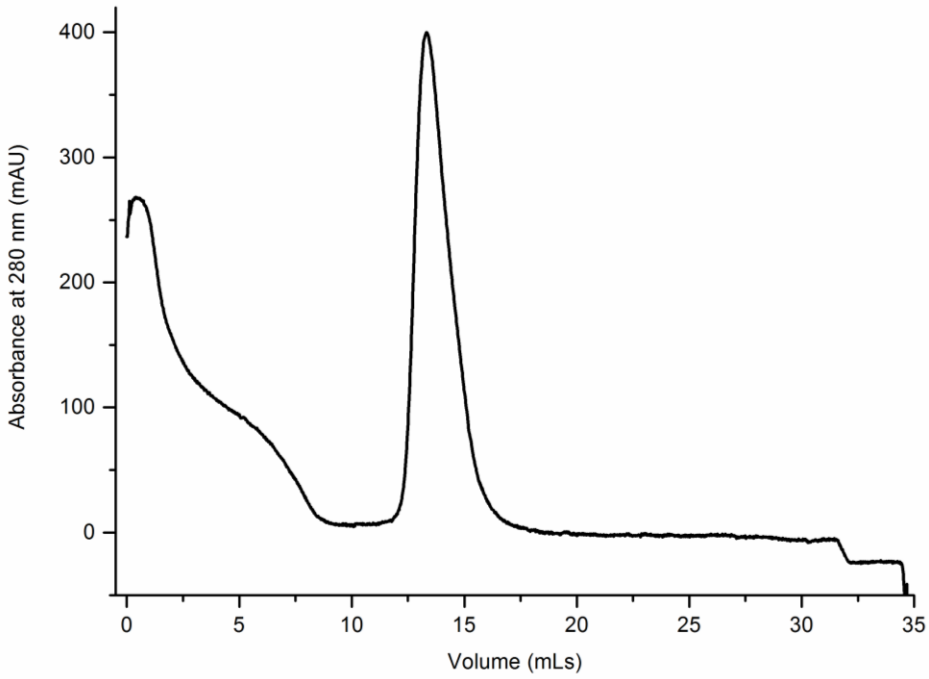


Figure 2.5 Fractionation of wild-type MS2 purified by CsCl gradient, trace shown is absorbance at 280 nm. Gradients were spun at 25,000 rpm, for 56 h at 4°C. The major band at ~14 mL corresponds to infectious virions as verified by plaque assay.

2.3.5. Bacteriophage Infection Assays

E. coli C3000 cells (100 µL overnight culture) were grown in minimal LB (10 mL) with MS2 supplement (100 µL) to $O.D_{650nm} = +0.5$, at 37°C and 180 rpm. They were infected with the desired MOI using bacteriophage stocks. All assays were left stationery for a min 15 min post-infection, this ensured bacteriophage adsorbed to F. pili. Successful infections caused cell lysis, as indicated by a drop in absorbance.

2.4. Plaque Assays: Quantitating Infectivity

All work was carried out aseptically, next to a Bunsen burner or in a laminar flow hood. Bacteriophage was diluted using MS2 dilution buffer (10 mM Tris-HCl, 100 mM NaCl, pH 7.6, autoclaved) and each dilution was performed in triplicate. For each plate C3000 (200 µL overnight culture) was added to bacteriophage dilution (100 µL) after gentle mixing, molten agarose top (9.7 mL) was added, mixed thoroughly and poured directly onto agar plate and left overnight at 37°C. The plaques were counted after a minimum of 12 h. See Figure 2.6 for examples of plaque assay results.

Agar plates: Tryptone (10 g), yeast (1 g), NaCl (8 g), agar (15 g) at pH 7.5, made up to 1 L and autoclaved, MS2 supplement (12 mL) added post autoclave. Plates allowed to dry at room temperature for 2 d before use.

Agarose top: Agarose (7 g), tryptone (5 g), yeast (0.5 g), NaCl (4 g), D-glucose (0.5 g), $CaCl_2.2H_2O$ (0.15 g) at pH 7.5, made up to 500 mL and autoclaved.

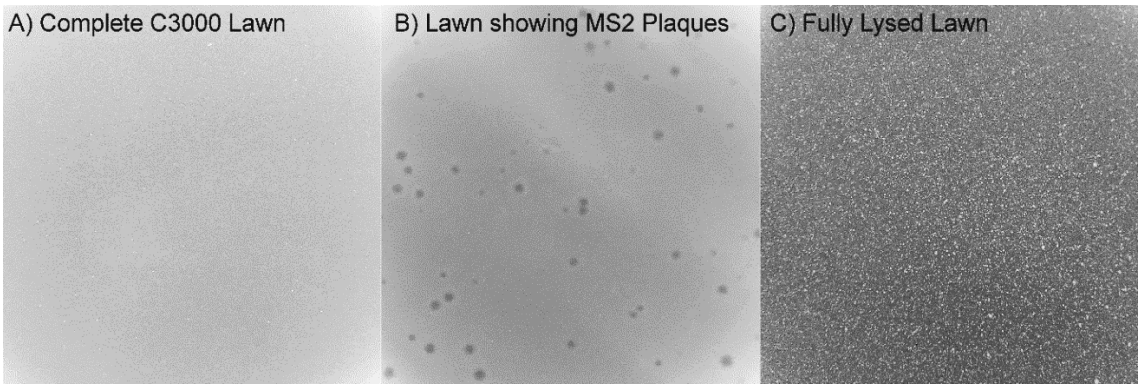


Figure 2.6 Photos of plaques assay demonstrating the effect of MS2 on C3000 agarose lawns. A) An uninfected E. coli. C3000 lawn is opaque and cell density even across the whole plate. B) C3000 infected with MS2 shows circular plaques, hazy in appearance. These can be clearly counted and used as a quantitative method to compare infectivity of different MS2 infection assays. C) C3000 cells infected with a high titre of MS2 shows a clear, transparent plate dotted with cell debris indicating the bacteriophage has lysed cells.

2.5. Bacterial Cell Growth Kinetics

Monitoring cell growth kinetics was done in a FLUOstar Galaxy plate reader with absorbance optics, using 96-well plates (Nuclon Delta) with maximum volume 175 µL. Light scattering was monitored at 650 nm, at 37°C and 100-200 rpm where applicable.

2.6. Western Blot to Identify wild-type MS2 Coat Protein

After an MS2 infection assay, lysates (30 µL) were denatured in SDS-PAGE buffer and run in duplicate on 15% (v/v) SDS-PAGE Gel at 200 V for 45 min, alongside a pre-stained protein ladder (BenchMark™ Invitrogen Life Technologies). One gel was stained with InstantBlue™ (Expedeon), the other prepared for Western Blot. Wadding and nitrocellulose paper were pre-soaked in transfer buffer. The second SDS PAGE gel was assembled in Western Blot apparatus and run in transfer buffer at 100 V for 1 h at room temperature. After successful protein transfer, the nitrocellulose was left in blocking buffer (50 mL) at 4°C overnight. Nitrocellulose membrane was washed with rinse buffer, 2 x 15 mL for 15 min. It was then washed in detergent buffer, 3 x 15 mL for 15 min. Primary antibody (Rabbit Ab MS2, 1 µL) was added to the last rinse and left for 1 h. The membrane was washed again with detergent buffer, 2 x 15 mL for 15 min). Secondary antibody (anti-Rabbit IgG isolated from goat 1 µL, Sigma). Antibodies were visualised using an

enhanced chemiluminescence (ECL) Kit (Pierce) and exposing to Kodak BioMax film (Perkin Elmer).

Buffer	Composition
Transfer buffer	MeOH (50mL), glycerine (2.9 g), Tris (5.8 g), SDS (0.37g), made up to 1 L with H_2O.
Blocking buffer	Tris (1.513 g), NaCl (2.19 g) at pH 7.5 and add dried milk powder (9.375 g) (Marvel), made up to 250 mL with H_2O.
Rinse buffer	Tris (1.513 g), NaCl (2.19 g) at pH 7.5 and made up to 250 mL with H_2O.
Detergent buffer	Rinse buffer with Tween20™ (Sigma) 0.1% (v/v).

2.7. Sedimentation Velocity Analytical Ultracentrifugation (svAUC) of RNAs

svAUC is a technique where molecules evenly distributed in a cell are subjected to a high angular velocity, causing the molecules to sediment towards the bottom of the cell. The molecules form a boundary in solution, showing concentration depletion at the solution meniscus and a concentration "plateau", see Figure 2.7. The velocity of the individual particles cannot be resolved, but the rate of movement of the boundary region can be measured. From this, the sedimentation coefficient (s) can be determined, which in turn depends on the mass and reciprocal of the frictional coefficient. See Equation 1 (Svedberg and Pederson 1940).

$$\frac{M(1-\tilde{\nu}\rho)}{Nf} = \frac{u}{\omega^2 r} = S \qquad \text{Equation 1.}$$

M = molecular weight ($g.mol^{-1}$)

$\tilde{\nu}$ = partial specific volume of molecules ($cm^3.g^{-1}$)

ρ = density of solvent ($g.mL^{-1}$)

N = Avogadro's number 6.022×10^{23} mol^{-1}

f = frictional coefficient of molecule

$\frac{u}{\omega^2 r}$ = velocity of the particle per unit of gravitational acceleration

s = sedimentation coefficient (Svedberg 10^{-13} seconds)

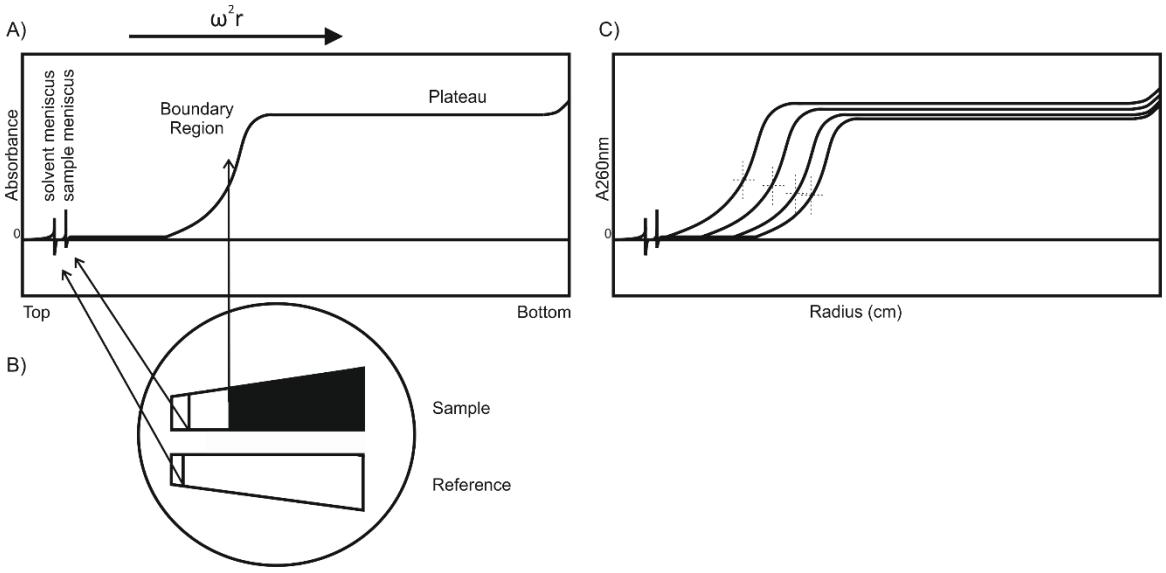


Figure 2.7 Schematic demonstrating solute boundary formation in a svAUC experiment. A) A sharp boundary forms when a high angular velocity is applied to a solution of molecules. B) Diagram of two-sector sedimentation velocity (sv) cell, showing concentration gradient in sample during boundary formation. C) Movement of a boundary in a svAUC experiment. Crosses indicate the mid points of each boundary.

Samples (0.32 mL) were placed in 1.2 cm pathlength 2-sector epon centrepiece cells built with sapphire windows. These were then centrifuged at 17,000 rpm at 20.0°C in an An50Ti analytical rotor in an Optima XL-I analytical ultracentrifuge (Beckman). Changes in solute concentration were detected by interference and absorbance scans at 260 nm. Buffer densities and viscosities were calculated using the program Sednterp version 1.09 (Laue, Shah et al. 1992). Radial absorbance plots were to fit for sedimentation profiles with the program Sedfit, version 12.1b 2010, using a continuous distribution c(s) Lamm equation model (Schuck 2000). The Lamm equation models how a macromolecular species sediments as a function of time and radial position, under the influence of sedimentation and diffusion in a sector shaped cell (Lamm 1929). For examples of raw data and c(s) analysis see results Chapter 3 section and Appendix 8.1.

2.8. Circular Dichroism Spectroscopy of RNAs

Circular Dichroism (CD) spectra were recorded using a Chirascan™ CD Spectrometer in a 1 mm path length quartz cell at various temperatures. The CD spectra were recorded over a range of

wavelength between 180–320 nm at 1 nm increments. Samples were 100 µL RNA (1 µM) in specified buffers and blank buffer spectra were collected for referencing.

2.9. Absorbance Spectroscopy of RNAs

Absorbance was measured using a UV-1800 UV spectrophotometer (Shimadzu) or a Photon Technology International QM-1 spectrofluorimeter. Data was collected at 20.0°C, in 1 cm pathlength quartz cuvette and scanned in 1 nm increments.

2.10. Mass Spectrometry (MS) Techniques

2.10.1. Native, Positive-ion Mode MS for Studying VLP Formation

Samples were analysed in aqueous 50 mM and 200 mM ammonium acetate buffers. Samples were sprayed in gold/palladium plated capillaries pulled in-house using a Model P-97 Laming/Brown micropipette puller (Sutter Instrument Co., Novato, CA, USA) and coated using an SC7620 sputter coater (Quorum Technologies Ltd, East Grinstead, West Sussex, UK). Ionisation was carried out by positive nano-electrospray ionisation (n-ESI) with a capillary voltage of 0.8 – 1.2 kV. Samples were analysed using a Synapt G2-S (Waters Corp., Manchester, UK), an orthogonal Time of Flight (ToF) instrument (Waters Corp., Manchester, UK) and a Synapt high definition mass spectrometer (HDMS) (Waters Corp., Manchester, UK). Ion mobility (IMS) experiments were carried out in the drift cell, using He as a sheath gas (2.39×10^{-2} bar) to induce collisional cooling, and Ar as the buffer gas (3.01 bar). Ion separations were achieved with a wave height of 40 V and a wave velocity of 1400 m/s.

2.10.2. Denaturing, Negative-ion Mode MS for RNA Mass Determination

RNA samples were analysed on the Synapt HDMS (Waters Corp., Manchester, UK) mass spectrometer. The sample was diluted to 70% (v/v) acetonitrile/1% (v/v) TEA for negative electrospray MS. The instrument was calibrated using a separate injection of sodium iodide at a concentration of 2 $\mu g.\mu L^{-1}$. The mass spectrum shows multiply charged ions and these data were

then transformed on to a molecular mass scale using the maximum entropy processing technique (Skilling *et al.* 1989*)*. Data were collected by Drs. Henry Fisher and James Ault, Astbury Centre, University of Leeds.

2.10.3. Ion Mobility MS to Study RNA:Ligand Interactions

Samples were analysed by Z-spray n-ESI MS using a quadrupole-IMS-orthogonal ToF MS (Synapt HDMS, Waters UK Ltd., Manchester, U.K.) using gold/palladium-coated nanospray tips prepared in-house. The MS was operated in positive-ion mobility-TOF mode using a capillary voltage of 1.2 kV, and cone voltage of 60 V. The source and desolvation temperatures were set at 80°C and 150°C, respectively. Nanoelectrospray gas pressure was 0.1 bar, source backing pressure was 4.16 mbar, trap and transfer argon gas pressure was 1.79 $x10^{-2}$ mbar and the IMS cell nitrogen gas pressure was 3.47 $x10^{-1}$ mbar. Trap collision energy was 6.0 V, transfer collision energy was 4.0 V and a trap bias of 20 V was used. The IMS travelling wave speed was 30 $m.s^{-1}$ and the wave height was 4.2 V. Mass calibration was performed by a separate injection of sodium iodide at a concentration of 2 $\mu g.\mu L^{-1}$. Data processing was performed using the MassLynx v4.1 (Waters UK Ltd.). Data were collected by Dr. James Ault, University of Leeds.

2.10.4. Denaturing, Negative-ion Mode MS for RNA:Cocktail MS Mass Determination

Ligands shortlisted were obtained commercially or gifted from MCCB Technology Group, School of Chemistry at the University of Leeds. Some of these were omitted from experimental work in Chapter 3 because of aqueous solubility issues. ESI-MS is compatible with a wide range of solvents such as DMSO, acetonitrile (MeCN) and methanol (MeOH), DMSO is frequently used in *in vivo* drug delivery, but it was omitted from use due to its capacity to “clog up” the MS when used in high-throughput mode. After some initial experiments using TR, it was decided to use MeCN as the ligand solvent because it showed no denaturing effects on RNA secondary structure and is compatible with the ligand purification protocol (Appendix 8.6). Ligands were purified using a C-18 matrix and buffer exchanged into 50% MeCN/H_2O (Pierce C-18 Tips, Thermo

Scientific) and divided up into 5 cocktails consisting of up to 5 ligands. The secondary structure of compounds in ligand cocktail #1, 2CL-1B-MECA, MTX, A3, Quercin and Terazosin are shown in Chapter 6, Figure 6.3 and ligand MS information shown in Appendix 8.7. PSs (34 μM) were heat annealed in the presence of $Mg(OAc)_2$ (1 mM) and exchanged into MS buffer, 200 mM ammonium acetate, pH 6.5 (Micro Bio-Spin 6, BioRad). RNAs and cocktails were mixed (1:20 molar ratio) in a 96-well plate, at room temperature, then loaded onto the MS auto-sampler to check for binding.

2.11. cDNA Preparation for NextGen Sequencing

RNA was extracted from infection assays as outlined in the protocol in Section 2.3.5. Reverse transcription, PCR, DNA isolation and gel extraction was carried out as described in Chapter 5. Agarose gels were run in order to verify that cDNA had been isolated, all contained approximately 10 $ng.\mu L^{-1}$, loading 10 μL.

2.12. MiSeq Data Analysis and Instrumentation

Illumina TruSeq Stranded Total RNA next generation libraries were generated from the purified cDNA and sequenced on an Illumina MiSeq sequencer to generate 150 bp paired-end sequence data. The data were aligned to the MS2 bacteriophage genome (NC ID 001417) using the BWA aligner (Li and Durbin 2009). Samtools (Li et al 2009) was used to sort the aligned data by position before converting the data to mpileup formatted files from which the variants were identified using the mpileup2snp and mpileup2indel methods of Varscan (Kobolt et al. 2009 and 2012).

2.13. X-ray Footprinting (XRF).

Experiments were carried out using pre-optimised protocols by Hulscher *et al.* (2016) at the National Synchroton Light Source Facility (NSLS-II, Brookhaven National Laboratory, Upton, NY, U.S.A.) working with Dr. Jen Bohon (Case Western University, Cleveland, OH) and Dr. Rebecca Chandler-Bostock (University of Leeds, U.K.). The beamline intensity was estimated at the start

of each run by exposing Alexa-488™ (2 μM in PBS) for specified times (10-50 ms) to give a standard curve. Experimental samples (0.5 $mg.mL^{-1}$, 5 μL in phosphate buffered saline (PBS)) were loaded into a pre-cooled sample holder, and exposed to beamline radiation for 50-200 ms. The samples were shipped back to Leeds for analysis. RNA was extracted from exposed samples using AVL Lysis buffer (Qiagen). First strand synthesis of a specified region of gRNA was performed using a high fidelity reverse transcriptase (Superscript IV, ThermoFisher™) and a DNA primer labelled with 5'-fluorescein to aid downstream processing. RNA was degraded (400 mM NaOH, 95°C, 5 min), DNA copies purified (Monarch PCR Clean Up Kit, NEB) and prepared for capillary electrophoresis. Sequencing ladder controls were prepared using DNA primers labelled with 5'-HEX™ and deoxyadenosine triphosphate:dideoxyadenosine triphosphate (dATP:ddATP, 1:8 molar ratio). Samples were spiked with sequencing ladder (10% mole/mole) and submitted for fragment analysis in HiDi Formamide (ThermoFisher™) to DNASeq (University of Dundee, U.K).

Chapter 3

3. Assaying Small Ligands Binding to TR

This page has been left intentionally blank.

3.1. Work Flow for Identifying RNA Ligands *in silico*

The use of *de novo* design to target specific viral RNA sequences and secondary structures is desirable, but requires specialist techniques outside the scope of this thesis. Instead an *in silico* approach was used to find novel RNA-binding molecules from current FDA-approved molecules. Not only does this cut down on synthesis optimization and experimental characterization, but success would allow pre-approved drugs to be "recycled" for different targets, namely as anti-virals. The plan of action is outlined in Figure 3.1.

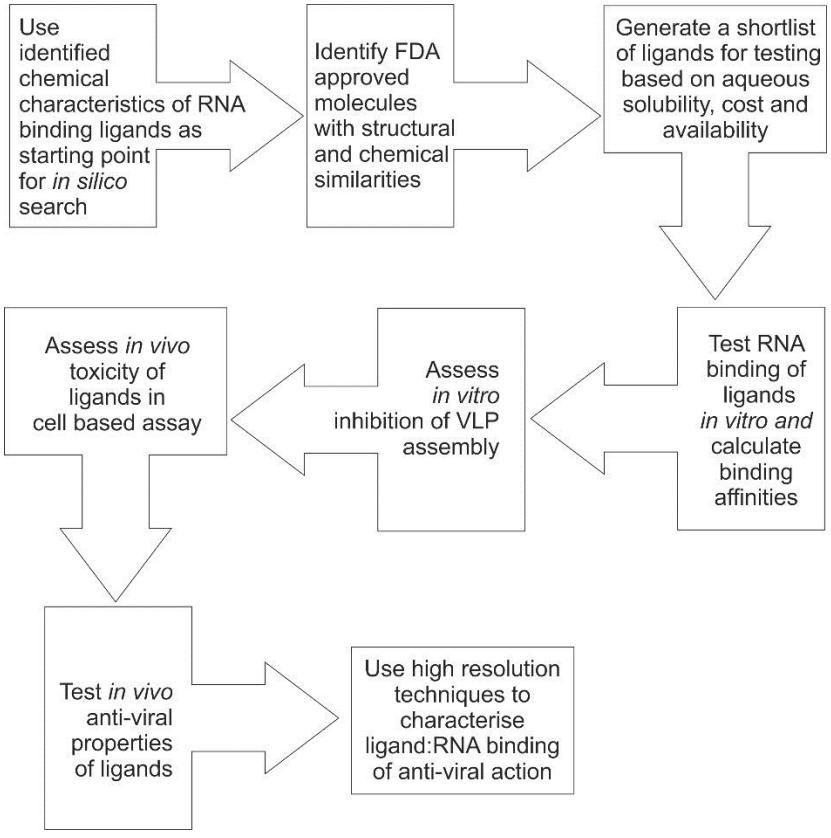


Figure 3.1 Schematic showing proposed workflow for identification and testing of anti-viral molecules.

The molecular structures of known RNA-binding ligands EDTA, spermidine, ethidium bromide, acridine orange, tetracycline, kanamycin and tobramycin as described in Section 1.6, are shown in Figure 3.2. These RNA-binding ligands (Figure 3.2) were assessed individually *in silico* using Rapid Overlay of Chemical Structures software ("ROCS" 3.2.0.4: OpenEye Scientific Software, Santa Fe, NM.) to look for trends in structural and electrostatic properties. This work was done

in collaboration with Dr Richard Foster, Medicinal Chemistry and Chemical Biology (MCCB) Technology Group, School of Chemistry at the University of Leeds.

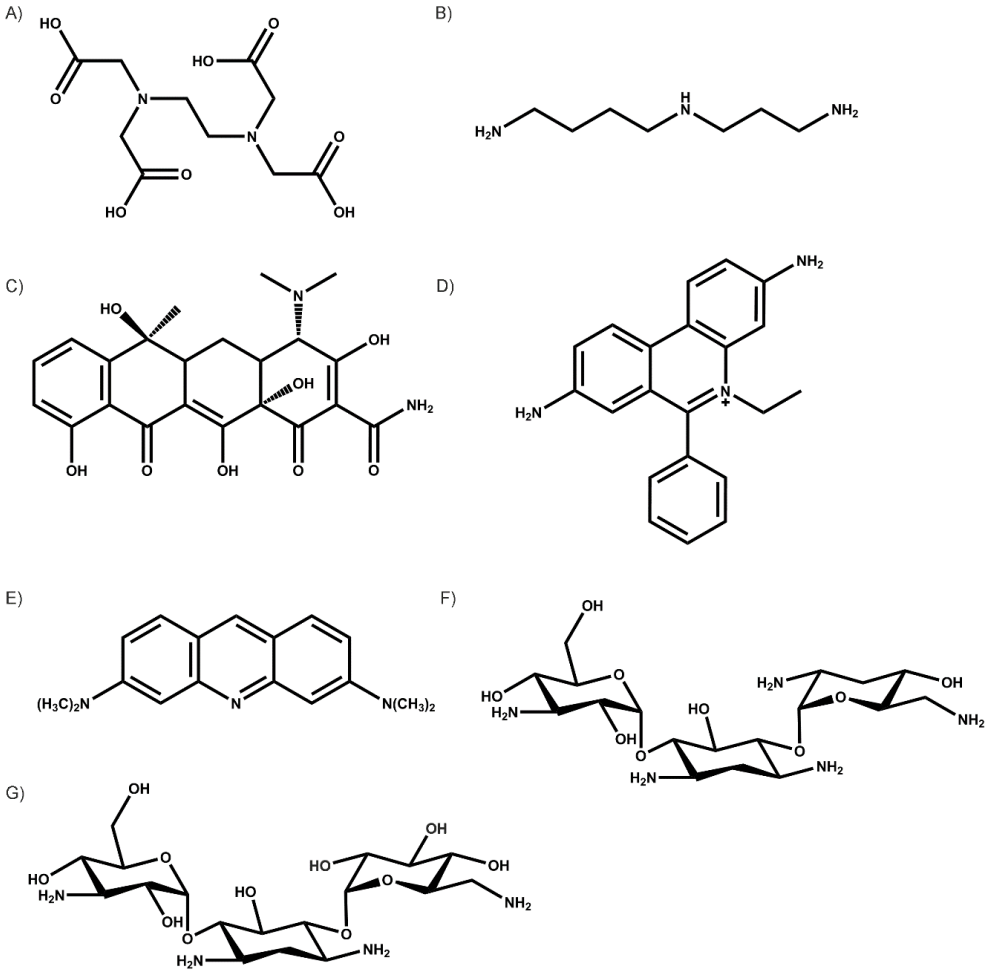


Figure 3.2 Examples of ligands shortlisted following literature review on RNA-binding molecules. A) EDTA. B) Spermidine. C) Ethidium bromide. D) Acridine Orange. E) Tetracycline. F) Kanamycin. G) Tobramycin. A) and B) Interact with RNA through electrostatic charges by aligning along the major groove of dsRNA helices. C) -E) Are molecules that use π-stacking interactions to get between planar base pair interactions. F) and G) are aminoglycosides that also bind through electrostatic charges.

ROCS is a fast shape comparison application and returns overlays based on matching both 3D shape and functional groups (Hawkins 2007). The principle is that molecules have similar shape if their overall volumes overlay well; any volume mismatch is a measure of dissimilarity. It queries shape around a grid of electron density (or arbitrary volume) reporting Tanimoto (coefficient of similarity between molecules) and Tversky (asymmetric similarity) measurements between shapes (Rogers 1960, Tanimoto 1957, Tversky 1977). A smooth Gaussian function

represents the molecular volume (Grant 1996), so it is possible to routinely minimize to the best global match. A match in chemical features can be defined from functional groups, including the parameters mentioned in Table 3.1.

Table 3.1 ROCS definitions of chemical functional groups used to facilitate the identification of compounds that are similar to the start library.

Functional Group	**Examples**
H-bond Donor	Acid hydroxyl groups
H-bond Acceptor	Carboxylate
Anion	Localised or delocalised negative charge e.g. tetrazole
Cation	Localised or delocalised positive charge e.g. guanidinium
Hydrophobic	Terminal or non-terminal aromatic or aliphatic groups e.g. phenyl
Rings	Rings of defined size e.g. 4-7 atoms

A ROCS search within the TOCRIS (chemical supplier) chemical library of 1,120 FDA-approved drug molecules checks for possible ligand matches. This *in silico* technology has the potential for further *de novo* design and has advanced to include physicochemical absorption, distribution, metabolism, excretion and toxicity (ADMET) property predictions, vital for repurposing existing drugs (Hawkins 2007).

3.1.1. ROCS Ligands Generated from *in silico* Search

ROCS generated a prioritised list of compounds from the library based on the ROCS scoring function (ROCS ComboScore). The top ranking compounds were selected for profiling, of these, 22 are available from a physical library for wet screening, gifted by the Medicinal Chemistry and Chemical Biology (MCCB) Technology Group, School of Chemistry at the University of Leeds, Figure 3.3. Five of the initial *in silico* screen were selected for experimental study based on their water solubility, cost and commercial availability (Figure 3.3). Appendix Table 8.8 summarises their important chemical and RNA-binding properties. Small quantities of the ligands were initially gifted from the MCCB Technology Group, then they were purchased on a larger scale from Sigma Aldrich®, Insight Biotechnology Ltd. and Santa Cruz Biotechnology® Inc.

A novel class of RNA-binding phenanthroindolizidines were shown to have anti-viral activity against Tobacco Mosaic Virus (Wang *et al.* 2012) during experimental design. Based on this study it was decided to include Tylophorine in the drug screening assay.

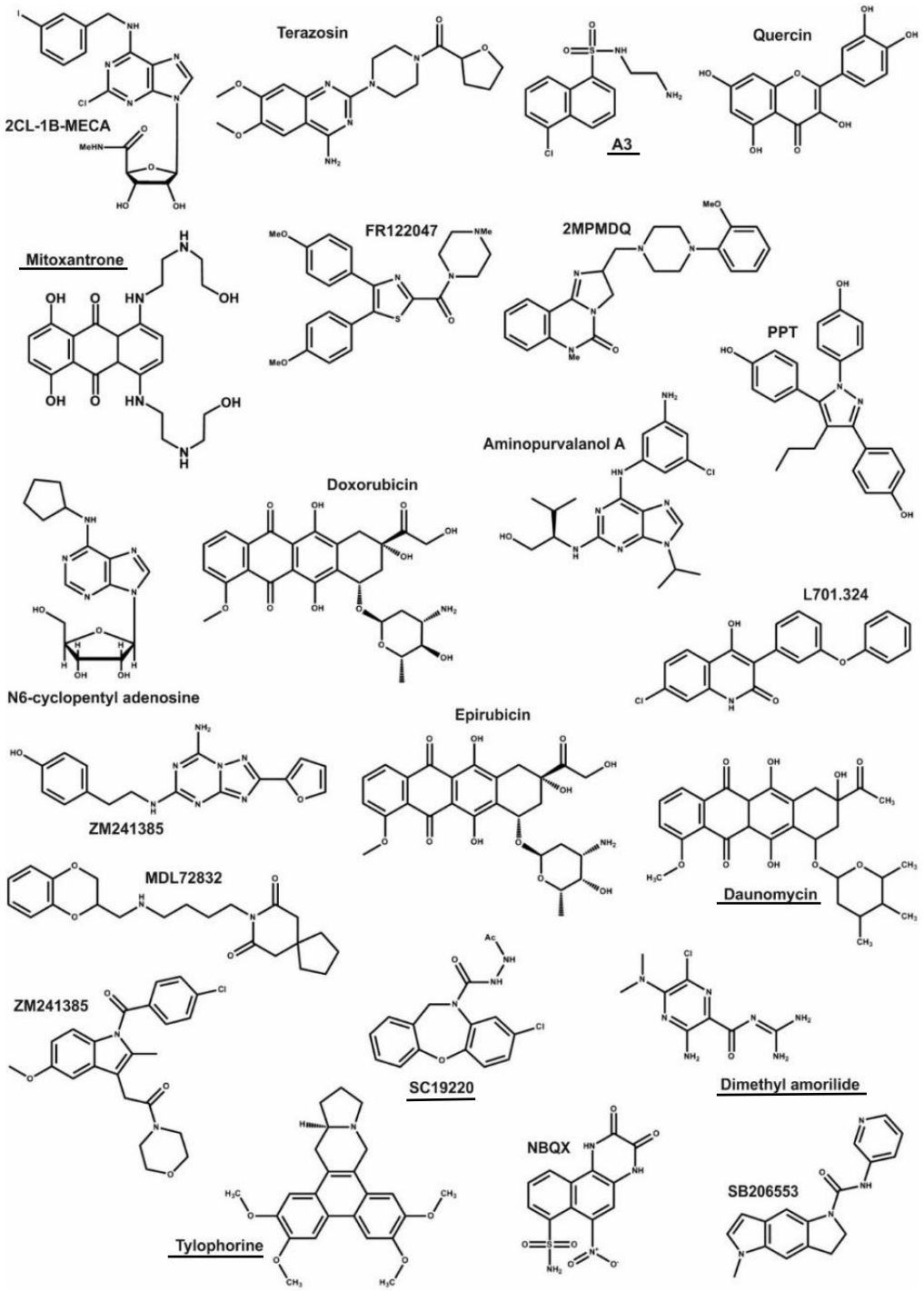


Figure 3.3 Chemical structures of the top ligands Identified by ROCS in silico screening, except for tylophorine which was added based on findings from Wang et al. (2012). Ligands used in RNA-binding assays are underlined. See Appendix Table 8.8 for information on biophysical and pharmacokinetic properties. Drawn in Chem3DPro® 13.0 (CambridgeSoft, Perkin Elmer®).

3.2. Method Optimisation for Studying RNA:Ligand Interactions

3.2.1. RNA:ligand binding by svAUC

MS2 full length and sub-section gRNA has been studied in previous svAUC experiments (Rolfsson 2010). This technique is a successful method for detecting ligand:RNA interactions, no change in RNA sedimentation properties is indicative of no ligand binding. The raw absorbance scans of TR at 260 nm, Figure 3.4 A, are fit to the Lamm Equation which calculates sedimentation coefficients (S-values) for the RNA molecules in solution (Lamm 1929, Schuck 2000). The buffer viscosity and density were calculated with the program Sednterp, 1.008 cP and 0.99834 $g.mL^{-1}$ respectively (Sednterp, Sedimentation Interpretation Program, BITC, University of New Hampshire, 2010). For all RNAs tested, a partial specific volume of 0.53 $mL.g^{-1}$ was used (Enger 1963). Experimental sedimentation coefficients can be corrected for buffer conditions by quoting extrapolated values in water at 20.0°C ($S_{20.w}$). Analysis using Sedfit (version 14.4d, Schuck NIH 2013) shows that TR sediments as a single species in solution with an $s_{20,w}$ of 1.4 ± 0.1 S, Figure 3.5 C. MS2 3' and 5' RNA fragments were also analysed. The 3' fragment is slightly larger than the 5' one (2577 versus 2469 nt) and this difference is reflected in their sedimentation coefficients, $S_{20.w}$ = 12.5 ± 0.6 and 11.6 ± 0.6, respectively, Figure 3.4 panel D. The broader distribution of the 3' fragment suggests it exists as multiple conformers in solution, agreeing with Rolfsson (2009). The negatively charged phosphate backbone of RNA molecules interacts electrostatically with cationic ligands. A svAUC experiment was set up to see if effects on secondary structures could be observed upon addition of multivalent cations, Table 3.2.

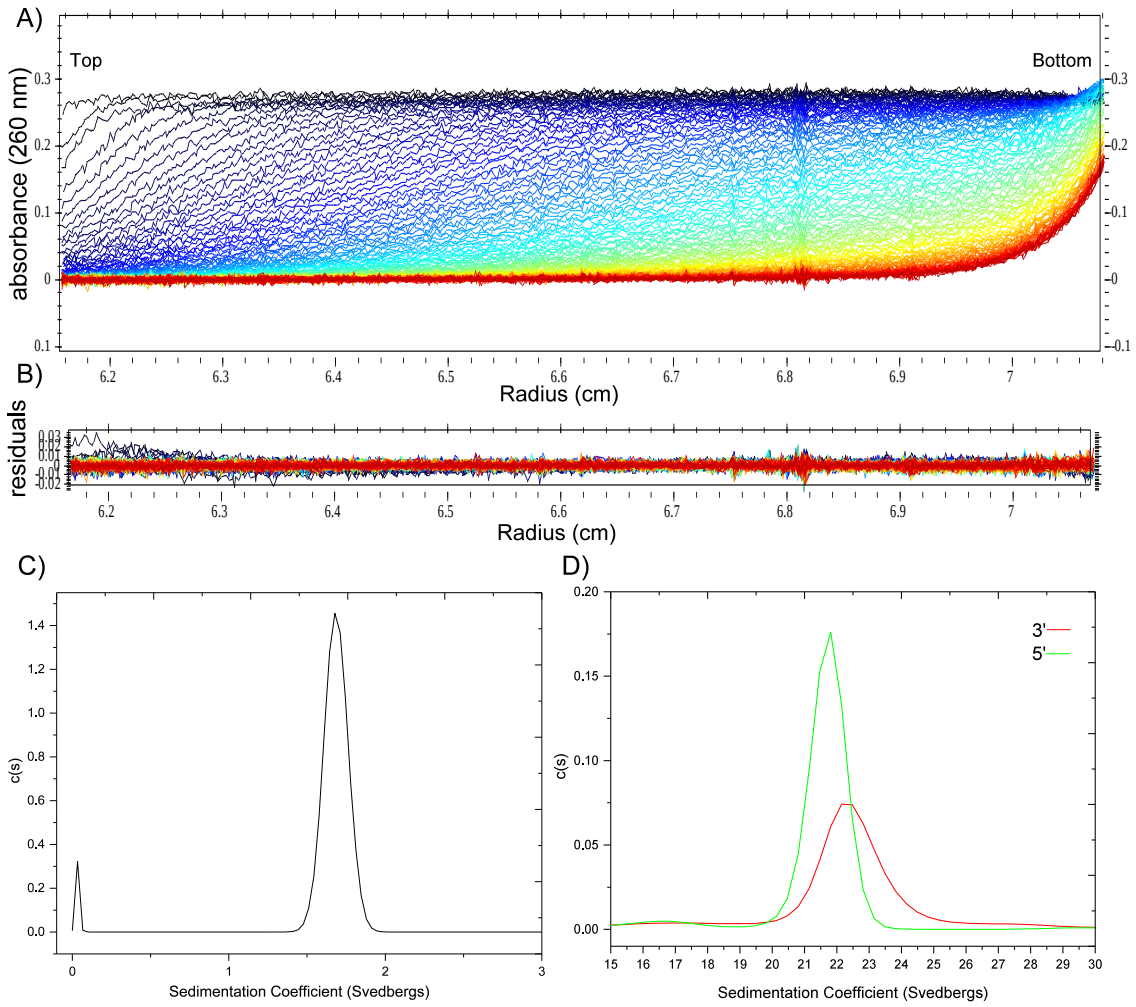


Figure 3.4 svAUC analysis of MS2 RNAs. A) 120 absorbance scans at 260 nm of raw data from TR [1 µM] experiment, scanning from the top of cell towards the outer rotor edge. The first scan is shown in dark blue through to the last in red. Data were collected over ~12 hours at 30,000 rpm with ~8 minutes between each scan. B) The residuals plot with data from A) shows a good fit to the Lamm equation, i.e. showing small deviations across the cell all equally distributed around zero. C) The sedimentation coefficient distribution plot determined by Sedfit (version 14.4d, Schuck, NIH 2013) shows TR S_{exp} = 1.7 ± 0.1, $S_{20.w}$ = 1.4 ± 0.1. The major peak is not perfectly symmetrical which could be due to the presence of different conformers in solution. The peak at ~0 S is due to a mis-match between properties of the reference/sample buffer. D) Sedimentation coefficient distribution plot for 3' and 5' MS2 RNA. 3' is the larger of the two fragments and the broader distribution suggests it exists as multiple conformers in solution. Data were collected over ~8 hours at 17,000 rpm with ~8 minutes between each scan. Raw data not shown. All experiments were run in aqueous Buffer A at 20.0°C.

Table 3.2 Hydrodynamic properties of MS2 RNAs in the presence of cations. svAUC experiment run in triplicate using conditions as described in Figure 3.5, with EDTA or spermidine (1 mM) added to Buffer A.

RNA fragment	# nt	1mM EDTA $S_{20.w}$	Buffer A $S_{20.w}$	1 mM Spermidine $S_{20.w}$
3'RNA	2577	8.9 ± 0.4	12.5 ± 0.6	13.7 ± 0.7
5'RNA	2469	9.5 ± 0.5	11.6 ± 0.6	12.3 ± 0.7
TR	19	0.93 ± 0.05	1.4 ± 0.1	3.5 ± 0.2

The addition of spermidine [1 mM] added mass due to ligand binding and increased S-values for all the RNAs tested. The addition of EDTA [1 mM] reduced S-values suggesting an increase in asymmetry of the sedimenting molecule. These results are consistent with smFCS assays by Borodavka *et al.* (2013), demonstrating the suitability of svAUC as a technique for identifying potential RNA-binding ligands. As synthesising larger RNA fragments by *in vitro* transcription used a large amount of expensive consumables; further ligand testing TR was used, as the shorter oligonucleotide is available at high purity on a micromole scale.

3.2.1.1. svAUC Analysis of TR RNA and ROCS Ligands

TR [1 µM] was incubated with the 6 shortlisted ligands from Figure 3.3 at stoichiometries detailed in Appendix Table 8.8, in Buffer A at 20.0˚C and subjected to svAUC. Sedimentation coefficients were calculated using a continuous distribution c(s) Lamm equation model (Schuck 2000), as described in Section 2.7. In order to reference between different experiments sedimentation values are adjusted to the standard density and viscosity of water, at 20.0 ˚C ($S_{20.w}$). Experimental differences were calculated for TR control, $S_{20.w}$ = 1.36 ± 0.45. All ligands were made up in diethylpyrocarbonate-treated water [1 mM], apart from tylophorine which was prepared as a stock in DMSO [1 mM]. All were then diluted to working concentrations with Buffer A. The results are shown in Figure 3.5 and Table 3.3. Examples of raw data analysis from the svAUC experiments is in Appendix Figure 8.1.

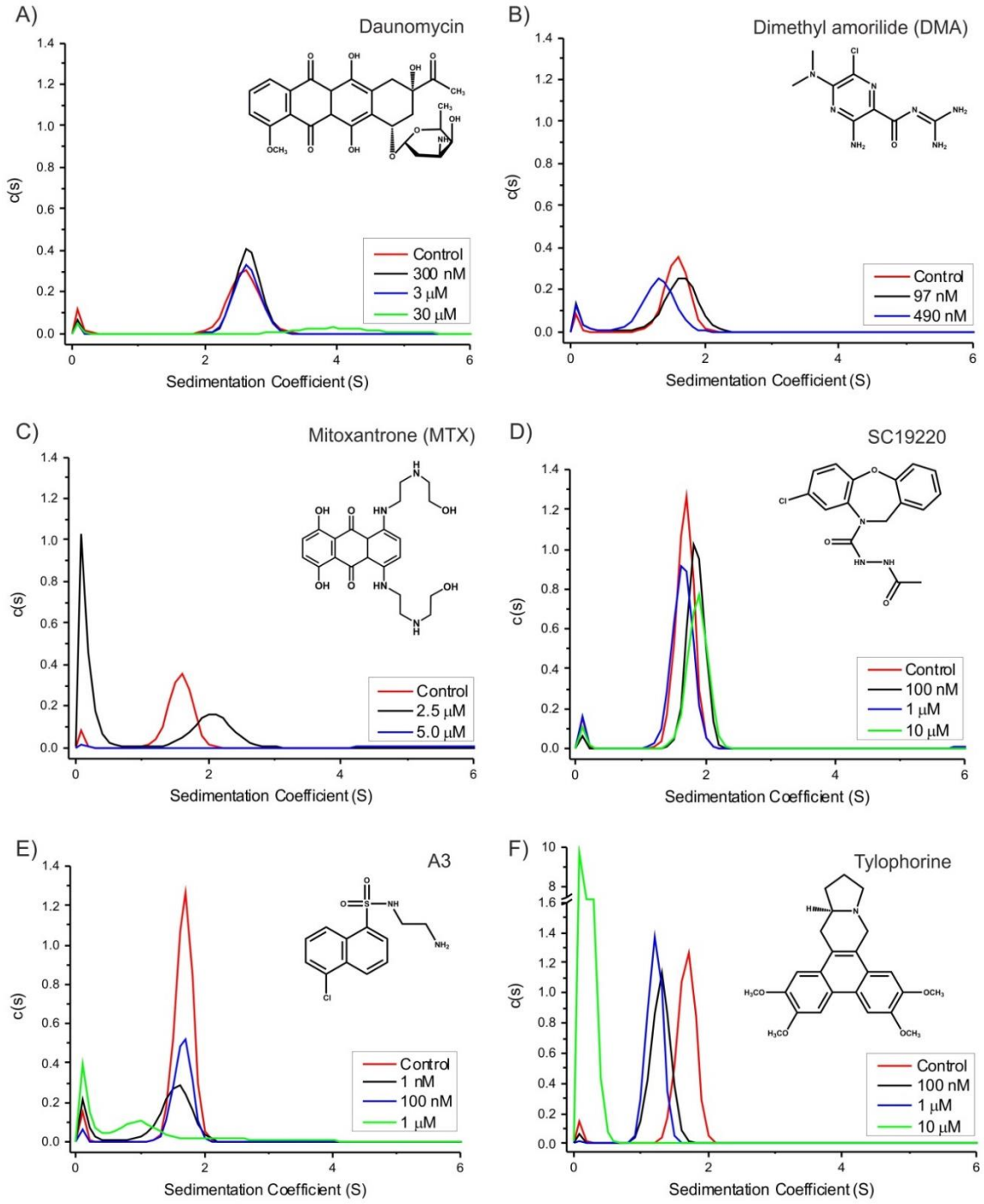


Figure 3.5 svAUC of MS2 TR RNA incubated with shortlisted ligands from ROCS screening. Ligands are A) Daunomycin, B) Dimethyl amorilide (DMA), C) Mitoxantrone (MTX), D) SC19220, E) A3 and F) Tylophorine. Experiment as described previously in Figure 3.4. Ligand:RNA ratios vary based on known binding affinities and stoichiometries of published data (see Appendix Table 8.8). Inset shows chemical structure of each ligand.

*Table 3.3 Effects of small ligands on the sedimentation properties of TR. All experiments carried out in triplicate as per the experimental method detailed in Figure 3.4. *Multiple peaks were shown at highest concentration tested.*

Ligand	Ligand [μM]	Sedimentation Coefficient of major species in solution $S_{20.w}$
Daunomycin	0	2.21 ± 0.13
	0.3	2.25 ± 0.13
	3	2.24 ± 0.13
	30	3.39 ± 0.20
DMA	0	1.34 ± 0.08
	0.097	1.40 ± 0.08
	0.49	1.12 ± 0.07
MTX	0	1.34 ±0.08
	2.5	1.75 ± 0.10
	5	RNA aggregation
SC19220	0	0.97 ± 0.08
	0.1	1.07 ± 0.09
	1	0.95 ± 0.08
	10	1.08 ± 0.09
A3	0	1.34 ± 0.08
	0.01	0.91 ± 0.08
	0.1	0.96 ± 0.08
	1	0.64 ± 0.05, 1.89 ± 0.11*
Tylophorine	0	0.97 ± 0.08
	0.1	0.78 ± 0.07
	1	0.70 ± 0.06
	10	0.25 ± 0.03

In the presence of low concentrations of daunomycin (1:0.3 and 1:3) there was very little shift in the sedimentation properties of TR. However at high ligand excess (RNA:daunomycin 1:30) the 260nm absorbance peak disappears suggesting precipitation of the RNA in solution, Figure 3.5 A, green trace.

Similar analysis in the presence of DMA shows broadening of the TR peak with increasing amounts of ligand, Figure 3.5 B. Even though DMA was included in the reference buffer, at

concentrations above 490 nM DMA, the absorbance properties of the free ligand obscured the absorbance scans and the data could not be used. The shift in TR S-value from $S_{20.w}$ = 1.4 to 1.1 at ~2:1 RNA:DMA suggests a change in RNA secondary structure.

At a 1:2.5 ratio RNA:MTX there is an increase in S-value of TR, Figure 3.5 C. However, at RNA:MTX ratio of 1:5 the peak corresponding to RNA almost disappears. At high concentrations of MTX the absorbance properties of the free ligand obscures the absorbance scans which is the reason for a reduction in peak intensity. The increase in S-value to $S_{20.w}$ ~4 could be due to MTX:TR multimerisation as MTX multimerisation is well documented above 1 µM (White *et al.* 1985).

SC19220 is predicted to bind RNA via electrostatics. At varying ratios of RNA:SC19220 there was no effect on the S-value suggesting that this ligand does not bind to TR stably under these conditions, Figure 3.5 D.

There is a decrease in S-value of TR when titrated with ligand A3 at 1:1 ratio, the peak at 0.1 S increases with A3 concentration. Negative mode electrospray ionisation mass spectrometry (ESI-MS) showed evidence for A3 dimer which would explain this peak, Figure 3.5 E. ESI-MS data of individual ligands is shown in Appendix Figure 8.7.1.

On increasing the relative amount of tylophorine the S-value of the RNA is reduced, suggesting RNA-binding and unfolding, Figure 3.5 F. svAUC highlights TR binding ligands from the *in silico* search, however for information on binding affinities a more sensitive spectroscopic approach is needed, one that is not complicated by absorbance properties of ligands.

3.2.2. Assessing ligand binding using CD

As RNA melts changes can be seen in the CD spectrum at λ 260 and 280nm, corresponding to unfolding of A-U and G-C bp, respectively (Jaeger 1990). If ligand binding causes RNA denaturation or stabilisation, then a change in CD signal would be expected. The thermal unfolding of TR monitored by CD, is shown in Figure 3.6 A. At 20.0°C the TR spectrum shows λ_{max}

267 nm, characteristic of dsRNA (Sosnick and Pan 2003). Raw CD data can be found in Appendix Figure 8.2. The CD signal change at 260nm (A-U bp) and 280nm (G-C bp) is used to calculate T_m 50.5°C, and 68.4°C respectively, Figure 3.6 B. It is expected that G-C bp melt at higher temperatures due to being stabilised by an extra hydrogen bond. The change in absorbance signal with increasing temperature shows hypochromicity at 260 and 280 nm, Figure 3.6 C. This confirms that TR denatures as it is heated. (Raw absorbance data can be found in Appendix Figure 8.3). This result correlates with previous data collected for heat denaturation of TR, T_m = 56.9°C (Parrot *et al.* 2000).

MTX bound to TR in the svAUC assay so it was decided to focus on collecting CD data for MTX alone. Promising results were shown by svAUC for tylophorine binding to TR, but as this ligand is only soluble in DMSO it was left out of this experiment. CD spectra were collected for TR RNA in the presence of increasing concentrations of MTX. Figure 3.7 panel A shows CD data of MTX only at concentrations 2.5, 5 and 10 µM, corrected for Buffer A. The CD spectrum of 5 µM MTX was lower than 10 µM, this result cannot be explained as the samples were (visibly) more dilute. Figure 3.7 panel B shows CD data of TR and MTX and Figure 3.7 panel C shows the same scans converted to molar ellipticity, all scans are corrected for ligand CD at the equivalent ligand concentration. Increasing the molar ratio of MTX:TR in solution shows a reduction in CD peak maxima relating to RNA (λ_{max} = 267 nm), which supports the idea that MTX binds to TR. Data suggests that an increase in TR concentration is needed in order to reduce noise from spectroscopic properties of the ligand, using more RNA makes this experimental method too expensive.

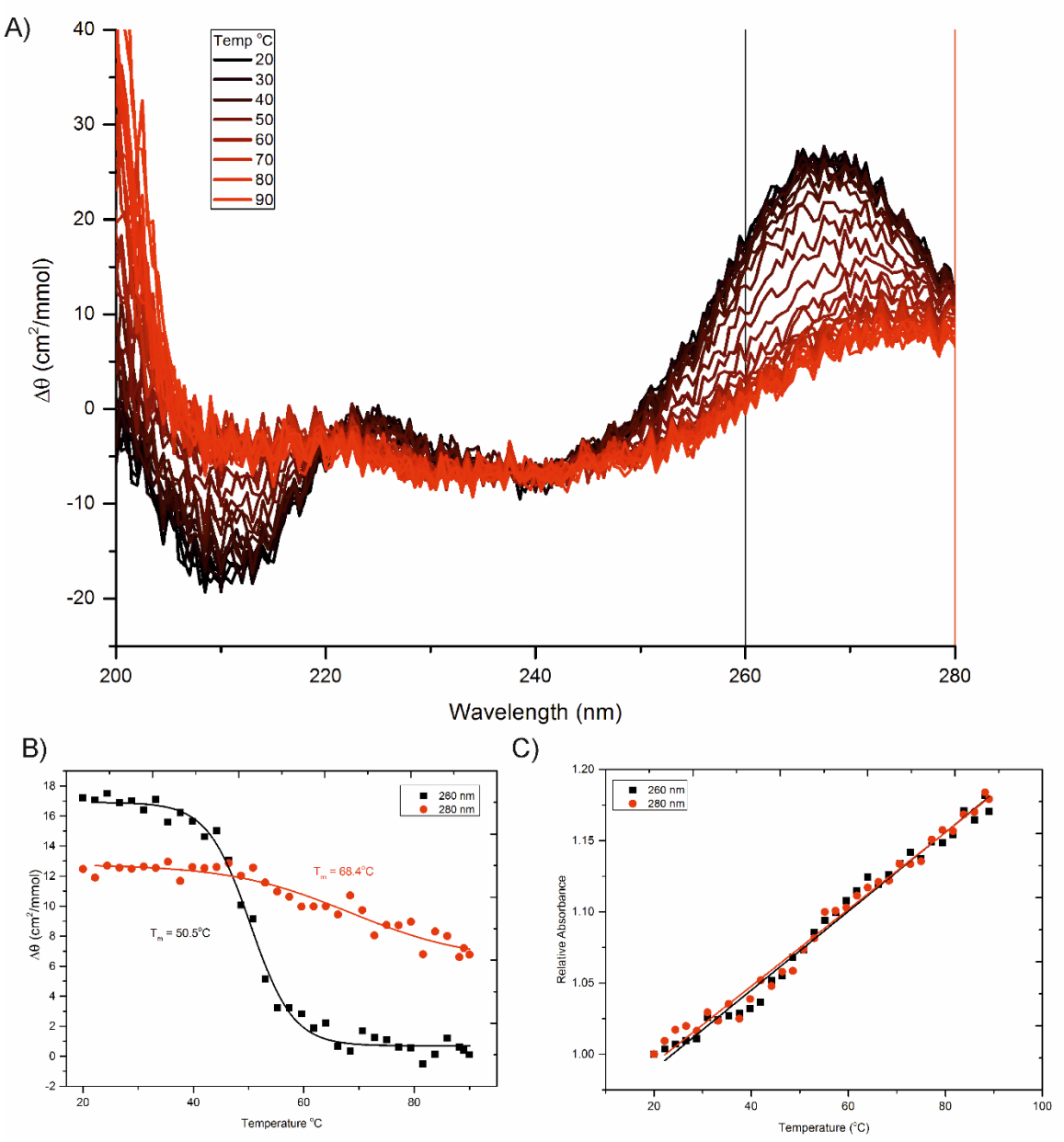


Figure 3.6 Thermal unfolding of TR RNA monitored by CD. A) The CD spectrum of TR in Buffer A. Scanning at 0.5 nm intervals with a heat ramp of 1.0 °C.minute-1, spectra were recorded from 20 °C (black) to 90 °C (red). Increasing the temperature generates a change in the CD signal denoting a change in the secondary structure of RNA. B) The change in molar ellipticity at 260 nm (black squares) and 280 nm (red circles) allows for calculation of Tm. C) Hypochromicity at 260 nm (black squares) and 280 nm (red circles) shows that with increasing temperature TR unfolds as it is heated. Raw CD data can be found in Appendix 8.2. Raw absorbance data can be found in Appendix 8.3.

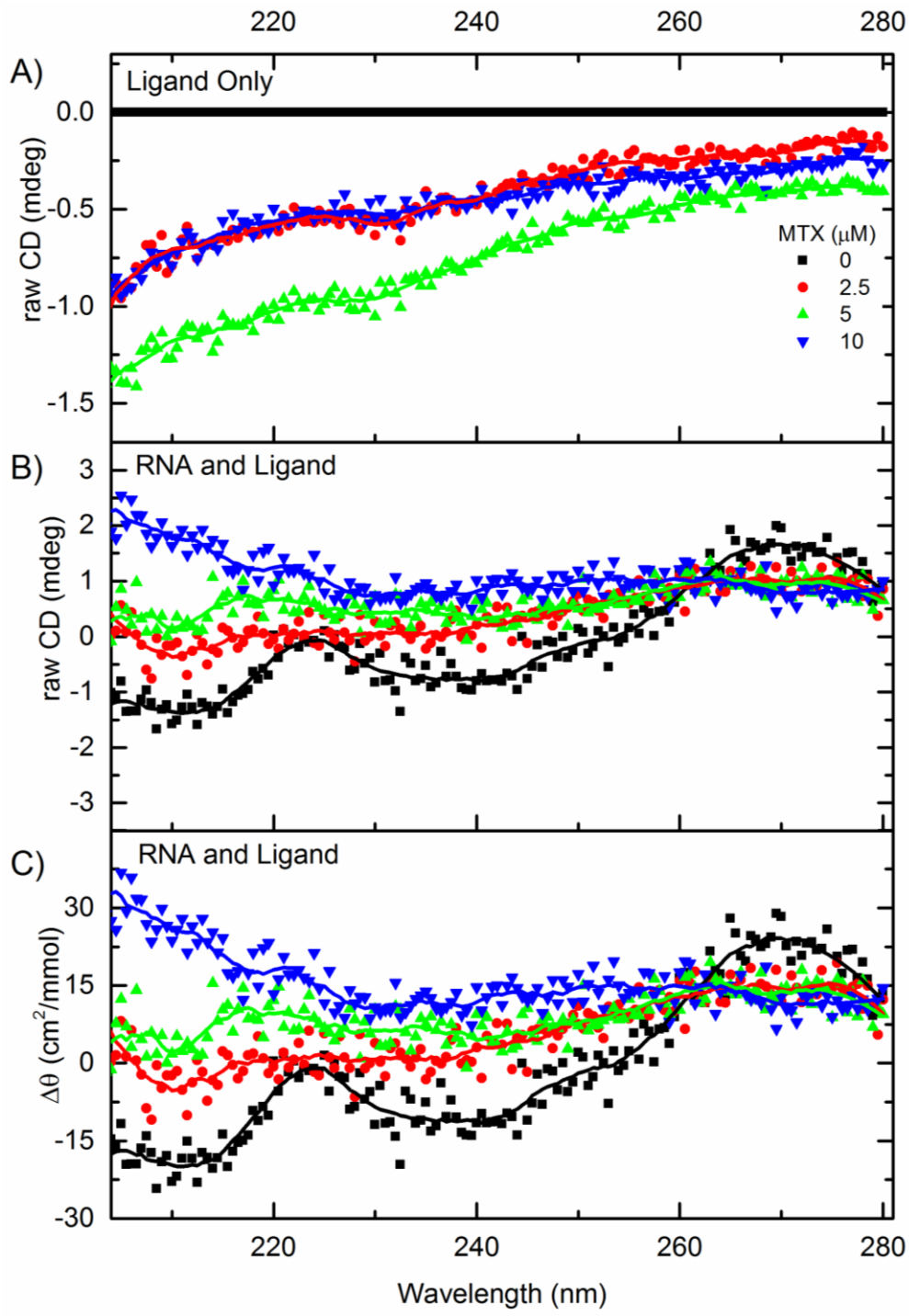


Figure 3.7 CD Spectra showing the effect of MTX on TR RNA. A) Raw CD of MTX at 0 (black), 2.5 (red), 5 (green) and 10 µM (blue), corrected for Buffer A. B) Raw CD of TR [1.1 µM] with increasing MTX. Corrected for MTX. C) Spectrum as in B) converted to molar ellipticity. An increase in MTX concentration shows a decrease in peak maxima at 270nm correlating to RNA secondary structure. Scanning stepwise by 0.5 nm, 205-280 nm, data recorded at 20°C.

3.2.3. Fluorescence Spectroscopy of MTX in the presence and absence of TR

MTX is fluorescent, excitation at 633 nm gives a λ_{max} emission that is concentration dependent as shown in Figure 3.8 (Mirela Enache 2010). MTX fluorescence emission increases with concentration, until approximately 11.2 µM (orange). Above this the emission spectrum is distorted by an inner filter effect (IFE), whereby the absorption of the sample reduces emission and shifts λ_{max} to a longer wavelength (Parker and Barnes, 1957). This IFE means that it would not be suitable to monitor fluorescence emission when studying MTX:TR.

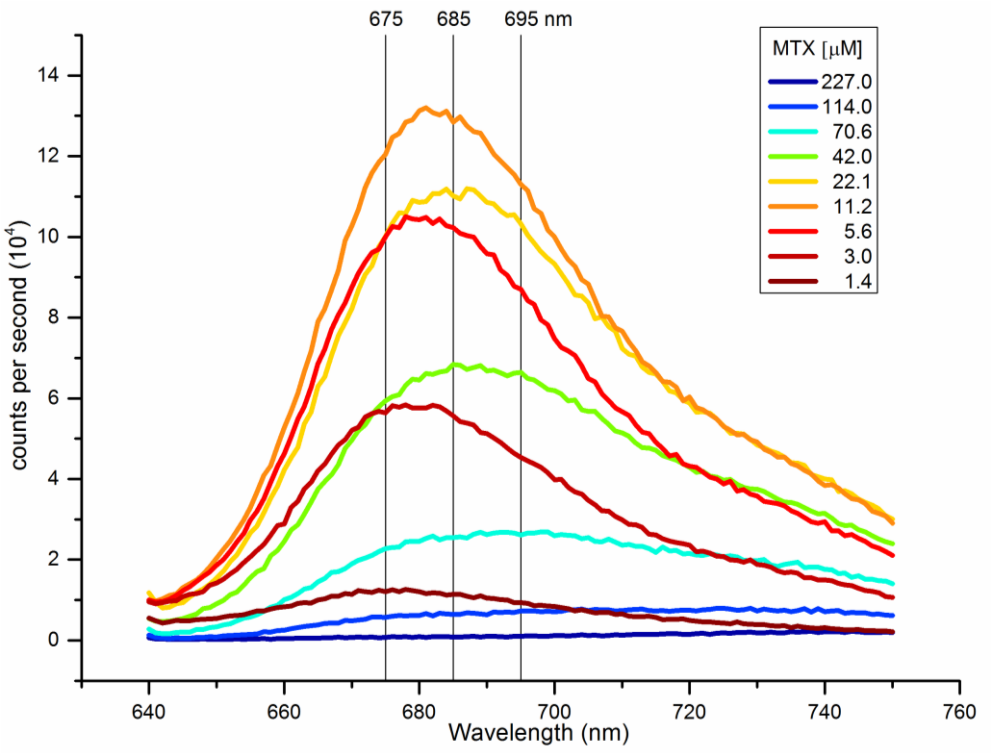


Figure 3.8 Fluorescent emission of MTX at different concentration. Excitation at 633 nm shows λ_{max} shifts from 675 to > 695 nm with increasing MTX concentration. Data collected at 20.0°C, 1 mm quartz cuvette, 1 nm stepwise in Buffer A.

3.2.4. UV-Visible Absorbance Spectroscopy of MTX +/- TR

The interaction of MTX with DNA causes a hypochromic and bathochromic (red) shift in the absorption λ_{max} of MTX. The binding of MTX to DNA oligonucleotides has been well characterised and MTX has a µM affinity through intercalation into base paired duplexes (Li *et al.* 2005). MTX alone shows multiple absorbance peak maxima, Figure 3.9. Characteristic λ_{max} values of MTX are shown at 242, 276, 610 and 660 nm (Kapuscinski 1981). There is an increase in absorbance

at all wavelengths upon increasing MTX concentration. λ_{max} at 660 and 610 nm correlate to monomer and dimer respectively (Enache and Volanski 2010). MTX's monomer concentrations were calculated using ε_{682nm} = 8,360 $M^{-1}.cm^{-1}$ (Kapuscinski 1981). Kapuscinski *et al.* determined concentrations below micromolar are primarily monomeric (Enache and Volanski 2010).

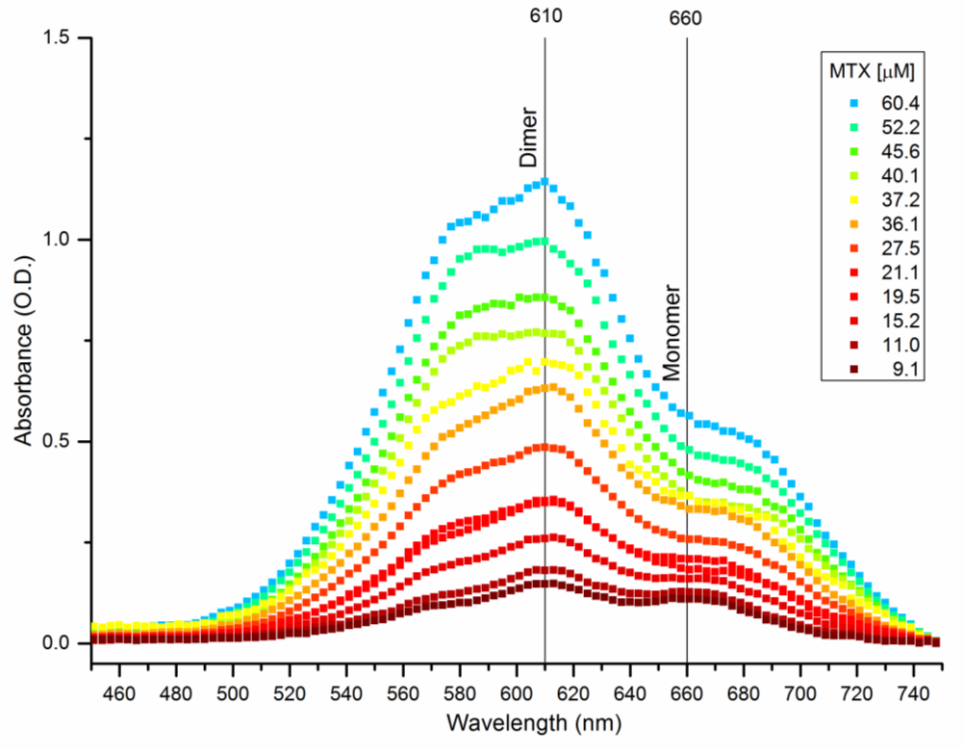


Figure 3.9 Absorbance spectra of MTX dilutions. Data collected at 20°C, 1 cm quartz cuvette, scanning 1 nm stepwise in Buffer A.

3.2.4.1 Titrating TR with MTX

The absorbance spectra of MTX in the presence and absence of TR are shown in Figure 3.10. The bathochromic shift of the peak maximum upon addition of TR RNA (purple) compared to MTX alone (black) is clear. There is a large drop in absorbance at both 610 and 660 nm, the reason behind this sudden step is unclear.

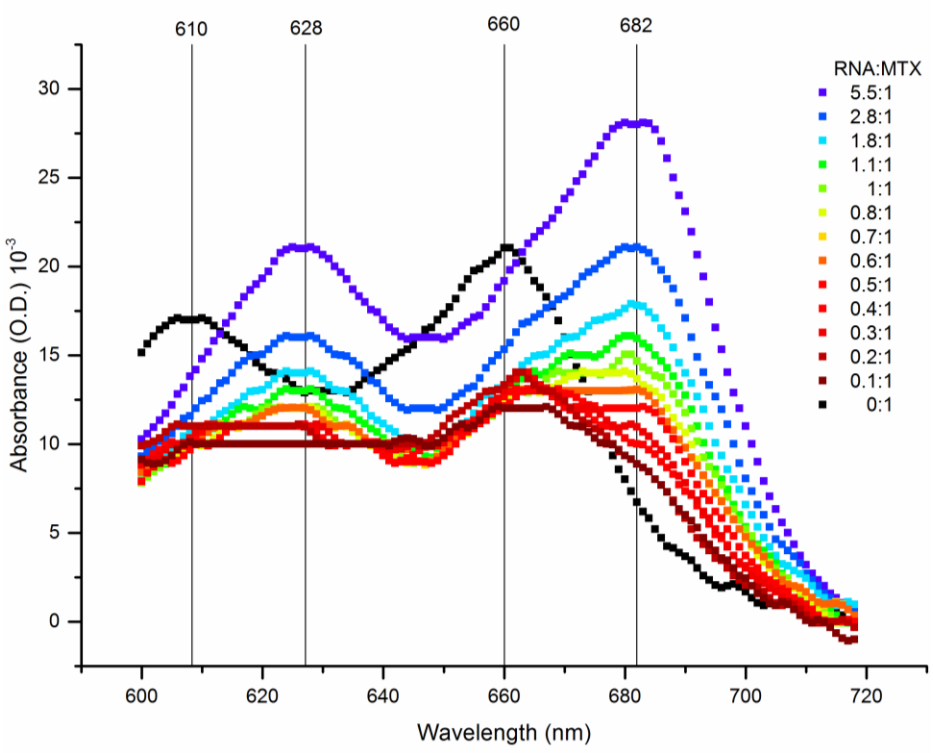


Figure 3.10 Absorbance spectra showing MTX titrated with TR. As the molar ratio of RNA decreases λmax shifts from 660 to 682 nm. MTX only (black). Data collected at 20 °C, 1 cm quartz cuvette, scanning 1 nm stepwise in Buffer A.

Using the Benesi-Hildebrand Equation (Equation 1) the bathochromic shift in λ_{682nm} in UV-Vis spectra upon RNA-binding can be used to calculate a binding constant, K (Benesi 1949).

$$\frac{A_0}{A-A_0} = \frac{\varepsilon_G}{\varepsilon_{H-G}-\varepsilon_G} + \frac{\varepsilon_G}{\varepsilon_{H-G}}\frac{1}{K[RNA]} \qquad \text{Equation (1)}$$

A_0 = absorbance of drug in the absence of RNA

A = absorbance of drug in presence of RNA

ε_G and $_{H-G}$ are the absorption coefficients of drug and its complex with DNA.

The plot of $A_0/(A-A_0)$ versus 1/[RNA] was constructed using the data from the absorbance titrations and a linear fitting of the data gave the binding constant. Figure 3.11 shows the raw data collected at λ_{682nm} and the subsequent Benesi-Hildebrand plot. The interaction of TR with MTX from the slope of the graph; $K_D = 3.71 \pm 0.22 \times 10^{-7}$ M. Previously cited work calculates that anthracycline molecules interact with dsDNA at $K_D \approx 10^{-5} – 10^{-6}$ M (Berg 1981, Ibrahim 2001). Examples with RNA in aqueous solution have not previously been quantitated.

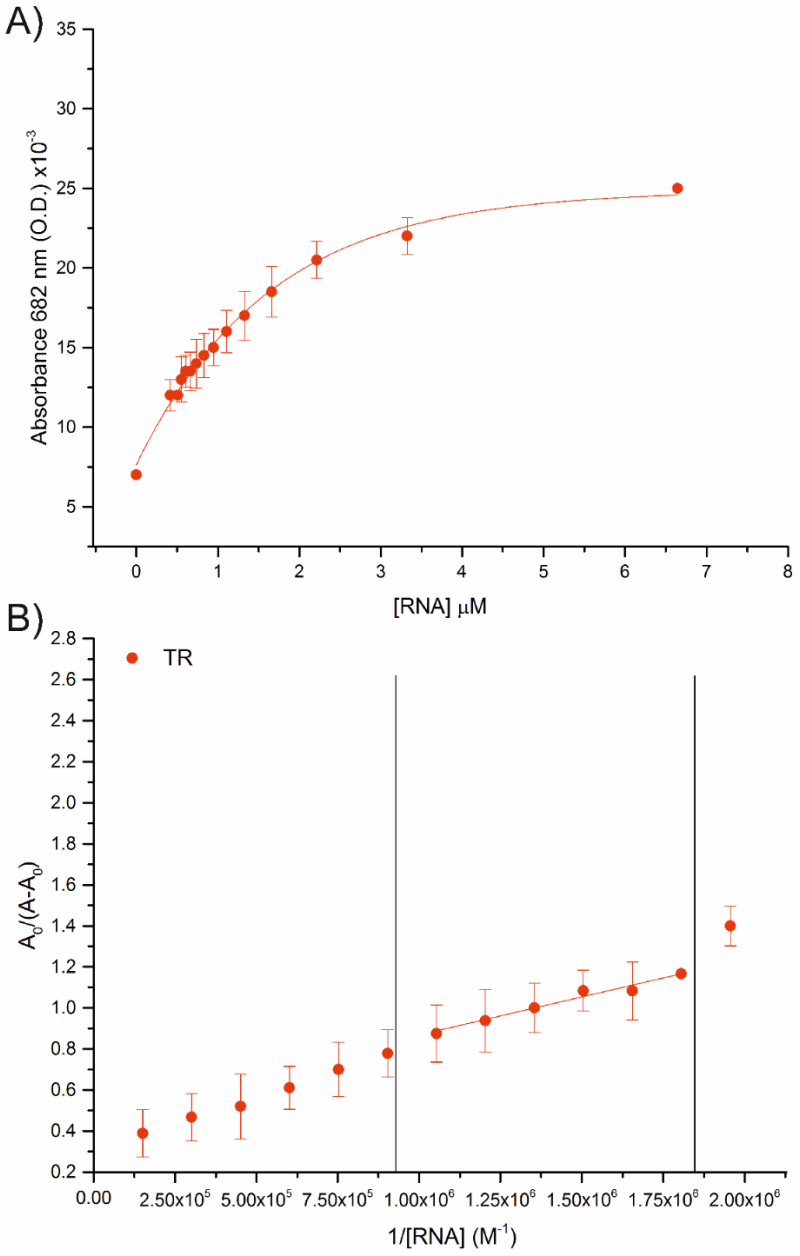


Figure 3.11 Determining the affinity of MTX for TR. A) Measuring A682nm of MTX upon TR titration. B) Plot showing $A_0/A-A_0$. A_0 = absorbance at 682nm of MTX only, A = absorbance at 682nm of MTX:TR complex. Calculation assumes 1:1 interaction therefore only data points within the black lines are used to calculate the K_D. Using a linear fit K_D = 3.71 ± 0.22 x 10^{-7} M, analysis done in OriginPro 9.1.0 (OriginLab© 2014).

3.2.4.2 Titrating TR-10A with MTX

MTX has been reported to bind to adenosine bulges in RNA stem loops (Zheng 2009), this would be consistent with the structure of TR. A TR RNA variant, TR-10A, was therefore produced to see if the removal of this base reduced the binding affinity (Figure 3.12).

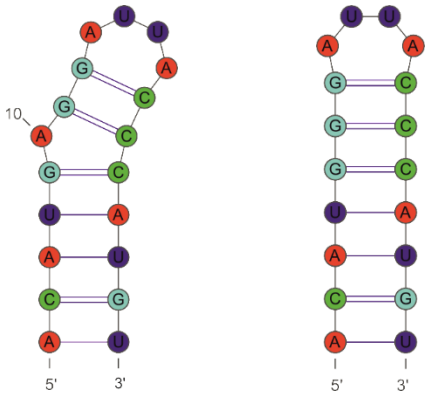


Figure 3.12 Secondary structures of TR (Valegård et al. 1994) and TR-10A (Helgstrand et al. 2002). Drawn in Varna (Darty 2009).

Its ability to bind MTX was assayed in the same way, TR-10A was titrated with MTX. The results showed that increasing the molar ratio of TR-10A in solution increases λ_{682nm} (Figure 3.14). Comparing TR and TR-10A binding to MTX using the Benesi-Hildebrand Equation shows that TR-10A binds MTX with a 3-fold lower affinity than TR, K_D = 1.11 ± 0.17 x 10^{-6} M versus 3.71 ± 0.22 x 10^{-7} M, Figure 3.13. This suggests that the adenosine bulge contributes to MTX binding. Note, using absorbance spectroscopy does not show the stoichiometry of the interaction and there is a possibility that multiple molecules of MTX bind to RNA.

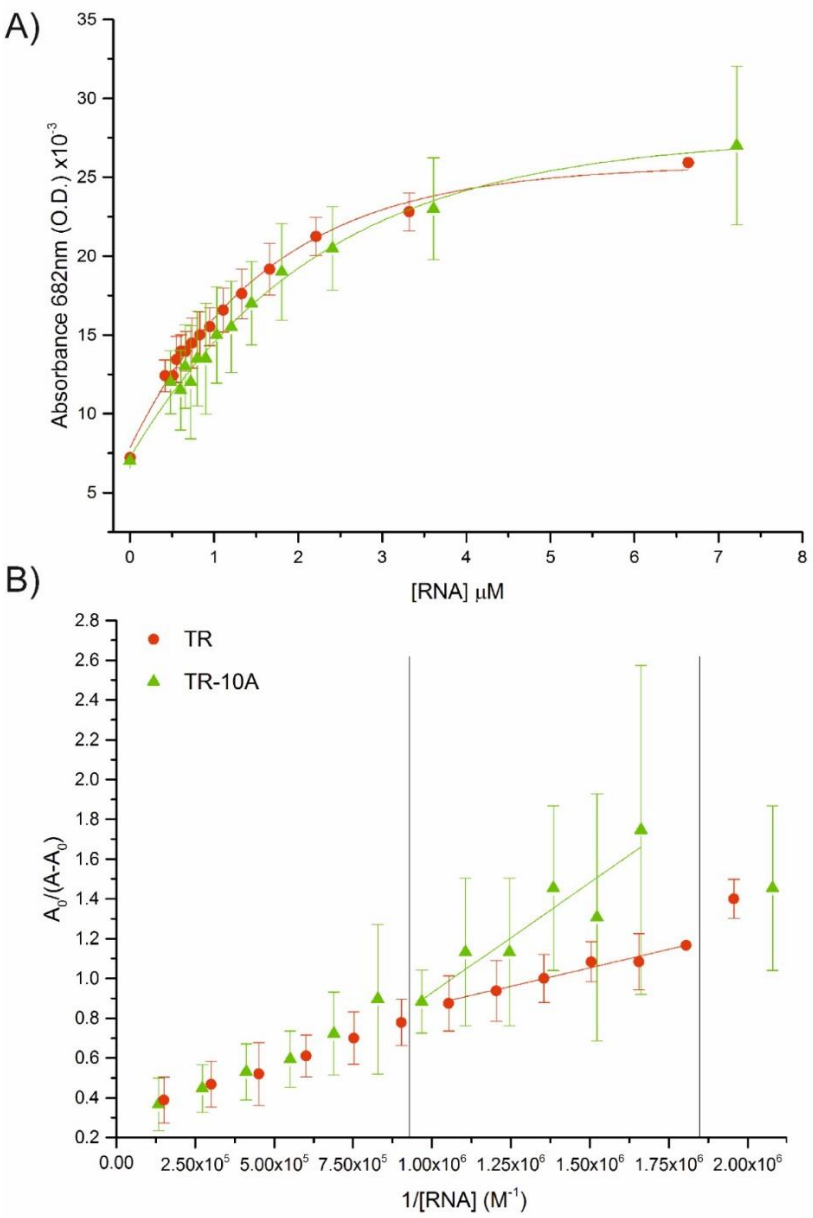


Figure 3.13 Comparing the affinity of MTX to TR (red circles) and TR-10A (green triangles). A) Plotting the A682nm of MTX upon RNA titration. Increasing the amount of RNA in solution increases λ682nm, suggesting binding. B) Plot showing $A_0/A\text{-}A_0$ versus 1/RNA. A_0 = absorbance at 682nm of MTX only, A = absorbance at 682nm of MTX:TR complex. Calculation assumes 1:1 interaction therefore only data points within the black lines are used to calculate the K_D. Using a linear fit $K_D = 1.11 \pm 0.17 \times 10^{-6}$ M, analysis done in OriginPro 9.1.0 (OriginLab© 2014).

3.2.5. Mass Spectrometry of RNAs in the presence and absence of MTX

To determine the stoichiometry of MTX bound to TR, this thesis predominately uses ESI-MS to generate multiple charge states of the molecules in question. It is more sensitive than MALDI and it can be run in positive or negative ion mode (charged by the addition or removal of proton(s)). TR binding to MTX was analysed using negative ion mode. The mass to charge ratios (*m/z*) of samples in a titration experiment revealed their binding stoichiometry and relative binding affinities. Note, the MS assays require a higher RNA and ligand concentration than the previous assays described above, as well as a volatile buffer. The experiments used ammonium acetate at pH 6.5 (MS Buffer) as used previously in many MS2 reassembly MS experiments, including those analysed by ESI-MS (Morton 2010, Morton 2008, Knapman 2010). This work was done in collaboration with Henry Fisher (Fisher 2014, PhD Thesis). There are different charge states generated when TR is ionised in negative ion mode, Figure 3.14. The major peak is $[TR]^{4-}$, the result of removing four phosphodiester protons. These data were then transformed to the molecular mass scale using the maximum entropy processing technique (Skilling, 1989), yielding an averaged mass for free TR RNA of 6045.49 ±0.06 Da. The expected mass from the nucleotide sequence is 6044.70 Da (OligoAnalyzer 3.1, Integrated DNA Technologies).

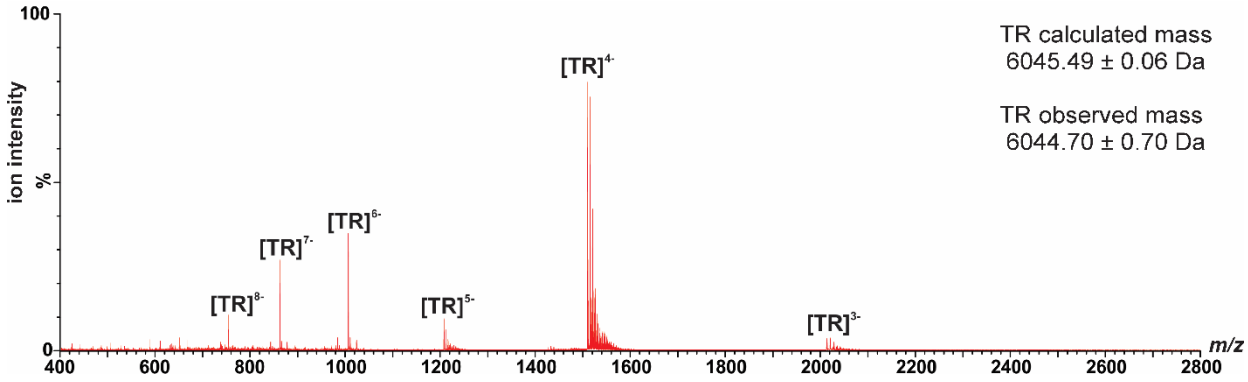


Figure 3.14 Mass spectrum of TR [14 µM] in 200 mM AmAc, pH 6.5.

Other ions were also detected, these were consistent with fragments of the parent MTX ion. Possible structures for the fragments are identified and are shown, Figure 3.15 A-E.

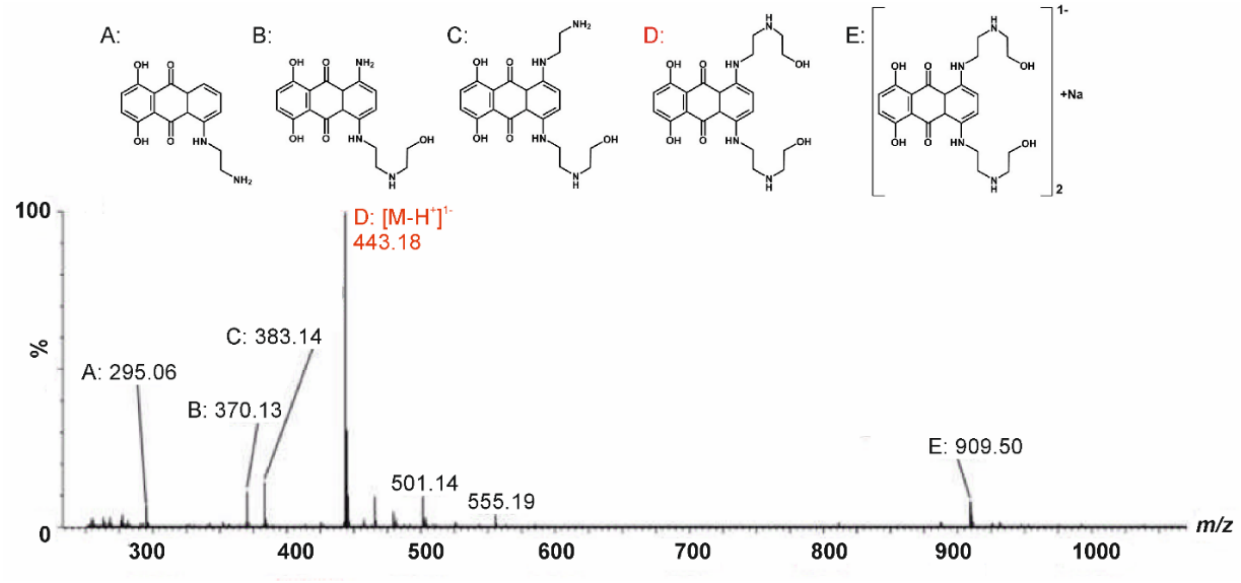


Figure 3.15 Negative ion MS of MTX. A- E denotes structures of fragments identified. D (red label) is MTX monomer detected at m/z 443.18.

The binding of MTX to TR was analysed by titration of the small ligand into a fixed concentration of TR and followed by MS under the same conditions, Figure 3.16.

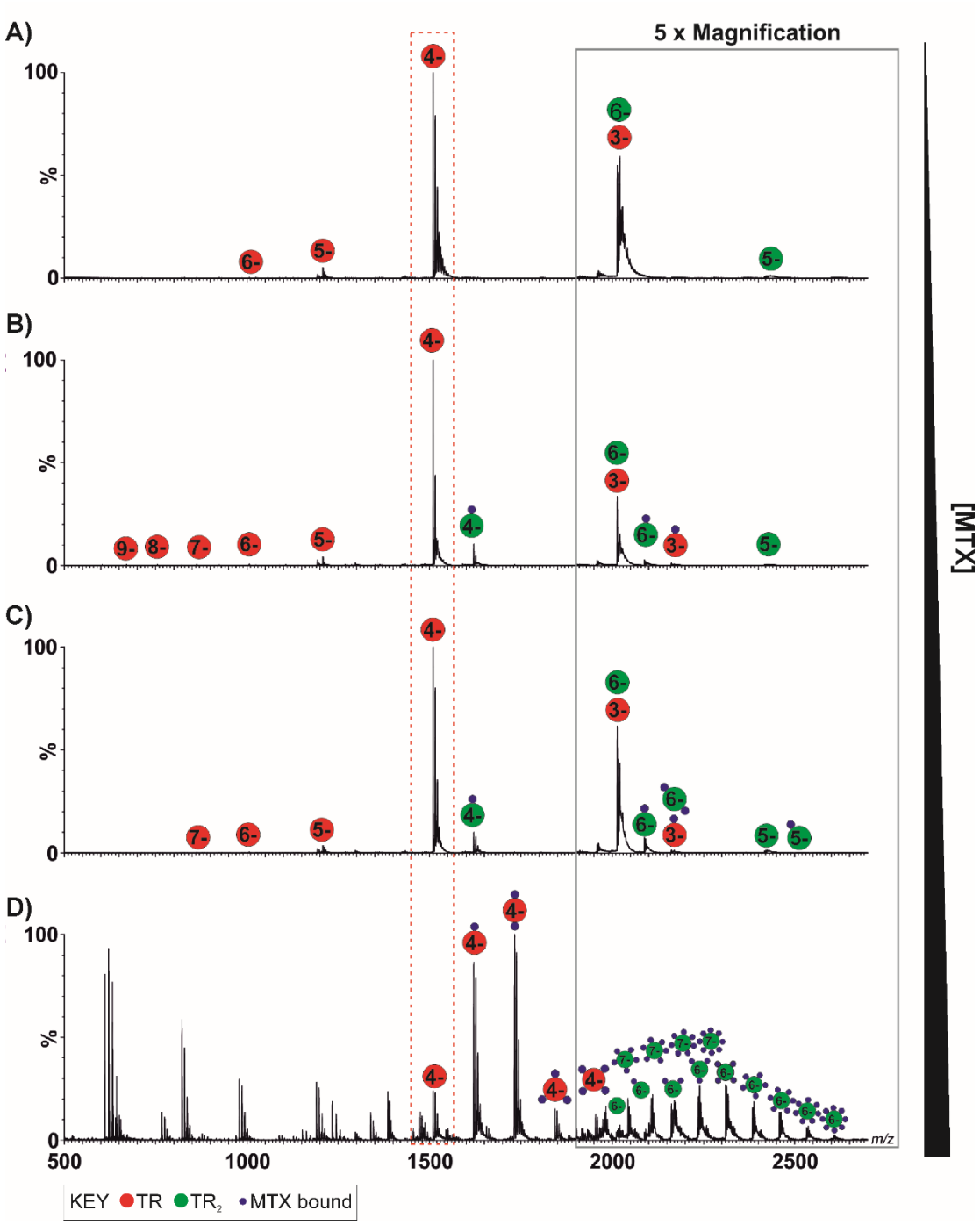


Figure 3.16 Negative Ion mode MS of MTX titrated into TR. Mass spectra of TR [7.5 µM] in MS Buffer in the presence of MTX at different molar ratios TR:MTX A) 50:1, B) 10:1, C) 5:1, and D) 1:1. Labels above peaks show TR monomer (red), TR dimer (green) and number of MTX bound (purple). Red dashed box highlights the ligand-free 4- charge state.

From this titration series, complexes of TR:MTX can be detected using negative ion MS. Using MassLynx™ software the mass spectra are smoothed using a Savitzky-Golay filter to optimize the signal-to-noise-ratio, the peaks of interest are integrated by measuring the centre of the

peak at half peak height. By comparing the integration data of the major species detected, monomer (4-) ion, it can be deduced that increasing MTX in solution increases the number of MTX bound per RNA, Table 3.6. Most notably at a 1:1 molar ratio up to four MTX molecules are bound.

Table 3.4 Comparing (4-) Ion intensities of monomeric TR with MTX. Integrated peak heights ion counts measured using MassLynx™, % shown in brackets.

TR:MTX	Free RNA	+1 MTX	+2 MTX	+3 MTX	+4 MTX	TOTAL IONS (%)
1:0.02	587,138 (100)	0	0	0	0	587,138 (100)
1:0.1	1,256,798 (90.9)	125,194 (9.1)	0	0	0	1,381,992 (100)
1:0.2	203,450 (91.2)	194,340 (8.3)	0	0	0	397,790 (100)
1:1	72,390 (10.5)	254,636 (37.1)	303,230 (44.2)	47,602 (6.9)	8,682 (1.3)	686,540 (100)

At the highest concentration of MTX multiple peaks appear in the range <1500 *m/z*, Figure 3.16 D. These can be attributed to contaminants such as plasticisers leaching from lab consumables (Prof. Alison E. Ashcroft, Personal Communication). Due to the high complementary of its secondary structure it is not uncommon for TR to exist as a dimer (TR_2) at micromolar concentrations. Figures 3.16 A-D shows evidence of TR_2, ΔG = -6.72 kcal.mole^{-1} (http://eu.idtdna.com/calc/analyzer). Figure 3.17 demonstrates the predicted secondary structure of the dimer. The presence of MTX increases the amount of TR_2, when compared to TR alone. By MS, TR_2 shows binding of up to eight MTX molecules. The alkyl-polyamine arms of MTX mean that each molecule may be able to interact electrostatically with two RNA SL at once. The self-associating nature of MTX at high concentrations could also be responsible for the high level of binding (Enache and Volanschi 2010).

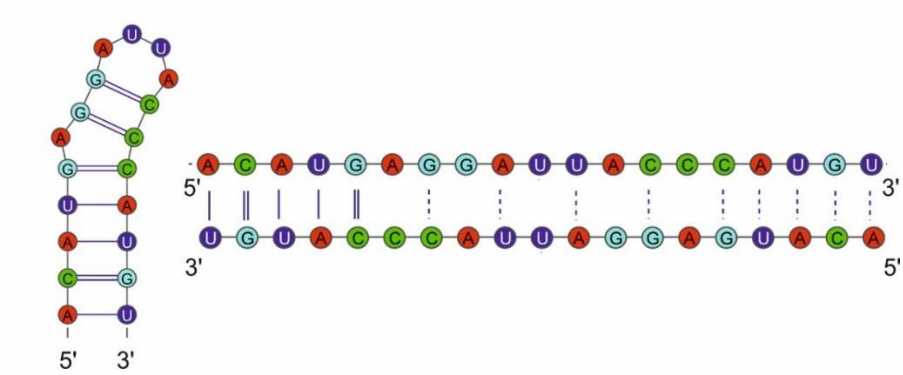


Figure 3.17 Predicted Secondary Structures of TR monomer ΔG = -7.09 kcal.mole^{-1} (left) and homodimer ΔG = -6.72 kcal.mole^{-1} (right). ΔG is calculated by taking into account the longest stretch of base pairs (bp). These bp are represented by a solid line. Dotted lines represent additional bp, but their presence does not impact calculated ΔG values. Actual ΔG values may vary based on presence of additional complementary bases. Calculated by http://eu.idtdna.com/calc/analyzer, drawn in Varna (Darty 2009).

To follow the effect of TR's adenosine bulge, the mutant RNA TR-10A was also titrated with MTX and subjected to ESI-MS. When ionised, TR-10A shows charge states 5-, 4- and 3- the most abundant being 4-, Figure 3.18. TR-10A RNA shows an observed mass = 5713.61 ±0.51 Da, correlating with the expected mass from the nucleotide sequence 5715.49 ±0.01 Da (OligoAnalyzer 3.1, Integrated DNA Technologies).

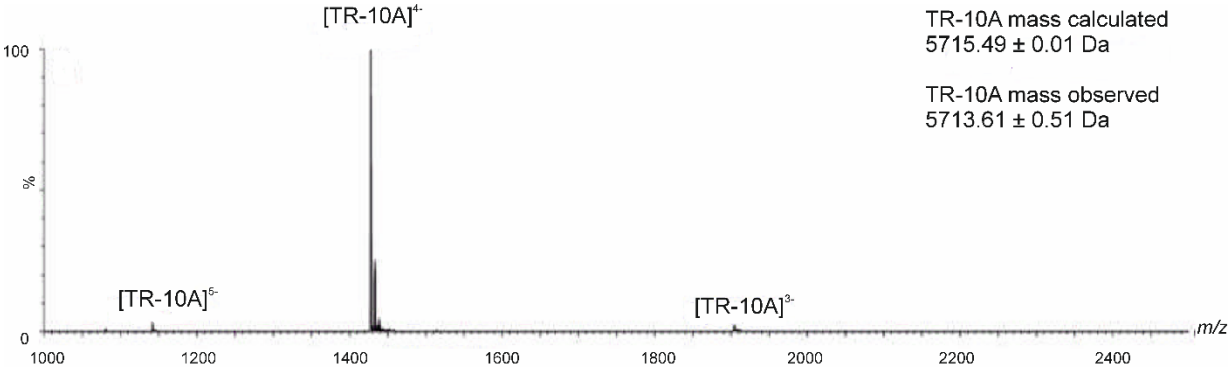


Figure 3.18 Mass spectrum of TR-10A RNA. RNA [14 µM] in 200 mM AmAc, pH 6.5. The calculated mass was 5715.49 ±0.01 Da which correlates with the observed mass of 5713.61 ±0.51 Da.

TR-10A was titrated with increasing amounts of MTX, see Figure 3.19. Again, the integration data of the major species detected, monomer (4-) ion, is compared in Table 3.7.

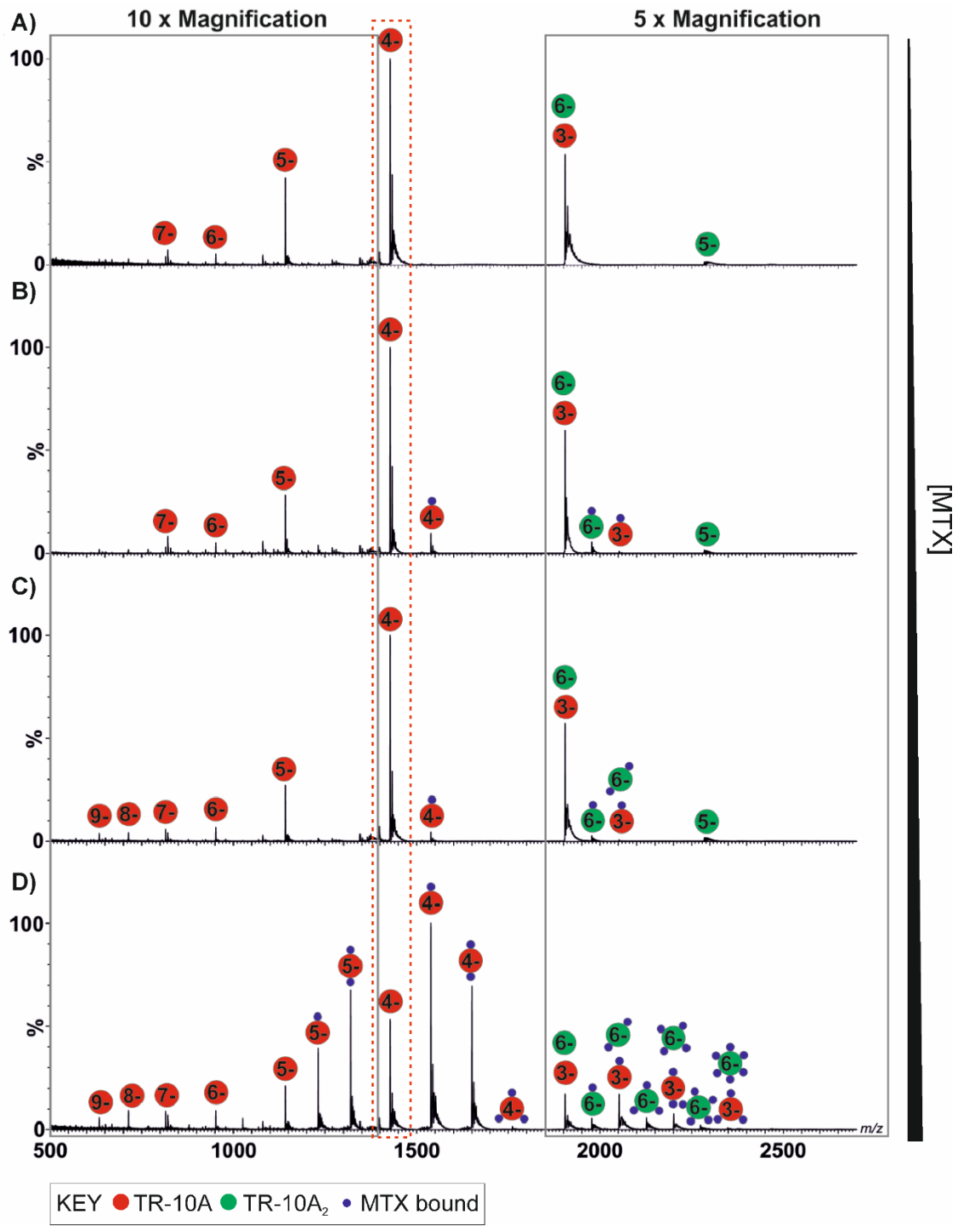


Figure 3.19 Negative Ion mode MS of MTX binding to TR-10A. Mass spectra of TR-10A [7.5 µM] in MS Buffer in the presence of MTX at different molar ratios TR-10A:MTX A) 50:1, B) 10:1, C) 5:1, and D) 1:1. Labels above peaks show TR monomer (red), TR dimer (green) and number of MTX bound (purple). Red box highlights the ligand-free 4- charge state.

Table 3.5 Comparing (4-) Ion intensities of monomeric TR-10A with MTX. Integrated peak heights (% shown in brackets) measured using MassLynx™. Increasing the ratio of MTX shows up to three MTX molecules bind per RNA.

TR-10A:MTX	Free RNA	+1 MTX	+2 MTX	+3 MTX	TOTAL IONS (%)
1:0.02	6,127 (100)	0	0	0	6,127 (100)
1:0.1	1,583,166 (96.5)	58,201 (3.5)	0	0	1,611,367 (100)
1:0.2	860,962 (95.7)	38,322 (4.3)	0	0	899,284 (100)
1:1	129,963 (23.8)	236,250 (43.25)	176,113 (32.24)	3,965 (7.3)	546,291 (100)

An interesting observation with this titration is the increase in the number of ionised charge states for TR-10A upon increasing MTX. In the region below 1500 *m/z* TR-10A shows charge states 9- through to 5-, highlighted with 10X magnification, Figure 3.19 D. In the absence of MTX, only ions 5- are generated, Figure 3.18. This implies that the presence of MTX allows for more phosphodiester protons to be lost during ionisation – this could be due to a higher degree of unfolded RNA (Banerjee 2012).

The ESI-MS of TR and TR-10A at a 1:1 RNA:MTX ratio, at the major monomeric (4-) charge state shows MTX binding, Figure 3.20. TR binds up to four MTX molecules, whereas TR-10A binds only three. There was less ligand-free TR in solution than ligand-free TR-10A, 10.5% versus 23.8 %. This result confirms that MTX has a greater affinity to TR than to TR-10A.

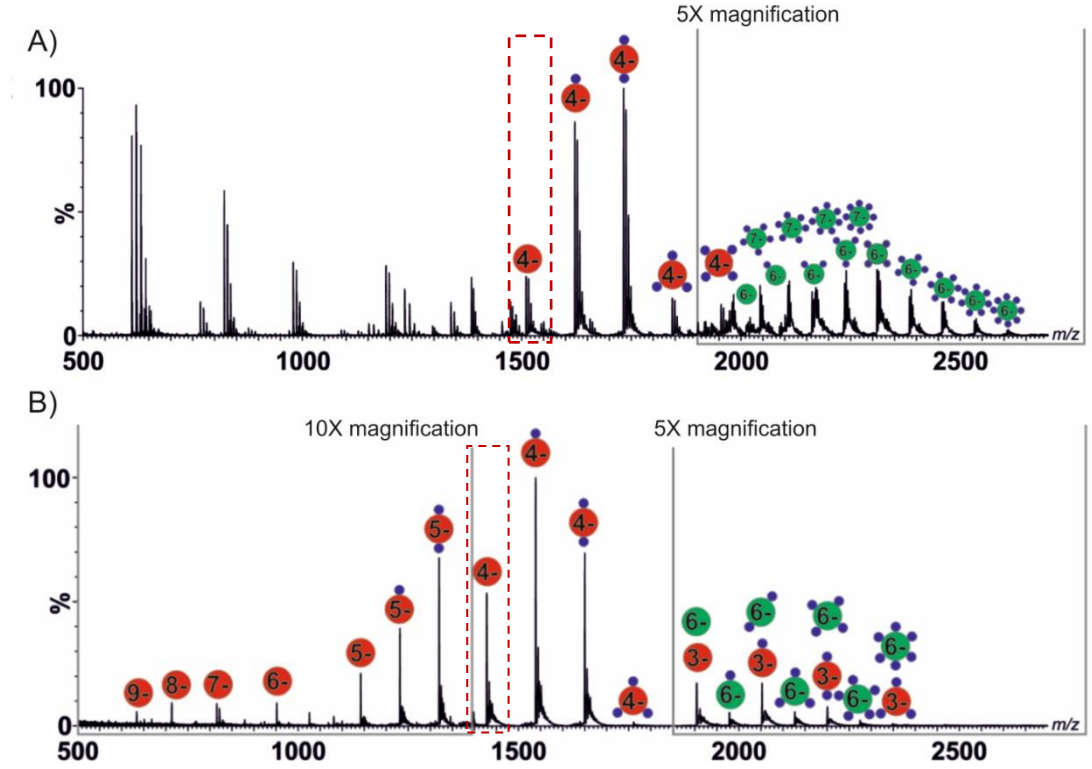


Figure 3.20 Assessing RNA:MTX binding interaction using ESI-MS. RNAs A) TR. B) TR-10A at 1:1 molar ratio with MTX. RNA monomer (red), RNA dimer (green) and MTX (purple) Observing the monomer 4- ion (highlighted by red box) shows the ligand-free 4- charge state. There is a higher amount of free TR-10A than free TR, suggesting that MTX has a greater affinity for binding to TR.

As mentioned in Section 3.2.1, svAUC showed evidence for DMA binding to TR and this interaction was also explored further by MS. DMA has been shown using fluorescence assays to specifically bind apical RNA stem loops (Stelzer 2011). Fluorescence binding studies showed that DMA binds the apical loop of HIV-1 TAR with µM affinity (Stelzer *et al.* 2011). DMA provides a rare example of a small molecule that exclusively targets an RNA apical loop (Moumné 2012). DMA was tested for binding to TR using the same MS assay described previously, results shown in Figure 3.21. At 10:1 ratio of DMA:TR, monomeric TR binds up to three DMA molecules, whereas at a 1:1 ratio there is no DMA binding but the presence of DMA shows an increase in the number of negative ion charges states of TR. The larger distribution of negative charge states in the region ≤ 1200 *m/z* of Figure 3.21 A, suggests an increase in unfolding of TR RNA as described in Banerjee *et al.* (2012).

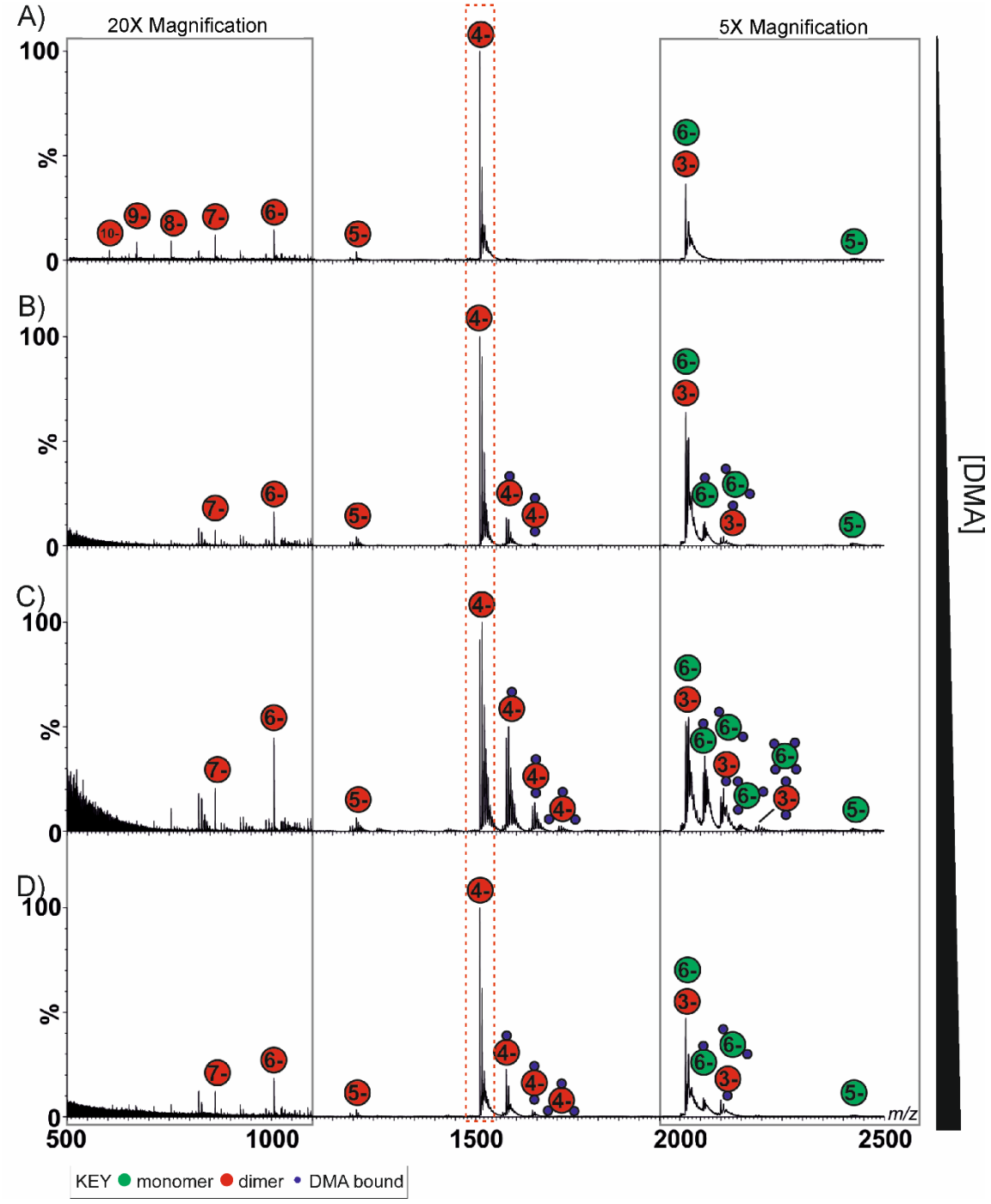


Figure 3.21 Negative Ion mode MS of DMA binding to TR. Molar Ratio TR:DMA A) 1:1 B) 1:5 C) 1:7 D) 1:10, TR [7.5 µM]. There is no binding at equi-molar concentrations however above this TR binds up to three DMA, and TR_2 can bind up to four. Red box denotes ligand-free 4- charge state.

Table 3.6 Comparing (4-) ion intensities of monomeric TR with DMA. Looking at the MS data in Figure 3.22, the major charge state of TR RNA [7.5 µM] is (4-), integrated peak areas were calculated as before (%s shown in brackets) (MassLynx™ Software, Waters). Increasing the ratio of DMA shows up to 3 molecules bind per TR RNA.

TR:DMA	Free RNA	+1 DMA	+2 DMA	+3 DMA	TOTAL IONS (%)
1:1	985,365 (100)	0	0	0	985,365 (100)
1:5	1,020,620 (87.1)	136,903 (11.8)	12,538 (1.1)	0	1,170,061 (100)
1:7	391,989 (59.6)	198,828 (30.2)	55,288 (8.4)	11,890 (1.8)	657,995 (100)
1:10	318,718 (81.3)	47,926 (12.2)	1,981 (5.1)	5,352 (1.4)	373,977 (100)

By comparing TR with both ligands at 1:1 molar ratio, Figure 3.23, the *m/z* region > 2000 has a peak distribution characteristic of non-specific binding (Rosu 2008). The difference between the binding affinities can be seen by comparing the amounts of ligand-free TR in the presence of these two ligands. As mentioned, DMA has been reported as binding to the apical stem-loop of HIV-1 TAR using NMR (Stelzer 2011). The sequence of the apical stem-loop of TR (-AUUA-), is different to HIV-1 TAR (-CUG GGA-), which could explain the difference in binding affinities.

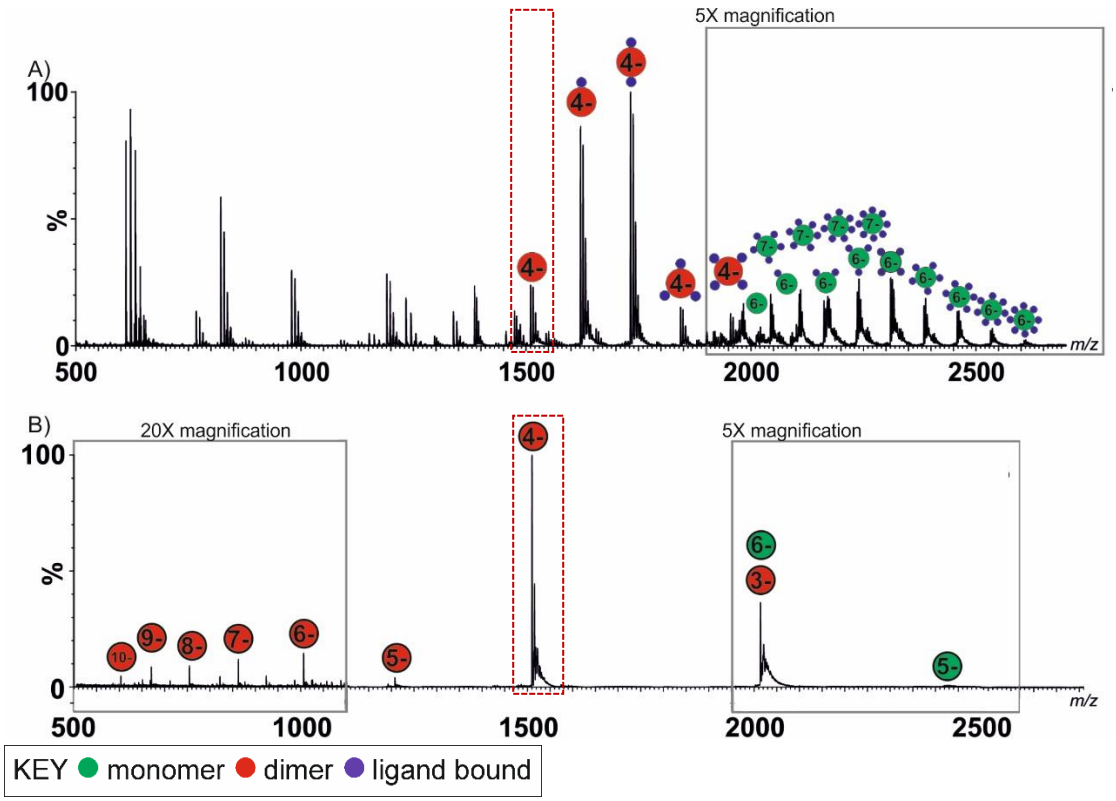


Figure 3.22 Negative ion mode MS of TR [7.5 µM] 1:1 molar ratio with ligands A) MTX and B) DMA. Red box denotes ligand-free 4- charge state. Comparing these two ligands shows that there is less TR:DMA than TR:MTX, suggesting TR has a higher affinity for MTX.

3.3. Discussion

From an initial literature search of RNA-binding molecules, ROCS was used to identify characteristics as a starting point for an *in silico* search. This included a wide range of structural features such as functional groups, aromatic rings and charge distributions. The molecules shortlisted for *in vitro* analysis looked promising from the start as some already were already known to be RNA-binding ligands. This technique could have been enhanced by classification of ligands based on their potential to bind RNA via electrostatics or intercalation, instead of taking a "one-molecule-fits-all" approach.

Firstly, svAUC identified changes in hydrodynamic properties of TR upon ligand binding. However, ligand-binding and RNA unfolding both increased and decreased the experimental sedimentation coefficient, which resulted in good binding molecules being missed. It was therefore decided to pursue a different technique. Ligand binding using changes in circular

dichroism or fluorescence spectroscopy showed that some molecules exhibited spectroscopic properties that complicated the analyses. Using this as an advantage, the unique absorbance properties of just one ligand, MTX, were assessed. The technique showed that MTX bound TR RNA and gave an insight into the stoichiometry and strength of the interaction. MTX binds TR with $K_D = 3.71 \pm 0.22 \times 10^{-7}$ M. This interaction is tighter than previously cited work for dsDNA which calculates that anthracycline molecules interacting with dsDNA are usually in the range $K_D \approx 10^{-5} - 10^{-6}$ M (Berg *et al.* 1981, Ibrahim 2001). Using mutant RNA "TR-10A" showed the importance of the bulged adenosine base within the RNA aptamer, removing this increased the K_D 3-fold to $1.11 \pm 0.17 \times 10^{-6}$ M. The significance of the adenosine bulge in TR is hugely important with respect to the binding capacity of TR to MS2 CP. The adenosine at position -10 makes hydrogen bonds to CP functional groups (Valegård *et al.* 1994). Substitution by deoxy-adenosine has no effect on CP affinity (Stockley *et al.* 1995) but deletion of this nucleotide results in no detectable affinity (Talbot *et al.* 1990). If MTX binds TR at the -10A position the RNA molecules may be unable to bind MS2 CP, suggesting MTX is a suitable anti-assembly ligand. Analysis by ESI-MS showed that the MTX used throughout these experiments was of questionable purity. Purification techniques, such as reverse-phase liquid chromatography, could be used to ensure further work is carried on only the intact monomer. Further work could be pursued using *de novo* synthesis to modify functional groups on MTX; changing charge distributions and 3D molecular shape will help to search for a higher affinity version of this ligand. ESI-MS showed that there are multiple ligands binding per SL, therefore better characterisation of the interaction should be pursued. Looking at the interaction at a molecular level could ultimately help design a better binding molecule. MTX has been characterised binding to RNA stem loops previously, more specifically it binds to un-paired adenosine bulges as demonstrated using NMR techniques (Zheng *et al.* 2009). NMR allows accounting for most of the hydrogen bonding involved in the interaction and the binding pocket that MTX fits into. RNA in the virus life cycle has many roles to perform in order for generation of successful progeny. The secondary

structure of RNA is believed to be crucial in ensuring the fulfilment at each stage from the initial virus infection through to mature virion release. This thesis is interested in the secondary structure of the viral RNA genome during the assembly of infectious particles and Chapter 4 investigates MTX disrupting virus assembly *in vitro.*

Chapter 4

4. Effects of TR Binding Ligands on MS2 Assembly

This page has been left intentionally blank.

4.1. Ligand Inhibition of TR-mediated Reassembly of MS2 *in vitro*

The work described in the previous Chapter identified several small molecular weight TR-binding ligands. The effects of those ligands on the *in vitro* reassembly of bacteriophage MS2 is described in this Chapter. For *in vitro* reassembly assays recombinant MS2 coat protein (CP) was prepared as described in Mastico (1993). The recombinant protein self assembles into VLPs, these are stored in saturated ammonium sulfate at 4°C until needed. VLPs are acid dissociated for 2 hours on ice (1:1 (v/v) glacial acetic acid) and then desalted into 20 mM acetic acid. Figure 4.1 shows an SDS PAGE of the fractions eluting from such a column. These are used within 2 hours of desalting in order to ensure minimal amounts of protein aggregation.

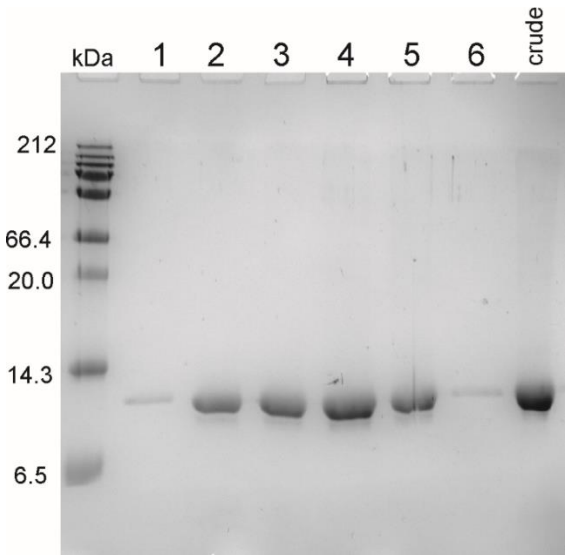


Figure 4.1 12% (w/v) SDS-PAGE of recombinant MS2 CP. Lanes: kDa Protein marker (Broad Range Protein Ladder 2-212 kDa, NEB). Lanes 1) - 6) Dissociated CP fractions after desalting into 20 mM acetic acid. Crude = recombinant MS2 VLP.

TR was synthesised and purified as per Section 2.1.5. TR [1 µM] was incubated with and without ligands [1 µM] in MS2 Buffer A. For reassembly reactions, MS2 CP was added to RNA using a 2-stage reassembly process; starting with equivalent ratios of TR:CP_2 to allow a 1:1 complex to form, then raising the CP_2 concentration to a final 1:2 ratio needed for VLP formation. There is no CP_2:RNA interaction at reduced pH so reassembles were spiked with Tris-HCl (final concentration 15 mM) to buffer at pH 6.5 and left at room temperature for at least two hours

prior to further work, this temperature allows for stable VLP formation. Concentrations were determined from absorbance measurements using molar absorption coefficients calculated from nucleotide or amino acid sequences. TR stock concentration 71.9 µM, see Section 2.1, Table 2.1 for sequence information. MS2 CP_2 stock concentration 24 µM, $\varepsilon_{280\,nm}$ = 33,420 $M^{-1}.cm^{-1}$ [1] (ProtParam, ExPASy using PDB file #1ZDH (Valegård *et al.* 1997)). *In vitro* TR-mediated MS2 reassemblies were assayed by native agarose gel electrophoresis and nsTEM, Figures 4.2 and 4.4 respectively.

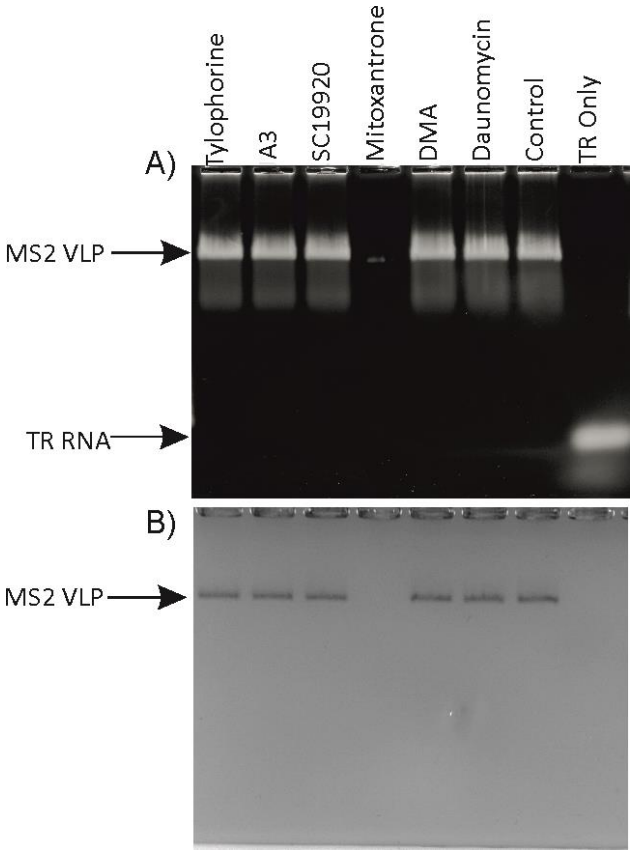


Figure 4.2 MS2 VLPs Reassembled with TR +/- Ligand. Native agarose gel electrophoresis of TR:ligand molar ratio 1:1 [1 µM] reassembled with MS2 CP_2 to equivalent molar ratio 1:2. "Control" denotes reassembly reaction with no ligand present. A) Stained for nucleic acid with EtBr. B) Stained for protein using InstantBlue™ Coomassie.

This experiment indicates that virus-like particles (VLPs) form in the presence of all the ligands except the ligand mitoxantrone (MTX) (Figure 4.2). Even with a lack of VLPs formed it was noted that there was no free TR RNA shown in the MTX reassembly sample. MTX was present at approximately twice the K_D for TR (Chapter 3) so it could be expected to bind the RNA. It is

however, possible that there was binding competition between the EtBr nucleic acid and MTX so the RNA was not well visualised. To test this idea 3'-FAM labelled TR (FAM-TR) was titrated with MTX and analysed by native agarose gel electrophoresis. The gel was stained using EtBr and even with increased incubation with EtBr and over exposure during visualisation, EtBr was poor at detecting TR in the gel (Figure 4.3 panel A). Scanning for FAM-TR showed that the decrease in RNA detection by EtBr staining is not associated with RNA degradation, but with competition for RNA intercalation of EtBr (Figure 4.3 B). The presence of MTX showed no change in the electrophoretic mobility of TR. nsTEM images of the reassembly products seen in Figure 4.2 are shown in Figure 4.4.

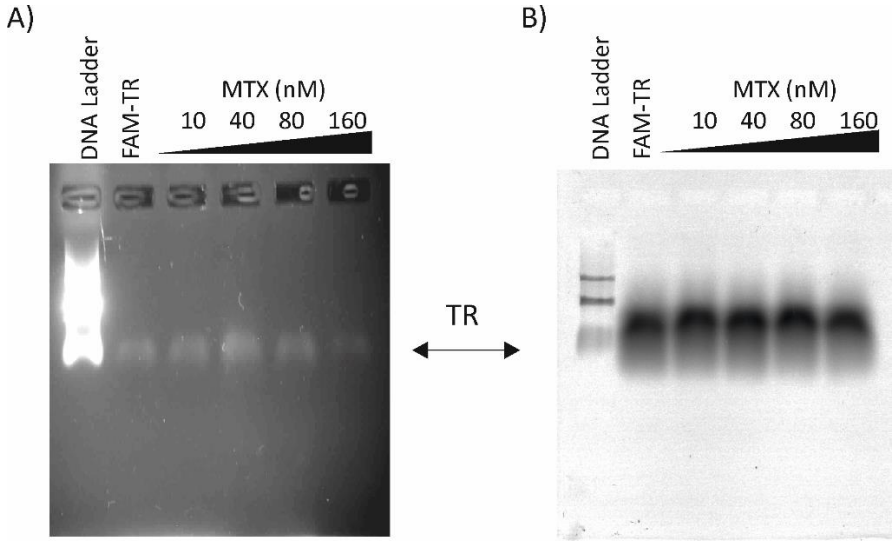


Figure 4.3 Binding of MTX to TR by native gel electrophoresis. A) TR RNA stained with ethidium bromide. Over-exposing with UV light detects some RNA is present. B) As in A) scanned for fluorescence, $Ex_{495\ nm}$, $Em_{525\ nm}$.

The results confirm that only MTX prevents the formation of VLPs under these conditions. However, the VLPs formed in the presence of A3, SC19220 or DMA are also less well ordered than with TR alone, suggesting that the ligands may have some inhibitory effects on assembly. Since the inhibition properties of MTX were clear, it was decided to concentrate on this ligand and to characterise its mode of action.

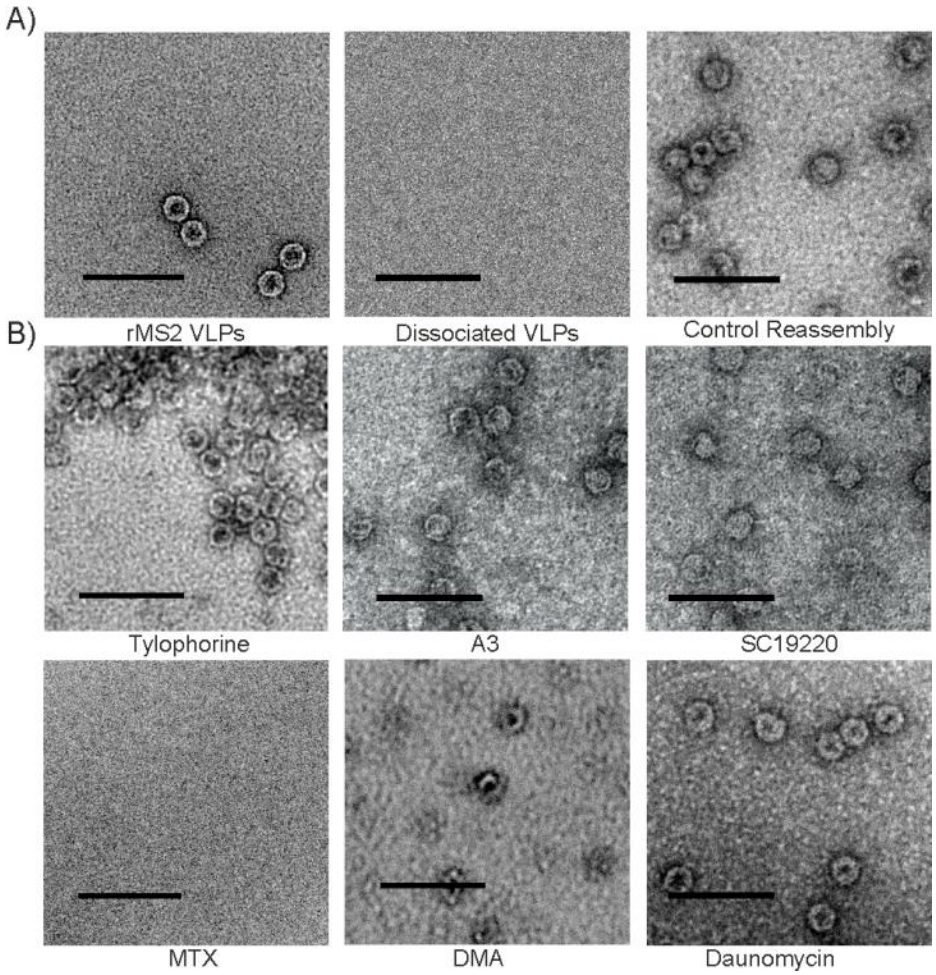


Figure 4.4 nsTEM of TR-mediated MS2 reassemblies +/- ligands. A) rMS2 VLPs, dissociated MS2 CP and TR-mediated MS2 reassembly. B) TR-mediated MS2 reassembly with various ligands. Images magnified 30k X, scale bar represents 100 nm, visualised with 2% w/v uranyl acetate.

4.2. Studying the Reassembly of MS2 with TR in the Presence and Absence of MTX using Native MS.

Previous work on MS2 reassembly using native MS monitored TR-mediated assembly and identified higher order intermediates corresponding to the assembly products that form around both the 3-fold and 5-fold axes, Figure 4.5 (Stockley *et al.* 2007).

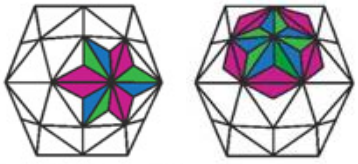


Figure 4.5 MS2 reassembly intermediates identified by MS showing the arrangement of the quasi-equivalent dimers, A/B (blue/green) and C/C (pink). Left [3(CP_2+TR)+3CP_2] 182.9 kDa. Right [5(CP_2+TR)+5CP_2] 304.8 kDa. Figure adapted from Morton et al. (2008).

Taking these experiments into consideration it was decided to monitor the reassembly of TR-mediated MS2 VLPs by native MS in the presence of MTX. The experimental MS work done in this Chapter was carried out in collaboration with Dr Henry Fisher, University of Leeds. At a 1:1 molar ratio of TR:CP_2 [3 µM] the major species is TR:CP_2 (green) with evidence for CP monomer (orange), Figure 4.6A. Note, it is not possible to record MS spectra at the lower TR concentrations used in the biochemical assays described in Section 4.1. A 10-fold magnification at *m/z* 6,000-10,000 shows a cluster of peaks corresponding to a 5-fold reassembly intermediate (dark blue). This result suggests that the experimental concentrations are not precisely 1:1 RNA:CP_2, and that there is slight excess CP_2. Increasing the molar ratio of TR:CP_2 to 1:2 reveals free CP monomer (orange) and a small amount of CP dimer (red), Figure 4.6 B. The loss of the peak corresponding to TR:CP_2 is due to the formation of higher-order reassembly intermediates, shown *inset* at 10-fold magnification [3(CP_2+TR)+3CP_2] (mauve), [5(CP_2+TR)+5CP_2] (dark blue) and 3[5(CP_2+TR)+5CP_2] (pink). The broad peak at m/z 23,000-26,500 (light blue) is indicative of VLPs (Stockley 2007). To measure the effect of MTX on reassembly this assay was repeated under identical experimental conditions but with a 20-fold excess of ligand over RNA. When TR:CP_2 is at a 1:1 ratio and in the presence of MTX, only peaks corresponding to CP monomer and CP_2 are observed suggesting that a) MTX does not bind to CP_2 and b) that MTX blocks TR binding by CP_2, Figure 4.7 A. Free MTX (green) and MTX dimer (pink) appear at <*m/z* 1000. Increasing the molar ratio of TR:CP_2 to 1:2 reveals a trace of [2(TR:CP_2)] (brown), but even increasing the *m/z* axis by a 10-fold magnification reveals no detectable species above ~67 kDa, Figure 4.7 B *inset*. These results confirm that MTX binds to TR and prevents its binding to CP_2. A native agarose gel of these MS samples was used to verify that the ionisation process used in MS analysis did not omit any other assembly intermediates, Figure 4.8. The results are consistent with those from MS. To examine if the increasing CP_2 concentration would overcome inhibition of a sample at a molar ratio TR:CP_2 at 1:4 was also examined, however it mainly formed aggregates that did not fly in the MS.

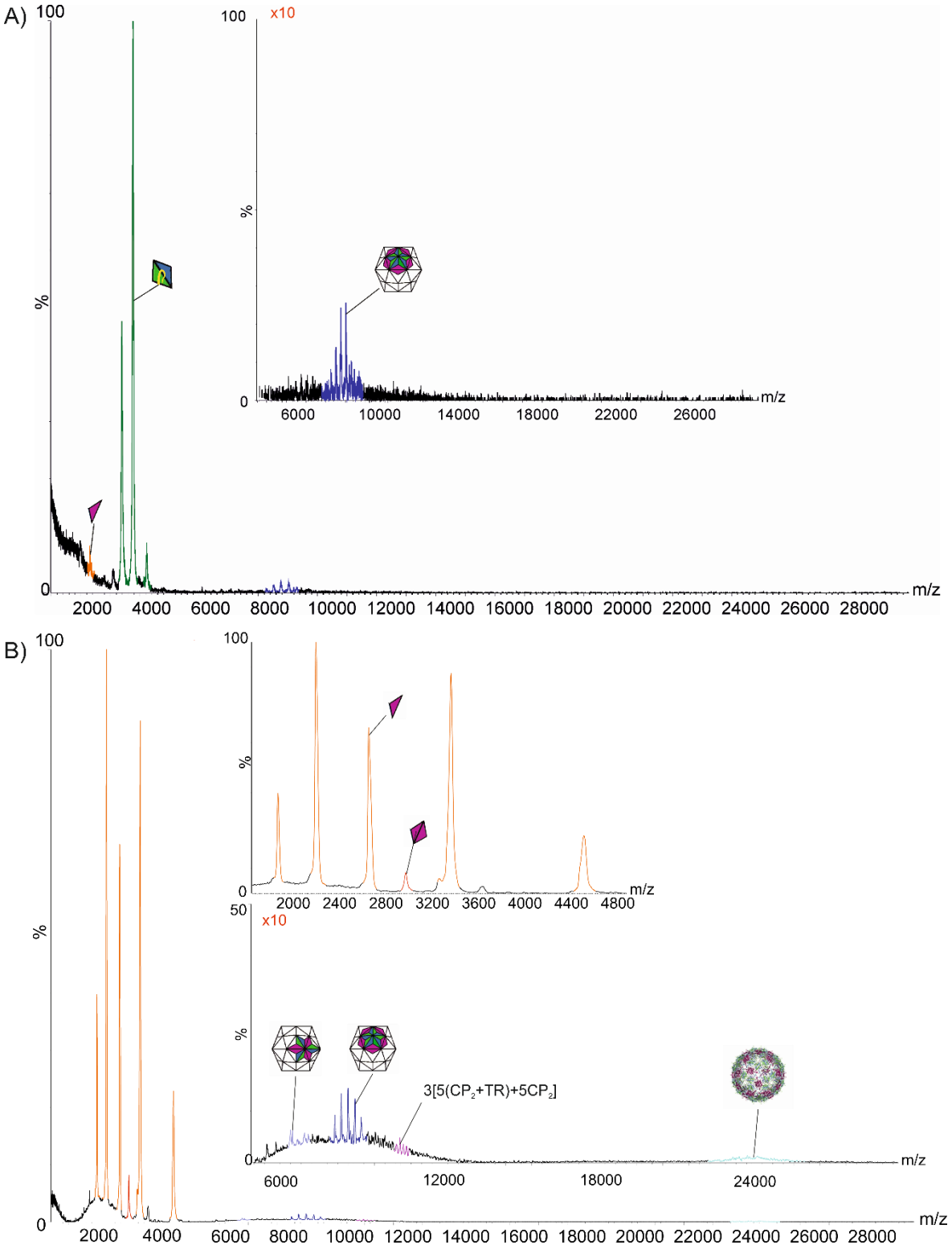


Figure 4.6 Native MS of TR and CP_2 reassembly reaction. A) At a 1:1 molar ratio of TR:CP_2 the major species is TR:CP_2 (green) with a small amount of monomeric CP (orange). Inset 10-fold magnification at m/z >6,000 shows a signal corresponding to the expected stoichiometry of the 5-fold assembly intermediate. This, and the presence of free CP, would suggest there is slight excess of CP_2 over TR. B) Changing the molar ratio of TR:CP_2 to 1:2, shows free CP (orange) in solution and the formation of reassembly intermediates. Inset m/z 1,000-4,800 shows only CP and CP_2. A 10-fold magnification of the region m/z >6,000 shows [3(CP_2+TR)+3CP_2] (mauve), [5(CP_2+TR)+5CP_2] (dark blue) and 3[5(CP_2+TR)+5CP_2] (pink). The broad peak at m/z 23,000-26,500 (light blue) is indicative of VLPs. Figure created in collaboration with Dr Henry Fisher.

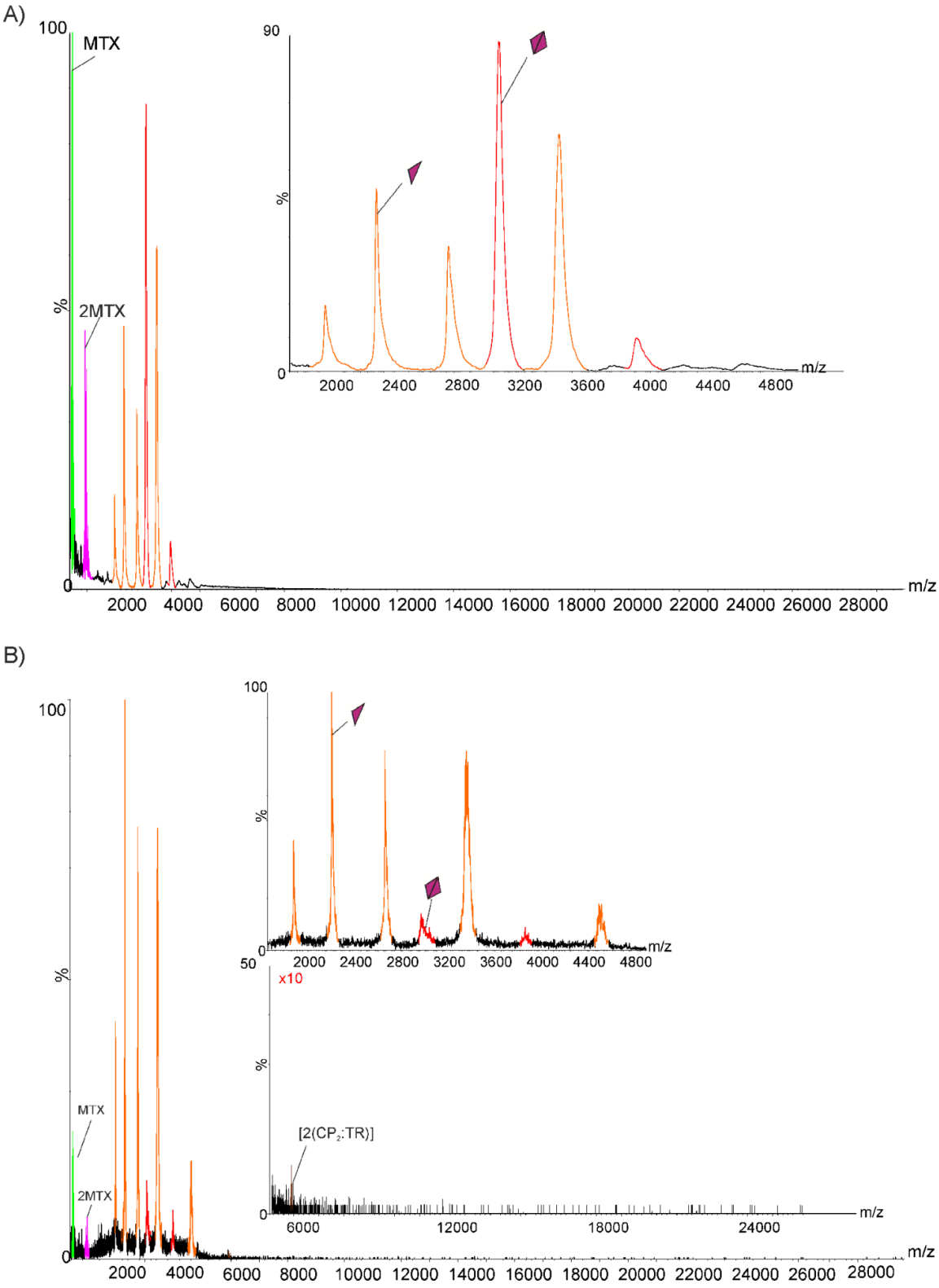


Figure 4.7 Native MS of TR and CP_2 reassembly reaction in the presence of MTX [60 µM]. A) At a molar ratio of TR:CP_2 only CP monomer (orange) and CP_2 (red) is shown. Free MTX (green) and MTX dimer (pink) appear at < m/z 1000, there is no evidence for MTX binding to CP. B) Increasing the molar ratio of TR:CP_2 1:2, free MTX (green) and MTX dimer (pink) appear at < m/z 1000 alongside CP monomer (orange) and CP_2 (red). Also shown inset bottom at 10-fold magnification is a species ~67 kDa [2(TR:CP_2)] (brown). There are no detectable species above this mass. Figure created in collaboration with Dr Henry Fisher.

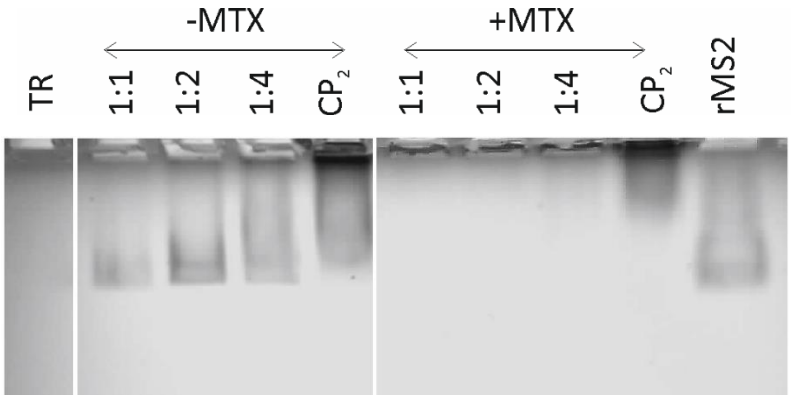


Figure 4.8 Native gel electrophoresis of MS reassembly samples. TR-mediated MS2 reassembly in the presence and absence of 60 µM MTX. Lanes denote molar ratios of TR:CP_2. CP_2 only and rMS2 = recombinant MS2 are included as controls. Stained for protein using InstantBlue™.

4.3. gRNA-Mediated MS2 Reassembly *in vitro*

There is only one copy of the TR nucleotide sequence in the MS2 genome, the sequences and locations of other PSs have been predicted as discussed in Chapter 1 (Dykeman *et al.* 2013). Their functions as PSs are consistent with recent cross-linking immunoprecipitation sequencing (CLIP-Seq) analysis of RNA content with CP of infectious MS2 virions (Rolfsson *et al.* 2015). Data from Rolfsson *et al.* (2015) suggests at least 8 unpaired adenosine sites in base-paired stems that bind to CP_2 (8/54), Appendix Figure 8.16. Previously, MTX has been shown to bind RNA at unpaired adenosine sites in base-paired stems (Zheng 2009), this is consistent with results from Chapter 3 using TR and TR-10A mutant. If MTX binds at these PS sites, their ability to bind CP could be inhibited. The effect of MTX on MS2 reassembly using gRNA was therefore assessed. Transcripts of gRNA were prepared as described in Borodavka et al. (2012), see Materials and Methods. Figure 4.10 B shows a denaturing gel of the purified gRNA.

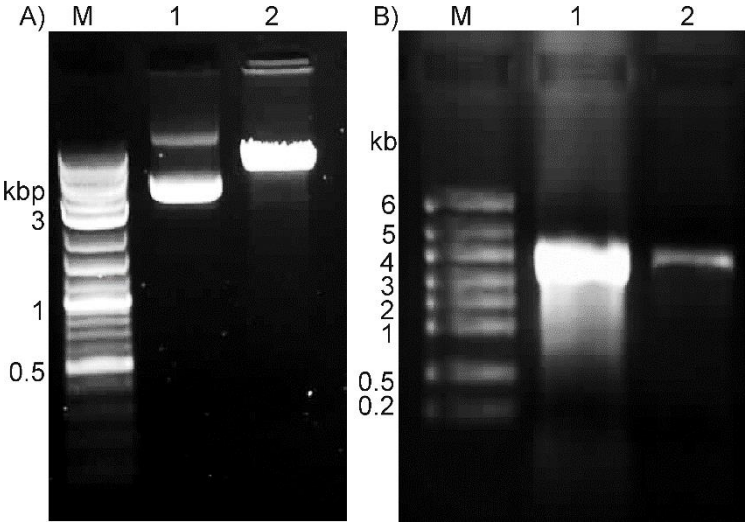


Figure 4.9 Native agarose gel analysis of MS2 gRNA preparations. A) 1.5 % (w/v) agarose Tris-Borate-EDTA gel of cDNA, run for 1 hour at 90V, 4 °C. Lanes: M = 2-log DNA ladder (0.1-10 kbp, New England Biolabs Inc.) Lanes 1) HindIII linearised MS2 cDNA plasmid. 2) Supercoiled plasmid MS2 cDNA. B) Denaturing formaldehyde/MOPS, 1.5% (w/v) agarose gel of in vitro transcription products of gRNA from sample in lane 1 in A). Lanes: M = RiboRuler High Range (0.2-6 kbp, ThermoFisher Scientific). Lanes 1) in vitro transcription product MS2 gRNA. 2) Purified MS2 gRNA at 1/20 dilution. Both gels visualised using EtBr stain.

Figure 4.9, panel A shows that the supercoiled plasmid cDNA was successfully digested into a linear 3,569 nt DNA template using *Hind*III. This template was transcribed with T7 RNA polymerase producing a band of the correct size (~3659 nt), Figure 4.10 B. The gRNA was purified to remove any residual dNTPs and enzymes (RNAEasy Kit, Qiagen), generating a stock for the all the reassemblies listed in this Chapter. Prior to the reassembly reactions, all gRNA was heated to 95°C for 1 minute, and then cooled to 4°C (1°C/minute) in order to promote secondary and tertiary intramolecular interactions. Native gel electrophoresis was used to see if there were any changes in electrophoretic mobility of gRNA when titrated with MTX. Titrating in MTX leads to a reduction in the amount of gRNA detected by EtBr (Figure 4.10), as expected if it binds. There was a slight decrease in the electrophoretic mobility of the RNA in the presence of MTX.

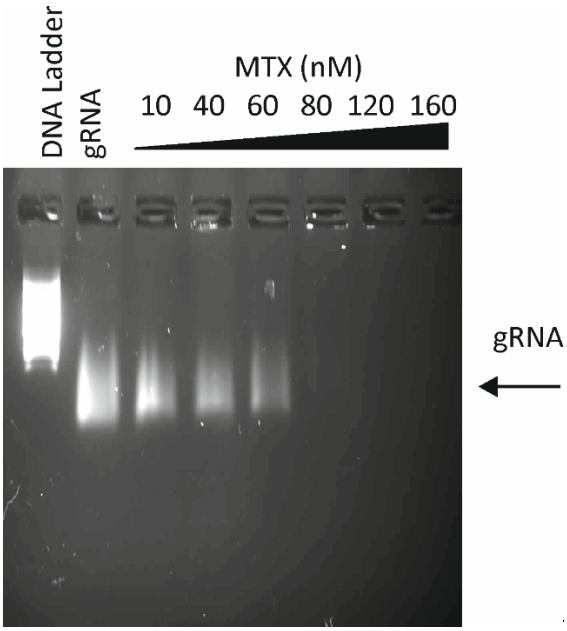


Figure 4.10 Effect of MTX on the electrophoretic mobility of gRNA. Lanes contain gRNA visualised by EtBr staining. There was a slight decrease in the electrophoretic mobility of gRNA and a reduction in the efficiency of EtBr stain. Both suggests that there is binding of MTX, with competition between MTX and EtBr.

4.3.1. Effects of MTX on Reassembly of MS2 VLPs with gRNA.

Due to the size of the MS2 gRNA VLP reassembly experiments can not be analysed by MS. Instead native agarose gels and nsTEM were used to assay the reactions. gRNA was incubated with MTX, as previously mentioned for TR, and then used in *in vitro* reassembly reactions with MS2 CP_2. The effect of titrating in MTX on MS2 gRNA reassembly reactions is shown in Figure 4.11. In order to avoid problems with EtBr staining in the presence of MTX, SybrGreen™ II (ThermoFisher Scientific) was used as an alternative for RNA visualisation. It binds RNA both electrostatically and by intercalation (Zipper, 2004). A control reassembly in the absence of MTX shows a band migrating at a similar position to the rMS2 VLP. A lower smear of potentially unincorporated gRNA is also present. Its intensity increases as the amount of MTX is increased, whilst simultaneously the amount of VLPs present is reduced. These results clearly suggest that increasing [MTX] reduces the amount of MS2 VLPs reassembled with gRNA *in vitro*.

Dykeman *et al.* (2013) have proposed that MS2 gRNA could have up to 60 PSs which could help to set the conformation of the 60 A/B CP_2 in the *T*=3 shell. CLIP-Seq data confirm that at least 54

of these are in contact with CP subunits (Rolfsson *et al.* 2015). In this experiment, even at the lowest concentration of MTX, 5 µM, there is an equivalent of ~2.5 MTX molecules per PS in a 60 PS gRNA. This should be sufficient to ensure an inhibitory effect on the formation of VLPs, especially if the 8 sites carry an unpaired adenosine, equivalent to the -10A site in TR, are the ligand binding target. The reassembly products of these experiments were visualised using nsTEM, see Section 4.3.2.

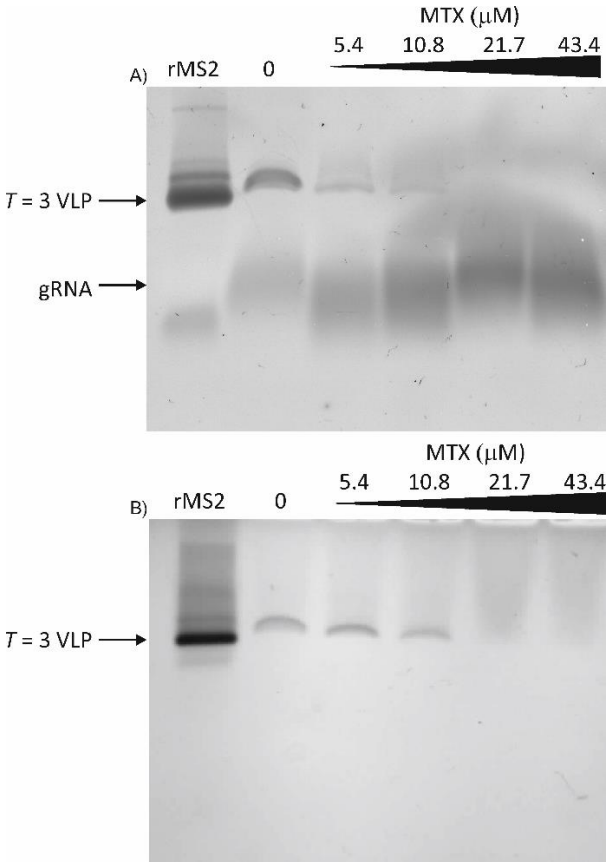


Figure 4.11 The Effects of MTX on Reassembly of MS2 VLPs with gRNA. Results were analysed by native agarose gel electrophoresis. MS2 VLP reassembled using gRNA [34 nM]. rMS2 is recombinant VLP CP_2 used as a control marker. A) Stained for nucleic acid using SybrGreen™ II. B) Stained for protein using InstantBlue™.

4.3.2. Using nsTEM to quantitate gRNA-Mediated MS2 Reassembly in the Presence and Absence of MTX

Aliquots of the samples from Figure 4.11 were analysed by nsTEM, ≥3 images from each EM grid were collected for each MTX titration and the numbers of partially and fully assembled VLPs counted. Comparing equivalent areas of prepared EM grids are demonstrated below. Comparing

nsTEM of MS2 reassemblies in the presence and absence of MTX are shown in Figure 4.12. In the presence of MTX various incomplete and mis-assembled capsids are seen.

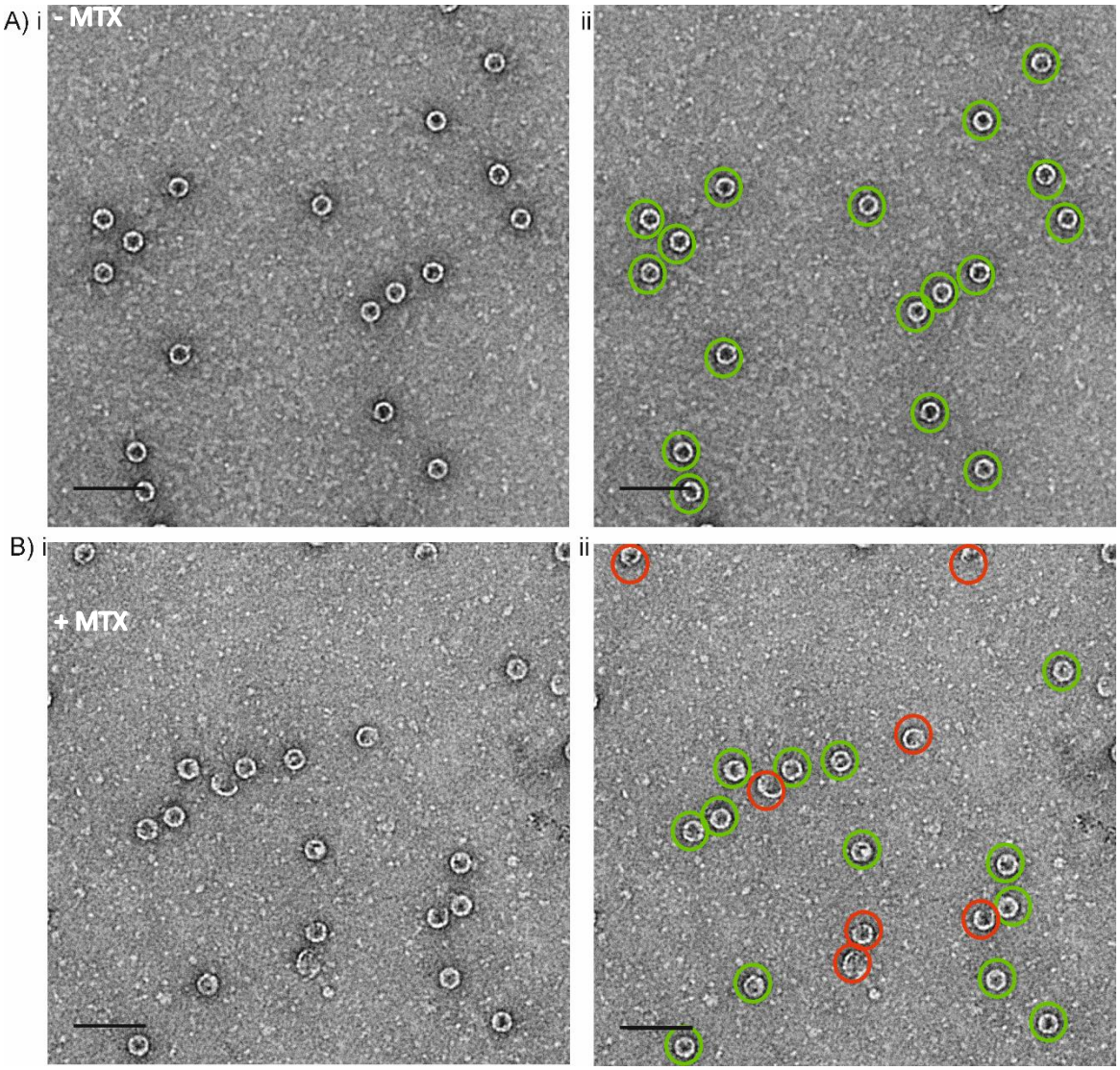


Figure 4.12 Analysing the effect of MTX on reassembly with gRNA by nsTEM; demonstrating VLP classification. A) gRNA MS2 VLP reassembly, no MTX. B) gRNA MS2 VLP reassembly in the presence of MTX. i) Original images. ii) Fully (green) and partially reassembled VLPs (red) are shown. Particles were classified as fully assembled if they consisted of stain-excluding complete protein rings in these projections. Incomplete or malformed species were assigned to the partially assembled category. EM images at 30k X magnification, scale bar 100 nm.

Quantification of these VLPs shows a clear trend, increasing MTX concentration reduces the number of completed VLPs formed whilst increasing the number of partial or malformed structures, as shown by Figure 4.13. These data confirm that MTX inhibits VLP assembly around

gRNA leading to the formation of malformed species and at higher concentrations completely oblates VLP assembly.

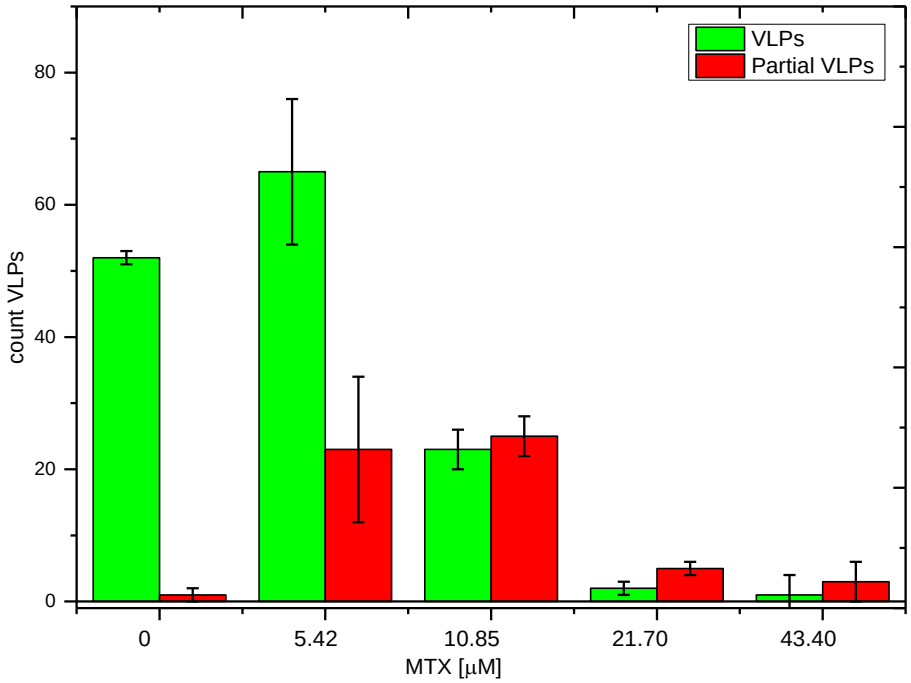


Figure 4.13 Quantification of fully and partially reassembled VLPs of gRNA reassembled MS2 in the presence and absence MTX.

4.4. Discussion

Native agarose gel data presented in this Chapter revealed that the formation of MS2 VLPs triggered by TR occurs in the presence of tylophorine, A3, SC19220, DMA and daunomycin and no VLPs were formed in the presence of MTX (1 μM), Figure 4.2. Further analysis by nsTEM suggested reassemblies with tylophorine, A3 and DMA appear less well ordered, which suggests that they may have some effect on VLP assembly. nsTEM confirmed no VLPs were seen in the presence of MTX, Figure 4.3. This effect was seen with MTX present at approximately three times the K_D for TR, as determined by Chapter 3, Figure 3.11. Both native MS and non-denaturing agarose gels showed that MTX inhibits the formation of higher-order, TR-mediated assembly species, Figure 4.6. Native MS showed no products over 67 kDa, suggesting that MTX effects occur early in the assembly mechanism. After evidence by native MS that MTX did not bind to either MS2 CP or CP_2 it was inferred that MTX exerts its anti-assembly effect(s) via RNA binding.

In gRNA-mediated MS2 reassemblies, titrating in MTX lowers the amount of reassembled VLPs *in vitro*, Figure 4.14. Quantification of reassembled VLPs shows a clear trend, increasing MTX concentration decreases the number of fully reassembled VLPs formed and increases the number of partially assembled VLPs, Figure 4.13. MTX binding to RNA PSs supports their use as anti-assembly viral drug targets, as outlined in Figure 4.14. PSs are sequence specific and this has been shown in other virus systems (Ford *et al.* 2013, Shakeel *et al.* 2017), not only this but the nucleotide spacing between them is important too (Patel *et al.* 2017). There are multiple SLs formed in the MS2 genome that do not function as PSs (Dai *et al.* 2016). RNA secondary structures form as RNA is transcribed *in vivo*, over time their structure can change to form a stable, long range interaction with other nucleotide sequences on an inter- or intramolecular level. How PSs interact with each other in three-dimensional space is important (Twarock *et al.* 2018), any molecule that can interfere with the formation of secondary structure motifs, prevent nascent folding or refolding of RNA could be used as a viral defence.

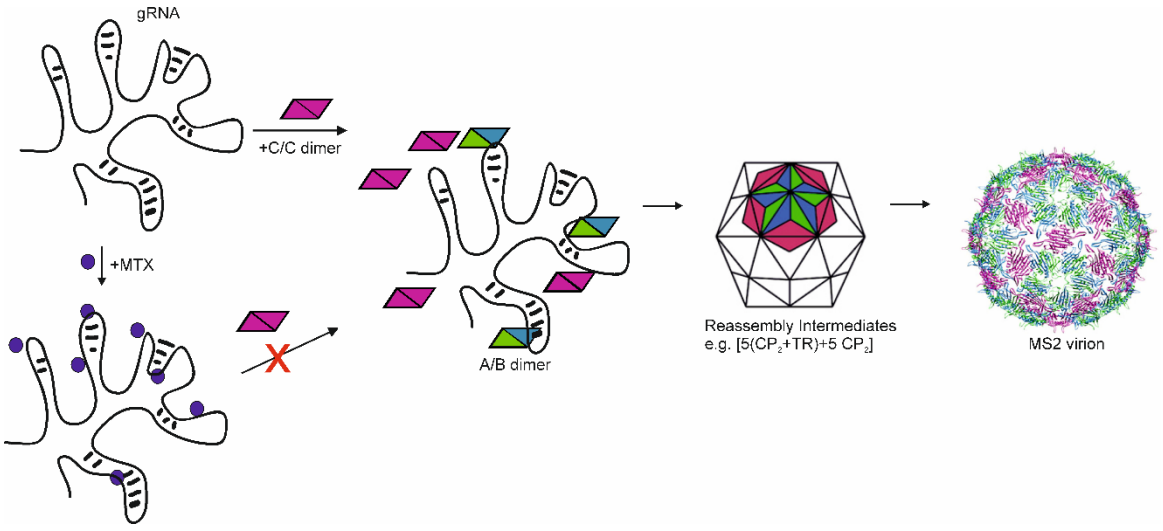


Figure 4.14 Schematic demonstrating the effect of MTX on the reassembly of gRNA-mediated MS2. gRNA incubated in the presence of MTX loses its ability to bind MS2 CP_2 and no assembly intermediates are formed.

To follow this hypothesis Chapter 5 investigates whether MTX inhibits phage products *in vivo.*

Chapter 5

5. Probing whether MTX Inhibits MS2 Production *in vivo*

This page has been left intentionally blank.

5.1. Establishing a Quantitative Assay of MS2 Infectivity

E. coli C3000 cells are commonly used hosts of MS2 (van Duin 1988). Cell growth is routinely monitored using absorbance at 600 nm, grown in minimal LB supplemented with glucose, thiamine and magnesium chloride (Jain 2009). Here, growth assays were performed aseptically in a low-volume, high-throughput 96-well plate format, monitoring at 650 nm rather than 600 nm due to limitations of equipment available at the time. Using alternative instrumentation appendix 8.4 compares growth curves monitored at 600 and 650 nm and calculates an average difference of -0.14 +0.07 O.D, showing that using either wavelength was suitable for monitoring cell growth. The test ligands described in Chapter 3 (daunomycin, MTX, tylophorine, DMA, A3 and SC19220) were added to C3000 starting cultures and their effects on cell growth monitored. Experiments were carried out at three concentrations: 16.7, 33.4 and 66.7 µM ligand. Figure 5.1 shows the resultant absorbance plots for the highest concentration only.

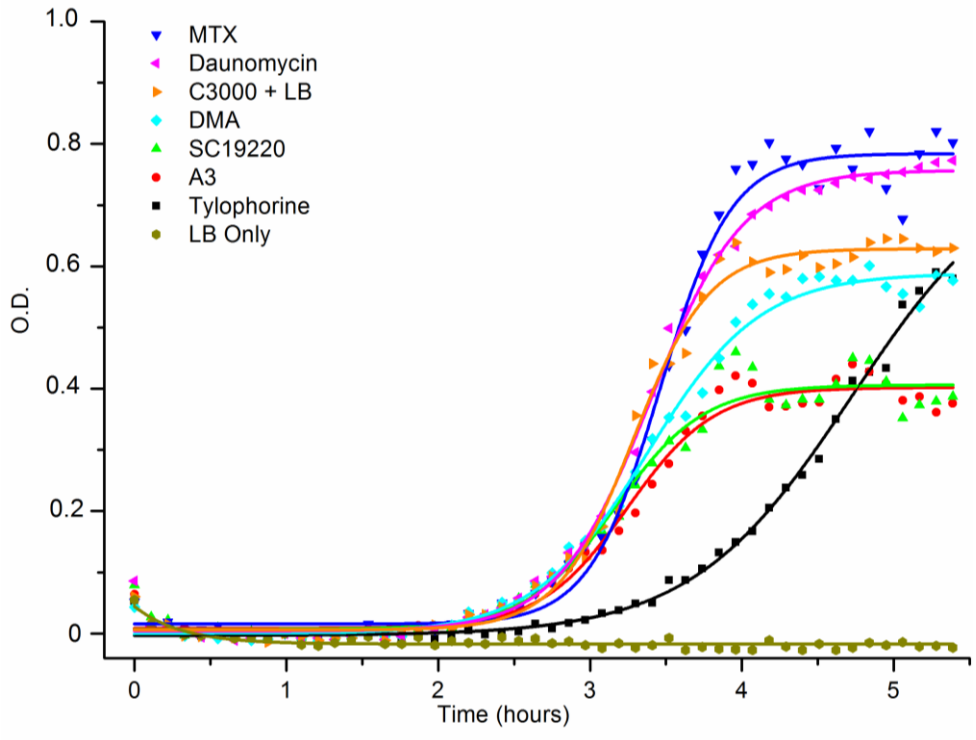


Figure 5.1 E. coli C3000 growth curves monitored by absorbance (O.D.) at 650 nm in the presence of 66.7 µM ligand. Control sample is C3000 in supplemented LB (blue triangle), here and throughout curves are recorded as single experiments in a 96-well plate format at 37°C with rotary shaking at 200 rpm, unless stated otherwise.

At the lowest concentration a similar result was shown, daunomycin and MTX enhanced cell growth as measured by light scattering. When compared to cells grown only in supplemented minimal LB, the addition of A3, DMA, SC19220 and tylophorine showed a reduction in cell density, suggesting that they are cytotoxic and unsuitable for further *in vivo* work. MTX and daunomycin resulted in a higher cell density, possibly due to the chemical structures of the ligands providing carbon/nitrogen source. MTX shows anti-assembly properties for MS2 *in vitro* (Chapter 4), so cells were grown in a wider concentration range in this ligand to check for toxicity. C3000 cells were grown overnight to allow all cultures to reach a stationary phase of cell growth and dilutions (100 µL) were plated onto minimal LB/agar to quantitate colony forming units. Figure 5.2 illustrates that over the concentration range tested, MTX has no effect on C3000 cell growth or viability.

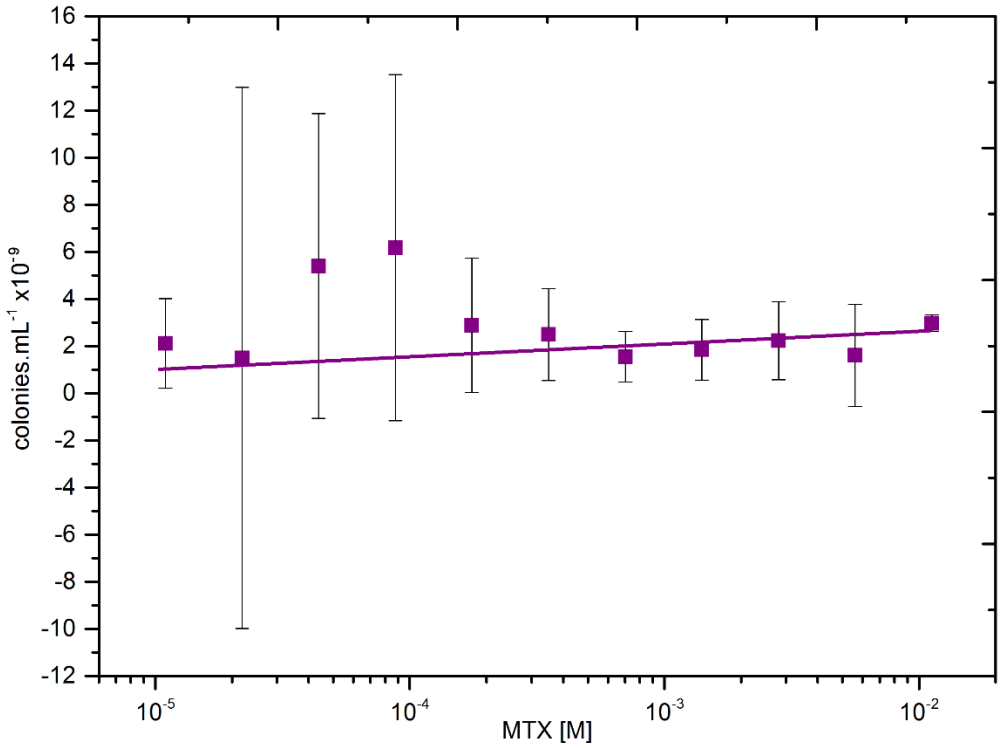


Figure 5.2 Number of E. coli C3000 colonies following overnight growth in supplemented LB in the presence of MTX. The average value of C3000 without MTX was calculated at 2.88 ± 2.50 x 10^8 colonies.mL^{-1}. All were repeated in triplicate. MTX has no detrimental effect on C3000 cells over this concentration range.

Using the same experimental set-up as Figure 5.1, C3000 cells were infected with bacteriophage MS2. The multiplicity of infection (MOI) of MS2 infections is usually <10 (Lima 2004). C3000 cells

were grown to +0.5 $O.D_{650}$ and infected with MS2 at 5 – 12 MOI (MS2 (5.1×10^9 $PFU.mL^{-1}$). Next, C3000 cells were infected with MS2 at OD_{650nm} = 0.5 Here they are held stationary at 37°C, 15 minutes at 0 rpm to allow MS2 adsorption to pili to occur. After this, plates are returned to 200 rpm and monitored at OD_{650nm}. Successful infections caused cell lysis, as indicated by a drop in 650 nm absorbance. Increasing the MOI shows an increase in cell lysis. An MOI of 10 was selected for all subsequent experiments.

5.2. Quantitating the Effect(s) of MTX on MS2 Infectivity

To test the effect(s) of MTX on MS2 infectivity, assays were set up as described in Section 5.1 but with the addition of MTX 15 minutes post MS2 infection. Measuring OD is only used as a visual check for cell lysis, final viral titres are measured using plaque assays. To quantitate this, two hours post infection lysozyme [3.5 μM] and EDTA [25 mM] were added to each reaction, allowing for complete release and desorption of virions from C3000 cells and the plaque forming units per mL ($PFU.mL^{-1}$) calculated. Cell lysates were purified by size exclusion into 15 mM Tris-HCl, 100 mM NaCl pH 7.6 (Illustra NAP™ 5 columns, GE Healthcare). This removed cell debris and excess lysozyme so plaque assays could be prepared. Minimal LB/agar plates were overlaid with C3000/agarose, gently mixed with separate dilutions of each purified sample. Figure 5.3 demonstrates the effect of MTX on MS2 and how it reduces its infectivity.

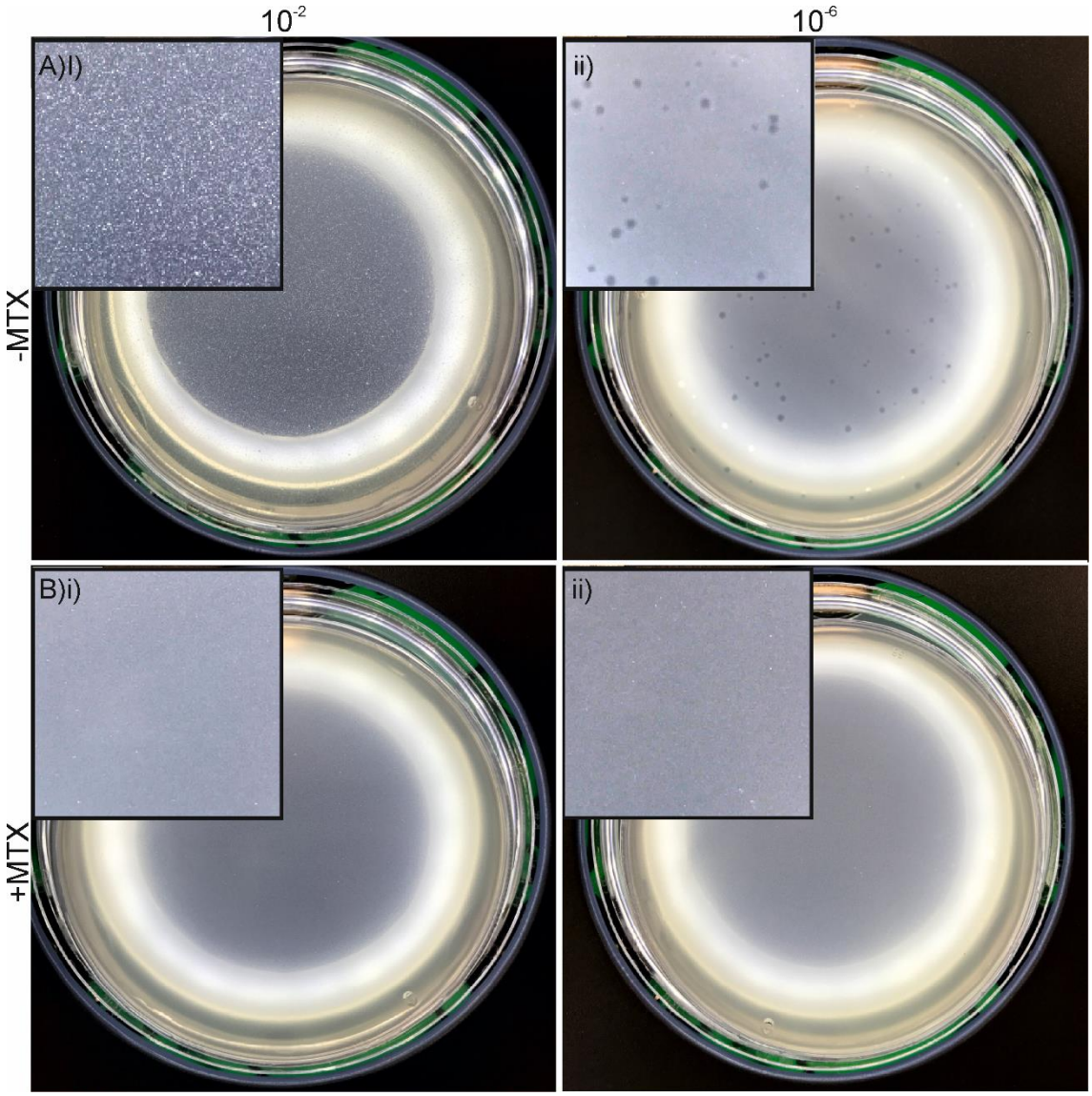


Figure 5.3 Plaque assay demonstrating the effect of MTX on MS2 infectivity. Supplemented minimal LB/agar plates overlaid with C3000/agarose mixed with purified MS2 from infection assay. A) In the absence of MTX. B) In the presence of MTX. Samples are diluted to i) 10^{-2} and ii) 10^{-6}. Inset At 2X magnification.

5.3. Testing MTX's Anti-Viral Effect on Other Bacteriophages.

To see if MTX showed anti-viral properties against other bacteriophages, GA and Qβ were tested using the same format. Like MS2, these bacteriophage have coat proteins that recognise and bind specific stem-loop (SL) structures in their gRNA (Horn 2005, Gott 1991). The experiment was repeated as before comparing the $PFU.mL^{-1}$ at each MTX concentration, results are shown in Figure 5.8.

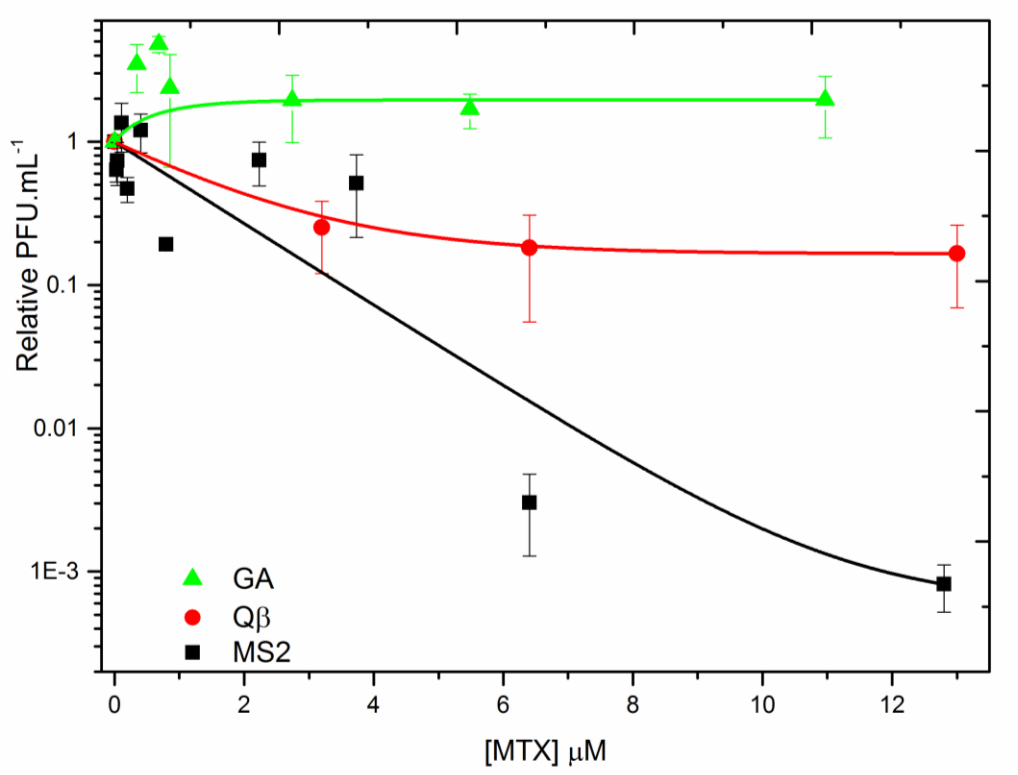


Figure 5.4 Comparing the effects of MTX on the viral titre of MS2, GA and Qβ. After infecting with bacteriophage, C3000 cells were treated with MTX. Experiments were repeated in triplicate in a 96-well plate format at 37°C with rotary shaking at 200 rpm. The PFU.mL^{-1} was determined using plaque assays.

In the presence of MTX, bacteriophage MS2 has the most reduced viral titre. At concentrations over 4 μM the PFU.mL^{-1} is reduced by approximately 100-fold, dropping further to 1000-fold at 12.8 μM. Qβ showed MTX caused a reduction in viral titre when infecting C3000 cells at the same MOI. Interestingly, GA showed a slightly higher titre when treated with MTX. These results suggest that the inhibitory effects of MTX are not completely specific to MS2.

5.4. Characterising the Effect of MTX on MS2 Infectivity

In view of the results with GA and Qβ experiments were carried out to test which processes in the MS2 lifecycle are being affected by MTX. The infection assay was repeated and the level of MS2 CP determined using a Western Blot. An aliquot of the crude cell lysate from each MTX concentration (10 μL) was analysed by Western Blot analysis (Figure 5.5).

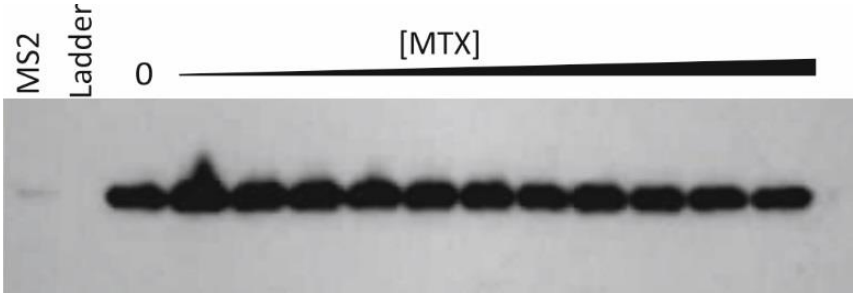


Figure 5.5 Western Blot of cell lysates following MS2 infection of C3000 cells in the presence and absence of MTX. Lane 1 = inoculum MS2. Lane 2 = MWt. Ladder. Lane 3 = "0" represents control with no MTX present. Lanes 3-13 = 0.5-1.1 µM MTX. Samples run initially on a 12% SDS-page, transferred to Whatman Protran 0.2 µm, and visualised using chemiluminescent HRP substrate.

All lanes showed a large increase in CP expression compared to the inoculum MS2, as expected. There is a 12% decrease in CP band intensity at the highest concentration of MTX, when compared to control without. See Materials and Methods for full protocol and Appendix 8.5 for densitometry lane analysis. Since CP production is lower following MTX addition the gRNA was assayed. gRNA was extracted from cell lysates in the presence and absence of MTX as above (RNAeasy Kit, Qiagen) and analysed using Reverse Transcription (RT) of a 620 nt fragment within the replicase gene and followed by 10 rounds of PCR (RT-PCR), Figure 5.6. See Chapter 2 for details of the primers and enzymes used, protocol adapted from (Shin and Sobsey, 2003).

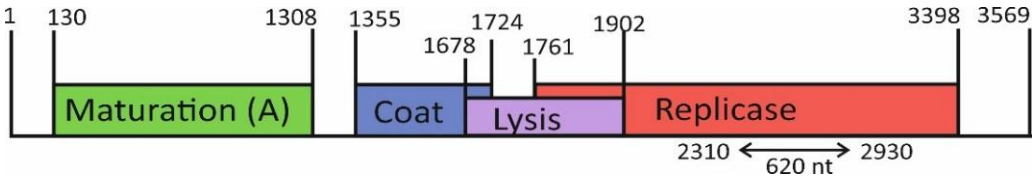


Figure 5.6 Schematic of MS2 genome. The arrows show the 620 bp region used to monitor gRNA replication.

The RT-PCR products were analysed by denaturing agarose gel electrophoresis, Figure 5.7. The major band at approximately 600 nt is the selected region amplified for this experiment. The lower bands are unincorporated primers.

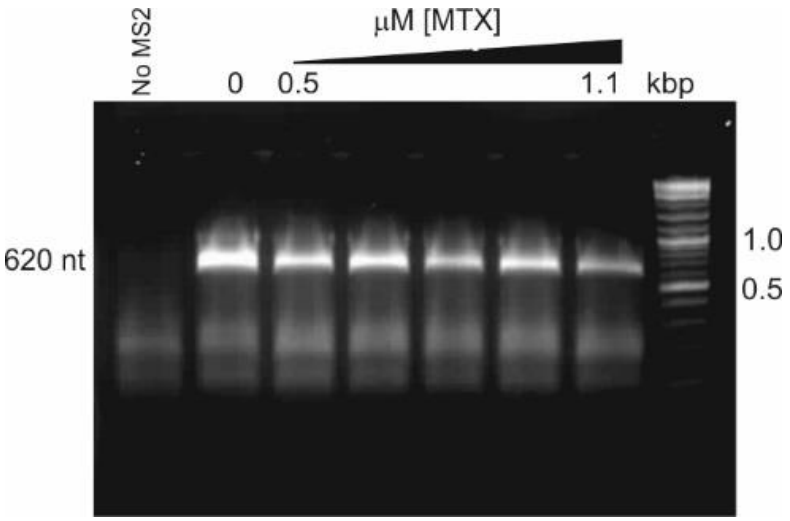


Figure 5.7 RT-PCR product of extracted RNA from MS2 infection in the presence and absence of MTX selecting for 620 bp region of MS2 genome. "0" represents control with no MTX present. "No MS2" is a sample from uninfected C3000 cells. Lanes 3-7 = 0.5 – 1.1 µM MTX. "kbp" corresponds to kilo base pair ladder.

There is a reduction in RT-PCR band intensity when comparing gRNA extracted from MS2 infections in the absence and presence of MTX, consistent with an MTX concentration dependent decrease in RT-PCR product. These experiments show that MTX does not prevent CP translation or gRNA replication, however they appear to be produced at lower levels with increasing MTX concentration.

5.5. MS2 Infectivity after pre-incubation with MTX

In prior experiments C3000 cells were infected with MS2 when they had reached OD_{650nm} = 0.5, and held stationary at 37°C, 15 minutes at 0 rpm to allow MS2 adsorption to pili to occur. After this period of incubation, MTX was added to cell cultures to look for an anti-viral effect. To investigate whether MTX interacts with cells or phage directly, MS2 was pre-incubated with MTX and added at OD_{650nm} = 0.5. Phage was pre-incubated with a higher concentration of MTX so that cell cultures had equivalent final concentrations. Results show there is little or no difference between the two conditions, Figure 5.8.

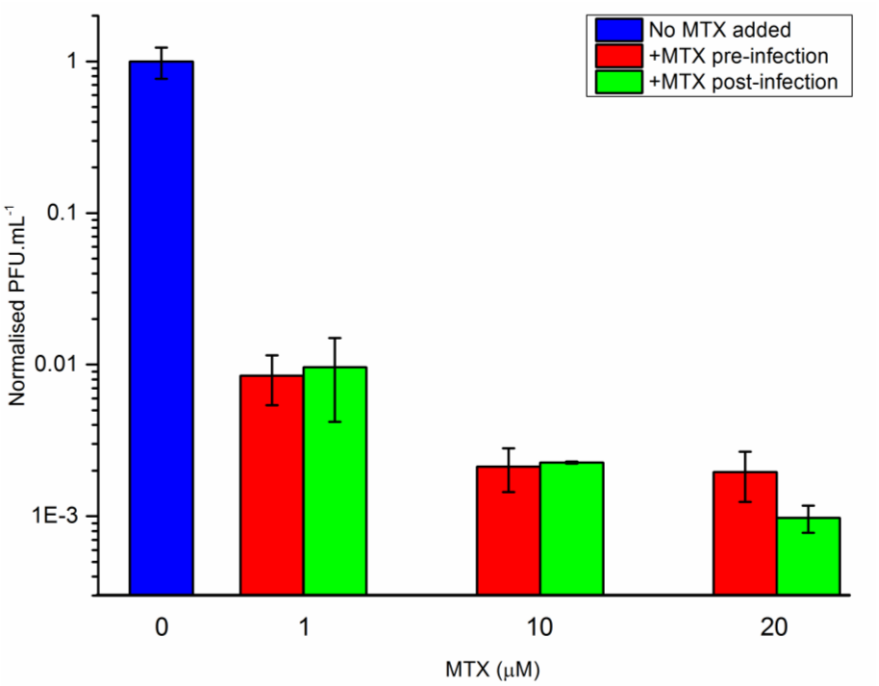


Figure 5.8 Comparing titres of phage purified from C3000 cells infected with MS2 pre-incubated with MTX (red) against infections inoculated with MTX 15 minutes post infection (green). PFU.mL^{-1} determined by plaque assay.

5.6. Characterisation of MTX-Resistant MS2

To study the effect of MTX on MS2, infections were repeatedly passaged in the presence and absence of MTX. Using 10 mL cultures and leaving infections overnight generated enough purified progeny virus for re-infection in the next passage. In the presence of MTX there was a reduction in infectious phage, it was not possible to concentrate recovered phage, or infect at a lower cell count so phage were re-infected into next passage at a reduced MOI. MOIs were kept equivalent for infections in the presence and absence of MTX and performed in triplicate and compared in Figure 5.9. Results from this assay showed MS2 passaged in the presence of MTX has a reduced titre which recovers by passage number 3. To examine the escape mutants gRNA sequences from MS2 inoculum stock and passage number 3 (presence and absence of MTX) were compared using Next-Generation sequence analysis (NGS).

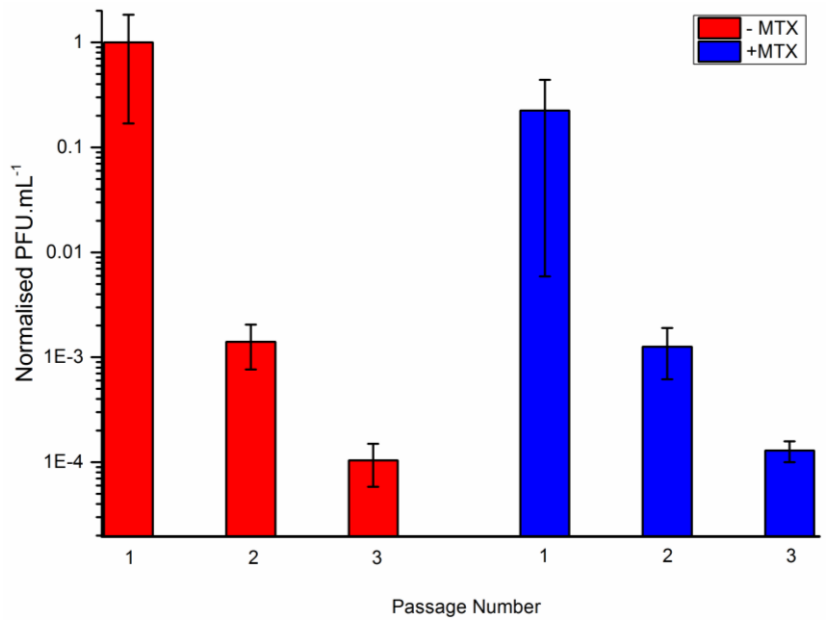


Figure 5.9 MS2 titres after passaging in the presence and absence of MTX [30 µM]. PFU.mL^{-1} determined by plaque assay. For passages 1, 2, 3 M.O.I. = 10, 1 and 1 respectively. Results normalised to allow for comparison to round 1, infecting with wtMS2 in the absence of MTX.

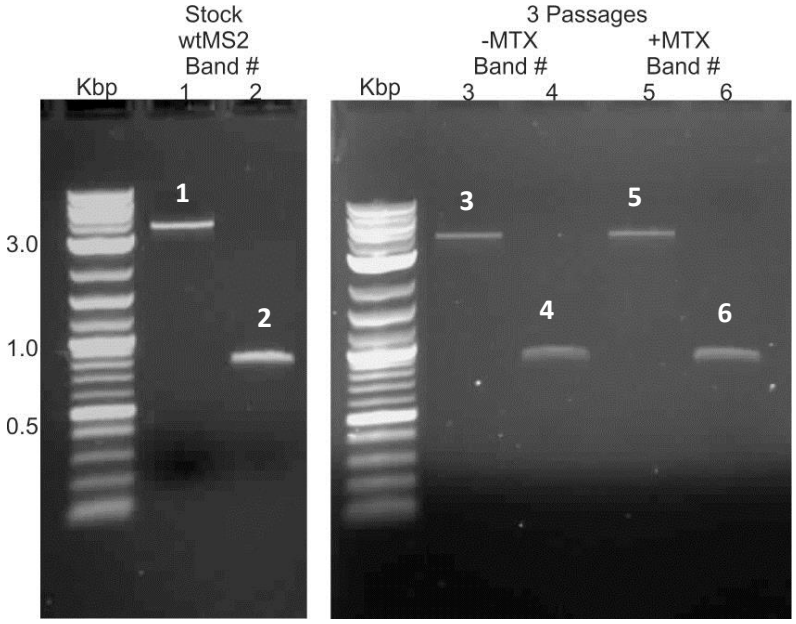


Figure 5.10 Purified bands of cDNA from PCR. cDNA copy of full length MS2 genome is 3569 nt. Denaturing 1% agarose formaldehyde/MOPS gel run in 1X MOPS, 90V for 1 hour.Visualised using EtBr. All samples (Bands # 1-6) were sent for NextGen sequencing.

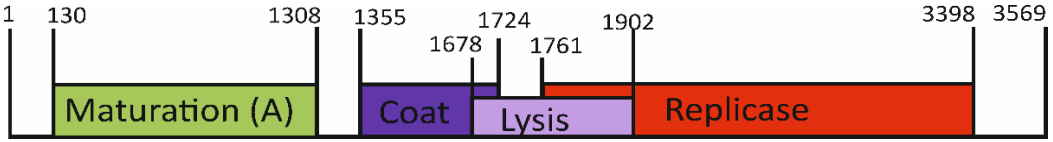
RNA was isolated from purified cell lysate (100 µL) using an RNA isolation kit (Qiagen®, RNAEasy MiniKit). Extracted gRNA was reverse transcribed using a high fidelity reverse transcriptase (Life Technologies®, Superscript III) and amplified using 30 rounds of PCR (KAPA BioSystems Inc., KAPA2G). For full details and primer sequence information see Section 2.1.7. The PCR products

were purified (Qiagen®, Qiaquick PCR Purification Kit) and run on a denaturing agarose gel, Figure 5.14. All PCR products from MS2 infections showed two major species (~0.8, ~3 kb), these bands were all excised for identification (Qiagen®, Qiaquick Gel Extraction Kit). Following a standardised protocol cDNA copies of the viral RNA were processed by St. James' University Hospital, Leeds, UK. Illumina TruSeq Stranded Total RNA Next Generation libraries were generated and sequenced on an Illumina MiSeq Sequencer to create 150 bp paired-end sequence data. The data was aligned to MS2 bacteriophage genome (NCBI NC_001417.2) using the Burrow-Wheels Aligner (Li *et al.* 2009). Samtools was used to sort the aligned data by position before converting the data to .mpileup formatted files from which the variants were identified using the .mpileup2snp and .mpileup2indel methods of Varscan (Koboldt 2009, 2012). The average read depth at each nucleotide position was >100,000 for each sample. This alignment work was done by Dr Ian Carr, St. James' University Hospital, Leeds, UK. For the full protocol see Section 2.12. Excised bands 1, 3 and 5 showed full length MS2 cDNA genomes, bands 2, 4, and 6 were 800 nt truncated genome fragments therefore sequence data is not shown. Bands 1, 3 and 5 are labelled as MS2 inoculum stock "*wt*MS2 No Passage", MS2 passaged three times "*wt*MS2 3X Passage" and MS2 passaged three times in the presence of MTX "MS2 + MTX 3X Passage", respectively. All three were shown to contain >98.3% MS2 cDNA and >0.5% PhiX174. PhiX174 is added to all Illumina sequencing as an internal control of the sequencing run. There are other species present which have been omitted from the analysis (<0.4%). For consensus sequence information and all identified cDNA fragments see Appendix Table 8.9.

Results from NGS show differences between samples analysed and the NCBI reference genome (NCBI 1417.2), see Table 5.1. Full length cDNA sequences with nucleic acid mutations highlighted can be found in Appendix Table 8.10. In Table 5.1 nucleic acid changes are shown in red (%), bold highlights no nucleic acid change between the reference genome and sequenced data, but are included for comparison between the samples.

Nucleic acid insertions and deletions were detected by NGS. In all three samples insertions were found at nts 2007 (97-99%) and 3450 (95%). Deletions were seen at nts 1999 (84-88%) and 3461 (97%). Unique to *wt*MS2 were insertions at nt 2008, 2009 and 2010 occurring at levels below 5%. wtMS2 also showed a deletion at nt 3642 (98%). Full information on identified nucleic acid insertions and deletions is available in Appendix Table 8.11. Some nucleic acid changes are silent mutations, causing no mutations at the amino acid level (**). See Appendix Table 8.12 for maturation protein (MP), CP and replicase amino acid sequences, translated using ExPASy Translate (SIB, Bioinformatics Research Tool). All amino acid mutations identified are shown in Figure 5.11. MS2 passaged 3X in MTX shows a unique mutation E89D (100%) at the NTD of MP, Figure 5.11, panel A. Also in the MP a mutation at A176T is present in both wtMS2 and MS2 passaged 3X, occurring 90% and 100% respectively. In the CP a mutation at A25S (9%) is located at the 3-fold axis between A, B and C CP subunits, Figure 5.11, panel B. Replicase shows the most amino acid mutations. Present in all samples are mutations P133S (≥99%) and G81E (49-67%), other mutations occur ≤4% but all are shown in Figure 5.11, panel C.

*Table 5.1 Results of NGS. Three samples shown are MS2 inoculum stock (wtMS2 No Passage), MS2 passaged three times (wtMS2 3X Passage) and MS2 passaged three times in the presence of MTX (MS2 + MTX 3X Passage). All are compared to MS2 NCBI ref NC_001417.2. Red text suggests there is a nucleic acid/amino acid mutation. ** denotes silent mutation at the amino acid level. Black text shows there is no mutation but are included as a comparison between the samples. MS2 + MTX 3X Passage shows a unique 100% mutation at nucleotide 393, G to T, amino acid E89D. Appendix Table 8.10 show full nucleic acid sequences.*



Sample	Nucleotide	393	652	1260	1391	1408	1547	2002	2004	2159	2160	2165	2426	2429	2591	2972	3092	3162	3477
		G	G	G	G	A	G	G	G	T	C	G	T	C	A	G	G	G	G
wtMS2 No Passage	Nucleotide (%)	G	A(90)	T(1)	T(1)	A	G	A(49)	C(4)	C(99)	T(99)	T(1)	C(98)	T(100)	T(100)	G	T(1)	T(1)	C(89)
	Amino acid mut.		A/T	M/I	**			G/D	D/H	**	P/S	M/I	**	**	**		**	G/W	
wtMS2 3X Passage	Nucleotide (%)	G	A(100)	G	G	G(1)	A(9)	A(67)	C(3)	C(100)	T(100)	G	C(99)	T(100)	T(100)	T(5)	G	G	C(100)
	Amino acid mut.		A/T			N/S	**	G/D	D/H	**	P/S		**	**	**	**			
MS2 + MTX 3X Passage	Nucleotide (%)	T(100)	G	G	G	A	G	A(66)	C(3)	C(100)	T(100)	G	C(99)	T(100)	T(100)	G	G	G	G
	Amino acid mut.	E/D						G/D	D/H	**	P/S		**	**	**				

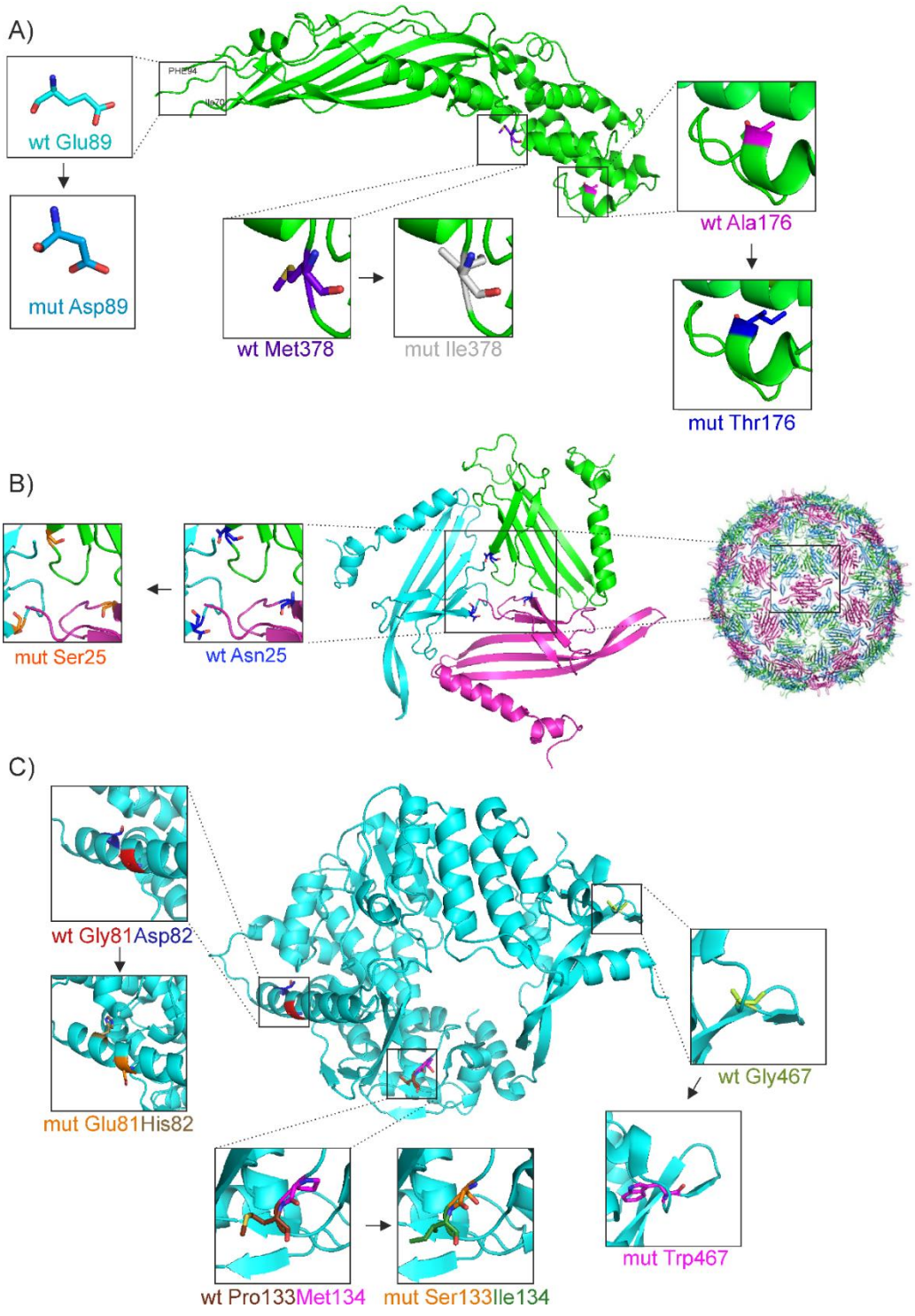


Figure 5.11 MS2 protein amino acid mutations identified by NGS Analysis. A) Maturation protein mutation in amino acids E89D, A176T, M378I. Note there is no known structure for region 70-94 (PDB 5TC1). B) Coat protein mutation S25R (PDB 2MS2) and C) Replicase amino acid mutations G81E, A82H, P133S, M134I and G467W. There is no structure for MS2's replicase, protein shown is replicase subunit of Qβ holoenzyme complex (PDB 4R71). Drawn in PyMOL (PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.).

Nucleotides 387-419 are in a stable hair pin loop as predicted experimentally (Fiers *et al.* 1975, Koning *et al.* 2016). This secondary structure, determined by chemical modification, identifies nucleotides that are likely to be base paired based on their chemical reactivity. Altering nucleotide at position G393U causes a change in the base pairing interactions in the local environment, MFOLD analysis of nt 387-419 predicts the change in secondary structure of the RNA, Figure 5.12. It should be noted also that nucleotides 395-410 have been shown experimentally to bind to protein *in virio*, termed PS P8 (Rolfsson *et al.* 2015).

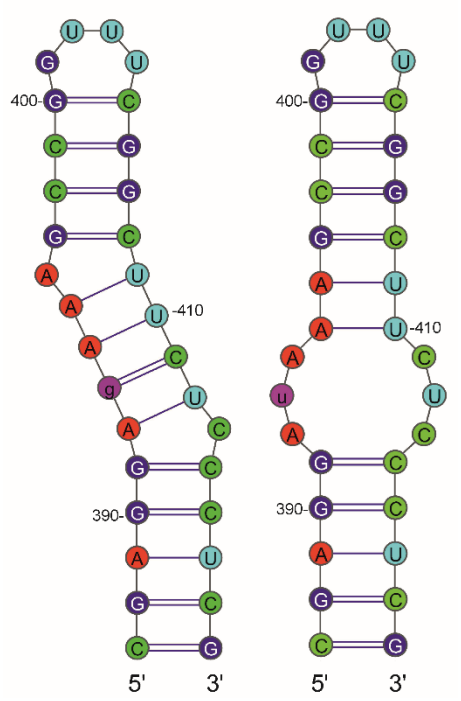


Figure 5.12 Secondary structure of MS2 nucleotides 387-419. Left 5'wt as determined by Fiers et al. 1975 and Koning et al. 2016. Right 5'mutant. G393U (purple) mutation shows a loss in the bulged adenosine base at position 10. Secondary structure of mutant predicted by MFOLD (Zuker, 2003), drawn in Varna (Darty, 2009).

To compare relative binding affinities, RNA stem loops from Figure 5.12 were incubated with MTX and subjected to negative mode ESI-MS, Figures 5.13-14. Each RNA SL shows experimental masses that correlate with the expected mass as calculated by OligoAnalyzer 3.1 (Integrated DNA Technologies Inc.). 5'*wt* estimated MWt. 10,231.2 $g.mol^{-1}$, experimental MWt. 10,231.4 $g.mol^{-1}$. 5'mutant estimated MWt. 10,192.1 $g.mol^{-1}$, experimental MWt. 10,191.4 $g.mol^{-1}$. When incubating at a ratio of 1:0.1 (RNA:MTX) up to three MTX molecules bind each SL. At 1:1 ratio there is no free-RNA in solution, both 5'*wt* and 5'*mutant* can bind up to 6 MTX per RNA SL. Peak integration data for the lowest concentration of MTX tested shows there is less 5'*wt* (24.0%)

free in solution compared to 5'*mutant* (49.5%), suggesting MTX has a higher affinity for 5'*wt*. For full peak integration see Appendix Table 8.16.

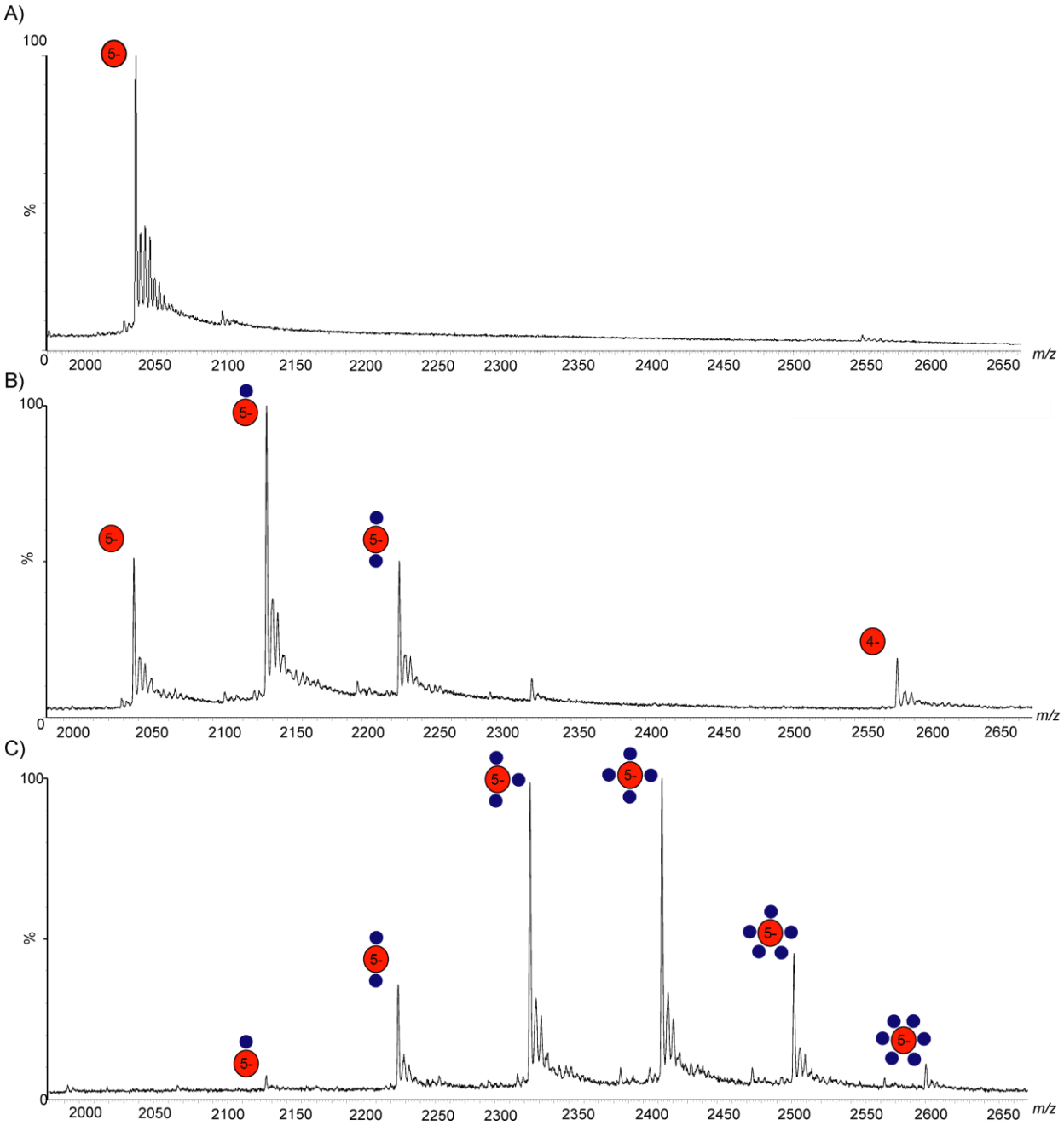


Figure 5.13 Negative ion mode MS of 5'wt RNA [1.6 µM] shown binding to MTX. A) RNA only. Estimated MWt. 10,231.2 g.mol^{-1} (Integrated DNA Technologies Inc., OligoAnalyzer 3.1), experimental MWt. 10,231.4 g.mol^{-1}. Major negative charge state for this RNA is 5-. B) Molar ratio 1:0.1, RNA:MTX. C) Molar ratio 1:1, RNA:MTX. Minor peaks at 2550 m/z are due to presence of 4- ion. For full peak integration see Appendix Table 8.16.

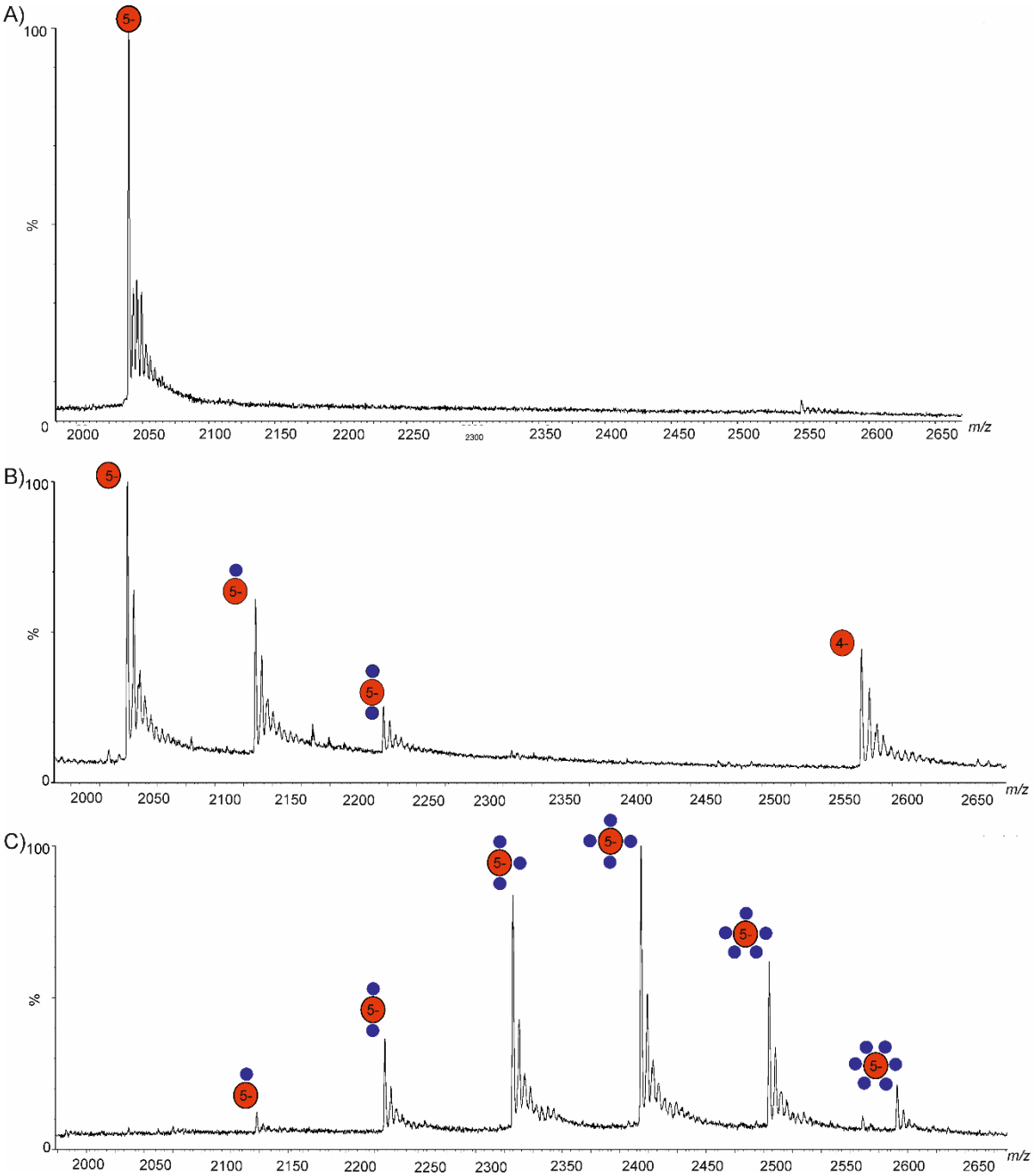


Figure 5.14 Negative ion mode MS of 5'mutant RNA [1.6 µM] shown binding to MTX. A) RNA only. Estimated MWt. 10,192.1 g.mol^{-1} (Integrated DNA Technologies Inc., OligoAnalyzer 3.1), experimental MWt. 10,191.4 g.mol^{-1}. Major negative ion for this RNA is 5-. B) Molar ratio 1:0.1, RNA:MTX. C) Molar ratio 1:1, RNA:MTX. Minor peaks at 2550 m/z are due to presence of 4- ion. For full peak integration see Appendix Table 8.16.

The secondary structures of *wt* and *mutant* gRNA from *in vitro* experiments and *in vivo* assays can be compared using hydroxyl radical X-ray footprinting (XRF) (Adilakshmi *et al.* 2009).

5.7. Probing gRNA in Virions using XRF.

Hydroxyl radicals can be generated *in situ* by photolysis of water (Hayes *et al.* 1990), causing RNA strands to break based on the solvent accessibility of the ribose ring (Tullius *et al.* 2009). The secondary structure of RNA and any tertiary interactions with proteins will affect the level of cleavage, giving structural data at nucleotide resolution. To deduce gRNA secondary structure in the presence and absence of protein; *wt*MS2 and extracted gRNA were compared, looking specifically at the nucleotide region encompassing TR, Figure 5.15. See Chapter 2, Section 2.12 for full XRF protocol and sample preparation. Work shown was done in collaboration with Dr Rebecca Chandler-Bostock (University of Leeds, U.K.).

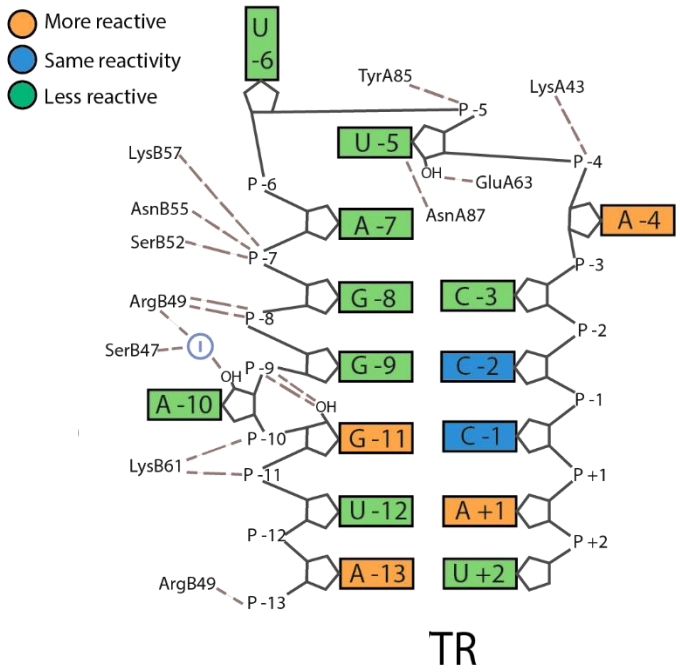


Figure 5.15 Hydroxyl radical footprint of TR (nts 1747-1763) shown as a ratio of wt virion gRNA to extracted gRNA. Dashed lines indicate the inter-molecular contacts seen in X-ray crystal structure (Valegård et al. 1997). Numbering is relative to AUG start codon. Figure provided by Rebecca Chandler-Bostock (University of Leeds, U.K.).

This experiment shows in this section of gRNA the majority of nucleotides are less reactive (green) towards hydroxyl radicals in the virion, than free gRNA. The reduction in reactivity correlates with the fact that nucleotides A-7, A-10 and U-5 are known to be important in the interaction with CP (Romaniuk *et al.* 1987).

A similar experiment focusing on nts 387-418 at the 5' end compares gRNA extracted from *wt* virions (extracted RNA), *wt* MS2 (RNA in virion) and *in vitro* gRNA-mediated reassembled MS2 (reassembly RNA), Figure 5.16. It has been planned to collect equivalent data for mutant MS2.

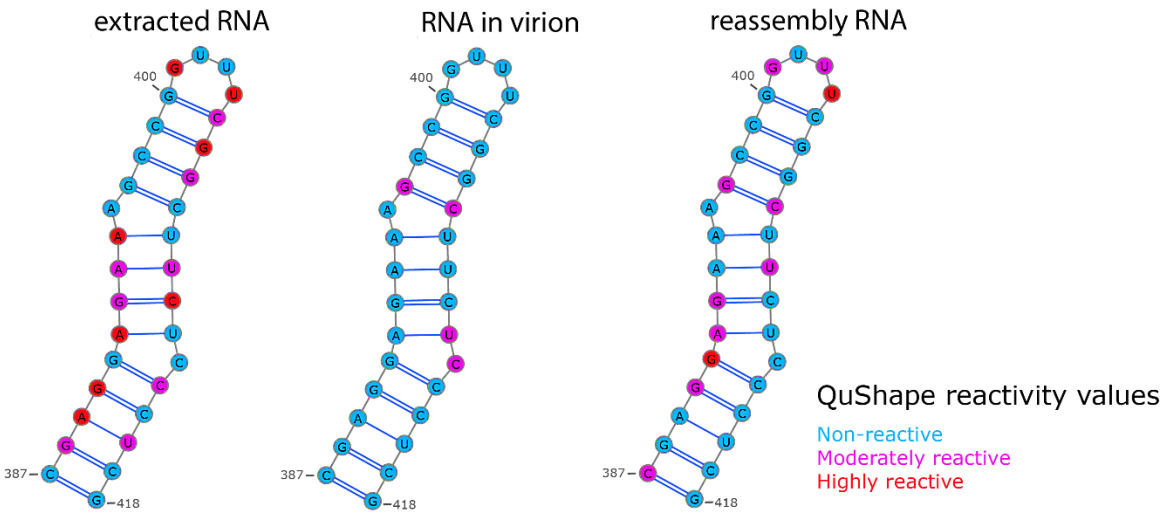


Figure 5.16 XRF reactivities and secondary structures around the putative MP binding site (nt 387-418) for extracted gRNA, virion and in vitro reassembled VLP. Figure provided by Rebecca Chandler Bostock, University of Leeds, U.K. Beamline exposure time 100 ms, samples compared relative to same gRNA with zero exposure.

Each RNA examined by XRF shows extracted RNA is more reactive than reassembled RNA, which is more reactive than in the RNA virion. As expected nucleotides that have more tertiary interactions are less reactive to hydroxyl radicals. This allows us to interpret the increase in reactivity seen for gRNA as evidence for interacting with proteins. Focussing on nucleotide G393, an increase in reactivity is shown in reassembly RNA when compared to RNA in virion, implying gRNA interacts with MP when in virion.

5.8. Discussion

This chapter optimised *in vivo* conditions for studying the inhibition of *wt*MS2 using RNA binding ligands. The effect of one ligand, MTX, showed a concentration dependent reduction in MS2 infectivity. There was a decrease in CP translation and RNA replication with increasing MTX concentration, however, virion production was never completely inhibited. Passage experiments, in the presence and absence of MTX, identified escape mutants. After 3 passages in MTX the escape mutant was at a higher titre relative to *wt*MS2.

Next Generation sequencing of the samples and referencing to NCBI reference genome (NCBI 1417.2) identified mutations across the whole genome. Comparing samples from passages in the presence and absence of MTX identified one unique mutation at G393U (100%). Nucleotides 387-419 (5'*wt*) are in a stable SL as shown experimentally (Fiers *et al.* 1975, Koning *et al.* 2016). MFOLD analysis of the mutated sequence shows G393U (5'*mutant*) causes a change in the base pairing interactions in the local environment, leading to the loss of an unpaired adenosine base. MS analysis of RNA:MTX titrations showed that 5'*wt* SL has a higher affinity towards MTX than 5'*mutant*.

Nucleotide 393 occurs in the MP coding region, within an RNA SL which binds the MP. The interaction between 393 and the MP is important in MS2 lifecycle; during an infection the gRNA–MP complex enters the cell (Shiba and Suzuki, 1981). This nucleotide mutation translates to an amino acid mutation E88D which is located in a flexible, unstructured region at the N-terminal domain, Figure 5.17. The latest cryo-EM structure of the MP *in situ* shows amino acid 89 is expected to sit in the β-sheet domain, protruding from the curvature of the CP shell (PDB 5TC1, Dai *et al.* 2016). Dai's model of MS2 shows RNA SL interacting with CP shell and MP, however it lacks detail of the SL encompassing nucleotide 393.

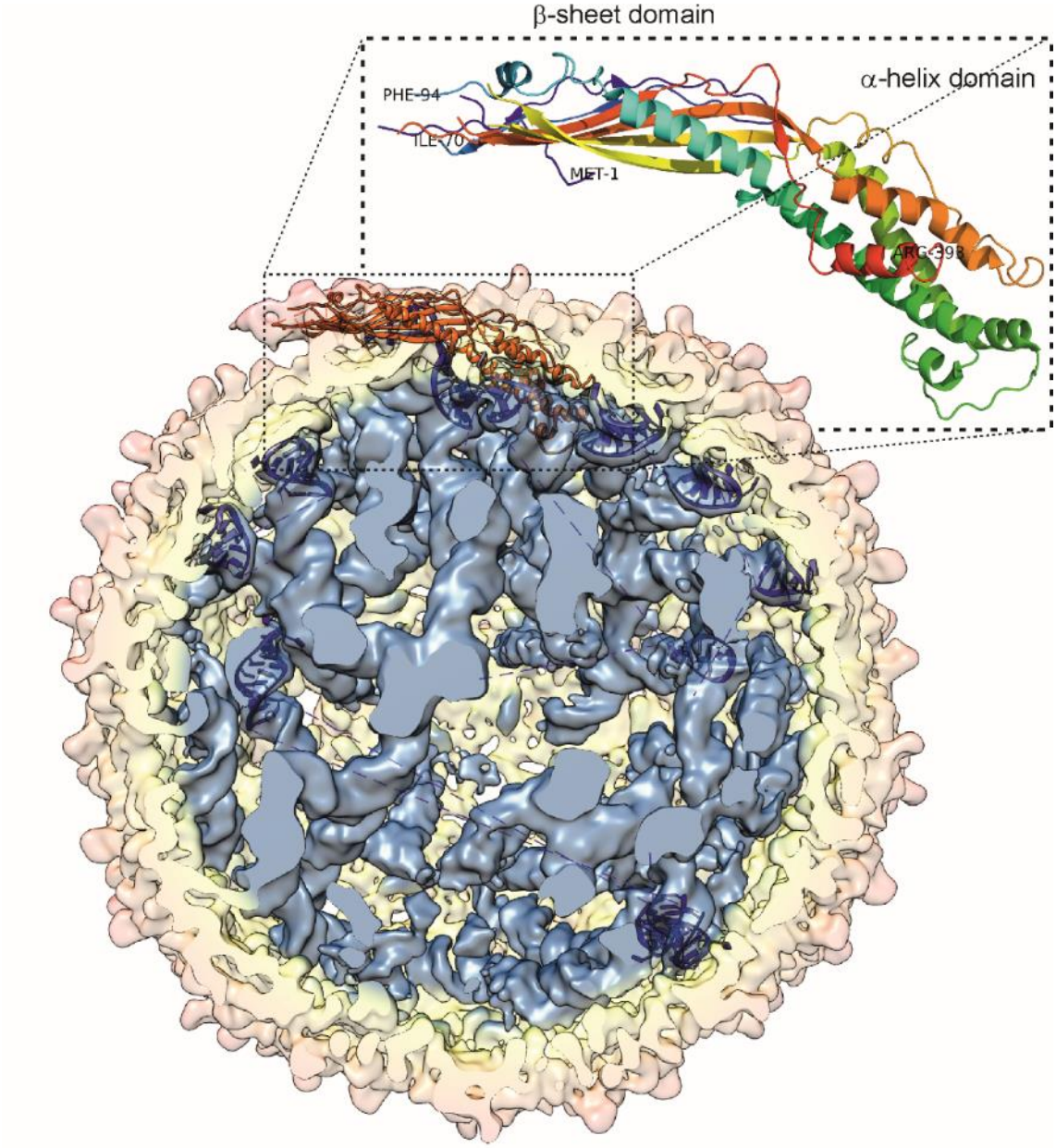


Figure 5.17 Structure of MS2 MP (orange) in situ as determined by cryo-EM. The β-sheet protrudes from the curvature of the coat protein shell (pink) and α-helix is buried within the interior, interacting with nucleic acid SLs (blue). Figure adapted from (Dai et al. 2016) drawn in Chimera (Pettersen et al. 2004). Expanded section shows high resolution structure of MP colour coded from CTD (red) to NTD (blue). Loop at NTD, Ile 70-Phe 94, highlights unresolved structure encompassing amino acid 89. Figure adapted from (Dai et al. 2016), drawn in PyMOL (PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.).

Koning *et al.* (2016) revealed an asymmetric EM reconstruction of MS2 bound to pilus at 4Å resolution. This model also generates proof that MP replaces one C/C CP_2 dimer in the capsid shell, allowing for MP to contact both gRNA and adsorb to pilus, Figure 5.18, panel A. Figure 5.18, panel B reveals how MP (yellow) interacts directly with gRNA (cyan), and identifies binding to the same 5' RNA SL reported before (Shiba and Suzuki, 1981; Rolfsson *et al.* 2015). Koning highlights the region within this SL that binds to MP (★), nucleotide 393 is also located in the same SL (★). G393U mutation changes the secondary structure, which may affect the SL's ability to bind MP, Figure 5.16.

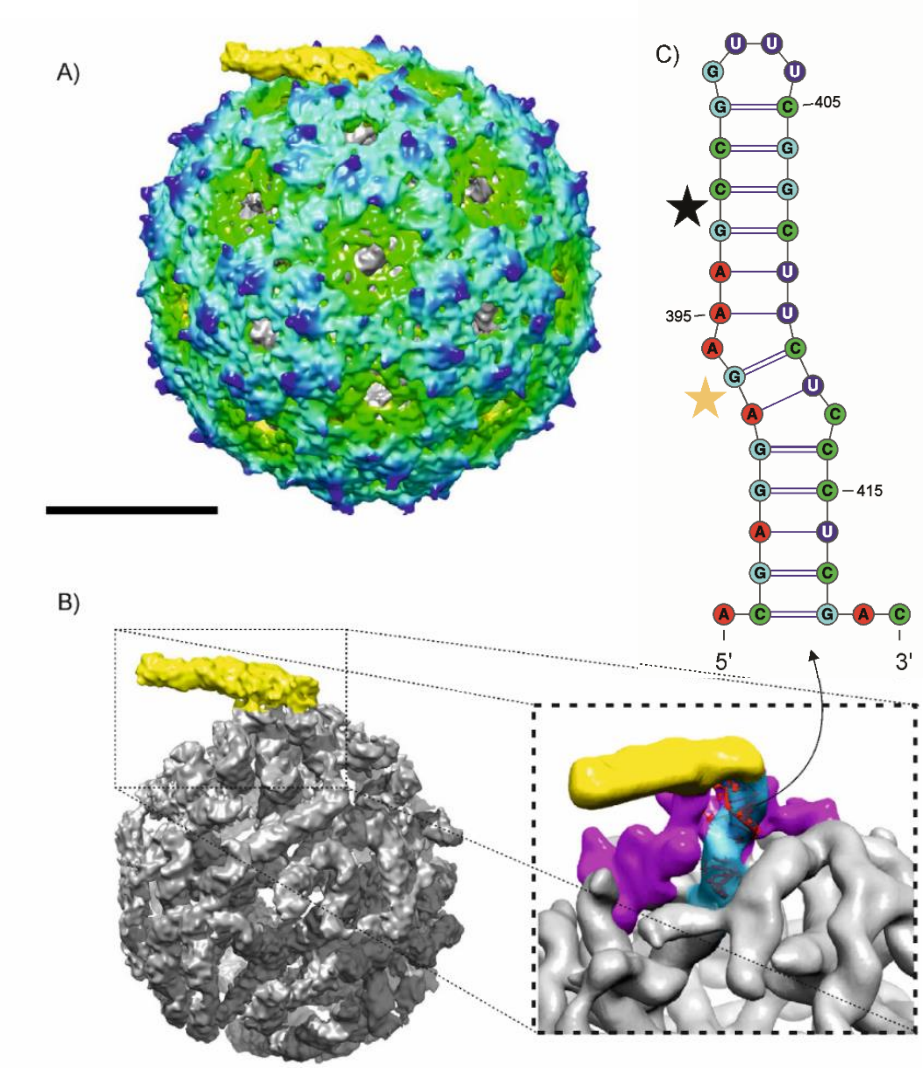


Figure 5.18 MS2 MP in situ adapted from Figures 1 and 2 in Koning et al. (2016). A) Asymmetric structure of MS2 (green-blue radially coloured) shows MP (yellow), which replaces one CP_2, scale bar is 100 Å. B) Inside CP shell showing the genome (grey), with expanded view highlighting neighbouring CP (magenta) and nucleotides 388-414 (cyan) with fit density (red). C) Predicted secondary structure of nts 386-421. ★MP binding site. ★Nucleotide 393.

The idea of using specific nucleic acid SLs as drug targets is not novel, but targeting ones involved in the structural formation of VLPs is. The results from this chapter support the hypothesis that targeting them does cause a drop in infectious titre. The original idea was to target the highest affinity PS, TR, but the results showed that MTX bound to SL 5'*wt* and caused a nucleic acid mutation at position 393. This region of RNA is very important in the lifecycle of MS2, and it was a promising result to see that this SL had been targeted. However, the ability of MS2 to form an escape mutant after just 3 passage experiments was a surprise, and suggests that there is a degree of vulnerability when using RNA SL as drug targets. The infectivity of mutant MS2 in the presence and absence of MTX is currently being studied *in vivo*; if the mutant form is only infectious in the presence of MTX then this would verify the suitability of testing with human pathogens. The results shown in this chapter do not explain whether MTX interacts with any other PSs in MS2 gRNA, or if these interactions caused any disruption in structural VLP formation. These should be assessed to generate a greater understanding of what is going on, there needs to be more structural clarification with regards to where and how MTX interacts with gRNA before it can be accepted as a drug target. This specific ligand has been identified as binding to bulged adenosine bases, if other molecules targeting other structural features (for example guanosine bulges) are used together, as a multi-pronged attack, they might have a better success rate with slower mutation rates.

This page has been left intentionally blank.

Chapter 6

6. Discussion and Future Work

This page has been left intentionally blank.

6.1. Discussion

There are currently a number of anti-viral drugs targeting virion adsorption, cell penetration, genome release, nucleic acid and viral protein synthesis and the assembly of virus-like particles (VLPs). The discovery of the evolutionarily conserved packaging signal (PS)-mediated assembly mechanism has been proposed as a novel anti-viral target and the work described here is the first to direct drugs towards the RNA stem-loops (SLs) involved.

Starting with a broad library, ligands were ranked based on their ability to bind RNA. Using bacteriophage MS2 as a model virus system, methods were identified for studying ligands binding to MS2's highest affinity RNA PS, TR. These assays identified the FDA-approved drug mitoxantrone (MTX) as a good binder, K_D = 3.71 ± 0.22 nM (Chapter 3, page 79). Given the NMR structure of TR (Borer *et al.* 1995) and MTX's affinity towards unpaired adenosine nucleotides (Zheng *et al.* 2009) this interaction makes sense. Deleting the adenosine bulge from TR increases the K_D 3-fold to MTX 1.11 ± 0.17 µM, showing its importance in this binding interaction (Chapter 3, page 81). Using this information MTX was tested further in *in vitro* assembly inhibition studies. MTX was shown to inhibit the assembly of TR-mediated MS2 VLPs (Chapter 4, page 98), and with increasing MTX concentration gRNA-mediated MS2 assembly reactions formed mal-assembled VLP intermediates, suggesting that the cooperativity of the individual PSs is compromised (Chapter 4, page 110). Multiple RNA SLs in MS2 gRNA have been identified as putative PSs, Table 6.1 (Dykeman *et al.* 2013, Rolfsson *et al.* 2015, Dai *et al.* 2016). Evidence of these by separate techniques verifies their importance in the architecture of the virion, and demonstrates how targeting conserved structural motifs is a viable anti-assembly route.

It was initially presumed that MTX was exerting it's *in vitro* anti-assembly activity by blocking key PS:CP interactions so *in vivo* experimental conditions were established to investigate further. MTX showed a concentration dependent inhibition of MS2 infectivity (Chapter 5, page 120).

Table 6.1 Comparing stem-loops identified within the MS2 genome using three different techniques. n/a indicates there is no matching sequence identified. Boxed area indicates the stem-loops containing an unpaired adenosine base.

Nucleotides	Dykeman *et al.* (2013)	Rolfsson *et al.* (2015)	Dai *et al.* (2016)
102-114	SL-27	n/a	12
179-200	n/a	n/a	15
593-606	SL-17	n/a	5
902-915	SL-14	P13	10
990-997	SL-12	P15	11
1460-1470	SL-3	n/a	9
1720-1731	SL-1	P26	8
1747-1763	TR	P27	1
1776-1791	SL+1	P28	2
2040-2053	SL+5	P31	13
2367-2374	n/a	n/a	3
2468-2481	SL+11	P37	14
2781-2796	SL+20	P44	4
2840-2852	SL+22	P45	6
3359-3372	SL+28	P54	7

The amounts of CP and RNA produced *in vivo* appeared reduced (Chapter 5 page 120), however even at the highest concentration of MTX tested MS2 was never completely inhibited implying that the effect could be via anti-assembly. Under the same *in vivo* conditions, the presence of MTX slightly increased the infectivity of a closely related bacteriophage GA, whereas a decrease in infectivity was seen with Qβ (Chapter 5, page 119). These bacteriophages have RNA SLs with similar secondary structures to MS2's PS TR, shown in Figure 6.1.

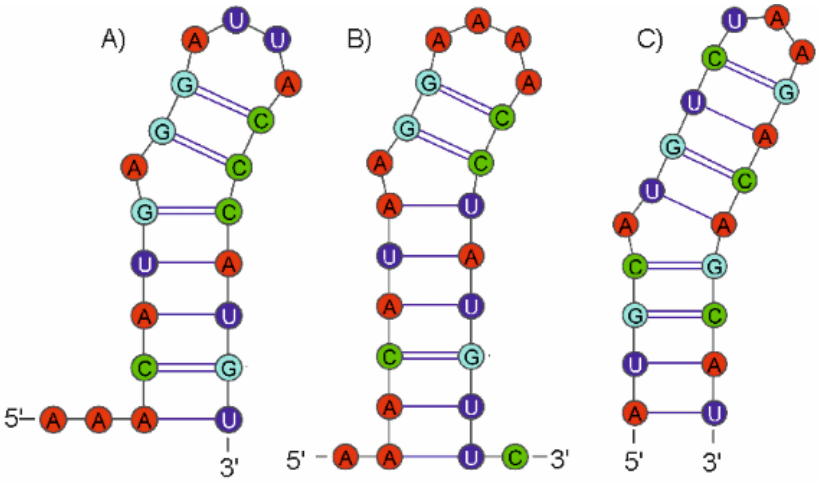


Figure 6.1 The secondary structures of Translational Repressors in bacteriophages A) MS2 (Carey et al. 1983), B) GA (Horn, 2005) and C) Qβ (Gott et al. 1991). All encompass an adenosine-bulge in the bp stem. Drawn in Varna (Darty et al. 2009).

After passaging 3 times in the presence of MTX, MS2 showed a higher titre than wild-type (Chapter 5, page 123). It was assumed that this was due to selection of an escape mutant. Next Generation Sequencing of the escape mutant confirmed only one site with 100% mutation, G393U (Chapter 5, Page 126). This nucleotide resides within SL 388-414 which has been shown by biochemical assays to bind to MP (Shiba and Suzuki 1981, Rolfsson *et al.* 2015, Koning *et al.* 2016). MFOLD analysis of G393U predicts a change in the secondary structure of SL 388-414, which reduces its affinity towards MTX (Chapter 5, Page 128). These suggest that the origin of inhibition *in vivo* is because of direct competition between MP and MTX for this RNA binding site. The recent asymmetric structure of *wt*MS2 reported by Dai *et al.* (2016) describes SLs interacting with MP, but omits SL 388-414, placed at the 5' end of the genome. This conflicts with our recent XRF of MS2 virions which clearly highlights these nucleotides are in contact with MP (Chapter 5, page 126). Bacteriophage Qβ has its MP in place of a CP dimer at a 2-fold axis (Gorzelnik *et al.* 2016), similar to MS2 (Dent *et al.* 2013, Dai *et al.* 2016, Koning *et al.* 2016). Gorzelnik *et al.* revealed gRNA is asymmetrically packaged, with higher RNA density around the MP, potentially to facilitate gRNA release upon infection. During MS2 infection, MP cleaves into two domains approximately 15 kDa NTD and 25 kDa CTD, both domains and gRNA enter the cell (Krahn *et al.* 1971).

Further work to resolve interactions between gRNA and MP, or indeed any structural proteins, would allow for direct drug-targeting of key SLs involved in the infection process. As it stands the results outlined in this thesis would have been very different if the initial ROCS analysis had not discovered MTX. The original aim was to use a "one-molecule-fits-all" approach, however after vigorous dissemination of MS2's gRNA and the variety of structural motifs underpinning protein:RNA interactions (Dai *et al.* 2016), it may have been better to look independently at different types of RNA-binding molecules (electrostatic versus intercalators) and to tailor molecules to target specific areas of interest. This approach will be considered when moving forwards with other virus systems and as highly-resolved nucleotide information become available.

6.2. Further Work

An important experiment would be reproducing *in vitro* reassembly experiments described in Chapter 4, page 108, testing the ability of gRNA encompassing mutant G393U to form VLPs in the presence and absence of MTX.

Comparing the affinities of *wt* and mutant gRNA binding to MP, in the presence and absence of MTX, would indicate if there is a specific MP binding site within G393U. Dr Iain Manfield (University of Leeds, U.K.) has recently developed a recombinant protein expression system for MP, which could be used with ^{32}P-labelled gRNA transcripts in *in vitro* filter-binding assays.

Infectious MS2 virions encompassing the G393U mutation, could be compared to *wt* in pulldown assays using *E. coli* F. pilus. This experiment would determine if the mutant behaves like *wt* in the absence of MTX.

Experiments outlined in Chapter 5, page 132, demonstrates the suitability of XRF in determining ssRNA secondary structure to nucleotide resolution. XRF could be used to look at the structure of RNA from infectious MS2 virions encompassing the G393U mutation, and to see if the mutation changes the secondary structure of SL 388-414 in virion. Comparing data from virions

to free-RNA and gRNA-mediated VLPs would give information on intra- and intermolecular structural features.

6.2.1. Developing High-throughput Screening of RNA Binding Ligands Using Mass Spectrometry

Work in Chapter 3 uses mass spectrometry (MS) to identify RNA binding ligands. This assay has been developed into a standardised protocol and applied to RNA PSs found in other virus systems. Work carried out by the Stockley group has identified a PS involved in the assembly of VLPs in a *T*=1 satellite plant virus, STNV (Bunka *et al.* 2011). Supporting sequence-specificity, PS "B3" presents nucleotides -ACAA- in an apical loop, interacting with high affinity to the positively charged CP NTD (Ford *et al.* 2013). PSs have also been identified in a human picornavirus, Parecho Virus (hPeV), "PS6" was identified as the highest affinity CP binding site on the HPeV genome. Mutations of the HPeV genome within this SL was shown to reduce hPeV infectivity (Shakeel *et al.* 2017). Most recently, "PS1" was characterised in the pre-gRNA of a dsDNA virus, HBV (Patel *et al.* 2017). These RNAs were synthesised and purified, as described in Chapter 2. All RNAs are <35-mers and suitable for negative mode MS analysis. The predicted MFOLD secondary structures of TR, PS1, PS6 and B3, are shown in Figure 6.2.

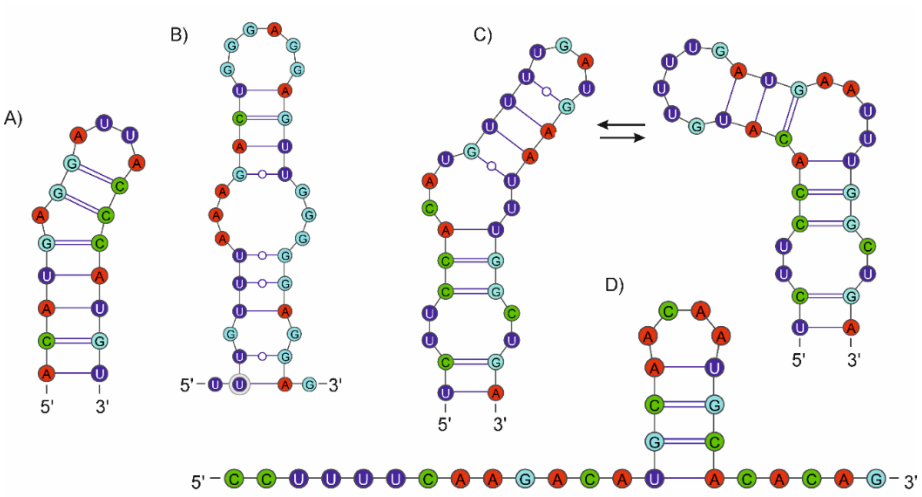


Figure 6.2 Secondary Structures of PS A) MS2 PS TR, B) HBV PS1, C) hPeV PS6 and D) STNV B3. Note there are two lowest-energy conformers of PS6. Secondary structures determined by MFOLD (Zuker 2003), drawn in Varna (Darty 2009).

RNAs were incubated with mixtures of ligands or “ligand cocktails” and ESI-MS was used to analyse changes in the measured mass of the RNA to identify successful binders. See Chapter 2, Section 2.10.4 for preparation of RNAs and ligands prior to MS. RNAs and cocktails were mixed (1:20 molar ratio) in a 96-well plate, at room temperature, then loaded onto the MS auto-sampler to check for binding. When TR is mixed with cocktail #1 and analysed by ESI-MS, the major peak at (4-) shows TR binds up to two MTX, Figure 6.3. There is no evidence for TR binding to any other ligands present in cocktail #1, suggesting this method is suitable for screening multiple ligands at once. Further work is being pursued screening multiple PSs in tandem, against much larger ligand libraries (~22,000 ligands). Using novel small molecule microarrays (SMM) developed by Connelly *et al.* (2017), more RNA binders have been identified and will be analysed by ESI-MS to verify any positive results. Using both SMM and ESI-MS it is hoped to find small molecular weight ligands that bind to RNA PS important in viral lifecycles, with enough specificity to exert an anti-viral effect.

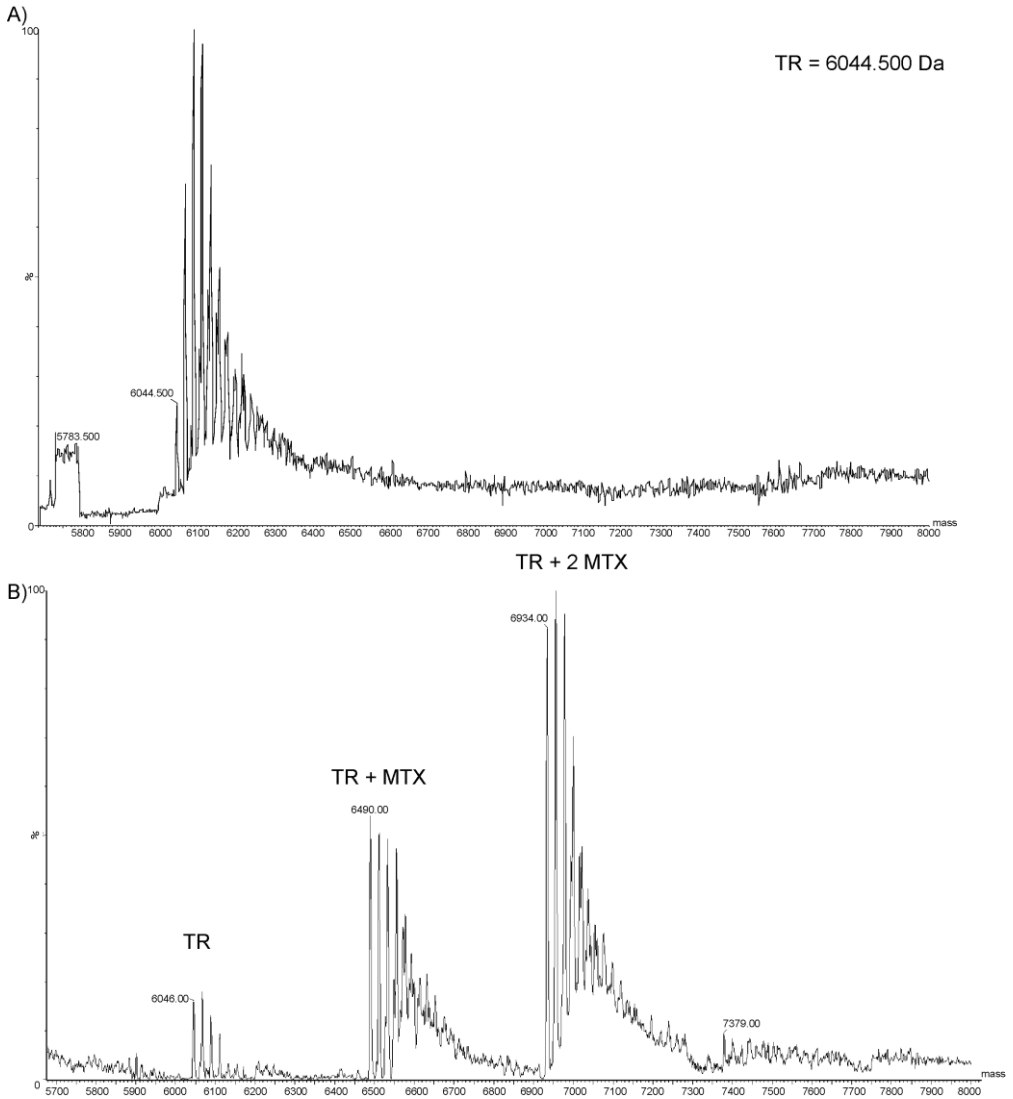


Figure 6.3 Negative-ion mode MS of TR +/- Cocktail #1. A) TR raw spectra m/z shows major charge state (4-). B) TR:Cocktail #1 1:20 molar ratio. Ligand cocktail dissolved in MeCN. (4-) peak shows up to 2 MTX bind per RNA.

This page has been left intentionally blank.

Chapter 7

7. References

Aboul-ela, F., Karn, J., and Varani, G. (1996). Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. Nucleic Acids Research 24, 3974-3981.

Aboul-ela, F. and Varani, G. (1998). Recognition of HIV-1 TAR RNA by Tat protein and Tat-derived peptides. Journal of Molecular Structure: THEOCHEM 423, 29-39.

Adams, M.J., Hendrickson, R.C., Dempsey, D.M., and Lefkowitz, E.J. (2015). Tracking the changes in virus taxonomy. Arch Virol 160, 1375-1383.

Adilakshmi, T., Soper, S.F., and Woodson, S.A. (2009). Structural analysis of RNA in living cells by in vivo synchrotron X-ray footprinting. Methods Enzymol 468, 239-258.

Adolph, K.W., and Butler, P.J.G. (1975). Reassembly of a spherical virus in mild conditions. Nature 255, 737-738.

Agnello, V., Abel, G., Elfahal, M., Knight, G.B., and Zhang, Q.-X. (1999). Hepatitis C virus and other Flaviviridae viruses enter cells via low density lipoprotein receptor. PNAS 96, 12766-12771.

Agol, V.I. (2006). Molecular mechanisms of poliovirus variation and evolution. Current topics in microbiology and immunology 299, 211-259.

Artigas, G., and Marchán, V. (2015). Synthesis and Tau RNA Binding Evaluation of Ametantrone-Containing Ligands. The Journal of Organic Chemistry 80, 2155-2164.

Arts, E.J., and Hazuda, D.J. (2012). HIV-1 Antiretroviral Drug Therapy. Cold Spring Harbor perspectives in medicine 2, a007161.

Balfour Jr., H.H. (1999). Antiviral Drugs. N Engl J Med 340, 1255-1268.

Baltimore D. (1971). Expression of animal virus genomes. Bacteriol Rev 35, 235-241.

Bancroft, J.B., Hiebert, E., and Bracker, C.E. (1969). The effects of various polyanions on shell formation of some spherical viruses. Virology 39, 924-930.

Bancroft, J.B., Hills, G.J., and Markham, R. (1967). A study of the self-assembly process in a small spherical virus formation of organized structures from protein subunits in vitro. . Virology 31, 354-379.

Banerjee, S., and Mazumdar, S. (2012). Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. International Journal of Analytical Chemistry 2012, 40.

Bardaro, M.F., Jr., Shajani, Z., Patora-Komisarska, K., Robinson, J.A., and Varani, G. (2009). How binding of small molecule and peptide ligands to HIV-1 TAR alters the RNA motional landscape. Nucleic Acids Res 37, 1529-1540.

Basnak, G., Morton, V.L., Rolfsson, O., Stonehouse, N.J., Ashcroft, A.E., and Stockley, P.G. (2010). Viral genomic single-stranded RNA directs the pathway toward a T=3 capsid. J Mol Biol 395, 924-936.

Bebenek, K. and Kunkel, T. A. Functions of DNA Polymerases, In Advances in Protein Chemistry, Academic Press, Volume 69, 2004, Pages 137-165.

Beckett, D., Wu, H.N., and Uhlenbeck, O.C. (1988). Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. J Mol Biol 204, 939-947.

Behbehani, A.M. (1983). The smallpox story: life and death of an old disease. Microbiological Reviews 47, 455-509.

Behe, M., and Felsenfeld, G. (1981). Effects of methylation on a synthetic polynucleotide: The B-Z transition in poly(dG-m5dC)poly(dG-m5dC). Proc NatL Acad Sci USA 78, 1619-1623.

Beijerinck, M.W. (1898). Über ein contagium vivum fluidum als Ursache der Fleckenkrankheit der Tabaksblätter. Verh K Akad Wet Amsterdam 65, 3-21.

Bella, J., and Rossmann, M.G. (2000). ICAM-1 receptors and cold viruses. Pharm Acta Helv 74, 291-297.

Benesi, H.A., and Hildebrand, J.H. (1949). A Spectrophotometric Investigation of the Interaction of Iodine with Aromatic Hydrocarbons. J Am Chem Soc 71, 2703.

Beremand, M.N., and Blumenthal, T. (1978). Overlapping genes in RNA phage: a new protein implicated in lysis. . Cell 18, 257-266.

Bernhardt, T.G., Wang, I.-N., Struck, D.K., and Young, R. (2002). Breaking free: "Protein antibiotics" and phage lysis. Research in Microbiology 153, 493-501.

Berg, H., Horn, G., Luthardt, U., and Ihn, W. (1981). 447 - Interaction of anthracycline antibiotics with biopolymers. Part V. Polarographic behavior and complexes with DNA. Bioelectrochemistry and Bioenergetics 8, 537-553.

Bergelson, J.M., Cunningham, J.A., Droguett, G., Kurt-Jones, E.A., Krithivas, A., Hong, J.S., Horwitz, M.S., Crowell, R.L., and Finberg, R.W. (1997). Isolation of a Common Receptor for Coxsackie B Viruses and Adenoviruses 2 and 5. Science 275, 1320-1323.

Bergh, O., Borsheim, K.Y., Bratbak, G., and Heldal, M. (1989). High abundance of viruses found in aquatic environments. Nature 340, 467-468.

Berkhout, B., Kastelein, R.A., and van Duin, J. (1985). Translational interference at overlapping reading frames in prokaryotic messenger RNA. Gene 37, 171-179.

Betts, R.F. (1989). Amantadine and rimantadine for the prevention of influenza A. Semin Respir Infect 4, 304-310.

Biemelt, S., Sonnewald, U., Galmbacher, P., Willmitzer, L., and Muller, M. (2003). Production of human papillomavirus type 16 virus-like particles in transgenic plants. J Virol 77, 9211-9220

Blakeley, B.D., M., D.S., Mohan, U., Burai, R., Tolbert, B.S., and McNaughton, B.R. (2012). Methods for identifying and characterizing interactions involving RNA. Tetrahedron 68, 8837-8855.

Boedtker, H. (1960). Configurational properties of tobacco mosaic virus ribonucleic acid. Journal of Molecular Biology 2, 171-188.

Boedtker, H. (1968). Dependence of the sedimentation coefficient on molecular weight of RNA after reaction with formaldehyde. Journal of Molecular Biology 35, 61-70.

Borer, P.N., Lin, Y., Wang, S., Roggenbuck, M.W., Gott, J.M., Uhlenbeck, O.C., and Pelczer, I. (1995). Proton NMR and structural features of a 24-nucleotide RNA hairpin. Biochemistry 34, 6488-6503.

Borodavka, A., Tuma, R., and Stockley, P.G. (2013). A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. RNA Biol 10, 481-489.

Borodavka, O. (2013). Understanding the dynamics of viral RNA genomes using single-molecule fluorescence. University of Leeds. PhD Thesis.

Botella, e.a. (1995). Receptor subtypes involved in dual effects induced by prostaglandin E2 in circular smooth muscle from dog colon. JPharmacolExpTher 273 1008.

Böttcher, B., and Crowther, R.A. (1996). Difference imaging reveals ordered regions of RNA in turnip yellow mosaic virus. Structure 4, 387-394.

Bourne, C., Lee, S., Venkataiah, B., Lee, A., Korba, B., Finn, M.G., and Zlotnick, A. (2008). Small-Molecule Effectors of Hepatitis B Virus Capsid Assembly Give Insight into Virus Life Cycle. Journal of Virology 82, 10262-10270.

Boutros, M., and Ahringer, J. (2008). The art and design of genetic screens: RNA interference. Nature reviews Genetics 9, 554-566.

Brandau, D.T., Jones, L.S., Wiethoff, C.M., Rexroad, J., and Middaugh, C.R. (2003). Thermal stability of vaccines. J Pharm Sci 92, 218-231.

Brenner, S., and Horne, R.W. (1959). A negative staining method for high resolution electron microscopy of viruses. Biochim Biophys Acta 34, 103-110.

Brito, A.F., and Pinney, J.W. (2017). Protein-Protein Interactions in Virus-Host Systems. Frontiers in Microbiology 8, 1557.

Brodersen, D.E., Clemons Jr., W.M., Carter, A.P., Morgan-Warren, R.J., Wimberly, B.T., and Ramakrishnan, V. (2000). The Structural Basis for the Action of the Antibiotics Tetracycline, Pactamycin, and Hygromycin B on the 30S Ribosomal Subunit. Cell 103, 1143-1154.

Buck, J., Fürtig, B., Noeske, J., Wöhnert, J., and Schwalbe, H. (2007). Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. Proceedings of the National Academy of Sciences 104, 15699-15704.

Bunka, D.H., Lane, S.W., Lane, C.L., Dykeman, E.C., Ford, R.J., Barker, A.M., Twarock, R., Phillips, S.E., and Stockley, P.G. (2011). Degenerate RNA packaging signals in the genome of Satellite Tobacco Necrosis Virus: implications for the assembly of a T=1 capsid. J Mol Biol 413, 51-65.

Butler, P.J. (1999). Self-assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed. Philos Trans R Soc Lond B Biol Sci 354, 537-550.

Cady, S.D., Schmidt-Rohr, K., Wang, J., Soto, C.S., DeGrado, W.F., and Hong, M. (2010). Structure of the Amantadine Binding Site of Influenza M2 Proton Channels In Lipid Bilayers. Nature 463, 689-693.

Carey, J., Cameron, V., Dehaseth, P.L., and Uhlenbeck, O.C. (1983). Sequence-specific interaction of R17-coat protein with its ribonucleic-acid binding-site. Biochemistry 22, 2601-2610.

Carnevali, M., Parsons, J., Wyles, D.L., and Hermann, T. (2010). A Modular Approach to Synthetic RNA Binders of the Hepatitis C Virus Internal Ribosome Entry Site. ChemBioChem 11, 1364-1367.

Casjens, S. and King J. (1974). P22 morphogenesis. I. Catalytic scaffolding protein in capsid assembly. J Supramol Struct 2, 202-224.

Caskey, M., Klein, F., Lorenzi, J.C.C., Seaman, M.S., West Jr, A.P., Buckley, N., Kremer, G., Nogueira, L., Braunschweig, M., Scheid, J.F., et al. (2015). Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. Nature 522, 487-491.

Calendar, R. (1988). The Bacteriophages, Vol 1 (New York and London: Plenum Press).

Caspar, D. L. (1956). Structure of bushy stunt virus. Nature 177, 475-476.

Caspar, D. L. (1975) Design principles in virus particle construction. In Horsfall FL, Tamm I (eds): Viral and Rickettsial Infections in Man. 4th Ed. JB Lippincott, Philadelphia.

Caspar, D.L.D., and Klug, A. (1962). Physical Principles in the Construction of Regular Viruses. Cold Spring Harb Sym 27, 1-24.

Chamakura, K.R., Tran, J.S., and Young, R. (2017). MS2 Lysis of Escherichia coli Depends on Host Chaperone DnaJ. J Bacteriol 199.

Champoux, J.J., and Schultz, S.J. (2009). Ribonuclease H: properties, substrate specificity and roles in retroviral reverse transcription. The FEBS journal 276, 1506-1516.

Cheng, S., and Brooks, C.L., 3rd (2015). Protein-Protein Interfaces in Viral Capsids Are Structurally Unique. J Mol Biol 427, 3613-3624.

Choudhury, M., and Basu, R. (1995). Studies of charge transfer interaction of nucleotides with proflavine. Journal of Photochemistry and Photobiology A: Chemistry 85, 89-92.

Clapham, P.R., and McKnight, A. (2001). HIV-1 receptors and cell tropism. British medical bulletin 58, 43-59.

Codutti, L., Leppek, K., Zálešák, J., Windeisen, V., Masiewicz, P., Stoecklin, G., and Carlomagno, T. (2015). A Distinct, Sequence-Induced Conformation Is Required for Recognition of the Constitutive Decay Element RNA by Roquin. Structure 23, 1437-1447.

Colman, P.M., and Ward, C.W. (1985). Structure and diversity of influenza virus neuraminidase. Current topics in microbiology and immunology 114, 177-255.

Connelly, C.M., Abulwerdi, F.A., and Schneekloth, J.S., Jr. (2017). Discovery of RNA Binding Small Molecules Using Small Molecule Microarrays. Methods Mol Biol 1518, 157-175.

Corey, L., Fife, K.H., Benedetti, J.K., Winter, C.A., Fahnlander, A., Connor, J.D., Hintz, M.A., and Holmes, K.K. (1983). Intravenous Acyclovir for the Treatment of Primary Genital Herpes. Ann Intern Med 98, 914-921.

Crespi, M.D., Ivanier, S.E., Genovese, J., and Baldi, A. (1986). Mitoxantrone affects topoisomerase activities in human breast cancer cells. Biochem Biophys Res Commun *136*, 521-528.

Creutzig, U., and al., e. (2013). Randomized trial comparing liposomal daunorubicin with idarubicin as induction for pediatric acute myeloid leukemia: results from Study AML-BFM. Blood 122, 37-43.

Crick, F.H.C., and Watson, J.D. (1956). Structure of small viruses. Nature 177, 473-475.

Curtiss, L.K., and Krueger, R.G. (1974). Localization of Coliphage MS2 A-Protein. J Virol 14, 503-508.

Dai, X., Li, Z., Lai, M., Shu, S., Du, Y., Zhou, Z.H., and Sun, R. (2016). In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. Nature advance online publication.

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics (Oxford, England) 25, 1974-1975.

Darzynkiewicz, Z. (1990). Differential staining of DNA and RNA in intact cells and isolated cell nuclei with acridine orange. . Methods in Cell Biology 33, 285-298.

Davidson, A., Leeper, T.C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J.A., and Varani, G. (2009). Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein. Proc Natl Acad Sci U S A 106, 11931-11936.

Davidson, A., Patora-Komisarska, K., Robinson, J.A., and Varani, G. (2011). Essential structural requirements for specific recognition of HIV TAR RNA by peptide mimetics of Tat protein. Nucleic Acids Res 39, 248-256.

Davies, J., Gorini, L., and Davis, B.D. (1965). Misreading of RNA codewords induced by aminoglycoside antibiotics. Mol Pharmacol, 93-106.

Davis, G.L. (1998). Interferon Alfa-2b Alone Or With Ribavirin For The Treatment Of Relapse Of Chronic Hepatitis C. New Engl J Med 339, 1493-1499.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System. (San Carlos, CA, USA: DeLano Scientific).

Dent, Kyle C., Thompson, R., Barker, Amy M., Hiscox, Julian A., Barr, John N., Stockley, Peter G., and Ranson, Neil A. (2013). The Asymmetric Structure of an Icosahedral Virus Bound to Its Receptor Suggests a Mechanism for Genome Release. Structure 21, 1225-1234.

Deval, J. (2009). Antimicrobial strategies: inhibition of viral polymerases by 3'-hydroxyl nucleosides. Drugs of today (Barcelona, Spain: 1998) 69, 151-166.

Dervan, P.B., Doss, R.M., and Marques, M.A. (2005). Programmable DNA binding oligomers for control of transcription. Current Medicinal Chemistry - Anti-Cancer Agents 5, 373-387.

Dixon, C.W. (1962). Smallpox (London: J. & A. Churchill).

Dragic, T., Litwin, V., Allaway, G.P., Martin, S.R., Huang, Y., Nagashima, K.A., Cayanan, C., Maddon, P.J., Koup, R.A., Moore, J.P., et al. (1996). HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature 381, 667-673.

Dykeman, E.C., Stockley, P.G., and Twarock, R. (2013). Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. J Mol Biol 425, 3235-3249.

Du, Y.H., Huang, J., Weng, X.C., and Zhou, X. (2010). Specific Recognition of DNA by Small Molecules. Current Medicinal Chemistry 17, 173-189.

Earl, L.A., and Subramaniam, S. (2016). Cryo-EM of viruses and vaccine design. Proceedings of the National Academy of Sciences 113, 8903-8905.

Edinger, T.O., Pohl, M.O., and Stertz, S. (2014). Entry of influenza A virus: host factors and antiviral targets. J Gen Virol 95, 263-277.

Eguchi, A., Lee, G.O., Wan, F., Erwin, G.S., and Ansari, A.Z. (2014). Controlling gene networks and cell fate with precision-targeted DNA-binding proteins and small-molecule-based genome readers. The Biochemical journal 462, 397-413.

Eigen, and Rigler (1994). Sorting Single Molecules: Application to Diagnostics and Evolutionary Biotechnology. PNAS 91, 5740-5747.

Eiserling, F.A., and Black, L.W. (1994). Pathways in T4 morphogenesis. In: Molecular Biology of Bacteriophage T4 (Washington, D. C.).

Enache, M., and Volanschi, E. (2010). Spectral Characterization Of Self-Association Of Antitumor Drug Mitoxantrone. Rev Roum Chim 55, 255-262.

Enger, M.D., Stubbs, E.A., Mitra, S., and Kaesberg, P. (1963). Biophysical Characteristics of the RNA-Containing Bacterial Virus R17. P N A S 49, 857.

Entzian, C., and Schubert, T. (2016). Studying small molecule-aptamer interactions using MicroScale Thermophoresis (MST). Methods (San Diego, Calif) 97, 27-34.

Faber, C., Sticht, H., Schweimer, K., and Rosch, P. (2000). Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. J Biol Chem 275, 20660-20666.

Fabris, D. (2010). A role for the MS analysis of nucleic acids in the post-genomics age. Journal of the American Society for Mass Spectrometry 21, 1-13.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science 246, 64-71.

Fiers, W., Contreras, R., Duerinck, F., Haegmean, G., Merregaert, J., Jou, W.M., Raeymakers, A., Volckaert, G., Ysebaert, M., Van de Kerckhove, J., et al. (1975). A-protein gene of bacteriophage MS2. Nature 256, 273-278.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van Den Berghe, A., et al. (1976). Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature 260, 500 - 507.

Fiers, W.e.a. (1975). A-protein gene of bacteriophage MS2. Nature 256, 273-278.

Fisher, A.J., McKinney, B.R., Schneemann, A., Rueckert, R.R., and Johnson, J.E. (1993). Crystallization of viruslike particles assembled from flock house virus coat protein expressed in a baculovirus system. Journal of Virology 67, 2950-2953.

Fisher, H.C. (2014). Chapter 4 Examination of Hepatitis B assembly and MS2 assembly inhibition. PhD Thesis. University of Leeds. PhD Thesis.

Ford, R.J. (2012). The Roles of RNA in the Assembly and Disassembly of Single-stranded RNA Icosahedral Viruses. University of Leeds. PhD Thesis.

Ford, R.J., Barker, A.M., Bakker, S.E., Coutts, R.H., Ranson, N.A., Phillips, S.E.V., Pearson, A.R., and Stockley, P.G. (2013). Sequence-Specific, RNA–Protein Interactions Overcome Electrostatic Barriers Preventing Assembly of Satellite Tobacco Necrosis Virus Coat Protein. Journal of Molecular Biology 425, 1050-1064.

Foster, T. L., Thompson, G. S., Kalverda, A. P., Kankanala, J., Bentham, M., Wetherill, L. F., Griffin, S. (2014). Structure-guided design affirms inhibitors of hepatitis C virus p7 as a viable class of antivirals targeting virion release. Hepatology, 59(2), 408-422.

Fouts, D.E., True, H.L., and Celander, D.W. (1997). Functional recognition of fragmented operator sites by R17/MS2 coat protein, a translational repressor. Nucleic Acids Research 25, 4464-4473.

François, B., Russell, R.J.M., Murray, J.B., Aboul-ela, F., Masquida, B., Vicens, Q., and Westhof, E. (2005). Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. Nucleic Acids Research 33, 5677-5690.

Franklin, R.E., and Gosling, R.G. (1953). The structure of sodium thymonucleate fibres. I. The influence of water content. Acta Crystallographica 6, 673-677.

Frydman, L., Rossomando, P.C., Frydman, V., Fernandez, C.O., Frydman, B., and Samejima, K. (1992). Interactions between natural polyamines and tRNA: an 15N NMR analysis. Proceedings of the National Academy of Sciences 89, 9186-9190.

Fuchs, R., and Blaas, D. (2012). Productive Entry Pathways of Human Rhinoviruses. Advances in Virology 2012, 13.

Furtig, B., Richter, C., Wohnert, J., and Schwalbe, H. (2003). NMR spectroscopy of RNA. Chembiochem 4, 936-962.

Ganisl, B., Taucher, M., Riml, C., and Breuker, K. (2012). Charge as you like! Efficient manipulation of negative ion net charge in electrospray ionization of proteins and nucleic acids. EMJS IMSC Sample Papers, 333-343.

Ganser-Pornillos, B.K., Yeager, M., and Sundquist, W.I. (2008). The structural biology of HIV assembly. Curr Opin Struct Biol 203–217.

Gao, W. (2004). Novel mode of action of tylophorine analogs as antitumor compounds. Cancer research 64, 678-688.

Gardner, M.R., Kattenhorn, L.M., Kondur, H.R., von Schaewen, M., Dorfman, T., Chiang, J.J., Haworth, K.G., Decker, J.M., Alpert, M.D., Bailey, C.C., et al. (2015). AAV-expressed eCD4-Ig provides durable protection from multiple SHIV challenges. Nature 519, 87-91.

Gell, C., Brockwell, D., and Smith, A. (2006). Handbook of Single Molecule Fluorescence Spectroscopy. (OXford: Oxford University Press).

Giacomoni, P.U., and Le Bret, M. (1973). Electronic structure of ethidium bromide. Febs Lett 29, 227-230.

Glebe, D., and Urban, S. (2007). Viral and cellular determinants involved in hepadnaviral entry. World J Gastroenterol 13, 22-38.

Goessens, W.H., Driessen, A.J., Wilschut, J., and van Duin, J. (1988). A synthetic peptide corresponding to the C-terminal 25 residues of phage MS2 coded lysis protein dissipates the protonmotive force in Escherichia coli membrane vesicles by generating hydrophilic pores. Embo J 7, 867-873.

Golan, D.A., Tashjian, A.H., Armstrong, E.J. (2011). Principles of Pharmacology: The Pathophysiologic Basis of Drug Therapy. International Edition. 1, 6.

Golmohammadi, R., Fridborg, K., Bundule, M., Valegard, K., and Liljas, L. (1996). The crystal structure of bacteriophage Q beta at 3.5 A resolution. Structure 4, 543-554.

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

Gorzelnik, K.V., Cui, Z., Reed, C.A., Jakana, J., Young, R., and Zhang, J. (2016). Asymmetric cryo-EM structure of the canonical Allolevivirus Qβ reveals a single maturation protein and the genomic ssRNA in situ. Proceedings of the National Academy of Sciences 113, 11519-11524.

Gott, J.M., Wilhelm, L.J., and Uhlenbeck, O.C. (1991). RNA binding properties of the coat protein from bacteriophage GA. Nucl Acids Res 19, 6499–6503.

Gottesfeld, J.M., Neely, L., Trauger, J.W., Baird, E.E., and Dervan, P.B. (1997). Regulation of gene expression by small molecules. Nature 387, 202-205.

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

Grant, J.A., Gallardo, M.A., and Pickup, B. (1996). A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. J Comp Chem 17, 1653.

Greve, J.M., Davis, G., Meyer, A.M., Forte, C.P., Yost, S.C., Marlor, C.W., Kamarck, M.E., and McClelland, A. (1989). The major human rhinovirus receptor is ICAM-1. Cell 56, 839-847.

Griffey, R.H., Greig, M.J., Haoyun, A., Sasmor, H., and Manalili, S. (1999a). Targeted Site-Specific Gas-Phase Cleavage of Oligoribonucleotides. Application in Mass Spectrometry-Based Identification of Ligand Binding Sites. J Am Chem Soc 121, 474–475.

Griffey, R.H., Hofstadler, S.A., Sannes-Lowery, K.A., Ecker, D.J., and Crooke, S.T. (1999b). Determinants of aminoglycoside-binding specificity for rRNA by using mass spectrometry. Proceedings of the National Academy of Sciences 96, 10129-10133.

Grillone, L.R., and Lanz, R. (2001). Fomivirsen. Drugs of today (Barcelona, Spain : 1998) 37, 245-255.

Havelaar, A.H., van Olphen, M., and Drost, Y.C. (1993). F-specific RNA bacteriophages are adequate model organisms for enteric viruses in fresh water. Appl Environ Microbiol 59, 2956-2962.

Hawkins, P.C.D., Skillman, A.G., and Nicholls, A. (2007). Comparison of Shape-Matching and Docking as Virtual Screening Tools. Journal of Medicinal Chemistry 50, 74-82.

Hayes, J.J., Kam, L., and Tullius, T.D. (1990). Footprinting protein-DNA complexes with gamma-rays. Methods Enzymol 186, 545-549.

Helgstrand, C., Grahn, E., Moss, T., Stonehouse, N. J., Tars, K., Stockley, P. G., & Liljas, L. (2002). Investigating the structural basis of purine specificity in the structures of MS2 coat protein RNA translational operator hairpins. *Nucleic Acids Research*, *30*(12), 2678–2685.

Highleyman, L. (1998). Fomivirsen. BETA, 29, 31.

Hirao, I., Spingola, M., Peabody, D.S., and Ellington, A.D. (1999). The limits of specificity: An experimental analysis with RNA aptamers to MS2 coat protein variants. Molecular Diversity 4, 75-89.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and VandePol, S. (1982). Rapid evolution of RNA genomes. Science 215, 1577-1585.

Horn, W.T., Convery, M.A., Stonehouse, N.J., Adams, C.J., Liljas, L., Phillips, S.E.V., and Stockley, P.G. (2004). The crystal structure of a high affinity RNA stem-loop complexed with the bacteriophage MS2 capsid: Further challenges in the modeling of ligand–RNA interactions. RNA 10, 1776-1782.

Horne, R.W. (1959). The icosahedral form of an adenovirus. J Mol Biol 1, 84-85.

Howe Grant, M. (1993). Chapter 41 Antiviral Agents. In Chemotherapeutics And Disease Control (John Wiley and Sons)

Hu, Y., Zandi, R., Anavitarte, A., Knobler, C.M., and Gelbart, W.M. (2008). Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size. Biophys J 94, 1428-1436.

Hulscher, R.M., Bohon, J., Rappe, M.C., Gupta, S., D'Mello, R., Sullivan, M., Ralston, C.Y., Chance, M.R., and Woodson, S.A. (2016). Probing the structure of ribosome assembly intermediates in vivo using DMS and hydroxyl radical footprinting. Methods (San Diego, Calif) 103, 49-56.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J Molec Graphics 14, 33–38.

Ibrahim, M.S. (2001). Voltammetric studies of the interaction of nogalamycin antitumor drug with DNA. Analytica Chimica Acta 443, 63-72.

ICTV (2016). International Committee on Virus Taxonomy: 2016 Release.

Inagaki, M., Kawamoto, S., Itoh, H., Saitoh, M., Hagiwara, M., Takahashi, J., and Hidaka, H. (1986). Naphthalenesulfonamides as calmodulin antagonists and protein kinase inhibitors. Mol Pharmacol 29, 577-581.

Jacobs, L.D. (2000). Intramuscular Interferon Beta-1a Therapy Initiated During A First Demyelinating Event In Multiple Sclerosis. New Engl J Med 343, 898-904.

Jaeger, J.A., Zuker, M., and Turner, D.H. (1990). Melting and chemical modification of a cyclized self-splicing group I intron: similarity of structures in 1 M sodium, in 10 mM magnesium and in the presence of substrate. Biochemistry 29, 10147-10158.

Jain, R. and Srivastava, R. (2009). Metabolic investigation of host/pathogen interaction using MS2-infected Escherichia coli. BMS Sys Biol, 120-130.

Jamieson, E.R. and Lippard, S.J. (1999). Structure, Recognition, and Processing of Cisplatin-DNA Adducts. Chem Rev. 9, 2467-2498.

Janne, J., Alhonen, L., Pietila, M., and Keinanen, T.A. (2004). Genetic approaches to the cellular functions of polyamines in mammals. Eur J Biochem 271, 877-894.

Jenner, E. (1798). An inquiry into the causes and effects of the variolae vaccinae : a disease discovered in some of the western counties of England, particularly Gloucestershire, and known by the name of the cow pox (Ashley & Brewer).

Jenner, E. (1801). The origin of the vaccine inoculation. (Berwick Street, Soho).

Kapuscinski, J., Darzynkiewicz, Z., Traganos, F., and Melamed, M.R. (1981). Interactions of a new antitumor agent, 1,4-dihydroxy-5,8-bis[[2-[(2-hydroxyethyl)amino]-ethyl]amino]-9,10-anthracenedione, with nucleic acids. Biochem Pharmacol 30, 231-240.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem 60, 2299-2301.

Keene, F.R. (2011). The shape of metallosupramolecular assemblies. Dalton Trans 40, 2405-2418.

Kegel, W.K., and van der Schoot, P. (2006). Physical regulation of the self-assembly of tobacco mosaic virus coat protein. Biophysical Journal 91, 1501-1512.

Kelly, B.N., Kyere, S., Kinde, I., Tang, C., Howard, B.R., Robinson, H., Sundquist, W.I., Summers, M.F., and Hill, C.P. (2007). Structure of the antiviral assembly inhibitor CAP-1 complex with the HIV-1 CA protein. J Mol Biol 373, 355-366.

Kielian, M., and Jungerwirth, S. (1990). Mechanisms of enveloped virus entry into cells. Molecular biology & medicine 7, 17-31

Kim, H.O., Ji, X.D., Siddiqi, S.M., Olah, M.E., Stiles, G.L., and Jacobson, K.A. (1994). 2-Substitution of N6-benzyladenosine-5'-uronamides enhances selectivity for A3 adenosine receptors. J Med Chem 37, 3614-3621.

Kleyman, T.R., and Cragoe Jr., E.J. (1988). Amiloride and its analogs as tools in the study of ion transport. J Membr Biol 105, 1-21.

Knapman, T.W., Morton, V.L., Stonehouse, N.J., Stockley, P.G., and Ashcroft, A.E. (2010). Determining the topology of virus assembly intermediates using ion mobility spectrometry–mass spectrometry. Rapid Communications in Mass Spectrometry 24, 3033-3042.

Knight-Jones, T.J.D., and Rushton, J. (2013). The economic impacts of foot and mouth disease – What are they, how big are they and where do they occur? Preventive Veterinary Medicine 112, 161-173.

Koboldt, D., Zhang, Q., Larson, D., Shen, D., McLellan, M., Lin, L., Miller, C., Mardis, E., Ding, L., and Wilson, R. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing Genome Research 22, 568-576.

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., and Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics (Oxford, England) 25, 2283-2285.

Koning, R.I., Gomez-Blanco, J., Akopjana, I., Vargas, J., Kazaks, A., Tars, K., Carazo, J.M., and Koster, A.J. (2016). Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. In Nature Communications. 7, Article number 12524.

Krahn, P.M., O'Callaghan, R.J., and Paranchych, W. (1972). Stages in phage R17 infection. Virology 47, 628-637.

Krumbholz, A., Schmidtke, M., Bergmann, S., Motzke, S., Bauer, K., Stech, J., Dürrwald, R., Wutzler, P., and Zell, R. (2009). High prevalence of amantadine resistance among circulating European porcine influenza A viruses. The Journal of general virology 90, 900-908.

Kyncl, J.J. (1986). Pharmacology of terazosin. The American journal of medicine 80, 12-19.

Lafon, M. (2005). Rabies virus receptors. J Neurovirol 11, 82-87.

Lago, H., Parrott, A.M., Moss, T., Stonehouse, N.J., and 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.

Lamm, O. (1929). Die Differentialgleichung der Ultrazentrifugierung. Ark Mat Astr Fys 21B, 1-4.

Lane, C.L. (2007). Structural studies of satellite tobacco necrosis virus (STNV) University of Leeds. PhD Thesis.

Lane, S. (2003). Structural and Functional Studies of Recombinant STNV capsids. University of Leeds. PhD Thesis.

Langner, K. (2006). https://commons.wikimedia.org/wiki/File:DNA_intercalation2.jpg.

Latham, M.P., Zimmermann, G.R., and Pardi, A. (2009). NMR chemical exchange as a probe for ligand-binding kinetics in a theophylline-binding RNA aptamer. J Am Chem Soc 131, 5052-5053.

Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In Analytical Ultracentrifugation in Biochemistry

and Polymer Science, S. Harding, Rowe, A., and Horton, J.,, ed. (London: Royal Society of Chemistry), pp. 90-125.

Lentz, T.L. (1990). The recognition event between virus and host cell receptor: a target for antiviral agents. J Gen Virol 71, 751-766.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler Transform. . Bioinformatics (Oxford, England) 25, 1754-1760.

Li, N., Ma, Y., Yang, C., Guo, L., and Yang, X. (2005). Interaction of anticancer drug mitoxantrone with DNA analyzed by electrochemical and spectroscopic methods. Biophysical Chemistry 116, 199-205.

Lima, S.M.B., Peabody, D.S., Silva, J.L., and de Oliveira, A.C. (2004). Mutations in the hydrophobic core and in the protein-RNA interface affect the packing and stability of icosahedral viruses. Eur J Biochem 271, 135-145.

Liu, F., Campagna, M., Qi, Y., Zhao, X., Guo, F., Xu, C., Li, S., Li, W., Block, T.M., Chang, J., et al. (2013). Alpha-interferon suppresses hepadnavirus transcription by altering epigenetic modification of cccDNA minichromosomes. PLoS Pathog 9, e1003613.

Lv, Z., Chu, Y., and Wang, Y. (2015). HIV protease inhibitors: a review of molecular selectivity and toxicity. HIV/AIDS (Auckland, NZ) 7, 95-104.

Ma, J. K., Drake, P. M. & Christou, P. (2003) The production of recombinant pharmaceutical proteins in plants. Nat. Rev. Genet. 4, 794–805.

Maiti, S., Haupts, U. and Webb, W.W. (1997). Fluorescence correlation spectroscopy-diagnostics for sparse molecules. Proc Natl Acad Sci U S A 94, 11753–11757.

Marsian, J., Fox, H., Bahar, M.W., Kotecha, A., Fry, E.E., Stuart, D.I., Macadam, A.J., Rowlands, D.J., and Lomonossoff, G.P. (2017). Plant-made polio type 3 stabilized VLPs—a candidate synthetic polio vaccine. Nature Communications 8, 245.

Mastico, R.A., Talbot, S.J., and Stockley, P.G. (1993). Multiple presentation of foreign peptides on the surface of an RNA-free spherical bacteriophage capsid. J Gen Virol 74 (Pt 4), 541-548.

Matrosovich, M., Herrler, G., and Klenk, H.D. (2013). Sialic Acid Receptors of Viruses. Topics in current chemistry.

Marchanka, A., Simon, B., Althoff-Ospelt, G., and Carlomagno, T. (2015). RNA structure determination by solid-state NMR spectroscopy. Nat Comms. 6, Article number 7024.

Mendelsohn, C.L., Wimmer, E., and Racaniello, V.R. (1989). Cellular receptor for poliovirus: Molecular cloning, nucleotide sequence, and expression of a new member of the immunoglobulin superfamily. . Cell 56, 855-865.

Min Jou, W., Haegeman, G., Ysebaert, M., and Fiers, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. Nature 237, 82-88.

Moitessier, N., Westhof, E., and Hanessian, S. (2006). Docking of Aminoglycosides to hydrated and flexible RNA. Journal of Medicinal Chemistry 49, 1023-1033.

Morton, V.L., Dykeman, E.C., Stonehouse, N.J., Ashcroft, A.E., Twarock, R., and Stockley, P.G. (2010). The Impact of Viral RNA on Assembly Pathway Selection. Journal of Molecular Biology 401, 298-308.

Morton, V.L., Stockley, P.G., Stonehouse, N.J., and Ashcroft, A.E. (2008). Insights Into Virus Capsid Assembly From Non-Covalent Mass Spectrometry. Mass Spectrometry Reviews 27, 575–595.

Moumné, R., Catala, M., Larue, V., Micouin, L., and Tisné, C. (2012). Fragment-based design of small RNA binders: promising developments and contribution of NMR. Biochimie. 94(7):1607-19.

Murchie, A.I., Davis, B., Isel, C., Afshar, M., Drysdale, M.J., Bower, J., Potter, A.J., Starkey, I.D., Swarbrick, T.M., Mirza, S., et al. (2004). Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. J Mol Biol 336, 625-638.

Namba, K., Pattanayek, R., and Stubbs, G. (1989). Visualization of protein–nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9Å resolution by x-ray fiber diffraction. J. Mol. Biol. 208, 307–325.

Neurath, A.R., and Strick, N. (1990). Antigenic mimicry of an immunoglobulin A epitope by a hepatitis B virus cell attachment site. Virology 178, 631-634.

NHS (2015). Vaccination Schedule.

Nicholson, K.G., Aoki, F.Y., Osterhaus, A., Trottier, S., Carewicz, O., Mercier, C.H., Rode, A., Kinnersley, N., and Ward, P. Efficacy and safety of oseltamivir in treatment of acute influenza: a randomised controlled trial. The Lancet 355, 1845-1850.

Nordén, B., and Tjerneld, F. (1976). Binding of Inert Metal Complexes to DNA Detected by Linear Dichroism. FEBS Journal Letters 67, 368-370.

Olmsted, J., and Kearns, D.R. (1977). Mechanism of Ethidium Bromide Fluorescence Enhancement on Binding to Nucleic Acids. Biochemistry 16, 3647.

Oxford, J., Kellam, P., and Collier, L. (2016). Human Virology (Oxford: Oxford University Press). 5th Edition.

Palmer, A.G., 3rd (2014). Chemical exchange in bio-macromolecules: past, present, and future. J. Magn. Reson. 241, 3-17.

Parker, C. A.; Barnes, W. J. (1957) Some experiments with spectrofluorimeters and filter fluorimeters. Analyst, 82, 606-617.

Parrott, A.M., Lago, H., Adams, C.J., Ashcroft, A.E., Stonehouse, N.J., and Stockley, P.G. (2000). RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behaviour. Nucleic Acids Research 28, 489-497.

Patel, N., Dykeman, E.C., Coutts, R.H.A., Lomonossoff, G.P., Rowlands, D.J., Phillips, S.E.V., Ranson, N.A., Twarock, R., Tuma, R., and Stockley, P.G. (2015). Revealing the density of encoded functions in a viral RNA. PNAS 112, 2227–2232.

Patel, N., White, S.J., Thompson, R.F., Bingham, R., Weiß, E.U., Maskell, D.P., Zlotnick, A., Dykeman, E.C., Tuma, R., Twarock, R., et al. (2017). HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. 2, 17098.

Patel, N., Wroblewski, E., Leonov, G., Phillips, S.E.V., Tuma, R., Twarock, R., and Stockley, P.G. (2017). Rewriting nature's assembly manual for a ssRNA virus. Proc Natl Acad Sci U S A 114, 12255-12260.

Perez-Gonzalez, C., Lafontaine, D.A., and Penedo, J.C. (2016). Fluorescence-Based Strategies to Investigate the Structure and Dynamics of Aptamer-Ligand Complexes. Frontiers in chemistry 4, 33.

Perlmutter, J.D., and Hagan, M.F. (2015). Mechanisms of Virus Assembly. Annual Review of Physical Chemistry 66, 217-239.

Perni, R.B., Almquist, S.J., Byrn, R.A., Chandorkar, G., Chaturvedi, P.R., Courtney, L.F., Decker, C.J., Dinehart, K., Gates, C.A., Harbeson, S.L., et al. (2006). Preclinical profile of VX-950, a potent, selective, and orally bioavailable inhibitor of hepatitis C virus NS3-4A serine protease. Antimicrob Agents Chemother. 50, 899-909.

Phillips, S.E.V., Lane, S.W., Dennis, C.A., Lane, C.L., Trinh, C.H., Rizkallah, P.J., Bunka, D.H., Dykeman, E.C., Ford, R., Barker, A., et al. (2012). Structure and RNA Recognition in Recombinant STNV Capsids. Biophysical Journal 102, 641a.

Pielak, R.M., Oxenoid, K., and Chou, J.J. (2011). Structural investigation of rimantadine inhibition of the AM2-BM2 chimera channel of influenza viruses. Structure (London, England) 19, 1655-1663.

Pohl, F.M. and Jovin, T.M. (1972). Salt-induced co-operative conformational change of a synthetic DNA: Equilibrium and kinetic studies with poly(dG-dC). Journal of Molecular Biology 67, 375-396.

Popov, O.V., Sumarokov, A.A., Shkol'nik, R., El'bert, L.B., and Vorob'eva, M.S. (1985). Reactogenicity and antigenic activity of a chromatographic cultured purified and concentrated inactivated dried vaccine against tick-borne encephalitis. Zhurnal mikrobiologii, epidemiologii, immunobiologii, 34-39

Prasad, B.V.V. and Schmid, M.F. (2012). Principles of Virus Structural Organization. Advances in experimental medicine and biology 726, 17-47.

Quigley, G.J., Teeter, M.M., and Rich, A. (1978). Structural analysis of spermine and magnesium ion binding to yeast phenylalanine transfer RNA. Proceedings of the National Academy of Sciences 75, 64-68.

Rana, T.M. (2007). Illuminating the silence: understanding the structure and function of small RNAs. Nature Reviews Molecular Cell Biology 8, 23.

Reha, D. Kabelac, M. Ryjacek, F. Sponer, J.E. Elstner, M. Suhai, S.and Hobza, P. (2002). Intercalators. 1. Nature of stacking interactions between intercalators (ethidium, daunomycin, ellipticine, and 4’,6-diaminide-2-phenylindole) and DNA base pairs. Ab initio quantum chemical, density functional theory, and empirical potential study. J Am Chem Soc, 124, 3366–3376.

Rigler, R. and Elson, E.S. (2001). Fluorescence Correlation Spectroscopy. (Berlin: Springer).

Roberts, J.W., and Steitz, J.E. (1967). The reconstitution of infective bacteriophage R17. Proceedings of the National Academy of Sciences 58, 1416-1421.

Rogers, D.J. and Tanimoto, T.T. (1960). A Computer Program for Classifying Plants. Science 132, 1115-1118.

Rolfsson, Ó., Middleton, S., Manfield, I.W., White, S.J., Fan, B., Vaughan, R., Ranson, N.A., Dykeman, E., Twarock, R., Ford, J., et al. (2015). Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2. Journal of Molecular Biology.

Rolfsson, Ó. (2009). The roles of MS2 RNA in MS2 capsid assembly. The University of Leeds. PhD Thesis.

Rolfsson, Ó., Toropova, K., Ranson, N.A., and Stockley, P.G. (2010). Mutually-induced Conformational Switching of RNA and Coat Protein Underpins Efficient Assembly of a Viral Capsid. Journal of Molecular Biology 401, 309-322.

Romaniuk, P.J., Lowary, P., Wu, H.N., Stormo, G., and Uhlenbeck, O.C. (1987). RNA binding site of R17 coat protein. Biochemistry 26, 1563-1568.

Rombaut, B., Foriers, A., and Boeyé, A. (1991). In vitro assembly of poliovirus 14 S subunits: Identification of the assembly promoting activity of infected cell extracts. Virology 180, 781-787.

Ross, J.I., Eady, E.A., Cove, J.H., and Cunliffe, W.J. (1998). 16S rRNA mutation associated with tetracycline resistance in a gram- positive bacterium. Antimicrob Agents Ch 42, 1702-1705.

Rosu, F., De Pauw, E., and Gabelica, V. (2008). Electrospray mass spectrometry to study drug-nucleic acids interactions. Biochimie 90, 1074-1087.

Rosu, F., Pirotte, S., Pauw, E.D., and Gabelica, V. (2006). Positive and negative ion mode ESI-MS and MS/MS for studying drug–DNA complexes. International Journal of Mass Spectrometry 253, 156-171.

Rowsell, S., Stonehouse, N.J., Convery, M.A., Adams, C.J., Ellington, A.D., Hirao, I., Peabody, D.S., Stockley, P.G., and Phillips, S.E.V. (1998). Crystal structures of a series of RNA aptamers complexed to the same protein target. Nature Structural Biology 5, 970 - 975.

Ruibal-Ares, B.H., Belmonte, L., Bare, P.C., Parodi, C.M., Massud, I., and de Bracco, M.M. (2004). HIV-1 infection and chemokine receptor modulation. Curr Hiv Res 2, 39-50.

Rupp, B. (2009). Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology (New York: Garland Science, Taylor and Francis Group).

Sabin, A.B. (1957). Present status of attenuated live-virus poliomyelitis vaccine. *Bull. N. Y. Acad. Med.* 33, 17–39.

Salk, J. E. et al. (1954) Formaldehyde treatment and safety testing of experimental poliomyelitis vaccines. *Am. J. Public Health Nations Health* 44, 563–570.

Sanjuán, R., and Domingo-Calap, P. (2016). Mechanisms of viral mutation. Cellular and Molecular Life Sciences 73, 4433-4448.

Sarrazin, C., Kieffer, T.L., Bartels, D., Hanzelka, B., Müh, U., Welker, M., Wincheringer, D., Zhou, Y., Chu, H.-M., Lin, C., et al. (2007). Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. Gastroenterology 132, 1767-1777.

Satyanarayana, S., Dabrowiak, J.C., and Chaires, J.B. (1993). Tris(phenanthroline)ruthenium(II) enantiomer interactions with DNA: Mode and specificity of binding. Biochemistry 32, 2573-2584.

Schneemann, A. (2006). The Structural and Functional Role of RNA in Icosahedral Virus Assembly. Annu Rev Microbiol 60, 51–67.

Scheid, J.F., Horwitz, J.A., Bar-On, Y., Kreider, E.F., Lu, C.-L., Lorenzi, J.C.C., Feldmann, A., Braunschweig, M., Nogueira, L., Oliveira, T., et al. (2016). HIV-1 antibody 3BNC117 suppresses viral rebound in humans during treatment interruption. Nature. 535, 556-560.

Schuck, P. (2000). Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. Biophysical Journal 78, 1606-1619.

Scott, L.J. and Figgitt, D.P. (2004). Mitoxantrone: a review of its use in multiple sclerosis. CNS drugs 18, 379-396.

Seitsonen, J., Susi, P., Heikkilä, O., Sinkovits, R.S., Laurinmäki, P., Hyypiä, T., and Butcher, S.J. (2010). Interaction of alphaVbeta3 and alphaVbeta6 integrins with human parechovirus. J Virol 84, 8509-8519.

Schaffer, M., Peng, G., Spingler, B., Schnabl, J., Wang, M., Olieric, V., and Sigel, R. (2016). The X-ray Structures of Six Octameric RNA Duplexes in the Presence of Different Di- and Trivalent Cations. International Journal of Molecular Sciences 17, 988.

Schuck, P. (2016). Sedimentation Velocity Analytical Ultracentrifugation Discrete Species and Size-Distributions of Macromolecules and Particles. (England: CRC Press).

Shakeel, S., Dykeman, E.C., White, S.J., Ora, A., Cockburn, J.J.B., Butcher, S.J., Stockley, P.G., and Twarock, R. (2017). Genomic RNA folding mediates assembly of human parechovirus. Nature Communications 8, 5.

Shiba, T. and Suzuki, Y. (1981). Localization of A protein in the RNA-A protein complex of RNA phage MS2. Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis 654, 249-255.

Shin, G.-A. and Sobsey, M.D. (2003). Reduction of Norwalk Virus, Poliovirus 1, and Bacteriophage MS2 by Ozone Disinfection of Water. Applied and Environmental Microbiology 69, 3975-3978.

Shtatland, T., Gill, S.C., Javornik, B.E., Johansson, H.E., Singer, B.S., Uhlenbeck, O.C., Zichi, D.A., and Gold, L. (2000). Interactions of Escherichia coli RNA with bacteriophage MS2 coat protein: genomic SELEX. Nucleic Acids Res 28, E93.

Sikkema, F.D., Comellas-Aragones, M., Fokkink, R.G., Verduin, B.J., Cornelissen, J.J., and Nolte, R.J. (2007). Monodisperse polymer-virus hybrid nanoparticles. Org Biomol Chem 5, 54-57.

Skilling, J. (1989). Classic Maximum Entropy (Kluwer).

Sosnick, T.R., Fang, X., and Shelton, V.M. (2000). Application of circular dichroism to study RNA folding transitions. In Method Enzymol (Academic Press), pp. 393-409.

Sosnick, T.R., and Pan, T. (2003). RNA folding: models and perspectives. Current Opinion in Structural Biology 13, 309-316.

Stanley, W.M. (1935). Isolation of a crystalline protein possessing the properties of tobacco mosaic virus. Science 81, 644-645.

Steinhauer, D.A., Domingo, E., and Holland, J.J. (1992). Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. Gene 122, 281-288.

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

Stelzer, A.C., Frank, A.T., Kratz, J.D., Swanson, M.D., Gonzalez-Hernandez, M.J., Lee, J., Andricioaei, I., Markovitz, D.M., and Al-Hashimi, H.M. (2011). Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. Nat Chem Biol 7, 553-559.

Stockley, P.G., Ashcroft, A.E., Francese, S., Thompson, G.S., Ranson, N.A., and Smith, A.M. (2005). Dissecting the fine details of assembly of a T=3 phage capsid. J Theor Med 6, 119-125.

Stockley, P. G., Twarock, R., Bakker, S. E., Barker, A. M., Borodavka, A., Dykeman, E. C., Ranson, N. A. (2013). Packaging signals in single-stranded RNA viruses: nature's alternative to a purely electrostatic assembly mechanism. Journal of Biological Physics, 39(2), 277-287.

Stockley, P.G., Rolfsson, O., Thompson, G.S., Basnak, G., Francese, S., Stonehouse, N.J., Homans, S.W., and 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, 541-552.

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

Stray, S.J., Bourne, C.R., Punna, S., Lewis, W.G., Finn, M.G., and Zlotnick, A. (2005). A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. . Proc Natl Acad Sci USA 102, 8138–8143.

Streatfield, S.J. (2005). Oral hepatitis B vaccine candidates produced and delivered in plant material. Immunol Cell Biol 83, 257-262.

Study, G.B.o.D. (2013). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study. In The Lancet (Mortality and Causes of Death Collaborators: Elsevier), pp. 117-171.

Su, R.T. and Taylor, M.W. (1976). Morphogenesis of Picornaviruses: Characterization and Assembly of Bovine Enterovirus Subviral Particles. J Gen Virol 30, 317-328.

Sugiyama, T., Hebert, R.R., and Hartman, K.A. (1967). Ribonucleoprotein complexes formed between bacteriophage MS2 RNA and MS2 protein in vitro. J Mol Biol 25, 455-463.

Sundquist, W.I., and Krausslich, H.G. (2012). HIV-1 assembly, budding, and maturation. Cold Spring Harbor perspectives in medicine 2, a006924.

Suttle, C.A. (2005). Viruses in the sea. Nature. 437, 356-361.

Svedberg, and Pedersen (1940). The Ultracentrifuge. Oxford University Press, London. Biol 375, 824-836.

Swalley, S.E., Baird, E.E., and Dervan, P.B. (1999). Effects of γ-Turn and β-Tail Amino Acids on Sequence-Specific Recognition of DNA by Hairpin Polyamides. J Am Chem Soc 121, 1113-1120.

Tabor, C.W. and Tabor, H. (1984). POLYAMINES. Annu Rev Biochem 53, 749-790.

Talbot, S.J., Goodman, S., Bates, S.R.E., Fishwick, C.W.G., and Stockley, P.G. (1990). Use of synthetic oligoribonucleotides to probe RNA- protein interactions in the MS2 translational operator complex. Nucl Acids Res 18, 3521.

Tan, W., and Yuan, G. (2013). Electrospray ionization mass spectrometric exploration of the high-affinity binding of three natural alkaloids with the mRNA G-quadruplex in the BCL2 5'-untranslated region. Rapid Commun Mass Spectrom 27, 560-564.

Tang, C., Loeliger, E., Kinde, I., Kyere, S., Mayo, K., Barklis, E., Sun, Y., Huang, M., and Summers, M.F. (2003). Antiviral Inhibition of the HIV-1 Capsid Protein. Journal of Molecular Biology 327, 1013-1020.

Tanimoto, T. (1957). An Elementary Mathematical theory of Classification and Prediction. Internal IBM Technical Report.

Tars, K., Bundule, M., Fridborg, K., and Liljas, L. (1997). The crystal structure of bacteriophage GA and a comparison of bacteriophages belonging to the major groups of Escherichia coli leviviruses1. Journal of Molecular Biology 271, 759-773.

Theiler, M., and Smith, H.H. (1937). The effect of prolonged cultivation in vitro upon the pathogenicity of yellow fever virus. The Journal of Experimental Medicine 65, 767-786

Thomas, B., and Akoulitchev, A.V. (2006). Mass spectrometry of RNA. Trends in Biochemical Sciences 31, 173-181.

Thomas, J.R., and Hergenrother, P.J. (2008). Targeting RNA with Small Molecules. Chemical Reviews 108, 1171-1224.

Thomas, T.J., Gunnia, U.B., and Thomas, T. (1991). Polyamine-induced B-DNA to Z-DNA Conformational Transition of a Plasmid DNA with (dG-dC), Insert. J Biol Chem 266, 6137-6141.

Toropova, K., Basnak, G., Twarock, R., Stockley, P.G., and Ranson, N.A. (2008). The three-dimensional structure of genomic RNA in bacteriophage MS2: Implications for assembly. J Mol Trieb, M., Rauch, C., Wellenzohn, B., FajarWibowo, Loerting, T., Mayer, E., and Liedl, K.R. (2012). Daunomycin Intercalation Stabilizes Distinct Backbone Conformations of DNA. Journal of Biomolecular Structure & Dynamics 21, 713-724.

Tullius, T. (2009). Structural biology: DNA binding shapes up. Nature 461, 1225-1226.

Tversky, A. (1977). Features of Similarity. Psychological Reviews 84, 327-352.

Valegård, K., Liljas, L., Fridborg, K., and Unge, T. (1990). The 3-dimensional structure of the bacterial-virus MS2. Nature 345, 36-41.

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

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

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

van Duin, J. (1988). The single stranded RNA bacteriophages., Vol I (New York: Plenum press).

Venkatakrishnan, B., Katen, S.P., Francis, S., Chirapu, S., Finn, M.G., and Zlotnick, A. (2016). Hepatitis B Virus Capsids Have Diverse Structural Responses to Small-Molecule Ligands Bound to the Heteroaryldihydropyrimidine Pocket. J Virol 90, 3994-4004.

Vicens, Q., and Westhof, E. (2003). RNA as a drug target: The case of aminoglycosides. ChemBioChem 4, 1018-1023.

Voet, D., Pratt, C.W., and Voet, J.G. (2013). Principles of Biochemistry, 4th Edition International Student Version, Vol 4 (Wiley).

von der Helm, K. (1996). Retroviral proteases: structure, function and inhibition from a non-anticipated viral enzyme to the target of a most promising HIV therapy. Biological chemistry 377, 765-774

Wade, W.S., Mrksich, M., and Dervan, P.B. (1992). Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. J Am Chem Soc 114, 8783-8794.

Wang, C., Takeuchi, K., Pinto, L.H., and Lamb, R.A. (1993). Ion channel activity of influenza A virus M2 protein: characterization of the amantadine block. J Virol 67, 5585–5594.

Wang, K., Su, B., Wang, Z., Wu, M., Li, Z., Hu, Y., Fan, Z., Mi, N., and Wang, Q. (2010). Synthesis and antiviral activities of phenanthroindolizidine alkaloids and their derivatives. J Agric Food Chem 58, 2703-2709.

Wang, Z., Wei, P., Wang, L., and Wang, Q. (2012). Design, Synthesis, and Anti-tobacco Mosaic Virus (TMV) Activity of Phenanthroindolizidines and Their Analogues. J Agr Food Chem 60, 10212-10219.

Ward, A.B., and Wilson, I.A. (2017). The HIV-1 envelope glycoprotein structure: nailing down a moving target. Immunological Reviews 275, 21-32.

Ward, P., Small, I., Smith, J., Suter, P., and Dutkowski, R. (2005). Oseltamivir (Tamiflu) and its potential for use in the event of an influenza pandemic. The Journal of antimicrobial chemotherapy 55 Suppl 1, i5-i21.

Waring, M.J. (1965). Complex formation between ethidium bromide and nucleic acids. Journal of Molecular Biology 13, 269–282.

Watson, J.D., and Crick, F.H.C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 171, 737-738.

Wei, X., Decker, J.M., Wang, S., Hui, H., Kappes, J.C., Wu, X., Salazar-Gonzalez, J.F., Salazar, M.G., Kilby, J.M., Saag, M.S., et al. (2003). Antibody neutralization and escape by HIV-1. Nature 422, 307-312.

Wienken, C.J., Baaske, P., Rothbauer, U., Braun, D., and Duhr, S. (2010). Protein-binding assays in biological liquids using microscale thermophoresis. Nat Commun 1, 100.

Westheimer, F. (1987). Why nature chose phosphates. Science 235, 1173-1178.

White, J.P., and Cantor, C.R. (1971). Role of magnesium in the binding of tetracycline to Escherichia coli ribosomes. Journal of Molecular Biology 58, 397-400.

White, R.J., and Durr, F.E. (1985). Development of mitoxantrone. Invest New Drugs 3, 85-93.

White, S., Baird, E.E., and B., D.P. (1997). Orientation Preferences of Pyrrole-Imidazole Polyamides in the Minor Groove of DNA. J. Am. Chem. Soc. 119, 8756-8765.

WHO (1998). Global Polio Eradication Initiative (GPEI).

WHO (2009). WHO Weekly Epidemiological Record. "Hepatitis B vaccines position paper". 40 (84): 405–420.

WHO (2013). WHO Expert Consultation on Rabies: second report. In WHO Technical Report Series, ed., ed. (Geneva).

Wilby, K.J., Partovi, N., Ford, J.-A.E., Greanya, E.D., and Yoshida, E.M. (2012). Review of boceprevir and telaprevir for the treatment of chronic hepatitis C. Can. J. Gastroenterol. 26, 205-210.

Wilkins, M.H.F., Stokes, A.R., and Wilson, H.R. (1953). Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. Nature 171, 738-740.

Witherell, G.W., Gott, J.M., and Uhlenbeck, O.C. (1991). Specific interaction between RNA phage coat proteins and RNA. Progr. Nucl. Acid Res. Mol. Biol. 40, 185–220.

Wu, H.N., and Uhlenbeck, O.C. (1987). Role of a bulged A residue in a specific RNA-protein interaction. Biochemistry 26, 8221–8227.

Xi, Z., Zhang, R., Yu, Z., Ouyanga, D., and Huang, R. (2005). Selective interaction between tylophorine B and bulged DNA. Bioorganic & Medicinal Chemistry Letters 15, 2673–2677.

Yamagishi, A. (1984). Electric Dichroism Evidence for Stereospecific Binding of Optically Active Tris Chelated Complexes to DNA. J. Phys. Chem. 88, 5709-5713.

Yang, W.-Y., He, F.J., Strack, R.L., Oh, S.Y., Frazer, M., Jaffrey, S.R., Todd, P.K., and Disney, M.D. (2016). Small Molecule Recognition and Tools to Study Modulation of r(CGG)exp in Fragile X-Associated Tremor Ataxia Syndrome. ACS Chemical Biology 11, 2456-2465.

Zander, C., Enderlein, J., and Keller, R.A. (2003). Single Molecule Detection in Solution: Methods and Applications. In Single Molecule Detection in Solution (Wiley-VCH Verlag GmbH & Co. KGaA), pp. i-xvi.

Zhao, H., Brautigam, C.A., Ghirlando, R., and Schuck, P. (2013). Current Methods in Sedimentation Velocity and Sedimentation Equilibrium Analytical Ultracentrifugation. Current protocols in protein science / editorial board, John E Coligan [et al] 0 20, 10.1002/0471140864.ps0471142012s0471140871.

Zhao, F., Zhao, Q., Blount, K.F., Han, Q., Tor, Y., and Hermann, T. (2005). Molecular recognition of RNA by neomycin and a restricted neomycin derivative. Angew Chem Int. Ed. Engl. 44, 5329-5334.

Zheng, S., Chen, Y., Donahue, C.P., Wolfe, M.S., and Varani, G. (2009). Structural basis for stabilization of the tau pre-mRNA splicing regulatory element by novantrone (mitoxantrone). Chem. Biol. 16, 557-566.

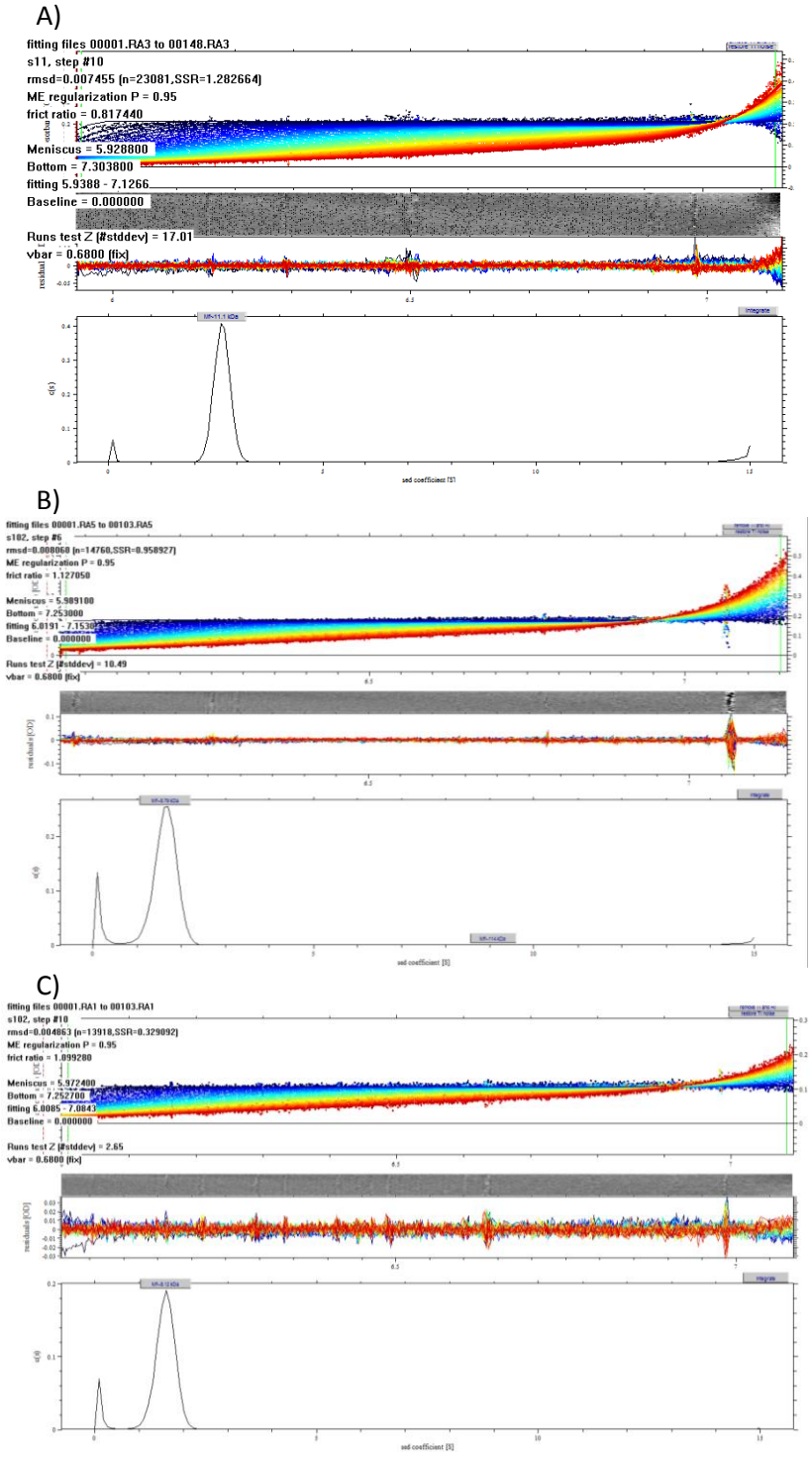
Zipper, H., Brunner, H., Bernhagen, J., and Vitzthum, F. (2004). Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. Nucleic Acids Res 32, e103.

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

Chapter 8

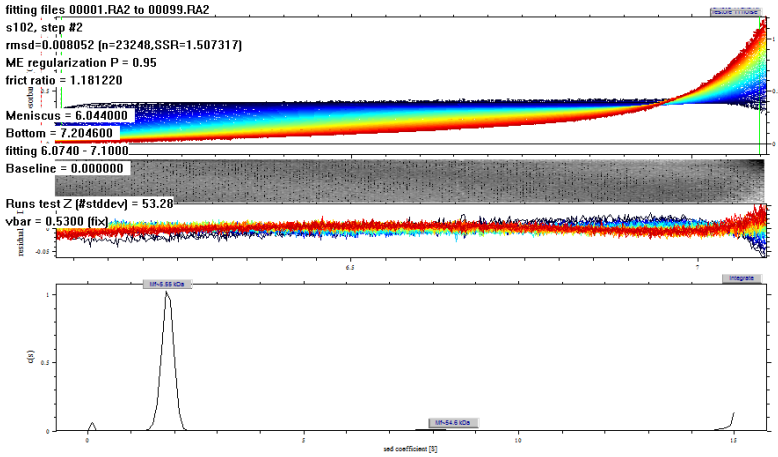
8. Appendix

This page has been left intentionally blank.

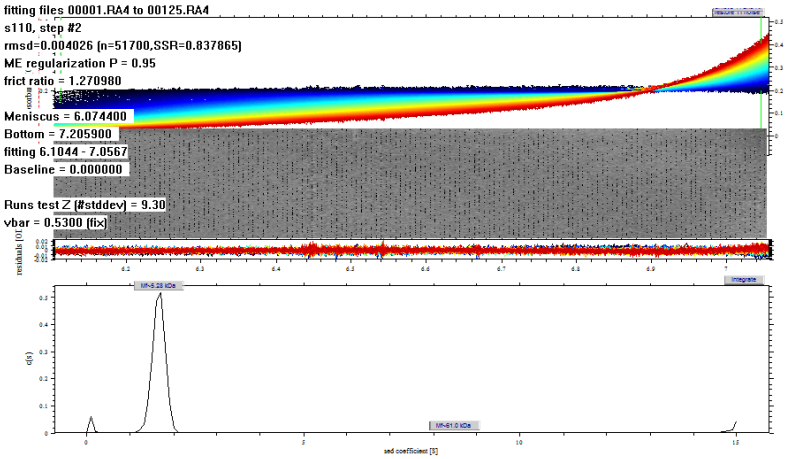


Appendix Figure 8.1.1 Screen shots of data analysis of svAUC of TR [1 µM] in the presence of ligands A) Daunomycin [300 nM], B) Dimethyl amorilide (DMA) [490 nM] and C) Mitoxantrone (MTX) [2.5 µM]. Collecting absorbance scans at 260 nm, scanning from the top of cell towards the outer rotor edge. The first scan is shown in dark blue through to the last in red. Data were collected over ~12 hours at 30,000 rpm with ~8 minutes between each scan. Raw data is shown only for the lowest concentration of ligand tested. Data analysis is carried out using Sednterp version 1.09 (Laue, Shah et al. 1992). Radial absorbance plots were fit to sedimentation profiles with the program Sedfit, version 12.1b 2010, using a continuous distribution c(s) Lamm equation model (Schuck 2000) as detailed in Section 2.7, Chapter 2.

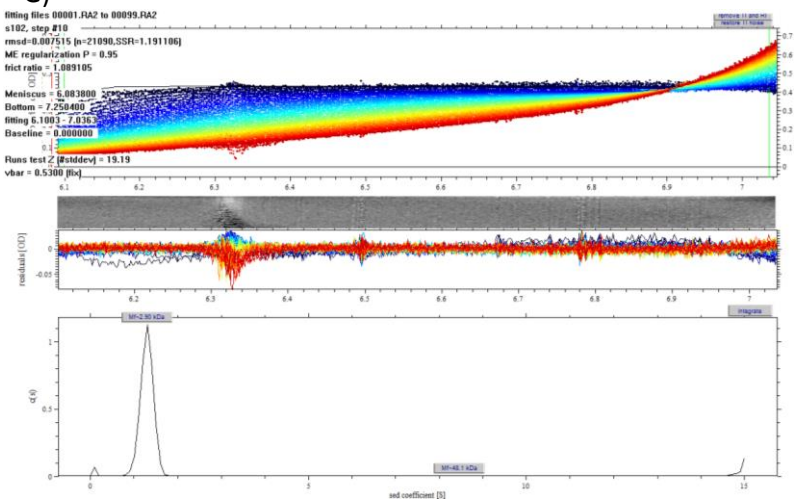
D)



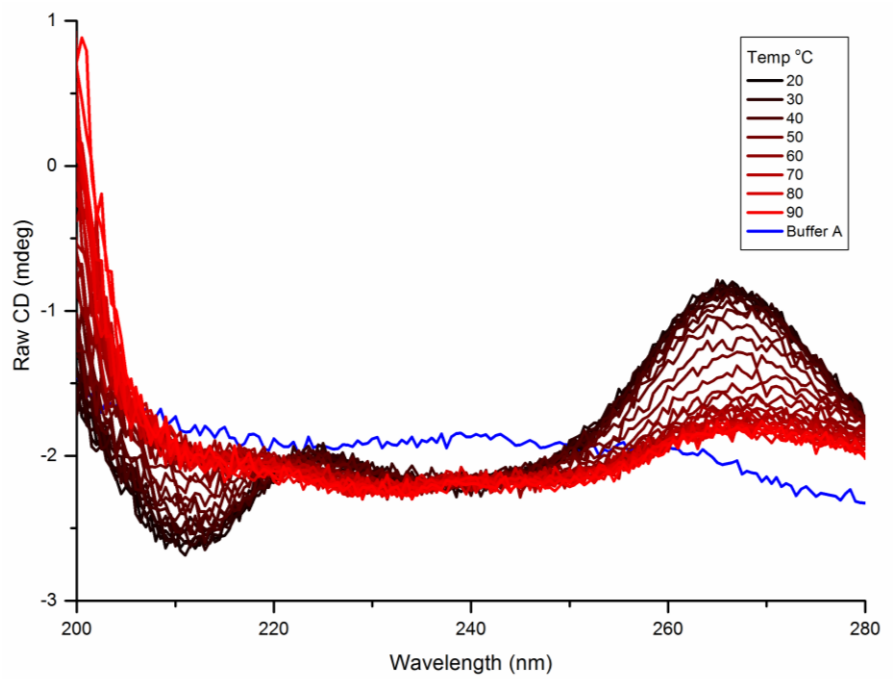
F)



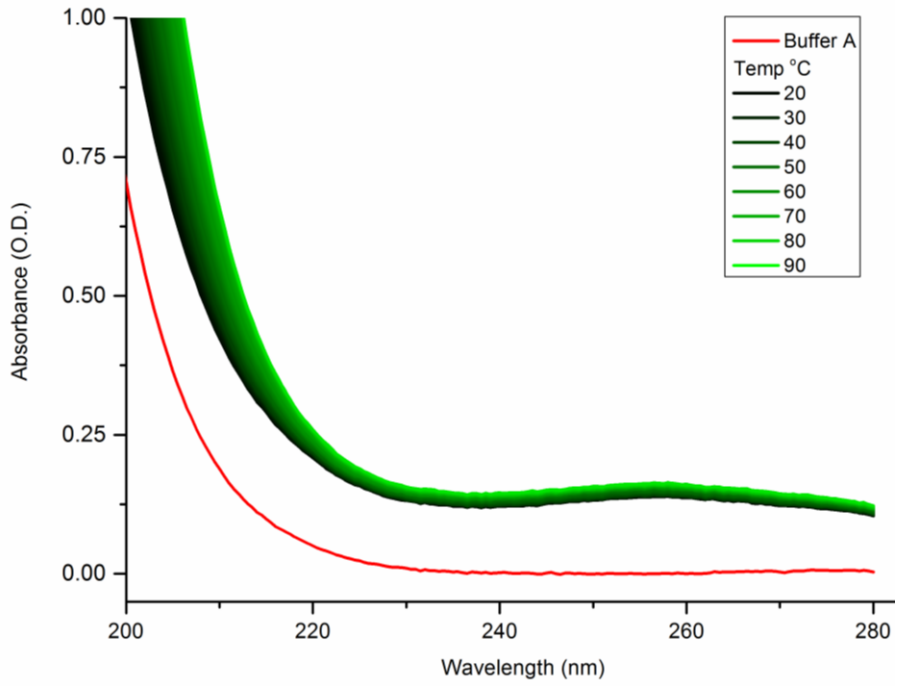
G)



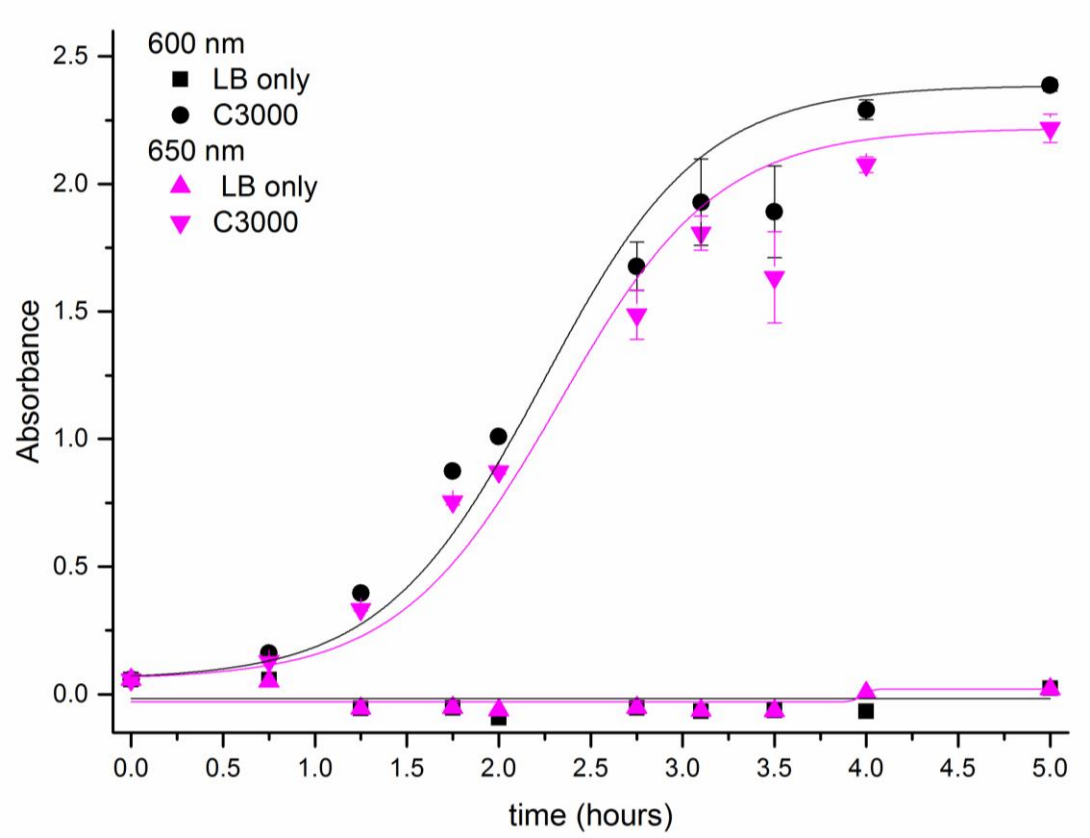
Appendix Figure 8.2.2 Screen shots of data analysis of svAUC of TR [1 µM] in the presence of ligands D) SC19220 [100 nM], E) A3 [1 nM] and F) Tylophorine [100 nM]. Collecting absorbance scans at 260 nm, scanning from the top of cell towards the outer rotor edge. The first scan is shown in dark blue through to the last in red. Data were collected over ~12 hours at 30,000 rpm with ~8 minutes between each scan. Raw data is shown only for the lowest concentration of ligand tested. Data analysis is carried out using Sednterp version 1.09 (Laue, Shah et al. 1992). Radial absorbance plots were fit to sedimentation profiles with the program Sedfit, version 12.1b 2010, using a continuous distribution c(s) Lamm equation model (Schuck 2000) as detailed in Section 2.7, Chapter 2.



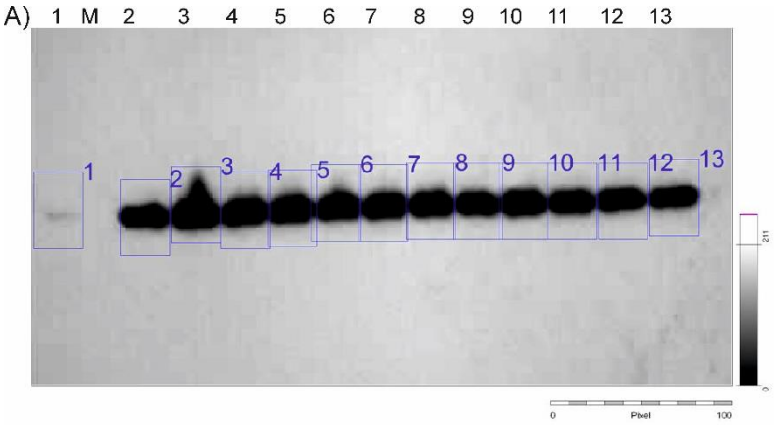
Appendix Figure 8.2 Thermal unfolding of TR RNA monitored by CD. The raw CD spectrum of TR [7.4 µM] in Buffer A. Scanning at 0.5 nm intervals with a heat ramp of 1.0 °C.minute^{-1}, spectra were recorded from 20 °C (black) to 90 °C (red). CD data for Buffer A only is shown (blue). Increasing the temperature generates a change in the CD signal denoting a change in the secondary structure of RNA.



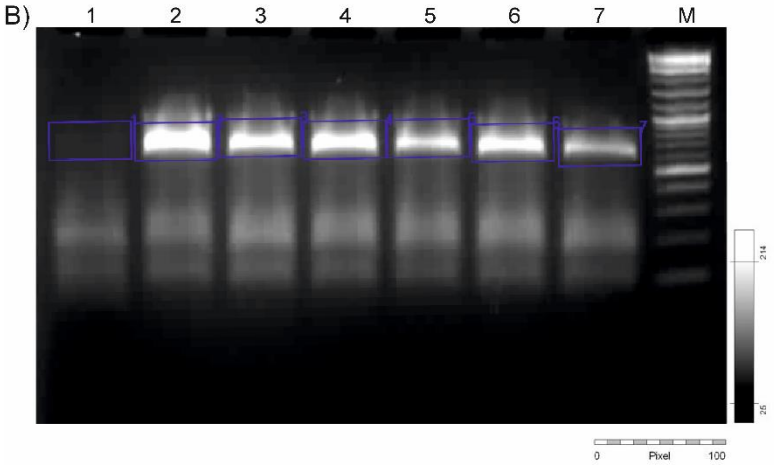
Appendix Figure 8.3 Thermal unfolding of TR [7.4 µM] monitored by absorbance in Buffer A. Scanning at 0.5 nm intervals with a heat ramp of 1.0 °C.minute^{-1}, spectra were recorded from 20 °C (black) to 90 °C (red). CD data for Buffer A only is shown (blue). Increasing the temperature generates hyperchromicity denoting a change in the secondary structure of RNA.



Appendix 8.4 E. coli C3000 cell growth monitored by absorbance at 600 and 650 nm. Data show similar values during growth and lag phases, suggesting either wavelength can be used as a guide for cell density. Cells grown in supplemented, minimal LB at 37.0°C with shaking at 160 rpm.

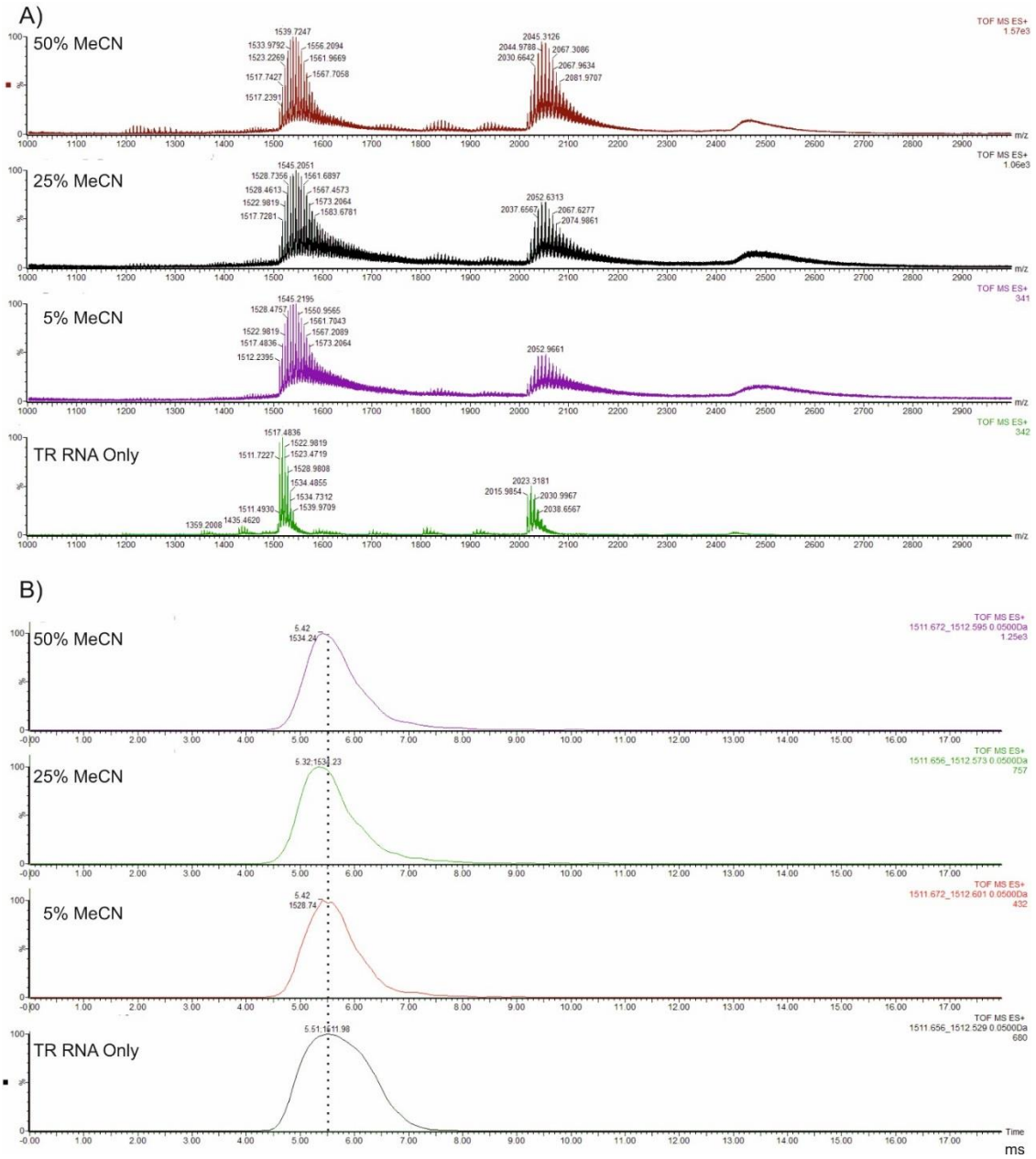


Lane	MTX [µM]	Intensity	Corrected for Background	Relative Intensity
1	0	100.411	13.618	-
2	0	142.098	55.305	1
3	0.50	168.11	81.317	1.470
4	0.54	150.01	63.217	1.143
5	0.58	146.041	59.248	1.071
6	0.63	140.991	54.198	0.980
7	0.68	136.54	49.747	0.900
8	0.73	133.873	47.080	0.851
9	0.79	133.88	47.087	0.851
10	0.86	138.16	51.367	0.929
11	0.93	137.817	51.024	0.923
12	1.00	137.29	50.497	0.913
13	1.10	135.689	48.896	0.884
Background	-	86.793	0	-

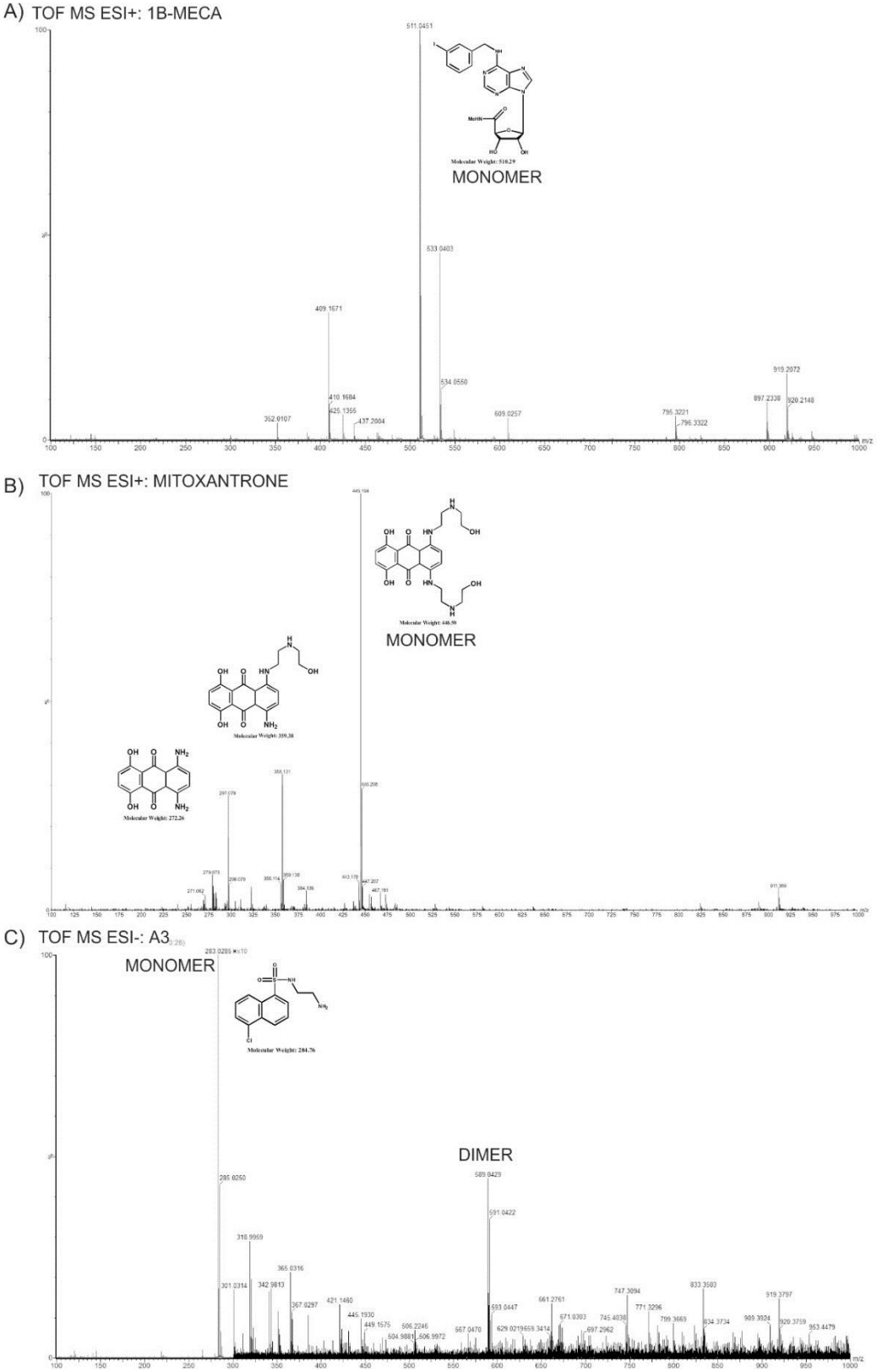


Lane	MTX [µM]	Intensity	Corrected for Background	Relative Intensity
1	0	64.651	0.069	-
2	0	129.385	64.803	1.000
3	0.50	115.128	50.546	0.780
4	0.63	116.197	51.615	0.796
5	0.73	102.959	38.377	0.592
6	0.93	110.000	45.418	0.701
7	1.10	85.969	21.387	0.330
Background	-	64.582	0	-

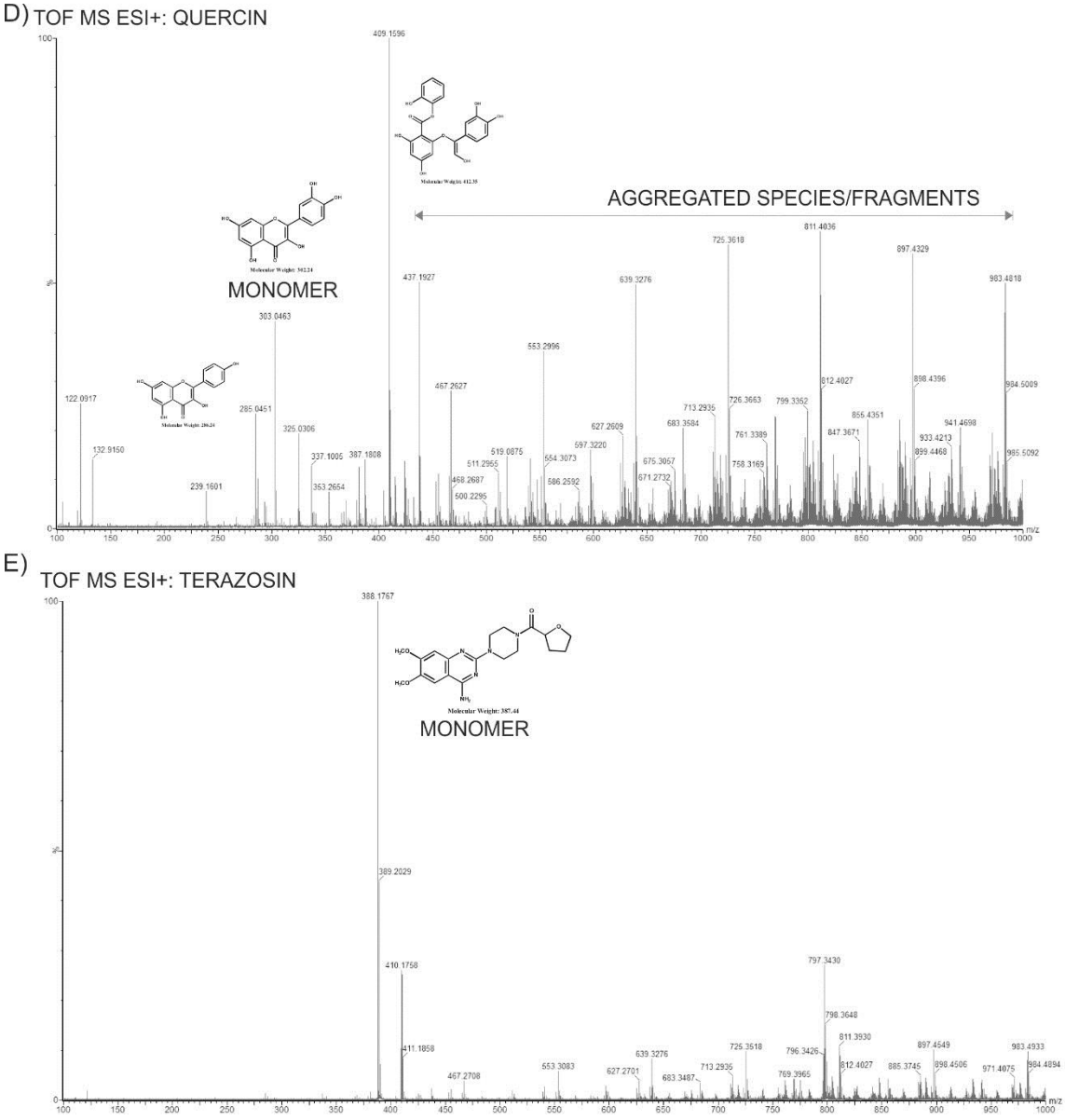
Appendix 8.5 Quantitating the effect of MTX on MS2 CP expression and gRNA replication. Gels analysed and band intensities compared using a defined area, against controls with no MTX. A) Western Blot analysis of lysates selecting for MS2 CP_2. 1) Inoculum MS2. M) MWt. Marker. 2) 0 MTX. 3)-13) 0.5-1.1 µM MTX. B) RT-PCR of infection lysates selecting for 620 bp region of MS2 genome. 1) no MS2. 2) 0 MTX. 3)-7) 0.5 – 1.1 µM MTX. M) MWt. marker. 2-log ladder (NEB). Analysed using AIDA Image Analyzer, Raytest, Version 4.14.



Appendix Figure 8.6 TR RNA Ion Mobility MS experiment in MeCN. A) Raw m/z spectra of positively charged ionised RNA. B) Arrival time distribution for major species $[M+4H]^{4+}$.Ion mobility is a technique that can separate gaseous ions based on their size and shape. Large ions experience more collisions with gases within the chamber and take more time (drift time) to traverse the chamber than smaller ions. Results show that increasing amounts of MeCN solvent (0 - 50%) has no effect on the secondary structure – as measured by drift time of the RNA.



Appendix Figure 8.7.1 ESI MS of Ligands in Cocktail #1. A) 2CL-1B-MECA monomer species, 511.05 Da. B) MTX shows monomer species 445.19 Da and smaller fragments, structures identified inset. C) A3 major species are monomer and dimer.



Appendix Figure 8.7.2 ESI MS of Ligands in Cocktail #1. D) MS of Quercin shows that it is a mix of fragments, monomer and aggregated species. The major species is monomer, 303.06 Da. E) Terazosin shows the major species is monomer, 388.18 Da.

Appendix 8.7.3 Ligand cocktail #1 information.

	Ligand	Mass (Da)	Comments
A	2CL-1B-MECA	511.05	Potent and selective A3 adenosine receptor agonist (Kim *et al.* 1994).
B	MTX	444.48	An antineoplastic, cytotoxic chemotherapy drug can also be used for treatment of multiple sclerosis (Scott and Figgitt 2004).
C	A3	285.03	Inhibits cAMP and cGMP dependent protein kinases (Inagaki *et al.* 1986).
D	Quercin	303.05	Competitive towards the nucleotides ATP and GTP as substrates for reverse transcription (Graziani *et al.* 1983).
E	Terazosin	388.18	α1 antagonist, blocks action of adrenaline on muscle of bladder and blood vessel walls (Kyncl 1986).

Appendix Table 8.8 Details of RNA binding ligands

Ligand **CAS#** **Supplier**	**MWt.**	**Solubility** **$mg.mL^{-1}$** **Aq,** **DMSO**	**Pharmacokinetic Information**	**Mode of Binding**	**Structural Features & Experimental Considerations**
Daunorubicin **23541-50-6** **LKT laboratories**	527.52	113,72	Suppresses acute leukaemia proliferation. Chemotherapeutic (Creutzig 2013)	DNA intercalator, prefers G-C bp and binds the minor groove of DNA	Anthracycline binds per 4-6 base pairs (Trieb 2012).
Dimethylamorilide **1214-79-5** **Sigma Aldrich**	294.19	50,50	Inhibitor of Na^+-H^+ exchange (Kleyman 1988)	Binds to apical RNA loops	Pyrazine ring binds apical RNA stem loops at 1:1 molar ratio (Moumné 2012).
Mitoxantrone.2HCl **70476-82-3** **Santa Cruz Biotech. Inc.**	517.40	1,89	Chemotherapeutic , inhibits DNA synthesis, Inhibits Topoisomerase II (Crespi 1986)	Binds RNA A-bulge stem loops by intercalator and electrostatics	Anthracenedione binds adenosine bulges in RNA stem loops (Zheng 2009).
SC19220 **19395-87-0** **Sigma Aldrich**	331.78	1,33	Prostaglandin antagonist (Botella 1995)	Electrostatics	Dibenzoxazepine hydrazide - exact ratio of binding unknown.
A3 hydrochloride **78957-85-4** **Santa Cruz Biotech. Inc.**	321.22	50,50	Protein kinase inhibitor (Masaki Inagaki 1986)	Electrostatics	Naphthalene sulphonamide - exact ratio of binding unknown.
(+)-S-Tylophorine **482-20-2** **Santa Cruz Biotech. Inc.**	393.48	1,30	Immunosuppressant and inflammation inhibitor. Chemotherapeutic through antiproliferative activity and induction of apoptosis (Gao 2004)	DNA intercalator and electrostatics	Phenanthra-indolizidine alkaloid binds bulged DNA stem loops, with a molar stoichiometry of 1:1 (Xi 2005).

Appendix Table 8.9 NextGen Sequencing Results showing identified, aligned genomes and % total. Phx174 (PhiX174) genome is from a library added to all Illumina sequencing as an internal control of the sequencing run. Aligned to MS2 NCBI_001417.2 using CLC Genome Workbench, Qiagen. Sequencing carried out by Dr. Ian Carr, St James' University Hospital, Leeds.

	Name	Consensus length	Total read count	% total
Sample 1				
wtMS2, no passage	MS2	3638	4195190	98.32196959
	Phx174	5342	20039	0.469650707
	Adenovirus type 2	1343	7836	0.183651027
	Human adenovirus C	2506	18260	0.427956580
	Unknown1	1747	11149	0.261297257
	Human adenovirus 2 pIIIa	3143	13635	0.319561225
	Unknown2	1526	220	0.005156103
	Unknown3	2064	295	0.006913866
	Unknown4	1212	164	0.003843641
			4266788	
Sample 3				
MS2 3X passage -MTX	MS2	3637	4448484	99.31396431
	Phx174	5386	28601	0.638527348
	Unknown5	2243	377	0.008416657
	Unknown6	2354	1527	0.034090810
	Unknown7	1402	224	0.005000879
			4479213	
Sample 5				
MS2 3X passage +MTX	MS2	3241	3543819	99.19653421
	phX174	5342	28704	0.803465786
			3572523	

Appendix Table 8.10 cDNA Sequences of MS2 Samples. Identified nucleic acid mutations are capitalised. Reading 5' to 3'.

Sequence: MS2 NCBI_001417.2
gggtgggacc cctttcgggg tcctgctcaa cttcctgtcg agctaatgcc atttttaatg tctttagcga gacgctacca tggctatcgc tgtaggtagc cggaattcca ttcctaggag gtttgacctg tgcgagcttt tagtaccctt gatagggaga acgagacctt cgtcccctcc gttcgcgttt acgcggacgg tgagactgaa gataactcat tctctttaaa atatcgttcg aactggactc ccggtcgttt taactcgact ggggccaaaa cgaaacagtg gcactacccc tctccgtatt cacgggggggc gttaagtgtc acatcgatag atcaaggtgc ctacaagcga agtgggtcat cgtggggtcg cccgtacgag gagaaagccg gtttcggctt ctccctcgac gcacgctcct gctacagcct cttccctgta agccaaaact tgacttacat cgaagtgccg cagaacgttg cgaaccgggc gtcgaccgaa gtcctgcaaa aggtcaccca gggtaatttt aaccttggtg ttgctttagc agaggccagg tcgacagcct cacaactcgc gacgcaaacc attgcgctcg tgaaggcgta cactgccgct cgtcgcggta attggcgcca ggcgctccgc taccttgccc taaacgaaga tcgaaagttt cgatcaaaac acgtggccgg caggtggttg gagttgcagt tcggttggtt accactaatg agtgatatcc agggtgcata tgagatgctt acgaaggttc accttcaaga gtttcttcct atgagagccg tacgtcaggt cggtactaac atcaagttag atggccgtct gtcgtatcca gctgcaaact tccagacaac gtgcaacata tcgcgacgta tcgtgatatg gttttacata aacgatgcac gtttggcatg gttgtcgtct ctaggtatct tgaacccact aggtatagtg tgggaaaagg tgcctttctc attcgttgtc gactggctcc tacctgtagg taacatgctc gagggcctta cggcccccgt gggatgctcc tacatgtcag gaacagttac tgacgtaata acgggtgagt ccatcataag cgttgacgct ccctacgggt ggactgtgga gagacagggc actgctaagg cccaaatctc agccatgcat cgaggggtac aatccgtatg gccaacaact ggcgcgtacg taaagtctcc tttctcgatg gtccatacct tagatgcgtt agcattaatc aggcaacggc tctctagata gagccctcaa ccggagtttg aagcatggct tctaacttta ctcagttcgt tctcgtcgac aatggcggaa ctggcgacgt gactgtcgcc ccaagcaact tcgctaacgg ggtcgctgaa tggatcagct ctaactcgcg ttcacaggct tacaaagtaa cctgtagcgt tcgtcagagc tctgcgcaga atcgcaaata caccatcaaa gtcgaggtgc ctaaagtggc aacccagact gttggtggtg tagagcttcc tgtagccgca tggcgttcgt acttaaatat ggaactaacc attccaattt tcgctacgaa ttccgactgc gagcttattg ttaaggcaat gcaaggtctc ctaaaagatg gaaacccgat tccctcagca atcgcagcaa actccggcat ctactaatag acgccggcca ttcaaacatg aggattaccc atgtcgaaga caacaaagaa gttcaactct ttatgtattg atcttcctcg cgatctttct ctcgaaattt accaatcaat tgcttctgtc gctactggaa gcggtgatcc gcacagtgac gactttacag caattgctta cttaagggac gaattgctca caaagcatcc gaccttaggt tctggtaatg acgaggcgac ccgtcgtacc ttagctatcg ctaagctacg ggaggcgaat ggtgatcgcg gtcagataaa tagagaaggt ttcttacatg acaaatcctt gtcatgggat ccggatgttt tacaaaccag catccgtagc cttattggca acctcctctc tggctaccga tcgtcgttgt ttgggcaatg cacgttctcc aacggtgctc ctatggggca caagttgcag gatgcagcgc cttacaagaa gttcgctgaa caagcaaccg ttacccccccg cgctctgaga gcggctctat tggtccgaga ccaatgtgcg ccgtggatca gacacgcggt ccgctataac gagtcatatg aatttaggct cgttgtaggg aacggagtgt ttacagttcc gaagaataat aaaatagatc gggctgcctg taaggagcct gatatgaata tgtacctcca gaaaggggtc ggtgctttca tcagacgccg gctcaaatcc gttggtatag acctgaatga tcaatcgatc aaccagcgtc tggctcagca gggcagcgta gatggttcgc ttgcgacgat agacttatcg tctgcatccg attccatctc cgatcgcctg gtgtggagtt ttctcccacc agagctatat tcatatctcg atcgtatccg ctcacactac ggaatcgtag atggcgagac gatacgatgg gaactatttt ccacaatggg aaatgggttc acatttgagc tagagtccat gatattctgg gcaatagtca aagcgaccca aatccatttt ggtaacgccg gaaccatagg catctacggg gacgatatta tatgtcccag tgagattgca ccccgtgtgc tagaggcact tgcctactac ggttttaaac cgaatcttcg taaaacgttc gtgtccgggc tctttcgcga gagctgcggc gcgcactttt accgtggtgt cgatgtcaaa ccgttttaca tcaagaaacc tgttgacaat ctcttcgccc tgatgctgat attaaatcgg ctacggggtt ggggagttgt cggaggtatg tcagatccac gcctctataa ggtgtgggta cggctctcct cccaggtgcc ttcgatgttc ttcggtggga cggacctcgc tgccgactac tacgtagtca gcccgcctac ggcagtctcg gtatacacca agactccgta cgggcggctg ctcgcggata cccgtacctc gggtttccgt cttgctcgta tcgctcgaga acgcaagttc ttcagcgaaa agcacgacag tggtcgctac atagcgtggt tccatactgg aggtgaaatc accgacagca tgaagtccgc cggcgtgcgc gttatacgca cttcggagtg gctaacgccg gttcccacat tccctcagga gtgtgggcca gcgagctctc ctcggtagct gaccgaggga cccccgtaaa cggggtgggt gtgctcgaaa gagcacgggt gcgaaagcgg tccggctcca ccgaaaggtg ggcgggcttc ggcccaggga cctcccccta aagagaggac ccgggattct cccgatttgg taactagctg cttggctagt taccaccca

Sequence: wtMS2, no passage

gggtgggacc cctttcgggg tcctgctcaa cttcctgtcg agctaatgcc atttttaatg tctttagcga gacgctacca tggctatcgc tgtaggtagc cggaattcca ttcctaggag gtttgacctg tgcgagcttt tagtaccctt gatagggaga acgagacctt cgtcccctcc gttcgcgttt acgcggacgg tgagactgaa gataactcat tctctttaaa atatcgttcg aactggactc ccggtcgttt taactcgact ggggccaaaa cgaaacagtg gcactacccc tctccgtatt cacgggggggc gttaagtgtc acatcgatag atcaaggtgc ctacaagcga agtgggtcat cgtggggtcg cccgtacgag gagaaagccg gtttcggctt ctccctcgac gcacgctcct gctacagcct cttccctgta agccaaaact tgacttacat cgaagtgccg cagaacgttg cgaaccgggc gtcgaccgaa gtcctgcaaa aggtcaccca gggtaatttt aaccttggtg ttgctttagc agaggccagg tcgacagcct cacaactcgc gacgcaaacc attgcgctcg tgaaggcgta cactgccgct cgtcgcggta attggcgcca gAcgctccgc taccttgccc taaacgaaga tcgaaagttt cgatcaaaac acgtggccgg caggtggttg gagttgcagt tcggttggtt accactaatg agtgatatcc agggtgcata tgagatgctt acgaaggttc accttcaaga gtttcttcct atgagagccg tacgtcaggt cggtactaac atcaagttag atggccgtct gtcgtatcca gctgcaaact tccagacaac gtgcaacata tcgcgacgta tcgtgatatg gttttacata aacgatgcac gtttggcatg gttgtcgtct ctaggtatct tgaacccact aggtatagtg tgggaaaagg tgcctttctc attcgttgtc gactggctcc tacctgtagg taacatgctc gagggcctta cggcccccgt gggatgctcc tacatgtcag gaacagttac tgacgtaata acgggtgagt ccatcataag cgttgacgct ccctacgggt ggactgtgga gagacagggc actgctaagg cccaaatctc agccatgcat cgaggggtac aatccgtatg gccaacaact ggcgcgtacg taaagtctcc tttctcgatT gtccatacct tagatgcgtt agcattaatc aggcaacggc tctctagata gagccctcaa ccggagtttg aagcatggct tctaacttta ctcagttcgt tctcgtcgac aatggcggaa ctggcgacgt Tactgtcgcc ccaagcaact tcgctaacgg ggtcgctgaa tggatcagct ctaactcgcg ttcacaggct tacaaagtaa cctgtagcgt tcgtcagagc tctgcgcaga atcgcaaata caccatcaaa gtcgaggtgc ctaaagtggc aacccagact gttggtggtg tagagcttcc tgtagccgca tggcgttcgt acttaaatat ggaactaacc attccaattt tcgctacgaa ttccgactgc gagcttattg ttaaggcaat gcaaggtctc ctaaaagatg gaaacccgat tccctcagca atcgcagcaa actccggcat ctactaatag acgccggcca ttcaaacatg aggattaccc atgtcgaaga caacaaagaa gttcaactct ttatgtattg atcttcctcg cgatctttct ctcgaaattt accaatcaat tgcttctgtc gctactggaa gcggtgatcc gcacagtgac gactttacag caattgctta cttaagggac gaattgctca caaagcatcc gaccttaggt tctggtaatg acgaggcgac ccgtcgtacc ttagctatcg ctaagctacg ggaggcgaat gAtCatcgcg gtcagataaa tagagaaggt ttcttacatg acaaatcctt gtcatgggat ccggatgttt tacaaaccag catccgtagc cttattggca acctcctctc tggctaccga tcgtcgttgt ttgggcaatg cacgttctcc aacggtgcCT ctatTgggca caagttgcag gatgcagcgc cttacaagaa gttcgctgaa caagcaaccg ttaccccccg cgctctgaga gcggctctat tggtccgaga ccaatgtgcg ccgtggatca gacacgcggt ccgctataac gagtcatatg aatttaggct cgttgtaggg aacggagtgt ttacagttcc gaagaataat aaaatagatc gggctgcctg taaggagcct gatatgaata tgtacctcca gaaaggggtc ggtgcCttTa tcagacgccg gctcaaatcc gttggtatag acctgaatga tcaatcgatc aaccagcgtc tggctcagca gggcagcgta gatggttcgc ttgcgacgat agacttatcg tctgcatccg attccatctc cgatcgcctg gtgtggagtt ttctcccacc Tgagctatat tcatatctcg atcgtatccg ctcacactac ggaatcgtag atggcgagac gatacgatgg gaactatttt ccacaatggg aaatgggttc acatttgagc tagagtccat gatattctgg gcaatagtca aagcgaccca aatccatttt ggtaacgccg gaaccatagg catctacggg gacgatatta tatgtcccag tgagattgca ccccgtgtgc tagaggcact tgcctactac ggttttaaac cgaatcttcg taaaacgttc gtgtccgggc tctttcgcga gagctgcggc gcgcactttt accgtggtgt cgatgtcaaa ccgttttaca tcaagaaacc tgttgacaat ctcttcgccc tgatgctgat attaaatcgg ctacggggtt ggggagttgt cggaggtatg tcagatccac gcctctataa ggtgtgggta cggctctcct cccaggtgcc ttcgatgttc ttcggtggga cTgacctcgc tgccgactac tacgtagtca gcccgcctac ggcagtctcg gtatacacca agactccgta cTggcggctg ctcgcggata cccgtacctc gggtttccgt cttgctcgta tcgctcgaga acgcaagttc ttcagcgaaa agcacgacag tggtcgctac atagcgtggt tccatactgg aggtgaaatc accgacagca tgaagtccgc cggcgtgcgc gttatacgca cttcggagtg gctaacgccg gttcccacat tccctcagga gtgtgggcca gcgagctctc ctcggtagct gaccgaggga cccccgtaaa cggggtgggt gtgctcgaaa gagcacgggt gcgaaagcgg tccggctcca ccgaaaCgtg ggcgggcttc ggcccaggga cctcccccta aagagaggac ccgggattct cccgatttgg taactagctg cttggctagt taccaccca

Sequence: MS2 3X Passage -MTX

gggtgggacc cctttcgggg tcctgctcaa cttcctgtcg agctaatgcc atttttaatg tctttagcga gacgctacca tggctatcgc tgtaggtagc cggaattcca ttcctaggag gtttgacctg tgcgagcttt tagtaccctt gatagggaga acgagacctt cgtcccctcc gttcgcgttt acgcggacgg tgagactgaa gataactcat tctctttaaa atatcgttcg aactggactc ccggtcgttt taactcgact ggggccaaaa cgaaacagtg gcactacccc tctccgtatt cacgggggggc gttaagtgtc acatcgatag atcaaggtgc ctacaagcga agtgggtcat cgtggggtcg cccgtacgag gagaaagccg gtttcggctt ctccctcgac gcacgctcct gctacagcct cttccctgta agccaaaact tgacttacat cgaagtgccg cagaacgttg cgaaccgggc gtcgaccgaa gtcctgcaaa aggtcaccca gggtaatttt aaccttggtg ttgctttagc agaggccagg tcgacagcct cacaactcgc gacgcaaacc attgcgctcg tgaaggcgta cactgccgct cgtcgcggta attggcgcca gAcgctccgc taccttgccc taaacgaaga tcgaaagttt cgatcaaaac acgtggccgg caggtggttg gagttgcagt tcggttggtt accactaatg agtgatatcc agggtgcata tgagatgctt acgaaggttc accttcaaga gtttcttcct atgagagccg tacgtcaggt cggtactaac atcaagttag atggccgtct gtcgtatcca gctgcaaact tccagacaac gtgcaacata tcgcgacgta tcgtgatatg gttttacata aacgatgcac gtttggcatg gttgtcgtct ctaggtatct tgaacccact aggtatagtg tgggaaaagg tgcctttctc attcgttgtc gactggctcc tacctgtagg taacatgctc gagggcctta cggcccccgt gggatgctcc tacatgtcag gaacagttac tgacgtaata acgggtgagt ccatcataag cgttgacgct ccctacgggt ggactgtgga gagacagggc actgctaagg cccaaatctc agccatgcat cgaggggtac aatccgtatg gccaacaact ggcgcgtacg taaagtctcc tttctcgatg gtccatacct tagatgcgtt agcattaatc aggcaacggc tctctagata gagccctcaa ccggagtttg aagcatggct tctaacttta ctcagttcgt tctcgtcgac aatggcggaa ctggcgacgt gactgtcgcc ccaagcaGct tcgctaacgg ggtcgctgaa tggatcagct ctaactcgcg ttcacaggct tacaaagtaa cctgtagcgt tcgtcagagc tctgcgcaga atcgcaaata caccatcaaa gtcgaggtgc ctaaagtggc aacccaAact gttggtggtg tagagcttcc tgtagccgca tggcgttcgt acttaaatat ggaactaacc attccaattt tcgctacgaa ttccgactgc gagcttattg ttaaggcaat gcaaggtctc ctaaaagatg gaaacccgat tccctcagca atcgcagcaa actccggcat ctactaatag acgccggcca ttcaaacatg aggattaccc atgtcgaaga caacaaagaa gttcaactct ttatgtattg atcttcctcg cgatctttct ctcgaaattt accaatcaat tgcttctgtc gctactggaa gcggtgatcc gcacagtgac gactttacag caattgctta cttaagggac gaattgctca caaagcatcc gaccttaggt tctggtaatg acgaggcgac ccgtcgtacc ttagctatcg ctaagctacg ggaggcgaat gAtCatcgcg gtcagataaa tagagaaggt ttcttacatg acaaatcctt gtcatgggat ccggatgttt tacaaaccag catccgtagc cttattggca acctcctctc tggctaccga tcgtcgttgt ttgggcaatg cacgttctcc aacggtgcCT ctatggggca caagttgcag gatgcagcgc cttacaagaa gttcgctgaa caagcaaccg ttacccccg cgctctgaga gcggctctat tggtccgaga ccaatgtgcg ccgtggatca gacacgcggt ccgctataac gagtcatatg aatttaggct cgttgtaggg aacggagtgt ttacagttcc gaagaataat aaaatagatc gggctgcctg taaggagcct gatatgaata tgtacctcca gaaaggggtc ggtgcCttTa tcagacgccg gctcaaatcc gttggtatag acctgaatga tcaatcgatc aaccagcgtc tggctcagca gggcagcgta gatggttcgc ttgcgacgat agacttatcg tctgcatccg attccatctc cgatcgcctg gtgtggagtt ttctcccacc Tgagctatat tcatatctcg atcgtatccg ctcacactac ggaatcgtag atggcgagac gatacgatgg gaactatttt ccacaatggg aaatgggttc acatttgagc tagagtccat gatattctgg gcaatagtca aagcgaccca aatccatttt ggtaacgccg gaaccatagg catctacggg gacgatatta tatgtcccag tgagattgca ccccgtgtgc tagaggcact tgcctactac ggttttaaac cgaatcttcg taaaacgttc gtgtccgggc tctttcgcga gagctgcggc gcgcactttt accgtggtgt cgatgtcaaa ccgttttaca tcaagaaacc tgttgacaat ctcttcgccc tTatgctgat attaaatcgg ctacggggtt ggggagttgt cggaggtatg tcagatccac gcctctataa ggtgtgggta cggctctcct cccaggtgcc ttcgatgttc ttcggtggga cggacctcgc tgccgactac tacgtagtca gcccgcctac ggcagtctcg gtatacacca agactccgta cgggcggctg ctcgcggata cccgtacctc gggtttccgt cttgctcgta tcgctcgaga acgcaagttc ttcagcgaaa agcacgacag tggtcgctac atagcgtggt tccatactgg aggtgaaatc accgacagca tgaagtccgc cggcgtgcgc gttatacgca cttcggagtg gctaacgccg gttcccacat tccctcagga gtgtgggcca gcgagctctc ctcggtagct gaccgaggga cccccgtaaa cggggtgggt gtgctcgaaa gagcacgggt gcgaaagcgg tccggctcca ccgaaaCgtg ggcgggcttc ggcccaggga cctcccccta aagagaggac ccgggattct cccgatttgg taactagctg cttggctagt taccaccca

Sequence: MS2 3X passage +MTX

gggtgggacc cctttcgggg tcctgctcaa cttcctgtcg agctaatgcc atttttaatg tctttagcga gacgctacca tggctatcgc tgtaggtagc cggaattcca ttcctaggag gtttgacctg tgcgagcttt tagtaccctt gatagggaga acgagacctt cgtcccctcc gttcgcgttt acgcggacgg tgagactgaa gataactcat tctctttaaa atatcgttcg aactggactc ccggtcgttt taactcgact ggggccaaaa cgaaacagtg gcactacccc tctccgtatt cacgggggggc gttaagtgtc acatcgatag atcaaggtgc ctacaagcga agtgggtcat cgtggggtcg cccgtacgag gaTaaagccg gtttcggctt ctccctcgac gcacgctcct gctacagcct cttccctgta agccaaaact tgacttacat cgaagtgccg cagaacgttg cgaaccgggc gtcgaccgaa gtcctgcaaa aggtcaccca gggtaatttt aaccttggtg ttgctttagc agaggccagg tcgacagcct cacaactcgc gacgcaaacc attgcgctcg tgaaggcgta cactgccgct cgtcgcggta attggcgcca ggcgctccgc taccttgccc taaacgaaga tcgaaagttt cgatcaaaac acgtggccgg caggtggttg gagttgcagt tcggttggtt accactaatg agtgatatcc agggtgcata tgagatgctt acgaaggttc accttcaaga gtttcttcct atgagagccg tacgtcaggt cggtactaac atcaagttag atggccgtct gtcgtatcca gctgcaaact tccagacaac gtgcaacata tcgcgacgta tcgtgatatg gttttacata aacgatgcac gtttggcatg gttgtcgtct ctaggtatct tgaacccact aggtatagtg tgggaaaagg tgcctttctc attcgttgtc gactggctcc tacctgtagg taacatgctc gagggcctta cggcccccgt gggatgctcc tacatgtcag gaacagttac tgacgtaata acgggtgagt ccatcataag cgttgacgct ccctacgggt ggactgtgga gagacagggc actgctaagg cccaaatctc agccatgcat cgaggggtac aatccgtatg gccaacaact ggcgcgtacg taaagtctcc tttctcgatg gtccatacct tagatgcgtt agcattaatc aggcaacggc tctctagata gagccctcaa ccggagtttg aagcatggct tctaacttta ctcagttcgt tctcgtcgac aatggcggaa ctggcgacgt gactgtcgcc ccaagcaact tcgctaacgg ggtcgctgaa tggatcagct ctaactcgcg ttcacaggct tacaaagtaa cctgtagcgt tcgtcagagc tctgcgcaga atcgcaaata caccatcaaa gtcgaggtgc ctaaagtggc aacccagact gttggtggtg tagagcttcc tgtagccgca tggcgttcgt acttaaatat ggaactaacc attccaattt tcgctacgaa ttccgactgc gagcttattg ttaaggcaat gcaaggtctc ctaaaagatg gaaacccgat tccctcagca atcgcagcaa actccggcat ctactaatag acgccggcca ttcaaacatg aggattaccc atgtcgaaga caacaaagaa gttcaactct ttatgtattg atcttcctcg cgatctttct ctcgaaattt accaatcaat tgcttctgtc gctactggaa gcggtgatcc gcacagtgac gactttacag caattgctta cttaagggac gaattgctca caaagcatcc gaccttaggt tctggtaatg acgaggcgac ccgtcgtacc ttagctatcg ctaagctacg ggaggcgaat gAtCatcgcg gtcagataaa tagagaaggt ttcttacatg acaaatcctt gtcatgggat ccggatgttt tacaaaccag catccgtagc cttattggca acctcctctc tggctaccga tcgtcgttgt ttgggcaatg cacgttctcc aacggtgcCT ctatggggca caagttgcag gatgcagcgc cttacaagaa gttcgctgaa caagcaaccg ttacccccg cgctctgaga gcggctctat tggtccgaga ccaatgtgcg ccgtggatca gacacgcggt ccgctataac gagtcatatg aatttaggct cgttgtaggg aacggagtgt ttacagttcc gaagaataat aaaatagatc gggctgcctg taaggagcct gatatgaata tgtacctcca gaaaggggtc ggtgcCttTa tcagacgccg gctcaaatcc gttggtatag acctgaatga tcaatcgatc aaccagcgtc tggctcagca gggcagcgta gatggttcgc ttgcgacgat agacttatcg tctgcatccg attccatctc cgatcgcctg gtgtggagtt ttctcccacc Tgagctatat tcatatctcg atcgtatccg ctcacactac ggaatcgtag atggcgagac gatacgatgg gaactatttt ccacaatggg aaatgggttc acatttgagc tagagtccat gatattctgg gcaatagtca aagcgaccca aatccatttt ggtaacgccg gaaccatagg catctacggg gacgatatta tatgtcccag tgagattgca ccccgtgtgc tagaggcact tgcctactac ggttttaaac cgaatcttcg taaaacgttc gtgtccgggc tctttcgcga gagctgcggc gcgcactttt accgtggtgt cgatgtcaaa ccgttttaca tcaagaaacc tgttgacaat ctcttcgccc tgatgctgat attaaatcgg ctacggggtt ggggagttgt cggaggtatg tcagatccac gcctctataa ggtgtgggta cggctctcct cccaggtgcc ttcgatgttc ttcggtggga cggacctcgc tgccgactac tacgtagtca gcccgcctac ggcagtctcg gtatacacca agactccgta cgggcggctg ctcgcggata cccgtacctc gggtttccgt cttgctcgta tcgctcgaga acgcaagttc ttcagcgaaa agcacgacag tggtcgctac atagcgtggt tccatactgg aggtgaaatc accgacagca tgaagtccgc cggcgtgcgc gttatacgca cttcggagtg gctaacgccg gttcccacat tccctcagga gtgtgggcca gcgagctctc ctcggtagct gaccgaggga cccccgtaaa cggggtgggt gtgctcgaaa gagcacgggt gcgaaagcgg tccggctcca ccgaaaggtg ggcgggcttc ggcccaggga cctcccccta aagagaggac ccgggattct cccgatttgg taactagctg cttggctagt taccaccca

Appendix Table 8.11 Insertion (+In) and Deletions (-Dels) identified from NGS Analysis. For example, +TGG would mean a whole codon has been inserted while -TGG would be a deletion of a codon. However, depending on how it fell in the protein sequence the mutation may affect a number of codon/amino acids. Similarly, +C and -C are insertion and deletion of a single C. It is important to note that some of these may be generated as an error in the SAMTOOLS aligner used. Black text shows there is no mutation. Red text suggests there is nucleotide(s) insertion or deletion. The details listed shows nucleotide Insertions (+In) and Deletions (-Dels).

+In -Dels	1999	2007	2008	2008	2008	3450	3461	3462
Nucleotide (%)	A	C	G	G	G	T	T	C
wtMS2 0X Passage	=-TGG(84)	=+GGT(97)	=+GT(1)	=+GTC(2)	=+GTT(5)	=+CC(95)	=-CC(97)	=-CG(92)
wtMS2 3X Passage	=-TGG(87)	=+GGT(99)	G	G	G	=+CC(95)	=-CC(97)	C
MTX MS2 3X Passage	=-TGG(88)	=+GGT(99)	G	G	G	=+CC(95)	=-CC(97)	C

Appendix Table 8.12 FASTA Protein Sequences of MS2. Identified protein mutations are highlighted. Reading N to C terminal.

Protein Sequences
(ORF1) Maturation
>MS2 NCBI_001417.2 PVRAFSTLDRENETFVPSVRVYADGETEDNSFSLKYRSNWTPGRFNSTGAKTKQWHYPSPYSRGAL SVTSIDQGAYKRSGSSWGRPYEEKAGFGFSLDARSCYSLFPVSQNLTYIEVPQNVANRASTEVLQKV TQGNFNLGVALAEARSTASQLATQTIALVKAYTAARRGNWRQALRYLALNEDRKFRSKHVAGRWL ELQFGWLPLMSDIQGAYEMLTKVHLQEFLPMRAVRQVGTNIKLDGRLSYPAANFQTTCNISRRIVI WFYINDARLAWLSSLGILNPLGIVWEKVPFSFVVDWLLPVGNMLEGLTAPVGCSYMSGTVTDVIT GESIISVDAPYGWTVERQGTAKAQISAMHRGVQSVWPTTGAYVKSPFSMVHTLDALALIRQRLSR **>wtMS2_0 passage** PVRAFSTLDRENETFVPSVRVYADGETEDNSFSLKYRSNWTPGRFNSTGAKTKQWHYPSPYSRGAL SVTSIDQGAYKRSGSSWGRPYEEKAGFGFSLDARSCYSLFPVSQNLTYIEVPQNVANRASTEVLQKV TQGNFNLGVALAEARSTASQLATQTIALVKAYTAARRGNWRQTLRYLALNEDRKFRSKHVAGRWL ELQFGWLPLMSDIQGAYEMLTKVHLQEFLPMRAVRQVGTNIKLDGRLSYPAANFQTTCNISRRIVI WFYINDARLAWLSSLGILNPLGIVWEKVPFSFVVDWLLPVGNMLEGLTAPVGCSYMSGTVTDVIT GESIISVDAPYGWTVERQGTAKAQISAMHRGVQSVWPTTGAYVKSPFSIVHTLDALALIRQRLSR **>wtMS2_3 passage** PVRAFSTLDRENETFVPSVRVYADGETEDNSFSLKYRSNWTPGRFNSTGAKTKQWHYPSPYSRGAL SVTSIDQGAYKRSGSSWGRPYEEKAGFGFSLDARSCYSLFPVSQNLTYIEVPQNVANRASTEVLQKV TQGNFNLGVALAEARSTASQLATQTIALVKAYTAARRGNWRQTLRYLALNEDRKFRSKHVAGRWL ELQFGWLPLMSDIQGAYEMLTKVHLQEFLPMRAVRQVGTNIKLDGRLSYPAANFQTTCNISRRIVI WFYINDARLAWLSSLGILNPLGIVWEKVPFSFVVDWLLPVGNMLEGLTAPVGCSYMSGTVTDVIT GESIISVDAPYGWTVERQGTAKAQISAMHRGVQSVWPTTGAYVKSPFSMVHTLDALALIRQRLSR **>MS2_3 passage + MTX** PVRAFSTLDRENETFVPSVRVYADGETEDNSFSLKYRSNWTPGRFNSTGAKTKQWHYPSPYSRGAL SVTSIDQGAYKRSGSSWGRPYEDKAGFGFSLDARSCYSLFPVSQNLTYIEVPQNVANRASTEVLQKV TQGNFNLGVALAEARSTASQLATQTIALVKAYTAARRGNWRQALRYLALNEDRKFRSKHVAGRWL ELQFGWLPLMSDIQGAYEMLTKVHLQEFLPMRAVRQVGTNIKLDGRLSYPAANFQTTCNISRRIVI WFYINDARLAWLSSLGILNPLGIVWEKVPFSFVVDWLLPVGNMLEGLTAPVGCSYMSGTVTDVIT GESIISVDAPYGWTVERQGTAKAQISAMHRGVQSVWPTTGAYVKSPFSMVHTLDALALIRQRLSR

(ORF3) Coat

> MS2 NCBI_001417.2

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRSSAQNRKYTIKVEVPKVATQTVGGVELP
VAAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIY

>wtMS2_0 passage

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKVATQTVGGVEL
PVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY

>wtMS2_3 passage

MASNFTQFVLVDNGGTGDVTVAPSSFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKVATQTVGGVEL
PVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY

>MS2_3 passage + MTX

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKVATQTVGGVEL
PVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY

(ORF4) Replicase

> MS2 NCBI_001417.2

MSKTTKKFNSLCIDLPRDLSLEIYQSIASVATGSGDPHSDDFTAIAYLRDELLTKHPTLGSGNDEATRRTLAIAKLREANGDR
GQINREGFLHDKSLSWDPDVLQTSIRSLIGNLLSGYRSSLFGQCTFSNGAPMGHKLQDAAPYKKFAEQATVTPRALRAAL
LVRDQCAPWIRHAVRYNESYEFRLVVGNGVFTVPKNNKIDRAACKEPDMNMYLQKGVGAFIRRRLKSVGIDLNDQSIN
QRLAQQGSVDGSLATIDLSSASDSISDRLVWSFLPPELYSYLDRIRSHYGIVDGETIRWELFSTMGNGFTFELESMIFWAIV
KATQIHFGNAGTIGIYGDDIICPSEIAPRVLEALAYYGFKPNLRKTFVSGLFRESCGAHFYRGVDVKPFYIKKPVDNLFALML
ILNRLRGWGVVGGMSDPRLYKVWVRLSSQVPSMFFGGTDLAADYYVVSPPTAVSVYTKTPYGRLLADTRTSGFRLARIA
RERKFFSEKHDSGRYIAWFHTGGEITDSMKSAGVRVIRTSEWLTPVPTFPQECGPASSPR

>wtMS2_0 passage

MSKTTKKFNSLCIDLPRDLSLEIYQSIASVATGSGDPHSDDFTAIAYLRDELLTKHPTLGSGNDEATRRTLAIAKLREANDHR
GQINREGFLHDKSLSWDPDVLQTSIRSLIGNLLSGYRSSLFGQCTFSNGASIGHKLQDAAPYKKFAEQATVTPRALRAALL
VRDQCAPWIRHAVRYNESYEFRLVVGNGVFTVPKNNKIDRAACKEPDMNMYLQKGVGAFIRRRLKSVGIDLNDQSINQ
RLAQQGSVDGSLATIDLSSASDSISDRLVWSFLPPELYSYLDRIRSHYGIVDGETIRWELFSTMGNGFTFELESMIFWAIVK
ATQIHFGNAGTIGIYGDDIICPSEIAPRVLEALAYYGFKPNLRKTFVSGLFRESCGAHFYRGVDVKPFYIKKPVDNLFALMLI
LNRLRGWGVVGGMSDPRLYKVWVRLSSQVPSMFFGGTDLAADYYVVSPPTAVSVYTKTPYWRLLADTRTSGFRLARIA
RERKFFSEKHDSGRYIAWFHTGGEITDSMKSAGVRVIRTSEWLTPVPTFPQECGPASSPR

>wtMS2_3 passage

MSKTTKKFNSLCIDLPRDLSLEIYQSIASVATGSGDPHSDDFTAIAYLRDELLTKHPTLGSGNDEATRRTLAIAKLREANDHR
GQINREGFLHDKSLSWDPDVLQTSIRSLIGNLLSGYRSSLFGQCTFSNGASMGHKLQDAAPYKKFAEQATVTPRALRAAL
LVRDQCAPWIRHAVRYNESYEFRLVVGNGVFTVPKNNKIDRAACKEPDMNMYLQKGVGAFIRRRLKSVGIDLNDQSIN
QRLAQQGSVDGSLATIDLSSASDSISDRLVWSFLPPELYSYLDRIRSHYGIVDGETIRWELFSTMGNGFTFELESMIFWAIV
KATQIHFGNAGTIGIYGDDIICPSEIAPRVLEALAYYGFKPNLRKTFVSGLFRESCGAHFYRGVDVKPFYIKKPVDNLFALML
ILNRLRGWGVVGGMSDPRLYKVWVRLSSQVPSMFFGGTDLAADYYVVSPPTAVSVYTKTPYGRLLADTRTSGFRLARIA
RERKFFSEKHDSGRYIAWFHTGGEITDSMKSAGVRVIRTSEWLTPVPTFPQECGPASSPR

>MS2_3 passage + MTX

MSKTTKKFNSLCIDLPRDLSLEIYQSIASVATGSGDPHSDDFTAIAYLRDELLTKHPTLGSGNDEATRRTLAIAKLREANDHR
GQINREGFLHDKSLSWDPDVLQTSIRSLIGNLLSGYRSSLFGQCTFSNGASMGHKLQDAAPYKKFAEQATVTPRALRAAL
LVRDQCAPWIRHAVRYNESYEFRLVVGNGVFTVPKNNKIDRAACKEPDMNMYLQKGVGAFIRRRLKSVGIDLNDQSIN
QRLAQQGSVDGSLATIDLSSASDSISDRLVWSFLPPELYSYLDRIRSHYGIVDGETIRWELFSTMGNGFTFELESMIFWAIV
KATQIHFGNAGTIGIYGDDIICPSEIAPRVLEALAYYGFKPNLRKTFVSGLFRESCGAHFYRGVDVKPFYIKKPVDNLFALML
ILNRLRGWGVVGGMSDPRLYKVWVRLSSQVPSMFFGGTDLAADYYVVSPPTAVSVYTKTPYGRLLADTRTSGFRLARIA
RERKFFSEKHDSGRYIAWFHTGGEITDSMKSAGVRVIRTSEWLTPVPTFPQECGPASSPR

Appendix Table 8.13 FASTA Protein Sequence of Qβ Replicase. Reading N to C terminal.

Qβ Replicase
> QBETA NCBI_39803 MSKTASSRNSLSAQLRRAANTRIEVEGNLALSIANDLLLAYGQSPFNSEAECISFSPRFDGTPDDFRINYLK AEIMSKYDDFSLGIDTEAVAWEKFLAAEAECALTNARLYRPDYSEDFNFSLGESCIHMARRKIAKLIGDVPS VEGMLRHCRFSGGATTTNNRSYGHPSFKFALPQACTPRALKYVLALRASTHFDIRISDISPFNKAVTVPKN SKTDRCIAIEPGWNMFFQLGIGGILRDRLRCWGIDLNDQTINQRRAHEGSVTNNLATVDLSAASDSISLA LCELLLPPGWFEVLMDLRSPKGRLPDGSVVTYEKISSMGNGYTFELESLIFASLARSVCEILDLDSSEVTVYG DDIILPSCAVPALREVFKYVGFTTNTKKTFSEGPFRESCGKHYYSGVDVTPFYIRHRIVSPADLILVLNNLYR WATIDGVWDPRAHSVYLKYRKLLPKQLQRNTIPDGYGDGALVGSVLINPFAKNRGWIRYVPVITDHTRD RERAELGSYLYDLFSRCLSESNDGLPLRGPSGCDSADLFAIDQLICRSNPTKISRSTGKFDIQYIACSSRVLAP YGVFQGTKVASLHEA

Appendix Table 8.14 Percent Identity Matrix Comparing the Replicase protein sequence in Qβ versus MS2. Created by Clustal2.1.

	Qβ NCBI_39803	**MS2 NCBI_001417.2**	**MS2_0 passage**	**MS2_3 passage**	**MS2_3 passage + MTX**
Qβ NCBI_39803	100.00	34.87	34.87	34.87	34.87
MS2 NCBI_001417.2	34.87	100.00	99.08	99.45	99.45
MS2_0 passage	34.87	99.08	100.00	99.63	99.63
MS2_3 passage	34.87	99.45	99.63	100.00	100.00
MS2_3 passage + MTX	34.87	99.45	99.63	100.00	100.00

*Appendix **Table 8.15** Multiple FASTA Replicase Protein Sequence Alignment of Qβ and MS2. Using CLUSTAL O (version 1.2.4).*

```
                          1                                                  60

Qβ NCBI_39803             MSKTASSRNSLSAQLRRAANTRIEVEGNLALSIANDLLLAYGQSPFNSEAECISFSPRFD
MS2 NCBI_001417.2         MSKTTKKFNSLCIDLPR----------DLSLEIYQSIAS-------------VATG-SGD
MS2_0 passage             MSKTTKKFNSLCIDLPR----------DLSLEIYQSIAS-------------VATG-SGD
MS2_3 passage             MSKTTKKFNSLCIDLPR----------DLSLEIYQSIAS-------------VATG-SGD
MS2_3 passage_MTX         MSKTTKKFNSLCIDLPR----------DLSLEIYQSIAS-------------VATG-SGD
                          61                                                 120
Qβ NCBI_39803             GTP-DDFRINYLKAEIMSKYDDFSLGIDTEAV---AWEKFLAAEAECALTNARLYRPDYS
MS2 NCBI_001417.2         PHSDDFTAIAYLRDELLTKHPTLGSGNDEATRRTLAIAKLREANG----DRGQINREGFL
MS2_0 passage             PHSDDFTAIAYLRDELLTKHPTLGSGNDEATRRTLAIAKLREAND----HRGQINREGFL
MS2_3 passage             PHSDDFTAIAYLRDELLTKHPTLGSGNDEATRRTLAIAKLREAND----HRGQINREGFL
MS2_3 passage_MTX         PHSDDFTAIAYLRDELLTKHPTLGSGNDEATRRTLAIAKLREAND----HRGQINREGFL
                          121                                                180
Qβ NCBI_39803             EDFNFSLGESCIHMARRKIAKLIGDVPS--VEGMLRHCRFSGGATTTNNRSYGHPSFKFA
MS2 NCBI_001417.2          HDKSLSWDPDVL---QTSIRSLIGNLLSGYRSSLFGQCTFSNGAPMGHKLQDAAPYKKFA
MS2_0 passage             HDKSLSWDPDVL---QTSIRSLIGNLLSGYRSSLFGQCTFSNGASIGHKLQDAAPYKKFA
MS2_3 passage             HDKSLSWDPDVL---QTSIRSLIGNLLSGYRSSLFGQCTFSNGASMGHKLQDAAPYKKFA
MS2_3 passage_MTX         HDKSLSWDPDVL---QTSIRSLIGNLLSGYRSSLFGQCTFSNGASMGHKLQDAAPYKKFA
                          181                                                240
Qβ NCBI_39803             LPQACTPRALKYVLALRASTHFDI----RIS-----DISPFNKAVTVPKNSKTDRCIAIE
MS2 NCBI_001417.2          EQATVTPRALRAALLVRDQCAPWIRHAVRYNESYEFRLVVGNGVFTVPKNNKIDRAACKE
MS2_0 passage             EQATVTPRALRAALLVRDQCAPWIRHAVRYNESYEFRLVVGNGVFTVPKNNKIDRAACKE
MS2_3 passage             EQATVTPRALRAALLVRDQCAPWIRHAVRYNESYEFRLVVGNGVFTVPKNNKIDRAACKE
MS2_3 passage_MTX         EQATVTPRALRAALLVRDQCAPWIRHAVRYNESYEFRLVVGNGVFTVPKNNKIDRAACKE
                          241                                                300
Qβ NCBI_39803             PGWNMFFQLGIGGILRDRLRCWGIDLNDQTINQRRAHEGSVTNNLATVDLSAASDSISLA
MS2 NCBI_001417.2          PDMNMYLQKGVGAFIRRRLKSVGIDLNDQSINQRLAQQGSVDGSLATIDLSSASDSISDR
MS2_0 passage             PDMNMYLQKGVGAFIRRRLKSVGIDLNDQSINQRLAQQGSVDGSLATIDLSSASDSISDR
MS2_3 passage             PDMNMYLQKGVGAFIRRRLKSVGIDLNDQSINQRLAQQGSVDGSLATIDLSSASDSISDR
MS2_3 passage_MTX         PDMNMYLQKGVGAFIRRRLKSVGIDLNDQSINQRLAQQGSVDGSLATIDLSSASDSISDR


                          301                                                360
```

```
Qβ NCBI_39803          LCELLLPPGWFEVLMDLRSPKGRLPDGSVVTYEKISSMGNGYTFELESLIFASLARSVCE
MS2 NCBI_001417.2       LVWSFLPPELYSYLDRIRSHYG-IVDGETIRWELFSTMGNGFTFELESMIFWAIVKATQ-
MS2_0 passage          LVWSFLPPELYSYLDRIRSHYG-IVDGETIRWELFSTMGNGFTFELESMIFWAIVKATQ-
MS2_3 passage          LVWSFLPPELYSYLDRIRSHYG-IVDGETIRWELFSTMGNGFTFELESMIFWAIVKATQ-
MS2_3 passage_MTX      LVWSFLPPELYSYLDRIRSHYG-IVDGETIRWELFSTMGNGFTFELESMIFWAIVKATQ-
                       361                                                   420
Qβ NCBI_39803          ILDLDSSEVTVYGDDIILPSCAVPALREVFKYVGFTTNTKKTFSEGPFRESCGKHYYSGV
MS2 NCBI_001417.2       IHFGNAGTIGIYGDDIICPSEIAPRVLEALAYYGFKPNLRKTFVSGLFRESCGAHFYRGV
MS2_0 passage          IHFGNAGTIGIYGDDIICPSEIAPRVLEALAYYGFKPNLRKTFVSGLFRESCGAHFYRGV
MS2_3 passage          IHFGNAGTIGIYGDDIICPSEIAPRVLEALAYYGFKPNLRKTFVSGLFRESCGAHFYRGV
MS2_3 passage_MTX      IHFGNAGTIGIYGDDIICPSEIAPRVLEALAYYGFKPNLRKTFVSGLFRESCGAHFYRGV
                       421                                                   480
Qβ NCBI_39803          DVTPFYIRHRIVSPADLILVLNNLYRWATIDGVWDPRAHSVYLKYRKLLPKQLQRNTI-P
MS2 NCBI_001417.2       DVKPFYIKKPVDNLFALMLILNRLRGWGVVGGMSDPRLYKVWVRLSSQVPSMFFGGTDLA
MS2_0 passage          DVKPFYIKKPVDNLFALMLILNRLRGWGVVGGMSDPRLYKVWVRLSSQVPSMFFGGTDLA
MS2_3 passage          DVKPFYIKKPVDNLFALMLILNRLRGWGVVGGMSDPRLYKVWVRLSSQVPSMFFGGTDLA
MS2_3 passage_MTX      DVKPFYIKKPVDNLFALMLILNRLRGWGVVGGMSDPRLYKVWVRLSSQVPSMFFGGTDLA
                       481                                                   540
Qβ NCBI_39803          DGYG-DGA--LVGSVLINPFAKN------RGWIRYVPVITDHTRDRERAELGSYLYDLFS
MS2 NCBI_001417.2       ADYYVVSPPTAVSVYTKTPYGRLLADTRTSGF-RLARIARERKFFSEKHDSGRYIAWFH-
MS2_0 passage          ADYYVVSPPTAVSVYTKTPYWRLLADTRTSGF-RLARIARERKFFSEKHDSGRYIAWFH-
MS2_3 passage          ADYYVVSPPTAVSVYTKTPYGRLLADTRTSGF-RLARIARERKFFSEKHDSGRYIAWFH-
MS2_3 passage_MTX      ADYYVVSPPTAVSVYTKTPYGRLLADTRTSGF-RLARIARERKFFSEKHDSGRYIAWFH-
                       541                                                   600              615
Qβ NCBI_39803          RCLSESNDGLPLRGPSGCDSADLF-AIDQLICRSNPTKISRSTGKFDIQYIACSSRVLAPYGVFQGTKVASLHEA
MS2 NCBI_001417.2       -TGGEITDSMKSAGVRVIRTSEWLTPVPTFPQECGPASSPR----------------------------------
MS2_0 passage          -TGGEITDSMKSAGVRVIRTSEWLTPVPTFPQECGPASSPR----------------------------------
MS2_3 passage          -TGGEITDSMKSAGVRVIRTSEWLTPVPTFPQECGPASSPR----------------------------------
MS2_3 passage_MTX      -TGGEITDSMKSAGVRVIRTSEWLTPVPTFPQECGPASSPR----------------------------------
```

Appendix Table 8.16.

A) Analysing 5'wt MS peaks from Figure 5.17. Integrated peak heights were measured using MassLynx™ Software, Waters.

Peak Label	**Mass (Da)**	**RNA:MTX % major ion**	
		1:0.1	1:1
RNA	10,231	24.0	0
+ 1 MTX	10,675	46.8	2.2
+ 2 MTX	11,119	23.5	12.3
+ 3 MTX	11,563	5.8	33.3
+ 4 MTX	12,008	0	33.5
+ 5 MTX	12,453	0	15.0
+ 6 MTX	12,898	0	3.8

B) Analysing 5'mutant MS peaks from Figure 5.18. Integrated peak heights were measured using MassLynx™ Software, Waters.

Peak Label	**Mass (Da)**	**RNA:MTX % major ion**	
		1:0.1	1:1
RNA	10,191	49.5	0
+ 1 MTX	10,636	30.5	4.1
+ 2 MTX	11,081	12.5	11.6
+ 3 MTX	11,525	0	27.7
+ 4 MTX	11,969	0	31.8
+ 5 MTX	12,414	0	20.2
+ 6 MTX	12,858	0	4.5

Appendix Figure 8.17. 54 Identified PSs within the MS2 genome by CLIP-Seq (Rolfsson et al. 2015). Eight have an adenosine-bulged base in their predicted secondary structure (red circles). Nucleotide positions in the genomic sequence are indicated in red text. Stem-loops are shown labelled from P1 to P54, representing 5' to 3'.