A Study of the Membrane Binding Properties of the Matrix Protein from Human Respiratory Syncytial Virus

HELEN KENNEDY MCPHEE

Durham University

A study of the membrane binding properties of the matrix protein from human respiratory syncytial virus

Helen Kennedy McPhee

Submitted in partial fulfilment of the requirements of the University of Durham for the degree of Doctor of Philosophy

Centre for Bioactive Chemistry
Department of Chemistry

September 2009

Abstract

This thesis describes the structure of the matrix (M) protein from human respiratory syncytial virus (RSV), and its interactions with model membranes composed of a range of lipids, which have been characterized using a variety of techniques.

The M protein was expressed in *E. coli* with a histidine tag and purified by immobilised metal affinity chromatography (IMAP). The protein was found to have acquired a mutation at position 254 from methionine to arginine, but this did not appear to affect its behaviour and both the native and mutated form of the protein were used in several experiments.

Efforts to produce 2D crystals of the M protein by incubation with lipid and detergent were unsuccessful, but in several detergent-free incubations, particularly where the lipid DPPE was included, extended helical structures were formed that bore a striking resemblance to the structure of filamentous RSV virions with the viral envelope removed. These helical assemblies appeared to grow from liposomes, and demonstrate the inherent ability of RSV M to self assemble into arrays of this nature.

Samples of purified protein were provided to Dr Victoria Money who was able to crystallize M and solve its structure to a resolution of 1.6 Å, revealing two domains linked by a region with little secondary structure that is potentially flexible, and a large positively charged area on the surface extending across both domains. The overall structure is somewhat similar to the crystal structure of the Ebola virus matrix protein VP40. Circular dichroism (CD) studies of M showed that it contains significantly less secondary structure in solution than in the crystal structure. Cross-linking experiments suggested that RSV M preferentially forms dimers, tetramers and hexamers in solution, in common with several other viral matrix proteins.

Experiments performed on a Langmuir trough revealed that the M protein partitions into phospholipid monolayers, regardless of whether the lipid head group is phosphocholine or phosphoethanolamine. It interacts differently with monolayers containing sphingomyelin, which in combination with cholesterol forms phase

separated liquid-ordered (L_o) microdomains. RSV M was found to bind extrinsically to the sphingomyelin-containing L_o domains. This did not occur when the protein was in contact with monolayers composed of only phospholipids and cholesterol that are also known to form L_o domains, showing that RSV M has a specific interaction with sphingomyelin. The evidence for this hypothesis obtained from Langmuir pressure-area isotherms was supported by direct visualization of the monolayer at the surface using Brewster angle microscopy (BAM), and examination of monolayers deposited on a modified silicon surface by atomic force microscopy (AFM).

The findings described herein suggest that specific interactions with sphingomyelin may provide a mechanism for localization of the matrix protein, and therefore targeting of the site of viral assembly, to particular regions of the host cell membrane. The structure of the protein, with its two domains linked by a potentially flexible region, would allow for changes in protein conformation whilst bound to the membrane: it is possible that such changes could play a key role in the function of RSV M.

Table of contents

List of tables

List of figures

List of abbreviations

AFM	Atomic force microscopy
BDV	Borna disease virus
CD	Circular dichroism (spectroscopy)
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulphonate
Chol	Cholesterol
Cryo EM	Electron cryo-microscopy
DAPC	1,2-diacetyl-*sn*-glycero-3-phosphocholine
DBPC	1,2-dibutyryl-*sn*-glycero-3-phosphocholine
DMPC	1,2-dimyristoyl-*sn*-glycero-3-phosphocholine
DNA	Deoxyribonucleic acid
DOGS-NTA-Ni	1,2-dioleoyl-*sn*-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] (nickel salt)
DOPC	1,2-dioleoyl-*sn*-glycero-3-phosphocholine
DOPS	1,2-dioleoyl-*sn*-glycero-3-phosphoserine
DPPA	1,2-dipalmitoyl-*sn*-glycero-3-phosphate
DPPC	1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine
DPPE	1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine
DRM	Detergent-resistant membrane
EM	Electron microscopy
ESI	Electrospray ionisation
F	Viral fusion protein
FRET	Fluorescence resonance energy transfer
G	Viral attachment glycoprotein
GalCer	Galactosyl ceramide, specifically D-galactosyl-β1-1′-N-nervonyl-D-erythro-sphingosine (C24:1 β-D-galactosyl ceramide)
HA	Viral haemagglutinating glycoprotein
HIV	Human immunodeficiency virus
hMPV	Human metapneumovirus
HN	Viral haemagglutinin-neuraminidase glycoprotein
ICAM	Intercellular adhesion molecule
IMAC	Immobilized metal affinity chromatography
L	L protein; viral RNA polymerase
M	Matrix protein from RSV (VSV and BDV matrix proteins are also termed M)
M1	Matrix protein from influenza virus
M^{254M}	Native form of RSV matrix protein
M^{254R}	RSV matrix protein expressed with a spontaneous mutation from methionine to arginine at the 254 position
MALDI-TOF MS	Matrix assisted laser desorption ionisation - time of flight mass spectrometry
MHis	RSV matrix protein with histidine tag
mRNA	Messenger RNA
MS	Mass spectrometry
MW	Molecular weight

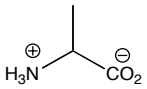
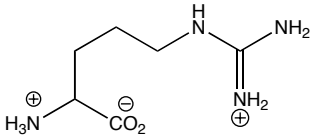
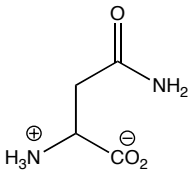
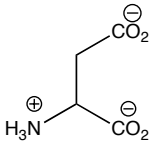
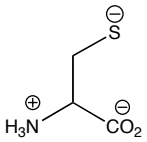
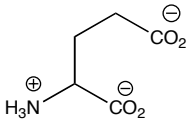
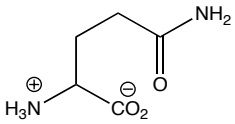
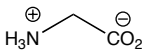
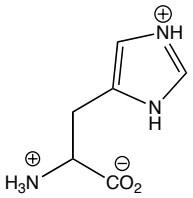
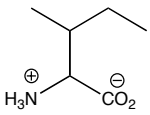
N	Viral nucleocapsid protein
NA	Viral neuraminidase glycoprotein
NHS	National Health Service
OTS	Octadecyl trichlorosilane
P	Viral phospho-protein
PA	Phosphatidic acid
PBS	Phosphate buffered saline
PC	Phosphocholine
PE	Phosphoethanolamine
PS	Phosphoserine
RMS	Root mean square
RNA	Ribonucleic acid
(-)RNA	Negative-sense RNA
RNP	Ribonucleoprotein (RNA complexed with protein/s)
RSV	Human respiratory syncytial virus
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SH	Small hydrophobic protein (in RSV)
SM	Sphingomyelin
SVD	Singular value decomposition
TEM	Transmission electron microscopy
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
US	United States (of America)
VLP	Virus-like particle
VP40	Matrix protein from Ebola or Marburg viruses
VSV	Vesicular stomatitis virus

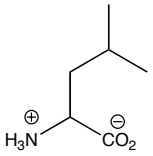
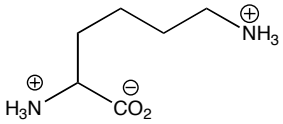
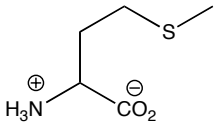
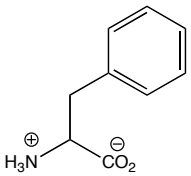
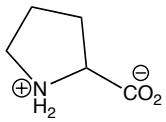
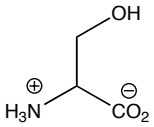
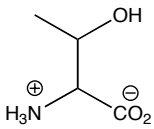
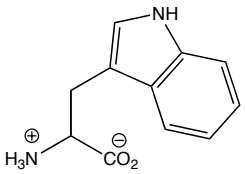
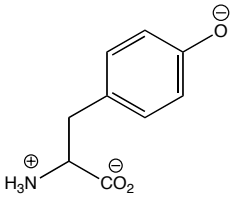
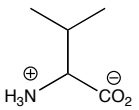
Units

º	Degrees of arc
ºC	Degrees Celcius
Da	Daltons (units of atomic mass)
kDa	Kilo-Daltons (10^3 Da)
cm	Centimetres (10^{-2} metres)
mm	Millimetres (10^{-3} metres)
µm	Micrometres (10^{-6} metres)
nm	Nanometres (10^{-9} metres)
Å	Ångströms (10^{-10} metres = 0.1 nm)
L	Litres
mL	Millilitres (10^{-3} litres)
µL	Microlitres (10^{-6} litres)
M	Molarity in moles per litre (= mol/dm^3)
mM	Molarity in millimoles (10^{-3} moles) per litre
µM	Molarity in micromoles (10^{-6} moles) per litre
s	Seconds
mNm^{-1} (mN/m)	Milli-Newtons per metre (units of surface tension)
dyn/cm	Dynes per centimetre (units of surface tension equal to mN/m)
Hz	Hertz (units of frequency)

kHz Kilohertz (10^3 Hertz)

The amino acids

Name	3-letter code	1-letter code	Structure
Alanine	Ala	A	
Arginine	Arg	R	
Asparagine	Asn	N	
Aspartic acid	Asp	D	
Cysteine	Cys	C	
Glutamic acid	Glu	E	
Glutamine	Gln	Q	
Glycine	Gly	G	
Histidine	His	H	
Isoleucine	Ile	I	

Leucine	Leu	L	
Lysine	Lys	K	
Methionine	Met	M	
Phenylalanine	Phe	F	
Proline	Pro	P	
Serine	Ser	S	
Threonine	Thr	T	
Tryptophan	Trp	W	
Tyrosine	Tyr	Y	
Valine	Val	V	

Acknowledgements

There are a great many people without whom this work would have been impossible.

First of all I must express my sincere thanks to Dr. John Sanderson; I couldn't have asked for a better supervisor. His support, enthusiasm and positive attitude have been an inspiration throughout this project. My co-supervisor Dr. Paul Yeo has also been brilliant, introducing me to the world of virology, and happy to let me loose in his lab any time.

Several people have provided invaluable practical assistance during the course of my experiments. I owe particular thanks to Dr. Jackie Mosely for expertly running all my mass spectroscopy samples, to Christine Richardson of the School of Biological and Biomedical Sciences for assisting me with my TEM work (even though I seemed to spell death to all her microscope filaments), to Dr. Scott Watson for going beyond the call of duty in helping me with atomic force microscopy, and to Dr. Andrew Beeby for his assistance (and patience) in setting up the Brewster angle microscope. I am indebted to Dr. Sharon Cooper for lending me her Langmuir trough, and never requesting its return, and to Prof. David Parker for allowing me to use his CD spectrometer. I would also like to say a big thank you to Prof. Rob Edwards and his group, especially Dr. Melissa Brazier-Hicks, Dr. Ian Cummins and Dr. David Dixon, for giving me (almost) free rein of their lab and equipment, and for their patient answering of many questions.

Dr. Victoria Money deserves a particularly special mention; her work in solving the crystal structure of the RSV matrix protein has been very important to this project. I fear I will never understand all the dark complexities of protein crystallography, but I am most grateful for her contribution, and for her continued interest in my work.

I am grateful for the support and friendship of all the staff and students, past and present, of the Sanderson group and lab CG216, in particular Dr. Liz Grayson, Dr. Georg Blaser, Helen Murphy, Natalie Cossar, Jennifer Tovey, Catherine Pridmore and Jenny Carlisle. I would also like to thank Dr. David Hodgson, and current and

previous members of his group, especially Dr. Mark Skipsey, Dr. Helen Watson and Milena Trmcic, for their encouragement and constructive criticism in research group meetings. All the staff and students in chemistry have made the Department a brilliant place to work, which I will miss very much.

Last (but not least) I could never have stuck it out to the end without all the support and encouragement from Andy, and my parents, who have both been here before me.

1 Introduction

1.1 Viruses

"...a group of minute infectious agents (20-300 nm long and/or wide), unable to multiply except inside a living cell of a host, of which they are obligate parasites and outside of which they are inert."[1]

Viruses are essentially collections of genetic information directed only towards the purpose of their own replication, the prototypical example of "selfish genes".[2] Each virus particle, or virion, consists of either a DNA or RNA genome contained within a protein coat called a capsid or nucleocapsid, which in some cases is also surrounded by a membrane composed of lipids, proteins and glycoproteins, referred to as an envelope. Viral DNA can be single-stranded or double-stranded, linear or circular, while RNA can be single stranded positive sense (+), like messenger RNA, negative sense (−), which requires a complementary copy to be produced before it can be replicated, double stranded, or ambisense (made up of + and – regions of RNA attached end to end). RNA genomes may also be segmented, with each piece encoding an individual gene.[3] Viruses are unable to make their own energy or proteins, and cannot replicate their genetic material except within a host cell. Additionally (and importantly for this project), viral reproduction occurs by self-assembly of viral components, rather than by the binary fission process seen in cells.[3]

Viruses with naked capsids are very resilient and able to withstand harsh environments during transmission between hosts. Enveloped viruses are more fragile because the integrity of the viral membrane can only be maintained in aqueous solutions. Such viruses are commonly transmitted in fluids, respiratory droplets, blood and tissue, and generally cannot withstand the acid conditions of the gastrointestinal tract.[3] A typical replication cycle for an enveloped RNA virus is shown in Figure 1.1.

To enter a suitable host cell, viruses must traverse the plasma membrane, a complex structure made up of a lipid bilayer containing intrinsic and extrinsic membrane proteins. The virus particle first has to bind to a specific protein or carbohydrate receptor on the cell surface (although of course such receptors are in fact present in the host cell for purposes other than the facilitation of viral attachment or entry).[2] Receptors are generally either molecules that are abundant on cell surfaces, or are only found on particular types of cells in which the virus can replicate. A wide range of different cell molecules are utilized as specific receptors for individual virus species. The availability of a particular receptor on a cell affects the ability of the virus to infect each type of cell. For example, poliovirus targets a type of intercellular adhesion molecule (ICAM) called CD155, which is found on many primate cells. However, poliovirus can only attack neurons and specific cells in the small intestine, since different membrane proteins seem to mask CD155 on cells from other types of tissue.[2] Interestingly, no receptors have been found that are involved in viral infection of plant cells, where instead physical damage to the cell seems to allow virus entry.[2, 4]

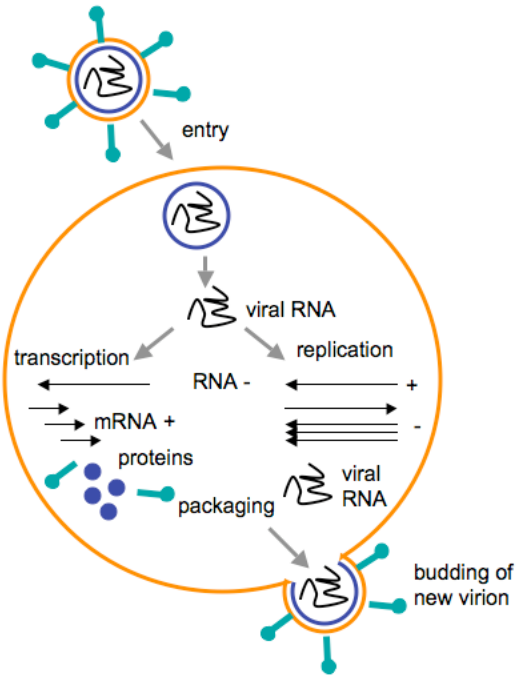


Figure 1.1: A typical replication cycle for an enveloped RNA virus

After binding, the viral genome is introduced into the cytoplasm of the host cell. In the case of enveloped viruses, this involves fusion of the viral lipid envelope with the either the plasma membrane or an endosomal membrane, a process promoted by

conformational change of a fusion domain or protein on the virus surface. In plasma membrane fusion, binding to receptors causes conformational change, while fusion with the endosomal membrane is triggered by the lower pH in this region of the cell.[2-4] For non-enveloped viruses the mechanism of entry is less well understood. The virus or subviral components somehow reach the far side of the plasma membrane, or the interior of an endosomal vesicle, but the process by which this is accomplished is not clear. One possibility is that interaction with a surface receptor causes the proteins of the viral capsid to rearrange themselves in a way that leads to penetration of the membrane.[4]

Once the membrane has been traversed, the viral genome is replicated. In many cases this happens after the genetic material is released from its protein coat, but in others the genome is retained within a nucleocapsid. Some DNA viruses rely on the existing cellular machinery in the host cell nucleus to replicate their DNA and transcribe their mRNA, while others replicate in the cytosol and usually encode their own DNA and RNA polymerases. When an RNA virus invades a cell, there are no existing enzymes present to allow replication of its genome, so these viruses must provide their own RNA polymerase.[3, 4] The virus may encode proteins that modify the transcription or translation apparatus of the host so that viral proteins are produced instead of those of the host cell; in other cases the sheer quantity of viral mRNA present is so large that it occupies most of the ribosomes, preventing efficient translation of cellular RNA.[3]

When all the viral structural proteins have been synthesized using the host cell translation machinery, the components self-assemble into progeny virus particles. This process occurs in a variety of ways depending on the structure of the resulting viral capsid. Enveloped viruses bud (Figure 1.3 A, p 9) from cellular membranes, either from the plasma membrane at the surface or from internal membranes such as those of the nucleus or the Golgi apparatus. They take with them part of the host cell membrane, which then becomes the lipid envelope of the new virion.[2-4] Viral budding is a complex process involving viral and cellular proteins that in many cases is not fully understood;[5] what is known about the mechanism as it occurs in human respiratory syncytial virus (RSV) and related viruses will be discussed later. The multiplication and release of large numbers of new virions often results in lysis and

death of the host cell; however, an enveloped virus can leave the cell by budding without destroying the plasma membrane or killing the host. Each infected cell can produce as many as 100,000 virions, but many of these may be defective, for a variety of reasons, and thus non-infectious.[3]

The major motivation for studying viruses is that so many cause serious disease in humans or animals, or damage important crop plants. More than 40 families of viruses have been implicated in human disease, with more being discovered each year.[3]

1.2 The negative-sense RNA viruses

Family	Genus	Genome size	Typical virus	Host(s)	Transmission
Mononegavirales (non-segemented)					
Rhabdoviridae	Vesiculovirus	13-16 kb	**VSV**	Vertebrates	Multiple
	Lyssavirus		Rabies	Vertebrates	Saliva
	Ephemerovirus		BEFV	Cattle	Arthropod-borne
	Novirhabdovirus		IHNV	Fish	-
	Cytorhabdovirus		MSSV	Plant	Arthropod-borne
	Nucleorhabdovirus		CCMoV	Plant	Arthropod-borne
Filoviridae	Filovirus	13 kb	Marburg	Primate	Contact with body fluids
			Ebola	Primate	Contact with body fluids
Paramyxoviridae	Respirovirus	16-20 kb	Sendai	Vertebrates	Respiratory
	Morbillivirus		Measles	Vertebrates	Respiratory
	Rubulavirus		Mumps	Vertebrates	Respiratory
	Megamyxovirus		Hendra	Vertebrates	Respiratory
	Pneumovirus		**RSV**	Vertebrates	Respiratory
	Metapneumovirus		AMPV	Vertebrates	Airborne
Bornaviridae	Bornavirus	~ 9	BDV	Vertebrates	
Segmented negative-strand RNA viruses					
Orthomyxoviridae	Influenza	13 kb in 7 or 8 segments	**Influenza A**	Vertebrates	Airborne
			Influenza B	Vertebrates	Airborne
			Influenza C	Vertebrates	Airborne
	Thogotovirus		Thogoto	Vertebrates	Arthropod-borne
Bunyaviridae	Bunyavirus	11-20 kb in 3 segments	Bunyamwera	Vertebrates	Arthropod-borne
	Hantavirus		Hantaan	Vertebrates	Faeces/urine/saliva
	Nairovirus		Dugbe	Vertebrates	Arthropod-borne
	Phlebovirus		Rift Valley Fever	Vertebrates	Arthropod-borne
	Tospovirus		TSWV	Plants	Thrips
Arenaviridae	Arenavirus	10-14 kb in 2 segments	Lassa	Vertebrates	Urine/saliva

Table 1.1: The negative-sense RNA virus families.[4]
In all cases vertebrates include humans. Abbreviations: VSIV, vesicular stomatitis Indiana virus, BEFV, bovine ephemeral fever virus, IHNV, infectious hematopoietic necrosis virus, VSV, vesicular stomatitis virus, RSV, respiratory syncytial virus, TRTV, turkey rhinotracheitis virus, BDV, Borna disease virus, TSWV, tomato spotted wilt virus, LCMV, lyphocytic choriomenigitis virus.

A significant number of single-stranded RNA viruses have a genome that is negative sense, that is, the RNA is complementary to mRNA. The (-)RNA viruses can be divided (see Table 1.1) into those with a single genome, the *Mononegavirales*, which have 6-11 cistrons that encode the individual proteins necessary for replication, and a second group whose RNA is segmented into 2-8 sections that each represent one cistron.[2] The two groups and the virus families they contain are shown in Table 1.1, along with examples of the most important (-)RNA viruses.

Many of these viruses are important human pathogens, causing diseases such as influenza (*Orthomyxoviridae*), mumps and measles (*Paramyxoviridae*), rabies (*Rhabdoviridae*), encephalitis (several members of the *Bunyaviridae*), upper and lower respiratory tract diseases (many viruses in the *Paramyxoviridae* family) and haemorrhagic fevers (numerous viruses of the *Bunyaviridae*, *Arenaviridae* and *Filoviridae* families), as well as others. Many (-)RNA viruses infect the entire human population at one time or another, for example RSV and influenza. Most of these very widespread viruses cause diseases that, although serious, are not often fatal. Other (-)RNA viruses such as rabies and Ebola viruses result in high mortality rates, but fortunately this means they are usually not able to infect large numbers of people.[4]

The (-)RNA viruses are morphologically diverse; the *Rhabdoviridae* are bullet-shaped, *Filoviridae* are filamentous, while the *Paramyxoviridae* are pleiomorphic; but they all share the same basic structure. All the (-)RNA viruses have a viral envelope and associated glycoproteins, a nucleocapsid protein, and the majority have a matrix protein; a typical virus particle (Figure 1.2, p 9) can be pictured as a lipid vesicle reinforced with matrix protein, possessing glycoprotein spikes and carrying a cargo of encapsidated viral genome associated with a viral polymerase complex.[6] The matrix protein, where it is present, plays a vital role in the assembly of new virus particles, and in defining the structure of the virion. The matrix protein from human respiratory syncytial virus (RSV), a member of the *Mononegavirales* family *Paramyxoviridae*, subfamily *Pneumovirinae*, and genus Pneumovirus, is the focus of the present study.

1.3 Respiratory Syncytial Virus

1.3.1 Clinical significance

RSV is the most common cause of fatal acute respiratory tract infection in infants and young children,[3] who are generally infected in winter between the ages of 6 weeks and 9 months. In the US, it accounts for half the hospital admissions of infants under 2 years old during January and February,[4] and in the UK around 20% of all admissions for infections of the lower respiratory tract in children are due to RSV.[7] The recently discovered (2001) human Metapneumovirus (hMPV) is another important cause of bronchiolitis, producing a very similar illness generally affecting slightly older children and the elderly; coinfection with both viruses may result in more severe disease.[8] hMPV is also a member of the *Pneumovirinae* and is very similar to RSV, with an analogous genome, further emphasising the need to study this family of viruses.[9]

The disease begins as an upper respiratory tract infection and progresses to the lower respiratory tract in 25-40% of cases.[4] Symptoms of wheezing and breathing difficulties stem from viral invasion of the respiratory epithelium, where the resulting immune response causes cell damage. Necrosis of the bronchioles can result in the formation of plugs composed of mucus, fibrin and dead tissue, which easily obstruct the narrow airways of young infants.[3, 10] Pneumonia, where the bronchioles become filled with fluid due to damage of the epithelia and leakage from plasma, can also occur as a result of RSV infection and/or secondary pulmonary bacterial coinfection.[11] RSV infection can often have long-term effects; up to 70 % of infants will be left with some type of respiratory morbidity for up to 10 years following RSV bronchiolitis,[9] which has been implicated by some studies in the induction and exacerbation of asthma.[9, 10]

In children that are otherwise healthy, RSV symptoms are treated by the administration of oxygen, intravenous fluids and nebulized cold steam.[3] Children and adults who are at risk of a more severe disease progression, such as premature infants, the immunocompromised, those undergoing chemotherapy or with

cardiopulmonary disease, are given the broad-spectrum antiviral ribavirin. This is the only drug currently available to treat RSV; it has been licensed for use in RSV respiratory disease in children since 1986, and for mechanically ventilated patients since 1993, and early treatment appears to reduce the severity of illness and help to prevent long-term complications.[9, 12-14] However, there are concerns about the use of ribavirin due to its high cost, the difficulty of administration (as an aerosol *via* a mist tent, mask, oxygen hood or ventilator for 12-20 hours a day), conflicting reports of its efficacy,[9] and possible problems with occupational exposure of medical staff, as ribavirin carries a risk of teratogenic effects and so is contra-indicated in pregnancy.[13, 14 15] It is also known to affect bone marrow function in patients, causing anaemia for the duration of treatment with the drug.[16, 17] Because of these issues, the treatment of RSV bronchiolitis, "remains a good example of therapeutic nihilism…"[7]

Most people will have been infected by RSV at least once by the time they are four years old; infection does not confer complete immunity and reinfections occur frequently throughout life, normally producing a less serious disease.[3, 4] However, the virus is increasingly being recognized as a significant problem in the elderly, especially in institutional settings such as care homes and hospitals, and is associated with an increased risk of hospitalisation and mortality, resulting in a disease burden similar to that of influenza.[18, 19] These adults, and other patients who are particularly at risk from serious RSV bronchiolitis, would benefit from an effective vaccine against RSV; unfortunately no such vaccine is currently available.

A formalin-inactivated whole virus vaccine (the same concept as the successful Salk polio vaccine) was trialled in the 1960s, but was found not to protect against RSV infection. In fact it increased the severity of illness upon challenge with RSV, especially in younger children; up to 80 % of vaccinated children who were subsequently infected with the virus were hospitalized, and two infants died because of lung damage.[20-22] This disastrous outcome has hampered subsequent vaccine development, along with the lack of an animal model that satisfactorily mimics the human disease, and the need to immunize young infants whose immune systems are not fully developed.[23]

The one preventative treatment available at present to protect vulnerable patients from RSV infection is palivizumab, a humanized monoclonal antibody to the viral fusion (F) protein.[24] This provides only short-term passive immunity, requiring several doses over the RSV season, but has been shown to decrease hospital admissions due to RSV in high-risk infants. In the US palivizumab is recommended for premature babies born at less than 32-35 weeks gestation, and for infants under 2 years with chronic lung disease. Around 100,000 children a year receive palivizumab in the US. Current NHS guidelines note the high cost of the drug (around £2500 for the five doses necessary), and UK studies have suggested that it only represents a good cost-benefit ratio when given to infants born prematurely who are receiving oxygen at home. Its benefits for infants with cardiac disease have also been studied, but in reality palivizumab is too expensive to have any real impact on hospital admissions caused by RSV, except in the very richest countries and for very small groups of patients.[7]

1.3.2 Structure and replication

RSV is the prototype member of the *Pneumovirinae*, a subfamily of the *Paramyxoviridae* in the order *Mononegavirales*. Its negative sense, non-segmented RNA genome is tightly bound to the N protein to form the helical nucleocapsid, which associates with the components of the viral RNA-dependent RNA polymerase (L, P, M2-1 and M2-2 proteins) to form the holo-nucleocapsid (or RNP complex). Together, P and L transcribe the viral RNA template; M2-1 is responsible for mRNA elongation and anti-termination during transcription of the viral genome, while the function of M2-2 is less well defined, although it may be involved with regulating RNA replication.[25] The virion is surrounded by a lipid bilayer envelope, which contains the fusion (F), attachment (G) and small hydrophobic (SH) proteins, necessary for the virus to attach to and enter a host cell. The G protein binds to receptors on the cell surface, then F initiates pH-independent fusion of the viral lipid envelope with the host plasma membrane.[26] Between the viral membrane and the holo-nucleocapsid is the matrix (M) protein. M is a peripheral membrane protein that is not thought to penetrate into the lipid bilayer, but forms a layer tightly associated with the underside of the envelope.[27] M is essential for assembly and

budding of progeny virions, and as the focus of this work will be discussed in greater detail later. While the general functions carried out by these proteins are known, the details of the processes that occur during RSV replication and assembly are not yet fully understood.[28] A schematic representation of an RSV virion is shown in Figure 1.2.

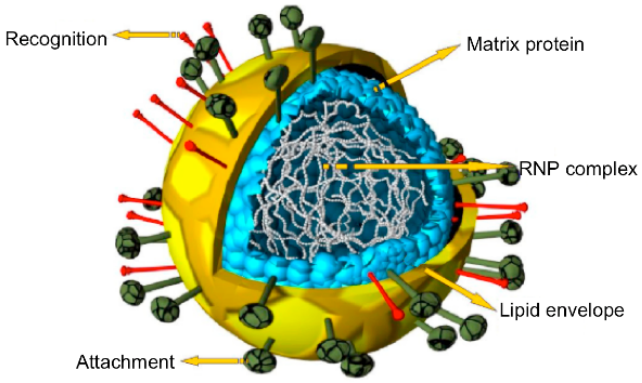


Figure 1.2: Schematic representation of an RSV virion, with the main features labelled.[§] The structure labelled 'RNP complex' is the holo-nucleocapsid described above. The membrane glycoproteins G, F and SH are not shown in detail but are represented by the structures labelled 'recognition' and 'attachment'.

As with other members of the *Paramyxoviridae*, viral replication occurs in the cytoplasm.[4, 26] RSV virions then assemble at, and bud from, the cell membrane, during which process they acquire their lipid envelope. Virions are often spherical and around 150-300 nm in diameter, in common with the rest of the *Paramyxoviridae*, but the virus can also form filamentous particles that are 60-100 nm in diameter and up to 10 µm in length.[26] Figure 1.3 shows a spherical RSV virion budding from a cell, and part of a filamentous virion.

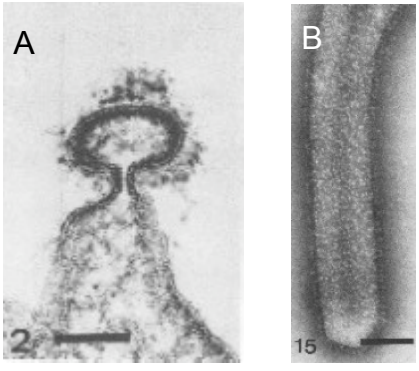


Figure 1.3: Spherical and filamentous RSV virions.
A: A spherical virion budding from the surface of a Vero cell 48h after infection. The matrix protein can be seen as a dark layer on the inner face of the membrane. B: An RSV virus filament produced from a Vero cell. Scale bars represent 100 nm.[29†]

[§] Drawn by Andrew Easton of Warwick University and used with his permission

Both spherical and filamentous particles without a complete internal structure have been observed in RSV cultures,[29] suggesting that a relatively high proportion of defective virions are produced during replication. Filamentous particles appear to have a striated appearance when the lipid envelope is removed by freeze-fracturing the sample, as shown in Figure 1.4. These striations, with a pitch of 54 nm, translate to a helical structure with an angle of 60º around the filament,[29] which may be inferred to represent the arrangement of the matrix protein in the virion.

Figure 1.4: A filamentous RSV virion with the membrane removed by freeze-fracture revealing the striations of the helical structure underneath. Scale bar is 100 nm.[29†]

1.4 Matrix proteins

1.4.1 Function

The matrix protein is a vital component of enveloped viruses, driving the assembly and packaging of viral components and budding of the new virion at the host cell membrane, as well as determining the morphology of the resulting virus particle.[26, 30-36] Many viruses are unable to produce progeny if the matrix protein is defective or not expressed. One example of the effects of this can be seen in subacute sclerosing panencephalitis (SSPE), a persistent measles infection of the brain causing a progressive fatal disease. The majority of the measles viruses replicating in the brain tissue do not express M protein effectively, so no infectious virions can be produced; however the ongoing infection, which cannot be resolved, results in destruction of brain cells leading to death.[26]

† Reproduced with permission from the American Society for Microbiology

Often over-expression of matrix protein alone is sufficient to induce the formation of virus-like particles (VLPs). Expression of the Ebola virus matrix protein, VP40, has been shown to result in the release of VLPs from mammalian cells.[37, 38] Expression of vesicular stomatatis virus (VSV) and Sendai virus matrix proteins had similar results.[39-41] Expression of various proteins from Nipah virus, another paramyxovirus, also caused the release of such particles. This study found that the Nipah virus M protein exhibited intrinsic budding ability and facilitated the inclusion of other viral proteins into the VLPs.[42] Figure 1.5 shows some of the particles produced by expressing matrix proteins in cells.

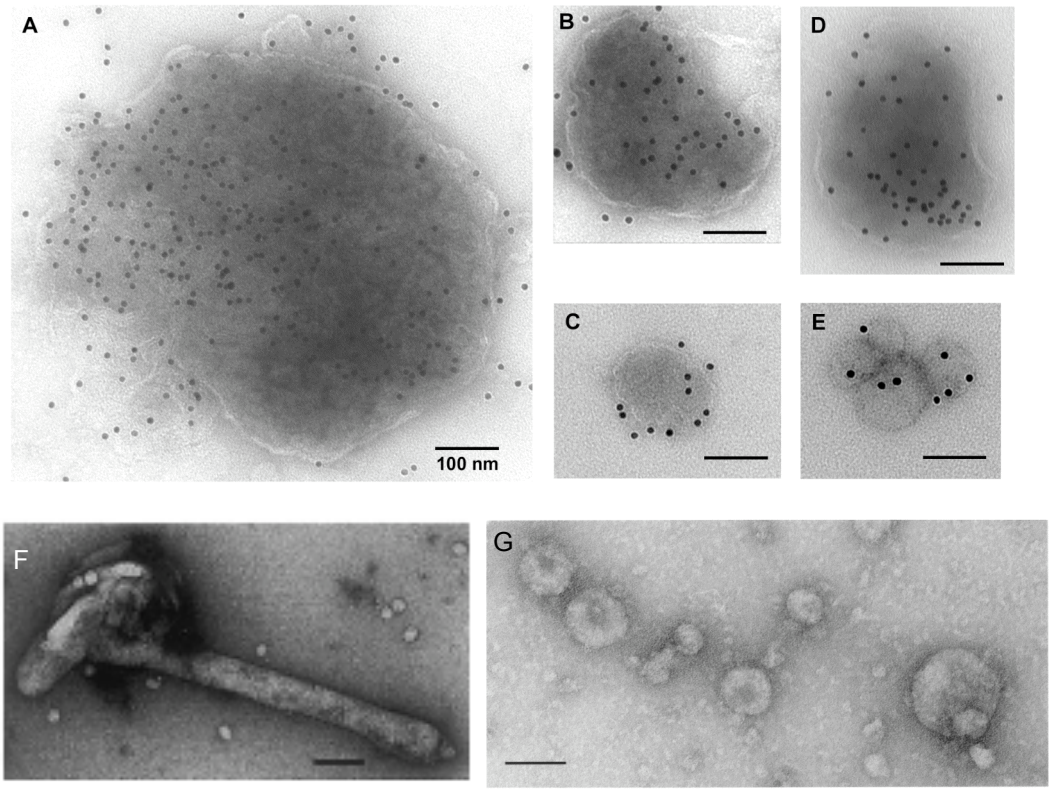


Figure 1.5: Electron micrographs showing virus-like particles produced from cells expressing viral matrix proteins.
A-C: VLPs from 293T cells expressing Nipah virus M with immuno-gold labelling using Mab F45G5; D-E: VLPs from 293T cells expressing N, M, F and G with immuno-gold labelling against G using mouse anti-HeV G polyclonal serum;[42] F: Filovirus-like particle released from a 293T cell upon expression of Ebola virus VP40;[37†] G: VLPs produced from 293T cells expressing the M protein from Sendai virus.[41§] Scale bars A-F: 100 nm, G: 82 nm.

[†] Reprinted from reference 37 J. Timmins, S. Scianimanico, G. Schoehn and W. Weissenhorn, *Virology*, 2001, **283**, 1-6. with permission from Academic Press.
[§] Reprinted from reference 40 F. Sugahara, T. Uchiyama, H. Watanabe, Y. Shimazu, M. Kuwayama, Y. Fujii, K. Kiyotani, A. Adachi, N. Kohno, T. Yoshida and T. Sakaguchi, *Virology*, 2004, **325**, 1-10. with permission from Elsevier.

1.4.2 Interaction with other viral components

Matrix proteins may control the assembly process by interacting with other viral components to recruit them into the virion as it forms. The ability of the N-terminal domain of VP40 to bind RNA may support the suggestion that matrix proteins provide a link between the membrane glycoproteins and the nucleocapsid complex,[30, 32] although in the *Mononegavirales* the genomic RNA is tightly encapsidated in the nucleocapsid complex and is unlikely to be available for interaction. The RSV M protein interacts with the cytoplasmic tail of the F protein, and other matrix proteins have been shown to associate with the cytoplasmic regions of viral fusion proteins.[43] Such interactions may influence the location at which assembly and budding occurs; this is discussed in greater detail in section 1.4.5 below.

For several enveloped virus matrix proteins, expression in conjunction with viral glycoproteins or nucleocapsids increases the yield of VLPs produced, suggesting that matrix proteins promote budding by linking the nucleocapsid to the viral envelope, facilitating transport of viral RNPs to the area of the plasma membrane where the glycoproteins are embedded.[44] Co-expression of human parainfluenza virus 1 matrix protein with the nucleocapsid protein (NP) demonstrated that the matrix protein was responsible for specifically incorporating the nucleocapsid into virions, interacting with the C-terminal domain of the NP protein.[34] Influenza M1 has also been shown to have RNP- as well as membrane-binding ability, and assists in RNP export.[45]

A further role of the matrix protein may be to stop the process of transcription and replication carried out by the viral polymerase, before the holo-nucleocapsid complex is incorporated into the assembling virion. It has been shown that matrix proteins from RSV, Newcastle disease virus and Sendai virus localize in the host cell nucleus early in infection, where they may perform the additional function of inhibiting transcription of the host cell genetic material. One study found that nuclear extracts from RSV infected cells were deficient in cellular transcriptional activity,[28] and the finding that RSV M binds RNA raised the possibility that this could occur through direct interaction with RNA.[32, 44]

1.4.3 Structures of other viral matrix proteins

To date, structural data obtained by X-ray crystallography have been published for four of the enveloped (-)RNA virus matrix proteins; three from the *Mononegavirales,* VP40 from Ebola virus,[46] M from VSV[47] and M from Borna disease virus (BDV),[48] and M1 from the orthomyxovirus influenza A.[49] Of these crystal structures only that of BDV M represents a full-length protein, although the data for VP40 are also fairly complete. The proteins have a low sequence identity of between 2 and 7 %, and their structures appear to have little in common, with no shared motif that might mediate their common membrane binding function, although the structure of BDV M does somewhat resemble the two domains of VP40.[30,48] Despite this structural diversity they share the same key functionalities, membrane binding, coordinating assembly and roles in budding and formation of mature virions.

1.4.3.1 Influenza M1

The first matrix protein structure published, in 1997, was a stable fragment (residues 2-158 of 252 in total) of the influenza M1 protein, which has been crystallized at both low[50] and neutral pH.[45] Both conditions gave an identical dimeric structure, (Figure 1.6), consisting of two highly helical subdomains that pack against each other, joined by a possibly flexible linker containing a short helix; there are 9 helices in total. Amino acid residues 88-111 were found to be necessary for RNA binding through protein mapping and deletion mutations; RNA binding is likely to have more importance for influenza M1 than for the *Mononegavirales* matrix proteins, as the RNA is found on the outside of the nucleocapsid complex and is available for interaction. Other notable features include a region of 10 positively charged side chains exposed on a surface formed by helices 6 and 7.

The area of the protein responsible for lipid binding was expected to be located in the N-terminal domain (residues 1-80). There is no exposed hydrophobic surface suitable for membrane insertion in this part of the structure, but a hydrophobic region, formed by helices 1 and 4, can be found at the interface between two

monomers. It was proposed that this region could be exposed by a conformational change, allowing it to insert into the inner leaflet of the lipid bilayer,[45, 50] but no evidence has been found for this mechanism. At the time of publication no structurally homologous proteins were found in the Protein Data Bank.[50]

```
1 MSLLTEVETYVLSIIPSGPLKAEIAQRLEDVFAGKNTDLEVLMEWLKTRPILSPL
56 TKGILGFVFTLTVPSERGLQRRRFVQNALNGNGDPNNMDKAVKLYRKLKREITF
110 HGAKEISLSYSAGALASCMGLIYNRMGAVTTEVAFGLVCATCEQIADSQ
```

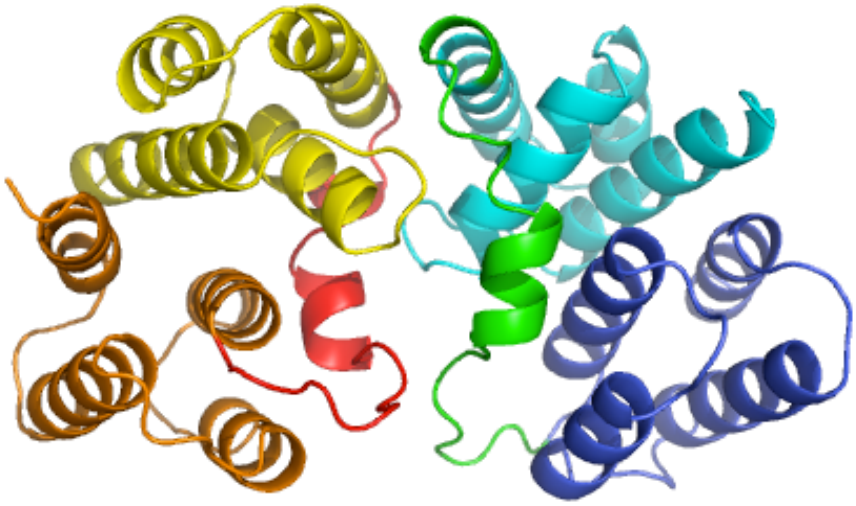


Figure 1.6: Sequence and structure of the crystallized dimer of part of influenza virus matrix protein M1. Each monomer consists of residues 2-158 of the 252-residue protein. Residues 88-111 (underlined) are responsible for RNA binding. The N-terminal domain (helices 1-4) is orange (blue) and the middle domain (helices 6-9) yellow (cyan), with the short linking helix (helix 5) in red (green). The cleaved residues 165-252 should form at least one other C-terminal domain.[50] Entry 1AA7 in the RSCB Protein Data Bank (http://www.rcsb.org/pdb).[†]

1.4.3.2 VSV M

A soluble thermolysin-resistant core of VSV M (residues 58-226 out of a total 229) was crystallized in 2002. This fragment does not self-associate or form aggregates like the full-length protein, and residues 119-128 were not resolved in the crystal structure. The published structure consists of two unequally sized domains; the N-terminal domain (residues 58-180) contains a large anti-parallel β-sheet made up of five strands, packed against two α-helices. The C-terminal part of the structure (residues 201-226) is made up of a small two-stranded anti-parallel β-sheet with an

[†] This image and those in Figures 1.7-1.9 were prepared using the program PyMol, DeLano Scientific, Palo Alto, CA, USA. http://www.pymol.org/

α-helix, and a 20-amino acid region of random coil joins the two domains. Again the protein fold is not found in any other published structures.[47]

The structure accounts for budding deficiencies seen in several temperature-sensitive mutants of VSV M. The mutated residues are positioned so as to disrupt efficient hydrogen bonding in the protein, rendering it less stable. The crystallized core does not bind to negatively charged liposomes as effectively as the full-length protein, suggesting that the missing lysine-rich N-terminal section is important for interaction with membranes. The disordered loop between residues 180 and 201 is also important in such interactions because of its hydrophobicity (especially the sequence PAVLA, which is responsible for self-association of the protein and is thus likely to be important for the role of M in the assembly process[51]), along with some small hydrophobic regions on the surface that are positioned close to the loop. Membrane binding is therefore likely to be mediated by a combination of hydrophobic and electrostatic interactions.[6, 47]

```
58 QLRYEKFFFTVKMTVRSNRPFRTYSDVAAAVSHWDHMYIGMAGKRPFYKILAFL
110 GSSNLKAT(PAVLADQGQP)EYHAHCEGRAYLPHRMGKTPPMLNVPEHFRRPF
163 NIGLYKGTVELTMTIYDDESLEAAPMIWDHFNSSKFSDFREKALMFGLIVEKK
214 ASGAWVLDSVSHFK
```

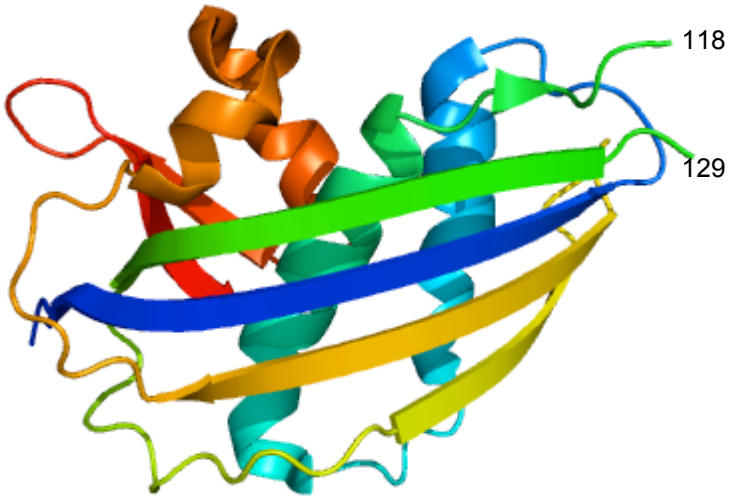


Figure 1.7: Amino acid sequence and ribbon diagram showing the crystallized core fragment (residues 58-226) of the VSV matrix protein.
The structure is coloured blue (N terminus) to red (C terminus). Note the unresolved region of structure between residues 118 and 129, bracketed in the sequence. Entry 1LG7 in the Protein Data Bank.[47]

1.4.3.3 Ebola VP40

A crystal structure of an incomplete Ebola virus matrix protein, VP40, published in 2000, (Figure 1.8) revealed a monomer consisting of two domains joined by a flexible linker. This linking region was not resolved in the published crystal structure. The N-terminal domain forms a β-sandwich structure composed of two sheets each containing three antiparallel strands each, and there are three loop regions at the furthest point from the C-terminal domain. This domain (residues 213-326) is required for membrane association *in vitro*, and consists of a three-stranded antiparallel β-sheet, with three other strands opposite forming a bent sheet. The two domains lie at about 60º to each other, and the lack of resolution in the linker region between them suggests possible flexibility in this area of the protein.

Large hydrophobic surfaces on the C-terminal domain would provide a possible mechanism for membrane association to occur through hydrophobic interaction, although preferential binding to negatively charged lipids suggests that there is also an electrostatic component to membrane binding.[46] Octameric VP40 (Figure 1.10) has been found to bind (and be stabilized by) an RNA triribonucleotide, which may indicate that the protein is involved in an as yet unknown regulatory step in the viral life cycle.[33]

```
44 GDTPSNPLRPIADDTIDHASHTPGSVSSAFILEAMVNVISGPKVLMKQIPIWLPLGVA
112 DQKTYSFDSTTAAIMLASYTITHFGKATNPLVRVNRLGPGIPDHPLRLLRIGNQAFL
159 QEFVLPPVQLPQYFTFDLTALKLITQPLPAATWTDD(TPTGS)NGALRPGISFHPKL
214 RPILLPNKSGKKGNSADLTSPEKIQAIMTSLQDFKIVPIDPTKNIMGIEVPETLVLK
270 LTGKKVTSKNGQPIIPVLLPKYIGLDPVAPGDLTMVITQDCDTCHSPASLPAVIEK
```

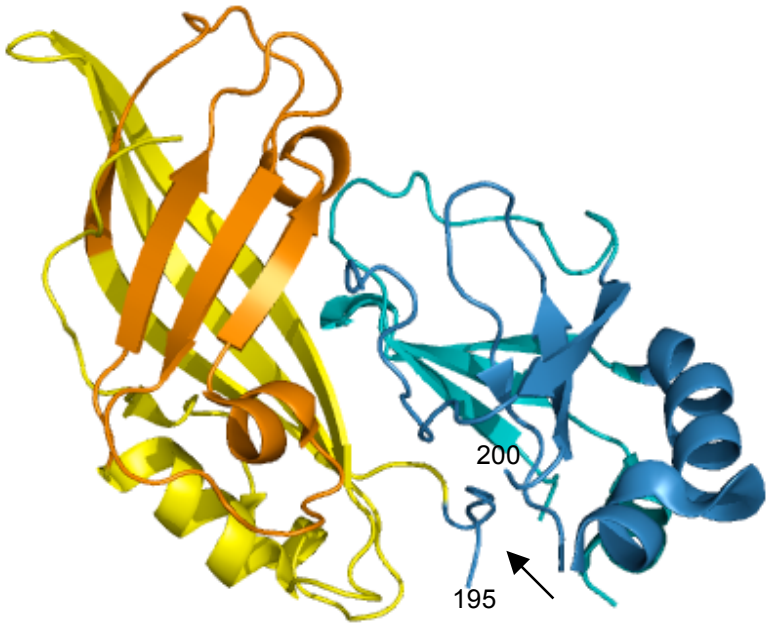


Figure 1.8: Amino acid sequence and ribbon diagram showing the incomplete (residues 44-326) crystal structure of Ebola virus VP40.
The N-terminal domain is shown in yellow/orange and the C-terminal domain in blue/cyan; the sequence is coloured to match the structure. Note the unresolved residues (195-200) in the region of random coil linking the two domains, indicated by an arrow, and bracketed in the sequence. Entry 1ES6 in the Protein Data Bank.[46]

1.4.3.3 Borna disease virus M

A crystal structure of the matrix protein from another member of the *Mononegavirales*, Borna disease virus, was published to 2.65 Å resolution in 2009 (Figure 1.9). BDV-M tetramers were crystallized and described.[48] The structure of each monomer is somewhat similar to the N-terminal domain of Ebola VP40 described above, although the sequence identity between the two proteins is only 10 %. The crystallized BDV-M monomer is 44 × 40 × 40 Å in size and forms an L-shaped β-sandwich made up of 6 antiparallel strands, arranged in 2 β-sheets of 3 strands each. The strands are bent and twisted slightly to give a concave shape. The 2 open sides of the β-sandwich are bridged on one side by helix α3, 2 loops and the N-terminus of the peptide chain, and on the other by helix α4.

The non-covalently linked BDV-M tetramer has a square planar arrangement, in which the L-shaped monomers tessellate. The C-terminal portion of one monomer interacts with the N-terminus and β-sheet II of the next monomer, with about 26 % of each monomer being buried in the tetramer. One face of the tetramer has a highly positively charged surface and is rich in hydrophobic and aromatic side chains, and it is suggested that this face is responsible for membrane interaction. In common with VP40 octamers,[33] BDV-M tetramers were crystallized with bound nucleic acid (cytidine-5'-monophosphate). However in BDV-M the binding site is on the opposite side of the monomer to the corresponding site in VP40, and the binding pocket is small so cannot accommodate bulkier bases such as adenine and guanine.

```
1 MNSKHSYVELKDKVIVPGWPTLMLEIDFVGGTSRNQFLNIPFLSVKEPLQLPREKKLTDY
61 FTIDVEPAGHSLVNIYFQIDDFLLLTLNSLSVYKDPIRKYMFLRLNKEQSHAINAAFNV
120 FSYRLRNIGVGPLGPDIRSSGP
```

Figure 1.9: Amino acid sequence and ribbon diagram showing the crystallized BDV matrix protein.
The structure is coloured blue (N terminus) to red (C terminus). This structure is entry 3F1J in the Protein Data Bank.[48]

1.4.4 Self-assembly of M proteins

During assembly of a virion, matrix proteins self-assemble into higher order oligomers at the host cell membrane, which provides the driving force for budding.[30, 47] Part of the N-terminal domain of VP40 was found to oligomerize into hexamers and octamers, when stabilized by binding a specific fragment of bacterial RNA (Figure 1.10),[33, 52] while the C-terminal domain is thought to simultaneously mediate membrane binding, which does not occur unless this domain is present.[30, 31] It has been suggested that the various oligomeric forms of VP40 might each play a distinct role in the viral life cycle.[52] The influenza M1 protein has a high propensity to oligomerize *in vitro*,[30] and assembled into elongated oligomers (or 'ribbons') in neutral conditions that have the potential to produce helices or coils that resembled the form taken by the protein in the virus particle.[45] The M protein from Sendai virus, a paramyxovirus, also self-assembled into tubes and sheets *in vitro*,[53] and VSV M has been found to self-associate into large multimers at physiological salt concentrations.[47] BDV-M forms oligomers *in vivo* and *in vitro*; the tetrameric form is the most stable, and further assembles into 2D lattice-like structures.[48]

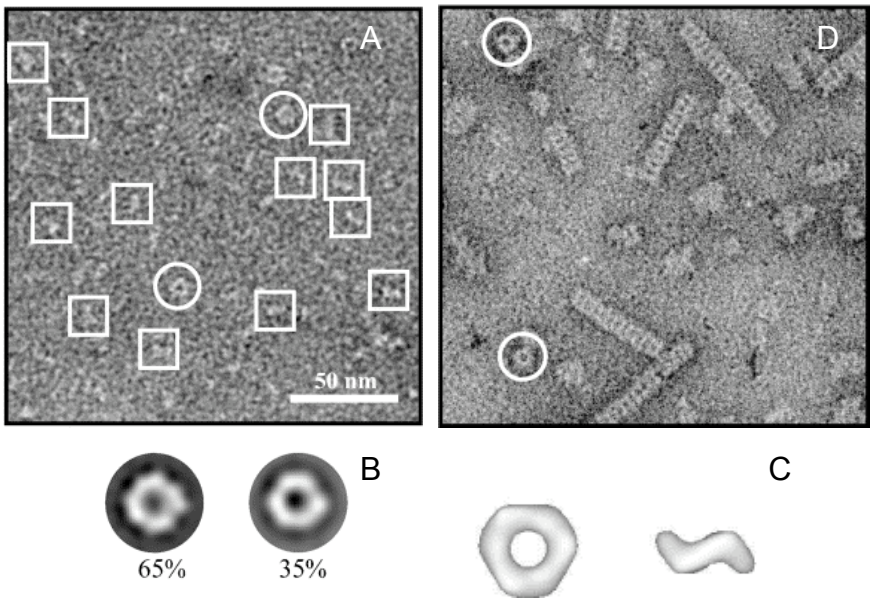


Figure 1.10: Oligomers and polymers of Ebola VP40
A: Ring like structures seen by negative-stain electron microscopy, viewed either side on (squares) or from above (circles). B: Approximately 65% of the rings appeared octameric and 35% hexameric. C: 3-dimensional reconstruction of a hexameric oligomer. D: Ring-like oligomers stacking to form polymers.[52†]

[†] Reproduced from reference 46 J. Timmins, G. Schoehn, C. Kohlhaas, H.-D. Klenk, R. W. H. Ruigrok and W. Weissenhorn, *Virology*, 2003, **312**, 359-368, with permission from Elsevier

1.4.5 Matrix proteins and membranes

Matrix proteins have an intrinsic ability to bind membranes, and when expressed using recombinant DNA methods in living cells, a significant fraction of the protein will be found tightly bound to membranes. Although they are peripheral membrane proteins, the membrane binding of several (-)RNA virus matrix proteins has been shown to be unusually stable when exposed to high salt concentrations or extremes of pH.[6]

Matrix proteins are thought to localize to and determine the site of assembly at the host cell lipid membrane *via* interactions with viral glycoproteins that have already been transported to the budding site, and whose cytoplasmic tail domains (CTDs) project through the membrane into the cytoplasm of the cell.[2, 46] In influenza virus it has been shown that such interactions are necessary in order for the viral nucleocapsid to be packaged correctly in the budding virion.[54] The composition of the lipid bilayer at the site of assembly also seems to be important.

It has been shown that several matrix proteins, including Ebola virus VP40, influenza virus M1 and VSV M, preferentially interact with membranes containing high levels of negatively charged lipids.[31, 46, 49, 55] This may occur *via* electrostatic interactions with positively charged areas on the surface of the matrix protein, found in influenza M1, and thought to be present in the N-terminal domain of VSV M, which was missing in the crystal structure. Matrix protein-membrane interactions are also likely to have a hydrophobic component, as hydrophobic regions are found in all the protein structures described above, which could associate with, or insert into, the hydrophobic interiors of membranes.

1.4.5.1 The structure of biological membranes

The Singer-Nicholson model of biological membranes is a useful starting point for understanding their structure. It consists of amphipathic lipid molecules arranged tail to tail in two leaflets, forming a lipid bilayer (Figure 1.11). The hydrophobic acyl chains, or 'tails' of the lipids associate with each other in the interior of the bilayer,

allowing them to be segregated from the aqueous media on either side of the membrane. The lipid 'head groups', charged hydrophilic groups such as choline, are arrayed on the outside of the bilayer where they can be solvated by water (more information about specific lipids is given in Chapters 2 and 3). In this model membrane proteins may be intrinsic, floating in a sea of lipid, or extrinsic, peripherally bound to the membrane surface.[56]

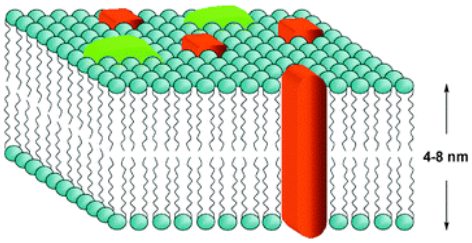


Figure 1.11: The Singer-Nicolson fluid mosaic model of a biological membrane. Peripheral proteins associated with the bilayer are shown in green, integral proteins in red.[56§]

The lipid bilayer exhibits liquid crystalline properties: pure samples of simple lipids go through clear transitions between gel ($L_{\beta'}$), rippled gel ($P_{\beta'}$) and liquid crystalline (L_{α}) phases with increasing temperature and/or pressure (Figure 1.12). Unsaturated lipids, with one or more double bonds in their acyl chains, tend to go through the main phase transition (T_m) between the $P_{\beta'}$ and L_{α} phases below 0°C and therefore membranes containing such lipids are more fluid under physiological conditions than membranes containing a greater proportion of saturated chain lipids.[56]

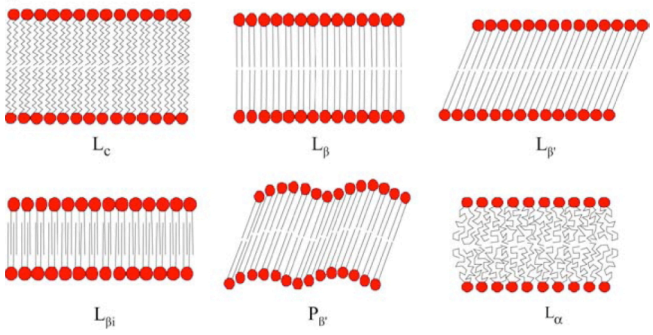


Figure 1.12: The arrangement of lipid molecules in typical lamellar phases that may be induced by changing temperature or pressure.[†]

[§] Reproduced by permission of the Royal Society of Chemistry
[†] Reprinted from, J. Eisenblätter and R. Winter, *Biophysical Journal*, 2006, **90**, 956-966, Copyright (2006), with permission from Elsevier.

Within the bilayer there is rapid lateral diffusion of lipid molecules in the plane of the membrane, but individual molecules do not readily 'flip-flop' from one leaflet of the bilayer to the other because of the thermodynamic constraints upon the movement of a charged headgroup through the lipophilic core.[57]

1.4.5.2 Lipid microdomains

Phase-separated lipid microdomains in the host cell membrane may help to determine where assembly occurs. Such domains are proposed to be areas within biological membranes, enriched in cholesterol, sphingolipids and certain proteins, that exist in a liquid-crystalline ordered (L_o) phase, surrounded by lipids in the liquid-crystalline disordered (L_d) phase. In the L_d phase lipids have full freedom of diffusional, rotational and conformational motion within the plane of the bilayer, while in the L_o phase intercalation of relatively flat, rigid cholesterol molecules keeps the lipid acyl chains in an extended, kink-free conformation, and movement is restricted to free diffusion in the bilayer plane (Figure 1.13).[58-60] Membrane domains with a L_o-phase-like structure are called 'lipid rafts' by many researchers, a label referring to the idea of these more ordered areas 'floating' within the dynamic sea of lipids described by the Singer-Nicholson fluid mosaic model of membranes.[60, 61] The existence of phase-separated domains has been demonstrated in model membranes *in vitro* by techniques such as fluorescence microscopy, NMR and atomic force microscopy (AFM), that allow areas of heterogeneous membrane to be detected and visualized.[62-65]

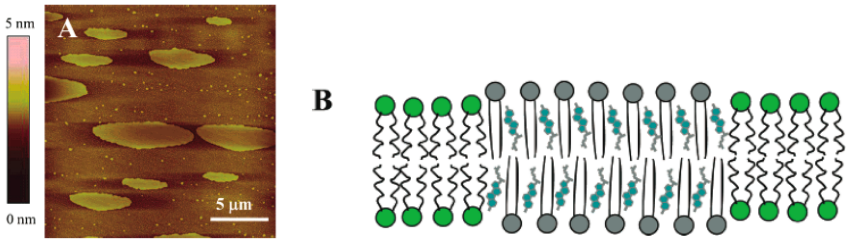


Figure 1.13: A: Microdomains in a supported lipid bilayer (DOPC/SM/cholesterol 2:2:1) imaged by AFM; B: Cartoon depiction of a lipid bilayer with coexisting L_o and L_d phases. Note the intercalating cholesterol molecules (in blue) in the L_o phase.[63†]

† Reprinted with permission from L. J. Johnston, *Langmuir*, 2007, **23**, 5886-5895, Copyright 2007 American Chemical Society.

Ternary mixtures of saturated lipids, unsaturated lipids and cholesterol are known to produce coexisting Lo and Ld phases below a miscibility transition temperature T_{mix}. Sphingomyelin is particularly suited as the saturated component because it can form hydrogen bonds with the 3-OH group of cholesterol, not only through its ester linkage, but also through its amine and hydroxyl groups. Figure 1.14 shows a miscibility phase diagram for mixtures of DOPC, palmitoyl sphingomyelin and cholesterol, and corresponding fluorescence microscopy images of giant vesicles displaying phase separated microdomains.

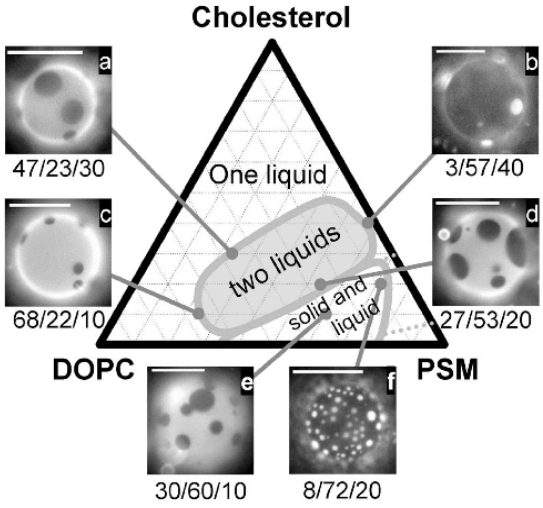


Figure 1.14: Miscibility phase diagram for mixtures of cholesterol, DOPC and sphingomyelin. Giant vesicles are shown corresponding to various proportions of the three components (shown under each image); scale bars are 20 μm.[66†]

Despite the evidence for their existence *in vitro*, the presence of ordered microdomains within cell membranes *in vivo* is a matter of ongoing debate.[67, 68] L_o phase domains are very often identified with detergent resistant membranes (DRMs), cell membrane fractions that are characterized by their insolubility in non-ionic detergents such as Triton X-100 at 4ºC, and the presence within them of lipid components such as cholesterol and sphingomyelin that are associated with microdomains *in vitro*.[60, 69-71] A number of membrane proteins are also considered to be representative of DRMs. It is not necessarily correct to think of lipid microdomains and DRMs as synonymous, however, and the idea is certainly not accepted by all researchers in this area.[59, 72] It is highly likely that treating membranes with detergent alters the number and size of ordered domains, or even

† Copyright (2005) by the American Physical Society

induces their appearance, as mixing of the detergent with the membrane is non-ideal and shifts lipid partitioning in favour of the ordered regions.[59] In addition, DRMs prepared using different detergents appear to contain different proportions of the various lipids; this may be because the putative L_o domains themselves are not homogenous, or could arise from the varying abilities of detergents to solubilize the wide range of lipids and proteins found in membranes.[58]

Localisation of the site of viral assembly to areas rich in certain membrane components, however such areas are defined, may ensure that specific lipids and/or host cell membrane proteins are incorporated into the viral membrane. Co-expression of Ebola VP40 and GP proteins resulted in the formation of virus-like particles that also contained the ganglioside G_{M1}, which is considered to be a marker for detergent resistant domains,[73, 74] suggesting that the particles were produced by budding from such domains.[75] RSV M is found in association with DRMs when expressed alongside the membrane fusion protein F, which itself associates with such membrane fractions without a requirement for any other viral proteins.[27, 76] RSV M, G and N proteins were also found in Triton X-100 resistant fractions isolated from infected cells, and the purified virus has been found to contain the proteins CD55, CD59 and caveolin-1 in its envelope, which again are often cited as markers for lipid microdomains.[77, 78] There is therefore evidence that phase-separated microdomains of specific composition within plasma membranes may act as platforms for assembly and budding of *Mononegavirales* virions, and that matrix proteins may help to localize the assembly process at these areas, or that the M protein needs to be localized to these regions to facilitate assembly.

1.5.4.3 Other peripheral membrane proteins

It is not only viral matrix proteins that bind peripherally to membranes. The total protein content of various types of cell membranes ranges from around 20 % (*e.g.* in the myelin sheath) to over 70 % (inner membranes of mitochondria), with around 50 % in plasma membranes,[57] and many cytosolic proteins are recruited to cellular membranes to form protein-protein and lipid-protein interactions during cell signalling and membrane trafficking. Such peripheral proteins use a range of

strategies to form reversible membrane interactions; some have one or more modular domains specializing in lipid binding, known as membrane targeting domains, while others utilise parts of their molecular surface (*e.g.* secretory phospholipase A_2), or amphipathic secondary structure regions (*e.g.* ADP-ribosylation factor), to interact with membranes. Others, such as Src and Ras proteins, possess covalently attached lipid anchors that embed themselves in the lipid bilayer.[79]

One group of membrane targeting domains are Bin-Amphiphysin-Rvs (BAR) domains, which have been shown to sense and induce membrane curvature. For example the BAR-domain containing proteins Amphiphysins mediate vesicle formation by assembling at plasma membranes, and along with another class of proteins, Arfaptins, form V-shaped dimers that have been proposed to detect and/or produce membrane curvature.[80] This type of protein function has clear parallels with the role of RSV M in viral assembly and budding.

1.5 The scope of this research

Several important features of matrix proteins, and specifically the RSV M protein, have been described above, but the picture is by no means complete and a number of important issues remain.

RSV M is known to localize to detergent resistant membrane fractions together with viral glycoproteins,[27, 77] although it is not clear whether it can associate with such membranes alone, or if the association with DRMs reflects binding of M to specific membrane microdomains. The crystal structures of other *Mononegavirales* matrix proteins, discussed in section 1.4.2 above, have revealed that they possess both hydrophobic and charged surface regions, allowing for membrane interaction *via* electrostatic and/or hydrophobic effects. RSV M is considered to be a peripheral membrane protein that does not significantly penetrate the lipid bilayer,[27] but little is really understood about the exact nature of its interactions with membranes and lipids, especially as its structure is not known.

It is therefore important to address the following questions:

- Does RSV M bind to phase-separated microdomains within membranes?
- If so, are there any particular lipids that are important?
- Does M bind to any other specific lipids?
- Is it possible to deduce information about the thermodynamics of binding in any of these cases?

In order to answer these questions, the work described in this thesis aims to develop methods to directly study matrix protein interaction with membranes. This has mainly been addressed so far using *in vitro* studies of matrix proteins associating with lipid vesicles,[31, 49, 55] but whilst these have given a qualitative picture of the types of membranes with which these proteins associate, they do not provide very much quantitative information about the binding process. In addition, the conclusion that matrix proteins bind to membrane microdomains has been reached only by detecting matrix proteins in association with detergent resistant membrane fractions. However, it does not necessarily follow that the lipids in such fractions have actually been in the form of phase-separated domains whilst binding to the matrix protein. Various techniques, including tensiometry, Brewster angle microscopy (BAM) and atomic force microscopy (AFM), will be used to visualize model membranes in association with RSV M, in order to test the hypothesis that it binds to distinct lipid microdomains. They will also aim to reveal to which, if any, lipids the M protein will preferentially bind, and provide quantitative data relating to the process.

In addition, efforts will be directed towards the problem of determining the 2D and 3D structures of the RSV matrix protein. This will reveal features that may shed light on how the protein is able to interact with membranes, individual lipids, other M protein molecules, or other viral components such as the N protein and the cytoplasmic tails of the F and G proteins. It will be interesting to determine whether the 3D structure of RSV M bears a resemblance to any of the other published matrix protein structures, which do not share any notable features, and have not yet provided a complete picture of the mechanism by which matrix proteins assemble and associate with membranes during viral replication.

1.6 References

1 M. Abercrombie, M. Hickman, M. L. Johnson and M. Thain, *The New Penguin Dictionary of Biology*, Penguin Books, London, 1992.

2 E. K. Wagner, M. J. Hewlett, D. C. Bloom and D. Camerini, *Basic Virology*, Blackwell Publishing, Oxford, 2008.

3 P. K. Murray, K. S. Rosenthal, G. S. Kobayashi and M. A. Pfaller, *Medical Microbiology*, Mosby Inc., St. Louis, 2002.

4 J. H. Strauss and E. G. Strauss, *Viruses and Human Disease*, Academic Press, London, 2002.

5 E. O. Freed, *Virus Research*, 2004, **106**, 85-86.

6 A. P. Schmitt and R. A. Lamb, in *Biology of Negative Strand RNA Viruses: The Power of Reverse Genetics*, ed. Y. Kawaoka, Springer Science & Business, 2004.

7 J. Handforth, M. Sharland and J. S. Friedland, *BMJ*, 2004, **328**, 1026-1027.

8 R. L. Smyth and P. J. M. Openshaw, *Lancet*, 2006, **368**, 312-322.

9 D. Edell, V. Khoshoo, G. Ross and K. Salter, *Chest*, 2002, **122**, 935-939.

10 N. G. Hansbro, J. C. Horvat, P. A. Wark and P. M. Hansbro, *Pharmacology and Therapeutics*, 2008, **117**, 313-353.

11 K. Thorburn, S. Harigopal, V. Reddy, N. Taylor and H. K. F. van Saene, *Thorax*, 2006, **61**, 611-615.

12 C.-H. Chen, Y.-T. Lin, Y.-H. Yang, L.-C. Wang, J.-H. Lee, C.-L. Kao and B. L. Chiang, *Pediatric Allergy and Immunology*, 2008, **19**, 166-172.

13 J. A. Englund, P. A. Piedra and E. Whimbey, *The American Journal of Medicine*, 1997, **102**, 61-70.

14 K. Ventre and A. Randolph, *Cochrane Database of Systematic Reviews*, 2007, CD000181.

15 D. K. Mehta ed., *The British National Formulary*, BMJ Publishing Group Ltd. and RPS Publishing, London, 2006.

16 N. Tanaka, F. Ishida and E. Tanaka, *The New England Journal of Medicine*, 2004, **350**, 1264-1265.

17 T. Oze, N. Hiramatsu, N. Kurashige, N. Tsuda, T. Yakushijin, T. Kanto, T. Takehara, A. Kasahara, M. Kato, H. Yoshihara, K. Katayama, S. Kubota, T. Hijioka, K. Ishibashi, M. Oshita, H. Hagiwara, Y. Haruna, E. Mita, S. Tamura and N. Hayashi, *Journal of Gastroenterology*, 2006, **41**, 862-872.

18 A. R. Falsey, P. A. Hennessy, M. A. Formica, C. Cox and E. E. Walsh, *The New England Journal of Medicine*, 2005, **352**, 1749-1759.

19 M. C. Zambon, J. D. Stockton, J. P. Clewley and D. M. Fleming, *Lancet*, 2001, **358**, 1410-1416.

20 H. W. Kim, J. G. Canchola, C. D. Brandt, G. Pyles, R. M. Chanock, K. Jensen and R. H. Parrott, *American Journal of Epidemiology*, 1969, **89**, 422-434.

21 J. Chin, R. L. Magoffin, L. A. Shearer, J. H. Schielble and E. H. Lennette, *American Journal of Epidemiology*, 1969, **89**, 449-463.

22 A. Z. Kapikian, R. H. Mitchell, R. M. Chanock, R. A. Shvedoff and C. E. Stewart, *American Journal of Epidemiology*, 1969, **89**, 405-421.

23 G. Meyer, M. Deplanche and F. Schelcher, *Comparative Immunology, Microbiology and Infectious Diseases*, 2008, **31**, 191-225.

24 U. F. Power, *Journal of Clinical Virology*, 2008, **41**, 38-44.

25 O. Dibben, L. C. Thorpe and A. J. Easton, *Virus Research*, 2007, **131**, 47-53.

26 T. Takimoto and A. Portner, *Virus Research*, 2004, **106**, 133-145.

27 G. Henderson, J. Murray and R. P. Yeo, *Virology*, 2002, **300**, 244-254.

28 G. Brown, C. E. Jeffree, T. McDonald, H. W. M. Rixon, J. D. Aitken and R. J. Sugrue, *Virology*, 2004, **327**, 175-185.

29 T. Bächi and C. Howe, *Journal of Virology*, 1973, **12**, 1173-1180.

30 J. Timmins, R. W. H. Ruigrok and W. Weissenhorn, *FEMS Microbiology Letters*, 2004, **233**, 179-186.

31 R. W. H. Ruigrok, G. Schoehn, A. Dessen, E. Forest, V. Volchkov, O. Dolnik, H. D. Klenk and W. Weissenhorn, *Journal of Molecular Biology*, 2000, **300**, 103-112.

32 R. Ghildyal, A. Ho, K. M. Wagstaff, M. M. Dias, C. L. Barton, P. Jans, P. Bardin and D. L. Jans, *Biochemistry*, 2005, **44**, 12887-12895.

33 F. X. Gomis-Rüth, A. Dessen, J. Timmins, A. Bracher, L. Kolesnikowa, S. Becker, H.-D. Klenk and W. Weissenhorn, *Structure*, 2003, **11**, 423-433.

34 E. C. Coronel, T. Takimoto, K. G. Murti, N. Varich and A. Portner, *Journal of Virology*, 2001, **75**, 1117-1123.

35 J. Solon, O. Gareil, P. Bassereau and Y. Gaudin, *Journal of General Virology*, 2005, **86**, 3357-3363.

36 J. H. Connor, M. O. McKenzie and D. S. Lyles, *Journal of Virology*, 2006, **80**, 3701-3711.

37 J. Timmins, S. Scianimanico, G. Schoehn and W. Weissenhorn, *Virology*, 2001, **283**, 1-6.

38 T. Noda, H. Sagara, E. Suzuki, A. Takada, H. Kida and Y. Kawaoka, *Journal of Virology*, 2002, **76**, 4855-4865.

39 P. A. Justice, W. Sun, Y. Li, Z. Ye, P. R. Grigera and R. R. Wagner, *Journal of Virology*, 1995, **69**, 3156-3160.

40 F. Sugahara, T. Uchiyama, H. Watanabe, Y. Shimazu, M. Kuwayama, Y. Fujii, K. Kiyotani, A. Adachi, N. Kohno, T. Yoshida and T. Sakaguchi, *Virology*, 2004, **325**, 1-10.

41 T. Takimoto, K. G. Murti, T. Bousse, R. A. Scroggs and A. Portner, *Journal of Virology*, 2001, **75**, 11384-11391.

42 J. R. Patch, G. Crameri, L.-F. Wang, B. T. Eaton and C. C. Broder, *Virology Journal*, 2007, **4**.

43 A. G. P. Oomens, K. P. Bevis and G. W. Wertz, *Journal of Virology*, 2006, **80**, 10465-10477.

44 L. Rodríguez, I. Cuesta, A. Asenjo and N. Villanueva, *Journal of General Virology*, 2004, **85**, 709-719.

45 A. Harris, F. Forouhar, S. Qiu, B. Sha and M. Luo, *Virology*, 2001, **289**, 34-44.

46 A. Dessen, V. Volchkov, O. Dolnik, H.-D. Klenk and W. Weissenhorn, *The EMBO Journal*, 2000, **19**, 4228-4236.

47 M. Gaudier, Y. Gaudin and M. Knossow, *The EMBO Journal*, 2002, **21**, 2886-2892.

48 P. Neumann, D. Lieber, S. Meyer, P. Dautel, A. Kerth, I. Kraus, W. Garten and M. T. Stubbs, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 3710-3715.

49 R. W. H. Ruigrok, A. Barge, P. Durrer, J. Brunner, K. Ma and G. R. Whittaker, *Virology*, 2000, **267**, 289-298.

50 B. Sha and M. Luo, *Nature Structural Biology*, 1997, **4**, 239-244.

51 M. Gaudier, Y. Gaudin and M. Knossow, *Virology*, 2001, **288**, 308-314.

52 J. Timmins, G. Schoehn, C. Kohlhaas, H.-D. Klenk, R. W. H. Ruigrok and W. Weissenhorn, *Virology*, 2003, **312**, 359-368.

53 M. H. Heggeness, P. R. Smith and P. W. Choppin, *Proceedings of the National Academy of Sciences of the United States of America*, 1982, **79**, 6232-6236.

54 J. Zhang, G. P. Leser, A. Pekosz and R. A. Lamb, *Virology*, 2000, **269**, 325-334.

55 J. J. Zakowski, W. A. Petri Jr. and R. R. Wagner, *Biochemistry*, 1981, **20**, 3902-3907.

56 J. M. Sanderson, *Organic & Biomolecular Chemistry*, 2005, **3**, 201-212.

57 T. M. Devlin, in *Textbook of Biochemistry with Clinical Correlations*, ed. T. M. Devlin, Wiley-Liss, New York, 5 edn., 2002.

58 T. P. W. McMullen, R. Lewis and R. N. McElhaney, *Current Opinion in Colloid & Interface Science*, 2004, **8**, 459-468.

59 D. Lichtenberg, F. M. Goñi and H. Heerklotz, *Trends in Biochemical Sciences*, 2005, **30**, 430-436.

60 N. Chazal and D. Gerlier, *Microbiology and Molecular Biology Reviews*, 2003, **67**, 226-237.

61 S. J. Singer and G. L. Nicolson, *Science*, 1972, **175**, 720-731.

62 S. L. Veatch, I. V. Polozov, K. Gawrisch and S. L. Keller, *Biophysical Journal*, 2004, **86**, 2910-2922.

63 L. J. Johnston, *Langmuir*, 2007, **23**, 5886-5895.

64 A. V. Samsonov, I. Mihalyov and F. S. Cohen, *Biophysical Journal*, 2001, **81**, 1486-1500.

65 H. A. Rinia, M. M. E. Snel, J. P. J. M. van der Eerden and B. de Kruijff, *FEBS Letters*, 2001, **501**, 92-96.

66 S. L. Veatch and S. L. Keller, *Physical Review Letters*, 2005, **94**, 148101.

67 J. Duggan, G. Jamal, M. Tilley, B. Davis, G. McKenzie, K. Vere, M. G. Somekh, P. O'Shea and H. Harris, *European Biophysics Journal*, 2008, **37**, 1279-1289.

68 M. Edidin, *Annual Reviews of Biophysics and Biomolecular Structure*, 2003, **32**, 257-283.

69 D. A. Brown and E. London, *Journal of Biological Chemistry*, 2000, **275**, 17221-17224.

70 X. Xu and E. London, *Biochemistry*, 2000, **39**, 844-849.

71 L. J. Pike, *Biochemical Journal*, 2004, **378**, 281-292.

72 J. Sot, L. A. Bagatolli, F. M. Goni and A. Alonso, *Biophysical Journal*, 2006, **90**, 903-914.

73 T. Harder, P. Scheiffele, P. Verkade and K. Simons, *Journal of Cell Biology*, 1998, **141**, 929-942.

74 C. Yuan, J. Furlong, P. Burgos and L. J. Johnston, *Biophysical Journal*, 2002, **82**, 2526-2535.

75 S. Bavari, C. M. Bosio, E. Wiegand, G. Ruthel, A. B. Will, T. W. Gisbert, M. Hevey, C. Schmaljohn, A. Schmaljohn and M. J. Aman, *Journal of Experimental Medicine*, 2002, **195**, 593-602.

76 E. H. Fleming, A. A. Kolokoltsov, R. A. Davey, J. E. Nichols and N. J. Roberts Jr., *Journal of Virology*, 2006, **80**, 12160-12170.

77 A. Marty, J. Meanger, J. Mills and R. Ghildyal, *Archives of Virology*, 2004, **149**, 199-210.

78 G. Brown, J. Aitken, H. W. M. Rixon and R. J. Sugrue, *Journal of General Virology*, 2002, **83**, 611-621.

79 W. Cho and R. V. Stahelin, *Annual Reviews of Biophysics and Biomolecular Structure*, 2005, **34**, 119-151.

80 B. Habermann, *EMBO Reports*, 2004, **5**, 250-255.

2 The structure of RSV M

2.1 Protein expression and purification

All the work described in this thesis was carried out using recombinant RSV matrix protein samples expressed by Dr. Paul Yeo at the Wolfson Research Institute, Durham University Queens Campus. Purification and further analysis was carried out by the author.

The amino acid sequence of the M protein from RSV strain A2 is shown in Figure 2.1. The expression protocol used a plasmid that incorporated a His tag at the N-terminus of the protein, to facilitate purification by immobilized metal affinity chromatography (IMAC). Without this tag sequence, the protein is 256 amino acids long, and has a calculated molecular mass[1] of 29250 Da (29.25 kDa). With the tag, the molecular mass is 31633.5 Da (31.6 kDa). Work was begun using the tagged protein, but after the first few months most experiments were carried out using untagged protein, to exclude the possibility that the behaviour of the protein might be altered by the presence of the extra amino acids. The His tag was removed using the enzyme Factor Xa; however, this does not cleave off the entire tag, but leaves four additional residues (shown in brackets in Figure 2.1) at the N-terminus of all 'untagged' M protein samples.

*MGHHHHHHHHHHSSGHIEG(HMLE)*METYVNKLHEGSTYTAAVQYNVLEKDDDP
30 ASLTIWVPMFQSSMPADLLIKELANVNILVKQISTPKGPSLRVMINSRSAV
81 LAQMPSKFTICANVSLDERSKLAYDVTTPCEIKACSLTCLKSKNMLTTVKD
132 LTMKTLNPTHDIIALCEFENIVTSKKVIIPTYLRSISVRNKDLNTLENIT
182 TTEFKNAITNAKIIPYSGLLLVITVTDNKGAFKYIKPQSQFIVDLGAYLE
232 KESIYYVTTNWKHTATRFAIKP***R***ED

Figure 2.1: The amino acid sequence of the RSV matrix protein.
The His tag is shown in italics; the four residues in brackets remain after Factor Xa cleavage. Residue number 254 (bold italics) is M in the native sequence and R in the mutated form M^{254R}. Sequence numbering does not include any part of the His tag.

An issue that arose some time into the project was the identity of residue number 254 (highlighted in bold italics in Figure 2.1). This residue is a methionine (M) in the

native matrix protein, but was found to have changed to arginine (R) during preparation of the transformation vector. Mass spectrometry of several protein samples showed a molecular mass that was consistently too high, and as a result of this the protein was re-sequenced and the change noted. For the sake of consistency, most of the work reported here was carried out using the mutated protein (hereafter called M^{254R}, or M^{254R}His if still possessing the histidine tag mentioned above), unless stated otherwise. The native protein (hereafter called M^{254M} or M^{254M}His) has also been expressed, and its crystal structure was found to be identical to that of M^{254R}; the structure is described in Section 2.2.1 below. Searching the Entrez protein sequence database[2] for RSV M protein sequences has shown that residue 254 varies in any case between leucine and methionine (see Table 3.1 below), although it is arginine in related pneumovirus proteins. However, since both leucine and methionine are hydrophobic, while arginine is hydrophilic, it was considered possible that this mutation to R could have an effect on protein function, and so several experiments were carried out using both the M^{254R} and M^{254M} proteins in order to look for any significant differences.

NCBI Entrez Protein Database entry	Protein & source organism	Residue 254
NP 056860	M, human respiratory syncytial virus	L
AAB82433	Matrix protein; M, human respiratory syncytial virus	L
AAR14263	M, human respiratory syncytial virus 9320 strain	L
AAC14898	Matrix protein M, human respiratory syncytial virus	M
AAB82444	Matrix protein, human respiratory syncytial virus	L
NP 044593	Matrix protein (M), respiratory syncytial virus	M
AAC57034	Matrix protein (M), respiratory syncytial virus	M
AAB86660	Matrix protein M, human respiratory syncytial virus	M
AAC55966	Matrix protein M, human respiratory syncytial virus	M
AAB86672	Matrix protein M, human respiratory syncytial virus	M
AAB59854	Matrix protein, human respiratory syncytial virus	M
AAC57024	Matrix protein, human respiratory syncytial virus S2	M
AAX23991	Matrix protein, human respiratory syncytial virus	M
P03419	Matrix protein, human respiratory syncytial virus A2	M

Table 2.1: RSV matrix protein sequences found in the NCBI Entrez Protein Database

Each new batch of protein produced was routinely checked for purity by SDS-PAGE, to check that only the desired band at around 30 kDa appeared on the gel (Figure 2.2), before being dialysed against PBS pH 7.4 ready for use in experiments.

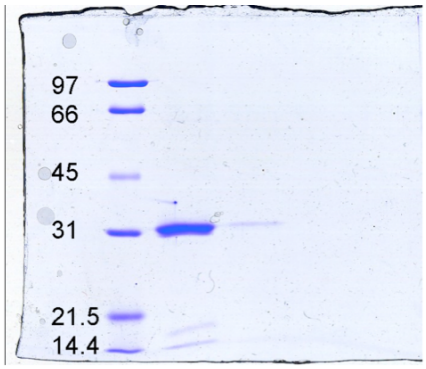


Figure 2.2: A typical 12% acrylamide SDS-PAGE gel showing purified M^{254R}.
By far the largest band is at the expected mass of around 30 kDa. The lane on the left contains markers at the masses shown.

2.1.1 Analysis by mass spectrometry

All protein mass spectrometry was carried out in the laboratory of Dr. Jackie Mosely, using a mixture of electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation (MALDI) techniques, which are accurate to greater than 1 Da. M^{254R} samples consistently produced molecular ions with masses of 29350 and 29430 Da, which are 100 and 180 Da higher than the theoretical mass of the protein based on the amino acid sequence. Initial ideas to account for the extra mass included possible binding of one or more Ca^{2+} or Zn^{2+} ions by the protein, which would add an extra 65.39 or 40.078 Da respectively to the mass. Many proteins contain such metal ions either for catalytic purposes, or to hold in place areas of specific tertiary structure. There is one mention in the literature[3] of a possible zinc finger domain in the RSV matrix protein, although no evidence was presented for this.

To test for the presence of bound metals, a small sample of M^{254R} was mixed with an equal volume of dilute pH 3 HCl, which would form a salt with any metal ions present and remove them from the protein. The solution was dialysed against water and then analysed by ESI mass spectrometry. No change in protein mass was observed, suggesting no metal ions had been bound in the original sample.

Protein mass spectrometry often gives results that differ slightly from the expected mass. Proteins can easily pick up contaminants in solution, for example polymers

such as PEG from plastic pipettes and sample containers, although the presence of a polymer is usually obvious from a mass spectrum as the repeating unit appears prominently. Alternatively, additional phosphate groups would add 78.97 Da each to the mass, although this is unlikely to be the cause of the mass increase since phosphorylation of proteins does not occur in bacterial cells such as the *E. coli* expression vector. Subsequent solution of the crystal structure of M conclusively showed that no bound metals or any other additional groups are present (see section 2.2.1). To exclude the possibility that the crystallization conditions had resulted in loss of a bound metal or other group, protein samples that had been incubated with the Tacsimate (Hampton Research) mixed salt solution used in crystallization (1.8305 M malonic acid, 0.25 M ammonium citrate tribasic, 0.12 M succinic acid, 0.3 M DL-malic acid, 0.4 M sodium acetate trihydrate, 0.5 M sodium formate, 0.16 M ammonium tartrate dibasic, pH adjusted with NaOH) were analysed by mass spectrometry. This showed no change in mass following the incubation.

Mass spectrometry also showed that M^{254R} frequently produced a fragment with a mass of 13810 Da, which corresponds to the mass of residues 136-256 at the C-terminus of the protein. Two bands at around 13-15 kDa were also frequently seen in SDS-PAGE analysis of protein samples that had been stored for long periods of time (Figure 2.3). These data suggest that M^{254R} is particularly susceptible to hydrolysis between residues Lys-135 and Thr-136, perhaps because this region is especially exposed in the folded protein structure. This site could represent a position where reaction or cleavage occurs *in vivo* to activate a specific aspect of matrix protein function.

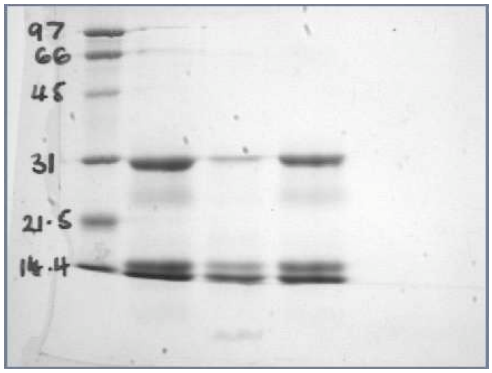


Figure 2.3: 12% SDS-PAGE gel showing several samples of M^{254R} with bands at 13-15 kDa.

2.2 Crystal structure

Several samples of matrix protein were provided to our collaborator Dr. Victoria Money. She subjected the samples to nanoscale crystallization trials, and was able to produce small needle-shaped crystals of M^{254R} (Figure 2.4) in a solution of 70% Tacsimate (Hampton research) at pH 7.0, conditions that were later refined to 55-65% Tacsimate at pH 7.0.

Figure 2.4: Example of M^{254R} crystals formed in 70% Tacsimate, pH 7.0

Dr. Money was able to obtain high quality X-ray crystallographic data from these crystals, which allowed the crystal structure of M^{254R} to be solved to 1.6 Å resolution, using MIRAS phasing techniques. This represents the first example of an intact matrix protein from the *Mononegavirales*.[4] The structure of native protein M^{254M} has also been solved, and found to be identical to that of M^{254R} described here.

The MIRAS method involves comparison of protein X-ray data with data from metal-derivitized protein crystals, and exploiting the differences between the two sets of information to calculate the structure. Often selenomethionine-substituted proteins are used, but in the case of M^{254R}, replacing sufficient methionine residues with selenomethionine (Se-Met), by growing cells expressing M^{254R} in media containing Se-Met, proved difficult. Some data from the crystallized Se-Met protein were eventually employed to solve the structure, supplemented by data from

mercury-derivitized protein crystals produced by soaking some of the original crystals in mercury.

X-ray crystallographic data were collected for the native crystals at the European Synchrotron Radiation Facility (ESRF) in Grenoble, and for the derivitized crystals at the Stanford Synchrotron Radiation Lightsource (SSRL), part of the SLAC National Accelerator Laboratory at Stanford University.

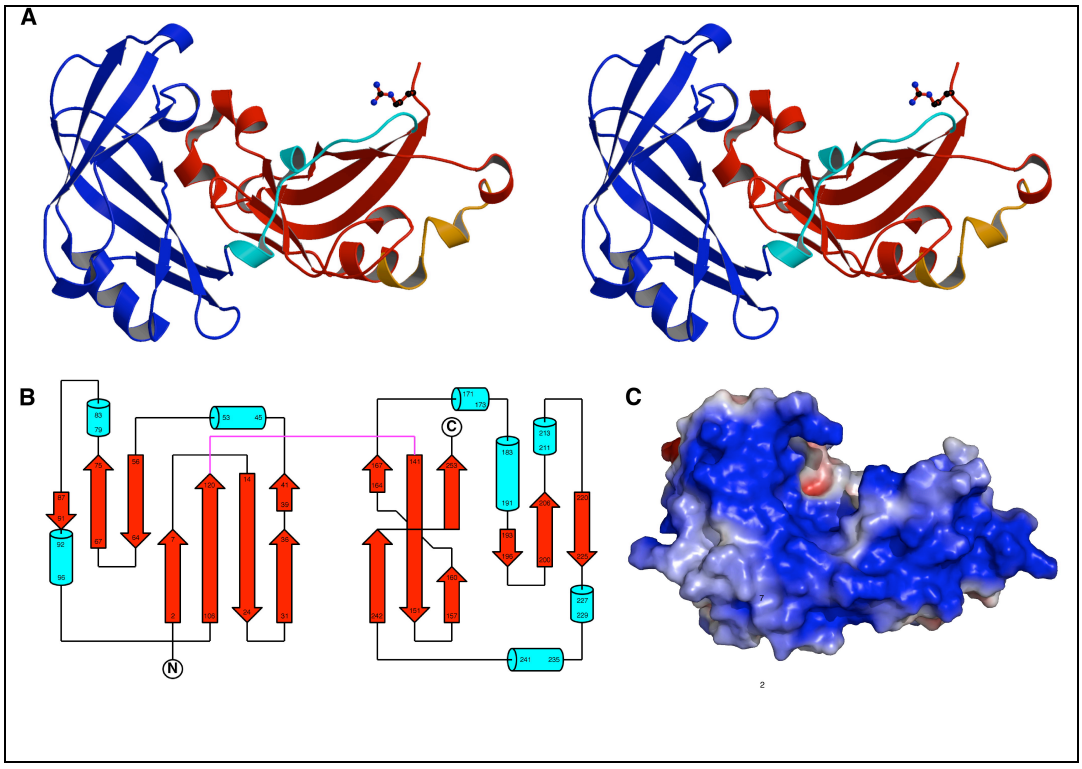


Figure 2.5: The three-dimensional structure of M^{254R}
A: Divergent stereo image, with the N-terminal domain in blue, the linker region in cyan and the C-terminal domain in red. Residue 254 is highlighted in ball-and-stick representation. B: Topology diagram, with the linker shown in magenta. β-sheets are represented by arrows, and helices by cylinders. C: Electrostatic surface potential (calculated using Adaptive Poisson-Boltzmann Solver software in PyMol); the range red to blue represents -5 to +5 kT/e, and uncharged residues are not coloured.[4]

The crystallized form of M^{254R} is monomeric, although there is evidence that oligomeric forms exist in solution (see Section 2.2.3 below). The overall structure of the protein consists of two clearly defined domains, joined by a 13-residue linker region. The N-terminal domain encompasses residues 1-126, while the C-terminal domain comprises residues 140-255 (Figure 2.5 A and B). All the residues were clearly resolved within the electron density, except for residues 99 and 100.

The N-terminal domain is a twisted β-sandwich made up of two β-sheets of 3 and 4 strands respectively, which are almost perpendicular to each other. The domain is horseshoe-shaped overall, with the first β-sheet forming the concave inner face and the second the convex outer face. The C-terminal domain is a flattened β-barrel comprising two 3-stranded anti-parallel β-sheets. The linking regions between strands 2 and 3, and 5 and 6, are predominantly helical. The 254R substitution is at the very end of the C-terminal domain where there is little secondary structure, and so is not likely to have a great influence on the structure or function of the protein.

Another notable feature of this structure is the large area (approximately 600 $Å^2$) of positively charged residues extending across both domains (Figure 2.5, C). This is likely to be important for the structural role of RSV M, which requires it to be able to form protein-protein and protein-lipid interactions. The binding of matrix proteins to membranes is thought to be mediated at least partially by electrostatic interactions,[5-7] and this positively charged region provides a mechanism by which it could bind to negatively charged lipids. A similar positively charged patch features in all the other matrix proteins for which structural information is available (see Chapter 1, Section 1.4.3). Another full length *Mononegavirales* matrix protein, that of Borna disease virus, was published very recently; this also exhibits a large positively charged patch on one face, which is suggested to be responsible for membrane binding.[8] Other distinctive areas of surface charge on the RSV M structure were a negatively charged area on the N-terminal domain and a positively charged pocket on the C-terminal domain, which could represent potential sites of interaction with binding partners, or for protein-protein interaction to take place. The charged patch is partially overlapped by a striking planar arrangement of six tyrosine residues that encircles most of the C-terminal domain. Tyrosine associates strongly with phospholipids in model systems,[9] and is often found in the interfacial region of membrane proteins;[10, 11] it is thus possible that this array of residues is important for binding.

The linker region between the two domains has no secondary structure except for a very short helical region. The domains are associated only *via* hydrophobic interactions and a few water-mediated hydrogen bonds, which, combined with the unstructured nature of the linker region, means that the domains may be able to

occupy different relative orientations. RSV M^{254R} undergoes proteolysis when stored in solution, as described in Section 2.1.1 above, and mass spectrometry of the fragmented samples showed that the cleavage site is between residues Lys-135 and Thr-136 in the linker region. Once such cleavage has occurred, the weak interdomain interactions are not sufficient to hold the two domains together. This raises the possibility that the protein may exist in alternative conformations in solution, the inter-domain packing in the crystal being metastable and driven by the behaviour of the hydrophobic interfacial residues.

The C-terminal domain of M^{254R} has several surface-exposed arginine and lysine residues that could contribute to interaction of this part of the protein with negatively charged membranes. If membrane binding were mediated by the C-terminal domain, largely through electrostatic interactions, the N-terminal domain would be free to perform more species-specific functions such as controlling the protein-protein interactions needed for virion assembly. This would be consistent with work done on Ebola VP40, which has shown that the N-terminal domain is sufficient for protein oligomerization, while the C-terminal domain must be present for membrane interaction to occur.[5]

2.2.1 Comparison with other matrix protein structures

As discussed at the outset of this work, four other matrix protein structures have been published. Three of these, VSV M, Ebola VP40 and BDV M, are from the *Mononegavirales*, although in the cases of VSV M and VP40 a proteolytically resistant core was crystallized rather than the full-length protein. The existing structures do not bear a great resemblance to each other at sequence or folding levels, but the new RSV M structure is broadly similar to Ebola VP40, which also consists of two domains joined by a potentially flexible linker (which was not resolved in the published structure). This structure also shares with the others regions of positive charge on the surface that are likely to be important in membrane binding.

It is possible to overlay the N- and C-terminal domains of M^{254R} onto the corresponding VP40 domains (Figure 2.6). This shows that although the overall fold

of the two proteins is closely related, the arrangement of β-sheets in each one differs; the N-terminal domain of M^{254R} contains one mixed 4-stranded and one anti-parallel 3-stranded β-sheet, compared to the two 3-stranded sheets in the same domain of VP40. Such a comparison proved impossible with VSV M, underlining the structural diversity among matrix proteins.

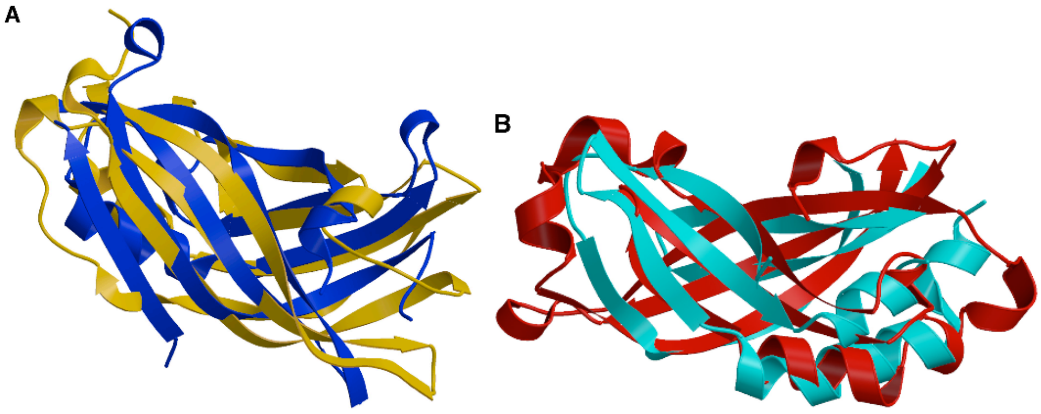


Figure 2.6: Comparison of the M^{254R} topology with that of Ebola VP40.
A: M^{254R} N-terminal domain in blue and VP40 in yellow; B: M^{254R} C-terminal domain in red and VP40 in cyan.[4]

2.3 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is a form of UV spectroscopy that is often used to probe the secondary structure of proteins.[12] The polypeptide backbone of a protein is chiral, so will selectively absorb the left- or right-handed circularly polarized components of polarized light. The absorption of proteins in the far-UV region is mainly due to the $n \rightarrow \pi^*$ (around 220 nm) and $\pi \rightarrow \pi^*$ (around 190 nm) transitions of the amide groups present, which are influenced by the geometry of the polypeptide backbone; thus the spectrum produced reflects the different types of secondary structure in the protein.[13]

The circular dichroism measured is the difference in extinction coefficients between the right- and left-handed circularly polarized light, in units of cm/M:[14]

$$\Delta\varepsilon = (\varepsilon_L - \varepsilon_R) \qquad (1)$$

and

$$\Delta\varepsilon = \frac{(A_L - A_R)}{cl} \quad (2)$$

where c = sample concentration and l = path length.

This is often reported in terms of ellipticity, θ, where

$$\theta = 33(A_L - A_R) \quad (3)$$

or molar ellipticity:

$$[\theta] = \frac{100\theta}{cl} \quad (4)$$

To convert spectrometer readings in units of millidegrees to molar ellipticity, the following equation was used:

$$[\Theta]_{MRW} = \frac{0.1 \times \theta_{obs}}{n \times c \times l} \quad (5)$$

where θ_{obs} is the observed ellipticity, *i.e.* the output from the spectrometer in millidegrees; n is the number of amino acid residues in the protein being studied; c is the concentration of the protein sample in M, and l is the pathlength of the sample cell in cm.

The degree of ellipticity at different wavelengths of UV light depends on the secondary structure of the protein; α-helices, β-sheets and random coils all produce characteristic CD spectra, as shown in Figure 2.7.

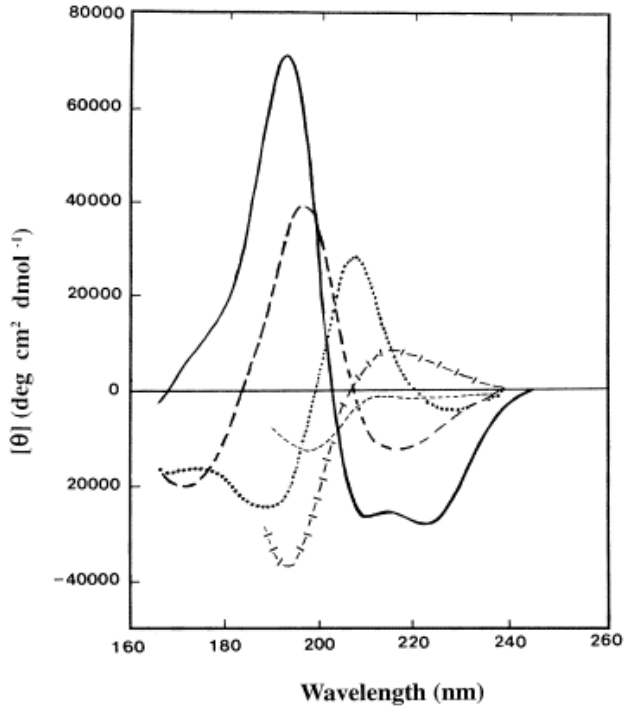


Figure 2.7: Characteristic CD spectra for various types of secondary structure.
Solid line, α-helix; long dashed line, anti-parallel β-sheet; dotted line, β-turn, cross dashed line, extended 3_{10}-helix; short dashed line, irregular structure.[12†]

A protein needs to be in its native folded conformation for analysis by CD spectroscopy, as this means the structure is fixed and the UV-absorbing groups are in asymmetric surroundings; in the denatured state the protein chain can move around freely so the signals from different areas will average out. Often, a protein must be kept in buffer solution to prevent it from unfolding, but the salts used frequently absorb UV light in the same region as proteins; for example, chloride ions absorb strongly below 200 nm.[12] Fortunately it was found that the RSV M protein is relatively stable in pure water, at least for a short time, so it was possible to measure its CD spectrum in the absence of any other species. Spectra were later measured in dilute phosphate buffer at pH 7.4 (which absorbs relatively weakly over the necessary wavelength range) to show that there were no major differences between data gathered in the absence and presence of pH control.

† Reprinted from S. M. Kelly, T. J. Jess and N. C. Price, *Biochimica et Biophysica Acta*, 2005, **1751**, 119-139, Copyright (2005), with permission from Elsevier

2.3.1 CD spectroscopy of M^{254R} in water

After dialysis of M^{254R} samples overnight against water, CD spectra were recorded between 350 nm and 240 nm using a 1 cm pathlength cell, and between 240 and 180 nm using a cell with a 2 mm pathlength. The shorter pathlength meant that the overall absorbance of the protein was kept below reasonable limits, to reduce noise, which can be a particular problem in CD spectroscopy at short wavelengths. The absorbance of a sample can be monitored whilst recording a spectrum by looking at the High Tension (HT) voltage trace, which displays the voltage applied to the photomultiplier in the spectropolarimeter. For this work, the maximum HT reading for usable data was taken to be 600 V. Any data points for which the HT value was above 600 V were disregarded.[12] In this case it was possible to measure the CD spectrum of M^{254R} (Figure 2.8) down as far as 182 nm before the HT reading became too high.

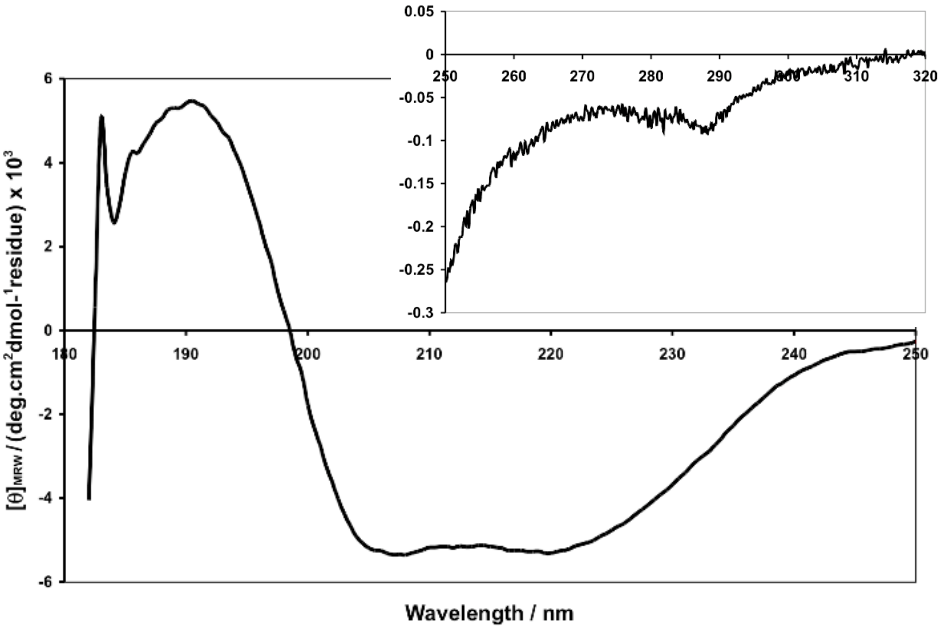


Figure 2.8: The CD spectrum of M^{254R} in water between 320 and 182 nm.
The main curve shows far UV data recorded using a 2 mm pathlength, while the inset box shows near UV data measured using a 1 cm pathlength.

The main feature of the spectrum is a large minimum between 198 and 240 nm, which by comparison with the spectra shown in Figure 2.7 is consistent with a protein containing a mixture of secondary structure types, although it is very difficult to say which if any structural features predominate simply by visual comparison of the spectrum with the model data. Also note the weaker peaks that can be seen

between 260 nm and 310 nm (inset box in Figure 2.8). Peaks in the region 260-320 nm arise from transitions in the aromatic amino acid side chains, with fine structure due to the involvement of different vibrational levels of the excited state. The largest peak in this region is close to 290 nm and probably represents tryptophan, as does the very small peak at about 305 nm. The weak feature closer to 280 nm is most likely due to tyrosine. Phenylalanine usually results in weak bands between 255 and 270 nm, which cannot be discerned here.[12]

2.3.1.1 Background to CD analysis

Further analysis of the spectrum shown in Figure 2.8 was undertaken in order to more accurately determine the secondary structure of M^{254R} in solution. Several empirical algorithms have been developed that provide an estimation of the secondary structure composition of a protein from its CD spectrum. These analyse the experimental CD data in relation to a basis dataset comprising a selection of proteins of known structure, in order to deconvolute the contributions to the spectrum from the various types of secondary structure present. The most widely used algorithms can be accessed *via* the Dichroweb online resource run by Birkbeck College,[13, 15, 16] or by downloading the software package CDPro.[17] Dichroweb uses CDSSTR, CONTIN/LL, SELCON3 and K2D, whilst CDPro includes the same methods with the exception of K2D.

All the analysis algorithms mentioned above rely on the assumption that the contributions from the different structural components are independent, and can be added linearly to produce the net spectrum C_λ,[13, 17] *i.e.*:

$$C_\lambda = \Sigma f_k B_{k\lambda} \qquad (6)$$

where f_k is the fraction of secondary structure k, and $B_{k\lambda}$ is the spectrum of the secondary structure component.[17]

The various analysis programs utilize different methods to analyse CD data. CDSSTR uses a variable selection method, in which an initial large database of

standard protein spectra is reduced to several smaller randomly selected sets of data (eight proteins) on which singular value decomposition (SVD) is performed. SVD extracts unique basis curves from a set of spectra from proteins of known structure, each curve being characterized by a particular mixture of secondary structures, and these are then used to analyse the conformation of the unknown protein. In CDSSTR, this selects the data sets fulfilling the criteria for a good fit (sum of fractions is between 0.95 and 1.05, each fraction is greater than -0.03, root mean square [RMS] between reconstructed and experimental data < 0.25 $\Delta\varepsilon$), which are then averaged. SVD gives the best estimates for α-helix content, but is less accurate for β-sheets and turns unless data are collected down to 184 nm. CDSSTR is thought to provide the best fits of CD data from globular proteins.[18, 19]

CONTIN uses the ridge regression method, fitting the spectrum of an unknown protein by comparing it to a linear combination of the spectra from a large database of proteins of known conformations. The contribution from each reference spectrum is kept small unless it contributes to a good agreement between the theoretical best-fit curve and the raw data. CONTIN gives good estimates of α-helices and β-sheets; different references are used each time, which helps in obtaining the best possible fit to the experimental data.[18] However, the program does not always give the best answer, and can sometimes reject solutions that are actually closer to the X-ray structure of a test protein than the solution it selects; the solution with the lowest standard error is often found to be the closest. Thus CONTIN/LL incorporates the variable selection method (as in CDSSTR) into the locally linearized model in CONTIN; the proteins where the RMS value is furthest from the spectrum being analysed are systematically deleted to create smaller reference sets. This gives a solution for each set, and the final answer is an average of all the solutions that fulfil the CONTIN selection rules.[19]

SELCON is the self-consistent method, in which the experimental spectrum is included in the basis set and an initial approximation of the unknown structure is made. The resulting matrix equation is solved using SVD, and the solution then replaces the initial approximation. This process is repeated until a self-consistent answer results. In SELCON3 (provided in CDPro), the α-helix and β-sheet

conformations are divided into regular and distorted fractions. SELCON provides very good estimates of the secondary structure content of globular proteins, but in SELCON3 the estimates of turns are poorer than in the other versions.[18]

K2D is a neural network program, where artificial intelligence is used to find correlations in data. The neural network is first trained using a set of known proteins until the CD input at each wavelength results in the output of the correct secondary structure, then the trained network is used to analyse the unknown CD data. K2D gives good estimates of helical and sheet contents of proteins and polypeptides, but does not estimate turns.[18]

The choice of protein reference set is also important. Dichroweb utilizes 8 different reference sets, whilst CDPro contains 10. Each set covers a specific wavelength range and contains a different number of reference proteins; normally the set with the largest number of proteins covering the appropriate wavelength range would be selected. It was fortunate that accurate CD data for M^{254R} in water was recorded down to 182 nm, as collecting data at wavelengths shorter than 190 nm significantly increases confidence in any structural information derived from analysis of the spectrum by comparison with data from reference proteins.[18]

2.3.1.2 CD analysis of M^{254R}

Analysis of the M^{254R} CD spectrum was carried out using both Dichroweb and CDPro, although only the data from CDPro are presented here, as it utilised more appropriate reference datasets. Most of the reference proteins generally used are globular, and may not provide an accurate analysis when used to deconvolute CD spectra from different types of proteins. It has been shown that the use of an extended reference set including membrane proteins improves the performance of the analysis overall,[17] and this might be expected to be particularly true for the analysis of M^{254R}, which is a peripheral membrane protein. CDPro datasets 4 and 10 (referred to as SP43 and SMP56 in reference 10) were used, where set 10 (SMP56) contained the same cytosolic reference proteins as were in set 4 (SP43) with the addition of 13 membrane proteins.[17]

The results from CDPro are shown in Figure 2.9 below. All three available algorithms were used, but only SELCON3 and CONTIN/LL gave meaningful results, with CONTIN/LL resulting in a significantly better fit.

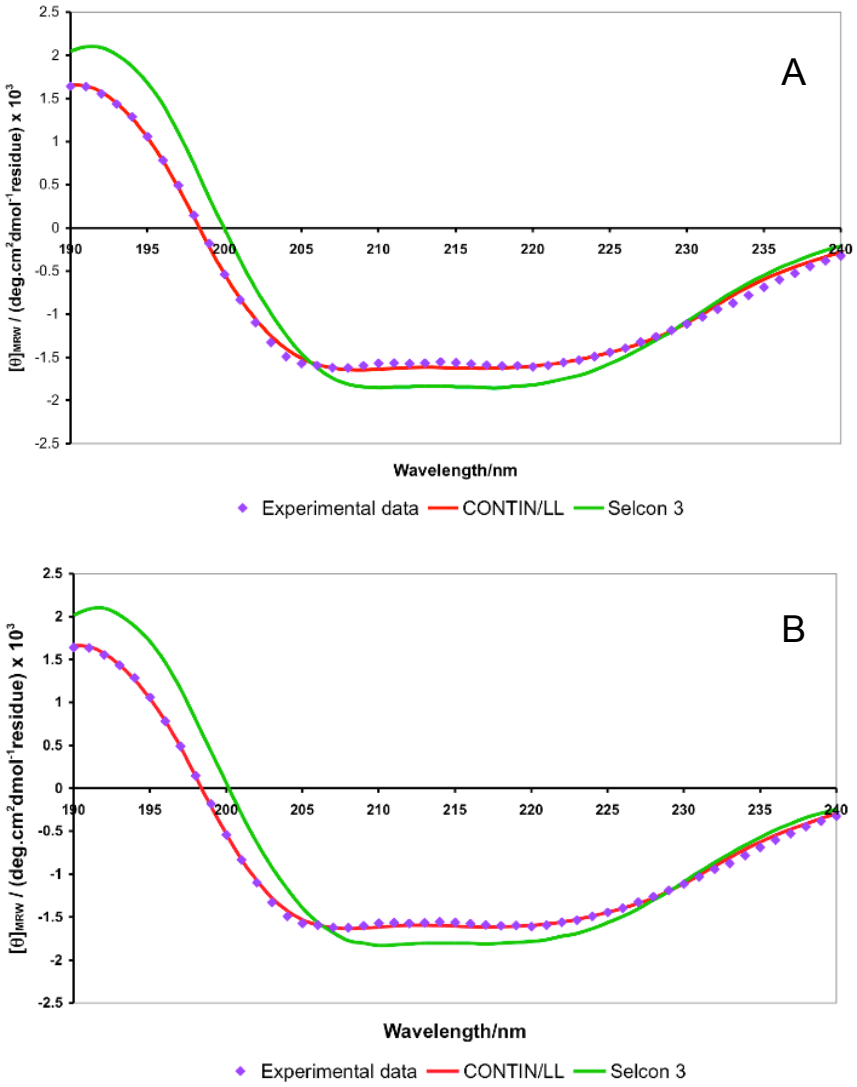


Figure 2.9: CD spectrum of M^{254R} in water (190-240 nm) fitted using the CONTIN/LL and Selcon3 algorithms in the CDPro package.
A: Using data set 4 (SP43); B: Using data set 10 (SMP56).

It may be seen from the chart that the two basis data sets give very similar results, and in both cases the spectrum calculated by CONTIN/LL appears to be a very good fit to the measured spectrum. The NRMSD (normalized root mean square deviation) can be a useful measure of how well the experimental data are matched by the theoretical CD spectrum calculated from the derived secondary structure composition. NRMSD is defined by,

$$\left[\frac{\Sigma(\theta_{\exp}-\theta_{calc})^2}{\Sigma(\theta_{\exp})^2}\right]^{1/2} \quad (7)$$

where θ_{exp} is the experimental molar ellipticity at a given wavelength and θ_{calc} is the calculated value. 0 is a perfect fit and 1 means there is no fit at all. Normally, an NRMSD value over 0.25 indicates that there has been an error in the analysis, while a value of 0.05 is ideal.[12] The values calculated here using the CONTIN/LL algorithm and dataset SMP56 gave an NRMSD of 0.020, which suggests a good fit. However, it must be borne in mind that out of the three possible algorithms used in CDPro, CONTIN/LL always gives the lowest NRMSD because of the way it is set up to minimize the difference between the experimental and calculated data.[17]

The secondary structure fractions from this analysis are shown in Table 2.2.

	a-Helix	3_{10}-Helix	Strand	Turn	Irregular[†]
% of total	3	4	32	16	45

Table 2:2 Secondary structure fractions resulting from analysis of the CD spectrum of M^{254R} in water using the CONTIN/LL algorithm with dataset SMP56.

2.3.2 CD spectroscopy of M^{254R} in buffer

Although the data resulting from CD spectroscopy of M^{254R} in water were excellent, the measurements were repeated in buffer solution to ensure that any changes in pH were not affecting protein folding, and hence the spectrum.

Many commonly used buffer solutions absorb significantly in the far UV region that is of most interest in protein CD spectroscopy.[12] Since $H_2PO_4^-$ is one of the less problematic species in this respect, and most of the other work on RSV M had been carried out in PBS pH 7.4, phosphate buffer was used for this CD work. A solution of NaH_2PO_4 pH 7.4 was made up and run as a blank in the CD spectrometer, and

[†] Secondary structure that does not fit into the other categories was originally termed 'random coil', but it is now more often referred to as irregular or intrinsically disordered, since it is neither random, nor coil, but covers a range of structures (which can be well-defined) with ϕ and ψ angles that do not match those adopted by helices, sheets and turns.[13]

progressively diluted until a relatively flat spectrum was obtained, at a concentration of 5 mM. A sample of M^{254R} was then dialysed against this solution and used to carry out CD spectroscopy, as described above for protein in water. The resulting spectrum is shown in Figure 2.10.

These data were then analysed over the 190-240 nm wavelength range using the CDPro package as before, and the same two datasets (4 and 10). The results are shown in Figure 2.11 below. The two sets of data gave very similar results in terms of the closeness of fit with the experimental data, and also in the calculated secondary structure fractions returned. CONTIN/LL again gave the lowest NRMSD value for both basis sets (0.023 for SMP56 and 0.019 for SP43). The secondary structure fractions resulting from using the CONTIN/LL algorithm with set SMP56 are reported in Table 2.3 below; as discussed above, it is not only the NRMSD that should be taken into account when considering the accuracy of the analysis, and since SMP56 includes membrane proteins it should produce the more accurate data.

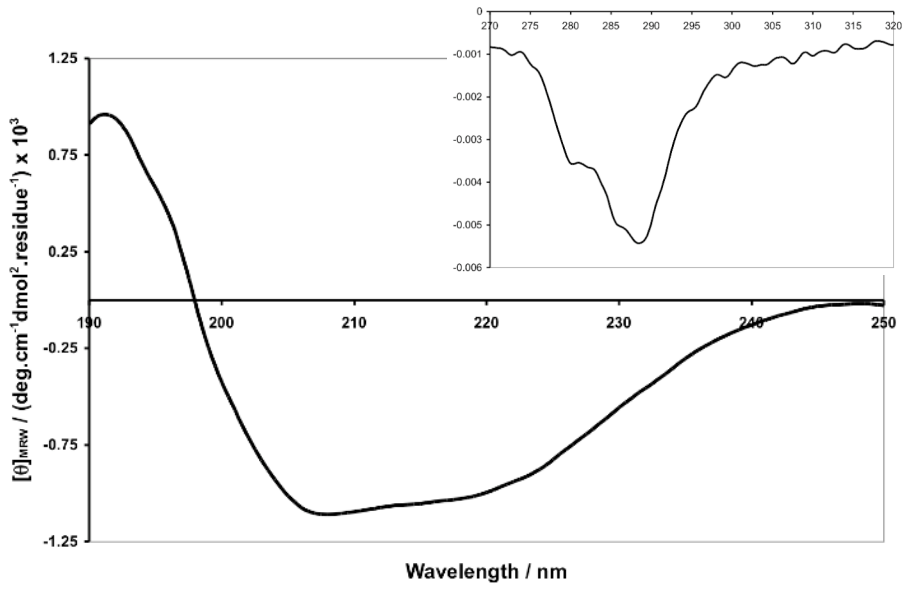


Figure 2.10: The CD spectrum of M^{254R} in 5 mM phosphate pH 7.4
The main plot shows far UV data recorded using a 2 mm pathlength, while the near UV data shown in the inset box were measured using a 1 cm pathlength.

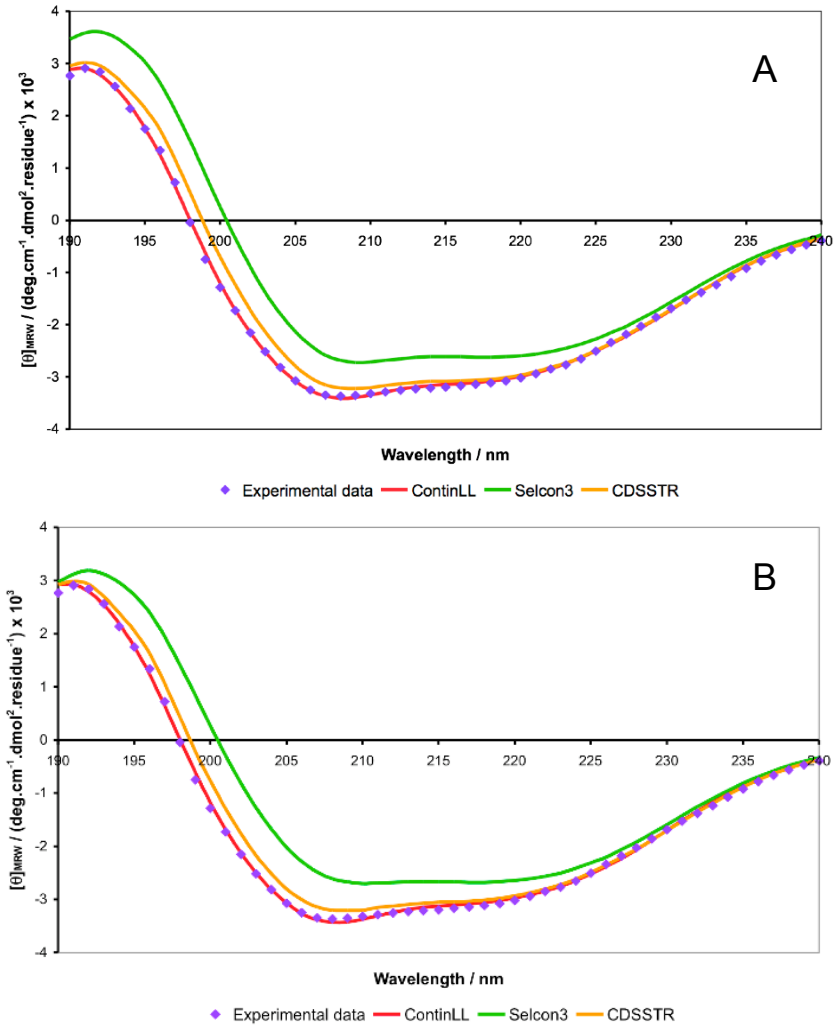


Figure 2.11: CD spectrum of M^{254R} in 5 mM phosphate pH 7.4 fitted using CDPro. A: Using dataset 4 (SP43); B: Using dataset 10 (SMP56)

	α-Helix	3_{10}-Helix	Strand	Turn	Irregular
% of total	12	12	22	22	32

Table 2:3 Secondary structure fractions resulting from analysis of the CD spectrum of M^{254R} in 5 mM phosphate pH 7.4 using the CONTIN/LL algorithm with dataset SMP56.

2.3.3 CD spectroscopy of M^{254M}

The native protein, M^{254M}, was subjected to CD analysis to verify that its spectrum was not significantly different to that of M^{254R}. The data are represented in Figure 2.12 below, with the results of the analysis using CDPro shown in Figure 2.13. Again the two possible basis data sets gave extremely similar results, but the best fit (NRMSD 0.023) was for CONTIN/LL using set 10 (SMP56). The secondary structure fractions from this are shown in Table 2.4.

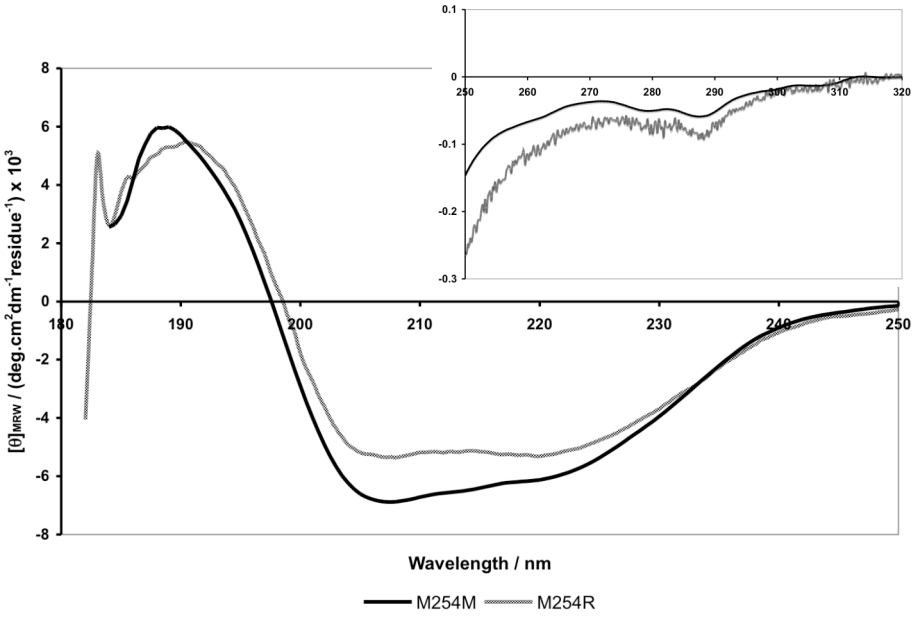


Figure 2.12: CD spectrum of M^{254M} in water, with the spectrum of M^{254R} for comparison. Inset box shows the 250-320 nm region.

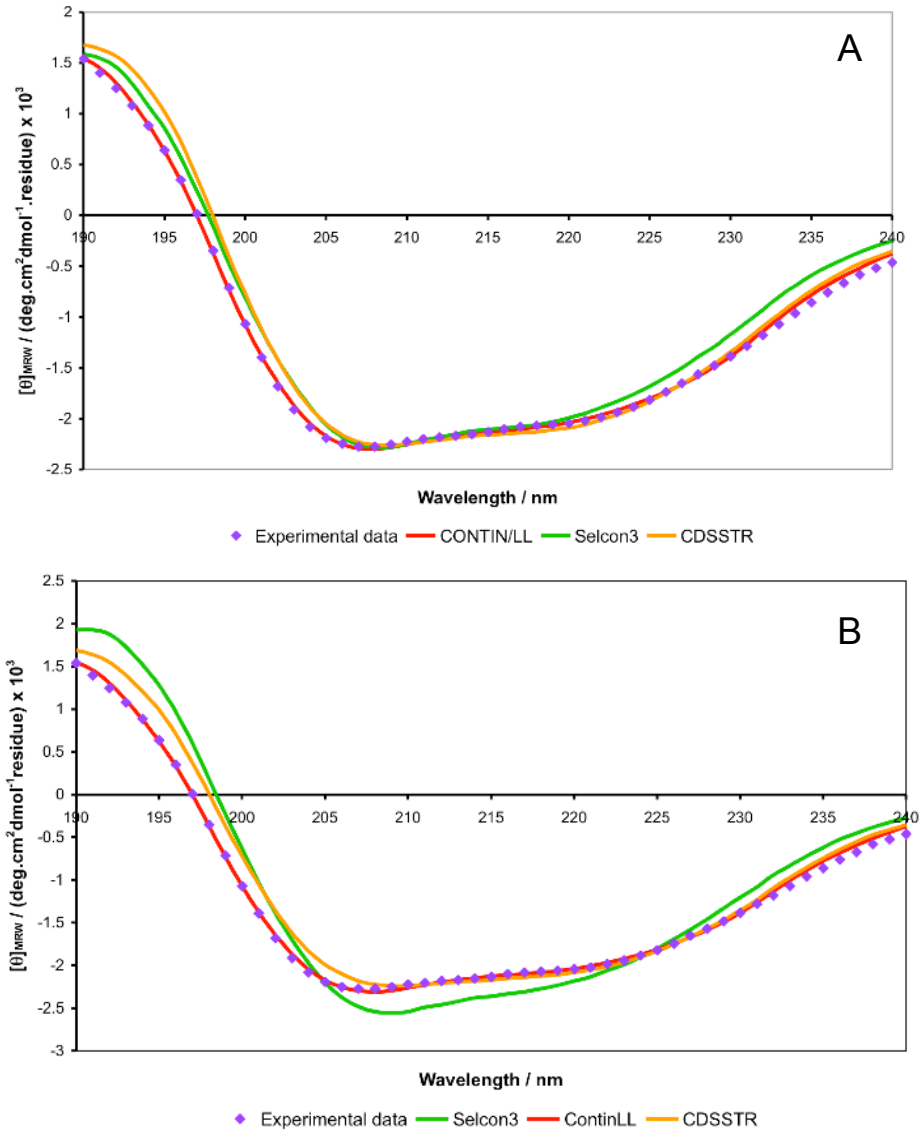


Figure 2.13: CD spectrum of M^{254M} in water fitted using CDPro.
A: Data set 4 (SP43); B: Data set 10 (SMP56).

	α-Helix	3_{10}-Helix	Strand	Turn	Irregular
% of total	8	10	27	24	31

Table 2.4: Secondary structure fractions resulting from analysis of the CD spectrum of M^{254M} in water using the CONTIN/LL algorithm with dataset SMP56.

2.4 Comparison of X-ray and CD structural data

The secondary structure fractions determined from the crystal structure (using the program Stride[20]) are shown in the following table, along with the values for M^{254R} and M^{254M} calculated from the best-fit circular dichroism data.

% of total	α-Helix	3_{10}-Helix	Strand	Turn	Irregular	Bridge
M^{254R} X-ray (using Stride)	12	8	47	21	12	1
M^{254R} CD in water	3	4	32	16	45	-
M^{254R} CD in buffer	12	12	22	22	32	-
M^{254M} CD in water	8	10	27	24	31	-

Table 2:5 The secondary structure content of M^{254R} using data from CD spectroscopy and X-ray crystallography.
All CD data are derived from the CONTIN/LL algorithm using basis set SMP56. Note that the Stride (structural identification) program assumes there is a disulphide bridge between 2 Cys residues, when in fact M^{254R} does not contain any such bridges.

The crystal structure of a protein is generally similar to the form the protein takes in solution, although where an equilibrium exists between several conformations of the protein in solution, crystallization may select out only one particular conformation.[14] CD spectroscopy will not reveal the presence of more than one protein structure in solution, since the spectrum produced is an average reading over the whole sample.

For all samples, RSV M appears to contain somewhat more regular secondary structure when in its crystalline form than it possesses in solution, which might imply that protein assembly, and/or membrane binding, incurs a loss of entropy arising from protein folding. This is seen in the case of intrinsically disordered proteins that adopt a well-defined functional conformation upon interaction with their substrates. In these systems, significant enthalpic contributions from electrostatic interactions offset the negative entropy to result in a favourable free energy of binding, often through having relatively large interaction surface areas, and many contacts between protein and substrate.[21, 22] Such thermodynamics may play a part in controlling the assembly of Ebola VP40 octamers, which is driven by RNA binding.[23]

Such a requirement for large areas of interaction between protein and binding partner is consistent with the surface features revealed by the crystal structure of M^{254R} described above. In particular, the extensive positively charged area spanning both domains potentially provides a large area over which protein-protein and protein-membrane interactions could occur.[4]

2.5 Cross-linking and mass spectrometry

A fragment of the Ebola virus matrix protein VP40 has been found to oligomerize in solution, forming hexamers and octamers; it has been suggested that these different oligomeric forms of the protein may perform distinct functions during virion assembly and release.[24] It was therefore of interest to determine whether RSV M also forms oligomers in solution, which might suggest that it operates in a similar way to VP40.

Samples of M^{254R} were reacted with glutaraldehyde over a range of absolute concentrations, and protein:glutaraldehyde molar ratios. At $3 \leq pH \geq 7\text{-}9$ the aldehyde groups react with the amine groups on lysine side chains in the protein, and because glutaraldehyde possesses two aldehydes, this often results in links forming between two lysine residues on adjacent protein molecules, joining the proteins together. The glutaraldehyde – lysine reaction (Figure 2.14), which forms a Schiff base containing imine bonds, is reversible, but adding sodium borohydride to the solution after a short time reduces the imines to amines and makes the links permanent. The $NaBH_4$ also reduces any remaining glutaraldehyde, quenching the reaction.

This type of cross-linking reaction will reveal whether any protein molecules are close together in solution, which would suggest that they are associated as oligomers. The mass of the cross-linked oligomers can be assessed in various ways, including gel filtration and electrophoresis, but in this case MALDI mass spectrometry was used.

The fact that the initial glutaraldehyde-lysine reaction is reversible means that there is an equilibrium between the initial reagents and the cross-linked products, the weighting of which can be affected by the conditions of the reaction. Consequently, the resulting cross-linked oligomers may be biased towards higher or lower masses depending on the position of equilibrium during the reaction. Le Chatelier's principle states that a system at equilibrium, when subjected to a change, will respond in a way that tends to minimize the effect of the change. In this case, a higher absolute

concentration of protein and/or glutaraldehyde, or a greater concentration of either reagent with respect to the other, will push the equilibrium over to the right, so that more cross-links are formed and higher order oligomers are produced. Quenching the reaction prevents such higher order oligomers from dissociating again, which further disrupts the multiple equilibria between protein and glutaraldehyde, also influencing the size of the eventual reaction products.

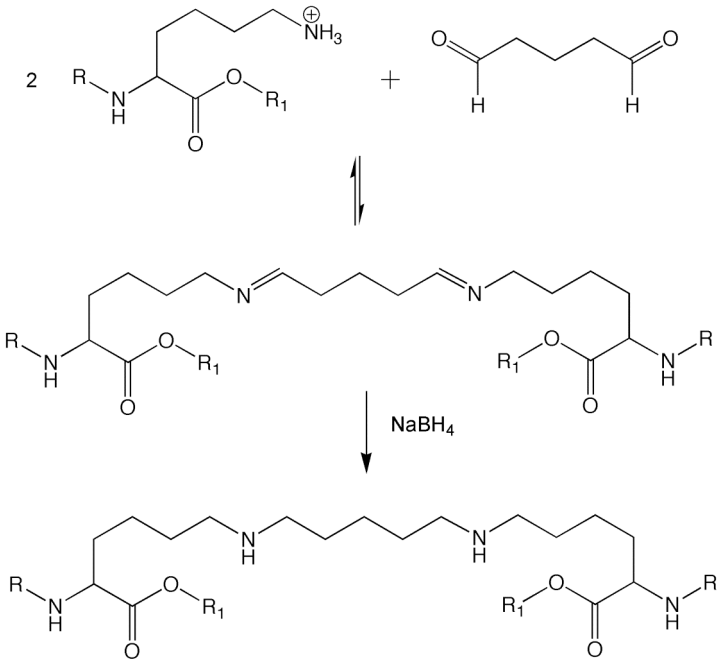


Figure 2.14: The reaction of glutaraldehyde with two lysine residues.

Figure 2.15 shows that the mass spectrometry detected a range of oligomers in the cross-linked samples. The mass of a M^{254R} monomer is 29250 Da, and various multiples of this monomeric mass were seen, representing dimers, tetramers and hexamers of the matrix protein. The exact mass of each oligomer varies between samples due to the different number of gluteraldehyde groups bound to each protein monomer.

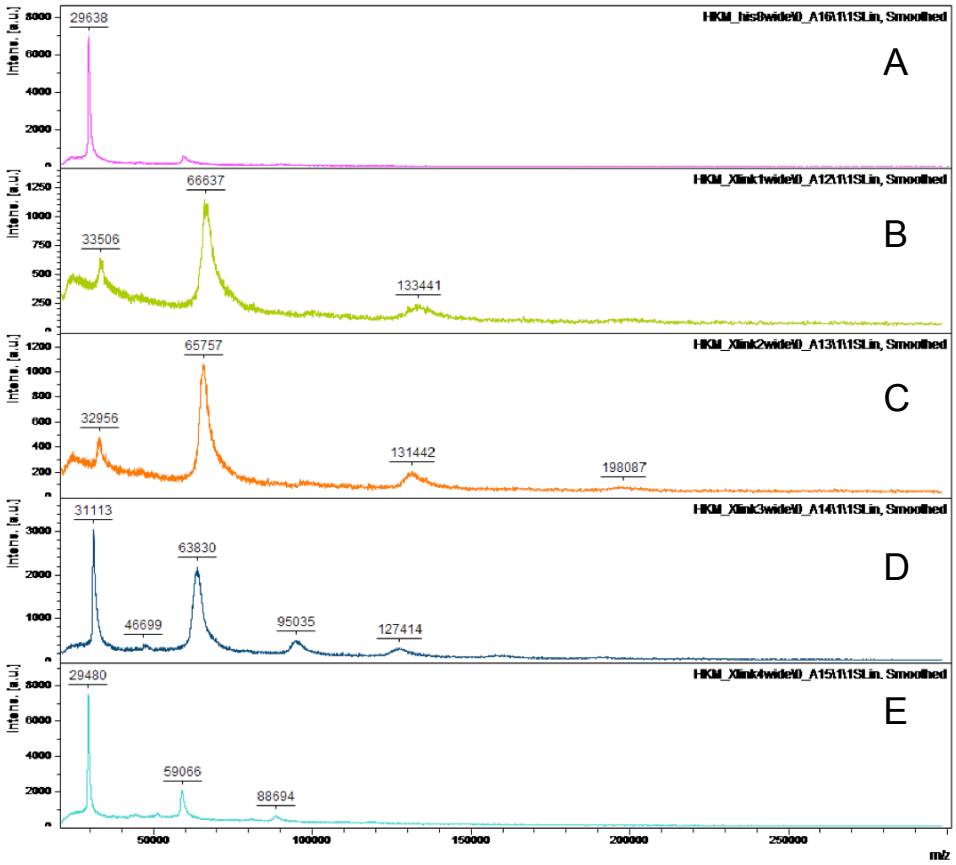


Figure 2.15: Mass spectra of samples from cross-linking experiments, after dialysis against water. The ratio of glutaraldehyde:protein increases from A to E.
A: M^{254R} alone; B: 14μM M^{254R}, 410 mM glutaraldehyde, 140 mM $NaBH_4$; C: 1.55 μM M^{254R}, 220 mM glutaraldehyde, 75.6 mM $NaBH_4$; D: 0.17 μM M^{254R}, 96 mM glutaraldehyde, 31.7 mM $NaBH_4$; E: 14 μM M^{254R}, no glutaraldehyde, 140 mM $NaBH_4$.

It should be noted that in this set of experiments, reactions containing higher ratios of gluteraldehyde also had a lower absolute concentration of reagents. Therefore in each sample the two factors affecting the equilibrium of the protein-glutaraldehyde reaction were acting against each other. In the sample containing only protein (Figure 2.15 A), monomers and some hydrogen-bonded dimers were apparent. In Figure 2.15 B, where the ratio of glutaraldehyde to protein was lowest, there were monomers, dimers and tetramers. In Figure 2.15 C, with a greater ratio of glutaraldehyde to protein, monomers, dimers, tetramers and hexamers were detected. In Figure 2.15 D, with a high glutaraldehyde:protein ratio, trimers and possibly pentamers were found, in addition to the oligomers composed of multiples of two protein molecules. The appearance of trimers and pentamers in Figure 2.15 D suggests that incomplete cross-linking occurred in this sample, probably because of the competing effects of having a high ratio of glutaraldehyde:protein, but a relatively low concentration of both reagents.

Overall, these results suggest that the RSV M protein preferentially self-associates in pairs of monomers; odd-numbered oligomers were not produced until the reaction conditions favoured incomplete cross-linking. It has been proposed that the tendency of matrix proteins to polymerize *in vitro* reflects their ability to self-assemble *in vivo* to form the protein scaffold needed for viral assembly.[24] These cross-linking data may therefore imply that dimers, and/or multiples of dimers, are important during M protein self-assembly when new virus particles are forming in the host cell. These findings are consistent with data from studies of the Ebola virus matrix protein VP40; it too has been found to form well-defined hexamers and octamers (although to form these it was necessary for the protein to bind a specific fragment of bacterial RNA) as well as dimers. [24, 25]

2.6 Protein assembly in the presence of lipids

Proteins, such as RSV M, that are associated with membranes *in vivo* are often difficult to characterize using standard crystallographic methods. The amphiphilic character of such proteins means that detergent is normally required for solubilisation, purification and crystallization, and it is frequently very challenging to grow well-ordered three-dimensional crystals suitable for analysis by X-ray diffraction.[26] Crystals formed in only two dimensions can now be studied by electron crystallography to give high resolution structural data; such crystals may be grown while allowing the protein to remain associated with lipid and detergents.[26, 27]

This type of two-dimensional crystallization strategy was initially adopted for the RSV M protein, before three-dimensional crystallization was found to be possible. M^{254R}His solution was mixed with lipid and detergent, before slow removal of the detergent by dialysis to leave a reconstituted artificial membrane hopefully containing crystallized protein. The lipid chosen for the first experiments was 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), as phosphocholine (PC) lipids are a major component of pulmonary surfactant, a lipid and protein complex secreted by epithelial cells in the lungs that stabilizes the alveoli at the air/water interface, and of lung cell plasma membranes.[28-30] 1,2-dioleoyl-*sn*-glycero-3-phosphocholine

(DOPC) and 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine (DPPE) mixtures were also used, as the small headgroup of DPPE means that it is more likely to form curved membranes. An increase in the degree of membrane curvature is implicated in the budding process of new virions.[31] The structures of these phospholipids, which were used extensively in this project, are shown in Figure 2.16.

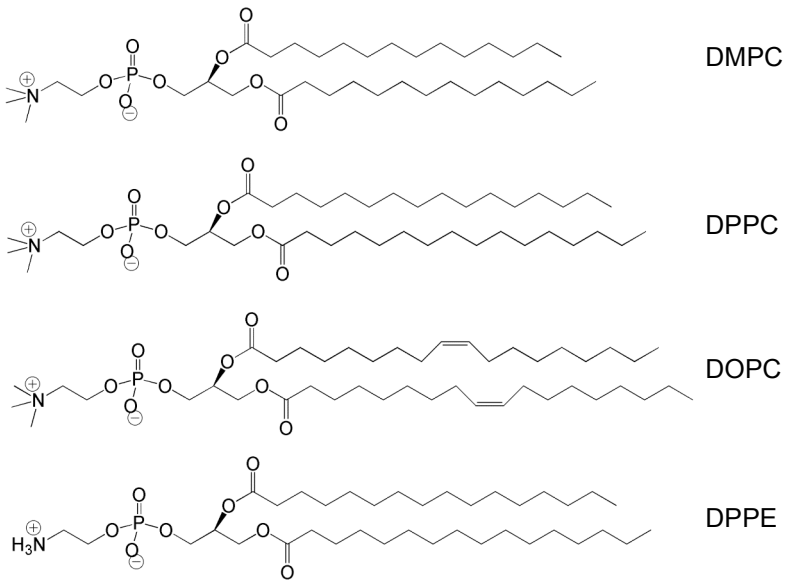


Figure 2.16: Some of the phospholipids used in this project.
Top to bottom: 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC); 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC); 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC); 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine (DPPE).

After detergent removal, samples were analysed by transmission electron microscopy (TEM) with negative staining, searching for areas of regular patterning that could represent 2D crystalline arrays. This process was repeated for several lipids (DMPC, DPPC, DPPE/DOPC mixture) and detergents (*n*-octyl-β-D-glucopyranoside, lauryldimethylamine oxide (LDAO) and Triton X-100) but no potential 2D crystals were observed.

In one particular sample however, other distinctive features were seen (Figure 2.17). These appeared to be extended structures, consistently around 30 nm in diameter and up to several microns in length. The dark line of stain down the centre of many of these structures shows that they are hollow. They also displayed a tendency to pair up or associate together in bundles. This sample contained no detergent, since it had been set up as a control, and hence did not actually undergo dialysis. This is a

particularly interesting result since it shows that M^{254R}His can form regular arrays with lipid without the need for detergent, or any other viral components.

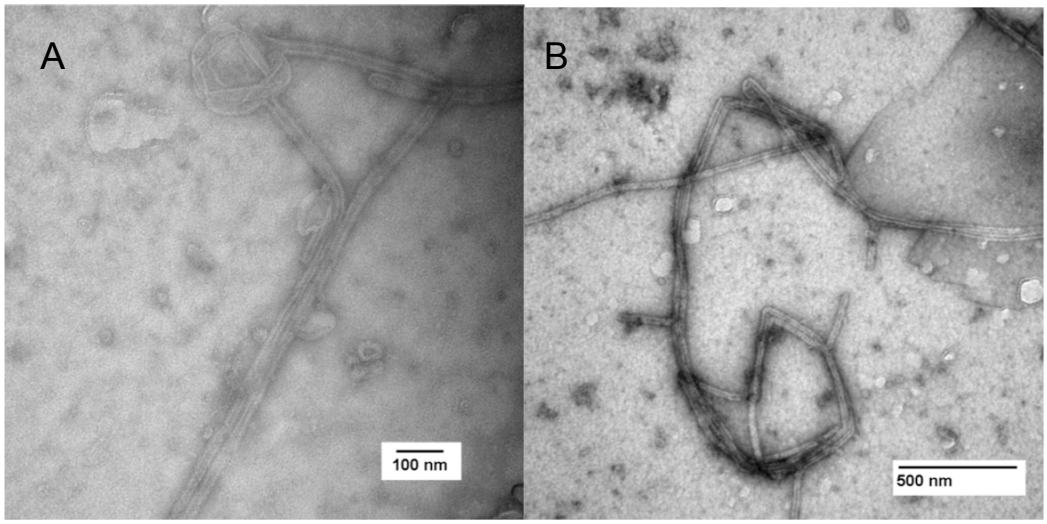


Figure 2.17: Extended hollow structures formed during incubation of M^{254R}His with lipid. DOPC/DPPC 8:2 (w/w), no detergent, lipid:protein ratio 1:15 (w/w), [protein] = 34.19, [lipid] = 106.21 μM

The sample also contained a significant quantity of other material such as excess protein and lipid, which obscured the areas of interest and made it difficult to obtain high-resolution images. A 50 – 0% sucrose gradient was therefore used to separate out the various components of the sample. Protein is much more dense than lipid, so excess protein equilibrated into the densest bottom layer of sucrose during centrifugation, while the lipid rose to the top of the gradient. The structures of interest contained both protein and lipid, and so were found in an intermediate position. 0.5 mL fractions were taken from the top of the gradient after ultra-centrifugation of the sample on the density gradient, and probed for protein content using a slot-blot and enhanced chemiluminescence detection.

Protein-containing fractions that were found towards the top of the gradient were dialysed against water, and re-examined by TEM. Eliminating unwanted components of the sample in this way allowed high-resolution images to be produced at the microscope's maximum magnification of 600,000 (Figure 2.18). These micrographs clearly show helical arrays of diameter 29 nm and pitch 33 nm.

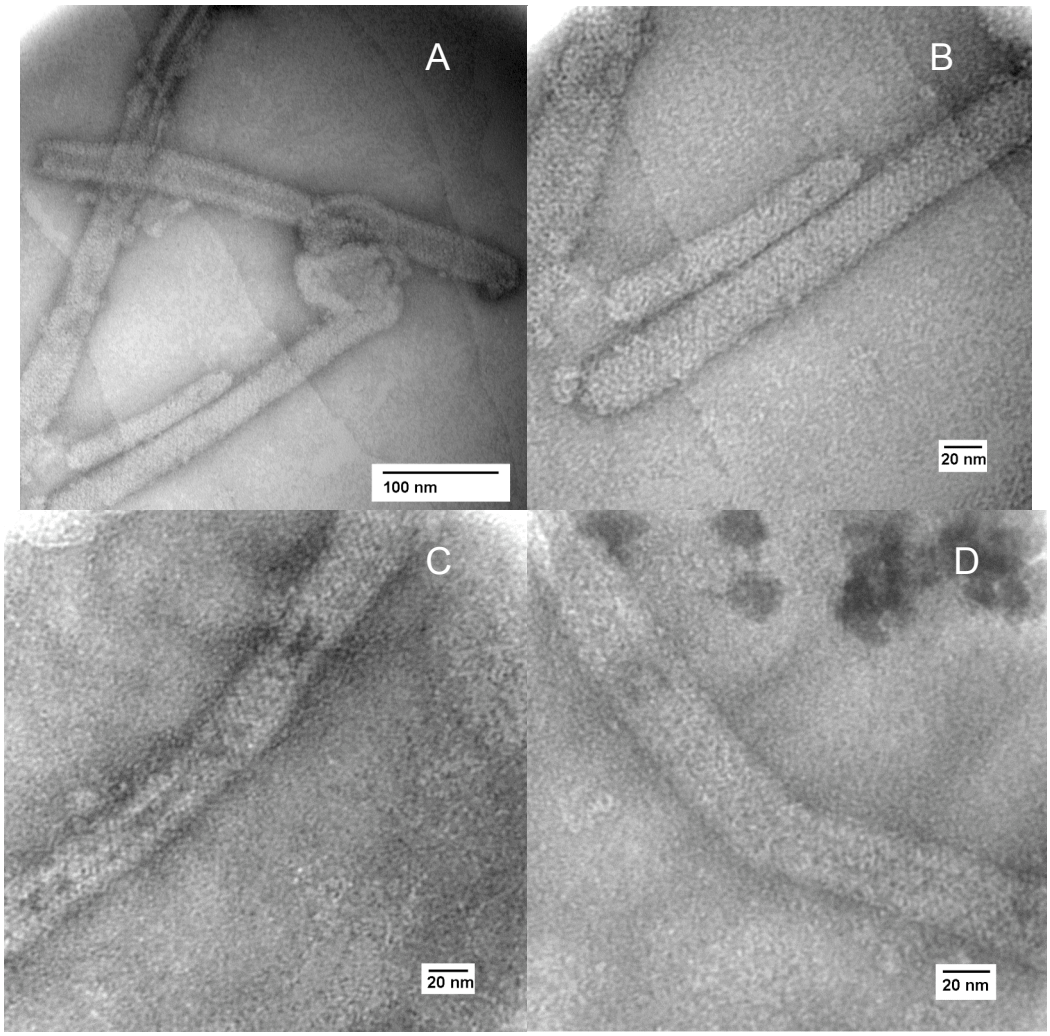


Figure 2.18: Extended helical structures isolated by sucrose gradient centrifugation. DOPC/DPPC 8:2 (w/w), no detergent, lipid:protein ratio 1:15 (w/w), [protein] = 34.19, [lipid] = 106.21 μM. The patterning is suggestive of a helical array.

These long helical structures display a striking similarity to helical assemblies of matrix protein seen in filamentous RSV virions (Figure 1.4, Chapter 1). These results indicate that the protein is able to self-assemble into the form it takes in the mature virus, in the absence of any other viral proteins. This strongly suggests that the RSV matrix protein, like the matrix proteins from Ebola, VSV, Sendai virus and influenza virus, has all the structural elements required for the formation of virus-like particles.[24, 33, 34]

In some images, for example Figure 2.17 A and Figure 2.18 A, tubules can be seen apparently growing from liposomes. Another example of this is shown in the image (from sample 1; see Appendix A) in Figure 2.19 below. Ordered tubules such as

these, growing out of liposomes, suggests a parallel with the budding of RSV particles from the host cell membrane, a process controlled by the matrix protein.

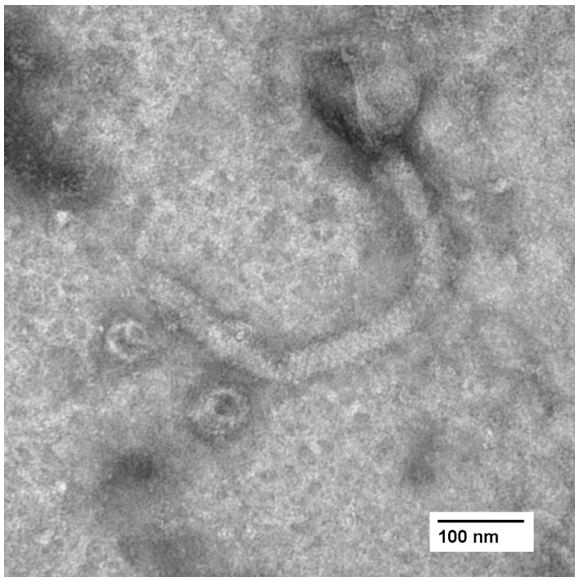


Figure 2.19: TEM negative stain image showing a tubule growing from a liposome. DPPE/DOPC 2:8, lipid:protein ratio 1:15 (w/w), [protein] = 17.09 μM, [lipid] = 54.06 μM.

The helical assemblies described above were produced without the use of detergent; it was therefore decided to discontinue the dialysis experiments and instead focus on replicating these interesting structures by incubating protein with lipid mixtures. The slow dialysis procedures had taken place over about a week at 37°C, so incubations were carried out under similar conditions. The table in Appendix A summarizes incubation experiments carried out; those in which helical structures were observed are highlighted. DPPA was used in some of the lipid mixtures, since the working hypothesis was that the original helical structures had formed very slowly as some of the lipids hydrolysed to phosphatidic acids. Cholesterol and sphingomyelin, which are other common components of cell membranes, were used in some of the later experiments, as concurrent Langmuir trough work (see Chapter 3, Section 3.2 for a full discussion) had suggested that there was some type of specific interaction between the M protein and these species; it was therefore of interest to see how the M protein behaved when incubated with them.

It may be seen from the summary table (Appendix A) that several samples containing different lipids produced some kind of tubular structures, but

unfortunately none of these contained helical assemblies that were quite as clearly defined as the original ones shown in Figure 2.18. It was noted that sample 1, and all the other samples in which structures resembling the original helical assemblies were seen, contained the lipid DPPE. This has a small headgroup compared to PC lipids, which allows membranes containing it to form curved structures. RSV M therefore appears to interact with membranes containing DPPE to form very ordered structures with a high degree of curvature; this supports the hypothesis that M preferentially interacts with such lipids in order to facilitate virion budding from the host cell membrane, a process that requires large changes in membrane curvature.

Several of the samples contained bundles of tubules of a smaller diameter (10 nm on average) than the helical assemblies (Figure 2.20). These were mainly observed in samples containing DPPA, which has an even smaller headgroup than DPPE, allowing greater curvature in the resulting structures thus producing narrower diameters.

The series of incubations also shows that the helical tubules were observed in both His-tagged and untagged protein samples, demonstrating that the native protein is sufficient to form these structures.

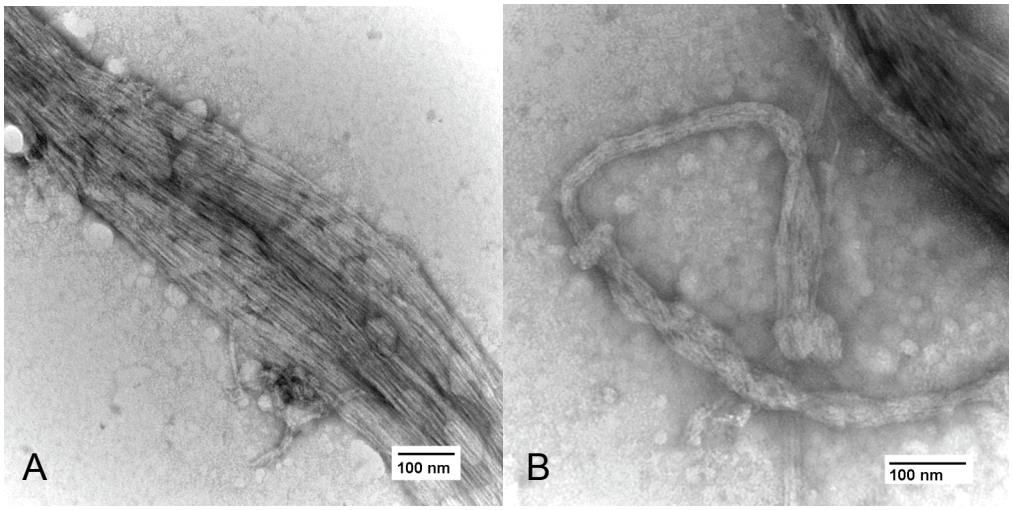


Figure 2.20: Negative-stained TEM images showing bundles of very narrow tubules. DOPC/DPPE/DPPA 8:1:1, lipid:protein ratio 1:50 (w/w), [protein] = 38.8 μM, [lipid] = 148.33 μM.

It must always be borne in mind that the manner of TEM sample preparation, and the drying that occurs in the vacuum of the microscope, may affect the appearance of the

samples being studied. It is possible that the apparent bundling together of the extended structures, seen in these images, is an artefact resulting from this drying. The next aim was therefore to further study these interesting structures whilst avoiding such artefacts resulting from the harsh sample preparation methods required for TEM.

The electron cryo-microscopy (cryo-EM) technique involves vary rapid freezing of the sample to be imaged, and observing it whilst still frozen. Sufficiently rapid freezing to below 273K leads to the formation of vitreous ice rather than ice crystals, allowing the sample to remain hydrated whilst being viewed under the microscope. This means the sample can be imaged in conditions closer to the native environment of a protein, and prevents the damage caused by flattening and dehydration in the vacuum of the microscope. It also removes the need for negative staining, which can limit the resolution obtained.[35]

High-quality cryo-EM pictures can ultimately be used to produce 3D images of single particles,[35] in this case the helical tubules. Given that X-ray structural data is now available for M^{254R}, the structure of the protein could be fitted to the 3D EM images to obtain a model for the arrangement of protein molecules within the tubules.

With this aim, several samples, from experiments 39, 41, 42 and 45, were sent to the laboratory of Prof. Alok Mitra at the University of Auckland, where the author hoped to learn more about cryo-EM and obtain, and possibly analyse, high resolution unstained images of the tubules. However, in Prof. Mitra's opinion the tubules in these samples were not of sufficient quality for such analysis to be carried out (this was probably not helped by the fact that the samples were delayed in transit). Unfortunately the sample (shown in Figure 2.18) in which the first, and best, helical structures had been seen was fairly small in volume and had long since been exhausted, mainly in TEM analysis.

Further incubation experiments were carried out on the suggestion of a member of Prof. Mitra's group, Jae-Kyung Hyung, who has had some success (as yet unpublished) using galactosyl ceramides (GalCers) in conjunction with nickel-

derivitized lipids to induce the formation of protein-lipid tubules. GalCers have a high propensity to form tubules when suspended in aqueous solutions and sonicated. When GalCers are doped with a nickel-containing lipid, the mixture should form tubules studded with Ni atoms, that can act as ligands for the histidine residues in His-tagged proteins, giving a template around which the protein can form a highly ordered helical array.[36] D-galactosyl-β1-1′-N-nervonyl-D-erythro-sphingosine (C24:1 β-D-galactosyl ceramide; GalCer) was mixed with 1,2-dioleoyl-*sn*-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] nickel salt (DOGS-NTA-Ni) in a 8:2 (w/w) ratio, suspended in PBS at pH 7.4 and sonicated for 2 minutes. The resulting suspension was mixed with M^{254R}His solution at lipid:protein ratios of 10:1 and 100:1, and incubated at room temperature for several days.

Portions of these two samples were examined by TEM with negative staining, as for the other incubation samples, but no tubules or other regular structures were observed. This does not necessarily mean that the method cannot produce good results for M^{254R}His, but more time is necessary for optimization of the conditions than was available at this stage of the project.

2.7 References

1 E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch, *Nucleic Acids Research*, 2003, **31**, 3784-3788.

2 Entrez Protein Sequence Database, NCBI.

3 R. Ghildyal, A. Ho, K. M. Wagstaff, M. M. Dias, C. L. Barton, P. Jans, P. Bardin and D. L. Jans, *Biochemistry*, 2005, **44**, 12887-12895.

4 V. A. Money, H. K. McPhee, J. A. Mosely, J. M. Sanderson and R. P. Yeo, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 4441-4446.

5 R. W. H. Ruigrok, G. Schoehn, A. Dessen, E. Forest, V. Volchkov, O. Dolnik, H. D. Klenk and W. Weissenhorn, *Journal of Molecular Biology*, 2000, **300**, 103-112.

6 J. J. Zakowski, W. A. Petri Jr. and R. R. Wagner, *Biochemistry*, 1981, **20**, 3902-3907.

7 R. W. H. Ruigrok, A. Barge, P. Durrer, J. Brunner, K. Ma and G. R. Whittaker, *Virology*, 2000, **267**, 289-298.

8 P. Neumann, D. Lieber, S. Meyer, P. Dautel, A. Kerth, I. Kraus, W. Garten and M. T. Stubbs, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 3710-3715.

9 J. M. Sanderson and E. J. Whelan, *Physical Chemistry and Chemical Physics*, 2004, **6**, 1012-1017.

10 J. M. Sanderson, *Organic & Biomolecular Chemistry*, 2005, **3**, 201-212.

11 M. B. Ulmschneider and M. S. P. Sansom, *Biochemica et Biophysica Acta*, 2001, **1512**, 1-14.

12 S. M. Kelly, T. J. Jess and N. C. Price, *Biochimica et Biophysica Acta*, 2005, **1751**, 119-139.

13 L. Whitmore and B. A. Wallace, *Biopolymers*, 2008, **89**, 392-400.

14 A. Fersht, *Structure and Mechanism in Protein Science*, W. H. Freeman and Company, New York, 1998.

15 A. Lobley, L. Whitmore and B. A. Wallace, *Bioinformatics*, 2002, **18**, 211-212.

16 L. Whitmore and B. A. Wallace, *Nucleic Acids Research*, 2004, **32**.

17 N. Sreerama and R. W. Woody, *Protein Science*, 2004, **13**, 100-112.

18 N. J. Greenfield, *Nature Protocols*, 2006, **1**, 2877-2890.

19 N. Sreerama and R. W. Woody, *Analytical Biochemistry*, 2000, **287**, 252-260.

20 D. Frishman and P. Argos, *Proteins: Structure, Function, and Genetics*, 1995, **23**, 566-579.

21 J. C. Hansen, X. Lu, E. D. Ross and R. W. Woody, *Journal of Biological Chemistry*, 2006, **281**, 1853-1856.

22 M. Sickmeier, J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic and A. K. Dunker, *Nucleic Acids Research*, 2007, **35**.

23 T. Hoenen, V. Volchkov, L. Kolesnikova, E. Mittler, J. Timmins, M. Ottmann, O. Reynard, S. Becker and W. Weissenhorn, *Journal of Virology*, 2005, **79**, 1898-1905.

24 J. Timmins, G. Schoehn, C. Kohlhaas, H.-D. Klenk, R. W. H. Ruigrok and W. Weissenhorn, *Virology*, 2003, **312**, 359-368.

25 F. X. Gomis-Rüth, A. Dessen, J. Timmins, A. Bracher, L. Kolesnikowa, S. Becker, H.-D. Klenk and W. Weissenhorn, *Structure*, 2003, **11**, 423-433.

26 J. L. Rigaud, M. Chami, O. Lambert, D. Levy and J. L. Ranck, *Biochimica Et Biophysica Acta-Biomembranes*, 2000, **1508**, 112-128.

27 G. Mosser, *Micron*, 2001, **32**, 517-540.

28 A. Saenz, O. Canadas, L. A. Bagatolli, M. E. Johnson and C. Casals, *FEBS Journal*, 2006, **273**, 2515-2527.

29 L. Balakireva, G. Schoehn, E. Thouvenin and J. Chroboczek, *Journal of Virology*, 2003, **77**, 4858-4866.

30 P. Palestini, C. Calvi, E. Conforti, L. Botto, C. Fenoglio and G. Miserocchi, *American Journal of Physiology: Lung Cellular and Molecular Physiology*, 2002, **282**, L1382-L1390.

31 H. T. McMahon and J. L. Gallop, *Nature*, 2005, **438**, 590-596.

32 T. Bächi and C. Howe, *Journal of Virology*, 1973, **12**, 1173-1180.

33 P. A. Justice, W. Sun, Y. Li, Z. Ye, P. R. Grigera and R. R. Wagner, *Journal of Virology*, 1995, **69**, 3156-3160.

34 T. Takimoto and A. Portner, *Virus Research*, 2004, **106**, 133-145.

35 J. Ruprecht and J. Nield, *Progress in Biophysics and Molecular Biology*, 2001, **75**, 121-164.

36 E. M. Wilson-Kubalek, R. E. Brown, H. Celia and R. A. Milligan, *Proceedings of the National Academy of Sciences of the United States of America*, 1998, **95**, 8040-8045.

3 Interactions of RSV M with lipid membranes

3.1 Choice of lipids

Eukaryotic cell membranes contain three main types of lipids: glycerophospholipids, sphingolipids and cholesterol or another similar sterol. Representatives of all three types of lipids were used in the following work; their structures are shown in Figure 3.1.

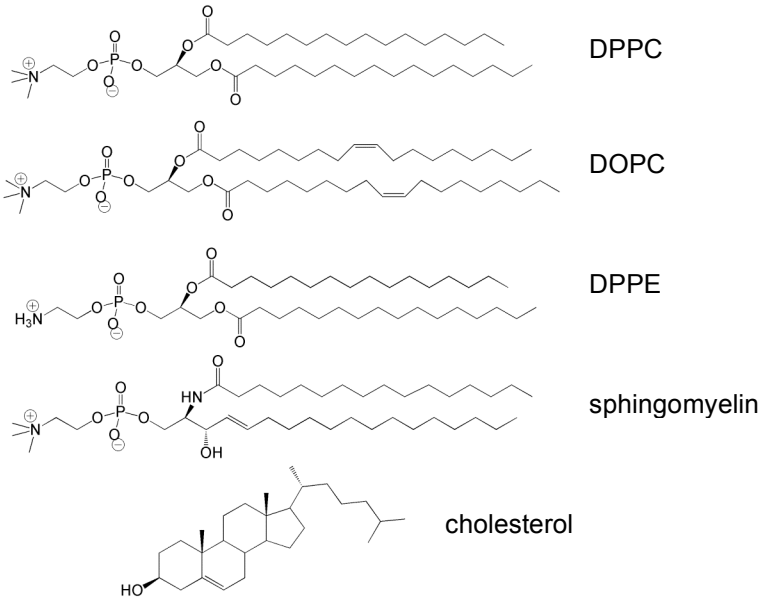


Figure 3.1: Some of the lipids used to model the interactions of RSV M with membranes. Top to bottom: 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC); 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC); 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine (DPPE); sphingomyelin (SM)[†]; cholesterol.

Glycerophospholipids (the most abundant) have a glycerol backbone ester-linked to two hydrophobic fatty acyl chains (or 'tail groups'), and a phosphorylated alcohol 'head group'. The acyl chains normally contain an even number of carbon atoms, and one or both may be unsaturated to some degree – naturally occurring lipids are often asymmetrically unsaturated. Sphingolipids are based around sphingosine rather than glycerol and contain one amide-linked saturated fatty acyl chain and a phosphorylated alcohol or sugar head group. The most common sphingolipids, sphingomyelins, have a phosphocholine head group, like the

[†] Note that carbon chain lengths vary in sphingomyelin; the figure shows C16-sphingomyelin, which has a molecular weight of 703.04. This corresponds to the average molecular weight of the Fluka egg sphingomyelin (product no. 85615) used in this work.

glycerophosphocholines. Sterols (steroid alcohols) are a subgroup of steroids, molecules consisting of four linked rings that are more planar in shape than either glycerophospholipids or sphingolipids. Sterols have a hydroxyl group at the 3-position on the A-ring, and often a side chain at position 17; cholesterol has an isooctyl side chain in this position.[1, 2]

There were several reasons for choosing these particular lipids. RSV targets the lungs, replicating in lung epithelial cells that contain relatively high concentrations of phosphocholine lipids, particularly PC and PE.[3, 4] Lung plasma membranes and lung surfactant contain particularly high levels of PC and PE lipids (Tables 3.1 and 3.2). PC and PE lipids are also abundant in the envelopes of *Paramyxoviridae* (Table 3.3), which are derived from the host cell membrane. DPPC and DPPE were therefore used in many of the following experiments as lipids representative of those found in lung epithelial cell membranes and the RSV viral envelope.

Lipid	Percentage of total phospholipids
Sphingomyelin	11
Phosphocholine	34
Phosphoserine	13
Phosphoinositide	9
Phosphoethanolamine	32

Table 3.1: Phospholipids found in the plasma membrane fraction of human lung tissue. Note the high levels of PC and PE lipids. The cholesterol:phospholipid ratio was 0.22.[3]

	Percentage w/w			
	Rat	Rabbit	Ox	Sheep
Total protein	10	8	18	14
Total lipid	88	90	79	86
	Percentage of total lipid w/w			
Triacylglycerol	5	4	15	7
Unesterified fatty acids	3	2	3	10
Cholesterol	3	3	1	3
Phosphoethanolamine	4	1	3	2
Phosphocholine	73	83	71	58
Phosphoglycerol	5	3	2	4
Phosphoinositide	5	2	3	4
Sphingomyelin	2	2	1	3
Lyso-phosphocholine	Trace	Trace	Trace	11

Table 3.2: Percentages of protein and lipids found in the lung surfactant of various species.[5]

Acyl chain	Percentage of total fatty acids
16:0 (palmitoyl)	4
18:0 (stearoyl)	20
18:1 (oleoyl)	40
18:2 (linoleoyl)	2
18:3	8

Table 3.3: Fatty acid composition in plasma membranes of A_{549} human lung epithelial cells.[6] Number of carbons in chain : number of unsaturated bonds

	Percentage of phospholipids				
	PC	**PE**	**SM**	**PS**	**PI**
RNA viruses					
Paramyxoviridae	8-53	10-41	12-30	2-17	0-11
Rhabdoviridae	16-38	20-34	16-31	7-20	0-10
Orthomyxoviridae	10-39	12-45	16-23	7-22	0-8
Togaviridae	21-49	19-35	7-29	9-21	0-9
Retroviridae	16-30	26-42	22-33	12-17	0-4
DNA viruses					
Herpetoviridae	51-57	21-25	4-21	4-13	1-13
Poxviridae	34-49	12-14	2-4	0-5	12-19

Table 3.4: Percentages of phospholipids found in the envelopes of various virus families. PC = phosphocholine, PE = phosphoethanolamine, SM = sphingomyelin, PS = phosphoserine, PI = phosphoinositide.[5]

DPPE is of additional interest because the ethanolamine head group is smaller than the choline head groups of PC lipids and thus packs differently in membranes, which can affect membrane curvature.[7] For example, a lipid bilayer with a greater proportion of PC lipids in one leaflet, and more PE lipids in the other leaflet, will curve towards the area where the PE lipids are more abundant because their head groups can pack together more closely (Figure 3.2).

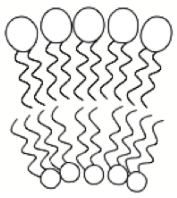


Figure 3.2: Diagram illustrating how lipid head group packing could influence membrane curvature.

Such effects may be important for the function of the RSV matrix protein during budding of virions from the host cell membrane, which involves major changes in membrane curvature (Figure 3.3). Matrix proteins have been shown to induce

changes in curvature in model membranes,[8] and altering the distribution of lipids within the membrane is one mechanism by which they could achieve this. If RSV M has a particular binding affinity for specific lipids (such as PEs) this might allow the protein to recruit these lipids to different areas of the membrane.

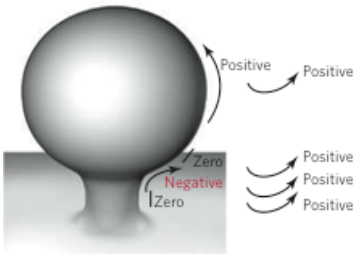


Figure 3.3: Membrane curvature during budding.[7†]

Cholesterol and sphingomyelin were used in order to study the interaction of RSV M with monolayers containing phase-separated microdomains. As discussed in the introduction (Section 1.4.5.1), there exist microdomains within membranes that are enriched in cholesterol and sphingolipids. The favourable interactions of cholesterol with glycosphingolipids, and its relatively planar shape, allow it to intercalate between them, forming a liquid crystalline ordered phase (L_o) surrounded by the liquid crystalline disordered (L_d) phase of the rest of the membrane. A lipid mixture containing equal quantities of DOPC, cholesterol, and either DPPC or sphingomyelin, provides a good model for this type of system.[9] RSV M has been found in association with detergent-resistant membrane fractions,[10] which may (or may not) be related to lipid microdomains; virus-like particles produced by expression of Ebola VP40 also appeared to bud from detergent-resistant domains.[11] Caveolin-1 and other cellular marker proteins attributed to microdomains have been identified in the viral envelopes of RSV and related species such as Newcastle disease virus; this finding has been used to suggest that such microdomains play a role in viral budding.[12-15] Studying RSV M in conjunction with model microdomain-containing membranes and monolayers will provide further insight into its interactions with these regions.

† Reprinted by permission from Macmillan Publishers Ltd: H. T. McMahon and J. L. Gallop, *Nature*, 2005, **438**, 590-596, copyright 2005

In addition to these synthetic lipids and lipid mixtures, a chloroform extract of mouse lung tissue was used for some experiments. The aim of this was to see how the complex natural mixture of lipids (and other chloroform-soluble components) from the mouse lung epithelial membrane behaved, in comparison with the simpler model lipid mixtures, and thus obtain some idea of how accurately the models reflect the situation for real membrane lipids. Of course cell plasma membranes also contain around 50 % protein,[2] which this experiment does not take into account.

3.2 Langmuir studies of protein-lipid monolayers

3.2.1 Modelling membranes as monolayers

Natural membranes are very complex structures containing a great variety of lipids and proteins, which means that studying living cells or isolated cellular membranes can only really provide useful information about global membrane phenomena. The easiest way to investigate specific aspects of biological processes that occur at the membrane is to use model systems. Lipid monolayers are widely studied as simple models of such cell membranes; experiments undertaken on a monolayer give a good indication of the behaviour of the membrane bilayer, which can be considered to consist of two weakly coupled monolayers.[16, 17] Lipids (and cholesterol) are amphiphilic molecules, with hydrophilic charged 'heads' containing a phosphate group, and hydrophobic alkyl 'tails'. This explains their tail-to-tail arrangement in the membrane bilayer, and means that they will be adsorbed at an air-water interface, spontaneously arranging into an essentially insoluble monolayer, with the charged head groups in the water and the hydrophobic tail groups in the air.

RSV M is a peripheral membrane protein, but the nature of its association with the plasma membrane is not well understood. Membrane proteins may associate with the surface of the membrane, or penetrate partially or completely into it. Such interactions may be studied by introducing proteins into the aqueous subphase below a lipid monolayer formed at the surface, and looking at the effect this has upon various measurable parameters.[18] The most convenient way to do this is to use a

Langmuir trough; the theory behind the behaviour of a monolayer at a surface is discussed in Section 3.2.2, while the method is described in Section 3.2.3.

3.2.2 Thermodynamics at the air-water interface

In practice there is no such thing as a free surface; it is more correct to instead talk about an interface, the boundary region between two phases, in this case liquid and gas. However, for simplicity, any water-gas interface is normally referred to as a water surface. At such an interface, there is a transition between the composition and properties of the two bulk phases on either side; on the one hand the air, with all its component molecules far apart and moving randomly, and on the other the water, where the molecules have much more limited freedom because of the strong hydrogen bonding interactions between them. The water surface always has excess free energy because of the difference in environment between the molecules at the surface and those in the bulk, the pattern of hydrogen bonding in the bulk water being disrupted near the surface. However, because such intermolecular forces only act over a very short range, the transition between the two phases occurs over a distance of only one or two molecular diameters, so the interfacial region is very thin.[19, 20]

When a liquid is in equilibrium with its vapour, the interfacial region (of volume V) is where the density, energy and entropy all undergo a transition. The thermodynamics of the interfacial region may be analysed by treating the system as though there is an imaginary mathematical dividing surface, and assuming that the properties of the two bulk phases do not change up to the point at which this surface divides them. The real system, containing the interface, will have different extensive quantities (energy U, entropy S, and moles of each component i, n_i) than does the reference system; the differences between these quantities are called the surface excesses and are denoted by a superscript s. The properties of the real system are thus given by:

$$V = V' + V''$$
$$U = U' + U'' + U^s$$

$$S = S' + S'' + S^s$$
$$n_i = n_i' + n_i'' + n_i^s \quad (1)$$

where ′ and ″ denote the two bulk phases of the reference system.

When the interface is plane, and the imaginary dividing surface is put in a defined position, the 'surface of tension', the following is true for any variation from equilibrium:

$$dU = TdS + \sum_i \mu_i dn_i - PdV + \gamma ds \quad (2)$$

where T is temperature, μ_i is the chemical potential of component i, s is the surface area and γ is the surface tension, the line force acting on the surface molecules; γ is the partial derivative of the Helmholtz free energy, F, of the system with respect to the surface area (*i.e.* the excess free energy per unit area), when the temperature, total volume and amounts of all components are kept constant:[20, 21]

$$dF = -SdT + \sum_i \mu_i dn_i - PdV + \gamma ds \quad (3)$$

and

$$\gamma = \left.\frac{\partial F}{\partial s}\right|_{T,V,n_i} \quad (4)$$

which may alternatively be expressed as:

$$\gamma = \left.\frac{\partial G}{\partial s}\right|_{T,P,n_i} \quad (5)$$

where G is the Gibbs free energy of the system.

There are several other relations between the surface excess quantities, which for a plane interface remain the same regardless of the position of the dividing surface, for example:

$$dU^s = TdS^s + \sum_i \mu_i dn_i^s + \gamma ds \qquad (6)$$

Integration and total differentiation of equation *(6)* gives the following expression, the surface analogue of the Gibbs-Duhem equation:

$$S^s dT + \sum_i n_i^s d\mu_i + s d\gamma = 0 \qquad (7)$$

and if temperature T is constant,

$$s d\gamma + \sum_i n_i^s d\mu_i = 0 \qquad (8)$$

For convenience, surface excess quantities can be expressed in terms of unit area; if $\Gamma_i = n_i^s / s$ (*i.e.* the surface excess concentration), equation *(8)* becomes:

$$d\gamma + \sum_i \Gamma_i d\mu_i = 0 \qquad (9)$$

For a two-component system,

$$d\gamma + \Gamma_1 d\mu_1 + \Gamma_2 d\mu_2 = 0 \qquad (10)$$

This is one form of the Gibbs adsorption equation. As mentioned above, it holds true regardless of the location of the imaginary dividing surface. If this is placed in such a position that the surface excess of solvent vanishes, *i.e.* so that $\Gamma_1 = 0$, then,

$$d\gamma = -\Gamma_2 d\mu_2 \qquad (11)$$

where $\mu_2 = \mu_2^0 + \mathrm{R}T\ln a_2$, a_2 being the activity of the solute and μ_2^0 its standard chemical potential. This gives:

$$\Gamma_2 = -\frac{1}{\mathrm{R}T}\left(\frac{d\gamma}{d\ln a_2}\right) \qquad (12)$$

This is the Gibbs adsorption isotherm, which very usefully allows the surface excess of a solute to be evaluated by measuring variation in the surface tension γ.[20]

When studying insoluble monolayers on liquid surfaces, the quantity that is normally of interest, and the one that can be measured, is the difference in surface tension between a clean liquid surface and one that is covered with a film. It was in fact the reduction in the tension of a clean water surface that led the early workers in this field Pockels, Rayleigh and Langmuir to their conclusions about the nature of films spread on liquid. Langmuir also made the analogy between this surface tension differential, and a force or pressure exerted by the molecules in the monolayer, allowing surface tension effects to be simply interpreted in terms of intermolecular forces. Since then, the force measured in monolayer experiments has been known as the surface pressure, Π, which is expressed as: [20]

$$\Pi = \gamma_0 - \gamma \qquad (13)$$

where γ_0 is the surface tension of the clean surface, and γ is the surface tension with the monolayer present.

The interaction of RSV M protein with lipid monolayers is an example of monolayer penetration, a term used to describe the interaction of an insoluble monolayer (the lipid) spread at a phase boundary with a soluble surface-active species (M protein) in one phase.[22] Naturally this is rather more complicated than the behaviour of a single-component monolayer, and several different approaches have been developed to model such systems. One of the most widely used is the Pethica modification of the Gibbs' adsorption isotherm, equation *(12)*, which can also be written as,

$$d\Pi = \sum_{i}^{n} \Gamma_i d\mu_i \qquad (14)$$

Assuming the adsorbed amounts of all the components, except the *j*th one, are constant, the partial derivative of *(14)* is,

$$\left(\frac{\partial \Pi}{\partial \mu_i}\right)_{\Gamma_{i\neq j}} = \Gamma_j + \sum_{i\neq j}^{n} \Gamma_i \left(\frac{\partial \mu_i}{\partial \mu_j}\right)_{\Gamma_{i\neq j}} \qquad (15)$$

Via a number of steps (see ref. 23 for the full details), this eventually leads to the generalized Pethica equation:

$$\left(\frac{\partial \Pi}{\partial \ln a_j}\right)_{\Gamma_{i\neq j}} = \mathrm{R}T\Gamma_j \left[1 - \sum_{i\neq j}^{n} \Theta_i\right]^{-1} \qquad (16)$$

where $\Theta_i = \Gamma_i / \Gamma_{\infty i}$ is the monolayer coverage. When there is one insoluble and one soluble component, denoted by 1 and 2, and an ideal bulk solution, this becomes the ordinary Pethica equation:

$$\left(\frac{\partial \Pi}{\partial \ln c_2}\right)_{\theta_1} = \frac{\mathrm{R}T\Gamma_2}{(1-\Theta_1)} \qquad (17)$$

where c_2 is the bulk concentration of the soluble component.[23]

The Pethica equation is considered to provide an adequate description of fairly complicated mixed monolayer systems, including the penetration of a soluble protein into a monolayer composed of insoluble phospholipid.[23] If the monolayer in question is taken to be an ideal mixture of soluble and insoluble surfactant, the behaviour of both components can be described by a generalized Szyszkowski-Langmuir equation:

$$\Pi = -\frac{RT}{\omega_\Sigma}\ln(1-\Theta_1-\Theta_2) \qquad (18)$$

where ω_Σ is the mean molar area defined by,

$$\omega_\Sigma = \frac{\Gamma_1\omega_1 + \Gamma_2\omega_2}{\Gamma_1 + \Gamma_2} \qquad (19)$$

Differentiating equation *(18)* with respect to Θ_2, assuming Θ_1 is constant (and also ω_Σ, which is approximately true) gives:

$$\left(\frac{\partial\Pi}{\partial\Theta_2}\right)_{\Theta_1} = \frac{RT}{\omega_\Sigma}\frac{1}{1-\Theta_1-\Theta_2} \qquad (20)$$

It follows from the Pethica equation *(17),* and equation *(20)*, that

$$d\ln c = \frac{1-\Theta_1}{\Theta_2(1-\Theta_1-\Theta_2)}d\Theta_2 \qquad (21)$$

Integration of the above expression gives the adsorption isotherm equation for a soluble surfactant and an insoluble monolayer, where K is the adsorption equilibrium constant:[23]

$$Kc_2 = \frac{\Theta_2}{1-\Theta_1-\Theta_2} \qquad (22)$$

Equation *(22),* which is a modified Langmuir binding isotherm, will be used later in this chapter to analyse experimental data obtained from lipid monolayers on the surface of a subphase containing dissolved M protein.

3.2.3 Pressure-area isotherms and the Langmuir trough

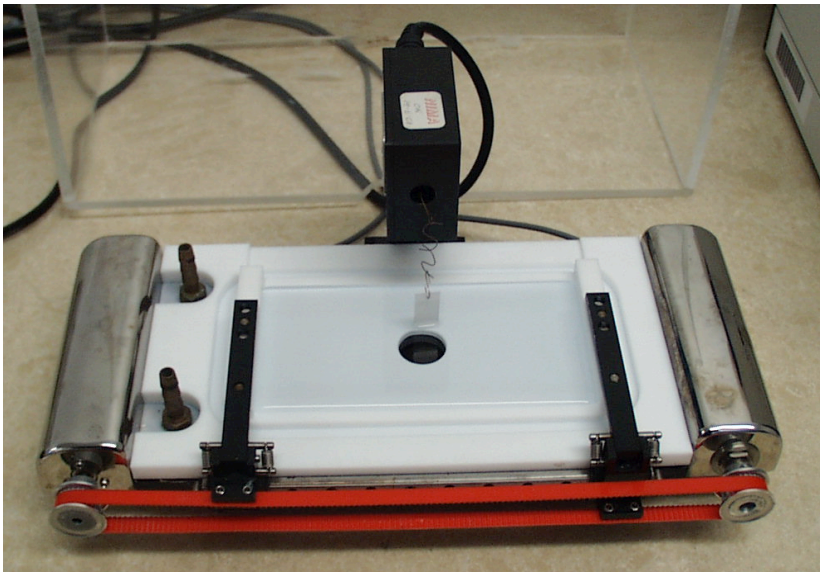


Figure 3.5: Photograph of the Langmuir trough (Nima Technology) used to carry out the work described in this chapter.

Monolayers can be conveniently studied using a Langmuir trough (Figure 3.5), which consists of a shallow trough made from a hydrophobic substance such as Teflon, with one or two moveable barriers. The trough is filled with a water or buffer subphase, and the monolayer is introduced by dropping a known quantity of lipid dissolved in a volatile solvent (usually chloroform) onto the subphase surface. The barriers are then moved across the surface to compress the monolayer while a Wilhelmy plate cut from filter paper monitors changes in surface pressure Π.

A diagram of a Wilhelmy plate is shown in Figure 3.6.[16] The forces acting on the plate are its weight P and surface tension γ downward, and Archimedes buoyancy A upwards, producing a net downward force f of:

$$f = P + 2\gamma(w + t)\cos\theta - A \qquad (23)$$

where w and t are the width and thickness of the plate ($t << w$), and θ is the contact angle of the plate with the liquid. If the plate is completely wetted, $\theta = 0$ and $\cos\theta = 1$, so that:

$$f = P + 2\gamma w - A \qquad (24)$$

P and A are constant even when the composition of the interface varies, as long as the plate is kept in the same position (which is achieved by the balance mechanism on the trough), so

$$\Delta f = 2w(\gamma - \gamma_0) = -2w\Pi \qquad (25)$$

and

$$\Pi = \frac{-\Delta f}{2w} \qquad (26)$$

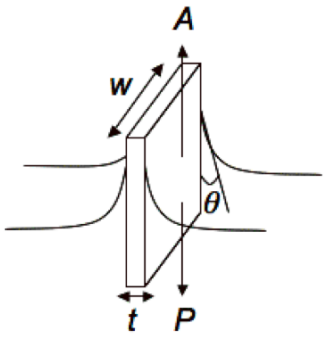


Figure 3.6: Diagram of a Wilhelmy plate, used to measure surface pressure Π.

The surface pressure may be plotted against the area per molecule to produce a pressure-area isotherm. Such isotherms exhibit changes in gradient when the monolayer goes through phase changes. For monolayers of simple amphiphiles such as straight-chain fatty acids, there are normally three distinct phases; the 'gas' phase, where the molecules are so far apart that they do not interact with each other; the 'liquid' phase, where the molecules are closer together but not aligned; and the 'solid' phase, where the molecules are tightly packed and ordered. When the monolayer is compressed beyond a certain point, known as the collapse pressure, the surface pressure abruptly drops because the monolayer has been compressed to its maximal extent and starts to fold in on itself, forming disordered multilayers. These phases can be seen in the pressure-area isotherm of stearic acid in Figure 3.7.

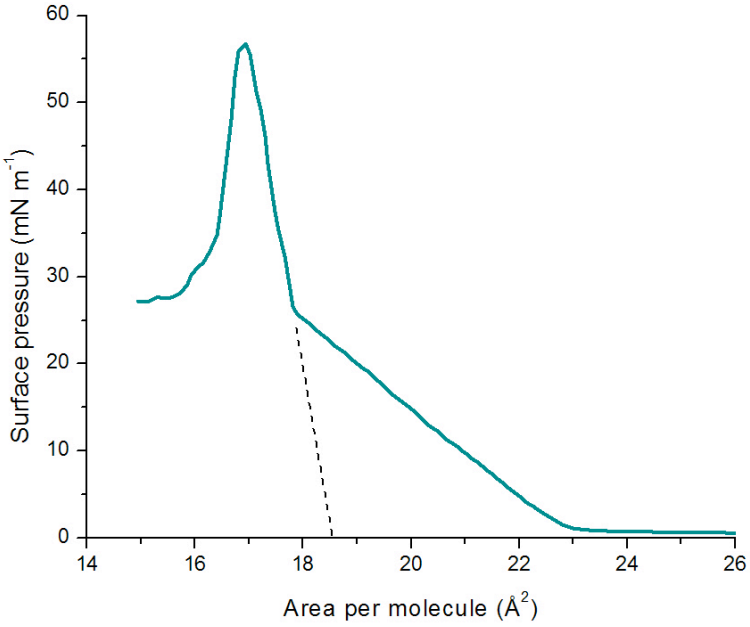


Figure 3.7: A pressure-area isotherm of stearic acid ($CH_3(CH_2)_{16}CO_2H$)*
At high areas the molecules are widely spaced in a 'gaseous' phase; when the gradient changes this represents the start of the 'liquid' phase, where the molecules are close enough together to interact, but are not orientated; the steepest part of the graph represents the 'solid' phase where the molecules are closely packed and aligned; at a surface pressure of around 55 mN m^{-1} the collapse pressure is reached.

Lipids behave in a slightly more complex way than conventional amphiphiles such as stearic acid. They may be found in one of several phases in the membrane bilayer depending on the composition, temperature or pressure (Figure 3.8).[24] Under physiological conditions mixed lipid bilayers exist predominantly in the liquid crystalline L_α phase, at least in the absence of cholesterol. The inclusion of cholesterol results in the formation of liquid crystalline ordered L_o domains within a liquid crystalline disordered phase L_d, as discussed in Section 1.4.5.1 in Chapter 1.[1] The main phase transition T_m is between the relatively fluid L_α and gel ($L_{\beta'}$) phases of a lipid bilayer, and occurs at significantly lower temperatures for unsaturated phospholipids than for saturated phospholipids (*e.g.* -20 °C for DOPC versus 41 °C for DPPC and 63°C for DPPE[25]). This is because *cis* double bonds in the unsaturated acyl chains prevent them packing tightly together. A solid crystalline (L_c) phase may form upon prolonged cooling, or at high pressures.

* Image provide courtesy of Nima Technology Ltd. (www.nima.co.uk)

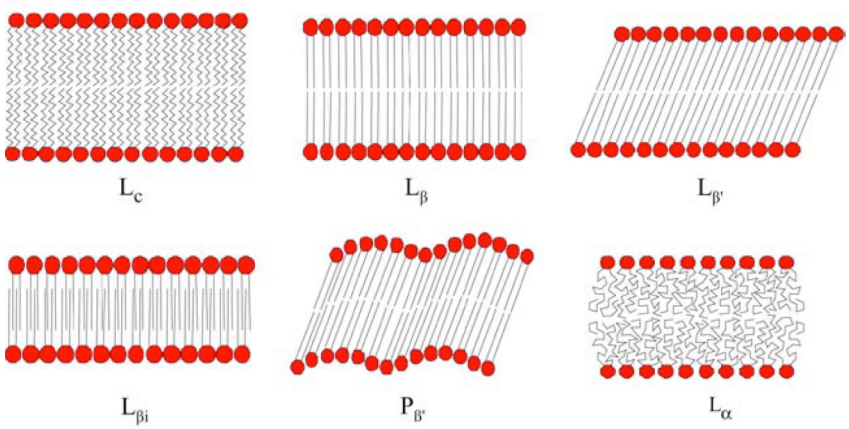


Figure 3.8: The arrangement of lipid molecules in various membrane phases.[†]

The general pressure-area isotherm shown in Figure 3.9 is similar to those produced from lipid monolayers, exhibiting changes in gradient that represent phase changes in the monolayer, as seen in the stearic acid isotherm in Figure 3.6. The main difference from the stearic acid isotherm is that the lipid monolayer goes through a period of phase coexistence, where the gradient of the isotherm is zero. The lipid L_α and $L_{\beta'}$ phases are roughly analogous to the 'liquid expanded' and 'tilted condensed' phases shown in Figure 3.9, while the 'untilted condensed' phase is similar to the lipid phase L_β or L_c (see Figure 3.8 above). It must be noted that only some of these possible phases can be seen in the experimental isotherms reproduced in the following sections, as it was difficult to compress one monolayer over the whole range of pressures that would be necessary to observe all of them. The surface pressure range in which lipid monolayers behave most like lipid bilayers is 30-35 mN/m,[26] and this range was covered in the experiments described here.

[†] Reprinted from, J. Eisenblätter and R. Winter, *Biophysical Journal*, 2006, **90**, 956-966, Copyright (2006), with permission from Elsevier.

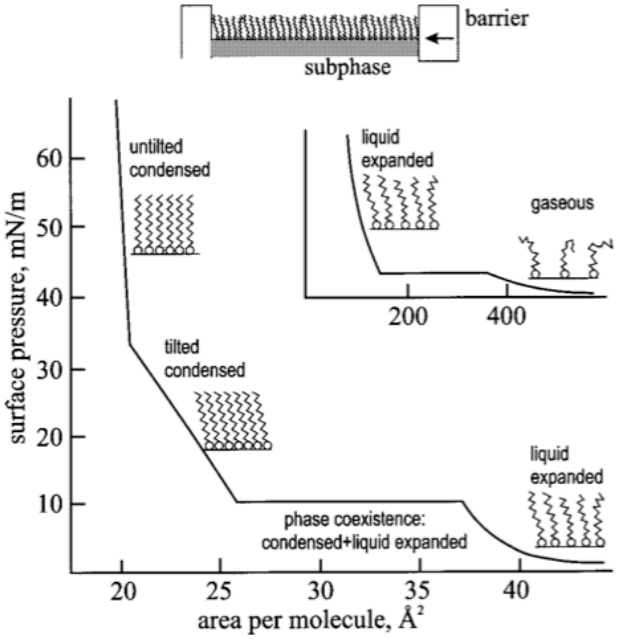


Figure 3.9: A schematic diagram showing a monolayer on a Langmuir trough, and a generalized Langmuir isotherm.
Horizontal sections of the isotherm represent regions of phase coexistence occurring during first order phase transitions, and sharp kinks indicate continuous transitions.*

A Langmuir trough (Figure 3.5) was used to study the behaviour of monolayers containing various mixtures of the lipids discussed in Section 3.1, in the presence of RSV M, to gather information on the protein's mode of interaction with the monolayer.

In all cases a lipid monolayer was spread by dropping lipid dissolved in chloroform on the surface of 52 mL of PBS subphase at pH 7.4 in the Langmuir trough. A surface pressure-area isotherm was recorded at a rate of 10 cm^2/min. This compression rate was found to give isotherms indistinguishable from those recorded more slowly, and meant that each experiment was completed within a reasonable length of time. Successive isotherms were recorded with the addition of increasing quantities of M^{254R} (dissolved in PBS at pH 7.4) at the surface, before the introduction of the lipid monolayer.

* From V. M. Kaganer, H. Möhwald and P. Dutta, Reviews of Modern Physics, 1999, 71, 779-816, Copyright (1999) by the American Physical Society

I am grateful to the 2008-9 M.Chem project student Jenny Carlisle for providing additional data for this section.

3.2.4 Surface adsorption of M

Before introducing any lipid monolayers to the Langmuir trough, the surface activity of RSV M on its own was studied. A surface adsorption isotherm for RSV was generated by dropping small aliquots of M^{254R} in solution onto a clean PBS surface, with the surface pressure initially reading 0. When the surface pressure had equilibrated (generally after about 20 minutes) the pressure reading was recorded, and the next aliquot of protein added. In common with many other proteins at the air-water interface,[27-29] the surface pressure Π displays a sigmoidal dependency on the logarithm of the bulk protein concentration C_B, and the following logistic equation *(27)*[27] could be fitted to the data (Figure 3.10).

$$\gamma_{lv} = \left(\frac{\gamma_{lv}^0 - \gamma_{lv}'}{1 + (\ln C_B^{\Pi/2} / \ln C_B)^N} \right) + \gamma_{lv}' \qquad (27)$$

where γ_{lv} is the liquid-vapour interfacial tension, γ_{lv}^0 and γ_{lv}' are the horizontal asymptotes of a concentration-dependent γ_{lv} curve (corresponding to the surface tension of the buffer in the absence of protein and the surface tension of the saturated interface, respectively), C_B is the molar concentration of protein in the bulk solution, Π is the surface pressure ($= \gamma_{lv}^0 - \gamma_{lv}'$), $C_B^{\Pi/2}$ is the concentration at half the maximum change in γ_{lv}, and N is a factor describing the gradient of the linear region of the curve. Note that Figure 3.10 is plotted in terms of Π rather than γ_{lv}, with values of Π obtained by subtracting the calculated γ_{lv} values from the surface tension of a clean buffer surface, γ_{lv}^0 (71.97 mN/m).[29]

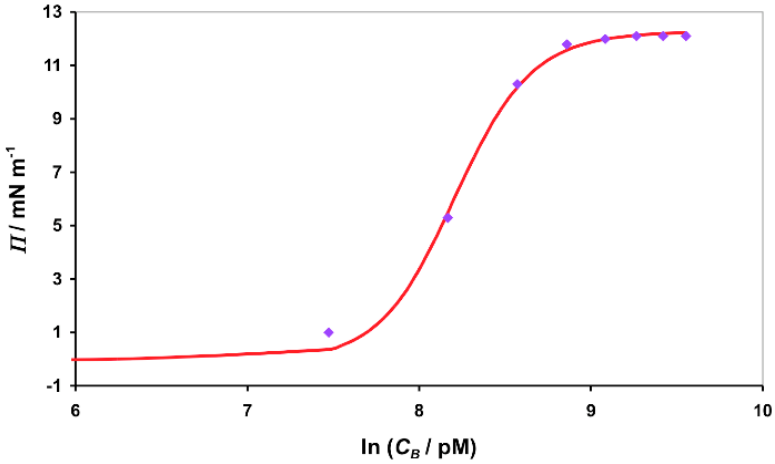


Figure 3.10: Surface pressure Π as a function of the natural log of the bulk protein concentration C_B.
The points show experimental data and the curve results from fitting to equation *(27)*.

Equation *(27)* forms part of an analysis of protein adsorption showing that proteins follow a 'Traube-like' rule. The Traube rule describes how the molar concentration of a linear hydrocarbon surfactant (such as the fatty acids discussed earlier), required to reduce the liquid-vapour interfacial tension to a given value, decreases regularly with each additional methylene unit in the hydrocarbon chain. Vogler *et al.*[27] have shown that proteins follow a similar pattern according to their size; that is, the molar concentration of protein required to reduce γ_{lv} to a specified value decreases with increasing molecular weight.

They explain this 'Traube-like' rule in terms of a relatively constant partition coefficient P that causes the protein molecules to enter a three-dimensional interphase region, separating the bulk solution from the bulk vapour phase. This contrasts with the more conventional model, described in the first part of section 3.2.2 above, where the surface is treated as a planar area with negligible thickness. The protein molecules are modelled as spheres, which have dimensions that scale in proportion to molecular weight because of the relationships between molecular weight, solvent-exposed area, volume, and packing density.[27, 29] This means that the interphase becomes thicker with increasing protein size, and the molar interphase concentration C_I is lower for larger proteins than for smaller ones; in fact it varies inversely with molecular weight.

Protein size and intermolecular repulsion within the interphase mean that there is an upper limit to interphase protein concentration, $C_I^{\max}$, at which interphase saturation is reached. This is equal to the bulk solution concentration $C_B^{\max}$ at which the limiting spreading pressure $\Pi^{\max}$ is reached. For M^{254R}, $\Pi^{\max}$ = 12.27 dyn/cm (Figure 3.10), determined from the fitted parameters in equation *(27)*. This is not particularly high when compared with $\Pi^{\max}$ figures for other proteins and surfactants (for example ubiquitin, 21-27 dyn/cm, and SDS, about 31 dyn/cm[29]), but is of a comparable magnitude to values obtained for membrane peptides,[30] and is consistent with preliminary data reported for the Borna disease virus matrix protein.[31]

As mentioned above, $C_B^{\Pi/2}$ is the bulk protein concentration at half the maximum change in γ_{lv}. This represents the mid-point of the linear-like region of the curve in Figure 3.10, and the gradient of this section, centred around $\ln C_B^{\Pi/2}$, gives the apparent Gibbs surface excess Γ_{lv} according to the Gibbs adsorption isotherm:[27]

$$\Gamma_{lv} = -\frac{1}{RT}\left(\frac{d\gamma_{lv}}{d\ln C_B}\right) \qquad (28)$$

For M^{254R}, $\ln C_B^{\Pi/2}$ = 8.66 and $C_B^{\Pi/2}$ = 3687 pM (3.687 nM), so that Γ_{lv} = 5.81 pmol/cm^2.[†] From this value for Γ_{lv} it is also possible to calculate the surface area occupied by one protein molecule, which is 2856.06 Å^2 (28.56 nm^2). The crystal structure of RSV M showed that the approximate dimensions of each of its two domains are 2 nm × 4 nm, so that each one has a maximum cross-sectional area of 8 nm^2. One domain at the interface could therefore occupy up to 16 nm^2 if a square lattice was formed, and if both domains were present at the surface with the linker region fully extended, the protein could occupy up to 64 nm^2, also assuming a square lattice formation. Even allowing for the possibility that the apparent and true surface excesses could differ substantially, the value for $1/\Gamma_{lv}$ of 28.56 nm^2 per molecule suggests that both domains of the protein are located at least partially at the surface. Significant interactions between protein molecules forming a lattice at the interface

[†] It should be noted that apparent surface excess values calculated in this way, using solute concentrations, may deviate substantially from actual surface excesses calculated using activity-corrected chemical potentials.[29]

would also be consistent with the rather low saturation concentration $C_B^{\max}$ of 5.789 nM, and account for the large molecular area if distal regions of the protein were in contact with each other.

Vogler *et al.*[29] produced a plot of $\ln C_B^{\max}$ versus $\ln MW$ for proteins covering a wide range of molecular weights, which showed that there is a linear relationship between the two quantities:

$$\ln C_B^{\max} = (-1.3 \pm 0.2) \ln MW + (19.8 \pm 1.0) \qquad (29)$$

According to this relationship, RSV M, which has a mass of 29.5 kDa, would be expected to have $C_B^{\max}$ = 4878228 pM, or 4.87 μM. This is three orders of magnitude higher than the value of 5.789 nM (5789 pM) obtained from the experimental data, which most probably indicates that RSV M does not behave as a hard sphere at the interface, as the model assumes. This reflects the fact that M is a peripheral membrane protein rather than a globular one; the crystal structure suggests that M has a relatively compact structure in solution, but it is reasonable to suppose that this could unfold significantly when the protein is at the interface (*c.f.* the CD data in section 2.4) to expose the hydrophobic surfaces buried between the two domains. This supports a model in which the region linking the domains is flexible, allowing domain movement, and provides a possible mechanism for membrane interaction.

There is another possible reason why a protein might adopt a different conformation in the adsorbed state, which is that it could simply denature at the interface. This is less likely to be the case for M^{254R}, however; additional experiments were carried out where protein was gradually added to the subphase and lipid monolayer *via* tubing, and changes in γ_{lv} took place over longer timescales (1-2 hours), and these gave very similar γ_{lv} - C_B plots to those presented here. Therefore any putative denaturation processes must either have had very little effect on the results, or else occurred significantly more quickly or slowly than the experimental time frame.

3.2.5 Isotherm reproducibility

The following experiments all rely on findings from Π-area isotherms for various lipid monolayers interacting with RSV M at the liquid-vapour interface. Several sets of Π-area isotherms from cycled compressions of the same monolayer were produced for M^{254R} and some of the lipid mixtures; examples of these are shown in Figure 3.11 below. In each case the surfactants were added to the surface, allowed to equilibrate for about 15 minutes, and the three isotherms recorded with no relaxation time in between each one.

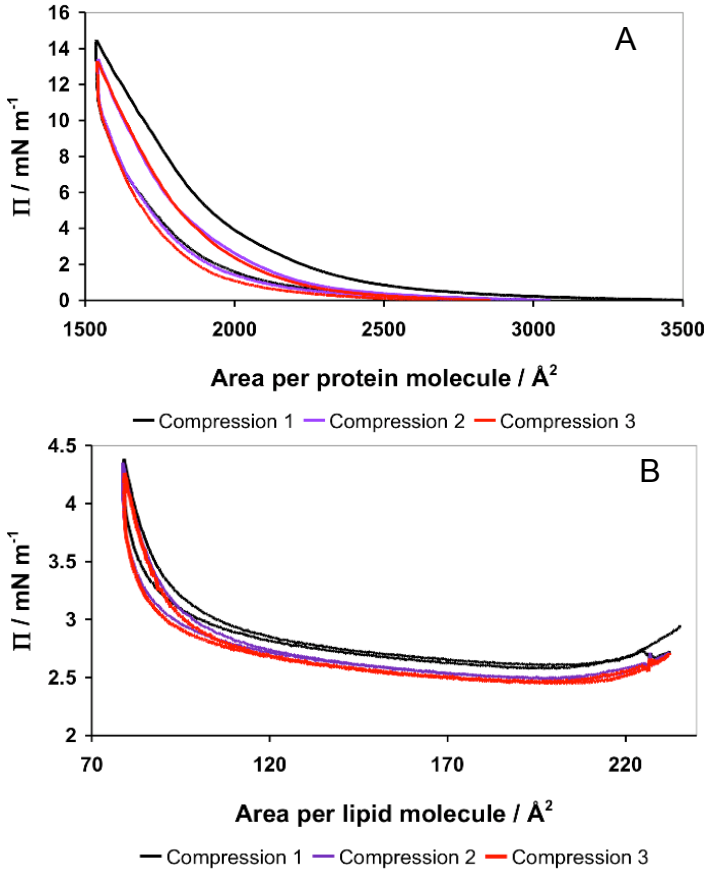


Figure 3.11: Π-A isotherms from cycled surface compressions.
A: Bulk M^{254R} concentration of 6.57 nM; B: 4 µg (5.26 nmol) of DOPC/DPPC 1:1 (w/w).

Both sets reveal some degree of hysteresis in the surface pressure values between isotherms (the average difference in Π values between the first isotherm and subsequent ones is 12 % for protein and 4.5 % for lipids), although the curves are broadly consistent with each other. It does however show in the case of the M protein, a small amount of material is lost from the surface during the first few

compressions, as one would expect for a soluble surfactant. The protein-lipid isotherms described below were recorded after 2-3 preliminary compressions to ensure that no additional material was lost from the surface during data collection.

3.2.6 Adsorption of RSV M to lipid monolayers

Figure 3.12 shows surface pressure-area isotherms for a range of lipid monolayers, in the absence of protein and with increasing quantities of M^{254R} in the subphase.

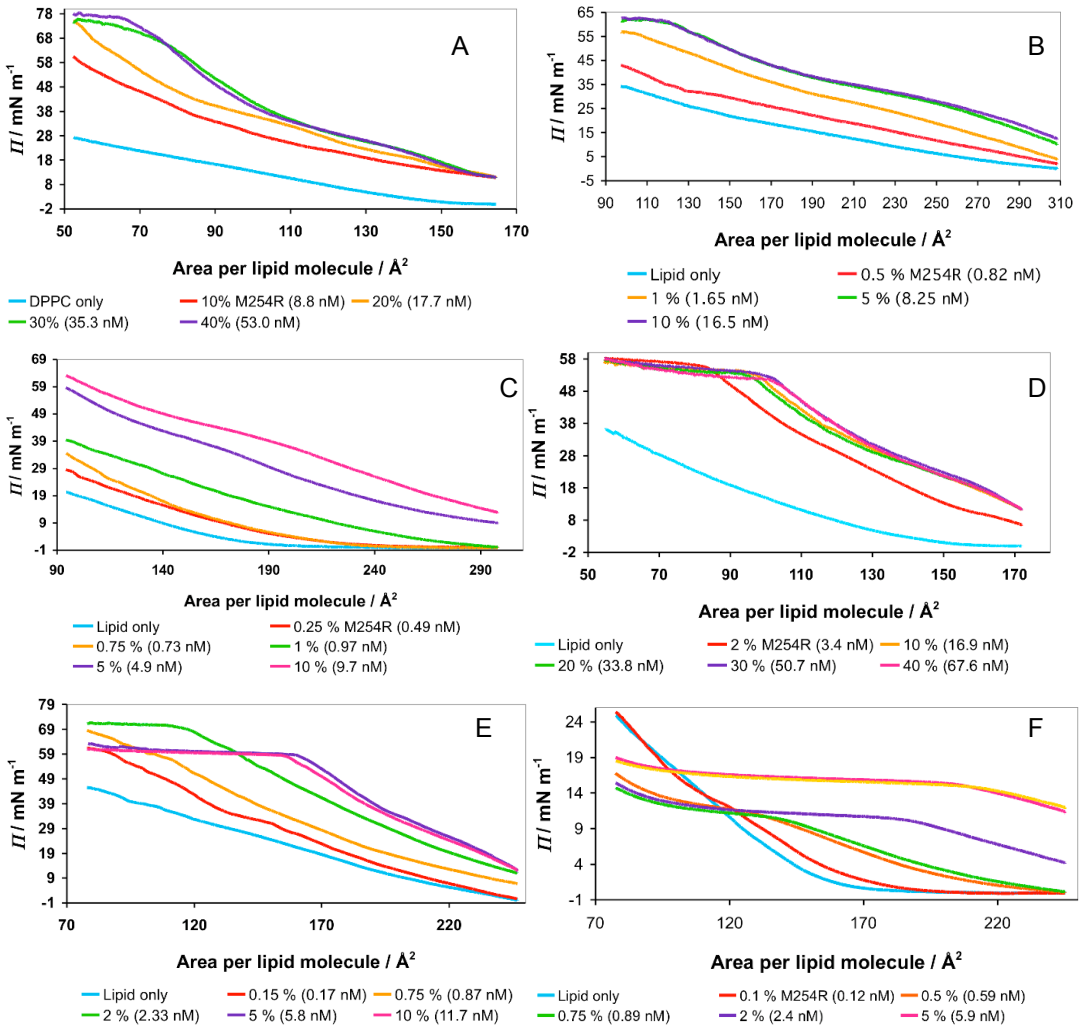


Figure 3.12: Π-A isotherms of lipid monolayers with increasing quantities of RSV M. Percentages refer to the number of moles of protein used with respect to the number of moles of lipid; figures in brackets give the subphase protein concentration. A: 7 µg DPPC; B: 4 µg DOPC; C: 4 µg DOPC/DPPC 1:1 (w/w); D: 4 µg DOPC/DPPE 4:1 (w/w); E: 4 µg DOPC/DPPC/cholesterol 1:1:1 (w/w); F: 4 µg DOPC/sphingomyelin/cholesterol 1:1:1 (w/w).

It is clear that isotherm sets A-E are very similar, and two key points are apparent. Firstly, in all these cases, an increase in the concentration of protein in the subphase C_B results in significant increases in both surface pressure and lipid molecular area. This is consistent with protein penetration of the monolayer. Secondly, for all monolayers except DOPC/DPPC, there is a value of C_B at which saturation of the monolayer with protein occurs, and there are no further changes to the Π-A isotherm upon increasing C_B (and one might expect to see this for DOPC/DPPC if a greater quantity of lipid were used). Any further protein added to the subphase after the interface is saturated will be in equilibrium with protein at the surface, while the quantity of lipid at the surface remains constant. In addition, for all sets of data A-E except DOPC/DPPC (again due to an insufficiency of lipid), a first-order phase transition produces a plateau in the isotherm, and the onset of this transition shifts to higher lipid molecular areas as the protein concentration is raised, which is also consistent with insertion of the protein into the monolayer.

For all the lipids except DOPC/SM/cholesterol, Π values for the mixed monolayer are greater than the sum of the Π values for similar concentrations of lipid and protein alone, suggesting that there is interaction between these lipids and RSV M because they are not behaving as separate entities in the monolayer.

Figure 3.13 shows cross-sections through the isotherms in Figure 3.12, at high areas per lipid molecule (*i.e.* outside the range of any phase transitions). Each point is the surface pressure reading for a different protein concentration at a constant area per lipid molecule. The curves drawn in red come from least-squares fitting of a modified Langmuir binding isotherm, equation *(22)*, to these data points, allowing calculation of the protein adsorption equilibrium constant K for each lipid monolayer, and also the mean molar area ω_{Σ}.

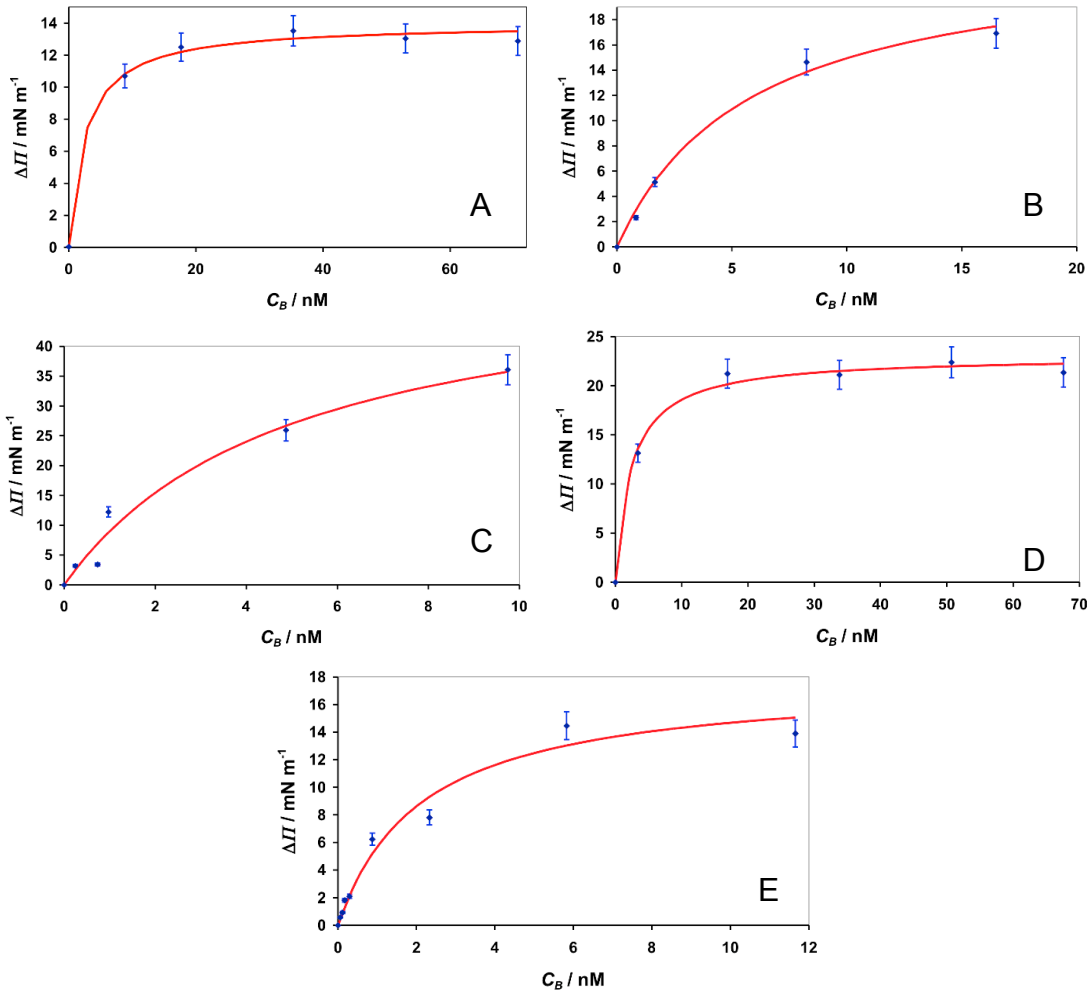


Figure 3.13: Adsorption isotherms corresponding to cross-sections through the isotherms in Figure 3.12 at constant areas per lipid molecule as a function of C_B.
Experimental data points are shown in blue and the curves in red result from fitting to equation *(22)*; error bars are 7 % (equal to the standard deviation in the maximum value of Π at equilibrium between protein and lipid – data not shown). A: DPPC, A = 160 Å^2 per lipid molecule, ω_Σ = 45.6 Å^2; B: DOPC, A = 280 Å^2 per lipid molecule, ω_Σ = 45.5 Å^2; C: DOPC/DPPC 1:1 (w/w), A = 200 Å^2 per lipid molecule, ω_Σ = 45 Å^2; D: DOPC/DPPE 4:1 (w/w), A = 150 Å^2 per lipid molecule, ω_Σ = 39 Å^2; E: DOPC/DPPC/cholesterol 1:1:1 (w/w), A = 240 Å^2 per lipid molecule, ω_Σ = 42.8 Å^2.

Lipid	C_B at surface saturation / nM	Adsorption equilibrium constant K / M^{-1}
DPPC	≥ 35	3.9×10^8
DOPC	≥ 8	1.7×10^8
DOPC/DPPC	-	2.0×10^8
DOPC/DPPE	≥ 3	4.2×10^8
DOPC/DPPC/chol	≥ 2	7.7×10^8

Table 3.5: Summary of parameters for RSV M adsorbing to lipid monolayers.

Surface saturation occurs at a much lower protein concentration for DOPC monolayers than for DPPC; 8 nM as opposed to 35 nM (Figures 3.12 A and 3.12 B, and Table 3.5). This probably reflects the fact that DOPC has two singly unsaturated acyl chains, making a DOPC monolayer significantly more amenable to lateral diffusion than one composed of DPPC. RSV M may therefore be able to insert more easily into a DOPC monolayer. The maximum change in surface pressure $\Delta\Pi$ is higher for DOPC than for DPPC (Figures 3.13 A and 3.13 B), which may result from the ability of the more fluid monolayer to accommodate greater quantities of protein.

In the case of DOPC/DPPC monolayers (Figures 3.12 C and 3.13 C), lower concentrations of lipid were used in order to allow observation of the lift-off area for each isotherm (at least at low C_B), but this also meant that the monolayers could not be compressed sufficiently to reach the maximum possible surface pressures. It is apparent from Figure 3.13 C that the maximum $\Delta\Pi$ is around 35-40 nM/m, which is approximately double the value for monolayers composed of DPPC or DOPC separately. This is most likely due to the greater errors inherent in fitting the Langmuir adsorption isotherm to the DOPC/DPPC data, where the surface saturation point was not reached. From Figure 3.12 C it is possible to envisage that the maximum pressure reached by a DOPC/DPPC monolayer could be around 70 nM/m, which is in between the two Π^{max} values for DPPC and DOPC. This suggests that the DPPC/DOPC monolayer behaves as an ideal mixture of the two components

Figures 3.12 D and 3.13 D show data from monolayers composed of a 4:1 (w/w) mixture of DOPC and DPPE. PE is a smaller headgroup than PC and thus can increase the curvature of lipid membranes. Particularly strong interactions between

RSV M and PE lipids could indicate that M preferentially binds to such lipids in order to recruit them into particular areas of the membrane so as to alter its degree of curvature, which may be important during RSV budding (see Figures 3.2 and 3.3). A cursory glance at Figure 3.12 D might suggest that monolayers of this lipid mixture become saturated particularly quickly. However, careful comparison of Figures 3.13 B and D reveals that saturation of the DOPC/DPPE monolayer occurs at a very similar protein concentration (and lipid:protein ratio) to saturation of the DOPC monolayer. The maximum surface pressure reached is 58 mN/m, and $\Delta\Pi^{max}$ is around 22 mN/m; these values are close to the corresponding values for DOPC. Therefore, the inclusion of DPPE does not appear to have a major effect upon the affinity of RSV M for the monolayer; the protein partitions into it as it does into DOPC and DPPC monolayers.

Monolayers containing DOPC/DPPC/cholesterol (Figure 3.12 E and 3.13 E) were chosen in order to provide a model of a membrane containing ordered microdomains, which are though to be targeted by RSV M during the assembly and budding of virions. These monolayer experiments were carried out at room temperatures of approximately 20ºC, under which conditions a 1:1:1 mixture of DPPC (or sphingomyelin), DOPC and cholesterol should form L_o domains composed mainly of DPPC (sphingomyelin) and cholesterol, surrounded by a more disordered L_d phase mostly made up of DOPC (Figure 3.14; also see Figure 1.14 in Chapter 1).

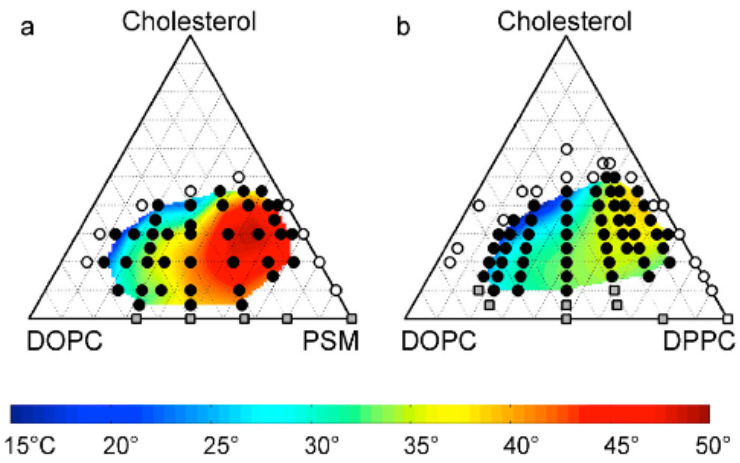


Figure 3.14: Miscibility phase diagrams for ternary mixtures of (a) DOPC/SM/cholesterol and (b) DOPC/DPPC/cholesterol at varying temperatures.[32†]

† Copyright (2005) by the American Physical Society

However, data from the DOPC/DPPC/cholesterol isotherms do not appear to be significantly different to those from the other lipids. Saturation of the surface occurs at a relatively low C_B of 2 nM, although the protein adsorption equilibrium constant K is a little higher than for the other lipids, at 7.7×10^8 M^{-1}, suggesting that the presence of lipid microdomains may slightly increase the affinity of RSV M for the monolayer. Unusually, the first order phase transition occurs at a lower pressure (60 mM/m as opposed to 70 nM/m) for $C_B \geq 6$ nM, reflecting a more condensed monolayer at these protein concentrations.

Monolayers composed of DOPC/SM/cholesterol (1:1:1), which were also expected to contain L_o microdomains, produced markedly different isotherms to the other lipid mixtures in the presence of RSV M protein (Figure 3.12 F). Three different types of behaviour appear to occur, depending on the protein concentration range. At low concentrations ($C_B \leq 0.1$ nM), the rise in pressure suggests that the protein exhibits simple partitioning to the surface in competition with the lipid. As the film is compressed the protein is expelled from the monolayer at the limiting spreading pressure of the protein (12-13 mN/m; see Section 3.2.4), leaving a film of insoluble lipid. At slightly higher concentrations (0.6 nM $\leq C_B \leq$ 2.5 nM), the protein partitions into the monolayer resulting in a first order phase transition, shown by a plateau, again at the limiting protein spreading pressure of around 13 mN/m. Notably, isotherms in this intermediate concentration range cross the lipid-only isotherm ($C_B = 0$) in the plateau region, resulting in smaller areas per lipid molecule than observed for the monolayer without protein at equivalent pressures. At the highest protein concentrations ($C_B \geq 6$ nM), saturation of the surface is seen, with a first order transition at $\Pi = 16$ mN/m. Although this is a higher pressure than that of the plateau at lower protein concentrations, the isotherms again cross the lipid-only isotherm, reflecting a reduction in the area per lipid molecule and the formation of a more condensed monolayer than is the case for the lipid alone. This may be due to extrinsic binding of protein to the monolayer, or certain areas of it, *via* specific interactions that cause the lipid molecules to pack more closely together.

It was not possible to fit the modified Langmuir binding isotherm (equation *(22)*) to data from cross-sections through the DOPC/SM/cholesterol isotherms. Equation

(22) holds as long as the protein-lipid interactions are similar for each lipid in the monolayer, which is unlikely to be true in this case given the complex behaviour described above.

3.2.7 Adsorption of RSV M to monolayers composed of mouse lung lipids

The mouse lung lipids used in this project were extracted from the lung tissue of several mice; only one extraction procedure was carried out due to time constraints, therefore the quantity of lung lipid extract available was limited. Some was saved for use in the experiments described in the next two sections, which meant that it was not possible to record a large number of isotherms using these lipids. Isotherms for a monolayer of mouse lung extract, with and without matrix protein, are shown in Figure 3.14 below.

It should be borne in mind that the extract used almost certainly contained other chloroform-soluble components of the lung tissue in addition to phospholipids, sphingolipids and cholesterol, such as cardiolipin, triacylglycerol and lyso-phosphocholine. TLC and NMR analysis of the mixture showed definitively that cholesterol was present, as well as PC lipids, and both saturated and unsaturated acyl chains were detected. No indication of the presence of sphingomyelin was observed. Because it was not possible to ascertain the exact composition of the mixture, it was assigned a nominal molecular mass of 600 Da, which seemed to be a reasonable figure given the masses of the most likely membrane components. Therefore the area per lipid molecule values are estimates, and the lipid:protein ratios are also approximations.

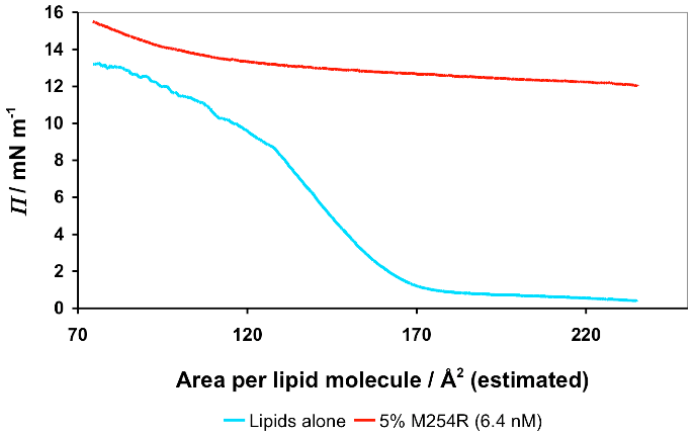


Figure 3.14: Π-A isotherms of a monolayer of mouse lung extract with and without RSV M. Percentages refer to the number of moles of protein used with respect to the (estimated) number of moles of lipid; figures in brackets give the subphase protein concentration.

The two isotherms in Figure 3.14 show that, in common with the majority of the model lipid mixtures discussed above, surface pressure readings increase with the addition of matrix protein to the subphase. There are too few data to ascertain the value of Π^{max} either with or without protein; however, from the appearance of each isotherm at low areas per lipid molecule, it seems reasonable to suppose that Π^{max} in the presence of protein will be significantly higher than in its absence, and could well be greater than the sum of the Π^{max} values for lipid and protein separately. The mouse lung extract therefore appears to interact with RSV M in a similar manner to the sphingomyelin-free lipid mixtures previously discussed, with the protein partitioning into the lipid monolayer. This would also support the finding from TLC and NMR that no sphingomyelin was present in the extract.

This does not necessarily imply that sphingomyelin is not present in mouse lung membranes, as there could have been some feature of the extraction procedure that meant sphingomyelin was not extracted from the tissue properly. Additionally, it is clear from Tables 3.1 and 3.2 earlier in this chapter that the proportion of sphingomyelin in animal lung cell membranes and lung surfactant is relatively low, and this may not have been detected by the analytical methods used. No single animal model duplicates all aspects of RSV infection, although mice have been used in many studies, and have provided useful insights into the immunology of RSV disease in particular.[33] Experimental findings using mouse lung tissue may not in fact be very close to the data that would be obtained in the same experiments if extracts

from human lung tissue were used. However, the similarity of the isotherms from this experiment to some of those from the simple lipid mixtures that were also tested, does support the use of such mixtures as reasonable models.

3.3 Brewster angle microscopy of protein-lipid monolayers

Brewster angle microscopy (BAM) provides a method for detailed imaging of monolayers at the air-water interface. It has certain advantages over other techniques used for this purpose, such as fluorescence microscopy, as it does not require marker species to be added to the system, which can alter its behaviour quite dramatically. Such fluorescent species mix more easily with fluid phases of the monolayer than with ordered ones, which is how the technique allows differentiation between different domains within the monolayer. However this also means that very highly ordered phases are difficult to see by fluorescence microscopy.[34]

In contrast, BAM allows monolayers to be seen simply by detecting the changes in the reflectivity of light from the surface caused by the presence of the surfactant. The air-water interface is an example of a Fresnel interface, which has negligible roughness and where there is a large change in refractive index between the two phases. When *p*-polarized light is shone onto such an interface, there is an angle of incidence, Brewster's angle, at which all of the light is perfectly transmitted through the interface with no reflection. At this specific angle, the dipole moments induced in the water by the incident beam point exactly in the direction of reflection.[34]

The images in this section were produced using a specially made Brewster angle microscope (Figure 3.15), which incorporated the Langmuir trough used to carry out the experiments in the previous section. *P*-polarized light from a low power helium-neon laser was shone onto the subphase surface *via* adjustable mirrors and a clean-up polarizer. A glass well was inserted into the floor of the trough to allow refracted light to escape, and reflected light was detected using a CCD video camera. This was connected to a video display unit, and to a computer program for capturing still images.

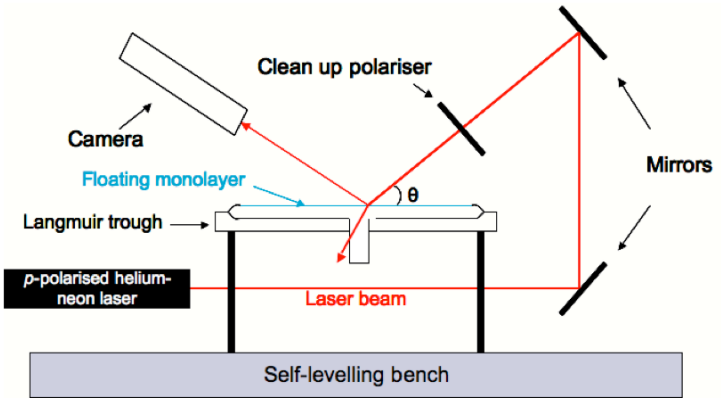


Figure 3.15: Diagram of the Brewster angle microscope used during this project. Angle θ is approximately 53°.

For an interface between air and pure water, Brewster's angle is 53°, but in this case PBS at pH 7.4 was used as the subphase. The incident angle of the polarized light was therefore adjusted slightly, from an initial angle of 53°, until no reflection was observed from the clean subphase surface. This technique allowed direct observation of RSV M interacting with the lipid monolayers described in the previous section.

Monolayers were spread as described previously; images were recorded during compression of the monolayer (see isotherms in previous section), although only part of each isotherm could be studied because from about halfway through each compression the barriers blocked the laser beam.

3.3.1 RSV M monolayers

The BAM images in Figure 3.16 below show a clean subphase surface, and monolayers composed of matrix protein alone. These pictures are included as a control to demonstrate that no features were seen at a range of protein subphase concentrations, and thus that a monolayer containing only RSV M is homogenous.

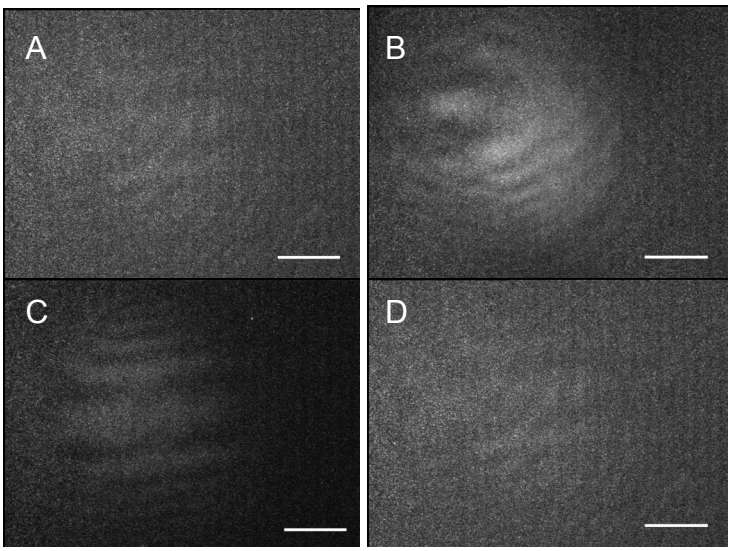


Figure 3.16: BAM images of a clean surface, and monolayers composed of RSV M alone. A: Clean surface; B: [M^{254R}] = 2.96 nM, Π = 13.5 mN m^{-1}; C: [M^{254R}] = 3.95 nM, Π = 9.4 mN m^{-1}; D: [M^{254R}] = 0.987 nM, Π = 0.5 mN m^{-1}. Scale bars are 500 µm.

3.3.2 Adsorption of RSV M to lipid monolayers

From the Π-A isotherms described in the previous section, it is clear that the RSV M protein interacts in a markedly different way with monolayers formed from DOPC/DPPC/cholesterol and DOPC/SM/cholesterol. Brewster angle microscopy was therefore conducted on films composed of these lipids in order to study their mixing behaviour in the presence of RSV matrix protein, in particular with regard to the formation of phase separated microdomains. BAM was also used to image films containing only PC lipids, for comparison. Samples of the images produced are shown in Figures 3.17 (DPPC and DOPC/DPPC), 3.18 (DOPC/DPPC/cholesterol) and 3.19 (DOPC/SM/cholesterol). Areas of condensed phase appear lighter and expanded phases are darker.

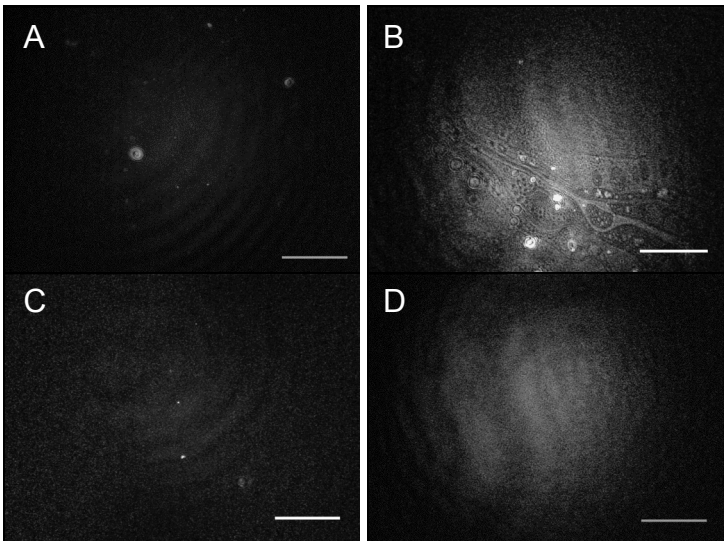


Figure 3.17: BAM images from compression isotherms of PC monolayers.
A and B: Lipid only films; A: DPPC ([lipid] = 102 nM), Π = 0 mN m^{-1}; B: DOPC/DPPC (1:1, [lipid] = 99.3 nM) Π = 0.3 mN m^{-1}; C and D: Mixed lipid/RSV M films, corresponding to films in A and B with the addition of protein; C: C_B = 5.14 nM, Π = 26.5 mN m^{-1}; D: C_B = 4.96 nM, Π = 20.1 mN m^{-1}. All scale bars are 500 µm.

Figure 3.17 demonstrates the lack of texture in films composed of PC lipids and RSV M, showing that the protein penetrates evenly into all regions of the monolayer. Some low-contrast phase separation is evident in the DOPC/DPPC film, but this disappears on interaction with the protein.

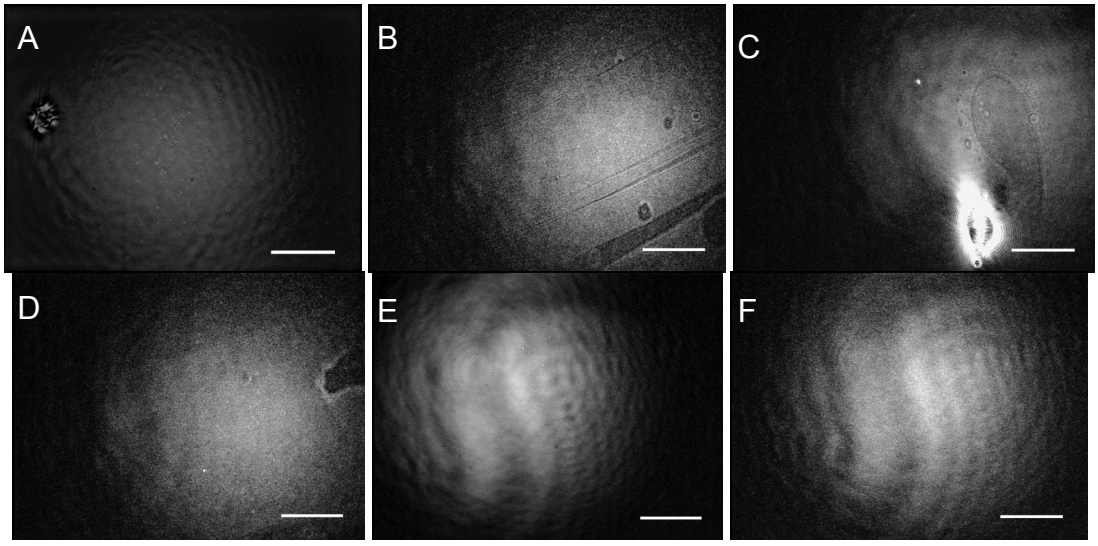


Figure 3.18: BAM images from compression isotherms of DOPC/DPPC/chol monolayers (1:1:1, 4 µg / 116 nM). A-D: Lipid only film; A: Π = 3.0 mN m^{-1}; B: Π = 4.5 mN m^{-1}, C: Π = 0.4 mN m^{-1}; D: Π = 9.3 mN m^{-1}; E-F: Mixed lipid/RSV M films; E: C_B = 1.16 nM, Π = 18.8 mN m^{-1}; F: C_B = 5.8 nM, Π = 22.3 mN m^{-1}. All scale bars are 500 µm.

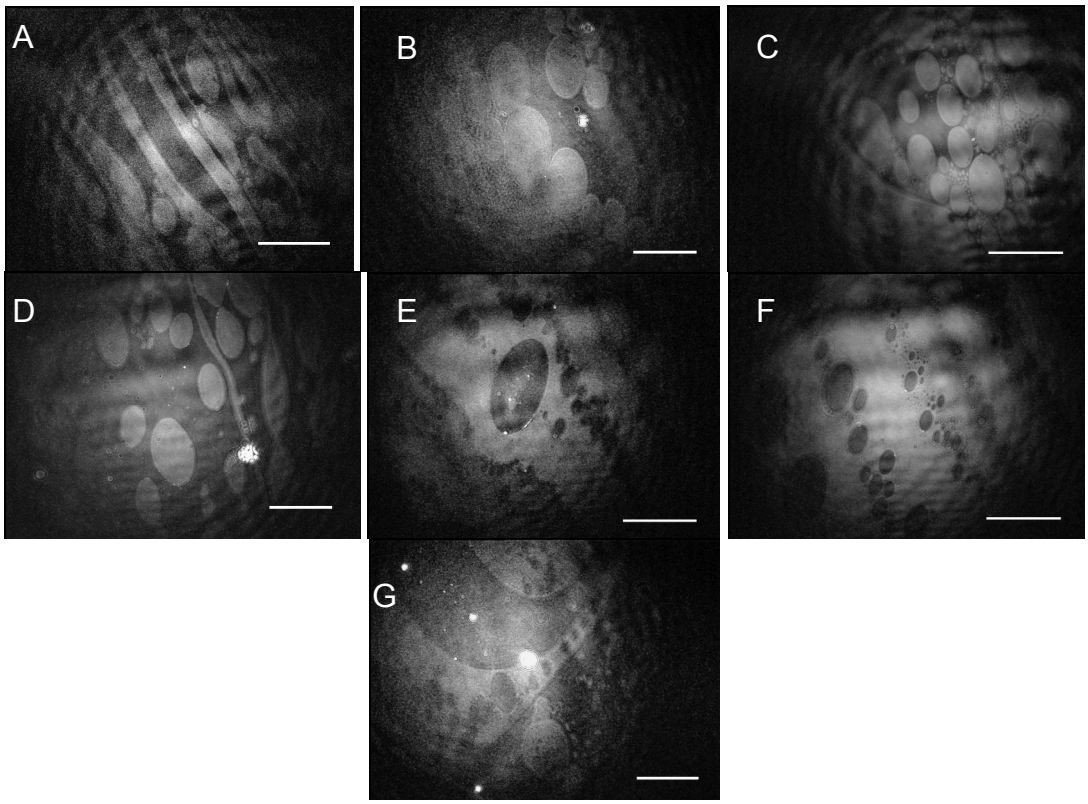


Figure 3.19: BAM images from compression isotherms of DOPC/SM/chol monolayers (1:1:1, 4 µg / 122 nM). A-C: Lipid only films; A: Π = -1.7 mN m^{-1} (prior to compression); B: Π = 0.1 mN m^{-1} C: Π = 4.1 mN m^{-1}; D-G: Mixed lipid/RSV M films (C_B = 5.9 nM); D: Π = 28.8 mN m^{-1}; E: Π = 23.1 mN m^{-1}; F: Π = 24.1 mN m^{-1}; G: Π = 24.8 mN m^{-1}All scale bars are 500 µm.

Films of DOPC/DPPC/cholesterol displayed little evidence of inhomogeneity at the pressures used, irrespective of the presence or absence of the protein (Figure 3.18), which is consistent with the sub-micron size of L_o domains expected in this mixture.[35] The elongated regions of expanded phase in Figure 3.18 B, C and D are seen because the monolayer had not been allowed sufficient time to equilibrate before the images were recorded.

In the case of DOPC/SM/cholesterol films, at low pressures and in the absence of protein, a mixture of fluid and more condensed phases is observed (Figure 3.19 A and B), the latter consisting of circular domains with diameters up to 400 µm. These domains become smaller and of higher contrast as the pressure is increased (Figure 3.19 C). In the presence of RSV M, the circular condensed phase domains persist at low pressures (Figure 3.19 D) but at higher pressures corresponding to the plateau region of the isotherms, extended areas of condensed phase are seen (Figures 3.19 E and 3.19 F). Following further compression to high surface pressures these extended

networks of condensed phase dissipate, resulting in more irregular patterns of phase separation (Figure 3.19 G).

These images indicate that at low pressures a mixture of DOPC/SM/cholesterol forms phase-separated domains that are not disrupted by the presence of RSV M. At higher pressures, smaller domains are formed in the absence of protein, and in the presence of protein these coalesce into extended networks of condensed phase, which eventually collapse into irregular patterns upon prolonged compression.

3.3.3 Adsorption of RSV M to monolayers composed of mouse lung lipids

BAM was conducted on monolayers composed of the chloroform-soluble components of mouse lung cells, to ascertain the behaviour of a natural lipid mixture when compared with the properties of the model lipid mixtures discussed above. Representative BAM images for monolayers of mouse lung extract are shown in Figure 3.20.

The quality of the images is not quite as high as for the images of model lipid mixtures, but it is apparent that there are similarities. In the absence of protein and at low pressures (Figure 3.20 A), a lighter condensed phase and a darker expanded phase can clearly be distinguished, even though the elongated shapes of the phases indicate that they are not quite in equilibrium. At a higher surface pressure (Figure 3.20 B) small circular condensed domains with diameters up to 100 μm can be seen, and the monolayer has an appearance similar to that of the DOPC/SM/cholesterol film.

In the presence of protein the monolayer was difficult to image clearly at low pressures, but at intermediate pressures (Figure 3.20 C and D) very small domains of condensed phase were observed. At higher pressures these expand into larger areas (Figure 3.20 E), which eventually form an extended network of condensed phase after prolonged compression (Figure 3.20 F). Monolayers of the chloroform-soluble

components of mouse lung cells therefore seem to behave in a similar way to monolayers of the DOPC/SM/cholesterol mixture.

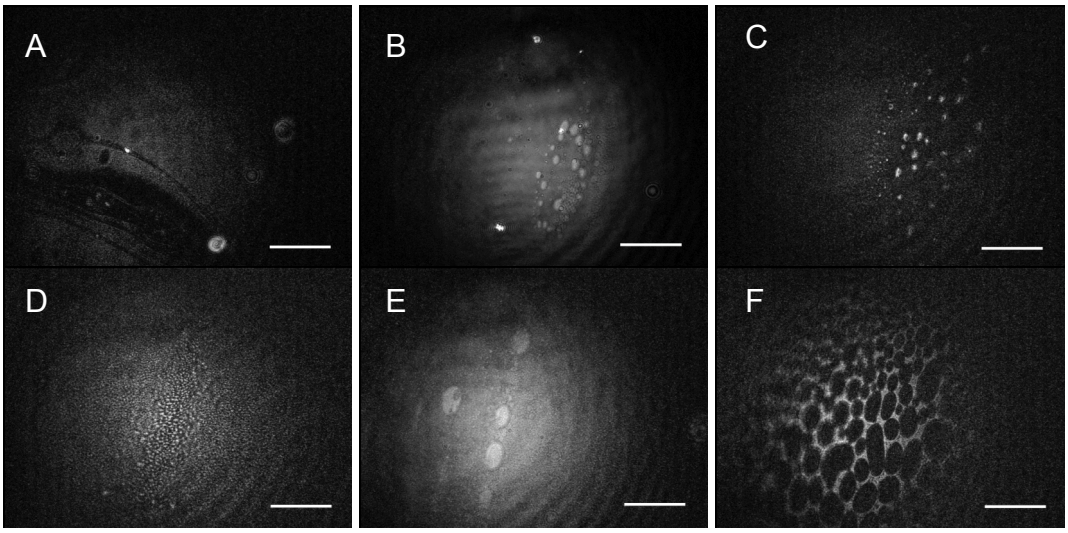


Figure 3.20: BAM images from compression of monolayers of mouse lung extract (4 µg / approx. 123 nM). A-B: Lipid-only films; A: Π = 0.4 mN m^{-1}; B: Π = 10 mN m^{-1}; C-G: Mixed lipid/RSV M films (CB = 6.4 nM); C: Π = 10.5 mN m^{-1}; D: Π = 11.6 mN m^{-1}; E: Π = 13.1 mN m^{-1}; F: Π = 13.3 mN m^{-1}; G: Π = 5.5 mN m^{-1}. All scale bars are 500 µm.

3.4 Atomic force microscopy of supported protein-lipid films

In order to gain further insight into the arrangement of lipids and protein within the monolayer systems discussed in the previous section, and to more fully characterize the phase-separated domains seen by BAM, protein-lipid films were deposited onto a solid support for examination by atomic force microscopy (AFM). This technique involves passing a cantilever with a very sharp tip across the surface of interest; the displacement of the cantilever by intermolecular forces is measured using a laser beam, giving a spatial resolution of around 0.1 nm vertically, and several nanometres in the lateral dimension.[36] AFM can be carried out in either contact or tapping mode: in contact mode a constant force between the tip and the sample is maintained by keeping the deflection of the tip constant. In tapping mode, the cantilever is made to oscillate near its resonance frequency, and intermolecular forces cause the amplitude of the oscillation to decrease as the tip approaches the sample surface. The height of the tip is adjusted so as to maintain a set oscillation amplitude whilst the tip is

scanned over the sample, and the image is produced by measuring the force between the cantilever and the surface. Tapping mode is therefore less damaging to delicate samples than contact mode, where the tip is dragged over the surface with constant force, and so this method is much more suitable for biological samples.

Tapping mode AFM can be carried out in solution, keeping the environment of biological samples as close as possible to their surroundings *in vivo*, and this type of AFM imaging has been widely used to study the distribution of domains within lipid monolayers and bilayers.[36-38] However in this case the underside of the lipid films was of most interest, as this is where the protein is proposed to bind peripherally. Therefore protein-lipid films were deposited at constant pressure onto an oxidised silicon surface that had been silanized by treatment with octadecyltrichlorosilane,[39] in such a manner that the face of the film in contact with the subphase became the uppermost surface of the sample. AFM images of protein-lipid films containing DOPC/DPPC/cholesterol (Figure 3.21) and DOPC/SM/cholesterol (Figures 3.22 and 3.23) are shown below.

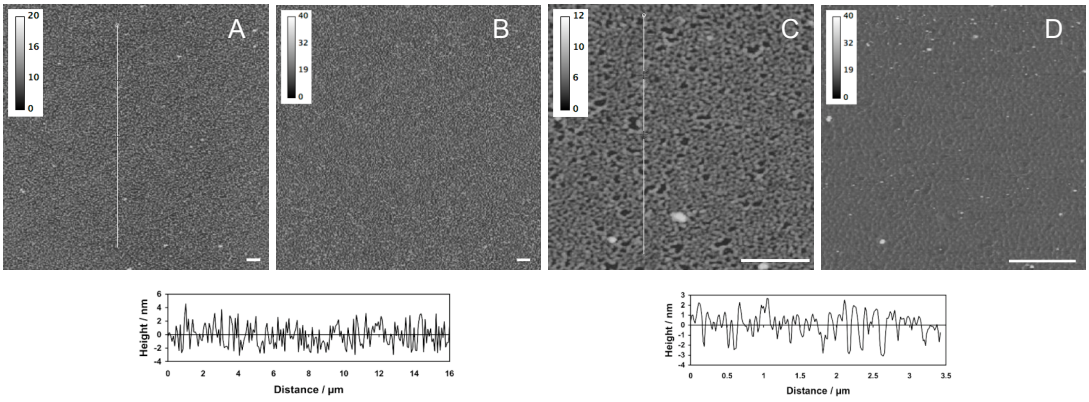


Figure 3.21: Tapping mode AFM height images of surfaces prepared by transfer of monolayers of DOPC/DPPC/cholesterol (1:1:1, [lipid] = 62.9 nM) and RSV M (C_B = 3.2 nM) spread on PBS subphases at 20°C and pH 7.4, to modified silicon surfaces.
Surfaces correspond to the face of the monolayer in contact with the subphase. The scale bars for the *z*-axis, in nm or degrees for height or phase images respectively, are in the top left corner of each image and scale bars of length 0.5 µm for the *x*- and *y*-axes are in the bottom right corner. Sections through the data are shown below the corresponding image. A and C: Height scans following deposition at Π = 29 mN/m. B and D: Phase images corresponding to A and C.

The surface profiles from the two lipid mixtures appear noticeably different. Deposition of DOPC/DPPC/cholesterol with RSV M at Π = 29 mN/m gave surfaces with relatively low roughness values (R_q = 1.0 ± 0.1 nm, R_a = 0.8 ± 0.1 nm) and a

uniform appearance, as shown by both the height (Figure 3.21 A) and phase (Figure 3.21 B) profiles. At higher resolution (Figure 3.21 C), pinhole defects can be seen in the film that allowed its thickness to be measured as 3.3 ± 1.6 nm.

Films deposited from DOPC/SM/cholesterol monolayers at a surface pressure of 8.8 nM/m posess some regions (Figures 3.22 A and 3.22 E) with a uniform distribution of surface features that have a similar pinholed appearance, and depth profile (4.0 ± 1.1 nm), to the DOPC/DPPC/cholesterol films, although the texture of the surface is rougher (R_q = 1.7 ± 0.2 nm, R_a = 1.4 ± 0.1 nm). Other regions of the samples appear quite different, however (Figures 3.22 B and 3.22 F), with denser surface coverage, an even rougher texture (R_q = 2.1 ± 0.3 nm, R_a = 1.7 ± 0.2 nm) and a much less even depth profile, with some features 6.5 ± 2.5 nm in height. The phase image (Figure 3.22 F) however is rather uniform, implying that there is a much higher density of material on the surface than is apparent from the height image.

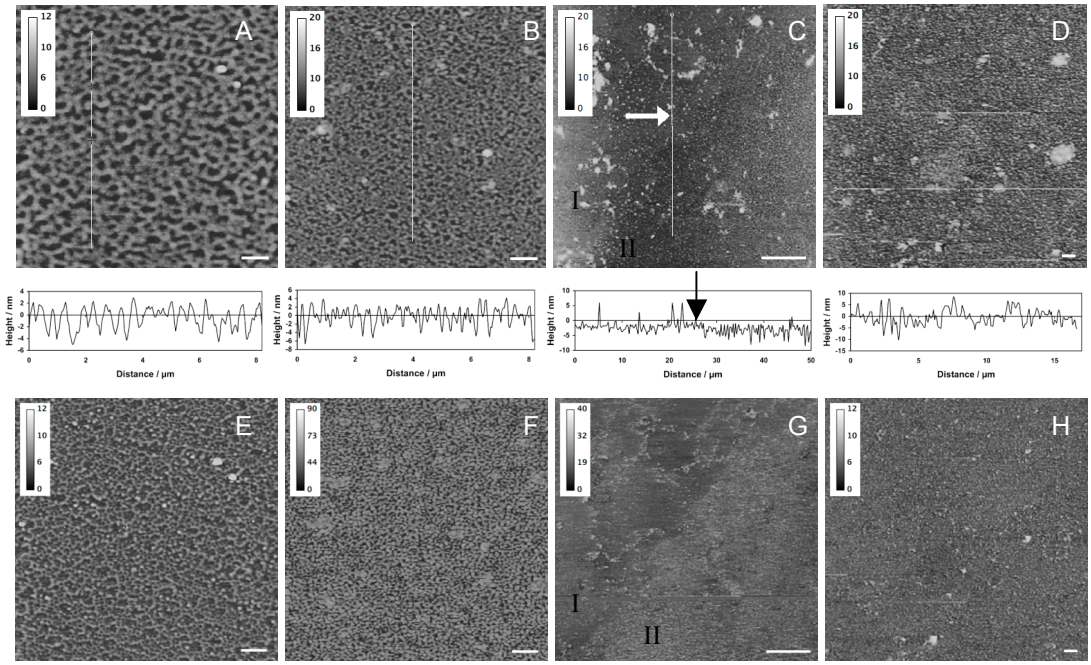


Figure 3.22: Tapping mode AFM images of surfaces prepared by transfer of monolayers of DOPC/SM/cholesterol (1:1:1, [lipid] = 65.8 nM) and RSV M (C_B = 6.6 nM) spread on PBS subphases at 20°C and pH 7.4, to modified silicon surfaces.
Surfaces correspond to the face of the monolayer in contact with the subphase. The scale bar (nm) for the *z*-axis, in nm or degrees for height or phase images respectively, are in the top left corner of each image and scale bars of length 0.5 µm (5 µm for C and G) for the *x*- and *y*-axes are in the bottom right corner. Sections through the data are shown below the corresponding image. A and B: Height scans following deposition at 8.8 mN/m; C: Height scan following film deposition at 17.3 mN/m. Two regions labelled I and II can be discerned; D: Height scan following deposition at 17.3 mN/m. E-H: Phase images corresponding to A-D.

A repeat experiment also revealed two distinct surface textures that can be seen particularly clearly in phase images (Figures 3.22 C and 3.22 G, where the two phases are labelled I and II). The surface in Figures 3.22 C and 3.22 G appears more disordered than the areas shown in Figures 3.22 A and 3.22 B, but even so there is a clear height difference of around 3 nm between regions I and II, which is consistent with the difference in feature depths between Figures 3.22 A and 3.22 B. The more densely covered region (I) contains circular features (white arrow in Figure 3.22 C) that would be consistent with this region being formed by annealing of separate microdomains at the surface. The area surrounding these circular features has a raised profile (arrow on height profile for Figure 3.22 C) but possesses similar viscoelastic properties to the material in region II, suggesting that the raised areas represent material that has been expelled during annealing.

Together, these data suggest that in this pressure range the protein is incorporated into L_d regions of the monolayer, inserting in between lipid molecules (Figures 3.22 A and 3.22 C region II), but binds extrinsically to phase-separated L_o domains (Figures 3.22 B and 3.22 C region I). Such a model would explain the depth profiles in these areas; the depths of the L_d phase are comparable to the thickness of the DOPC/DPPC/cholesterol film, while the L_o phase is deeper by 2.5-3 nm. This depth increase will be partially due to the protein, with the rest arising from the approximately 1 nm increase in height of L_o domains compared to L_d domains.[36, 40]

Films deposited from DOPC/SM/cholesterol at 17.3 mN/m produced surfaces that were consistently rough (R_q = 2.8 ± 0.4 nm, R_a = 2.2 ± 0.3 nm, Figures 3.22 D, 3.22 H and 3.23 A-H), with no pinhole defects (just as one might expect at higher pressures). Several areas of heterogenous film can be seen, containing features of height 5-10 nm. In some images features of a similar height appear to take the form of extended linear structures (Figures 3.23 A and 3.23 B). This would be explained by a structural rearrangement of RSV M upon compression that results in one domain being expelled from the interface.

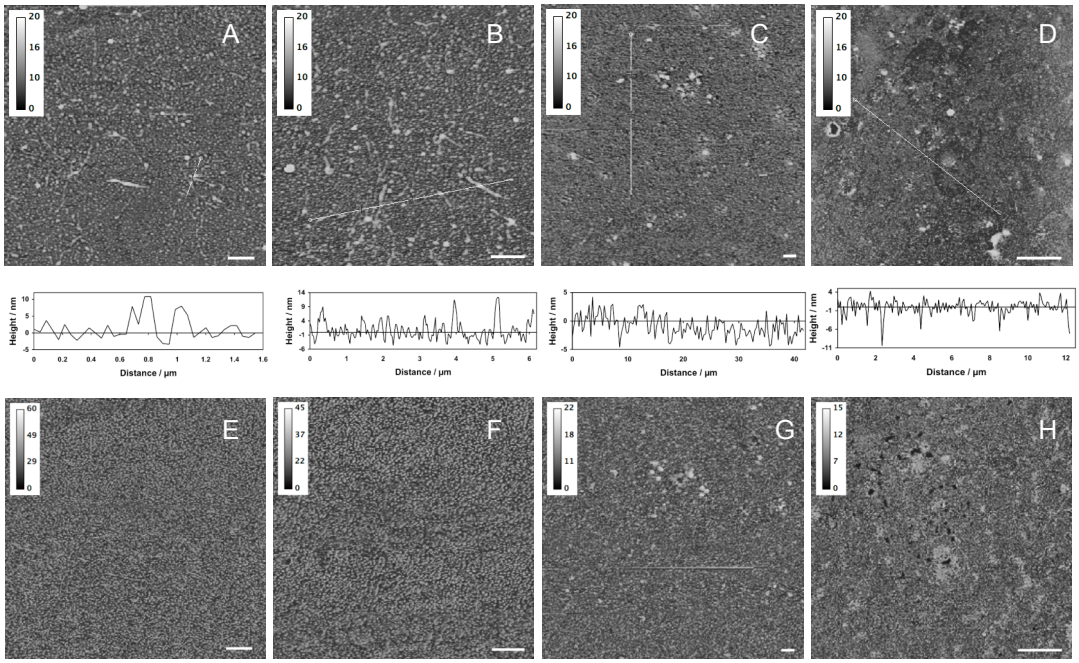


Figure 3.23: Additional tapping mode AFM images of surfaces prepared by transfer of monolayers of DOPC/SM/cholesterol (1:1:1, [lipid] = 65.8 nM) and RSV M (C_B = 6.6 nM) spread on PBS subphases at Π = 17.3 mN/m, 20°C and pH 7.4, to modified silicon surfaces. Surfaces correspond to the face of the monolayer in contact with the subphase. The scale bar (nm) for the *z*-axis, in nm or degrees for height or phase images respectively, are in the top left corner of each image and scale bars of length 0.5 µm (5 µm for D and H) for the *x*- and *y*-axes are in the bottom right corner. Sections through the data are shown below the corresponding image. A-D: Height scans; E-H: Phase images corresponding to A-D. A and B are scans of a different surface region to Figure 3.22 C. C-D are from a repeat experiment but show similar features to Figure 3.22 C.

3.5 References

1 T. P. W. McMullen, R. Lewis and R. N. McElhaney, *Current Opinion in Colloid & Interface Science*, 2004, **8**, 459-468.

2 T. M. Devlin, in *Textbook of Biochemistry with Clinical Correlations*, ed. T. M. Devlin, Wiley-Liss, New York, 5 edn., 2002.

3 P. Palestini, C. Calvi, E. Conforti, L. Botto, C. Fenoglio and G. Miserocchi, *American Journal of Physiology: Lung Cellular and Molecular Physiology*, 2002, **282**, L1382-L1390.

4 H. Dombrowsky, G. T. Clark, G. A. Rau, W. Bernhard and A. D. Postle, *Pediatric Research*, 2003, **53**, 447-454.

5 F. D. Grunstone, J. L. Harwood and F. B. Padley, *The Lipid Handbook*, Chapman and Hall, London, 1995.

6 X. Liang and Y. Huang, *The International Journal of Biochemistry & Cell Biology*, 2002, **34**, 1248-1255.

7 H. T. McMahon and J. L. Gallop, *Nature*, 2005, **438**, 590-596.

8 J. Solon, O. Gareil, P. Bassereau and Y. Gaudin, *Journal of General Virology*, 2005, **86**, 3357-3363.

9 S. N. Ahmed, D. A. Brown and E. London, *Biochemistry*, 1997, **36**, 10944-10953.

10 G. Henderson, J. Murray and R. P. Yeo, *Virology*, 2002, **300**, 244-254.

11 S. Bavari, C. M. Bosio, E. Wiegand, G. Ruthel, A. B. Will, T. W. Gisbert, M. Hevey, C. Schmaljohn, A. Schmaljohn and M. J. Aman, *Journal of Experimental Medicine*, 2002, **195**, 593-602.

12 G. Brown, J. Aitken, H. W. M. Rixon and R. J. Sugrue, *Journal of General Virology*, 2002, **83**, 611-621.

13 A. Marty, J. Meanger, J. Mills and R. Ghildyal, *Archives of Virology*, 2004, **149**, 199-210.

14 G. Brown, C. E. Jeffree, T. McDonald, H. W. M. Rixon, J. D. Aitken and R. J. Sugrue, *Virology*, 2004, **327**, 175-185.

15 J. P. Laliberte, L. W. McGinnes, M. E. Peeples and T. G. Morrison, *Journal of Virology*, 2006, **80**, 10652-10662.

16 R. Maget-Dana, *Biochimica et Biophysica Acta-Biomembranes*, 1999, **1462**, 109-140.

17 V. M. Kaganer, H. Möhwald and P. Dutta, *Reviews of Modern Physics*, 1999, **71**, 779-819.

18 R. M. Swart, in *Langmuir-Blodgett Films*, ed. G. Roberts, Plenum Press, New York, 1990.

19 R. A. Hann, in *Langmuir-Blodgett Films*, ed. G. Roberts, Plenum Press, New York, 1990.

20 G. L. Gaines, *Insoluble Monolayers at Liquid-Gas Interfaces*, Wiley Interscience, New York, 1966.

21 M. C. Petty, *Langmuir-Blodgett films: an introduction*, Cambridge University Press, Cambridge, 1996.

22 B. A. Pethica, *Transactions of the Faraday Society*, 1955, **51**, 1402-1411.

23 D. Vollhardt and V. B. Fainerman, *Advances in Colloid and Interface Science*, 2000, **86**, 103-151.

24 J. Eisenblätter and R. Winter, *Biophysical Journal*, 2006, **90**, 956-966.

25 Phase transition temperatures for glycerophospholipids, Avanti Polar Lipids (www.avantilipids.com/index.php?option=com_content&view=article&id=1700&itemid=419).

26 D. Marsh, *Biochemica et Biophysica Acta*, 1996, **1286**, 183-223.

27 A. Krishnan, C. A. Siedlecki and E. A. Vogler, *Langmuir*, 2003, **19**, 10342-10352.

28 R. R. Niño, C. C. Sánchez, M. C. Fernández and J. M. R. Patino, *Journal of the American Oil Chemists Society*, 2001, **78**, 873-879.

29 A. Krishnan, Y.-H. Liu, P. Cha, D. Allara and E. A. Vogler, *Journal of the Royal Society Interface*, 2006, **3**, 283-301.

30 O. Fontvila, C. Mestres, M. Muñoz, I. Haro, M. A. Alsina and M. Pujol, *Colloids and Surfaces A-Physicochemical and Engineering Aspects*, 2008, **321**, 175-180.

31 P. Neumann, D. Lieber, S. Meyer, P. Dautel, A. Kerth, I. Kraus, W. Garten and M. T. Stubbs, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 3710-3715.

32 S. L. Veatch and S. L. Keller, *Physical Review Letters*, 2005, **94**, 148101.

33 L. G. Byrd and G. A. Prince, *Clinical Infectious Diseases*, 1997, **25**, 1363-1368.

34 M. M. Hossain, K.-I. Iimura and T. Kato, *Journal of Colloid and Interface Science*, 2008, **319**, 295-301.

35 J. Duggan, G. Jamal, M. Tilley, B. Davis, G. McKenzie, K. Vere, M. G. Somekh, P. O'Shea and H. Harris, *European Biophysics Journal*, 2008, **37**, 1279-1289.

36 L. J. Johnston, *Langmuir*, 2007, **23**, 5886-5895.

37 C. Yuan, J. Furlong, P. Burgos and L. J. Johnston, *Biophysical Journal*, 2002, **82**, 2526-2535.

38 H. A. Rinia, M. M. E. Snel, J. P. J. M. van der Eerden and B. de Kruijff, *FEBS Letters*, 2001, **501**, 92-96.

39 S. M. D. Watson, K. S. Coleman and A. K. Chakraborty, *ACS Nano*, 2008, **2**, 643-650.

40 E. Keating, L. Rahman, J. Francis, A. Petersen, F. Possmayer, R. Veldhuizen and N. O. Petersen, *Biophysical Journal*, 2007, **93**, 1391-1401.

4 Conclusions

There has been relatively little previous research that addresses the interaction between matrix proteins and membranes in a quantitative fashion. The work described in this thesis has utilised a range of techniques, including some that are more usually employed in physical chemistry, to provide quantitative data regarding the binding of RSV M to specific membrane components. This, together with the determination of the first crystal structure of RSV M, has enabled new light to be shed upon several aspects of the protein. The main conclusions that can be drawn are summarised below.

4.1 Structure

The elucidation of the protein's structure by Dr. Victoria Money[1] showed that it bears some similarity to the structure of the Ebola virus matrix protein VP40,[2] having two domains linked by an unstructured region. RSV M also possesses a large positively charged patch on its surface, in common with VP40, influenza virus M1, VSV M, and Borna disease virus M.[2-6] Such areas of surface charge have been implicated in binding of matrix proteins to membranes, particularly those containing negatively charged lipids.[6, 7] The crystal structure of RSV M also revealed a negatively charged lobe on the N-terminal domain that could potentially interact with a positively charged pocket on the C-terminal domain to mediate protein-protein binding, or take part in interactions with other binding partners. Protein cross-linking experiments produced dimers, tetramers and hexamers of M, suggesting that these are important during self-assembly of the protein *in vivo*, consistent with oligomerization studies on Ebola VP40 and BDV M where dimers, tetramers, hexamers and octamers were isolated.[6, 8, 9]

CD spectra of both the native (M^{254M}) and mutated (M^{254R}) forms of RSV M revealed that the protein contains significantly less ordered secondary structure in solution than is apparent from the crystal structure. This may imply that assembly of protein monomers and/or membrane binding during virion formation incurs an entropic

penalty due to protein folding. Large enthalpic contributions from strong electrostatic interactions would counteract this negative entropy to give a favourable free energy of binding. Such a situation would be favoured by the extensive positive patch on the surface of RSV M, which provides a large potential interaction surface area.

4.2 Interactions with membrane lipids

Several experiments described herein have together provided a more detailed picture than was previously available of the way RSV M interacts with membranes. Cholesterol-enriched domains within membranes have been proposed to provide a platform for virion budding by concentrating viral components to specific locations on the plasma membrane.[10, 11] It has been suggested that the isolation of viral matrix proteins in conjunction with detergent-resistant membrane fractions and cellular marker proteins, indicates that matrix proteins interact with such membrane microdomains.[12-15]

Using lipid monolayers to model plasma membranes, this thesis demonstrates that RSV M can associate with lipids in the absence of other viral components such as the F protein or viral glycoproteins, which is consistent with earlier observations.[12] RSV M does appear to interact very differently with certain monolayers containing lipid microdomains, but it does not seem to be the domains *per se* that cause this. The protein partitions into monolayers composed of PC, PC/PE and PC/cholesterol, as demonstrated by the pressure and area increases in lipid isotherms, and supported by the relatively uniform surface textures seen in AFM images of DOPC/DPPC/cholesterol films, and the low level of phase separation observed by BAM. The pressure-area isotherms show that the protein stays in the monolayer above its equilibrium pressure of 12.3 mN/m, and well beyond the pressure of 35 mN/m, which is considered to be representative of the pressure within cell membranes.[16, 17] This, along with the increased surface pressures in mixed protein-lipid monolayers when compared to the pressure readings expected for either lipid or protein alone, provides evidence that interactions exist between the protein and the

lipids. The inclusion of DPPE does not appear to alter the behaviour of the protein in relation to the monolayer in these experiments.

When sphingomyelin is included in the lipid mixture, the protein interacts with the monolayer in two separate ways. It appears to penetrate into the L_d domains as described above for monolayers without sphingomyelin, causing pressure and area increases in Π-A isotherms and producing surface textures in AFM images similar to those seen for PC monolayers. At the same time, extrinsic binding to L_o domains produces isotherms where, after monolayer compression, the protein-lipid monolayer occupies smaller areas per molecule than does a monolayer of lipid alone. These results can be explained by specific interaction of extrinsically bound RSV M with L_o domains that result in reduced areas per lipid molecule, together with expulsion of the protein from L_d domains on compression. Since it is known that mixtures of DOPC/DPPC/cholesterol also form L_o domains, this implies that RSV M has specific interactions with sphingomyelin, rather than simply binding to L_o domains. In lipid mixtures such as DOPC/SM/cholesterol, it has been shown that the sphingomyelin does not reside entirely in the L_o domains, with up to 10 % being found in L_d domains.[18] A low concentration of SM in these regions would be enough to explain the loss of protein from the L_d phase, as the monolayer area is reduced and the density of sphingomyelin increases. The differences between the Π-A isotherms at different protein concentrations may be explained by alterations in the relative occupancy of the available protein adsorption sites in each phase in different concentration ranges.

Incubation of RSV M with these same lipid mixtures gave additional interesting results. When the protein was incubated with DOPC/DPPE over long periods of time, extended (up to several μm long) helical assemblies of uniform diameter were produced. In many cases partially formed structures appeared to extend from the surface of liposomes, showing that lipid is a key requirement for their development. Similar assemblies were less frequently produced in the presence of DOPC/DPPC/cholesterol and DOPC/SM/cholesterol mixtures. These structures share a notable resemblance to helical assemblies that have been described following removal of the viral enveloped from filamentous RSV virions,[19] and demonstrate that RSV M has an inherent tendency to self-assemble into helical arrays of this nature.

When considered in conjunction with the binding data from lipid monolayers, it is apparent that this assembly must occur from protein that is integrated into the membrane, rather than extrinsically bound. These data support a model in which formation of the nascent assembly removes protein from the membrane, disturbing the binding equilibria, and leading to partitioning of more protein into the membrane from solution to provide the materials for further assembly. In this case PE would only be required to promote negative curvature at the site of budding. This model contrasts with previous suggestions that such helical assemblies form upon vesicle fusion, which was thought to provide the lipid and protein needed for assembly.[20]

Cholesterol depletion studies have been used to suggest that cholesterol is a necessary component in the formation of viral filaments.[21] The results presented here do not necessarily contradict such findings, since RSV M does partition into membranes containing cholesterol. However, the increased incidence of highly ordered assemblies forming from membranes containing DPPE indicates that its presence is key, and also that membrane microdomains are not required for RSV M to self assemble into filaments. It is not yet clear why DPPE appears to be necessary for the formation of these helical assemblies, while its inclusion in lipid mixtures had no apparent effect in the monolayer experiments discussed above; the different conditions (of temperature *etc.*) under which the two sets of experiments were conducted may have had some effect, for example by altering lipid distribution in the model membranes.

These data suggest a potential model for the interaction of RSV M with membranes *in vivo*. As M protein is synthesised in the host cell, it will be incorporated into sphingomyelin-deficient areas of membrane until saturation occurs, and accumulate peripherally on SM-rich membranes including L_o domains. A specific binding interaction with membranes rich in SM, as has been demonstrated, would provide a mechanism for targeting the site of viral assembly to such membrane regions. In fact RSV assembly has been found to occur in the apical region of HEp-2 cells, although of the F, SH and G viral proteins studied, none appear to have any particular affinity for apical regions of the cell.[22] Other viral components, including the M protein, are therefore potentially responsible for targeting. Apical membranes contain high levels of sphingomyelin,[23] and although there are usually greater quantities in the

outer membrane leaflet than in the inner, exact levels of SM in the inner leaflet during RSV infection are not well characterized. The possibility that specific binding of M to SM performs a role in targeting RSV assembly to apical regions therefore cannot be ruled out, and would be supported by the work presented here.

The presence of cholesterol in the viral envelope has been shown to be essential for the infectivity of other members of the *Paramxyoviridae*,[24] but this may be simply because cholesterol is necessary to form L_o domains that contain sufficient sphingomyelin to allow peripheral binding of the matrix protein.

4.3 Future work

A simple way to further support the conclusions set out above would be to carry out CD spectroscopy of RSV M in the presence of liposomes containing SM, which would reveal any change in the protein's secondary structure upon binding to sphingomyelin, providing more evidence of a specific interaction between M and sphingomyelin. The use of fluorescence microscopy, and fluorescently labelled lipids and/or protein, would add to the information already obtained through work on lipid monolayers, by allowing specific species to be located within the monolayer during imaging. This would, for example, confirm whether M partitions into L_d domains and is expelled upon compression.

It would also be very interesting to repeat the monolayer and incubation experiments using DOPS in place of DPPE. DOPS is a major component of cell membranes (see Tables 3.1-3.4 in Chapter 3) and has a similarly small headgroup to DPPE, but is negatively charged. Since matrix proteins are proposed to bind preferentially to negatively charged lipids, and given the large positively charged patch apparent on the surface of the RSV M crystal structure, specific binding to DOPS might be expected. If mixtures including DOPS behaved in a similar manner to those containing DPPE, especially upon incubation with M, this might provide further evidence that it is the small headgroup of such lipids, and the corresponding increase in membrane negative curvature, that is important for the formation of helical assemblies.

Additionally, the question of the protein's conformation during membrane binding remains. The crystal structure of the protein, with two clearly defined domains linked by an unstructured region that is almost certainly flexible, provides great potential for conformational change. There is some evidence that such a change occurs during binding, for example some of the AFM images of DOPC/SM/cholesterol-protein films display surface features that would be consistent with one domain of the protein being expelled from the monolayer under certain conditions. One method that may be useful in determining the movement of the N- and C-terminal domains under various conditions is fluorescence resonance energy transfer (FRET) imaging. FRET is a distance-dependent process where energy is transferred from an excited fluorophore (the donor) to another fluorophore (the acceptor) *via* long-range dipole-dipole coupling.[25] Changes in the fluorescence emitted from each fluorophore allow the distance between them to be estimated. FRET imaging is often used to study intermolecular interactions, for example between two proteins, but in this case the donor species could be positioned on one domain of the protein, with the acceptor on the other. This would allow detection of any movement of the domains in relation to one another during a conformational change.

From the crystal structure there is a convenient tryptophan residue in the N-terminal domain (position 35) opposite a serine in the C-terminal domain (position 198), which could be modified to a cysteine. Although M already has 5 cysteines, none are disulphide bridged and only one is conserved throughout the *Pneumovirinae*. It is therefore feasible to convert these existing cysteines to serines prior to changing ser198 to cysteine. Analysis of the modified protein by CD would reveal any significant changes in secondary structure, and testing in lipid binding assays would ensure its functionality was retained. The sole remaining cysteine could then be tagged with an acceptor moiety such as 5-({[(2-iodoacetyl)amino]ethyl}amino)naphthalene-1-sulfonic acid (1, 5-IAEDANS) and utilized in fluorescence spectroscopy experiments. For example, modified M, with and without the fluorescence acceptor as a baseline control, could be added to giant unilamellar vesicles (to mimic a membrane's inner surface), and any changes in fluorescence would act as a surrogate marker for domain movement.

The use of neutron reflectivity would complement the BAM work previously carried out, as well as the FRET experiments described above. In neutron reflectivity a highly collimated neutron beam is shone onto the surface of interest, and the intensity of reflected radiation is measured as a function of wavelength. If such an experiment were performed on a supported lipid-protein film, the resulting reflectivity profile would give detailed information about the film structure, including its thickness and density, with far less potential for sample damage than is the case for AFM or X-ray reflectometry. This would help to determine how far RSV M penetrates into monolayers, and whether one domain is expelled under certain conditions.

Another experiment that should be performed is to express the two domains of RSV M separately, and repeat some of the work described here using the individual domains. This would help to determine whether one domain is absolutely necessary for membrane binding, as was found to be the case for the C-terminal domain of Ebola VP40.[7] This could imply that the other domain is important for self-assembly and oligomerization (as with the N-terminal domain of VP40),[8] and/or interaction with other viral components such as the RNP complex and incorporating them into the nascent virion.[4, 26] Additionally, the crystal structure showed that RSV M has a distinctive ring of six planar tyrosine residues at the surface of the C-terminal domain, which may be utilized in membrane binding.[1] Mutagenesis studies involving the replacement of one, several or all of these residues may reveal whether they are in fact necessary for efficient association of the M protein with membranes.

4.4 References

1 V. A. Money, H. K. McPhee, J. A. Mosely, J. M. Sanderson and R. P. Yeo, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 4441-4446.

2 A. Dessen, V. Volchkov, O. Dolnik, H.-D. Klenk and W. Weissenhorn, *The EMBO Journal*, 2000, **19**, 4228-4236.

3 B. Sha and M. Luo, *Nature Structural Biology*, 1997, **4**, 239-244.

4 A. Harris, F. Forouhar, S. Qiu, B. Sha and M. Luo, *Virology*, 2001, **289**, 34-44.

5 M. Gaudier, Y. Gaudin and M. Knossow, *The EMBO Journal*, 2002, **21**, 2886-2892.

6 P. Neumann, D. Lieber, S. Meyer, P. Dautel, A. Kerth, I. Kraus, W. Garten and M. T. Stubbs, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 3710-3715.

7 R. W. H. Ruigrok, G. Schoehn, A. Dessen, E. Forest, V. Volchkov, O. Dolnik, H. D. Klenk and W. Weissenhorn, *Journal of Molecular Biology*, 2000, **300**, 103-112.

8 J. Timmins, G. Schoehn, C. Kohlhaas, H.-D. Klenk, R. W. H. Ruigrok and W. Weissenhorn, *Virology*, 2003, **312**, 359-368.

9 F. X. Gomis-Rüth, A. Dessen, J. Timmins, A. Bracher, L. Kolesnikowa, S. Becker, H.-D. Klenk and W. Weissenhorn, *Structure*, 2003, **11**, 423-433.

10 N. Chazal and D. Gerlier, *Microbiology and Molecular Biology Reviews*, 2003, **67**, 226-237.

11 T. Suzuki and Y. Suzuki, *Biological and Pharmaceutical Bulletin*, 2006, **29**, 1538-1541.

12 G. Henderson, J. Murray and R. P. Yeo, *Virology*, 2002, **300**, 244-254.

13 A. Marty, J. Meanger, J. Mills and R. Ghildyal, *Archives of Virology*, 2004, **149**, 199-210.

14 G. Brown, C. E. Jeffree, T. McDonald, H. W. M. Rixon, J. D. Aitken and R. J. Sugrue, *Virology*, 2004, **327**, 175-185.

15 J. P. Laliberte, L. W. McGinnes, M. E. Peeples and T. G. Morrison, *Journal of Virology*, 2006, **80**, 10652-10662.

16 L. J. Johnston, *Langmuir*, 2007, **23**, 5886-5895.

17 D. Marsh, *Biochemica et Biophysica Acta*, 1996, **1286**, 183-223.

18 T.-Y. Wang and J. R. Silvius, *Biophysical Journal*, 2003, **84**, 367-378.

19 T. Bächi and C. Howe, *Journal of Virology*, 1973, **12**, 1173-1180.

20 H. S. Young, J. L. Rigaud, J. J. Lacapere, L. G. Reddy and D. L. Stokes, *Biophysical Journal*, 1997, **72**, 2545-2558.

21 D. S.-Y. Yeo, R. Chan, G. Brown, L. Ying, R. Sutejo, J. Aitken, B.-H. Tan, M. R. Wenk and R. J. Sugrue, *Virology*, 2009, **386**, 168-182.

22 M. Batonick, A. G. P. Oomens and G. W. Wertz, *Journal of Virology*, 2008, **82**, 8664-8672.

23 S. Schuck and K. Simons, *Journal of Cell Science*, 2004, **117**, 5955-5964.

24 H. Imhoff, V. von Messling, G. Herrler and L. Haas, *Journal of Virology*, 2007, **81**, 4158-4165.

25 R. B. Sekar and A. Periasamy, *Journal of Cell Biology*, 2003, **160**, 629-633.

26 E. C. Coronel, T. Takimoto, K. G. Murti, N. Varich and A. Portner, *Journal of Virology*, 2001, **75**, 1117-1123.

5 Experimental details

5.1 General

All chemicals were used as received unless stated otherwise. All solvents used were reagent grade from Fisher Scientific, Loughborough, UK and used without further purification. Water was purified using a Milli-Q 5 system (Millipore, Billerica, MA, USA) to give a resistivity of 18.2 MΩ. Phosphate buffered saline (PBS) pH 7.4 was obtained from Invitrogen, Paisley, UK at 10 × concentration (105 mM KH_2PO_4, 305mM Na_2HPO_4, 1.54 M NaCl), diluted with Milli-Q water and adjusted to the correct pH using concentrated HCl. Matrix protein samples were stored (at 4 °C) and used in PBS pH 7.4, unless otherwise stated.

All lipids were obtained as lyophilized solids unless stated otherwise. DPPC (1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine) and DMPC (1,2-dimyristoyl-*sn*-glycero-3-phosphocholine) were purchased from Bachem AG, Bubendorf, Switzerland, while DPPE (1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine) and DOPC (1,2-dioleoyl-*sn*-glycero-3-phosphocholine) were Sigma brand (Sigma-Aldrich Ltd., Dorset, UK). DPPA (1,2-dipalmitoyl-*sn*-glycero-3-phosphate, monosodium salt) and sphingomyelin from chicken egg yolk were Fluka brand (Sigma-Aldrich). Cholesterol came from Lancaster (Alfa Aesar, Heysham, UK). Galactosyl ceramide (GalCer) and 1,2-dioleoyl-*sn*-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] (Nickel salt) (DOGS-NTA), 5 mg/mL in chloroform, were obtained from Avanti Polar Lipids, Instruchemie BV, The Netherlands. Prior to use lipids were dissolved in chloroform or 2:1 chloroform: methanol and stored at -20 °C.

All enzymes were sourced from Sigma.

5.2 Expression and purification of matrix protein

The full length M ORF was cloned into the pET16b expression vector and expressed in *E. coli* strain BL21 (Al). The cloning procedure added an N-terminal His-tag that allowed protein purification by immobilized metal affinity chromatography (IMAC). The bacteria were grown for 3-4 hours at 37 °C in the presence of ampicillin (50 μg/ml) before expression was induced by addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG, Melford Laboratories Ltd., Suffolk, UK) and 0.2 % (w/v) arabinose (Sigma) for 4 hours at 28 °C. The culture was harvested by centrifugation at 5000 × g for 10 minutes, and the pellet stored at –70 °C until needed. Expression was carried out by Dr. R. Paul Yeo at the Wolfson Research Institute, Durham University Queens Campus.

To extract the MHis protein, the sample was thawed and suspended in phosphate buffer (0.05 M NaH_2PO_4, 0.3 M NaCl, pH 7.2-8.0), lysozyme (1 μg/mL) added, and the solution cooled on ice for 15-30 minutes. After addition of DNase 1 (1000 u) and RNase A (1 μg/mL) the cells were lysed using a French press if volume >100 mL, or a sonicator for smaller volumes. CHAPS detergent (Sigma) was added to a final concentration of 8 mM. After cell lysis the suspension was centrifuged at 50,000 × g for 30 minutes at 4 °C, and the pellet discarded. Imidazole (1M in phosphate buffer pH 7.4) was added to the supernatant to a concentration of 0.05 M, and a 50 % suspension of Ni-NTA agarose beads added (Qiagen Ltd., Crawley, West Sussex; 0.75 mL per L of supernatant). The mixture was placed on a carousel for 30-60 minutes at 4 °C before centrifugation at 100 × g for 2 minutes. The supernatant was removed and the pelleted beads suspended in 50 mL phosphate buffer containing 0.05 M imidazole and 8 mM CHAPS. After centrifuging a second time the pellet of Ni-Agarose beads was suspended in 10 mL phosphate buffer containing 0.05 M imidazole and 1 % CHAPS, and loaded onto a column. The liquid was allowed to flow through before addition of 50 mL phosphate buffer containing 0.1 M imidazole and 1 % CHAPS, which was also allowed to flow through. The MHis protein was eluted by application of a 0.2–0.5 M imidazole gradient in phosphate buffer, and the fractions analysed by SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis). Fractions containing the protein were combined and dialysed

against PBS pH 7.4, then stored at 4 °C. For the majority of experiments the His tag was removed using Factor Xa enzyme and Xarrest agarose beads (Novagen, Merck Chemicals Ltd., Nottingham, UK) as per the manufacturer's instructions.

5.2.1 SDS-PAGE

Purified M protein solutions were analysed on 12 % acrylamide separating gels made up using 2 mL ProtoFlow 30 % acrylamide gel (Flowgen Bioscience Ltd., Nottingham, UK), 1.25 mL resolving buffer (1.5 M tris, 0.4 % (w/v) SDS, pH 8.8), 1.7 mL water, 5 mL TEMED and 50 µL 10 % (w/v) ammonium persulphate solution in water. Stacking gels consisted of 0.65 mL ProtoFlow gel, 1.25 mL stacking buffer (0.5 M tris, 0.4 % (w/v) SDS, pH 6.8), 3.05 mL water, 5 µL TEMED and 25 µL 10 % APS solution. Gel mixtures were degassed under vacuum before cross-linking was initiated by the APS addition. Samples were mixed with an equal volume of loading buffer (40 % (v/v) glycerol, 1.5 M tris, 4 % (w/v) SDS, 0.05% (w/v) bromophenol blue, 4.75 % (v/v) 2-mercaptoethanol, pH 6.8) and heated at 95 °C for 5 minutes before being loaded onto the gel and run in buffer containing 0.025 M tris, 0.192 M glycine, 0.1 % (w/v) SDS, with a pH of 8.3-8.6. SeeBlue molecular weight standard (Invitrogen) was also run on each gel. Gels were stained for a minimum of 15 minutes with a fresh solution of 0.1 % (w/v) Coomassie Blue dissolved in 50 % (v/v) methanol and 7 % (v/v) acetic acid, and destained with two 20-30 minute washes of a 40 % methanol (v/v) 10 % (v/v) glacial acetic acid solution.

5.2.2 Determining protein concentration

A quartz cuvette was filled with 3 mL 6 M aqueous guanidine hydrochloride (Fluka; recrystallized from methanol) and run as a blank in a UV spectrometer over a wavelength range of 200 to 350 nm. 5 µL of the protein solution to be tested was added to the cuvette and after mixing the UV spectrum was recorded again. The guanidine hydrochloride ensured that the protein was fully unfolded, so that all residues could contribute to the UV signal. The measurement was repeated after the addition of 10, 15 and 20 µL of solution. The absorption values at 280 nm in each

case were used to calculate the concentration of the protein solution as follows, where V_0 is the initial volume in the cuvette, V is the volume of protein added each time, and x is the unknown concentration of the protein solution. All volumes should be in mL, and concentrations in M (mol.dm^{-3}); absorbance values have no units.

If A = observed absorbance at 280 nm, then $A(corr)$, the corrected absorbance allowing for the change in volume on addition of protein solution, is:

$$A(corr) = A\left(\frac{V_0 + V}{V_0}\right) \qquad (30)$$

and the concentration of the protein solution, x, is the overall concentration in the cuvette, c, multiplied by the dilution factor:

$$x = c \times \frac{V}{V_0}$$

$$c = \frac{xV}{V_0} \qquad (31)$$

Combining equation *(31)* with the Beer-Lambert law ($A = \varepsilon cl$):

$$A(corr) = \frac{\varepsilon x V}{V_0} \qquad (33)$$

Therefore a plot of $A(corr)$ against V has a gradient of $\frac{\varepsilon x}{V_0}$, from which x, the concentration of protein solution, can be calculated if ε and V_0 are known. The absorption coefficient ε for the matrix protein was calculated from the number of tyrosine (Tyr), tryptophan (Trp) and cystine (formed from a cysteine-cysteine disulphide bond) residues in the amino acid sequence, using the formula:[1, 2]

$$\varepsilon(protein) = [N(Tyr) \times \varepsilon(Tyr)] + [N(Trp) \times \varepsilon(Trp)] + [N(Cystine) \times \varepsilon(Cystine)] \quad (34)$$

where $\varepsilon(Tyr) = 1285$, $\varepsilon(Trp) = 5685$ and $\varepsilon(Cystine) = 125$ M^{-1}cm^{-1}

This gave a value for ε of 24220 $M^{-1}cm^{-1}$, since in the RSV M protein none of the cysteine residues appear as half-cystines.

5.3 Circular dichroism spectroscopy

Spectra were recorded between 180 and 320 nm using a Jasco J-810 Spectropolarimeter. Protein samples were dialysed against 5 mM phosphate buffer pH 7.4 overnight at 4°C before use, and the same buffer was used as a blank. Far-UV spectra, and the corresponding blanks, were recorded by averaging 8 accumulations recorded at a rate of 10 nm/min using a path length of 0.2 cm with a pitch of 0.5 nm, a bandwidth of 1 nm and a response time of 2 s. Near-UV spectra were recorded using a 1 cm path length cell, a pitch of 0.2 nm and a response time of 1 s. Protein samples were serially diluted with buffer until a spectrum could be measured down to at least 190 nm without saturation of the detector. After subtraction of the appropriate blank, spectral smoothing was carried out in the Jasco Spectra Analysis program using the Savitsky-Golay method. Smoothed data were analysed for protein secondary structure using the CDSSTR, SELCON3 and CONTIN/LL programs[3] accessed either *via* the Dichroweb service[4,5] or the CDPro package. Both a general protein reference set (SP43/dataset 4) and one including membrane proteins (SMP56/dataset 10) were used.[3,6]

5.4 Protein cross linking

A solution of hRSV M protein (1.55 μM) and glutaraldehyde (0.275 M) in 0.55 ml water (equivalent to 0.4 mg/mL protein and 5% (v/v) glutaraldehyde) was incubated for 2 minutes at room temperature before the addition of sodium borohydride (1 M in 0.1 M NaOH; 45 μl). This was repeated several times with differing ratios of protein:glutaraldehyde, with the volume of $NaBH_4$ needed to quench calculated from the volume of glutaraldehyde used.

5.5 Mass spectrometry

Samples were dialysed overnight against water and analysed by MALDI-TOF mass spectrometry with a sinapinic acid matrix on a Bruker Daltonics Autoflex TOF/TOF instrument.

5.6 Extraction of mouse lung lipids

Lungs from 5 CD-1 mice, freshly culled by cervical dislocation, were frozen in liquid nitrogen and homogenized using a pestle and mortar with 2:1 (v/v) chloroform/methanol mixture, sufficient to give a final volume approximately 15 times the volume of the lung tissue. The homogenate was filtered (Whatman grade no. 1 paper) then the biphasic solution was washed with 1-2 mL of distilled water. Most of the lower lipid-containing organic phase was drained off and retained, and the remaining interfacial liquid washed with water a second time, before the final few drops of clean organic phase were run off and added to the rest.[7] This chloroform solution containing the lipids was stored at -20°C until needed.

5.7 Incubation of matrix protein with lipids

The appropriate volume of lipid stock solution in chloroform was syringed into a round-bottomed flask, and the solvent removed on a rotary evaporator to leave a thin lipid film. This was dried in a desiccator overnight, and then re-hydrated with a suitable volume of protein in PBS, with additional buffer solution if necessary. The resulting suspension was mixed on a vortex mixer before being incubated at 37 °C for 5-7 days, followed by at least 2 days at 4 C. Samples were examined by transmission electron microscopy (TEM) to look for any ordered structures.

5.7.1 Transmission electron microscopy

Samples were prepared on 400 mesh carbon film coated copper grids (Agar Scientific, Stansted, Essex), which were used as supplied. A 5 μL droplet of the sample solution was deposited onto the grid and left for approximately 1 minute, before the excess liquid was carefully removed from the edge of the grid using filter paper. An equal volume of negative stain solution (1% uranyl acetate in water) was added to the grid in the same way. Grids were examined using a Hitachi TEM-H7600 transmission electron microscope in high-resolution mode, with an accelerating voltage of 80-100 kV.

5.8 Studies using lipid monolayers

5.8.1 Langmuir isotherms

Work was carried out on a 150 x 70 mm Langmuir trough (model 102M, Nima Technology Ltd., Coventry, UK) with two moving barriers. Before and after each use, and between runs, the trough was cleaned with isopropyl alcohol to remove any contaminants. The trough was filled with PBS pH 7.4, the barriers closed, and the surface between them cleaned with an aspirator. After opening the barriers a surface pressure-area isotherm was recorded, and the cleaning process repeated until the isotherm obtained was flat to within 0.1 mNm^{-1}. The required quantity of lipid/s dissolved in chloroform was deposited onto the surface using a glass microsyringe (Hamilton, Reno, NV, USA) and left for 15 minutes to equilibrate. Π-A isotherms were recorded at 50 cm^2/minute; the monolayer was compressed part-way and then expanded 3 or 4 times before being compressed as far as possible. This was intended to ensure that no lipid was lost to the subphase during the final compression, improving reproducibility.[8] In the case of protein (with or without lipid) monolayers, protein dissolved in PBS pH 7.4 was deposited onto the surface using an automatic pipette, with the lipid solution, if used, added immediately after *via* microsyringe. The system was left to equilibrate for 15 minutes, after which isotherms were recorded in the manner described above.

5.8.2 Brewster angle microscopy

The equipment used for Brewster angle microscopy is shown in Figure 3.15 in Chapter 3. The dual-barrier Nima Langmuir trough mentioned above was modified by replacing the window in the centre of its base with a cylindrical glass well approximately 3 cm deep. A *p*-polarized 633 nm helium-neon laser (CVI Melles-Griot, Leicester, UK) was passed through a Glan-Taylor polarizer so that it was incident on the surface of the subphase in the trough at an angle (θ in Figure 3.15) of approximately 53° to the normal; the angle was then adjusted so that no reflection could be detected from the clean surface, ensuring that all the light was being refracted and exiting through the glass well. Lipid and/or protein monolayers were spread in the fashion described in section 4.8.1, and were observed *via* reflected light collected using a CCD video camera (Optem Zoom 70, Qioptik Imaging Solutions, Newport, NY, USA) and digitized using LabView (National Instruments Corporation, Austin, TX, USA).

5.8.3 Preparation of supported protein-lipid films

Silicon wafers cut to approximately 1 cm^2 were cleaned by immersion in a 3:1 (v/v) mixture of 30% hydrogen peroxide and concentrated sulphuric acid for 45 minutes at 80°C, followed by treatment in a 1:1 mixture of 30% hydrogen peroxide and concentrated HCl at room temperature for 15 minutes. The clean wafers were rinsed with Milli-Q water and dried in an oven in a sample vial covered with foil to prevent contamination from dust. The wafers were then added to a 1 mM solution of octadecytrichlorosilane (OTS) (Aldrich) in dry toluene, in a glass vial sealed with Teflon tape to exclude moisture, for 24-36 hours. The wafers were next sonicated for 2 × 15 minutes in toluene, followed by 2 × 15 minutes in chloroform, before a final rinse with Milli-Q water.[9]

To deposit each protein-lipid film, a monolayer of protein and lipid was prepared as described in section 5.8.1, this time on a larger Langmuir trough incorporating a

dipping mechanism (Nima Technology Ltd., model number 612D). The monolayer was compressed to the required pressure, which was maintained whilst dipping took place. The OTS-coated wafer was dipped downwards through the monolayer from a position above the surface, before being released into a clean sample vial submerged in the subphase below the dipping mechanism. The vial was then carefully retrieved, and the top few millimetres of liquid suctioned off to remove any remaining surfactant. The deposited protein was stabilized by making the solution in the vial 250 mM with respect to gluteraldehyde (Fluka *BioChemika,* 25 % in water) for 2 minutes before quenching with sodium borohydride sufficient to give a 6.25:1 molar ratio of glutaraldehyde:sodium borohydride (4 mL of 1 M sodium borohydride in 0.2 M sodium hydroxide, in 14 mL total volume). This cross-linked the protein molecules so that the film could be exposed to air for examination by AFM.

Some films were also deposited onto silicon wafers that had been cleaned but were not coated in an OTS monolayer. In this case the procedure was followed exactly as described above, with only the treatment in 1 mM OTS omitted.

5.8.4 Atomic force microscopy

AFM imaging was carried out in tapping mode on a MultiMode AFM equipped with a NanoScope IV controller (Veeco Instruments Inc., California, USA) using etched silicon probes, with a spring constant of 40 N/m (Budget Sensors Tap300, Windsor Scientific, Slough), and a drive frequency of 300 kHz. Images were recorded at typical scan rates of 0.5-1.0 Hz, with integral and proportional gains adjusted for optimum imaging (integral ~0.2-0.4, proportional ~0.3-0.5). The amplitude setpoint was typically 1.5-1.6. Each image was recorded with 256 scan lines. Images were analysed using the Veeco NanoScope software version 6.13r1.

5.9 References

1 H. Edelhoch, *Biochemistry*, 1967, **6**, 1948-1954.

2 C. N. Pace, F. Vajdos, L. Fee, G. Grimsley and T. Gray, *Protein Science*, 1995, **4**, 2411-2423.

3 N. Sreerama and R. W. Woody, *Analytical Biochemistry*, 2000, **287**, 252-260.

4 L. Whitmore and B. A. Wallace, *Nucleic Acids Research*, 2004, **32**.

5 A. Lobley, L. Whitmore and B. A. Wallace, *Bioinformatics*, 2002, **18**, 211-212.

6 N. Sreerama and R. W. Woody, *Protein Science*, 2004, **13**, 100-112.

7 J. Folch, M. Lees and G. H. Sloane Stanley, *Journal of Biological Chemistry*, 1957, **226**, 497-509.

8 S. M. Ennaceur and J. M. Sanderson, *Langmuir*, 2005, **21**, 552-561.

9 S. M. D. Watson, K. S. Coleman and A. K. Chakraborty, *ACS Nano*, 2008, **2**, 643-650.

Appendix A: Protein-lipid incubation experiments

Expt.	Protein	Lipid	Lipid:protein (w/w)	[Total lipid] / μM	[P] / μM	Incubation
Original	M254RHIS	DPPE/DOPC 2:8	1:15	106.21	34.19	WR
A	M254R	DPPE/DOPC 2:8	1:15	182.06	58.53	WR
B	M254RHIS	DPPE/DOPC 2:8	1:15	75.29	24.34	WR
C	M254R	DPPE	1:15	165.47	58.53	WR
D	M254R	DOPC	1:15	145.65	58.53	WR
1	M254R	DPPE/DOPC 2:8	1:15	54.06	17.09	WR
2	M254R	DPPE/DOPC 2:8	1:15	54.06	17.09	WR
3	M254RHIS	DPPE/DOPC 2:8	1:15	54.06	17.09	WR
4	M254RHIS	DPPE/DOPC 2:8	1:15	54.06	34.19	WB
6	M254R	DPPE/DOPC 2:8	1:10	159.00	34.19	WB
7	M254R	DPPE/DOPC 2:8	1:10	159.00	34.19	WB
8	M254R	DPPE/DOPC 2:8	1:1	1590.02	34.19	WB
9	M254R	DPPE/DOPC 2:8	1:1	1590.02	34.19	WB
10	M254R	DPPE/DOPC 2:8	5:1	7950.14	34.19	WB
11	M254R	DPPE/DOPC 2:8	5:1	7950.14	34.19	WB
12	M254R	DPPE/DOPC 2:8	1:10	318.01	68.37	RT
13	M254R	DPPE/DOPC 2:8	1:10	179.67	38.63	RT
14	M254R	DOPC/DPPE/DPPA 8:1:1	1:10	147.67	38.63	RT
15	M254RHIS	DPPE/DOPC 2:8	1:10	240.09	51.56	RT
16	M254RHIS	DOPC/DPPE/DPPA 8:1:1	1:10	197.33	51.56	RT
17	M254R	DPPE/DOPC 2:8	1:10	228.96	49.13	RT
18	M254R	DOPC/DPPE/DPPA 8:1:1	1:10	188.19	49.13	WR
19	M254RHIS	DPPE/DOPC 2:8	1:10	397.51	85.47	WR
20	M254RHIS	DOPC/DPPE/DPPA 8:1:1	1:10	326.71	85.47	WR
21	M254R	DOPC/DPPE/DPPA 8:1:1	1:2	741.64	38.80	HB, fridge
22	M254R	DOPC/DPPE/DPPA 8:1:1	1:20	74.16	38.80	HB, fridge
23	M254R	DOPC/DPPE/DPPA 8:1:1	1:50	29.73	38.80	HB, fridge
24	M254R	DOPC/DPPE/DPPA 8:1.75:0.25	1:10	148.33	38.80	HB, fridge
25	M254R	DPPE/DOPC 2:8	1:50	16.79	18.09	HB, fridge
26	M254R	DOPC/DPPE/DPPA 8:1:1	1:50	13.80	18.09	HB, fridge
27	M254R	DOPC/DPPE/DPPA 8:1:1	1:75	9.20	18.09	HB, fridge
28	M254R	DOPC/DPPE/DPPA 8:1:1	1:100	6.90	18.09	HB, fridge
29	M254R	DOPC/DPPE/DPPA 8:1:1	1:10	177.73	47.19	WR 10 d, CR 7 d
30	M254R	DOPC/DPPE/DPPA 8:1:1	1:15	120.23	47.19	WR 10 d, CR 7 d
31	M254R	DOPC/DPPE/DPPA 8:1:1	1:20	88.87	47.19	WR 10 d, CR 7 d
32	M254R	DOPC/DPPE/DPPA 8:1:1	1:30	62.72	47.19	WR 10 d, CR 7 d
33	M254R	DOPC/DPPE/DPPA 8:1:1	1:13	95.14	32.33	WR 10 d, CR 7 d
34	M254R	DOPC/DPPE/DPPA 8:1:1	1:15	82.33	32.33	WR 10 d, CR 7 d
35	M254R	DOPC/DPPE/DPPA 8:1:1	1:17	72.66	32.33	WR 10 d, CR 7 d
36	M254R	DOPC/DPPE/DPPA 8:1:1	1:19	65.08	32.33	WR 10 d, CR 7 d
37	M254R	DOPC/DPPE/DPPA	1:17	64.30	28.58	WR 14 d,

		8:1:1				CR 7 d
38	M254R	DOPC/DPPE/DPPA 8:1:2	1:19	57.50	28.58	WR 14 d, CR 7 d
39	M254RHIS	DPPE/DOPC 2:8	1:15	115.12	37.17	WR 7 d, CR 7 d
40	M254RHIS	DOPC/DPPC/chol 1:1:1	1:15	113.90	37.17	WR 7 d, CR 7 d
41	M254RHIS	DOPC/SM/chol 1:1:1	1:15	146.78	47.07	WR 7 d, CR 7 d
42	M254R	DPPE/DOPC 2:8	1:15	65.32	21.07	WR 7 d, CR 7 d
43	M254R	DPPE/DOPC 2:8	1:10	97.98	21.07	WR 7 d, CR 7 d
44	M254R	DOPC/DPPE/DPPA 8:1:1	1:15	53.69	21.07	WR 7 d, CR 7 d
45	M254R	DOPC/DPPC/chol 1:1:1	1:15	64.63	21.07	WR 7 d, CR 7 d
46	M254R	DPPE/DOPC 2:8	1:15	53.11	17.09	WR 7 d, CR 7 d
47	M254R	DOPC/SM/chol 1:1:1	1:15	53.42	17.09	WR 7 d, CR 7 d

Table A: Protein-lipid incubation experiments carried out with the aim of forming tubular structures.
Experiments where tubules (of varying quality) were seen are highlighted. Note that SM = sphingomyelin, chol = cholesterol, HB = heating block (37 ºC), WB = water bath (37 ºC), WR = warm room (37 ºC), CR = cold room (4 ºC), RT = room temp (20-25 ºC), d = days, and ratios are w/w.

Appendix B: Paper describing the structure of the RSV matrix protein

Surface features of a *Mononegavirales* matrix protein indicate sites of membrane interaction

Victoria A. Money[a,b], Helen K. McPhee[c,1], Jackie A. Mosely[c,1], John M. Sanderson[c,2], and Robert P. Yeo[b,d,2]

[a]The Maurice Wilkins Centre for Molecular Biodiscovery and the School of Biological Sciences, University of Auckland, Thomas Building, 3a Symonds Street, Auckland Central 1010, New Zealand; [c]Department of Chemistry and [b]Centre for Bioactive Chemistry, University Science Laboratories, Durham University, South Road, Durham, DH1 3LE, United Kingdom; and [d]School for Medicine and Health, Queen's Campus, Durham University, Stockton-on-Tees, TS17 6BH, United Kingdom

Edited by Robert A. Lamb, Northwestern University, Evanston, IL, and approved January 14, 2009 (received for review June 13, 2008)

The matrix protein (M) of respiratory syncytial virus (RSV), the prototype viral member of the *Pneumovirinae* (family *Paramyxoviridae*, order *Mononegavirales*), has been crystallized and the structure determined to a resolution of 1.6 Å. The structure comprises 2 compact β-rich domains connected by a relatively unstructured linker region. Due to the high degree of side-chain order in the structure, an extensive contiguous area of positive surface charge covering $\approx$600 Å^2 can be resolved. This unusually large patch of positive surface potential spans both domains and the linker, and provides a mechanism for driving the interaction of the protein with a negatively-charged membrane surface or other virion components such as the nucleocapsid. This patch is complemented by regions of high hydrophobicity and a striking planar arrangement of tyrosine residues encircling the C-terminal domain. Comparison of the RSV M sequence with other members of the *Pneumovirinae* shows that regions of divergence correspond to surface exposed loops in the M structure, with the majority of viral species-specific differences occurring in the N-terminal domain.

CD | crystal structure | respiratory syncytial virus | sequence alignment

Respiratory syncytial virus (RSV) is the prototype member of the *Pneumovirinae*, a subfamily of the *Paramyxoviridae* (order *Mononegavirales*). Morphologically, the extracellular virion consists of a lipid bilayer envelope, within which are embedded 3 glycoproteins, 2 of which (F and G) are important in cell attachment and viral entry into target cells. The third, the SH protein, contributes to pathology in the host (1). Internally, virions contain helical nucleocapsids that consist of N protein tightly bound to the negative-sense nonsegmented genomic RNA. The nucleocapsid in turn is associated with components of the viral RNA-dependent RNA polymerase (L, P, M2-1, and M2-2 proteins), forming the holo-nucleocapsid (2–4). Between the holo-nucleocapsid and the outer envelope there is a layer of matrix protein (M), which is associated peripherally with the membrane (5). The other family members of the *Mononegavirales* (*Rhabdoviridae*, *Filoviridae*, and *Bornaviridae*) all subscribe to this basic arrangement of the virion, although the overall morphology can vary between the families. For example, *Paramyxoviridae* virions are pleiomorphic, whereas the *Rhabdoviridae* have a regular bullet shape structure, and the *Filoviridae* have a more filamentous shape.

Extracellular RSV virions form by a budding process that occurs at the plasma membrane within specialized lipid domains (5, 6) and M appears to drive the final assembly process, which is the incorporation of the holo-nucleocapsid and initiation of the budding process (7, 8). Before budding, there is a coordinated assembly of viral components; and it is evident that the glycoproteins and M proteins are important determinants of the location on the plasma membrane at which the virus buds (9). It is also possible that the interaction between M and the glycoproteins, possibly mediated with the cytoplasmic tails, are important in the budding process. Genome silencing to prevent transcription and replication by the viral polymerase before incorporation of the holo-nucleocapsid in the nascent virion is a related function of the M. M are also implicated in host cell transcriptional cut-off (10), possibly via a direct interaction with RNA (11).

A number of matrix-like proteins are known to bind membranes or lipid vesicles in vitro, most likely through a combination of hydrophobic and electrostatic interactions (12–14). Expression of certain Ms in eukaryotic cells in the absence of other viral proteins can induce formation of virus-like particles (VLPs). The efficiency of VLP generation can be increased if the M is coexpressed with a viral glycoprotein (15–18). Ms share a tendency to oligomerize, a feature likely to be important in the self-assembly and budding processes (19). In tissue culture, RSV induces formation of long slender projections from the surface of the cell known as viral filaments. It was found that removal of the lipid membrane from viral filaments left an M containing sheath (5); it is possible that M oligomerization and self-assembly is the driving force behind the formation of viral filaments. In this article we report the structure of the full-length RSV M protein solved at a resolution of 1.6 Å, and discuss the implications of this structure for the function of the protein.

Results and Discussion

Structure of the RSV M. The RSV M was purified by using nickel affinity chromatography. During the cloning process, a methionine to arginine change occurred, and we refer to the resultant form of M as M^{254R}. The crystal structure of M^{254R} was solved by using MIRAS techniques to a resolution of 1.6 Å, representing the first example of an intact M from the *Mononegavirale*s. Despite evidence for higher order oligomers [supporting information (SI) Fig. S1] such as dimers, tetramers, and hexamers, in solution, the crystallized form is monomeric. Crystallographic data are presented in Table S1. The overall fold consists of 2 clear domains connected by a 13-residue linker region. The N-terminal domain comprises residues 1 to 126, whereas the C-terminal domain consists of residues 140 to 255 (Fig. 1 *A* and *B*). Only residues 99 and 100 could not be assigned a clear location within the electron density. The N-terminal domain consists of a twisted β-sandwich comprised of 2 β-sheets, 1 of 3 and 1 of 4 strands, positioned almost perpendicular to each other. The overall topology of this domain is of a curved

Author contributions: V.A.M., J.M.S., and R.P.Y. designed research; V.A.M., H.K.M., J.A.M., J.M.S., and R.P.Y. performed research; V.A.M., H.K.M., J.A.M., J.M.S., and R.P.Y. analyzed data; and V.A.M., J.M.S., and R.P.Y. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors for the RSV M protein have been deposited in the Protein Data Bank, www.rcsb.org [PDB ID codes 2vqp (coordinates) and r2vqpsf (structure factors)].

[1]H.K.M. and J.A.M. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: j.m.sanderson@durham.ac.uk or r.p.yeo@durham.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0805740106/DCSupplemental.

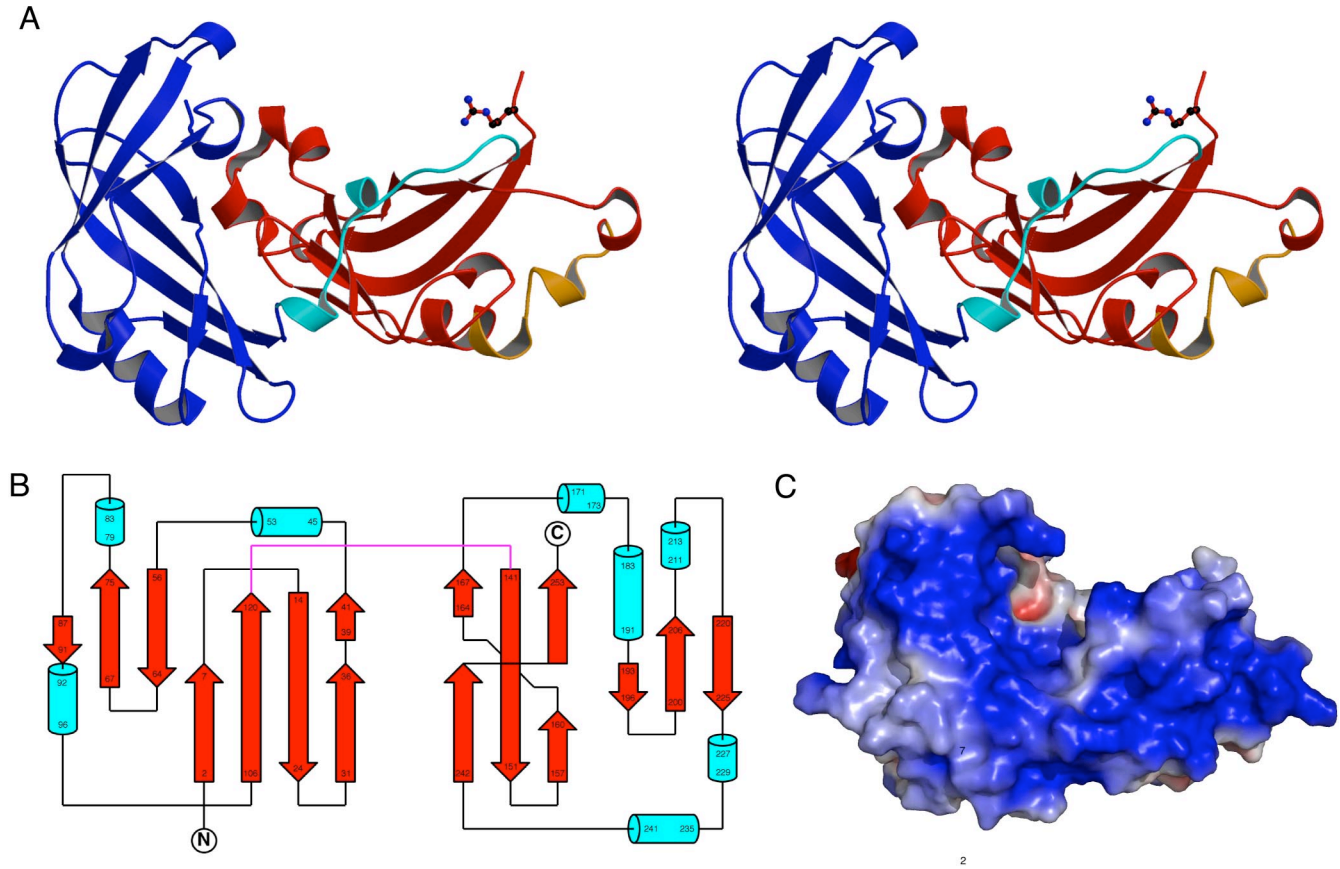


Fig. 1. Three-dimensional structure of the RSV M protein. The crystal structure of M^{254R} (resolution 1.6 Å) shows 2 domains composed largely of β-sheets. Statistical information for the X-ray data are provided in *SI Materials and Methods*. (*A*) Divergent (wall eyed) stereoview of M^{254R} colored according to domain with the linker shown in cyan, the N-terminal domain in blue and the C-terminal domain in red. Residue R254 is shown in ball-and-stick representation. (*B*) A topology diagram of the protein. The linker between the N- and C-terminal domains is shown in magenta. Residues (numbers refer to Met as +1) in β-sheets are represented by broad arrows and helices as cylinders. (*C*) Electrostatic surface potential (calculated with APBS) for M^{254R}, presented in a color range from red to blue (−5 to +5 kT/e); uncharged residues are uncolored.

horseshoe-like arrangement with β-sheet 1 forming the concave, inner face flanked by loop regions, and β-sheet 2 forming the convex, outer face. The C-terminal domain consists of a flattened β-barrel, comprising 2 3-stranded anti-parallel β-sheets. The regions linking the sheets between strands 2 and 3, and strands 5 and 6, are largely helical in nature (Fig. 1*B*). There is no evidence for any complexed metal ions or for a potential zinc finger motif. The 254R substitution in the protein lies at the very end of the C-terminal domain in an area largely devoid of secondary structure; therefore, it is unlikely to have a significant effect on the protein structure or function.

The linker region is largely lacking in secondary structure features, with the exception of a short helical region. The presence of this linker is consistent with structures obtained for fragments of the Ebola virus (EBOV) VP40 and vesicular stomatitis virus (VSV) Ms, and from capsid proteins of retroviral origin that fulfil similar in vivo roles (20–24), and its unstructured nature suggests that the N- and C-terminal domains may be able to occupy different orientations, relative to each other, than that observed in the crystal. The 2 domains are only loosely associated, with the major interactions between them being hydrophobic in nature, supported by a small number of water-mediated hydrogen bonds. We have observed that, in common with EBOV VP40, proteolysis of RSV M^{254R} occurs in solution and results in the dissociation of the N- and C-terminal domains, the weak interdomain interactions being insufficient to hold them together. We have mapped the cleavage site of M^{254R} by limited mass spectrometry to the linker region between amino acids Thr-136 and Leu-137 (see Fig. S1). These weak interdomain interactions further suggest that the protein may exist in alternative quaternary structures in solution, and that the interdomain packing observed in the crystal may be metastable, driven mainly by sequestering of the hydrophobic residues, which form the major part of the interface, away from the bulk solvent. Flexibility of this type has been speculated as being important for the M of the influenza virus, as well as for EBOV VP40 and VSV M, and is thought to be necessary to accommodate the different functions of the M throughout the viral life cycle.

Structure Comparison. The tertiary structure of M^{254R} is globally similar to that of the EBOV VP40 protein (24), showing the same overall fold with a Z score of 6.6 and a rmsd of 3.7 Å. Searches of the PDB for structurally homologous proteins to M^{254R} using the programs DALI, VAST, and SSM (25–27) yield no further significantly similar structures. Performing the same searches using the N- and C-terminal domains independently demonstrates that, in addition to being closely related to each other, the N-terminal domain shows slight similarities to parts of 2 DNA topoisomerases from *Escherichia coli*. These similarities are restricted to the shape of the β-sheets in domain 2 of the DNA binding proteins, and are unlikely to be relevant to the function of the N-terminal domain.

Structural information on Ms is scarce partly due to the intrinsic difficulty of working with these hydrophobic proteins

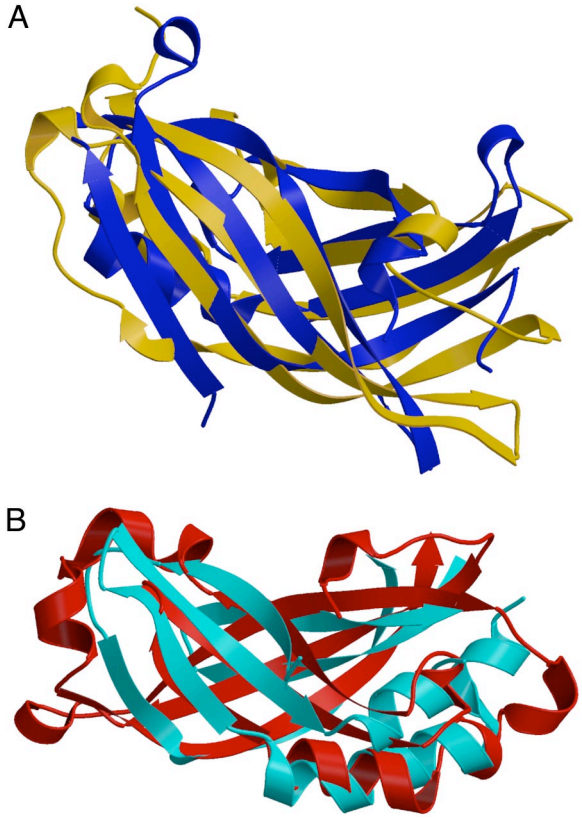


Fig. 2. Comparison of the RSV M protein topology with that of EBOV VP40. The diagrams show an overlay of the β-sheet arrangements of M^{254R} with EBOV VP40 (PDB code 1ES6). (*A*) RSV M protein N-terminal domain in blue and VP40 in yellow; and (*B*) M^{254R} protein C-terminal domain in red and VP40 in cyan. The same images are presented in Fig. S2, in stereoscopic views.

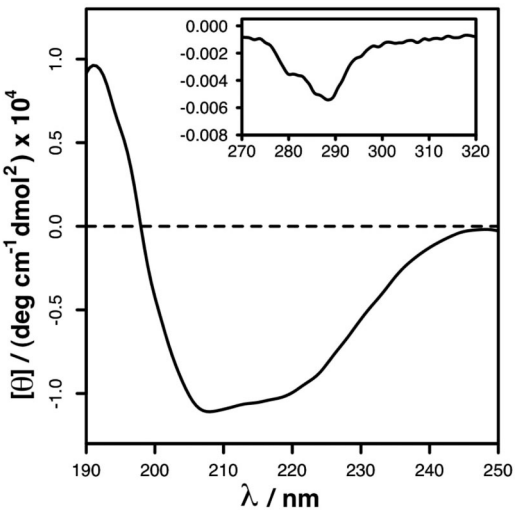


Fig. 3. CD spectrum of RSV M. Datasets of the far-UV (main box, 190–250 nm) and near- UV (inset box, 270–320 nm, axes are the same as the main plot) spectra were collected as indicated in *Experimental Procedures*. The data were analyzed for secondary structure information by using the CDSSTR, SELCON3, and CONTIN/LL programs. A more comprehensive dataset on *M254R* structure in solution is provided in Fig. S3, Fig. S4, Table S2, and Table S3; also, see Table 1. A comparison with X-ray data with calculated CD spectra indicates that the crystal is more ordered than the solution structure.

that are prone to self-aggregation. Only 2 Ms from the *Mononegavirales* have been subjected to high-resolution structure determination. For both published structures, a proteolytically resistant core was crystallized as opposed to the full-length structure. Separate overlay of the N- and C-terminal domains of M^{254R} with those of EBOV VP40 (PDB code 1ES6, see Fig. 2 and Fig. S2), allows the difference in angle between the domains in the 2 proteins to be accounted for, and shows the similarities in the protein core of each domain. Despite the clear close relation between the folds of the 2 proteins, the topological arrangement of the 2 differs. The N-terminal domain of RSV M^{254R} contains a mixed 4-stranded and 1 anti-parallel 3-stranded β-sheet as opposed to the 2 3-stranded sheets found in VP40. A similar comparison of M^{254R} and VSV M could not be performed, emphasising the structural, but not functional, diversity of the 2 Ms.

Electrostatic Surface of the RSV M. To fulfil its structural role, RSV M must be able to form protein–protein and protein–lipid interactions. Consequently, one would expect to observe surface areas with significant hydrophobic patches, as well as positively charged regions, that would favor protein–membrane association. Examination of the surface of M reveals an extensive positively charged area of ≈600 Å^2, extending across both N- and C-terminal domains encompassing, and including, a significant contribution from the linker (Fig. 1 *C*). Because the binding of M to cell membranes is thought to be mediated largely by electrostatic contacts (12, 28), this region provides a mechanism by which the protein is able to associate with negatively charged host membranes. Comparison of the electrostatic surface of M^{254} with VP40 is hampered by the fact that in the latter the linker region is not modeled. However, it does appear that a significant positively-charged patch is a feature present on all Ms for which there is structural information available, including retroviral homologues. Looking at the rest of M^{254R}, there is a distinctive negatively charged lobe on the N-terminal domain and a positively charged pocket on the C-terminal domain, which are potential sites for directing interactions with binding partners. Work on the isolated N-terminal domain of EBOV M, VP40 (29), indicates that this domain has the capacity to oligomerize in the presence of nucleic acids, pointing to a more prominent role for this domain in RNP association, with the C-terminal domain interacting predominantly with the membrane. The precise role of each domain in the full-length proteins (RSV M and EBOV VP40) has yet to be determined, although the extent of the positively-charged surface of RSV M indicates that functional surfaces have the potential to extend across the domain boundary and the linker to some degree.

Comparison of Structures in Solution and in Crystal. To determine whether the crystal structure was also that of the protein in solution, we performed a CD analysis of the protein. Data collected included wavelengths down to 190 nm, which significantly increases confidence in the interpretation of CD derived structural information for comparison with crystallographic data (29). Evaluation of CD (Fig. 3 and Table 1; for more details of CD spectrum deconvolution, see Figs. S3 and S4 and Tables S2 and S3) and X-crystallographic data indicate that the secondary structure content of M^{254R} in solution is lower than that observed in the crystalline state. This observation implies that assembly of the protein is associated with significant unfavorable entropic contribution from protein folding, in common with the binding of disordered proteins to their counterparts (30–33). Such proteins achieve a favorable binding free energy through large enthalpic contributions from intramolecular electrostatic interactions to offset negative entropic contributions from protein folding. In order for this to be achieved, the number of contacts between the protein and its binding partners are numerous. In the case of M, analogy with disordered proteins would suggest that the contact area for protein–protein and protein–membrane

Table 1. Comparison of the solution and crystallographic structure of the the RSV M protein

	α-Helix	3_{10}-Helix	Strand	Turn	Coil	Bridge
% of total, CD*	12	12	23	22	31	—
% of total, X-ray†	12	8	47	21	12	1

Summary of the output resulting from analysis of the M^{254R} structure using the Stride program. Comparison of the CD and X-ray data. The output from the Contin/LL program is presented as this gave the best fit between experimental and computed structures. For more details of the various programs used to fit the CD data to known structures, see *SI Materials and Methods*. The CD spectrum shows that the solution structure is more disordered that that of the crystallized form, compare "strand" (23 vs. 47%) and "coil" (31 vs. 12%).
*Output from the CONTIN/LL program.
†Output from Stride analysis.

interactions is similarly large. These requirements are consistent with the observed distribution of aromatic and positively charged residues in the protein, particularly with regard to the large patch of positive electrostatic potential described above. Intrinsic disorder has been proposed as a mechanism by which viral proteins are able to form multiple binding interactions with different partners; thus, expanding protein functionality without a concomitant increase in the size of the genome. RSV M fits this description having multiple interactions, such as with itself, the nucleocapsid, the viral glycoproteins, such as the F protein, via their cytoplasmic tails and with the host cell and viral membranes. All of these interactions are essential for the assembly and budding of a virion particle, and could contribute to stabilizing the structure.

Comparison with Other Pneumovirinae Ms. Analysis of sequence alignments of a number of pneumovirus and metapneumovirus M proteins with M^{254R} reveals that the majority of significant amino acid sequence diversity (Q-scores of 20 and above) can be mapped to external loop and edge regions of the β-sheets and to the linker region (Fig. 4 *A* and *B*; for alignments, see Fig. S5). An alignment of the more closely related bovine and ovine RSV M proteins with human RSV M proteins demonstrates that the major variations in amino acid sequence are mostly found in regions at each end of the horseshoe structure, with the remainder occurring in the linker and the ends of the helix that lies on the outer surface of the N-terminal domain. Inclusion of the more distantly related metapneumoviral Ms in the alignment, such as those from human and avian metapneumoviruses, produces a similar pattern for the C-terminal domain; however, the differences are much more pronounced in the N-terminal domain, with significant sequence diversity occurring in the linker and in the regions adjacent to the linker. We hypothesize that these differences correlate with surface residues that mediate species-specific interactions, such as M with its cognitive nucleocapsid or viral glycoproteins. Given that membrane interactions are unlikely to be sequence-specific, and that protein activity is not transferred between species outside experimental conditions, this hypothesis is consistent with the involvement of loop regions in protein–protein and/or protein–RNA interactions, particularly for the N-terminal domain, where sequence diversity is greatest.

Model for Membrane Binding. The large positively-charged area on the surface of RSV M, which spans both domains, is consistent with the role of this protein in membrane association. We hypothesize that it is likely that this patch will be the driving force for association with the negatively-charged lung membrane (34, 35). This deduction would be consistent with biochemical observations of other matrix-like proteins, where electrostatic charge is the major component of the interaction between protein and membrane, at least in vitro. Association of the positively charged patch to the membrane, particularly at N- and C-terminal domain orientations other than those observed in the crystal structure, would leave a significant hydrophobic area on the protein exposed that may drive interactions with other viral components, or become buried at the protein–membrane interface.

Considering the C-terminal domain alone, although a significant proportion of the surface residues are hydrophobic (36), it is pertinent that this domain has a number of surface exposed arginine and lysine residues that are able to contribute to a favorable interaction with negatively-charged membranes. Membrane binding by the C-terminal domain, driven largely by electrostatic interactions, with a contribution from hydrophobic residues, would leave the N terminus free to perform more species-specific functions, such as protein–protein interactions needed for virion assembly. This interpretation would be consistent with observations on EBOV VP40 (30). We also note that

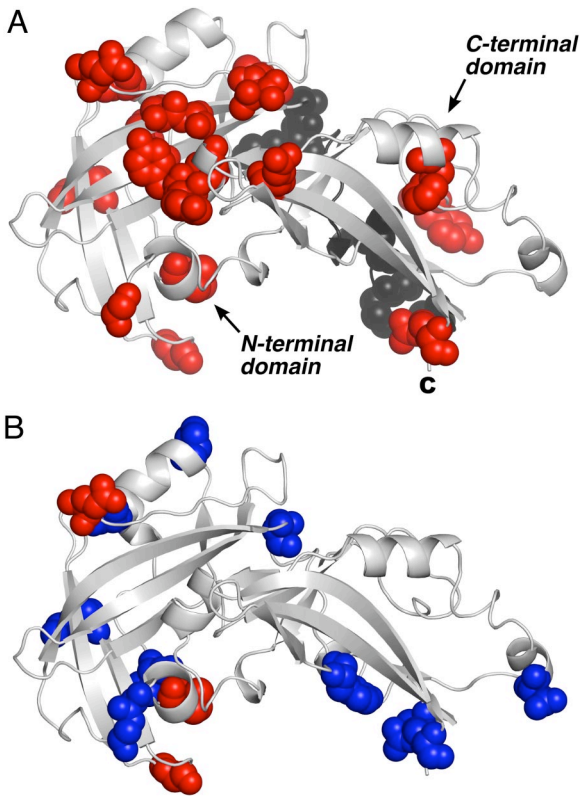


Fig. 4. Distribution of residues over the surface of RSV M that display significant sequence diversity from related proteins. (*A*) Residues with a Q score ≤20 after alignment of all 8 pneumovirus proteins to M^{254R} are displayed in red (nonlinker residues) or black (linker residues). The C terminus is indicated by C. (*B*) Residues with a Q score ≤20 or ≤50 after alignment of all RSV proteins are shown in red and blue, respectively. The orientation of the structure is the same as in *A*. Alignments are presented in Fig. S5.

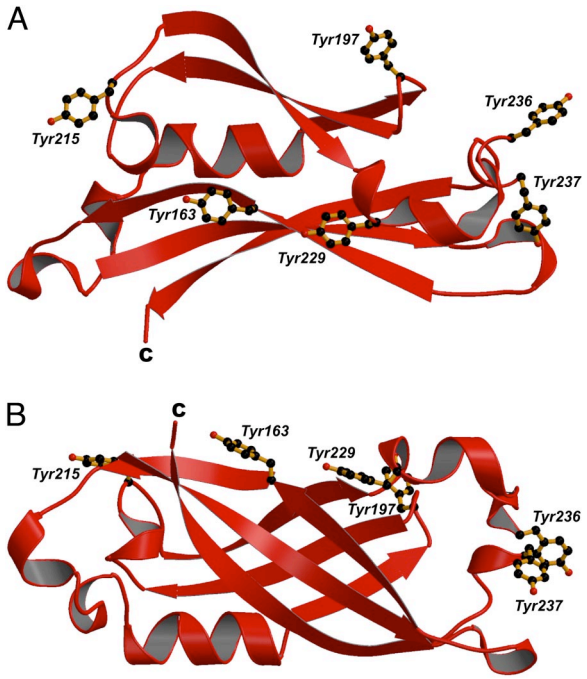


Fig. 5. Distribution of tyrosine residues in the C-terminal domain of the RSV M. The N-terminal domain of M^{254R} has been omitted for clarity. The planar distribution of the residues (shown as ball-and-stick representations) on the surface of the M protein is readily apparent. The C-terminal residue is indicated by C; *A* and *B* show orthogonal views of this domain.

the C-terminal domain has a striking arrangement of tyrosine residues (Y163, Y197, Y215, Y229, Y236, and Y237) forming a planar distribution that encircles the majority of the C-terminal domain (Fig. 5 *A* and *B*) with a degree of overlap with the positively-charged patch. Tyrosine, among other aromatic amino acids, has been demonstrated to occur with high probability in the interfacial region of membrane proteins (37–39), and in model systems has been shown to associate strongly with phospholipids (40). The functionality of the tyrosine residues has yet to be investigated, but the motif is conserved within all RSV Ms. The metapneumoviral Ms do not have this motif, although, in these proteins, functionally equivalent residues replace the tyrosines. Mutagenesis studies based on the structure presented within this article will further our understanding of the contribution of particular residues or features to the biological functionality of this protein.

Conclusion

The structure of the M of RSV has been solved by X-ray crystallography to a resolution of 1.6 Å. The high resolution and high degree of crystallographic order observed in the structure allow us to throw light on the mode of membrane binding and the mechanism by which this protein performs its varied and critical roles. Central to these roles, there is a significant area (≈600 Å^2) of positive electrostatic potential that forms an extended surface for interaction with the membrane, which carries a complementary negative potential. Because this area spans both domains and the linker, the geometry of surface contacts made by the protein will depend on the relative positioning of the domains, should those contacts involve contributions from both domains.

Experimental Procedures

Protein Expression Purification and Crystallization. A histidine tagged version of M^{254R} was expressed in *E. coli* strain BL21 (CodonPlus). Cells were lyzed by sonication and M^{254R} purified by nickel-affinity chromatography. After removal of insoluble material, the protein was subjected to crystallization trials. Successful crystallization was obtained in 70% Tacsimate, pH 7.0; later, the conditions were optimized to 55–65% Tacsimate, pH 7.0.

X-Ray Data Collection. Data from native, seleno-methionine derivatized, and mercury soaked crystals were collected at the European Synchrotron Radiation (Grenoble, France) (native crystal) and Stanford Synchrotron Radiation Lightsource (derivatized forms) facilities. The crystal statistics are presented in the *SI Materials and Methods*. The 3D figures presented were generated by using Molscript (Fig. 1*A*), PyMol (Fig. 1*B*), and the topology diagram rendered in TopDraw (see *SI Materials and Methods*).

Circular Dichroism Spectroscopy. Protein samples were dialysed against 5 mM phosphate buffer overnight at 4 °C. Far-UV CD spectra and the corresponding blanks were recorded between 190 and 250 nm in a cuvette of path length of 0.2 cm by using a Jasco J-810 Spectropolarimeter by averaging 8 accumulations recorded at a rate of 10 nm/min, with a pitch of 0.5 nm, a bandwidth of 1 nm, and a response time of 2 s. Near-UV spectra were recorded by using a 1-cm cell, with a pitch 0.2 nm and a response time of 1 s. After subtraction of the appropriate blank, binominal smoothing was carried out within the Jasco Spectra Analysis program. Smoothed data were analyzed for protein secondary structure by using the CDSSTR, SELCON3, and CONTIN/LL programs (41), accessed either via the Dichroweb service (42, 43), or the CDPro package. Both a general protein (SP43/dataset 4) and a membrane protein (SMP56/dataset 10) reference set were used (41, 44).

Full methods are available in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Professor E. N. Baker (University of Auckland) for support and advice on the preparation of this manuscript; the European Synchrotron Radiation Facility (Grenoble, France) and Stanford Synchrotron Radiation Lightsource for X-ray data collection facilities; and Dr. G. Sharples and Dr. E. Pohl (Durham University, Durham, United Kingdom) for critical reading of the manuscript. V.A.M. was supported by a European Molecular Biology Organization long-term fellowship. This work was also supported in part by contributions from OneNorthEast and from the Royal Society.

1. Fuentes S, Tran KC, Luthra P, Teng MN, He B (2007) Function of the respiratory syncytial virus small hydrophobic protein. *J Virol* 81:8361–8366.
2. Collins PL, Chanock RM, Murphy BR (2001) in *Fields Virology*, eds Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (Lippincott-Raven, Philadelphia), Vol 1, pp 1443–1486.
3. Murphy LB, *et al.* (2003) Investigations into the amino-terminal domain of the respiratory syncytial virus nucleocapsid protein reveal elements important for nucleocapsid formation and interaction with the phosphoprotein. *Virology* 307:143–153.
4. Murray J, Loney C, Murphy LB, Graham S, Yeo RP (2001) Characterization of monoclonal antibodies raised against recombinant respiratory syncytial virus nucleocapsid (N) protein: Identification of a region in the carboxy terminus of N involved in the interaction with P protein. *Virology* 289:252–261.
5. Henderson G, Murray J, Yeo RP (2002) Sorting of the respiratory syncytial virus matrix protein into detergent-resistant structures is dependent on cell-surface expression of the glycoproteins. *Virology* 300:244–254.
6. Brown G, *et al.* (2004) Analysis of the interaction between respiratory syncytial virus and lipid-rafts in Hep2 cells during infection. *Virology* 327:175–185.
7. Takimoto T, Portner A (2004) Molecular mechanism of paramyxovirus budding. *Virus Res* 106:133–145.
8. Peebles ME (1991) in *The Paramyxoviruses* (Plenum, New York), pp 427–456.
9. Chazal N, Gerlier D (2003) Virus entry, assembly, budding, and membrane rafts. *Microbiol Mol Biol R* 67:226–237.
10. Ghildyal R, Baulch-Brown C, Mills J, Meanger J (2003) The matrix protein of Human respiratory syncytial virus localises to the nucleus of infected cells and inhibits transcription. *Arch Virol* 148:1419–1429.
11. Rodriguez L, Cuesta I, Asenjo A, Villanueva N (2004) Human respiratory syncytial virus matrix protein is an RNA-binding protein: Binding properties, location and identity of the RNA contact residues. *J Gen Virol* 85:709–719.
12. Ruigrok RW, *et al.* (2000) Structural characterization and membrane binding properties of the matrix protein VP40 of Ebola virus. *J Mol Biol* 300:103–112.
13. Scianimanico S, *et al.* (2000) Membrane association induces a conformational change in the Ebola virus matrix protein. *EMBO J* 19:6732–6741.
14. Timmins J, Ruigrok RW, Weissenhorn W (2004) Structural studies on the Ebola virus matrix protein VP40 indicate that matrix proteins of enveloped RNA viruses are analogues but not homologues. *FEMS Microbiol Lett* 233:179–186.
15. Han Z, *et al.* (2003) Biochemical and functional characterization of the Ebola virus VP24 protein: Implications for a role in virus assembly and budding. *J Virol* 77:1793–1800.
16. Jayakar HR, Jeetendra E, Whitt MA (2004) Rhabdovirus assembly and budding. *Virus Res* 106:117–132.
17. Schmitt AP, Leser GP, Waning DL, Lamb RA (2002) Requirements for budding of paramyxovirus simian virus 5 virus-like particles. *J Virol* 76:3952–3964.

18. Sugahara F, *et al.* (2004) Paramyxovirus Sendai virus-like particle formation by expression of multiple viral proteins and acceleration of its release by C protein. *Virology* 325:1–10.
19. Gaudin Y, Barge A, Ebel C, Ruigrok RW (1995) Aggregation of VSV M protein is reversible and mediated by nucleation sites: Implications for viral assembly. *Virology* 206:28–37.
20. Riffel N, *et al.* (2002) Atomic Resolution Structure of Moloney Murine Leukemia Virus Matrix Protein and Its Relationship to Other Retroviral Matrix Proteins. *Structure* 10:1627–1636.
21. Hatanaka H, *et al.* (2002) Structure of equine infectious anemia virus matrix protein. *J Virol* 76:1876–1883.
22. Gaudier M, Gaudin Y, Knossow M (2002) Crystal structure of vesicular stomatitis virus matrix protein. *EMBO J* 21:2886–2892.
23. Dessen A, *et al.* (2000) Crystallization and preliminary X-ray analysis of the matrix protein from Ebola virus. *Acta Crystallogr D* 56:758–760.
24. Dessen A, Volchkov V, Dolnik O, Klenk HD, Weissenhorn W (2000) Crystal structure of the matrix protein VP40 from Ebola virus. *EMBO J* 19:4228–4236.
25. Holm L, Sander C (1996) Mapping the protein universe. *Science* 273:595–603.
26. Gibrat J, Made T, Bryant S (1996) Surprising similarities in structure comparison. *Curr Opin Struct Biol* 6:377–385.
27. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D* 60:2256–2268.
28. Faaberg KS, Peeples ME (1988) Association of soluble matrix protein of Newcastle disease virus with liposomes is independent of ionic conditions. *Virology* 166:123–132.
29. Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* 1:2876–2890.
30. Hoenen T, *et al.* (2005) VP40 Octamers Are Essential for Ebola Virus Replication. *J Virol* 79:1898–1905.
31. Hansen JC, Lu X, Ross ED, Woody RW (2006) Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J Biol Chem* 281:1853–1856.
32. Radivojac P, *et al.* (2007) Intrinsic disorder and functional proteomics. *Biophys J* 92:1439–1456.
33. Sickmeier M, *et al.* (2007) DisProt: The Database of Disordered Proteins. *Nucleic Acids Res* 35:D786–D793.
34. Dombrowsky H, Clark GT, Rau GA, Bernhard W, Postle AD (2003) Molecular species compositions of lung and pancreas phospholipids in the cftr (tm1HGU/tm1HGU) cystic fibrosis mouse. *Pediatr Res* 53:447–454.
35. Palestini P, *et al.* (2002) Composition, biophysical properties, and morphometry of plasma membranes in pulmonary interstitial edema. *Am J Physiol* 282:L1382–L1390.
36. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M (1995) Calculated using the ASC program. *J Comp Chem* 16:273–284.
37. Landolt-Marticorena C, Williams KA, Deber CM, Reithmeier RA (1993) Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins. *J Mol Biol* 229:602–608.
38. Sanderson JM (2005) Peptide-lipid interactions: Insights and perspectives. *Org Biomol Chem* 3:201–212.
39. Ulmschneider MB, Sansom MS (2001) Amino acid distributions in integral membrane protein structures. *Biochim Biophys Acta* 1512:1–14.
40. Sanderson JM, Whelan EJ (2004) Characterisation of the interactions of aromatic amino acids with diacetyl phosphatidylcholiney. *Phys Chem Chem Phys* 6:1012–1017.
41. Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* 287:252–260.
42. Lobley A, Whitmore L, Wallace BA (2002) DICHROWEB: An interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* 18:211–212.
43. Whitmore L, Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* 32:W668–W673.
44. Sreerama N, Woody RW (2004) On the analysis of membrane protein circular dichroism spectra. *Protein Sci* 13:100–112.