Recognition of HIV-1 Genomic RNA by Gag

Submitted for the degree of PhD

Ben Maddison B.Sc. (hons), (Warwick University)

August 1999

Department of Microbiology and Immunology
University of Leicester
UNIVERSITY OF LEICESTER

STATEMENT OF ORIGINALITY

The accompanying thesis submitted for the degree of PhD entitled

**Recognition of HIV-1 Genomic RNA by Gag**

is based on work conducted by the author in the Department of Microbiology and Immunology at the University of Leicester during the period between September 1996 and June 1999.

All the work described in this thesis is original unless otherwise acknowledged in the text or by references. The following plasmid constructs, AL-2, B1-1, Cl-4, D-1, pI+, pJ+, pG+, pL, pM, and pN were constructed by Dr Pankaj Marya, pMH93.010, pMH93.017, pMH94.004, pMH95.001 and pGEX-p55 were constructed by Dr Malcolm Haddrick, pSH91.200 was constructed by Dr Shaun Heaphy, pJR was constructed by Dr Jackie Russell and pAL94.113 by Dr Andrew Lear.

None of the work has been submitted for another degree in this or any other university.

Signed: ___________________________ Date: 29th Aug 1999
Abstract

In the HIV-1 infected cell, a specific recognition event has to take place between the structural polyprotein precursor known as gag, and the virion genomic RNA. This interaction is essential for ensuring that the assembling virus packages a genome. The interaction is specific as cellular RNAs and spliced genomic RNAs are largely excluded from the virus particle. Several published reports have suggested that the p55 gag protein specifically interacts with RNA sequences in the 5' untranslated region of the virion genome. This thesis describes *in-vitro* investigations designed to further understand this interaction with the aim of determining a minimal binding site for the HIV-1 gag protein on the genomic RNA. Using a deletion type approach the interaction was analysed by filter binding techniques with a GST-p55 gag fusion protein and *in-vitro* transcribed RNA. This interaction was then subsequently investigated by ribonuclease footprinting. These methodologies identified regions of RNA rich in secondary structure, between the 5' splice donor and up to and including the gag start codon, which interact with GST-p55 in solution.

The location of the gag-RNA interaction in infected cells had not been established. I investigated this using subcellular fractionation techniques, UV photo affinity cross-linking of RNA-protein interactions, immunoprecipitation and RT-PCR. These experiments show that gag-RNA interactions do not occur in the nuclei of the infected cell. This interaction probably occurs at some location in the soluble cytoplasm, as gag-RNA complexes could be identified in the soluble cytoplasmic fractions after fractionation on sucrose gradients which separate out cytosol and membrane fractions. Preliminary experiments suggest that gag interacts with HIV-1 genomic RNA sequences in the cytoplasm and not subgenomic HIV-1 RNAs, as required for virion morphogenesis.
Publication

Acknowledgements

Acknowledgements are due to all the staff past and present, of the Department of Microbiology and Immunology at the University of Leicester who have offered assistance during this project. I am specifically grateful to my supervisors Shaun Heaphy and Alan Cann for their guidance during this project. I would also like to thank members of Lab 136 for their encouragement and help, notably Pankaj Marya for filter binding work, Alan Lau for tissue culture and CIII training, Malcolm Haddrick for general help and advice and Sarah Rogan for protein expression and analysis. Most importantly I am grateful to Susie who, behind the scenes has housed, fed and watered me, and has given me the support and encouragement needed during the last 3 years.
For my Parents
Contents

Title page. i
Statement of originality. ii
Abstract. iii
Publication. iv
Acknowledgements. v
Dedication v
Contents. vii
Abbreviations. xii
List of figures. xiv
List of tables. xvi

Chapter 1

1.1 Retroviruses. 1
1.2 Discovery of HIV and AIDS. 4
1.3 HIV-1 Pathogenesis. 6
Clinical aspects of infection. 6
1.4 Molecular biology of HIV-1. 10
1.4.1 Physical structure. 10
1.4.2 The structure of HIV-1 genome. 12
1.4.3 The infectious cycle. 16
1.4.4 Structure of genomic RNA. 20
1.5 RNA dimerisation. 23
1.5.1 Retroviral RNA dimers. 23
1.5.2 Advantages of a dimeric genome. 25
1.5.3 Mechanism of RNA dimerisation. 26
1.5.4 In-vitro analysis of RNA dimerisation. 27
1.6 Gag proteins and their role in virion morphogenesis. 33
1.6.1 Features of the structural proteins. 33
1.7 RNA packaging. 40
1.7.1 Cis acting signals. 40
1.7.2 Trans acting factors. 41
1.7.3 Gag-RNA interactions analysed in-vitro. 42
1.7.4 In-vivo packaging studies. 43
1.7.5 Regulation of RNA packaging. 45
1.8 Project outline. 46

Chapter 2

2.1 DNA techniques. 47
2.1.1 Agarose gel electrophoresis. 47
2.1.2 Restriction digestion. 47
2.1.3 Phenol-chloroform extractions. 48
2.1.4 Ethanol precipitation. 48
2.1.5 Gel purification of DNA. 48
2.1.6 Ligations. 49
2.1.7 Preparation of transformation competent \textit{E.coli} cells. 49
2.1.8 Transformation of \textit{E.coli}. 49
2.1.9 Isolation of plasmid DNA. 50
    Miniprep procedure. 50
    Caesium chloride maxipreps. 51
2.1.10 Polymerase chain reaction. 51
2.1.11 DNA sequencing. 52
2.1.12 Plasmid construction. 53
2.1.13 Oligonucleotides. 53
2.1.14 Estimation of primer concentration. 55
2.2 RNA work. 55
    2.2.1 Precautions taken when working with RNA. 55
    2.2.2 \textit{In-vitro} transcription. 55
    2.2.3 5' end labelling of RNA. 57
    2.2.4 \textit{In-vitro} RNA dimerisation. 57
    2.2.5 Extraction of virion genomic RNA. 58
    2.2.6 Northern blot analysis of vRNA. 59
    2.2.7 Random prime DNA probes. 59
    2.2.8 RNA filter binding assay. 60
    2.2.9 RNAse footprinting RNA-protein interactions. 61
    2.2.10 3' boundary analysis of H4/GST-p55 interaction. 61
    2.2.11 UV-cross linking RNA-protein interactions. 62
    2.2.12 Identification of gag in protein-RNA complexes. 62
    2.2.13 Sucrose density gradient purification of membranes. 62
    2.2.14 cDNA synthesis. 63
    2.2.15 Immunoprecipitation of gag-RNA complexes. 63
2.3 Cell culture and virus techniques. 64
    2.3.1 Cell culture. 64
    2.3.2 Counting cells. 64
    2.3.3 Storing cells. 64
    2.3.4 Handling cells. 64
    2.3.5 Virus stocks. 65
    2.3.6 Virus infection of C8166 cells. 65
    2.3.7 Nuclei purification and extraction. 66
        NP-40 extraction. 66
        Dignam extraction. 67
    2.3.8 \textsuperscript{3}H myristic acid labelling of infected cells. 67
2.4 Protein Analysis. 67
    2.4.1 SDS polyacrylamide gel electrophoresis. 67
    2.4.2 Coomassie blue staining of protein gels. 68
    2.4.3 Western blotting of protein gels. 68
    2.4.4 ECL detection of western blots. 69
    2.4.5 Alkaline phosphatase detection of protein in western blots. 70
2.5 Expression of HIV-1 p55 in \textit{E.coli}. 70
    2.5.1 Protein expression. 70
    2.5.2 Protein extraction. 71
    2.5.3 Protein purification. 71
        Binding to Glutathione sepharose 4B. 72
        FXa cleavage. 72
Chapter 3

3.1 Introduction to kissing loop dimer work. 75
3.2 Mechanisms of in-vitro RNA dimerisation. 76
3.3 Development of an in-vitro dimerisation competition assay. 80
3.4 Competition experiments with extracted virion genomic RNA. 87
3.5 Summary 91
3.5.1 Kissing loop dimers can be dissociated by competing oligonucleotides in-vitro. 91
3.5.2 Dimers of extended conformation can be dissociated by oligonucleotides in formamide buffers. 91
3.5.3 Genomic RNA is stable to conditions that dissociate in-vitro RNA dimers. 92
3.5.5 Possible reasons for genomic RNA stability. 92

Chapter 4

4.1 Introduction to in-vitro identification of RNA binding sites for p55 on the leader sequence of HIV-1. 95
4.2 GST-p55 protein purification. 97
4.3 Deletion analysis of nucleotides in the 5' leader sequence that are required for GST-p55 binding. 98
4.4 5' and 3' deletions centre binding on SL4. 100
4.5 RNA sequences flanking SL4 don’t bind GST-p55. 100
4.6 Minimal binding RNAs are centred on SL4. 100
4.7 Leader RNAs deleted in SL4 bind GST-p55. 101
4.8 H4 and H21 have similar binding affinity for GST-p55. 101
4.9 A minimal SL4 RNA requires small 5' or 3' sequence. 103
4.10 Boundary determination for H4/GST-p55 interaction. 103
4.11 Role of kissing loop dimer in RNA-gag recognition. 105
4.12 Ribonuclease footprinting identifies further p55-RNA binding sites. 106
4.13 T1 cleavage of GST-p55/H18 doesn’t reveal any points of differential cleavage. 107
4.14 V1 cleavage of GST-p55/H18 shows that GST-p55 may bind to the stem sequences in SL4. 108
4.15 RNAse T1 footprinting of GST-p55/H4 reveals sites of protection in SL3. 109
4.16 RNAse V1 footprinting of GST-p55/H4 reveals sites of protection in SL3/SL4. 111
4.17 Summary of ribonuclease footprinting. 112
4.18 Northwestern analysis of p55-RNA interactions. 114
4.19 Northwestern binding of HIV-1 RNA to immobilised GST-p55 is dependent on KCl concentration. 115
4.20 RNA binding to immobilised p55 is not specific.
4.21 GST-p55 shows no difference in affinity towards H4 or C1 when immobilised to a nitrocellulose filter.
4.22 Northwestern blotting using extracts from infected cells.
4.23 Summary.
4.23.1 Filter binding.
4.23.2 RNA dimerisation is not required for RNA binding.
4.23.3 RNAse footprinting identifies RNA binding sites 5' of SL4.
4.23.4 Northwestern blotting.
4.23.5 Future experiments.

Chapter 5

5.1 Introduction to p55 expression.
5.2 Cleavage of GST-p55 by FXa.
5.3 Cleavage of GST-p55 on GS4B Matrix.
5.4 Detergent activated cleavage of GST-p55.
5.5 Cleavage and immunoaffinity purification of p55.
5.6 Expression of p55 from pET16B-p55 construct.
5.7 Expression and purification of p55 from pTYB4-p55.
5.8 Induction of p55 from pTYB4-p55 in E. coli.
5.9 Immunoaffinity purification of p55.
5.10 Summary.
5.10.1 P55 purification from FXa cleaved GST-p55.
5.10.2 Expression of p55 from pET16B-p55.
5.10.3 Expression of p55 from pTYB4-p55.
5.10.4 Future experiments.

Chapter 6

6.1 Introduction to in-vivo packaging studies.
6.2 Infected cells can be fractionated into nuclear and cytoplasmic fractions.
6.3 Gag proteins localise to the nucleus of the infected cell.
6.4 Gag appears in the nucleus 24hours post infection.
6.5 p55 gag builds up in the nuclei of the infected cell.
6.6 Myristylation state of nuclear p55.
6.7 Nuclear p55 does not bind RNA, extranuclear p55 does.
6.8 Membrane and cytoplasmic fractionation.
6.9 Extraction and analysis of RNA-protein interactions isolated.
6.10 Confirmation of membrane fractionation by gp120/160 distribution.
6.11 p55 UV-cross links with RNA from the cytoplasmic fraction.
6.12 p55 gag binds genomic RNA and not subgenomic RNA.
6.13 Summary.
6.13.1 HIV-1 p55 co-purifies with nuclear and cytoplasmic fractions.
6.13.2 p55 accumulates with the nuclear fraction over time.
6.13.3 Myristylated p55 can be detected in the nucleus.
6.13.4 Photoaffinity cross linking detects p55-RNA complexes in the
Cytoplasm and not the nucleus.

6.13.5 Analysis of membrane bound and cytoplasmic p55-RNA interactions.

6.13.6 RT-PCR analysis.

6.13.7 Future experiments.

Chapter 7

7.1 RNA Dimerisation
7.2 RNA-Protein interactions in-vitro
7.3 RNA-Protein interactions in-vivo
7.4 Genome packaging in other viruses
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACS</td>
<td>Autocomplementary sequence</td>
</tr>
<tr>
<td>AIDS</td>
<td>Acquired Immune Deficiency Syndrome</td>
</tr>
<tr>
<td>ASLV</td>
<td>Avian Sarcoma Leukosis Virus</td>
</tr>
<tr>
<td>BLV</td>
<td>Bovine leukaemia virus</td>
</tr>
<tr>
<td>CA</td>
<td>Capsid protein</td>
</tr>
<tr>
<td>CBD</td>
<td>Chitin binding domain</td>
</tr>
<tr>
<td>CIP</td>
<td>Calf intestinal alkaline phosphatase</td>
</tr>
<tr>
<td>CMV</td>
<td>Cytomegalovirus</td>
</tr>
<tr>
<td>cpm</td>
<td>Counts per minute</td>
</tr>
<tr>
<td>DLS</td>
<td>Dimer Linkage site</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl Sulphoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxy ribonucleotide triphosphate</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethyl pyrocarbonate</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediamine Tetra-acetic Acid</td>
</tr>
<tr>
<td>EM</td>
<td>Electron microscopy</td>
</tr>
<tr>
<td>FCS</td>
<td>Foetal calf serum</td>
</tr>
<tr>
<td>FXa</td>
<td>Factor Xa</td>
</tr>
<tr>
<td>GS4B</td>
<td>Glutathione sepharose 4B</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione-S-transferase</td>
</tr>
<tr>
<td>HaSV</td>
<td>Harveys sarcoma virus</td>
</tr>
<tr>
<td>HEPES</td>
<td>N-2-Hydroxyethylene piperazine-N-2-ethane sulphonic acid</td>
</tr>
<tr>
<td>HFV</td>
<td>Human Foamy Virus</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>HTLV</td>
<td>Human T-Cell Leukaemia virus</td>
</tr>
<tr>
<td>IN</td>
<td>Integrase</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>KL</td>
<td>kissing-loop</td>
</tr>
<tr>
<td>KiSV</td>
<td>Kirsten sarcoma virus</td>
</tr>
<tr>
<td>LTR</td>
<td>Long Terminal Repeat</td>
</tr>
<tr>
<td>MA</td>
<td>Matrix protein</td>
</tr>
<tr>
<td>MMTV</td>
<td>Mouse mammary tumour virus</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MuLV</td>
<td>Murine Leukaemia virus</td>
</tr>
<tr>
<td>NC</td>
<td>Nucleocapsid protein</td>
</tr>
<tr>
<td>NLS</td>
<td>Nuclear localisation signal</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PBS</td>
<td>Primer binding site</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PMSF</td>
<td>Phenyl-methyl Sulphonyl Fluoride</td>
</tr>
<tr>
<td>PR</td>
<td>Protease</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinyl Difluoride</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>REV-A</td>
<td>Reticulo-Endothelial Virus (A)</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>rNTP</td>
<td>ribonucleotide triphosphate</td>
</tr>
<tr>
<td>RT</td>
<td>Reverse transcriptase</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse-transcription polymerase chain reaction</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulphate</td>
</tr>
<tr>
<td>SIV</td>
<td>Simian immunodeficiency virus</td>
</tr>
<tr>
<td>SL(1-4)</td>
<td>Stem-loop (1-4)</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-Acetate EDTA</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-Borate EDTA</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris buffered saline</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA</td>
</tr>
<tr>
<td>Tm</td>
<td>melting temperature</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer Ribonucleic acid</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>v/v</td>
<td>Volume by volume</td>
</tr>
<tr>
<td>w/v</td>
<td>Weight by volume</td>
</tr>
</tbody>
</table>
List of Figures

Chapter 1

Figure 1.1  Formation and release of retroviral particles.  4
1.2  Budding and maturation of HIV-1.  10
1.3  Structure of mature HIV-1.  11
1.4  Genetic organisation of HIV-1.  12
1.5  The infectious cycle.  16
1.6  RNA splicing.  19
1.7  HIV-1 genomic RNA structure.  21
1.8  Secondary structure of the Leader RNA.  22
1.9  Electron micrograph of RSV dimeric RNA.  24
1.10  The kissing loop interaction.  29
1.11  Location of the gag gene.  35
1.12  p17 Matrix functional domains.  35
1.13  p24 Capsid functional domains.  36
1.14  p6 protein domains.  39
1.15  Sequence and zinc finger domains of p7 Nucleocapsid.  41

Chapter 2

Figure 2.1  Random prime probe preparations.  60

Chapter 3

Figure 3.1  *In-vitro* transcribed RNAs.  76
3.2  Agarose gel of phenol dimerised RNA.  77
3.3  RNA dimer structures.  78
3.4  Effect of TBE and TBM$_{0.1}$ on kissing loop dimers.  79
3.5  Effect of TBM$_{0.01}$ on kissing loop dimer.  80
3.6  Annealing positions of competing oligonucleotides.  81
3.7A  Kissing loop dimer competition with MH14 control.  82
B  Kissing loop dimer competition with MH20.  83
3.8  Graph showing kissing loop dimer competition by oligonucleotides.  84
3.9  Competition using BM5 on phenol dimer at elevated temperature.  85
3.10  Phenol dimer stability in formamide buffers.  86
3.11  Melting phenol dimer with BM5.  87
3.12  Stability of HIV-1 genomic RNA to formamide buffers.  88
3.13  Competition for kissing loop and mature dimers in virion RNA.  89
3.14  Extended competition incubation for virion dimer.  90
3.15  Proposed further points of contact described from EM studies of HIV-1 RNA.  93
Chapter 4

Figure 4.1A Filter binding curve for GST-p55 and GST with HIV RNA. 96
B Filter binding curve showing specificity of GST-p55 for HIV-1 RNA.
C Gel shift experiments. 96
4.2 GST-p55 purification. 97
4.3 Structure of leader RNA, and RNAs constructed. 99
4.4 Filter binding curve comparing H3 with H21. 101
4.5 tRNA competition filter binding. 102
4.6 Minimal RNA binding domains. 103
4.7 3' boundary analysis. 104
4.8 Filter binding curve for KL+/-RNA. 105
4.9 Ribonuclease footprint H18/GST-p55 using T1 RNAse. 107
4.10 Ribonuclease footprint H18/GST-p55 using V1 RNAse. 108
4.11 Ribonuclease footprint H4/GST-p55 using T1 RNAse. 110
4.12 Ribonuclease footprint H4/GST-p55 using V1 RNAse. 111
4.13A Sites of V1 and T1 cleavage in RNA278-362. 112
B Sites of V1 and T1 protection in RNA278-362. 113
4.14 Northwestern blot GST-p55/H4. 115
4.15 Comparison of northwestern blot using C1 and H4 RNA probes. 116
4.16 Binding of RNA in northwestern blot using increasing amounts tRNA competitor. 117
4.17 Northwestern blot probing infected cell extracts. 118

Chapter 5

Figure 5.1 Protease cleavage sites in GST-p55. 124
5.2 Solution cleavage of GST-p55 with FXa. 124
5.3 Cleavage GST-p55 bound to GS4B Matrix using FXa. 125
5.4A Cleavage including 0.5% SDS. 126
B Cleavage including 0.1% SDS. 127
C Detergent comparison in cleavage reaction. 127
5.5 Attempt at immunoaffinity purification FXa cut GST-p55. 128
5.6 Induction p55 from pET16B-p55 in BL21 cells. 129
5.7 Expression p55 in BL21 vs B834 cells. 130
5.8A Diagram showing p55 expression from pTYB4-p55. 131
5.8B Induction p55 from pTYB4-p55. 132
5.9 Solubility of p55 expressed in B834/BL21 cells. 133
5.10A Immunoaffinity purification p55 fractions. 134
B Protein purification steps. 134

Chapter 6

Figure 6.1A Vinculin does not contaminate nuclear fractions. 138
B gp120 does not contaminate nuclear fractions. 138
6.2A Nuclear localisation of HIV-1 gag proteins. 139
B Possible gag cleavage products. 140
6.3 Time course of nuclear gag accumulation.

6.4 Growth curve for p55 nuclear accumulation experiment.

6.5 Nuclear-cytoplasmic accumulation of gag over extended time course.

6.6 p24 titres during extended time course experiment.

6.7 Myristylated proteins in cytoplasmic and nuclear fractions.

6.8A Immunoblot of p55 band in myristic acid labelling experiment.

B Plot of p55 pixel density in nuclear and cytoplasmic fractions.

C Plot of 3H distribution in cytoplasmic and nuclear fractions.

6.9A Extracted RNA run on agarose gel.

B Distribution of p55 gag in UV cross linking experiment.

C p55 UV cross links with RNA in the cytoplasm and not the nucleus.

6.10 Membrane fractionation immunoblot.

6.11A method to identify RNA-p55 complexes in cytoplasm and membrane.

B Sucrose density gradient plot.

C p24/vinculin immunoblot of fractions from sucrose gradient.

6.12 p17 immunoblot of fractions from sucrose gradient.

6.13 gp120/160 distribution in cytoplasmic and membrane fractions.

6.14 p55 UV-cross links with RNA in the cytoplasmic fractions.

6.15 Diagram showing RT-PCR amplifications.

6.16 RT-PCR results from p55 immunoprecipitation.

**Chapter 7**

Figure 7.1  *In-vivo* interactions identified.

**Tables**

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>2</td>
</tr>
<tr>
<td>2.1</td>
<td>57</td>
</tr>
<tr>
<td>2.2</td>
<td>68</td>
</tr>
<tr>
<td>3.1</td>
<td>81</td>
</tr>
</tbody>
</table>

**Tables**

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>2</td>
</tr>
<tr>
<td>2.1</td>
<td>57</td>
</tr>
<tr>
<td>2.2</td>
<td>68</td>
</tr>
<tr>
<td>3.1</td>
<td>81</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Retroviruses
Retroviruses are classified into 3 main families by virtue of their pathogenicity (see table 1). They are associated with many diverse diseases, which include both rapid and long-latency malignancies, immunodeficiencies, and infections that are characterised by lifelong viraemias with the absence of any apparent ill effect. These 3 families are known as the Oncoretroviruses, which include Rous Sarcoma virus (RSV) and Murine Leukaemia Virus (MuLV), the Lentiviruses which include Human Immunodeficiency Virus (HIV), Simian Immunodeficiency Virus (SIV) and Visna (from sheep) and the Spumaretroviruses, of which there are many human and primate isolates. Acutely transforming oncoretroviruses such as RSV are able to cause cancer by the virtue of an acquired oncogene. Non-acutely transforming oncoretroviruses occasionally produce disease by insertional activation of cellular oncogenes (e.g. mouse mammary tumor virus, MMTV). The pathogenicity of spumaretroviruses is not well defined. These viruses typically cause vacuolation or ‘foaming’ of cells in tissue culture. They have been isolated from a number of mammalian species, however the persistent infection by these agents is not associated with disease (Weiss et al 1996). Lentiviruses on the otherhand are characterised by slow disease typified by long latency.

Over recent years there has been intensive study of lentiviruses after the discovery of the important human pathogens Human T-cell Leukaemia Virus-I (HTLV-1) (Poiesz et al 1981) and Human Imunodeficiency Virus (HIV) (Gallo et al 1984, Popovic et al 1984). These two viruses are of important medical significance due to the severity of the diseases that they cause, and the number of infected individuals. This has meant that over the last 15 years these viruses have been extensively studied.
### Table 1 Retrovirus Classification

<table>
<thead>
<tr>
<th>Subfamily</th>
<th>Group</th>
<th>Example Isolates</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oncovirinae</td>
<td>Avian Leukosis</td>
<td>Rous Sarcoma Virus</td>
<td>Transforming virus contains Src</td>
</tr>
<tr>
<td></td>
<td>Sarcoma</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mammalian C-type</td>
<td>Murine Leukaemia Virus</td>
<td>Causes T-cell lymphoma</td>
</tr>
<tr>
<td></td>
<td>B-type viruses</td>
<td>Mouse Mammary Tumour Virus</td>
<td>Causes mammary carcinoma</td>
</tr>
<tr>
<td></td>
<td>D-type viruses</td>
<td>Mason Pfizer Monkey Virus</td>
<td>Unknown pathogenicity</td>
</tr>
<tr>
<td></td>
<td>HTLV-BLV groups</td>
<td>Human T-cell Leukaemia Virus</td>
<td>Causes T-cell lymphoma</td>
</tr>
<tr>
<td>Lentivirinae</td>
<td>Lentivirus</td>
<td>Human Immunodeficiency viruses 1 and 2</td>
<td>Causes AIDS in humans</td>
</tr>
<tr>
<td>Spumavirinae</td>
<td>Foamy virus</td>
<td>Simian Foamy Virus</td>
<td>Cytopathic and syncitium inducing in-vitro</td>
</tr>
</tbody>
</table>

Retroviruses belong to the group 6 virus family according to the Baltimore virus classification scheme. These viruses possess a diploid RNA genome. The two RNAs that make up the genome are transcribed from a proviral sequence which is integrated within the host cell chromosome. The RNAs are typically around 9.5kb in length and are both capped and polyadenylated like any other cellular mRNA. Unlike other viral RNA genomes, these RNAs are not translated into protein on infection of a new cell. Instead, the RNA is copied by a viral associated RNA dependent DNA polymerase (reverse transcriptase) into a double stranded DNA copy of the virion genome. This DNA copy is then inserted into the chromosome of the infected cell. These steps
(Baltimore 1970), peculiar to retroviruses, are an exception to the central dogma describing the flow of genetic information from DNA to RNA to protein.

Retroviruses can also be distinguished by their mechanism of assembly into an virus particle, and their acquisition of an outer envelope. Oncoviruses comprise particles of the types A to D. Type A structures represent non-infectious cytoplasmic assembled gag shells most often derived from endogenous retroviral sequences. B-type retroviral particles assemble into gag cores which migrate to the membrane and bud through it acquiring their plasma membrane derived envelope. C-type oncoretroviruses form a crescent like shell of gag precursor protein at the plasma membrane through which the assembling particle buds and gains its lipid membrane (see figure 1.1) type C assembly also describes HIV assembly. D type viruses such as Mason Pfizer monkey virus (MPMV) also form intracytoplasmic cores as in type B particles, which also bud through the plasma membrane and acquire their envelope, the shape of the virion core differs to that of type B virus and the surface projections of the D type virion are longer than those of type B. This method of virus characterisation is primarily by electron microscopy.

The formation of the virus particle from the gag precursor molecules is driven by self assembly properties of the proteins (e.g. Campbell and Vogt 1997). A single point mutation within the Matrix domain of MPMV was able to switch the assembly of a type D retrovirus to that resembling assembly of a C type virus (Rhee et al 1990). This was thought to be due to a putative intracellular targeting motif in type D retroviruses, recognised in the 3D NMR structure of MPMV Matrix protein and not present in type C retroviruses. (Conte et al 1997). This has recently been confirmed by inserting an 18 amino acid sequence from this region of MPMV to the same region in a type C retrovirus, and this caused cytoplasmic targeting and retention of this protein (Choi et al 1999).
1.2 Discovery of HIV and AIDS.

In 1981 a collection of conditions associated with iatrogenic immunosuppression such as Kaposi's sarcoma, oral thrush and pneumocystic pneumonia were described in a group of young men. This collection of symptoms in these individuals became referred to as the Acquired Immune Deficiency Syndrome or AIDS. It was soon established that there were high risk activities recognised in those men who showed symptoms of AIDS. The epidemiology, or distribution of cases was characteristic of transmission by blood or seminal fluid contact, with high risk groups including intravenous drug abusers, recipients of donated blood products and a high incidence in homosexual men.
In 1983 a reverse transcriptase-containing retrovirus was isolated from a patient suffering with lymphadenopathy (Barre-Sinousses et al. 1983). This virus caused cytopathic changes to T-cells in culture and was named Lymphadenopathy Associated Virus or LAV. This virus showed similar properties to a virus previously isolated, Human T-cell leukaemia virus I (HTLV-I), normally associated with T-cell leukaemia's in man. The type of infection characterised by HTLV-I was different to LAV. HTLV had been shown to immortalise T-cells in tissue culture and was primarily cell associated. Another retrovirus, named HTLV-III was subsequently isolated from a patient with AIDS (Popovic et al. 1984). This isolate showed some antigenic cross reactivity to HTLV-I and HTLV-II (a retrovirus related to HTLV-I) but like LAV caused cytopathic effects in T-lymphocytes. At this time a further retrovirus was isolated from an AIDS patient and was given the name AIDS associated retrovirus or ARV (Levy et al. 1984). This virus also caused cytopathic effects on T-cells and cross reacted with antisera to LAV.

Thus LAV, HTLV-III and ARV all showed similar characteristics and were distinct from HTLV-I. By 1986 a proposal was put forward, now universally accepted, to call these viruses the Human Immunodeficiency Virus or HIV (Coffin et al. 1986). Due to the high levels of homology between HTLV-III and LAV, not generally seen between two different retroviral isolates, archived material was re-examined and it is now accepted that LAV and HTLV-III were probably the same virus and that laboratory culture contamination had probably occurred with these specimens (Chang et al. 1993).

Another retrovirus, similar to but antigenically distinct from the isolates previously described was isolated from patients in West Africa (Clavel et al. 1986). Infection by this virus also leads to AIDS. The original isolates were thus named HIV-1 and the newly recognised strains with epicentres in West Africa, HIV-2.

**Origin of HIV Infection in Man**

Before the start of the present HIV pandemic there may have been isolated cases of the disease in humans dating back to the late 1950s (Zhu et al. 1998). The virus had generally been thought of as a simian virus which may have crossed species at some
point, and in the case of HIV-2 does show antigenic cross reactivity to SIVmac (Clavel et al 1986). The primate reservoir of HIV-2 was identified as the Sooty Mangabey (Gao et al 1992). Recently a natural host for an HIV-1 related virus has been described in a West African population of chimpanzee (Gao et al 1999). It is likely that the zoonotic transmission of these viruses to man has been brought about by the fact that chimpanzees are commonly hunted for food, especially in West Africa.

1.3 HIV-1 Pathogenesis

Clinical Aspects of Infection

It was realised that in addition to AIDS there was a number of clearly defined clinical stages associated with HIV infection. After each clinical stage has been diagnosed patients do not revert to earlier stages if the signs of symptoms settle. Diagnosis is based on the Centre for Disease Control (CDC) classification system, and white blood cell count.

Initial infection (CDC stage I)

A self limiting illness whilst seroconverting to HIV. This primary infection may be associated with acute clinical signs such as a glandular fever like illness, including sore throat, swollen lymph nodes, and non-specific flu-like symptoms. Seroconversion usually occurs between 1-10 weeks post infection. After this point antibody to the virus can be detected in the serum.

Asymptomatic Infection (CDC stage II)

Following the acute infection a complete clinical recovery occurs where an infected person has no symptoms. The asymptomatic time period of the infection may last from months to years.

Symptomatic Infection (CDC stage III)

Infected individuals at this stage of the infection show enlarged lymph nodes in the absence of any other illness. Further clinical signs may also include tiredness, lethargy, sweating and muscular pains.
AIDS (CDC stage IV)

This stage is divided into five subsections, each stage indicating the addition of more serious problems associated with immunosuppression. A variety of symptoms are present at this stage e.g. opportunistic infections such as candida, toxoplasma and CMV. Certain malignancies also manifest themselves, such as Kaposi sarcoma. General complications also arise in the kidneys, heart, lung, blood and nervous system. Survival times for the patient once diagnosed with AIDS, in the absence of any drug therapy regime, is typically less that one year.

Therapy

Efforts to control HIV in the infected patient have so far been concerned with anti-retroviral drugs targeted towards virus specific enzymatic activities. The first licensed drugs such as AZT were targeted towards the reverse transcriptase. These drugs delay the onset of disease but do not prevent it. More recently several protease inhibitor drugs have been licensed, which specifically target the viral protease. Due to the extremely high mutation rate of virus within an individual (Coffin 1995), there is a generation of drug resistant viruses in the infected patient and then a progression to disease. The introduction of drug cocktails have proved effective at reducing viral loads to undetectable levels for up to 2 years (Gulick et al 1997). There are problems associated with this regime however, as virus can apparently be maintained in reservoirs in the host that are not sensitive to antiretroviral drugs (Wong et al 1997). In these cases selection of drug resistant strains and progression to disease in the absence of further therapy, may in the longer term, be a matter of course. The intensive study of HIV molecular biology has, and will, yield further targets for the rational design of therapeutic compounds or methodologies.

Epidemiology of HIV

At the beginning of 1998 the World Health Organisation predicted that over 12 million people had already lost their life to AIDS. At this time a further 30 million people were probably infected with HIV. These infections are not evenly distributed globally. The highest concentrations of infected cases are in the countries of the developing world, mostly in those countries that cannot afford to care for the infected
people. It is estimated that 89% of people infected with HIV live in sub-Saharan Africa and in the developing countries of Asia.

Pathogenesis of HIV infection
Patients infected with HIV and those patients in AIDS were originally noted as being depleted in a subset of T-lymphocyte positive for the CD4 surface antigen. It was later shown that the natural host cell in the infected individual is the CD4+ T helper cell (Klatzmann et al 1984). CD4 was shown to be the primary component of the HIV-1 cell surface receptor (Dalgliesh et al 1984). Other CD4+ cells can also be infected by HIV, such as macrophages (Gendelman et al 1989) and also neuronal cells such as astrocytes and ganglia cells (Levy et al 1994).

Primary infection by HIV results in a high level of viraemia which disseminates virion throughout the body and the lymphoid organs. A powerful immune response is mounted, consisting of both humoral and cell mediated immune responses (Borrow et al 1994, Moore et al 1994). This reduces the amount of viraemia in the patient to about 10% of the peak viraemic load, however the virus is not completely eradicated. It is thought that during the period of clinical latency the virus is active and continuously replicates at high levels. The lymphoid organs are regarded as the primary sites of virus replication and sequestration (Pantaleo et al 1991, 1993). Thus clinical latency does not correlate with viral latency.

The amount of virus turned over per day is extremely high (Ho et al 1995), there is a high turn over of CD4+ T-cells, with a large number of these cells being lost and replaced daily. The mechanisms of T-cell depletion are still unclear, no main mechanism for T-cell depletion has been recognised in-vivo, although cytopathicity is readily observed in-vitro. Viral infection of CD4+ cells is thought to result in depletion of CD4+ T-helper cells. These cells recognise antigen complexed with MHC-II molecules on the surface of macrophages or antigen presenting cells and are instrumental in activation of cell mediated immune responses. CD4+ cell depletion may be caused by a number of mechanisms. These may include syncitium formation between infected and non-infected cells, the selective infection and destruction of
memory T cells and inappropriate killing of uninfected cells. It is also possible that Th cell death may occur by apoptosis (Gougon and Montagnier 1993).

Ultimately the HIV specific immune responses fail to clear the virus. This is determined by a few clearly distinct mechanisms. The virus is able to form pools of latently infected CD4+ cells, which don’t get eliminated by cytotoxic T-lymphocyte (CTL) responses. These cells may lie dormant until activation at some point. Virus particles also evade an efficient immune response by association with the follicular dendritic cell network (Pantaleo et al 1994). This may be an important mechanism for the maintenance of HIV infection over time by acting as a source of virus for infection of cells that reside in or pass through the lymphoid tissue.

Change in viral genotype is important for viral escape from the immune response. If changes occur in those epitopes that are recognised by CTLs, or by antibody, then HIV specific CTL or antibody would not recognise these virus mutants and the virus would thus evade the immune response elicited by these components (Philips et al 1991). Due to the high error rate during reverse transcription by the viral reverse transcriptase, it is estimated that viral genomes containing one of every possible single point mutation are generated daily in the infected individual (Coffin et al 1995). This high rate of mutant generation has enormous implications for the understanding of the evolution of viral drug resistance and viral escape from the immune effector mechanisms.

During the advanced stages of disease when a patient is in AIDS, the HIV specific cytotoxic effects are not found (Haynes et al 1996). Neutralising antibody is not detected and there are reduced amounts of antibody to HIV proteins. At this stage of the disease there are increased levels of expression of some cytokines. This must be due to the fact that the majority of the circulating mononuclear cells are CD8+ cells, monocytes or macrophages, after the depletion of the majority of CD4+ cells. It is thought that the primary mechanism responsible for the immunosuppression that is associated with HIV infection is due to the destruction of the lymphoid tissue.
1.4 The Molecular Biology of HIV-1

1.4.1 Physical structure

HIV-1 buds from the surface of an infected cell as an immature particle composed of a shell of p55 gag (and gag-pol) protein associated with two copies of genomic RNA contained within a host derived lipid membrane envelope. After particle release there is an activation of the viral protease which is packaged into the virus as a fusion to gag protein, produced by a -1 frameshift during translation of gag. Maturation proceeds with the condensation of a cone shaped core made up of Capsid protein which contains the nucleocapsid, a complex of the RNA genome together with nucleocapsid protein. Figure 1.2 shows the steps in budding and maturation of HIV-1.

Figure 1.2

- **I** shows p55 gag targeted to the inner surface of the plasma membrane at regions of gp160 insertion, this creates a bulge on the surface of the cell. At some point during this two copies of genomic RNA are recruited for packaging.
- **II**. The p6 region in gag probably assists in the closing off of the core shell.
- **III** The immature virus buds from the cell.
- **IV**. At the same time or shortly after release the viral protease is activated and brings about the maturation of the particle by the cleavage of the p55 gag protein. In the immature virus the genomic RNA may be associated with membrane bound gag rather than free in the viroplasm.

The mature HIV-1 particle is about 110nm in diameter. The outer most layer, known as the envelope, is made up of host cell derived plasma membrane which contains virus glycoprotein gp120, linked to the membrane via a transmembrane gp41 protein. The membrane glycoprotein gp160 is important for cellular attachment. Below the outer membrane is a layer of Matrix protein which comprised the N-terminal
sequence of p55 in the immature particle. This protein is associated with the envelope by virtue of a membrane targeting signal and the addition of a myristic acid residue at the extreme N terminus of the protein (Gottlinger et al 1989). Figure 1.3 shows the structure of mature HIV-1.

**Figure 1.3**

![Diagram of mature HIV-1 virion](image)

**Figure 1.3** Basic structure of the mature HIV-1 virion, showing the essential virion architecture and the location of the major components.

Within the virus particle is a cone shaped Capsid structure which is comprised of p24 Capsid protein. Contained within the Capsid structure are two copies of positive sense genomic RNA which reside physically linked as a dimer, complexed with p7 nucleocapsid protein. In the case of HIV the genomic RNAs are around 9.5kb in size. Each RNA molecule has annealed to it a tRNA$_{byc}$ which acts as a primer for reverse transcription of the RNA in the newly infected cell. Also associated with the nucleocapsid are the enzymatic properties that are associated with the virus. These include the reverse transcriptase (p66/51) which is essential for the formation of proviral DNA on infection. This proviral DNA is inserted into the host cell chromosome by another enzymatic activity contained within the virus particle known as integrase or p32. The protease is thought to reside in the space between the Capsid
and the Matrix layer. The proteins p11, p66/51 and p32 are all incorporated into the virion during the assembly of the HIV gag proteins at the plasma membrane of the infected cell. About 1 in 20 translations of gag mRNA result in a -1 frameshift to produce a gag-protease-polymerase-integrase fusion protein that becomes incorporated into an assembling virus during gag assembly, thereby ensuring that the virus has its full complement of enzymatic proteins. The virus also contains some cellular proteins e.g. Cyclophilin, and some cellular derived tRNA.

1.4.2 The Structure of the HIV-1 genome

The standard retrovirus proviral genomic structure comprises three genes arranged between two long terminal repeat elements. The genes are, gag, encoding the virus structural proteins, pol, encoding the enzymatic activities of the virus and env which encodes the envelope glycoproteins. HIV-1 is an example of a complex retrovirus, which contains in addition to these genes, coding sequence for two regulatory proteins tat and rev, and the auxiliary proteins nef, vif, vpr and vpu. These smaller proteins are made from singly and multiply spliced messages and serve to modulate gene expression and virus growth. Figure 1.4 shows the genetic organisation of the HIV-1 genome.

Figure 1.4

Figure 1.4 The genetic organisation of HIV-1. Depicted are the Long Terminal Repeat elements (LTR), the gag, pol and env genes and the spliced gene products. Also shown is the position of the putative packaging signal ψ. Pol gene products PR - protease, RT: Reverse Transcriptase, IN-Integrase. Env gene, SU - Surface (exposed) region and TM, Transmembrane components of gp160.
**Gag gene**
As the gag protein is important to the work described in this thesis, this protein is discussed in detail with respect to packaging and morphogenesis in section 1.6.

**Pol gene**
The pol gene is expressed via a -1 ribosomal frameshift during the translation of gag protein. The amounts of pol are thus regulated by this frequency of frameshifting and are kept at a constant ratio to the amount of gag that is translated. The 5’ coding region of the pol gene, which becomes directly fused to the gag protein during translation is the viral protease or p11. Retroviral proteases are examples of aspartic proteases. These enzymes are characterised by an active site containing two aspartate residues, each of which are located in a separate polypeptide domain. The active site is thus between the two domains, thus to be functional in the immature virus the protein has to dimerise (Kaplan *et al* 1994). This activity is concentration dependent whereby, it is only activated in the assembled particle when gag is at a high localised concentration.

Fused to the C-terminus of protease is the pol protein. This protein consists of two enzymatic activities, reverse transcriptase (p66) and an RNase H (p51) subunits. The functional reverse transcriptase is a heterodimer of these proteins (leGrise *et al* 1991, Hostomsky *et al* 1992).

The C terminal end of the gag-pol fusion protein contains the integrase protein. This is a 32kDa protein which forms part of the preintegration complex that is responsible for the transfer of freshly reverse transcribed proviral DNA to the nucleus where the protein then catalyses the integration of the protein randomly within the host cell chromosome.

**Env gene**
The env gene encodes the envelope associated viral glycoproteins. The protein is translated as a polyprotein precursor on polyribosomes associated with the rough endoplasmic reticulum. The protein oligomerises in the rough endoplasmic reticulum (Earl *et al* 1990). The assembly into oligomers, either trimers or tetraders, is
necessary for their subsequent transport into the Golgi apparatus. It is in the Golgi that the protein is cleaved into membrane associated gp41, and surface gp120 proteins. The Golgi then transports the protein to the cell surface to form the site of virus assembly. gp160 is the virus receptor that interacts with the surface of CD4+ cell, it is an important antigenic determinant in the infected person to which a powerful immune response is elicited.

**Auxiliary gene products**
Following infection and integration of HIV-1 in a permissive cell, transcription from the viral LTR results in the production of smaller multiply spliced protein products tat, rev, nef, vif, vpr and vpu.

**Tat**
Initially in the infected cell small amounts tat protein are produced from multiply spliced message. The amount of full length transcripts are low due to the low processivity of the transcription complex (Kao *et al* 1987). The tat protein is able to cause the transactivation of transcription by binding to a bulge in a stem loop commonly known as the transactivation response hairpin (Dingwall *et al* 1990). The binding of this protein to this region increases the processivity of the complex which is transcribing from the 5' LTR (Kao *et al* 1987, Laspia *et al* 1990). The effect of tat is to cause a 100 fold increase in the amount of processive transcription complexes originating from the LTR. This increases the amount of tat and other small proteins that are translated from spliced message, and ultimately all viral proteins

**Rev**
Rev is also translated from a spliced message during the early infection. The concentration of rev gradually builds up and carries out its function later on in the infectious cycle. Rev protein multimerises on a purine rich bubble contained in a sequence of extensive secondary structure known as the rev responsive element, (RRE), which is located in the genomic RNA (Heaphy *et al* 1990). The action of rev is to promote the nuclear to cytoplasmic transport of unspliced and singly spliced RNA, this also has the effect of decreasing the amount of tat and rev in the cell (Cullen 1992), thus facilitating late gene expression of gag, pol and env.
**Vif**

The vif protein is expressed from a singly spliced mRNA whose expression is rev dependent, and hence appears in high levels late in the infection. A substantial amount of vif colocalises with gag at the membrane of the infected cell, and a small number of vif molecules are packaged into the released virus. The only known biological function of vif is to increase the infectivity of virions up to 100 fold, when the virus is produced from non-permissive cells (Gabuzda *et al* 1992). Vif negative virus show some defects in their ability to fully reverse transcribe and integrate a full proviral DNA (Simon and Malim 1996), thus vif may have important roles in the early stages of the infectious process.

**Nef**

Nef is present in the early stages of infection at far higher levels than tat/rev (typically 80% of early spliced RNA- Cullen *et al* 1994) and probably has many distinct roles in the infected cell. Nef is able to downregulate the cell surface expression of CD4 (Garcia and Miller 1991). This presumably reduces interactions between CD4 and newly translated env protein, and also facilitates the release of mature virions. Nef is also able to downregulate the cell surface expression of MHC class 1 receptors, thus having the effect of inhibiting CTL mediated lysis of infected cells (Collins *et al* 1998). There is some evidence that nef is able to exert effects on cellular signal transduction and activation, although its exact role in this context is unclear.

**Vpr**

Vpr is another late gene product which is packaged into the virion. The protein is found here in similar molar amounts to gag and probably interacts with the p6 domain of p55 for particle targeting (Paxton *et al* 1993). Vpr is thought to have a role in the nuclear import of HIV-1 preintegration complexes, of particular importance in non-dividing cells (Heinzinger *et al* 1994, Vodicka *et al* 1998).

**Vpu**

Vpu is a protein unique to HIV and is not seen in any of the other lentiviruses. In the infected cell this protein is able to enhance the virion release and cause the selective degradation of CD4 in the cell endoplasmic reticulum. Enhanced virion release
mediated by vpu apparently occurs by the promotion of budding from the plasma membrane rather than intracellular membrane structures, and by facilitating the release of budding virions from membranes. The degradation of CD4 is brought about by the binding of vpu to CD4, and a subsequent recruitment of proteins which cause the selective proteosome degradation of CD4 and the release of vpu (Margottin et al 1998).

1.4.3 The Infectious Cycle

Figure 1.5 shows the steps that are typical to an HIV, or any retroviral infection.

**Figure 1.5**

A diagram illustrating the cross section of an infected cell, showing each stage in the infection process, from entry to progeny virus release.
Attachment and Entry

HIV-1 binds to cells expressing the CD4 molecule on their cell surface via an interaction with the virus gp120 protein (Maddon et al. 1986, McDougal et al. 1986). This was the first retroviral cellular receptor described. It soon became apparent that although CD4 was the primary receptor, this molecule was not sufficient to permit infection of some other cell types (Chesebro et al. 1990). Co-receptors for HIV, essential in addition to CD4, have recently been identified as a range of chemokine receptors (Feng et al. 1996). It is now thought, that after the initial CD4-gp120 interaction there is an exposure of the V3 loop in the gp120 subunit. It is this region that interacts with an appropriate chemokine co-receptor extracellular domain. This is thought to induce conformational changes in env which allows the fusion peptide in gp41 to insert into the cell membrane (reviewed in Doms and Peipert 1997) and allow cellular entry of the virus core.

Reverse Transcription

An essential step in the virus lifecycle which occurs after the virus core has entered the cell cytoplasm is the conversion of the retrovirus RNA genome into the proviral DNA by the virally encoded RNA dependent DNA polymerase. The primer for reverse transcription in HIV-1 is a tRNA<sub>lys</sub> (Mak and Kleiman 1997). This is carried into the infected cell bound to a target sequence in the viral RNA known as the primer binding site, a region about 200 nucleotides from the 5' end of the genomic RNA. There is evidence that some reverse transcription may occur inside the virus particle (Huang et al. 1997), the remainder of the reverse transcription, a complex process involving two strand transfers in order to construct two functional LTRs must occur in the cytoplasm after uncoating. For a detailed review of reverse transcription refer to Gotte et al. (1999).

DNA Integration

After reverse transcription the proviral DNA is transferred to the nucleus of the infected cell by a preintegration complex, a large nucleoprotein complex, most likely made up proviral DNA, integrase, Matrix and vpr (Heinzinger et al. 1994). HIV is able to infect non-dividing cells due to this nuclear transfer mechanism, which is not dependent on the breakdown of nuclear membrane at cell division as in other
retroviruses (Bukrinsky et al 1993). Integration can occur at many sites in the host genome, it is thought that preferential sites of integration are those at kinked or distorted regions such as that found at nucleosomes (Miller et al 1995). The process of integration is carried out by the viral integrase protein which catalyses three distinct reactions. Firstly overhanging $\text{CA}_{\text{OH}}$ ends are produced in the linear viral DNA. In the second step these 3' ends are joined to the 5' ends of the target DNA, and in the third step which is probably assisted by cellular proteins, unpaired nucleotides at the 5' ends of the viral genome are removed and joined to the target site 3' ends. This generates the integrated provirus which is flanked by five base pair repeats of the target site DNA.

**Transcription**

Once integrated into the host chromosome, the HIV provirus initially acts as an RNA polymerase II responsive gene. The functional promoter region is contained in the 5' LTR. The elements indicative of an RNA polymerase II gene are the presence of a TATA box located at $-22$ to $-27$ from the +1 start residue (455 in HIV$_{NL43}$ proviral DNA). The promoter also contains distal enhancer elements with a complex array of binding sites for cellular transcription factors, such as NFkB, and SP1 (Nabel and Baltimore 1987, Coffin 1991). Under normal cellular conditions this promoter acts to direct the transcription of a low basal level of RNA transcripts, which result in the production of a low amount of viral RNA. In the early infection this RNA is multiply spliced and leads to the production of the early gene products tat, rev and nef. The production of tat, as described earlier, has the effect of increasing the apparent efficiency of transcription by interacting with the tar hairpin in nascent transcripts and altering the processivity of the transcription complex. This results in an increase in the amount of early gene products. As the amount of rev protein rises, so the cytoplasmic accumulation of full length RNA transcripts occurs and hence the production of the structural and envelope proteins, as described earlier.

**RNA Splicing**

HIV transcription essentially occurs from a single promoter and results in the production of a full length genomic RNA. To express all of the genes in this RNA the virus makes use of the cells splicing machinery to regulate the temporal expression of
the viral genes. The RNA sequence contains a 5' splice donor site within the leader sequence and further splice donor and acceptor sites within the genome such that all RNAs whether spliced or unspliced contain the same 5' and 3' sequences. RNAs can thus be unspliced, such as gag and pol message, which also acts as genomic RNA targeted for an assembling virus. Singly spliced message, such as Env message and multiply spliced such as tat or rev. Examples of HIV-1 mRNA are shown below in figure 1.6.

**Figure 1.6**

![Diagram showing examples of the different mRNAs that are made in the infected cell.](image)

The unique sequence 3' of the 5' splice donor to the start of gag (ψ) is only contained in the genomic RNA. This is important for the work discussed in this thesis as it is likely that this sequence contains the putative RNA packaging signals that are responsible for the targeting of a genomic RNA into an assembling virus.

**Translation**

All HIV-1 RNAs are capped and translation proceeds via a cap-dependent translation mechanism. Translation actually exerts a further regulatory mechanism to ensure that all the viral genes are translated in the right amounts. The translation of gag from the full length genomic RNA can result in either the translation of p55 gag, or the production of a gag-pol fusion protein consisting of gag up to and including the p7 domain fused to the pol open reading frame. This results in a protein of 160kDa. The production of this gag-protease-polymerase-integrase fusion results from a -1
ribosomal frameshift as the ribosome encounters a U/A slippery sequence during gag translation (Wilson et al 1988). Ribosomal frameshifting ensures that on average 1 in 20 gag translations that initiate will result in the production of gag-pol proteins. This ensures that these proteins are synthesised in the correct amounts and allows a mechanism of targeting into assembling virus that is dependent on the self assembly of the gag domain of the fusion protein.

**Virus Assembly**

The p55 gag protein is synthesised on free ribosomes within the cytoplasm of the infected cell (Hunter 1994). Common to many retroviral gag proteins this is then posttranslationally modified by covalent addition of a myristic acid residue to the N-terminal glycine residue (Gottlinger et al 1989), which together with an N-terminal membrane targeting sequence (see section 1.6) directs the protein to the inner surface of the plasma membrane. Electron micrographs show that prior to particle budding gag forms crescent shaped complexes just below the plasma membrane (Gelderblom 1991). Somehow, (discussed later) two full length genomic RNAs are selected from the homogeneous mix of cellular RNAs and the immature particle buds from the surface of the cell acquiring the virus glycoproteins with the plasma membrane derived envelope on release.

**Virus Maturation**

An immature virus is released from the surface of the infected cell. Concomitant with release or shortly afterwards the viral protease autocatalytically cleaves itself from the gag-pol fusion protein. Cleavage of the viral polyprotein precursor, p55 gag, causes a rearrangement of the virion architecture to that of a mature virus. Maturation of the virion RNA also occurs to that of a more heat stable RNA dimer than that contained in the immature virus, (Fu and Rein 1993).

**1.4.4 Structure of the Genomic RNA**

Retroviral genomes vary in length from between 9-10kb. The genome consists of a dimer of RNA held together by non-covalent interactions (discussed later). The RNA genome consists of the 3 main gene blocks gag, pol and env bordered 5' and 3' by untranslated regions as shown in figure 1.7.
Like any mRNA transcribed in a eukaryotic cell, the HIV RNA genome is capped at its 5' end by a 7-methyl guanosine structure, and at the 3' end of the RNA there is a poly adenylated tail. The genome is terminally redundant with a repeated sequence known as R at both ends of the RNA. This R sequence is important during the reverse transcription process, as it facilitates the strand transfer mechanism following the reverse transcription of the U5 region.

The 5' UTR in all retroviruses is extremely rich in RNA secondary structure, and this is highly conserved between strains of the same virus (Berkhout 1996). Several publications have predicted the RNA secondary structure in this region, based on a variety of techniques and all agree on the type of structures that are depicted in figure 1.8 (Baudin et al 1993, Sakaguchi et al 1993, Clever et al 1995).
Figure 1.8

Diagram showing those structures that are expected to fold in the 5' leader region of HIV-1NL43. The structures are labelled with their putative functions. The numbers labelled on the RNA correspond to the 5' and 3' boundaries of RNAs constructed in chapter 4, and serve for reference purposes. SL2 carries the splice donor site between G residues 289+290. SL3 has previously implicated in packaging and is also known as \( \psi \). SL4 carries the Gag AUG start at 336-338. The numbers under each individual domain of secondary structure represent the Mulfold energy data for each structure in Kcal/mol at 37°C in 1M NaCl.

This region holds many important regulatory and functional signals which govern essential phases of the virus lifecycle. It is likely that this structure may be a gross over simplification as it is likely that there are structural or functional interactions between the different RNA domains, and that the actual RNA is structurally dynamic. However to date tertiary interactions within this region have not been studied. The most 5' region of secondary structure is the tar hairpin. This is responsible for the binding of the viral transactivator, tat, at a bulge domain in the hairpin as this sequence forms part of the nascent transcript (Dingwall et al 1989). Due to the terminal redundancy of the retroviral genome both tar and the next region of secondary structure, the poly A hairpin, are reiterated at the 3' end of the genome. It is thought that the role of tar as the transactivation response signal is only active in the context of the 5' end of the genomic RNA. The next stable region of secondary structure is known as the poly A hairpin. This structure in the context of the 3' end of the virion RNA directs the polyadenylation of the transcribed RNA (Keller 1995). In the context
of the 5' end of the viral RNA this signal is not active. This signal may be silenced in the context of the leader RNA by its close proximity to the mRNA start site or cap structure (Cherrington and Ganem 1992). The next region of distinct secondary structure is known as the primer binding site. A tRNA$_{pr}$ binds to the loop region in this sequence and acts as the primer for the viral reverse transcriptase (Jiang et al. 1993). 3' of the primer binding site are four smaller stem loop structures. SL1 or the dimer initiation site is thought to be important for the dimerisation of the retroviral RNA genome, this is fully discussed in the next section. SL2 contains the major first splice donor at the $^{267}_{\text{CUG}} \downarrow GUG_{292}$ sequence. This site is used to make most of the subgenomic spliced RNAs. The next stem loop structure, known as SL3 or $\psi$ has previously been implicated as a signal in the RNA that is recognised by the gag protein, important in RNA packaging. This is discussed in the context of packaging in section 1.7. The last stem loop, SL4 extends into the gag coding region and actually contains the start codon for gag translation. The positioning of the start codon in a region of RNA secondary structure may be thought of as having a negative effect on translation. The significance of the positioning of AUG in a region of secondary structure needs further investigation. The sequence 3' of the gag start codon does not fold to form extensive secondary structure.

1.5 RNA Dimerisation

1.5.1 Retroviral RNA Dimers

All retroviruses carry two copies of genomic RNA in a dimeric structure. These two RNAs resemble full length unspliced mRNA and are both capped and polyadenylated. The dimeric nature of the genome was originally observed during sedimentation analysis of the viral genome. The genome of some RNA tumour viruses such as RSV were shown to exist in a complex that sedimented at around 60-70S, together with some smaller RNA molecules of around 5S (Robinson et al. 1965). If these complexes were subject to denaturation by heating then the larger complex was shown to dissociate giving RNA species which would sediment at around 36S. In addition to the 36S species there was a mix of heterogeneous material that would sediment between 7-30S (Duesberg 1968). Also noted in these early studies was the fact that the structure of the genomic RNA in mature and freshly harvested viruses differed with respect to thermal stability, (Stolfus and Snyder 1975, Cheung et al. 1972, Brahic
and Vigne 1975). Thus viral RNA could be seen to undergo a maturation step as the virus particle got older. Estimations of the molecular weights of the genomic RNA were made (Kung et al 1975) and predicted that the higher molecular weight complexes must be made up of two monomeric genomic RNA species.

The dimeric nature of the retroviral RNA genome was confirmed by numerous electron microscopy studies using a diverse range of viruses species. Figure 1.9 shows an electron micrograph from Murti et al (1981), showing the dimer point of contact in the genome of RSV.

Figure 1.9

Electron microscopy of the RNA genomes of, endogenous cat virus RD-114 (Bender et al 1976), Friend murine virus (Dube et al 1976), MuLV (Hu et al 1977), AKR, NZB, and wild mouse viruses (Bender et al 1978), MoLV and RSV (Murti et al 1981) all showed the physical link between the retroviral genomic RNA. Electron microscopy further showed that there was a stable point of contact between the two RNAs at the 5' end of the genome, which was present even under mild denaturing conditions (Murti et al 1981). Under less stringent conditions the genome appears to have multiple points of contact (Mangel et al 1974). The presence of a stable point of contact under partially denaturing conditions became referred to as the Dimer linkage site (DLS) (Murti et al 1981). This stable point of contact was shown to be located in the 5' leader region, 3' of the primer binding site in each RNA and could not be re-formed after heat denaturation. Recently the genome of HIV-1 was also analysed by electron microscopy (Hoglund et al 1997). This work showed more than one stable point of contact in the virion RNA and has further enhanced the understanding of RNA dimerisation in this virus. A further technique for the analysis of monomeric and dimeric RNA was described by Khandjian and Meric (1986), this was based on
non-denaturing northern blotting of the virion RNA. This technique allows the visualization of genomic dimer and monomer positions on an autoradiograph blot. Once again this showed the dimeric nature of RSV RNA. This method has become widely adopted for analysing the oligomeric structure of retroviral genomes, in studies as to what factors promote and contribute to dimerisation in-vivo.

1.5.2 The Advantages of a Dimeric Genome
The conservation of a dimeric genome amongst diverse retroviruses suggests that the possession of two genomic size RNAs is essential for the virus lifecycle.

Reverse transcription
During reverse transcription of the retroviral RNA to make the proviral DNA sequence, two functional LTRs need to be generated to produce a functional provirus. The production of two functional LTRs from reverse transcription of the viral RNA is accomplished by a strand transfer from the 5' end of the initiating RNA to the 3' end of either the same or the other RNA molecule. This can be either an inter or intra molecular strand transfer (Panganiban and Fiore 1988, Hu and Temin 1990). These different strand jumps may occur at the same frequencies (Van-Wamel and Berkhout 1998). It is possible then, that two molecules are needed in some instances for the production of proviral DNA containing two functional LTRs. This is not an absolute requirement as the synthesis of a viral DNA can occur from a single viral RNA.

Viral Recombination and Template Switching
The two retroviral RNA strands are commonly nicked by nucleases that are generally regarded as being associated with the virus particle. A dimeric genome will function to keep the RNA sequence together in the presence of nicking. It is thought that the dimeric diploid genome may provide the opportunity for the reverse transcriptase to switch template strands from a damaged molecule to an intact template. Frequent template switching during reverse transcription can result in high levels of recombination. This is probably an asset to a virus such as HIV where genetic variability is closely linked to pathogenicity and continuation of disease (Hu and Temin 1990). The virus may generate recombinants this way as proposed by the forced copy choice model (Coffin et al 1980). Interstrand transfer has actually been
show to occur during the initial stages of reverse transcription in the small amount of sequence 5' of the primer binding site (Panganiban and Fiore 1988), and probably indicates that there are a large number of strand transfers during the reverse transcription of an entire genome.

Translation
It is still not clear where dimerisation of genomic RNA actually takes place in the process of virion morphogenesis. It is likely that an immature dimer forms in the infected cell cytoplasm and this is then recognised by the viral genomic RNA packaging mechanisms. It is possible that dimerisation of genomic RNA may be able to regulate translation to a certain extent. In the case of RSV monomeric RNA transcripts and extracted virion RNA direct the translation of large amounts of gag protein whilst the dimeric or native RNA does not when assayed in in-vitro translation systems (Bieth et al 1990, Darlix 1986). This must presumably be due to the inability of the translation machinery to either initiate a translation when an RNA is part of a dimer or the translation machinery may be stalled by the presence of secondary or tertiary structure in an RNA. If this is the case genomic RNA dimerisation may negatively regulate translation.

Mechanism of RNA Dimerisation
In early electron microscopy studies of retroviral RNA and analysis by sedimentation techniques, it was apparent that whatever was holding the RNA dimer together was not sensitive to a wide range of chemical treatments. These included mild denaturants such as urea and formamide, detergents such as SDS, proteases such as proteinase K, and solvents such as phenol and chloroform (e.g. Bader and Steck 1969, Dube et al 1976). This property of the viral RNA suggested that this was due to the RNA molecules themselves and not the presence of protein factors that may have held the RNA dimers together. Furthermore Bader and Steck (1969), suggested that hydrogen bonds must be involved in holding the two molecules together at many sites along the genome of MuLV. The observation that heat denatured RNA consists of a heterogeneous mix of fragment sizes suggested that an extensively nicked RNA could be held as a dimer structure by multiple points of Watson-Crick base pairing along the genome of the virus. The nature of the interaction was thought to be different from
the expected conventional Watson-Crick basepairing, as the two genomic RNAs were seen to lie in a parallel arrangement i.e. 5' opposite 5' ends when the RNA was visualised by electron microscopy (Kung et al 1975).

1.54 In-vitro Analysis of RNA Dimerisation

Although dimer linkage had been shown to occur in genomic RNA as visualised by EM, it was not until around ten years ago that groups began to look at the localisation and mechanism of RNA dimerisation in in-vitro derived model systems. Focusing on the 5' regions located by electron microscopy, to be involved in dimerisation in the genomic RNA, Bieth et al (1990) showed that short transcripts from RSV could spontaneously dimerise in-vitro. By deletion analysis this group showed RNAs containing some of the first 600 nucleotides of the RNA would dimerise. This was followed by studies in MLV which showed in-vitro dimer formation within the first 1600 nucleotides of MoMV in a region previously implicated in genomic RNA packaging. Interestingly when these RNAs were observed by EM they appeared just like those isolated from mature virion (Prats et al 1990). Darlix et al (1990) were able to show that a 100bp RNA sequence in HIV-1, from a similar region to those in the other viruses, was also able to dimerise spontaneously in-vitro. These three studies also recorded an enhanced ability for RNA dimerisation in the presence of nucleocapsid p7 protein. The in-vitro dimerisation ability of RNA derived from the 5' region of numerous other retroviruses has now been documented, from REV-A (Darlix et al 1992), HaSV (Feng et al 1995), BLV (Katoh et al 1993), HIV-2 (Berkhout et al 1993), Rat retrotransposon VL30 (Torrent et al 1994), ASLV (Fosse et al 1996) and HFV (Erlwein et al 1997). Again it is HIV-1 which has received the most attention in terms of genomic RNA dimerisation and has become the paradigm for retroviral RNA dimerisation study in recent years.

Relevance of In-vitro RNA Models

It is thought that in-vitro dimerisation models are good representatives of what may occur in-vivo since the thermal stabilities of some in-vitro model dimers and those extracted from virion are approximately the same e.g. in Mo-MuLV (Roy et al 1990). The structure of RNA from this region, based on chemical probing of virion extracted and in-vitro transcribed sequences seems to be identical (Tounekti et al 1991).
Further regions of RNA, such as those previously implicated in packaging (Hayashi et al 1992) show the same structures both in synthetic RNA fragments (Baudin et al 1993) and RNA isolated from infected cells (Hayashi et al 1992). These analyses go some way to showing that in-vitro studies are probably relevant to what actually goes on in the virion particle, and that in-vitro studies can offer good biochemical evidence as to mechanisms that may be occurring in-vivo.

Identification of Mechanisms Involved in RNA Dimerisation
Dimerisation of in-vitro transcribed RNA is concentration dependent and may account for the inability of extracted genomic RNA to re-dimerise after heat denaturation. RNA dimerisation was also shown to be dependent on the size of the monovalent cation available in solution, temperature and ionic strength (Marquet et al 1991). In addition the antisense of these RNAs also did not dimerise. The presence of a conserved consensus sequence in a diverse range of retroviruses led this group to propose that RNA dimers were probably not forming by Watson-Crick basepairing mechanisms, and may involved purine quartets. These structures are proposed to arise where regions of four purine molecules, two from each RNA monomer, act to co-ordinate one monovalent cation. Polypurine tracts sharing the 6 nucleotide consensus PuGGAPuA lie in the regions that had been proposed to be important for RNA dimerisation in, M-MuLV (Prats et al 1990), HIV-1 (Darlix et al 1990) and REV (Darlix et al 1992). Dimers composed of guanine quartets derived from HIV-1 sequence showed a higher thermal stability in the presence of potassium than other monovalent cations, and that under these condition a 98mer RNA dimer spanning the HIV-1 gag start codon was dependent on guanine quartets, as the deletion of one of the of the polypurine tracts ablated dimerisation (Sundquist and Heaphy 1993). This was also shown using a similar HIV-1 RNA (Awang and Sen 1993). The contribution to dimerisation of retroviral RNA by guanine quartets has been studied in-vivo. Mutations in the polypurine tract that disrupt dimerisation by this mechanism in-vitro have no effect on the stability of the dimeric structure of RNA isolated from virions (Haddrick et al 1996), although this may be a contributory interaction within the mature dimer.

Further analysis of the sequences in the dimer linkage region first described by Darlix et al (1990), showed that additional sequences 5' of the DLS dimerise in-vitro. These
RNAs dimerise in the presence of lower strength ionic buffers (Paillart et al. 1994, Marquet et al. 1994, Skripkin et al. 1994). The location of this additional dimerisation site was referred to as the Dimer Initiation Site or DIS. The discovery of this additional dimerisation domain allowed nucleotides 1-311 to dimerise in-vitro. This dimer dissociates during electrophoresis if the buffer does not contain magnesium (Marquet et al. 1994). The properties of the two forms of dimer that could be induced in-vitro were therefore different. The 5’DIS dimerised rapidly, was dependent on divalent cations and had a melting temperature of 47°C (HIV\textsubscript{MAL}), whereas the 3’ DLS dimerised slowly, showed greater thermal stability and dimerised in a cation dependent manner, presumably by forming a guanine quartet structure.

The Kissing Loop Interaction

Computer modelling and chemical/nuclease accessibility mapping of this region of RNA showed the leader region is rich in secondary structure (see figure 1.8 and Baudin et al. 1993, Clever et al. 1996, Berkhout 1996). Chemical modification interference analysis of an HIV-1 RNA spanning the RNA +1 to 707, showed that a palindromic sequence (in HIV\textsubscript{MAL} 274GUGCAC279) upstream of the major slice donor, and downstream of the primer binding site was essential for the dimerisation of RNA upstream of the 3’DLS (Skripkin et al. 1994). This palindromic sequence was located in the loop of a short hairpin structure. A mechanism proposed for the dimerisation of the RNA at this structure was described as a kissing loop. This describes a model whereby the palindromic nature of the sequences in each of the stem loop monomers interact via a stem loop interaction, see figure 1.10.

Figure 1.10

![Figure 1.10](image-url)  
Figure 1.10 The proposed mechanism for HIV RNA dimerisation via a kissing loop interaction at stem loop 1. (HIV\textsubscript{NL43} sequence).
Disruption of the autocomplementary loop sequences (ACS) resulted in the loss of in-vitro dimerisation. The proposed kissing loop mechanism (Laughrea and Jette 1994) suggested that the 6 palindromic nucleotides from one RNA bind to the same 6 nucleotides on the other RNA in an antiparallel manner (see figure 1.10). It is then thought that the stem of each stem loop dissociates and re-anneals with the complementary sequence from the other stem, forming an extended complex with a change in interstrand base pairing from 6bp to 28bp. This model has been supported by carrying out compensatory mutational studies, within the loop region and making sequence mutations in the stem and loop regions that abolish dimerisation in-vitro (e.g. Paillart et al 1994, Haddrick et al 1996, Clever et al 1996).

The dimerisation of RNA by this mechanism in-vitro is specific and is not the result of simple annealing of the two denatured molecules. The essential component is the loop-loop kissing interaction, and under conditions where the kissing loop complex forms in-vitro, hybridisation of the loops does not propagate down the DIS stems to form extended dimeric complexes (Paillart et al 1996). Support for the kissing loop dimer and the high stability extended dimer of HIV-1 RNA have been described. These studies showed that the less stable dimer forms under physiological temperatures, while the more stable conformation would form with the presence of nucleocapsid protein (Fu et al 1994, Feng et al 1996, Muriaux et al 1996) and after incubation at higher temperatures (Laughrea and Jette 1996, Muriaux et al 1996). It is also likely that the kissing loop dimer collapses and forms the expected extended dimer during phenol extraction (Haddrick et al 1996), as a dimer can be formed that is stable to electrophoresis in the absence of magnesium.

The nature of the kissing loop interaction has been extensively studied and refined over the last few years. Sequence comparison of more than 30 HIV-1 and SIV isolates reveals that only a fraction of the possible palindromic sequences at the DIS are represented (Berkhout 1996). The 3 nucleotides that flank the palindromic sequence are highly conserved, most often adenine and always purines. The importance of this is suggested by the fact that an antisense DIS RNA which contains the same self complementary sequence in the loop, does not dimerise suggesting that the structure of the DIS stem loop is important for dimerisation (Clever et al 1996, Skripkin et al 1996). The flanking purine residues were thought to contribute to non-canonical
interactions in the stem loop which must contribute to dimer stability (Paillart et al 1998).

The solution structure of the kissing loop complex formed from small RNAs just resembling the SL1 stem loop have been solved by 2D NMR (Dardel et al 1998, Mujeeb et al 1998). These two groups propose that the SL1-SL1 interaction is atypical of kissing loop hairpin models, in that the loops are distorted by the interstrand stacking of the flanking adenosines. These may contribute to an inherent instability of the structure which may bring about the melting of the initial contact and lead to the more stable structure by extensive inter strand base pairing possibly via an interaction with p7 Nucleocapsid.

Effect of Nucleocapsid protein on RNA Dimerisation

RNA dimer stability in viral particles had been observed to increase over time after particle release in various retroviruses (Canaani et al 1973, Cheung et al 1972, Stolfus and Snyder 1975). This change in thermostability was observed to take place after particle maturation (Cheung et al 1972). A study in RSV implicated the nucleocapsid protein p12 in RNA dimerisation (Meric and Spahr 1986). This was further suggested after the observation that a protease deficient strain of RSV, which could not process the gag polyprotein precursor contained an immature genomic RNA (Oertle and Spahr 1990). A study in MoMuLV confirmed that this maturation event is dependent on the functional processing of the gag precursor and in KiSV this maturation event is abrogated by mutations in the nucleocapsid cysteine array sequence (Fu and Rein 1993). Lear et al (1995) showed that RNA initially packaged into virions, as analysed from rapid harvest extracts, appears to be monomeric even in the presence of cleaved gag. This suggests that rearrangement of the dimer into the stable structure mediated by nucleocapsid takes time, and may occur after the morphological changes associated with proteolysis.

Initial studies on RNA dimerisation in-vitro had shown that the dimerisation of RNA oligomers was enhanced by the presence of nucleocapsid protein. Two studies (Feng et al 1996, Muriaux et al 1996) showed that the addition of HIV-1 p7 nucleocapsid protein to kissing loop dimers of HIV-1 and HaSV RNA promoted the conversion of the transient kissing loop dimer into a more stable form consistent with an RNA of the
extended structure. p7 Nucleocapsid is reported to possess nucleic acid unwinding properties (Khan and Giedroc 1992) and RNA renaturation capabilities (Dib-Hajj et al 1993), thus both these activities must probably work together to open up both stems of the SL1-SL1 complex and then permit the formation of the double stranded extended dimer (Muriaux et al 1996).

The Role of Kissing Loop Dimerisation in-vivo

Mutations have been introduced into the DIS of infectious molecular clones of HIV and these show profound effects on the virus growth efficiency and 'fitness'. Mutations in the kissing loop palindrome show delayed virus growth kinetics (Haddrick et al 1996, Harrison et al 1998) and a reduction in the amount of RNA dimer (Haddrick et al 1996, Laughrea et al 1997). In other studies the genomic dimer was shown to have similar thermal stability to wild type RNA (Berkhout and Van-Wamel 1996, Clever and Parslow 1997, Sakuragi and Panganiban 1997). These two observations would tend to suggest that dimerisation by the ACS sequence in SL1 is important for dimer formation but it is not an absolute requirement as dimeric RNA can still form. This therefore suggests that SL1 may be important either in genomic RNA dimer initiation, or in dimer stability. Further features of palindrome disruption have shown between a 10-1000x decrease in the infectivity of the virus, (Paillart et al 1996, Berkhout and Van-Wamel 1996).

The low infectivity of these viruses may be due to a defect in the second strand transfer during reverse transcription with these viruses (Paillart et al 1996). This group proposes that the DIS increases the proximity of the redundant nucleic acid ends which may favour the strand transfer. Berkhout et al (1998) further analysed the role of dimerisation in reverse transcription and have shown that a dimeric RNA is actually needed for the first strand transfer as shown in-vitro using denatured and native purified viral RNA.

Mutations in the kissing loop region have also been shown to result in a reduction of RNA packaging by viruses (Clever and Parslow 1997, Paillart et al 1996, Laughrea et al 1997). It is possible that this motif is recognised as part of a packaging signal during the packaging of the genomic RNA into an assembling virus.
The necessity for an auto complementary sequence in SL1 has been demonstrated by passing autocomplementary sequence mutant virus for long term in tissue culture. Recovered outgrowing virus all contained palindromic sequences in SL1, so the ACS was selected for over time (Berkhout et al 1996). Furthermore using RSV with a randomised sequence in this region, after growth in tissue culture, analysis of outgrowing virus showed the presence of a functional ACS in the RSV equivalent of SL1 (Doria-Rose and Vogt 1998). These experiments show that an ACS is actually selected for in-vivo to allow efficient replication of the virion.

Detailed analysis of the gag polyprotein precursor processing in HIV-1 has revealed that the cleavage of the p2 peptide from p24 occurs inefficiently in viruses that have deletions in SL1. This suggests that altered growth kinetics associated with SL1-SL1 mediated RNA dimerisation may also be due, in part to aberrant processing of the viral structural protein (Liang et al 1999). This is likely to be linked in part to the increase in protease processing of gag proteins in the presence of RNA (Sheng and Erikson-Viitanen 1994, Sheng et al 1997). Furthermore deletions in SL1 which impair virus infectiveness can be rescued by long term culture of virus which results in the accumulation of point mutations in gag (Liang et al 1999). These mutations are consistent with a role for SL1 in gag protein assembly and processing, in addition to its roles in RNA dimerisation and putative roles in packaging.

1.6 Gag proteins and Their role in Virion Morphogenesis

1.6.1 Features of the structural protein

The gag polyprotein precursor of HIV, p55 is synthesised from full length, unspliced HIV RNA. This protein directs the assembly of the virus and is sufficient on its own to produce non-infectious virus like particles, in either baculovirus expression systems (e.g. Hughes et al 1993, Gheyson et al 1989), vaccinia expression systems (e.g. Hoshikawa et al 1991, Karacostas et al 1993), in-vitro from purified protein (Campbell and Vogt 1995), and even in crude reticulocyte lysates (Spearman and Ratner 1996). This shows that HIV capsids can assemble from p55 in a cell free reaction mix. The expression of gag-pol fusion protein does not lead to efficient virus-like particie (VLP) production (Karakostas et al 1993).
Essential for the formation of a virus particle in an infected cell, the p55 gag protein becomes N-terminally myristylated by the addition of a myristic acid residue to the N-terminal glycine residue (Gheyson et al 1989, Gottlinger et al 1989, Bryant and Ratner 1990, Freed et al 1994). Mutations which block this interaction block virus assembly and membrane association of the protein. HIV and other retroviruses are not lytic and generally cause long term infections, so assembly of the virus requires that components are transported to a site in the infected cell, where assembly and budding may occur. This targeting of p55 to the cell membrane or the site of assembly may be due to sequences within p17 Matrix (Rhee and Hunter 1990). It is possible that gag may actually translocate to the plasma membrane by interactions with cytoskeletal proteins such as β-actin. The successful targeting and association of p55 gag with the membranes is not only due to the myristic acid moiety as some retroviruses such as Visna and ALV are not myristylated on the p55 gag protein. In the case of HIV and a variety of SIV isolates there is a basic stretch of amino acids which are thought to make up a membrane targeting signal (Zhou et al 1994). This basic region together with the myristic acid residue is predicted to interact with anionic phosphatidyl serine and phosphatidyl inositol head groups, which concentrate on the inner side of the lipid bilayer. Three dimensional modelling of the p17 Matrix shows that the basic stretch of amino-acids forms part of a β-pleated sheet which protrudes from the protein exposing 4 of the basic residues to the environment (Massiah et al 1994, Matthews et al 1994). This projection is surrounded by further basic amino-acids.

The products of the gag gene are depicted in the following diagrams to illustrate those domains that are described in the text.
Figure 1.11 Location of the gag gene

Figure 1.11 The localisation of the gag gene in the genome of HIV-1. Also shown are the separate domains of p55.

Figure 1.12 p17 Matrix

Figure 1.12 Representation of functional domains in the HIV-1 p17 Matrix protein.
Figure 1.13 p24 Capsid

On binding to the plasma membrane it is thought that the myristate moiety and the basic surface charges are neutralised by their respective partners in the lipid bilayer. It has been suggested, based on the low hydrophobicity of the myristic acid compared to other long chain fatty acids that myristic acid may be preferentially found in those proteins that interact with membranes in a reversible manner.

An observation that p17 Matrix binds membrane with seemingly less affinity than p55 (Zhou and Resh 1996, Spearman et al 1997) led to the proposal that in p55 the myristate group must be exposed, whilst in p17 Matrix it is sequestered within protein folds. This idea has been described as the conformational myristyl switch model (Zhou and Resh 1996). In this model HIV gag reversibly binds with plasma membrane, whereby in p17 Matrix the myristylated N-terminus may adopt a hidden, membrane inaccessible conformation (Spearman et al 1997). This model has been based on examples of myristylated cellular proteins such as recoverin and ADP ribosylation factor 1, which show a regulated exposure of the myristate group in membrane binding. In the absence of structural data on myristylated p55 or p17 Matrix, predictions on the nature of myristate sequestration have been based on protein mutagenesis strategies. Mutations near the N-terminus of p17 Matrix are proposed to reduce the availability of the myristyl group for insertion, while compensatory mutations within the globular core of the protein compensate for these N-terminal mutations presumably by the prevention of sequestration of the myristyl group (Paillart and Gottlinger 1999, Spearman et al 1997, Ono and Freed 1999).
Mechanisms which may govern membrane association of these two proteins have been proposed to involve the phosphorylation of gag. Yu et al (1995) observed that after phorbol ester treatment and induction of PKC in transfected cells, p17 Matrix rapidly translocated to the membranes, in a manner consistent with an apparent exposure of the myristate group. It is also possible that the cleavage of p55 by the viral protease may trigger the sequestration of the myristate group in the released p17 Matrix protein (Hermida-Matsumoto and Resh 1999).

It is not clear what mechanisms govern the specific binding of p55 to the plasma membrane, rather than other cellular membranes. Large deletions in p17 Matrix can be made which still allow the production and release of virus but at the expense of promiscuous membrane assembly, i.e. at both plasma and intracellular membranes (Reil et al 1998). Plasma membrane targeting and membrane assembly is likely to be a specific function of the matrix domain of gag, as single point mutations in a type D retrovirus (MPMV) Matrix, can change the assembly to a type C virus (Rhee and Hunter 1990, Conte et al 1997, Choi et al 1999). Further domains in p17 Matrix are thought to interact directly with the cytoplasmic tail of the transmembrane gp41 glycoprotein (Freed and Martin 1996), this must presumably have a function in ensuring particle assembly occurs at those regions in the cell membrane where gp120/41 is inserted, and thus ensuring that the virus glycoproteins are incorporated into the virus.

**Gag Assembly**

It is not clear whether the gag protein starts to self assemble in the cytoplasm or if this only occurs at the plasma membrane. Gag protein can be seen to aggregate at the membrane in small electron dense patches (Gelderblom 1991), which as the assembly of the particle progresses buds outwards until such a point, when an enveloped particle is released from the cell surface. For the gag protein to form these structures which bud from the cell surface, the proteins must be capable of assembling with each other to form the immature virus. It is widely thought that gag-gag (p55) interactions are mediated through multiple domains of the protein. Mutational analysis of Capsid has defined specific regions which directly affect assembly. It is thought that the C-terminal domain of p24 Capsid, may be involved in gag-gag recognition (Reicin et al 1995, Kattenbeck et al 1997). In addition to Capsid-Capsid interactions the Matrix
domain of p55 must contribute to the assembly of the immature HIV particle. The p17 Matrix protein of SIV and HIV crystallises as a trimer (Hill et al 1996, Rao et al 1995). If this is physiologically significant these must contribute to gag-gag interaction stability in the assembling particle as interactions taking place in p17 Matrix monomers have recently been shown to induce a trimeric form of p55 gag (Morikawa et al 1998). Domains in p7 Nucleocapsid protein, have also been implicated in having an important role in assembly, identified using deletion studies. C-terminal deletions of gag, removing all of the nucleocapsid domains prevents particle assembly and release (Jowett et al 1992, Carriere et al 1995), a truncation of p55 containing only the first zinc finger of p7 Nucleocapsid could assemble into virions (Dawson and Yu 1998) suggesting that this domain is the one important for assembly.

Proteolytic Cleavage of Gag

Each HIV virion is thought to be composed of about 2000 molecules of gag and about 100 molecules of the gag-pol fusion protein. Concomitant with, or shortly after the release of the virus particle from the infected cell, the viral protease becomes activated. This activation is most likely due to the increased concentration of protease in the virus particle resulting in the dimerisation of two protease monomers. The protease autocatalytically cleaves itself from the rest of the fusion protein and initiates a processing cascade of the p55 gag molecules by recognition of specific sites in gag resulting in an ordered sequence of cleavage (Pettit et al 1998). Following cleavage of the gag polyproteins there is a maturation process within the virion. During this time a Capsid core is formed, which characteristically stains dark when viewed by electron microscopy (Kaplan et al 1993). The proteolytic cleavage is not required for the assembly or release of the virus particle (Khol et al 1998), but is a requirement for the formation of an infectious virus particle (Kaplan et al 1993). In the mature virus particle the gag proteins are cleaved into their constituent parts and are poised for a role in infection, rather than virus assembly.

The p17 Matrix protein, previously the domain responsible for driving particle formation forms a layer of protein associated with the lipid envelope. On infection of a new cell, Matrix has a pivotal role in the early stages of infection. This protein has been implicated in the nuclear targeting of the viral preintegration complex
(Bukrinsky et al 1993, Gallay et al 1995). p17 Matrix has a nuclear localisation signal which it is thought may act in concert with similar activities of Vpr (Heinzinger et al 1994) and Integrase (Gallay et al 1997), however the precise role of p17 Matrix in this respect is unclear. This nuclear targeting of the preintegration complex is thought to be essential for HIV infection of terminally differentiated cells such as macrophages. This nuclear targeting would agree with a conformational myristyl switch mechanism for membrane binding as Gallay et al (1995) reported phosphorylation of p17 Matrix on a tyrosine residue induced the protein to dissociate from the plasma membrane and associate with the preintegration complex.

In the early stages of infection the p24 Capsid is not known to have any other function other than the delivery of the nucleocapsid to the cell cytoplasm. It has been proposed that the dissociation of the Capsid structure may be facilitated by destabilising effects of bound cyclophilin A (Gross et al 1998).

Gag p6

This small protein comprises the C-terminal 51 amino-acids of p55 gag. It is this protein domain in p55 that is thought to be involved in the release of viral particles from infected cells, as deletion of this domain causes the retention of particles at the membrane (Gottlinger et al 1991, Paxton et al 1993). It is possible that p6 interacts with host proteins at the plasma membrane to facilitate the virus particle release (Garnier et al 1996).

Figure 1.14 p6

A further function of the p6 domain during assembly is to bind to and hence cause the incorporation of Vpr into the virus particle, this interaction occurs at the C-terminus.
of p6 (Paxton et al 1993). In the mature particle however it is not clear if this interaction still takes place, and the localisation of p6 within the virus is unclear.

### 1.7 RNA Packaging

During the assembly of an infectious virus, two copies of full length unspliced viral RNA must be selected from a pool of cellular RNAs. This selection must be a specific process as the genomic RNA only accounts for about 1% of the total RNA in the infected cell. Subgenomic and non-viral RNAs are largely excluded from the process of encapsidation (Luban and Goff 1994). There is now unequivocal evidence that RNA packaging in retroviruses involves cis acting packaging signals within the RNA sequence and a trans-acting factor responsible for the recognition of these RNA elements, the gag protein.

#### 1.7.1 Cis-acting Signals

From the fact that genomic RNAs are packaged and spliced messages are not, one might expect that the retroviral packaging signals reside in intron sequence which is not present in those subgenomic messages. The first characterised packaging defects were characterised in a RSV mutant, SE21Qb, which packaged RNA at less than 1% of wild type (Linial et al 1978). This virus had the full complement of genes but lacked a 179bp segment near the 5' end of the genome (Anderson et al 1992). Further packaging defects mutations have subsequently been analysed in ALV (Stoker and Bissel 1988), MMuLV (Mann et al 1983), REV (Watanabe and Temin 1982), and these also map to regions in the 5' UTR. As the region for a packaging signal (psi or ψ) is located in the 5' UTR in other retroviruses, most of the studies in HIV have centred on this region. Initial studies showed that deletions of only 20 nucleotides up to and including the gag start codon could cause a large reduction in RNA packaging (Lever et al 1989, Aldovoni and Young 1990). Other groups reported smaller packaging defects in this region (Luban and Goff 1994, Kim et al 1994). Differences in observed packaging efficiencies may be attributable to the differences in RNA expression in the cell lines used for these studies. Whatever sequence was recognised in the RNA was cis acting as a sequence of 46 nucleotides derived from the genome immediately upstream of the gag start site could be packaged by virus like particles produced from a recombinant vaccinia virus (Hayashi 1992). The structure or motif
required for packaging the RNA genome maps to the leader sequence of the genome in HIV, and other retroviruses.

1.7.2 Trans-acting Factors – The Role of Gag in RNA Packaging.

Virus like particles consisting of just the gag protein are able to assemble and selectively package genomic RNA (Oertle and Spahr 1990, Sakalian et al 1994). This shows that gag must contain all the RNA recognition and RNA binding determinants. It is widely accepted that the nucleocapsid domain of p55 is the region responsible for selectively binding RNA. In the mature infectious particle the NCp7 protein forms a RNP complex with the genome, whilst in the immature virus it must be p55 that makes the initial interaction with the RNA.

Nucleocapsid p7

The Nucleocapsid p7 protein is a proteolytic fragment of the p55 gag protein. It is characterised by a conserved sequence, common to all retroviruses (excluding the spumaretroviridae), known as a Zinc finger motif.

Figure 1.15 Nucleocapsid p7

![Diagram showing domains in p7 responsible for co-ordinating Zinc ions](image)

The sequence Cys-X_2-Cys-X_4-His-X_4-Cys is contained twice in the nucleocapsid of HIV-1 and each independently co-ordinates a Zn^{2+} ion, to form a mini-globular domain (Chance et al 1992, Summers et al 1990, Summers et al 1992), essential for the binding of nucleic acid. In HIV the two zinc finger domains are linked via a small flexible linker sequence (see figure 1.15), each finger domain is not functionally
equivalent. Mutations which disrupt the first (N-terminal) zinc finger motif result in virus particles that do not package genomic RNA as efficiently as wild type and show an increase in subgenomic RNA packaging. The second cysteine-histidine box can be mutated and not significantly affect the specificity of RNA encapsidation (Schwartz et al 1997, Gorlick et al 1993) however as discussed earlier, further stages of the virus lifecycle are probably affected by mutations such as these. As these two motifs only differ in the non-conserved regions, these regions must effect the differences seen in packaging between the two sequences.

*In-vitro,* the p7 Nucleocapsid protein has been shown to exhibit a non-specific RNA binding activity, which may be expected, as early studies showed the genome in the mature virus to be tightly associated with the nucleocapsid protein (Bolognesi et al 1973, Quigley et al 1972). p7 Nucleocapsid binds RNA at a predicted saturation density of 4-7 nucleotides per nucleocapsid protein (Khan and Giedroc 1994, You and Mchenry 1993), with no sequence specificity, consistent with approximately 2000 p7 Nucleocapsid molecules coating the virion genome.

### 1.7.3 Gag –RNA interactions analysed *in-vitro*

The p55 gag protein shows sequence specific RNA binding activity as evidenced by the deletion mutagenesis in viruses previously described. The regions to which gag must recognise are located in the 5' UTR or leader, the sequence described earlier for other important aspects such as dimerisation. Several groups have investigated the interaction between the gag proteins and HIV-1 RNA. The interaction was investigated using northwestern analysis by Luban and Goff (1991), who detected specific RNA binding to p55, by two sequences within 335 nucleotides of the gag coding region. Using a gel shift assay (Geigenmuller et al 1996) a Matrix deleted gag protein bound with greatest affinity to sequence 5' of nucleotide 261 in an HIV RNA transcript, i.e. an RNA containing tar, the poly A hairpin and the primer binding site. Further binding of this protein to RNA was seen using sequence containing the stem loops 2-4. Using transcripts containing SL1-SL4 nucleotides 76-508, p7 Nucleocapsid and GST-p55 both showed specific binding in filter binding assays (Dannull et al 1994, Schmalzbauer et al 1996). The transcripts containing SL1 and small amounts of SL2 did not bind to GST-p55, in contradiction to Geigenmuller et al (1996). Using gel shift experiments Berkowitz and Goff (1994) showed that GST-
p55 bound optimally to RNA containing SL2, SL3 and SL4 when some flanking sequences were included in the transcripts. These collective stem loops on their own in the absence of flanking sequence showed much reduced binding, and the stem loops on their own showed no binding. However SL4 with a 3', 25 nucleotide extension did bind GST-p55. GST-p55 and GST-p15 were both used in filter binding analysis by Clever et al (1995), both proteins specifically bound to transcripts containing SL1,2,3,4 and stem loops SL1,3 and 4 also bound on their own but with reduced affinity compared to the full length transcript. These publications have all shown that gag proteins can bind specifically to several RNA elements in the 5'UTR of HIV-1. Contradictory evidence for the binding of p55 to some sequences and the lack of clarity in this area shows that this interaction deserves further study.

**In-vitro Selected Evolution of RNA Ligands.**

Several groups have attempted SELEX type experiments to analyse RNAs which can be selected for by virtue of their affinity to gag protein. Binding of RNA by p7 Nucleocapsid was analysed by Allen et al (1996) who showed that a stem loop structure with a stem bulge was essential for the specific binding of p7 Nucleocapsid, which although doesn't resemble sequence from HIV-1 may be structurally significant. p7 Nucleocapsid was used by Berglund et al (1997), to show by SELEX that stem loop structures rich in G and U residues, also possessing a bulge region but not resembling HIV sequence bound the protein with high affinity. The fact none of these RNAs resemble HIV sequence is probably due to the fact that in the case of the virus particle and virus lifecycle p7 Nucleocapsid protein has multiple roles, which must require a compromise between specific and non-specific RNA binding. One other such experiment has been carried out by SELEX, comparing p55 gag with p17 Matrix, p24 Capsid and p7 Nucleocapsid. Lochrie et al (1997), showed that GST-p55 bound to multiple ligands with high affinity, the ligands isolated bound to two domains, either p17 Matrix or p7 Nucleocapsid independently, again selected RNA sequence showed no similarity to that from the viral genome.

1.7.4 In-vivo packaging Studies

*In-vitro* studies allow accurate biochemical analysis of interactions that may be occurring during the packaging process, however these theories also need to be
analysed in-vivo so as to assess their significance. The emerging picture of packaging in virions has centred on the role of the elements in the 5' UTR. It is thought that three of the hairpins SL1, SL2 and SL3 are important for encapsidation. McBride and Panganiban (1997) showed that at least two of these, SL1 and SL3 function in RNA packaging, the presence of one of them being essential for efficient packaging. It was shown that the removal of both of these sequences does not completely abolish packaging, suggesting that some further packaging motifs may be recognised by gag. In another study the structures SL1-SL4 were compared in a competitive packaging assay, which utilised point mutations to disrupt the formation of stem loop structures. Using this methodology SL2 was seen to have only slight effects on the relative encapsidation efficiency of genomic RNA, whilst SL4 and SL1 have fairly comparable effects. Disruption of SL3 (as shown by Lever et al 1989) showed the greatest affect on RNA packaging (McBride and Panganiban 1997). Functionally SL1 and SL3 are dependent on their location within the leader RNA, which may be consistent with the packaging signal consisting of a higher ordered structure involving tertiary interactions between SL1, 3 and 4.

A deletion of SL3 and SL4 by Luban and Goff (1994) resulted in the packaging of both genomic and spliced sequences, with no enrichment of the genomic RNA, consistent with the positioning of SL3 and 4 in intron sequence. Sequences in the 5' region of the gag coding region have also been shown to have an effect on RNA packaging (Luban and Goff 1994, Buchschacher and Panganiban 1992) although the significance of this sequence and its exact location is not known. As this sequence is only present in genomic RNA it is likely that this may have a role in RNA packaging.

Sequences at the 5' end of the viral genomic RNA, 5' to SL1, have been analysed using vectors deleted in Tar, Poly A hairpin and the primer binding site hairpin. Deletion of any three of these sequences results in a substantial reduction in encapsidation (McBride et al 1997). This region has been shown to bind to gag in-vitro (Geigenmuller and Linial 1996). Recently it has been shown in-vivo that the region responsible for packaging in tar, maps to the lower portion of the tar stem, showing that the stem integrity is important for packaging (Helga-Maria et al 1999).

Some work has proposed that sequences within env, spanning the RRE, may be required for packaging, (Kaye et al 1995, Richardson et al 1993), but it is likely that
any increased packaging effect due to these sequences is due to an indirect effect on packaging and probably results from enhanced transport of RRE containing RNA from the nucleus to the cytoplasm (McBride et al 1997).

1.7.5 Regulation of RNA Packaging

Little is known mechanistically about how an RNA is selected for packaging into an assembling particle. It is possible that there is no regulatory mechanism governing this and packaging and translation of RNA is a random event. It is also possible that RNA may be targeted for packaging or translation. It may be that there are two ‘pools’ of genomic RNA, whereby one is destined for packaging and one is destined for translation. It has been shown in MuLV infected cells that if RNA synthesis is arrested by the addition of Actinomycin D, viral RNA incorporation is arrested after about 3 hours, however virus particles lacking RNA were still produced after 12 hours. This is consistent with an idea that there may be two pools of RNA in the infected cell, one pool for packaging and one pool for translation (Messer et al 1981).

It is possible that genome dimerisation causes a switch from translation to packaging. Bieth et al (1990) showed the inability of dimeric RNA to be translated in vitro, this may cause a switch (possibly RNA concentration dependent) which causes the RNA to be packaged rather than translated.

A further method which may produce an RNA packaging pool in an infected cell, could occur if RNA was selected for packaging straight after transcription, before splicing or recognition by rev. This situation may arise in the infected cell as there have been reports of retroviral gag proteins residing in the nucleus (Royer et al 1991, Nash et al 1993, Schliephake and Rethwilm 1994). Gag may bind genomic RNA in the nucleus and ‘select’ the RNA to be packaged.

Recent evidence has been put forward suggesting that HIV-1 and HIV-2 may differ in their mechanisms of RNA selection and packaging. HIV-1 RNA can act as both mRNA and genomic RNA, the translation of gag in cis is not necessary for efficient encapsidation (McBride et al 1997) and HIV-1 can be efficiently used as a vector suggesting that RNA translation is not a requirement for packaging. HIV-2 on the otherhand shows a requirement for packaging by gag, in cis, whereby translated gag
only binds to its own message (Kaye and Lever 1999), which maybe forms a site for gag nucleation.

RNA domains in SL1 which stop dimerisation, such as mutations at the ACS, cause a reduction in RNA packaging (Laughrea et al 1997, Clever and Parslow 1997). This may be due to either the dimer structure being recognised by gag for packaging or by the fact that if gag recognises a structure in monomeric RNA, then an ACS mutation will only allow the packaging of one RNA and the observed packaging effect would be reduced. In this respect it is likely that dimerisation is not a pre-requisite for packaging RNA.

1.8 Project outline
The aims of this project were to further characterise the interactions which contribute to the process of genomic RNA packaging in HIV-1 using both in-vivo and in-vitro methods. In addition, the role of the kissing loop complex in RNA dimerisation which may form at the DLS was to be further investigated using a direct approach on extracted virion RNA.

The HIV-1 leader RNA was subject to a deletion analysis in an attempt to identify regions that were important in RNA packaging, by binding to the p55 gag protein. This identified structures which are important for the recognition by gag in some RNAs. Using more sensitive approaches, some residues implicated in the binding of p55 were identified, which up to that time had not been defined. Using subcellular fractionation techniques the location of RNA packaging selection was investigated, and has shown that selection of RNA is likely to occur in the cytoplasm by the binding of p55 gag to genomic RNA transcripts.

At the outset of this project the first published results on kissing loop deficient virion had shown that although this interaction was involved in the formation of virion RNA dimers, its requirement was not absolute. The fact that some RNA is still dimeric under these situations suggested a role for the kissing loop interaction in maintaining dimer stability. The nature of this interaction requires biochemical characterisation in virion extracted RNA dimers, which was carried out using a competitive oligonucleotide strategy.
Chapter 2

Materials and Methods

2.1 DNA techniques

2.1.1 Agarose Gel Electrophoresis of Nucleic Acids

<table>
<thead>
<tr>
<th>1X TBE</th>
<th>1X TAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 mM Tris pH 8.0</td>
<td>40 mM Tris acetate pH 8.3</td>
</tr>
<tr>
<td>90 mM boric acid</td>
<td>1mM EDTA</td>
</tr>
<tr>
<td>2 mM EDTA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1X TB</th>
<th>Loading Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 mM Tris pH 8.0</td>
<td>0.25 % w/v bromophenol blue</td>
</tr>
<tr>
<td>90 mM boric acid</td>
<td>40 % w/v sucrose in water</td>
</tr>
</tbody>
</table>

Gels were prepared using ICN Genetic Technology grade agarose containing 0.5μg/ml ethidium bromide. For analytical gels, these were prepared with, and run in 1X TBE buffer at a constant voltage of 40-90 V for a 75mm x 85mm minigel. Preparative gels were made and run in 1X TAE buffer. DNA samples were run, after mixing loading dye to 25% v/v to the sample. Molecular weight markers were run along side the sample lanes.

λ Hind III markers
23130; 9416; 6557; 4361; 2322; 2027; 564 and 125 bp.

φX174 Hae III Markers
1353; 1078; 872; 603; 310; 281; 271; 234; 194; 118 and 72 bp.

2.1.2 Restriction Digestion.
Restriction enzymes were purchased from Gibco BRL or New England Biolabs. Restriction digestions were carried out, typically in a reaction containing a five fold excess of enzyme to DNA at 37°C for 1-2 hours, in the diluted 10X buffer supplied with the enzyme. After digestion samples were either subjected to analytical
electrophoresis in agarose gels, or were heat inactivated at 65°C for 10 minutes. Then the DNA was recovered by phenol chloroform extraction and ethanol precipitation.

<table>
<thead>
<tr>
<th>Enzymes Used</th>
<th>Buffers Used at 1X concentration (Gibco)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BamHI</td>
<td>React 2 (50mM Tris pH 8.0, 10mM MgCl₂, 50mM NaCl)</td>
</tr>
<tr>
<td>EcoR I</td>
<td></td>
</tr>
<tr>
<td>Hind III</td>
<td>React 3 (50mM Tris pH 8.0, 10mM MgCl₂, 100mM NaCl)</td>
</tr>
<tr>
<td>Pvu II</td>
<td></td>
</tr>
<tr>
<td>Sph I</td>
<td>React 6 (50mM Tris 7.4, 6mM MgCl₂)</td>
</tr>
<tr>
<td>Xho I</td>
<td>50mM KCl, 50mM NaCl)</td>
</tr>
<tr>
<td>Nco I</td>
<td></td>
</tr>
</tbody>
</table>

2.1.3 Phenol / Phenol Chloroform extraction

To remove DNA from associated proteins in a sample, an equal volume of phenol (equilibrated with 0.1M Tris pH 8.0), or a solution of phenol / chloroform / isoamyl alcohol (25:24:1) was added to the sample. The mixture was vortexed vigorously for 30 seconds then the solution spun in a microcentrifuge at 13000rpm for 5 minutes. The top aqueous layer containing nucleic acid was then removed and this procedure repeated as required. RNA was also extracted by phenol extraction using the same technique only citrate equilibrated phenol (pH 4.0) was used.

2.1.4 Ethanol precipitation

DNA and RNA was precipitated from aqueous solution by the addition of 0.1 volumes of 3M sodium acetate (pH 5.2) and 2.5 volumes of 95% ethanol. The tube was then vortex mixed and incubated on dry ice for 15 minutes or at -20°C for 30 minutes. The precipitated nucleic acid was recovered by spinning in a microcentrifuge for 15 minutes at 13000rpm, or for 15 minutes at 2500rpm in a mistral 3000i centrifuge. The supernatant was removed and the pellet was allowed to air dry for 15 minutes. Dried pellets were resuspended in sterile nanopure (np) water (DNA), or DEPC treated water (RNA). Nucleic acids were quantitated by UV absorbance at 260nm assuming that the an OD of 1 is equal to 50μg/ml for DNA, or 40μg/ml for RNA samples.

2.1.5 Gel purification of DNA

DNA was recovered from TAE agarose gels using the Geneclean II kit (Bio 101) following the manufacturers protocol. Briefly the DNA band of interest was located using a hand held UV transilluminator. The band was carefully excised using a clean scalpel blade, and excess agarose was trimmed off. Three volumes of 3M sodium iodide was added to the gel and this was heated to 60°C for 5 minutes.
suspension was then added to the solution and incubated at room temperature for 5 minutes. The resin was then recovered by centrifugation for 30 seconds at 13000 rpm in a microcentrifuge. The pellet was then washed using New Wash buffer. This step was repeated a further two times. The DNA was eluted from the beads by the addition of nanopure water and incubation at 65°C for 5 minutes.

2.1.6 DNA Ligations
Ligations were set up in 20μl volumes of 1X DNA ligase buffer, (50mM Tris-HCl pH 7.6, 10mM MgCl₂, 1mM ATP, 1mM DTT, 5% PEG), (Gibco BRL), containing 50ng vector DNA, (appropriately restricted). Insert DNA was added, usually giving a 3:1 ratio molar excess, insert : vector. One unit of T4 DNA ligase was added and the reaction incubated at room temperature for 16 hours. Generally half of the reaction mix was directly transformed into transformation competent E. coli.

2.1.7 Preparation of Transformation Competent E. coli

Transformation competent E. coli were prepared by inoculating a 250ml flask of 2X YT with a 1:50 volume of an overnight culture of E. coli DH5α. These cells were grown to an OD₅₉₅ of 0.6 in a 37°C shaking incubator set at 200rpm. The cells were pelleted at 4°C / 5000rpm in a Sorval Superspeed centrifuge using a GS3 rotor. The cell pellet was then resuspended on ice in 125 mls of sterile ice cold 10mM NaCl. The cells were again pelleted then resuspended in 62.5 mls ice cold 30mM CaCl₂. The cells were again pelleted then resuspended in 10 mls 30mM CaCl₂/15% glycerol and left on ice for 10 minutes. The cells were then aliquotted in 200μl volumes into sterile eppendorfs and snap frozen in an ethanol dry ice bath. Competent E. coli were stored at -80°C for up to six months.

2.1.8 Transformation of E. coli

<table>
<thead>
<tr>
<th>LB (litre)</th>
<th>10 g Bacto Tryptone</th>
</tr>
</thead>
<tbody>
<tr>
<td>5g Bacto yeast extract</td>
<td></td>
</tr>
<tr>
<td>10g NaCl</td>
<td></td>
</tr>
</tbody>
</table>
200μl of frozen competent *E. coli* were thawed on ice. DNA was added to the cells in a volume less than 20μl and mixed. The cells were left on ice for 10 minutes, then heat shocked for exactly 2 minutes by placing in a water bath set at 42°C, then the tubes were returned to ice. After 10 minutes 200μl of LB was added to the cells and aliquots of this transformation preparation were plated onto LB agar plates containing 100μg Ampicillin, and the plates incubated overnight at 37°C. All plasmids used in this thesis contained an ampicillin resistance gene, and were grown in media containing 100μg/ml ampicillin.

2.1.9 Isolation of Plasmid DNA

Miniprep Procedure

<table>
<thead>
<tr>
<th>Solution 1</th>
<th>Solution 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>50mM Tris pH 7.5</td>
<td>0.2M NaOH</td>
</tr>
<tr>
<td>10mM EDTA</td>
<td>1% w/v SDS</td>
</tr>
<tr>
<td>100μg/ml RNase A (Sigma)</td>
<td></td>
</tr>
</tbody>
</table>

**Diatomaceous Earth Resin**

- 1% w/v prewashed diatomaceous earth resin (Sigma)
- 4M guanidine hydrochloride
- 50mM Tris-HCl pH 7.0
- 10mM EDTA

For small scale preparation of plasmid DNA, or when screening transformants for inserts, a colony from a fresh transformation was grown up overnight in 2mls of LB media containing 100μg/ml ampicillin. 1.5 mls of this culture was then transferred to an eppendorf and the cells pelleted by centrifugation for 5 minutes at 13000rpm in a microcentrifuge. The supernatant was removed by aspiration and the pellet then resuspended in 150μl of solution 1 by vortexing. 150μl of solution 2 was then added and the mixture mixed gently by inversion until the solution had cleared. 150μl of 3M potassium acetate was then added and mixed by inversion and precipitate pelleted in a 5 minute spin at 13000rpm in a microcentrifuge. The supernatant was removed to a fresh tube, and 1ml of the resin preparation was added, this was mixed by inversion and the DNA allowed to bind to the resin during an incubation at room temperature for 1 minute. The tube was then centrifuged for 1 minute at 13000 rpm in a microcentrifuge and the supernatant discarded. 1ml of 80% isopropanol was then added and the tube vortex mixed. The resin was again pelleted and the supernatant removed. The resin pellet was then washed with acetone and again pelleted and the acetone removed. The
pellet was then allowed to air dry for 10 minutes, before being resuspended in sterile nanopure water. The DNA was eluted from the resin by heating for 10 minutes at 60°C, then removing the supernatant from the resin pellet after a 1 minute spin at 13000 rpm in a microcentrifuge. For restriction digestion analysis typically 5-10 μl of this preparation was used. For manual sequencing 50-70μl was used directly.

**Caesium Chloride preparation of plasmid DNA - Maxiprep**

<table>
<thead>
<tr>
<th>ST</th>
<th>Triton lysis buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>25% w/v sucrose</td>
<td>1.5% v/v Triton X100</td>
</tr>
<tr>
<td>50mM Tris-HCl pH8.0</td>
<td>50mM Tris-HCl pH 8.0</td>
</tr>
<tr>
<td></td>
<td>50mM EDTA pH 8.0</td>
</tr>
</tbody>
</table>

A two litre culture flask containing 500mls 2xYT media, 100μg/ml ampicillin was inoculated with a 1:50 dilution of an overnight *E. coli* culture and grown overnight. The cells were recovered by centrifugation in a Sorval GS3 rotor at 5000rpm for 10 minutes. The cell pellet was resuspended in 2.5mls of ST buffer. 1ml of a 20mg/ml stock of lysozyme (BDH) in ST was added and the sample incubated on ice for 10 minutes. Following this step 1.5mls of 0.5M EDTA was added to the mix and incubated on ice for a further 10 minutes. The cells were then lysed by the addition of 2.5mls Triton lysis mix, added with gentle mixing. The mix was then centrifuged for 45 minutes at 20000rpm in a Sorval SS34 rotor, and the clear supernatant was carefully decanted. The volume of the supernatant was measured and caesium chloride was added to 1g/ml. The caesium chloride was dissolved by incubation at 37°C for 10 minutes, then the tubes were centrifuged at 2500rpm in a MSE Mistral 3000i centrifuge for 5 minutes. The supernatants were removed and added to Beckman 3.9ml ultracentrifuge heat seal tubes together with 25μl 10mg/ml ethidium bromide. After heat sealing, the tubes were spun at 80000 rpm overnight or at 100000 rpm for 5 hours. Plasmid bands were recovered by side puncture of the tubes using a 2 ml syringe and a 20g needle. Ethidium bromide was removed from the sample by extraction with an equal volume of water / caesium chloride saturated propan-2-ol until the solution was colourless. The extracted DNA was then diluted with 2 volumes of TE pH 7.5 and the DNA recovered by ethanol precipitation. DNA was resuspended in np water and quantitated by optical density at 260nm.

### 2.1.10 Polymerase Chain Reaction (PCR)

All reactions used vent DNA polymerase from New England Biolabs (NEB). Vent was used because of its higher fidelity and lower misincorporation rate than other
thermostable polymerases. Reactions were optimised for primer annealing temperature and in some instances the magnesium concentration was optimised.

**Typical Reaction**

- 200µM each dNTP
- 25µM each primer
- 2.5 units vent DNA polymerase
- 1-20ng DNA template

In 1X PCR buffer : 10mM KCl, 20mM Tris pH 8.0, 0.1% Triton X100, 2mM MgSO₄

Final reaction volume 50µl

Reactions were set up on ice and then overlain with mineral oil (Sigma). PCR cycling was then performed. Typically: 95°C for 1 minute

45-60°C for 1 minute

72°C for 1 minute

Generally PCR was performed for 30 cycles. 10µl of the PCR reaction was analysed by TBE agarose gel electrophoresis. PCR products were then recovered from the reaction mix by phenol-chloroform extraction and ethanol precipitation.

### 2.1.11 DNA Sequencing

Double stranded DNA was sequenced by the chain termination dideoxy method (Sanger et al 1977) using the Amersham Sequenase kit. Briefly, double stranded plasmid DNA 5-10µg was denatured by the addition of NaOH and EDTA to 0.4M and 0.25mM respectively. This was incubated for 30 minutes in a 37°C waterbath. Denatured DNA was then ethanol precipitated and the DNA recovered by centrifugation at 13000rpm in a microcentrifuge. After washing with 70% ethanol and air drying the DNA was resuspended and used in sequencing reactions as per the manufacturers instructions. Sequencing reactions were run on 6% polyacrylamide TBE gradient gels.

<table>
<thead>
<tr>
<th>40% Acrylamide stock (per l)</th>
<th>6% Top (per l)</th>
<th>6% Bottom (per l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>380g acrylamide</td>
<td>150ml 40% stock</td>
<td>75ml 40% stock</td>
</tr>
<tr>
<td>20g bis-acrylamide</td>
<td>50ml 10X TBE</td>
<td>125mls 10X TBE</td>
</tr>
<tr>
<td></td>
<td>460g urea</td>
<td>230g urea</td>
</tr>
</tbody>
</table>

Two clean 20x39cm glass plates were separated by two 0.3mm spacers and sealed with gel tape. The plates were clamped together with bulldog clips. 32mls of top solution and 14 mls of bottom solution were measured out, and 70µl TEMED and 70µl 25% w/v ammonium persulphate was added to the top solution. 30µl of TEMED and 30µl
of 25% APS was added to the bottom solution, both solutions were mixed thoroughly. 25mls of the top solution and 7 mls of the bottom solution were then drawn into a pipette and a buffer gradient was formed by drawing up a few bubbles into the pipette. The acrylamide mix was then poured between the sequencing gel plates and the comb inserted between the plates. The gel was allowed to polymerise for 1 hour then the gel tape was removed and the gel was placed in the sequencing gel apparatus. The comb was removed and the wells were washed thoroughly with running buffer (1XTBE). 4μl of each sequencing reaction was then loaded per lane then the gels run at 37W. After the gels had run, the plates were separated and the gel fixed in a 10% IMS 10% acetic acid bath for 10 minutes. The gel was transferred to Watmann 3MM paper and then dried on a gel vacuum drier set at 80°C for 1 hour. Gels were visualised by autoradiography, usually involving an overnight exposure to GRI X-ray film.

2.1.12 Plasmid construction
Constructs for the transcription of viral RNA sequences were made by PCR from cloned plasmid sequences. The PCR primers contained restriction sites to facilitate cloning and bacteriophage RNA polymerase promoter sequences 5' to the HIV-1 sequences. Generally 5' primers contained an EcoR I restriction site, whilst 3' primers contained Hind III sites (see 2.1.13). Amplified products were ligated into an appropriately restricted vector. For RNA transcription pUC119 was used, here the bacteriophage promoter sequence was incorporated into the PCR product on the forward primer. The cloned sequences all utilised transcription starts at authentic HIV-1 sequences. All plasmid inserts were fully manually sequenced using sequenase (Amersham).

2.1.13 Oligonucleotides
All oligonucleotides were produced by the Protein and Nucleic Acid Chemistry Laboratory at the University of Leicester. All oligonucleotides were ethanol precipitated and then resuspended in nanopure water. Some oligonucleotides were additionally purified on denaturing polyacrylamide gels. The oligonucleotides shown overleaf are written 5'-3' and the positions they anneal to in the virion RNA genome of HIV-1NL-43 are highlighted. Sequence between highlighted residues anneals to genomic sequence. Forward primers are indicated by increasing numerical positions. For backward primers the order is reversed. Other sequence in these oligonucleotidess corresponds to restriction site or bacteriophage RNA polymerase promoter element.
e.g. BM09

5' TCAGAATTCTAATCCGACTCACTATAGAGATGGGTGCGAGAGCGTC 3'

Δ Eco RI Δ T7 RNA Polymerase Δ Target Primer sequence Δ Promoter

Oligonucleotide sequence 5'-3'

BM01 GCGTAATACGACTCACTATAG (T7 RNA polymerase promoter 'top' oligo)
BM02 AAGCGGCACGGCTTATAGTGAGTCGTATTACGC (255-267)
BM04 TGCCTGCGCTT (263-255)
BM05 CTTGCCTGTGCGCTTCAGCAAG (270-248)
BM06 ATTTCCGACGGCTTATAGTGAGTCGTATTACGC (255-267)
BM07 CTTGCCCTAGAGATTACGCAAG (270-267)
BM08 AGGATCTTGGCCTTAACCGAATTTTTTCCCATATTTATCTAATTCTC CCCCCGTTTCCTCCTTCTTCAGCGCTGGATCCAGAAGGGGGCCGAGGACG (406-297)
BM09 TCAGAATTCTAATCCGACTCATTTAGAGATGGGTGCGAGAGCGTC (338-352)
BM10 TCGAAGCTTTAATACCGACGCTCTCG (360-343)
BM11 TCAGAATTCTAATCCGACTCATTTAGAGATGGGTGCGAGAGCGTC GTC (338-352)
BM12 TCGAAGCTTTGATACCGACGCTCTCG (357-344)
BM14 TATAACCGTGGGTGCGAGAGCGTCTAGTATTAAGC (454-480)
BM15 CTTGTTTAATCTACTGCCCCTTC (4532-4510)
MH14 CAGTATACGCGTGAGAATTTGATAG (352-380 CSF)
MH17 CGCCTCTTGGCGTCCTTCTCTCTCTCTCTCTGCAAGCCGAGGCTAG (275-238)
MH20 GCTGTCGCGCTTCAGCAAGCC (263-246)
AL17 TCAGAATTCTAATACGACTCAGCTCTATAAGGGCGTCTGTTAAG ACC (1-18)
AL18 TGAGGATCCCTGTGCTTAAACCAGAATTTTTTCCC (406-382)
PM02 TCGAAGCTTTCTCTCTCTCTCTAGCAGCCTCG (335-316)
SH232 TCGGAATCTAATACGACTCAGCTCTATAAGGGCGTGAAGAAG AGAAG (317-335)
SH241 TCGAAGCTTTAATACGACGCTCTCG (813-797)

Sequencing Primer (M13 -40 primer) GTTTTCCCAGTCAGGAC
2.1.14 Estimation of Primer concentration

Usually DNA oligonucleotides were purified from the crude preparation by ethanol precipitation, or gel purification. DNA oligonucleotide concentration was estimated by determining the $A_{260}$ of an appropriate dilution of the oligonucleotide preparation, to within the linear range of the spectrophotometer i.e. between 0.1 and 1. For each oligonucleotide the number of A,T,C and G residues was counted, and the concentration in pmol/μl determined from the absorbance of a 1000pmol/μl solution of each base.

e.g. BM05

<table>
<thead>
<tr>
<th>Residue</th>
<th>Count</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3x15.4</td>
<td>46.2</td>
</tr>
<tr>
<td>C</td>
<td>8x7.3</td>
<td>140.8</td>
</tr>
<tr>
<td>G</td>
<td>7x11.7</td>
<td>81.9</td>
</tr>
<tr>
<td>T</td>
<td>5x8.8</td>
<td>44</td>
</tr>
</tbody>
</table>

Σ = 312.9

The $A_{260}$ of a 1:100 dilution of the oligonucleotide was 0.103

so: \[ \frac{10.3 \times 1000}{312.9} = 32.9 \text{ pmol/μl} \]

2.2 RNA Work

2.2.1 Precautions taken when working with RNA

All RNA work was carried out using solutions prepared with 0.1% v/v DEPC treated nanopure water. Whilst carrying out all procedures latex gloves were worn and these were changed regularly. Glassware and plasticware such as gel tanks was rinsed with 3% hydrogen peroxide then DEPC treated water before use. RNA was generally stored as an ethanol precipitate at -80°C. RNasin (Promega) or Prime (5'-3' labs), recombinant protein inhibitors of ribonuclease, were included in some reactions.

2.2.2 In-vitro Transcription

Large amounts of cold RNA were produced by transcription using Ribomax kits from Promega. Briefly, plasmids were linearised by digestion with Hind III. The linear DNA was then gel purified and used in the reactions as per the manufacturers instructions.

Smaller transcripts were made by 'Ulhenbeck' transcription. Here transcription by T7 polymerase was initiated from two annealed overlapping oligonucleotides. BM1 was the top oligonucleotide and the bottom oligonucleotide contained the homologous T7 RNA polymerase promoter sequence and the antisense sequence to be transcribed (BM2 and BM6). These oligonucleotides were designed so that transcription would start at a G residue, at an authentic HIV-1 sequence, this has been demonstrated to give better yields of RNA during transcription (Milligan et al 1987). Briefly 40pmol of each
oligonucleotide were mixed in a 20μl reaction mix (40mM Tris-HCl pH 8.1, 5mM DTT, 1mM Spermidine and 0.01 % Triton X100). This was heated to 95°C and allowed to cool to room temperature in a beaker of water. PEG 8000 was added to 40mgs/ml and rNTPs added to a concentration of 2mM. 2μl of T7 RNA polymerase (purified in house) was added together with 10u RNase inhibitor (Rnasin Promega) and the reaction was incubated for 4 hours at 37°C. EDTA was then added to a final concentration of 30mM. An equal volume of formamide loading dye was added and the RNA was heated at 95°C for 2 minutes then run on a polyacrylamide gel and gel purified. Radiolabelled RNA was transcribed from linearised plasmid templates in 10 or 20μl reaction volumes in 1x transcription buffer (80mM HEPES-KOH pH 7.5, 24mM MgCl₂, 2mM Spermidine, 40mM DTT). Reactions contained 1-2μg DNA, 500μM rATP rGTP rCTP and 2.5-5μl of α³²P-UTP (25-50μCi specific activity 160Ci/mmol) cold UTP was included in the reactions at a concentration of 15μM. 2μl of T7 RNA polymerase from an in house purified stock was used per transcription. Transcription was carried out at 37°C for 1 hour then an equal volume of formamide loading dye was added, the sample heated to 95°C for 2 minutes then the RNA gel purified. RNA was stored in ethanol at -20°C, and recovered as and when required. Radiolabelled RNA was generally only used within 48 hours of transcription before being discarded.
## Table 1

**Plasmids used for the Transcription of RNA**

<table>
<thead>
<tr>
<th>RNA</th>
<th>Vector</th>
<th>Cloning sites</th>
<th>HIV RNA Sequence</th>
<th>Promoter</th>
<th>Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>pUC119</td>
<td>EcoR1/BamH1</td>
<td>1-406</td>
<td>T7</td>
<td>pAL94.113</td>
</tr>
<tr>
<td>H2</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>195-404</td>
<td>T7</td>
<td>pMH93.010</td>
</tr>
<tr>
<td>H3</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>254-404</td>
<td>SP6</td>
<td>pMH93.017</td>
</tr>
<tr>
<td>H4</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>278-404</td>
<td>T7</td>
<td>pSH91.200</td>
</tr>
<tr>
<td>H5</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>195-362</td>
<td>T7</td>
<td>pMH94.004</td>
</tr>
<tr>
<td>H6</td>
<td>pUC119</td>
<td>EcoR1/HindII</td>
<td>278-336</td>
<td>T7</td>
<td>AL.2</td>
</tr>
<tr>
<td>H7</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>317-404</td>
<td>T7</td>
<td>B1-1</td>
</tr>
<tr>
<td>H8</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>338-404</td>
<td>T7</td>
<td>CL-4</td>
</tr>
<tr>
<td>H9</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>338-352</td>
<td>T7</td>
<td>D-1</td>
</tr>
<tr>
<td>H10</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>338-362</td>
<td>T7</td>
<td>pBM96.001</td>
</tr>
<tr>
<td>H11</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>338-379</td>
<td>T7</td>
<td>pI+</td>
</tr>
<tr>
<td>H12</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>338-389</td>
<td>T7</td>
<td>pJ+</td>
</tr>
<tr>
<td>H13</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>353-404</td>
<td>T7</td>
<td>pG+</td>
</tr>
<tr>
<td>H14</td>
<td>pUC119</td>
<td>EcoR1/BamH1</td>
<td>1-335</td>
<td>T7</td>
<td>pL</td>
</tr>
<tr>
<td>H15</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>317-362</td>
<td>T7</td>
<td>pM1</td>
</tr>
<tr>
<td>H17</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>317-379</td>
<td>T7</td>
<td>pBM97.001</td>
</tr>
<tr>
<td>H18</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>317-389</td>
<td>T7</td>
<td>pN1</td>
</tr>
<tr>
<td>H19</td>
<td>pUC119</td>
<td>EcoR1/BamH1</td>
<td>1-406 (ΔSL4)</td>
<td>T7</td>
<td>pBM97.002</td>
</tr>
<tr>
<td>H20</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>195-404 (Δ)</td>
<td>T7</td>
<td>pBM97.005</td>
</tr>
<tr>
<td>H21</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>254-404 (Δ)</td>
<td>T7</td>
<td>pBM97.006</td>
</tr>
<tr>
<td>H22</td>
<td>pUC119</td>
<td>EcoR1/BamH1</td>
<td>278-404 (Δ)</td>
<td>T7</td>
<td>pBM97.004</td>
</tr>
<tr>
<td>H23</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>325-389</td>
<td>T7</td>
<td>pBM97.008</td>
</tr>
<tr>
<td>H24</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>333-389</td>
<td>T7</td>
<td>pBM97.009</td>
</tr>
<tr>
<td>H25</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>325-362</td>
<td>T7</td>
<td>pBM97.010</td>
</tr>
<tr>
<td>H26</td>
<td>pUC119</td>
<td>EcoR1/HindIII</td>
<td>333-362</td>
<td>T7</td>
<td>pBM97.011</td>
</tr>
<tr>
<td>C1</td>
<td>pBSsk</td>
<td>SmaI</td>
<td></td>
<td>T7</td>
<td>pJR</td>
</tr>
</tbody>
</table>

### 2.2.3 5' End Labeling of RNA

RNA was 5' labelled with $^{32}$P by incubating 200pmol unlabelled *in-vitro* transcribed RNA with 4units calf intestinal alkaline phosphatase (CIP), (Gibco-BRL) in 1X buffer (1mM ZnCl$_2$, 1mM MgCl$_2$, 10mM Tris-HCl pH 8.3), at 50°C for 30 minutes. The reaction was phenol-chloroform extracted and then ethanol precipitated. This dephosphorylated RNA was then kinased using 20units polynucleotide kinase (New England Biolabs) in 20μl 1X buffer (50mM Tris-HCl pH 8.2, 10mM MgCl$_2$, 0.1mM EDTA), containing 5μl $\gamma^{32}$P-dATP (370MBq/ml;111TBq/mmol - NEN), for 30 minutes at 37°C. This labelled RNA was then gel purified. Alternatively the unincorporated counts were removed by a double ethanol precipitation. The gel purified RNA was generally stored at -20°C in the gel elution buffer until needed.

### 2.2.4 In-vitro RNA Dimerisation

**Cacodylate buffer**

50mM sodium cacodylate pH 7.5
Two types of dimer can be formed with *in vitro* transcribed RNA containing the kissing loop hairpin (SL1), nucleotides 198-280. Haddrick *et al* (1996) describe formation of RNA dimers with these RNAs by phenol extraction. These dimers are TBE gel stable, magnesium independent and are most likely of the extended conformation, (see diagram 1.10). The kissing loop dimer as described by (Skripkin *et al* 1994) was formed by heat denaturing the RNA in Cacodylate buffer for 1 minute at 90°C. This was followed by a slow cool to room temperature. These RNA dimers are not TBE stable and require the presence of Mg\(^2+\) in the agarose gel to stay together. These kissing loop dimers were run on GTG (FMC) agarose gels made with TBM\(_{0.1}\) (Tris Borate buffer with MgCl\(_2\) at 0.1mM).

### 2.2.5 Extraction of Virion Genomic RNA

<table>
<thead>
<tr>
<th>20% Sucrose TNE</th>
<th>Lysis Mix</th>
</tr>
</thead>
<tbody>
<tr>
<td>20% w/v sucrose (BDH)</td>
<td>1% SDS</td>
</tr>
<tr>
<td>10mM Tris-HCl pH 8.0</td>
<td>50mM Tris-HCl pH 7.5</td>
</tr>
<tr>
<td>100mM NaCl</td>
<td>100mM NaCl</td>
</tr>
<tr>
<td>1mM EDTA</td>
<td></td>
</tr>
</tbody>
</table>

Virion RNA was harvested from infected cell culture supernatants typically about 72 hours post infection and 16 hours post media change. Cells were pelleted by a 10 minute spin in a mistral centrifuge at 1250rpm and the culture supernatant was filtered through a 0.45μm Acrodisk. The supernatant was split between polypropylene centrifuge tubes and 3 mls of sucrose /TNE was carefully layered under the culture supernatant. The tubes were spun at 26000rpm in a Sorval AH629 rotor. The media was removed by pipette and the sucrose cushion was then poured off to waste. The tubes were then left inverted for 5 minutes to drain and then 300μl lysis solution was added and the resuspended pellets were transferred to 1.5ml eppendorf tubes. Proteinase K was then added to 200ng/μl along with 10μg/μl yeast tRNA. The tubes were then incubated at 37°C for 30 minutes. Viral RNA was then recovered by one phenol and two phenol-chloroform extractions. The RNA was then stored as an ethanol precipitate at -80°C until use.
2.2.6 Northern Blot Analysis of Virion RNA

Sodium Phosphate transfer buffer
25mM Na$_2$HPO$_4$ pH 6.5
25mM Na$_2$H$_2$PO$_4$

20X SSC
3M NaCl
0.3M sodium citrate

Prehybridisation Buffer
15mls formaldehyde (BDH)
6mls 50% w/v dextran sulphate
6mls TP5
3ml water
1.74g NaCl

TP5
1.0% w/v poly vinyl pyrolidine
1.0% w/v bovine serum albumin
1.0% w/v ficoll (Mwt 400,000)
0.5% w/v sodium pyrophosphate
5.0% w/v SDS
250mM Tris-HCl pH7.5

Virion RNA was resuspended from ethanol precipitates in TE pH 7.5. This was mixed with bromophenol blue loading dye and then run on 0.8% agarose 1X TBE, 19cm x 20cm gels, run at a constant 50V at 4°C overnight. The RNA was then transferred onto a Hybond N (Amersham) nylon membrane, by electrophoresis in a Hoefer electroblotter. Transfer was carried out overnight at 4°C in sodium phosphate transfer buffer with the powerpack set at 200mA. The membrane was recovered and the RNA cross linked to it using a UV cross linker (Stratagene) on the auto cross link function. The membrane was then washed 3 times in fresh sodium phosphate buffer (15 minutes for each wash), followed by one 5 minute wash in 1% w/v SDS. The membrane was then incubated for 4 hours at 42°C in a rotary hybridiser in 10mls of hybridisation buffer. A freshly prepared random primed 32P labelled DNA probe was denatured for 5 minutes in a 100°C waterbath, then immediately place on ice. The denatured probe was added to the membrane in a volume of hybridisation buffer to give at least 1x10^6 cpm / ml. Hybridisation was carried out overnight at 42°C in a rotating hybridisation oven. The probe solution was then removed and the membrane washed, initially in 2xSSC/1% SDS for 5 minutes at room temperature, then at 42°C for 30 minutes in 2xSSC/1% SDS. The membrane was then washed at 65°C for 30 minutes, then one final high stringency wash in 0.2X SSC/1% SDS for 15 minutes. The filter was then allowed to air dry, it was covered in cling film and the blot was visualised by autoradiography.

2.2.7 Random Primed DNA probes
Radioactively labelled DNA probes were made from full length HIV-1 genomic DNA. This was presented as two fragments of a SacI digest from an infectious molecular clone of HIV-1 NL43 as template in the labeling reaction, prepared using the
'Radprime' system (Gibco) according to the manufacturers instructions. To assess the labelling efficiency and the probe quality as well as to remove the unincorporated counts, the reaction mix was purified by gel filtration chromatography. A Pasteur pipette was drawn out to form a fine capillary, and the end was blocked with polyalomer wool. The pipette was filled with Sephadex G100/TE pH 7.5 slurry and the column was washed several times with TE. The crude reaction was then applied to the top of the column together with 150μl TE buffer. Fractions (typically 15) were collected in eppendorf tubes. The activity in each tube was assessed using a hand held Geiger counter, and additional aliquots of 150μl TE were applied to the column. The first high activity fractions typically up to tube 6 were used as the probe. This probe is made up of the largest random primed fragments which are unable to penetrate the pores in the resin. It is these fractions that hybridise best to the RNA. Shown are the typical activities of the fractions in a high activity probe and a low activity probe (figure 2.1).

Figure 2.1

![Random Prime DNA Probe Preparations](image)

**Figure 2.1** Graph illustrating the distribution of radioactivity distribution in two different probe preparations.

### 2.2.8 RNA Filter Binding Assay

Binding reactions were set up on ice. A 30μl reaction contained approximately 0.1nM (20,000cpm) RNA that had been allowed to renature during a slow cool in a glass beaker from 90°C to room temperature. The 1 times reaction binding buffer was composed of 30mM Tris -HCl pH 7.1, 30mM KCl, 2mM MgCl₂, 10mM ZnCl₂. The
reactions were prepared by adding 20u RNasin (or 1u Prime), 2µg sheared salmon sperm DNA, 0.8µg tRNA, 5mM DTT and 0-137nM GST-p55 protein to the RNA. Reactions were incubated on ice for 15 minutes. To analyse RNA bound to GST-p55, the reactions were applied under suction to a prewetted nitrocellulose filter (HAWP type- Millipore), with a 0.45µm pore size. The filters were washed with 1ml binding buffer 3 times. Filters were allowed to air dry and were then placed in a scintillation vial. 4mls of Scintillation fluid was added to the vial and the radioactivity that was retained on the filter was quantitated by liquid scintillation counting. Reaction series were generally compared to RNAs H1 and H4 as positive control and C1 as negative controls.

2.2.9 Ribonuclease Footprinting of RNA-Protein interactions

32P 5' end labelled RNA that had been gel purified, 50000 cpm, was heated to 90°C and then allowed to slowly cool to room temperature in 30mM Tris-HCl pH 7.1, 30mM KCl, 2mM MgCl2, 10µM ZnCl2. The RNA (0.63nM) was then incubated on ice for 15 minutes in the presence of 1µg yeast tRNA, 5mM DTT, and the presence or absence of 171nM GST-p55 in a total volume of 12µl. Ribonucleases V1 or T1 were then added and the reactions incubated at 37°C for 10 minutes. Reactions were stopped by the addition of formamide loading dye. The samples were analysed on a 12% polyacrylamide, 7M urea gel. Gels were fixed, dried and subject to autoradiography and phosphorimage analysis. RNA fragments were sized by running RNA sizing ladders produced by heating a small sample of RNA (20000cpm) for 1 minute in 50mM NaCO3, 1mM EDTA. The reactions were stopped by snap freezing on dry ice.

2.2.10 3' Boundary analysis of H4 RNA / p55-GST interaction

Cold H4 RNA was 5' kinased with 32P 1x10⁶ cpm of the gel purified RNA was then hydrolysed to make an RNA ladder as previously described. The RNA was recovered by ethanol precipitation and resuspended in 50µl 1X binding Buffer. The RNA was then folded by a slow cool from 90°C to room temperature. The RNA was then incubated in a filter binding type reaction, containing in 1X Binding Buffer, 0.8µg tRNA, 5mM DTT, 2units 'Prime' RNAse inhibitor, and either 0, 68.5 or 137nM p55-GST, with 50000 cpm RNA ladder. Binding was allowed to proceed for 20 minutes. The reaction mixtures were then applied to prewetted nitrocellulose filters, and processed as for a filter binding reaction. The RNA was eluted from the filter by boiling in formamide loading dye. This was then loaded onto a 12% polyacrylamide gel, 7M urea gel.
2.2.11 UV- Cross linking RNA-Protein Interactions
Lysate produced by Dignam extraction or NP-40 lysis (both nuclear and cytoplasmic fractions were kept on ice and divided into 2. Ribonuclease inhibitor (Prime) was added and mixed to half of each fraction. Lysate for cross linking was placed in a well of a Nunclon 24 well tissue culture dish. RNA-protein contacts were then cross linked by placing the dish in a Stratagene UV stratalinker 2200, set to 10J/cm². After cross linking the lysate was removed and the RNA was purified on Qiagen RNeasy spin columns using the cytoplasmic RNA purification protocol.

2.2.12 Identification of Gag protein in RNP complexes
RNA from cross linked samples was ethanol precipitated and resuspended in 10μl TE pH7.5. 2μl of a 10μg/μl stock of RNAse A was then added and the RNA digested at 37°C for 90 minutes. The sample was then boiled in SDS buffer and run on a 12% SDS PAGE gel. Gag protein was then identified by Western blotting and ECL detection. RNA from UV irradiated extracts was then compared to RNA that had been extracted from the non-irradiated extracts.

2.2.13 Sucrose gradient purification of membranes
Cytoplasmic fractions were further analysed by sucrose density centrifugation by the following method. Infected cells were washed with PBS and then allowed to swell for 15 minutes in ice cold 10mM Tris pH7.6, 1mM EDTA. The cells were homogenised using the mechanical homogeniser previously described, then the nuclei were removed by centrifugation at 13000 rpm for 5 minutes in a microcentrifuge. The lysate was then split into two, and one half was UV irradiated as for the cytoplasmic and nuclear fractions. Each fraction was then made up to 80% w/v sucrose, in NTE buffer (100mM NaCl, 10mM Tris pH 7.4, 1mM EDTA) and placed in the bottom of a centrifuge tube in a final volume of 3 mls. This was then overlain with 9mls of 65% w/v sucrose in NTE, and the tubes were then topped up with 10% w/v sucrose in NTE to form a 80%, 65%, 10% step gradient. The tubes were spun to equilibrium during an 18 hour spin at 100,000g in a Sorvall Ultraspeed centrifuge. 1ml fractions from the gradient were collected from the top of the tube by gilson pipette and samples of these fractions were weighed to create a profile of the sucrose density down the tube. The fractions containing the visible membrane at the 10% / 65% sucrose interface were diluted 10 fold with NTE buffer and the membrane was pelleted during a 30 minute spin at 100,000g. The membrane pellet was resuspended in 175μl NTE containing 0.1% NP-40, and the RNA was extracted on Qiagen spin columns using the cytoplasmic RNA extraction protocol. The rest of the fractions containing p55 gag and cellular proteins were ethanol precipitated by the addition of 10X volume of 100% ice cold ethanol and storage at -80°C for 30 minutes. Protein and nucleic acid was
recovered by centrifugation and the pellet resuspended in NTE/0.01% NP-40. RNA was then extracted as for the membrane pellet. Gag bound RNA was analysed as previously described for the nuclear and extranuclear fractions.

### 2.2.14 cDNA Synthesis

To identify RNA species by PCR, RT-PCR was carried out. RNA purified by phenol extraction of immunoprecipitates was reverse transcribed. This was carried out using random hexamer primers and AMV reverse transcriptase. The following reactions were typically assembled in a 0.5ml eppendorf.

**1X reaction**

- 1mM dNTP mix
- 500ng random hexamers
- 1-5μg total RNA
- 1 unit Prime RNAs inhibitor
- 2.5 units AMV reverse transcriptase (Promega)
- In 30μl 1X RT buffer

RNA and primers were heated to 70°C for 5 minutes, then cooled on ice. The reaction mix was then assembled and incubated at 42°C for 1hour. From this reaction 1μl was PCR amplified to identify spliced and unspliced HIV RNAs.

### 2.2.15 Immunoprecipitation of gag-RNA complexes

To confirm that gag proteins interact with HIV-1 genomic RNA in the infected cell and to show that this is probably what is being cross linked to in the UV-cross linking experiments, p55 was immunoprecipitated from an infected cell lysate. RNA was phenol chloroform extracted from the immunoprecipitate and then ethanol precipitated. Briefly, 5x10^6 cells were lysed in 500μl ice cold NP-40 buffer and the nuclei removed by a 5 minute spin at 13000 rpm in a microcentrifuge. To the crude cytoplasmic lysate 100μl of partially purified (50% ammonium sulphate precipitated serum) p6 polyclonal antibody, with 20 units of 'Prime' RNAs inhibitor and 100μl 50%v/v protein A sepharose slurry was added. The mix was allowed to immune precipitate for 4 hours, then the beads were separated by a 20 second spin in the microcentrifuge. The supernatant was removed, phenol-chloroform extracted and ethanol precipitated to give a cytoplasmic RNA fraction. The protein A beads were washed 4X with 1ml of ice cold NP-40 lysis buffer and the RNA from the immune precipitate was then purified by a phenol-chloroform extraction and an ethanol precipitation.
2.3 Cell Culture and Virus techniques

Mammalian tissue culture cells used are listed below. Culture stocks originated from the MRC AIDS directed Program, Potters Bar, NIBSC

C8166 : Human T-Lymphoblastoid cell line. (Salahuddin et al 1983)

2.3.1 Cell Culture

Cultures were initiated from liquid Nitrogen frozen stocks. Cells were thawed at room temperature and added dropwise to excess media. Cells were pelleted, then resuspended in fresh media two more times to remove traces of DMSO.

Cell cultures were maintained under Category III laboratory conditions. They were grown in Nunclon tissue culture flasks at 37°C / 5% CO₂ in 1X RPMI 1640 (Gibco BRL) supplemented with 10%v/v heat inactivated Foetal calf serum, 100units/ml penicillin, 100mgs/ml streptomycin and 2mM glutamine, all from Gibco BRL.

2.3.2 Counting Cells

Cells were maintained at a viable cell density of between 0.2-1x10⁶/ml. Viable cells were counted by mixing 200μl of cells with 200μl Trypan blue solution (Sigma). The mix was then placed on a haemocytometer and the cell density worked out by counting the cell number, in 4 quadrants. To calculate the number per ml this number was then multiplied by 5000.

2.3.3 Storing Cells

Periodically fresh cells were frozen down to make cell stocks. Cells were pelleted at 1250rpm for 5 mins in a mistral 3000i centrifuge, washed once in PBS, then pelleted again. The cell pellet was then resuspended in 1ml of freezing media, at room temperature, (fresh media containing 20% FCS, and 10% DMSO). The cells were placed in a Biofreeze (Costar) vial and were frozen slowly in a foam box, on dry ice overnight. These were then placed in liquid nitrogen for long term storage.

2.3.4 Handling Virus

Virus work was carried out in the category III containment facility in the department of Microbiology and Immunology at the University of Leicester, under the correct safety conditions (as detailed in the Department of Microbiology and Immunology safety manual p46-50).
2.3.5 Virus Stocks

Virus stocks were made from the electroporation of a circularly permuted HIV-1 plasmid. The cells were placed in fresh tissue culture medium 24 hours prior to electroporation and the cell density was adjusted to 5x10^5/ml. Immediately before electroporation the cells were washed with 1x PBS then pelleted and resuspended at a cell density of 8.0x10^6/ml. 250μl of cells were then placed in an ice cold electroporation cuvette, along with 25μg pMAJ-1 DNA (a molecular clone of HIV-1 NL4-3). The cuvette was left on ice for 10 minutes, then the cells were electroporated using a Biorad Gene Pulser set at 960μF/ 150mV. After 10 minutes on ice the cells were removed from the cuvette and added to fresh medium in a fresh tissue culture flask. 24 hours after electroporation the cell density was adjusted to 5x10^5/ml in fresh media. For a period of one week tissue culture supernatant samples were taken every day and assayed for p24. Tissue culture supernatants were frozen in liquid nitrogen generally as 1ml aliquots in 2ml Biofreeze vials (Costar).

2.3.6 Virus infection of C8166 cells

Virus stocks were removed from liquid nitrogen storage and allowed to slowly thaw to room temperature. Infections were initiated by adding infectious tissue culture supernatant containing approximately 5-20ng p24, (as detected by ELISA assay), to 1 ml of C8166 cells (5x10^6) in RPMI. Virus was allowed to infect by incubation at 37°C for 1-2 hours. The cells were then washed 2X in fresh media and then left at a concentration of 5x10^5/ml in RPMI at 37°C. The culture was inspected daily for cytopathic effects characteristic of HIV-1 infection, such as cell clumping, rosetting and the formation of large ballooning cytoplasms. The cell density was kept between 5-10x10^5 cells/ml.

2.3.7 HIV-1 p24 Elisa Assay

<table>
<thead>
<tr>
<th>TBS</th>
<th>Diluent</th>
<th>TMT-SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>144mM NaCl</td>
<td>TBS</td>
<td>2.0% Skimmed Milk</td>
</tr>
<tr>
<td>25mM Tris-HCl pH 7.5</td>
<td>0.1% Empigen</td>
<td>20% Sheep Serum</td>
</tr>
<tr>
<td></td>
<td>10% Sheep serum</td>
<td>0.5% Tween 20 in TBS</td>
</tr>
</tbody>
</table>

This method utilises a double antibody 'sandwich' for the quantitation of viral p24 capsid protein against external standards (Conley et al 1993). The capture antibody (D7320) was diluted from the stock (in water) to 1mg/ml in 100mM NaHCO3, pH 8.5. This antibody was then added to the wells of a microtitre plate and left overnight at room temperature to adsorb. The wells of the plate were then washed with 200μl TBS,
three times. The wells were then blocked by the addition of 200μl of 2% skimmed milk powder in TBS then incubated for a further 30 minutes. The wells were again washed three times with 200μl TBS. The culture supernatants that contained virus were inactivated by the addition of Empigen (Calbiochem), to a final concentration of 1% and incubation for a further 30 minutes at 56°C. Dilutions were then made of the supernatant in diluent. p24 standards were prepared using antigen standards from the MRC ADP reagent repository. These were prepared using the stock solution at 10μgs/ml. p24 standard was then diluted to give 3.1, 1.0, 0.31, 0.1 and 0.003ng /ml standard in diluent. Duplicate 100μl aliquots were loaded into the wells of a microtitre plate. Diluent was used as a negative control. 100μl of the diluted samples containing the culture supernatant was also added to the wells of the microtitre plate and the antigen capture was allowed to proceed for 3 hours at 30°C. The microtitre plate was then washed three times with 200μl TBS. The detection antibody (EH12E1-AP) was used at a concentration of 0.5μg/ml in TMT/SS, and 100μl of this solution was then added to each well. The plate was then incubated at 30°C for one hour. The wells were then washed with 1ml of AMPAK (Novo Biolabs) wash buffer. 50μl of AMPAK substrate was then added to each well and the plate incubated for a further hour at 30°C. 50μl of amplifier was then added, and the intensity of color generated in each well read on a plate counter by absorbance at 490nm.

2.3.8 Nuclei purification and extraction

NP-40 extraction

NP-40 lysis buffer

0.65 % Nonidet P40
150mM NaCl
10mM Tris-HCl pH7.8
1.5mM MgCl2

Infected C8166 cells were lysed by the addition of 100μl ice cold NP-40 buffer per 1x10^6 cells. Cells were left on ice for 5 minutes, then the cells passed through a Gilson pipette tip 10 times. The nuclei were then pelleted in a 5 minute spin in a microcentrifuge running at 13000rpm. The supernatant was removed and kept on ice as the cytoplasmic fraction. The nuclear fraction was washed a further 3 times with fresh NP-40 buffer, and the nuclear pellet retrieved by centrifugation at 13000rpm for 5 minutes in a microcentrifuge. The fractions were then boiled in SDS buffer and run on polyacrylamide gels for protein analysis.
Dignam Extraction

This method for the extraction of nuclei is as essentially described by Dignam et al (1983). PBS washed cells were resuspended in 5 cell volumes of buffer A. The cells were left for 10 minutes and then pelleted in a five minute spin at 1250rpm (Mistral). The cells were then resuspended in two cell volumes of buffer A. Due to the confined guidelines for working in a category III containment lab, a glass homogeniser could not be used for Dounce homogenisation. The cell suspension was placed in a 2ml screw capped vial. To homogenise these cells a 2ml syringe plunger (Becton Dickinson) was modified by cutting channels in the rubber seal to give four vertical grooves. This plunger was then placed in the tube and 10 strokes used to homogenise the sample. The homogenised cells were centrifuged and the cytoplasmic fraction (supernatant) removed. The nuclear pellet was then washed in buffer A and again pelleted. The nuclear pellet was then resuspended in a volume equal to 5 original cell volumes of buffer C. The nuclei were then homogenised as before. The homogenised nuclei were then incubated on a rotary mixer for 30 minutes at 4°C. The tube was spun in a microcentrifuge for 5 minutes at 13000rpm and the supernatant removed as the nuclear extract.

2.3.9 3H Myristic acid labeling of infected cells

To determine whether the state of p55 N-terminal acylation, has any effect on the nuclear localisation of p55 gag protein, infected cells were labelled with 3H 9,10-myristic acid (Amersham). 24 hours post infection 5x10⁶ cells were removed and resuspended in 5mls of 1X RPMI, 10% delipidated calf serum (Sigma) containing 50μCi / ml myristic acid. Myristic acid was dissolved in DMSO, to give a final concentration of 0.25% v/v DMSO and the cells were labelled for 16 or 36 hours. At harvest, cells were fractionated into nuclear and extranuclear fractions using the NP-40 lysis method. For determining the extent of myristylation of nuclear gag proteins the fractions were run out on an SDS protein gel and western blotted using the alkaline phosphatase method of detection. p55 Gag bands were excised from the nitrocellulose membrane and added to 4mls of scintillation fluid, the bands were left to dissolve in the fluid for one hour then the tubes counted by liquid scintillation. For the direct visualisation of myristylated proteins, the membrane was sprayed 3 times with 'Enhance' (NEN) and then exposed to X ray film for flourography at -80°C overnight.

2.4. Protein Analysis

2.4.1 SDS Polyacrylamide Gel Electrophoresis

Ten pairs of 100mm x100mm glass plates were stacked with 0.4mm spacers in the minigel casting apparatus. The resolving gel was poured into the apparatus, leaving 10mm for the stacking gel and the comb length. The resolving gel was covered with
water saturated isobutanol and then allowed to polymerise. The isobutanol was poured off and the gel combs inserted between the plates. The stacking gel was then poured on top of the resolving gel. Protein samples were prepared by boiling a sample in 1X SDS buffer for 3-4 minutes. The samples were loaded into the wells of the gel using a Hamilton syringe. The gels were then run at 15mA until the bromophenol blue dye front had reached the stack/resolving gel interface then the sample was run through the rest of the gel at 25mA. Proteins were sized by running the samples against Biorad protein molecular weight standards.

Biorad Protein molecular weight standards: 94kDa, 67kDa, 43kDa, 30kDa, 20kDa, 14kDa.

2.4.2 Coomassie Blue staining of Protein gels

<table>
<thead>
<tr>
<th>Coomassie Stain</th>
<th>Destain solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>200mls methanol</td>
<td>800mls methanol</td>
</tr>
<tr>
<td>40mls glacial acetic acid</td>
<td>160mls glacial acetic</td>
</tr>
<tr>
<td>200mls water</td>
<td>800mls water</td>
</tr>
<tr>
<td>1.1g coomassie brilliant blue.</td>
<td></td>
</tr>
</tbody>
</table>

Protein gels were visualised by coomassie staining. The gels were incubated in stain for 30 minutes then excess stain was washed away with destain solution until the gel background was again colourless.

2.4.3 Western Blotting of Protein Gels

Primary antibodies for use in western blots

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Epitope</th>
<th>Species</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARP 308</td>
<td>p55/p24 aas 355-377</td>
<td>Mouse (monoclonal)</td>
<td>MRC-ADP</td>
<td>Haaheim et al 1991</td>
</tr>
<tr>
<td>ARP 315</td>
<td>p55/p15 aas 121-134</td>
<td>Mouse (monoclonal)</td>
<td>MRC-ADP</td>
<td>Ferns et al 1989</td>
</tr>
<tr>
<td>ARP 332</td>
<td>gp120 aas 101-120</td>
<td>Mouse (monoclonal)</td>
<td>MRC-ADP</td>
<td>Thiriart et al 1989</td>
</tr>
</tbody>
</table>
Vinculin  NA  Mouse  Sigma  NA
(monoclonal)

Transfer Buffer (per l)  Ponceau Stain
3g Tris  0.2% Ponceau
14.4g glycine  7% acetic acid
200mls methanol

Following electrophoresis, gels were equilibrated in transfer buffer for 10 minutes. A sheet of Hybond C+ (Amersham) nitrocellulose paper was cut to the size of the gel and also incubated in the transfer buffer for 10 minutes. The transfer apparatus (Cambridge Electrophoresis) was assembled using Watmann 3MM paper. Three sheets of pre-soaked paper were used each side of the gel with the Hybond C+ paper in direct contact with the gel on the anode side of the gel. This ‘sandwich’ was then loaded into the electrophoresis tank and run at 20V overnight at 4°C, in transfer buffer. Blotted nitrocellulose paper was then stained with Ponceau stain for 5 minutes then washed with water once. Positions of the stained marker bands were marked with pencil and the membrane was washed thoroughly with water. Membranes were blocked with TMT-SS for 1 hour then diluted primary antibody in TMT-SS was added. The membrane was then incubated for a further hour with gentle rocking. After 3x 5 minute washes with PBS-T (PBS +0.1% Tween), the membrane was incubated with an appropriate dilution of secondary antibody (Sigma 1/12000, Serotech 1/1000) in TMT-SS for 1 hour. The membrane was extensively washed in PBS-T for 10 minutes, 3 times. A final wash then done in PBS for a further 10 minutes.

2.4.4 Enhanced Chemiluminescence Detection of Western blots
ECL reagents (Amersham) were used as substrate for the secondary horse radish peroxidase conjugated antibodies. Briefly reagent 2 was added in enough volume to cover the membrane. Then an equal volume of reagent 1 was added and the membrane incubated in this solution for 1 minute. Visualisation of blot was carried out by autoradiography, generally using multiple time points.
2.4.5 Alkaline Phosphatase Detection of Protein in Western Blots

NBT
Nitroblue tetrazolium (50mgs /ml) in 70% dimethyl formamide (DMF)

X-Phos
12.5 mg/ml 5-bromo-4chloro-3indoyl-phosphate (BCIP) in 50% DMF

Alkaline phosphatase buffer
100mM Tris-HCl pH9.5, 100mM NaCl, 5mM MgCl₂

Proteins were blotted onto Hybond C membrane (Amersham). The secondary antibodies were alkaline phosphatase conjugates (Dako). After the final washes the membrane was rinsed in alkaline phosphatase buffer. The membrane was then incubated with 10 mls of alkaline phosphatase buffer containing 44µl NBT, and 33µl X-phos. When the colour had developed, typically after 5-10 minutes, the membrane was washed in water and then dried.

2.5 Expression of HIV-1 p55 protein in E.coli

2.5.1 Protein expression

<table>
<thead>
<tr>
<th>E. coli strains</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topp II</td>
<td>F’, proAB, lac⁹Z15, TΔMn10, (tet⁵)</td>
</tr>
<tr>
<td>BL21</td>
<td>F-ompT hsdSB (rB⁻mB⁻)gal dcm (DE3)</td>
</tr>
<tr>
<td>B834</td>
<td>F-ompT hsdSB (rB⁻mB⁻)gal dcm met (DE3)</td>
</tr>
</tbody>
</table>

The HIV-1 nucleocapsid p55 protein was expressed in E. coli as a fusion to the C terminus of Glutathione-S-Transferase, using the plasmid pGEX-gag which is comprised of the entire gag coding region of HIV-1\textsubscript{NL4-3} gag cloned into the pGEX-1 vector at the Eco R1 and Bam H1 sites. When expressed in a wide variety of E.coli strains the protein produced is soluble but suffers from extensive C-terminal truncation. The GST-p55 protein was purified initially on a glutathione sepharose 4B (Pharmacia) column to enrich proteins containing the GST tag. As this protein is not pure, because protein truncates also existed in this preparation, the fusion protein was purified on an antibody column. The antibody column used antibody from polyclonal sera raised against a peptide derived from the p6 region of nucleocapsid protein to enrich for the intact protein as described in Maddison et al (1998). The GST-p55 preparations used in the filter binding and ribonuclease footprinting were prepared by Dr Pankaj Marya and were stored at -80°C until use.
To make p55 protein, the BamH I/Hind III fragment from the GST-gag construct was cut out and blunt ended with Klenow fragment from DNA polymerase 1. This fragment was ligated into the pET 16B vector also blunt ended with Klenow at the Bam HI and NcoI blunt end sites. When expressed in *E. coli* this protein should consist of p55 with the following N-terminal amino acids MDPIQGR. As detailed in the results section, expression of this protein from this construct resulted in both N and C terminal proteolysis. The lack of an N-terminal tag (GST) would have made purification of this protein very difficult without use of a N terminal antibody.

The intact Nucleocapsid protein was eventually purified by initially expressing the GST-p55 fusion from GST-Gag, and then purification of this protein on Glutathione sepharose 4B (Pharmacia) in a batch purification. The protein was then digested with Factor Xa and released protein was purified by immunoaffinity chromatography (see results).

*E. coli* Topp II cells containing the pGEX-Gag construct were grown in 2XYT media containing 100µg/ml ampicillin at 37°C in a shaking incubator. When the optical density of the culture reached OD$_{595}$ = 1.0 IPTG (Calbiochem) was added to a final concentration of 0.2mM and the cultures were incubated for a further 5 hours. The cells were then pelleted in a Sorval GS-3 rotor at 5000rpm for 10 minutes. Cell pellets were stored at -20°C until use.

### 2.5.2 Protein Extraction

GST-p55 is expressed as a soluble protein in *E. coli* (Maddison *et al* 1998), and was released from the cells by sonication. Cells were resuspended in ice cold PBS containing 2mM PMSF, 2µg/ml aprotinin, and 2mM benzamidine. Cells were sonicated in 3x15 second bursts with the sonicator set at an amplitude of 8µm. Insoluble material and unlysed cells were removed by centrifugation at 10000rpm for 10 minutes in a Sorval GS3 rotor. The clear supernatant was then removed and kept on ice.

### 2.5.3 Protein Purification

**Binding to Glutathione Sepharose 4B**

GST tagged protein in the cleared sonicate was then incubated with glutathione sepharose 4B beads (Pharmacia) in a batch purification using a 0.1 volume of beads which had been prewashed in PBS. Binding was allowed to take place for 30 minutes on a plate rocker, then the beads were recovered by centrifugation. The beads were washed 3 times in 10 mls of PBS then once in Factor Xa (FXa) cleavage buffer (50mM Tris-HCl pH7.5, 150mM NaCl, 1mM CaCl$_2$).
Factor Xa Cleavage

The cleavage of the p55 protein from the fusion was performed using FXa to a final concentration of 0.01 mgs/ml in 1ml for every 5mls of sonicate 1x FXa buffer containing 0.1% (w/v) SDS. The solution was periodically agitated to stop the beads from settling and after 3 hours the beads pelleted in 2500rpm spin for 10 minutes in the Mistral centrifuge. The supernatant was removed and protease inhibitors PMSF, benzamidine and aproteinin added to 2mM, 2µg/ml and 2mM respectively to inactivate the FXa. The supernatant contains the cleaved p55 protein. This was then subsequently mixed with the antibody column matrix for binding and immunoaffinity purification.

2.5.4 Antibody Purification of p55

Rabbit polyclonal antisera was raised to the C terminal nucleocapsid peptide KELYPLASLRSLSFGS as described by Maddison et al (1998). The antibodies were partially purified by a 50% ammonium sulphate precipitation. Briefly 2.5mls of serum was made up to 10 mls with PBS. The solution was placed in a beaker with a magnetic stirrer and 10 mls of saturated ammonium chloride was added slowly. The solution was left to stir overnight at 4°C and the precipitated proteins were then recovered by a 20 minute spin at 20000rpm in a Sorval ss34 rotor. The precipitate was then resuspended in 750µl PBS and placed in a dialysis bag. The precipitate was then extensively dialysed against 3x300ml volumes of PBS overnight. The concentration of the protein solution then estimated by absorbance at 260 and 280nm using the following equation:

\[(1.55 \times A_{280}) - (0.76 \times A_{260}) = \text{Protein concentration in mg/ml.} \ (\text{Harlow and Lane})\]

The antibody in the protein precipitate was then bound to protein A sepharose CL4B (Sigma). The beads were washed with sterile nanopure water on a sintered glass filter. 20mgs of antibody per ml gel volume was added, and the slurry incubated mixing at room temperature for 3 hours. The beads were then washed twice in 10mls 0.2M sodium borate pH 9.0. The beads were resuspended in 10mls of sodium borate and solid dimethylpimelimidate powder was added to a final concentration to 20mM. The antibody was cross linked to the protein A during a 30 minute incubation at room temperature on a rotary mixer. The coupling reaction was stopped by washing the beads in 0.2M glycine pH 8.0, then incubation for two hours at room temperature. The matrix was recovered by centrifugation at 2500rpm (Mistral MSE centrifuge) for 5 minutes then washed twice in PBS. The matrix was stored at 4°C until use.
2.5.5 Antigen Binding and Elution

<table>
<thead>
<tr>
<th>Protein Binding Buffer</th>
<th>Wash Buffer A</th>
<th>Wash Buffer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>40% v/v glycerol</td>
<td>25mM Tris pH 7.5</td>
<td>10mM Tris pH 6.5</td>
</tr>
<tr>
<td>300mM NaCl</td>
<td>400mM NaCl</td>
<td>200mM NaCl</td>
</tr>
<tr>
<td>100mM KCl</td>
<td>50mM KCl</td>
<td></td>
</tr>
</tbody>
</table>

The protein sample was mixed with an equal volume of protein binding buffer and then added to the antibody matrix. The antigen was allowed to bind to the antibody during a two hour incubation on a plate rocker at room temperature. This mix was then added to a disposable 10ml column and the flow through collected. The flow through was then used to rinse out the tube that contained the binding reaction. This was also applied to the column. This process was repeated until all the protein A beads had been transferred to the column. The column was washed with 10 times the volume of the beads wash buffer A. This was followed by five times the bead volume with buffer B. The protein was then eluted from the antibody column by successive additions of 0.2M glycine pH 2.5. The eluted protein was collected in eppendorfs containing 200μl 1.5M Tris pH 9.5 for every 1ml of eluate. The approximate concentration of protein in each of the fractions was estimated by measurement of absorbance of the protein mix at 280nm and 260nm.

2.5.6 Expression of p55 from pTYB4-p55 in E.coli

Following the attempts at purification of p55 from its GST fusion by Factor Xa cleavage, the p55 sequence was amplified by PCR using primers BM14 and MH12 using the pGEX-gag vector as template. This fragment was restricted with EcoR I and Nco I and ligated into the IMPACT T7 (NEB) intein fusion system at the EcoR I and Nco I sites of the construct pTYB4. The resulting plasmid, pBM 99.001 (pTYB4-p55) was then transformed into BL21 and B834 E. coli cells. p55 inductions were carried out by the addition of IPTG.

The p55 protein expressed from this construct was immunoaffinity purified. The crude sonicate was bound to the antibody matrix (3mls matrix + 5mls sonicate) in 20mM Tris pH7.5, 50mM KCl, 40mM NaCl and 1% Triton X-100. This was mixed overnight at 4°C. Once applied to the column the matrix was washed with 300mls of 0.1M Tris pH 7.5, 0.14M NaCl and 0.5% Triton X-100. This was then washed with 50mls of 50mM Tris 7.9, 0.5M NaCl. The p55 protein was eluted from the column in 0.2M glycine pH 2.5. Fractions were collected and their absorbance at A260 and A280 measured.
2.5.7 N-Terminal sequencing of p55 Gag

The protein purified from the Impact vector system was N-terminally sequenced to verify its authenticity. Briefly a 20μl aliquot of the protein was loaded onto a 12% SDS PAGE gel in 1X SDS loading buffer and run as for a standard protein gel. The gel was equilibrated in 1X transfer buffer (48mM Tris, 39mM glycine, 10% methanol, 0.03% SDS) for 10 minutes. Protein was transferred to a PVDF membrane (Biorad), which had been soaked in 100% methanol for 5 minutes, by electroblotting at 50V for 2 hours at 4°C, in 1X transfer buffer. Protein was visualised on the PVDF membrane by staining with Coomassie Blue stain (0.1%w/v coomassie brilliant blue, in 50% v/v methanol / water) for 5 minutes at room temperature. The blot was then destained for 10 minutes by the addition destain solution, (50% methanol, 10% acetic acid, 40% water). The protein band was carefully excised using a scalpel, and given to the protein and nucleic acids chemistry laboratories at Leicester University for N-terminal sequencing.
Chapter 3

Kissing Loop Mediated Dimerisation of Genomic RNA in HIV-1

3.1 Introduction

Current views on the mechanisms of RNA dimerisation are summarised in the introduction. A kissing loop mechanism for dimerisation has been shown to occur in-vitro. Studies in-vivo using virus with mutations in the autocomplementary loop sequence on SL1 have unequivocally shown the importance of SL1 in RNA dimerisation. SL1 in vivo probably has a multifunctional role, in that mutations in SL1 affect packaging (Clever et al 1997), and first strand transfer (Paillart et al 1997). SL1 interactions may also just be the initial or primary interactions that occur in the dimerisation of the genomic RNA, and without these the dimerisation of genomic RNA may not happen.

There is therefore, direct evidence by use of in-vitro models that a kissing loop interaction facilitates RNA dimerisation in transcripts containing the SL1 autocomplementary sequence. Indirect evidence from the mutagenesis of molecular clones also shows that this structure is important for genomic RNA dimerisation.

It would therefore be of use to investigate the role of SL1 mediated dimerisation, biochemically, analysing mature genomic RNA dimers extracted from virions. This would give direct information as to the significance of this structure in the mature RNA dimer. It may be argued that the kissing loop dimer, existing in whichever conformation, either open or extended (see figure 1.10), contributes to dimer stability in the HIV-1 genome. This might be suggested by the fact that virion RNA dimers isolated from viruses mutated in their autocomplementary sequence show much lower amounts of dimeric RNA (Haddrick et al 1996). The development of an in-vitro assay that is able to specifically disrupt pre-formed RNA dimers at the DLS would be useful in determining if this site of dimerisation contributes to the stability or maintenance of the mature dimeric genome.
3.2 Mechanisms of in-vitro RNA dimerisation

The dimerisation of RNAs containing the dimer initiation site (DIS) nucleotides in the context of SL1 (nucleotides 247-271) was studied using the following RNAs, transcribed in-vitro by run off transcription using T7 RNA polymerase from the following plasmids.

<table>
<thead>
<tr>
<th>RNA</th>
<th>Plasmid</th>
<th>RNA transcribed (vRNA sequence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA 1</td>
<td>pMH94.004</td>
<td>195-359</td>
</tr>
<tr>
<td>RNA 2</td>
<td>pMH95.001</td>
<td>195-359 SL1 mutation</td>
</tr>
<tr>
<td>RNA 3</td>
<td>pMH93.010</td>
<td>195-404</td>
</tr>
</tbody>
</table>

**Figure 3.1** shows diagrammatically the in-vitro transcribed RNAs used in this chapter, based on a sequence from the RNA leader of HIV-1NL43 shown. Depicted are those regions that are expected to form regions of secondary structure namely the small stem loops SL1, SL2, SL3 and SL4.

**Phenol assisted Dimerisation of RNA Transcripts**

It is clear that two different dimeric structures are obtainable using these RNAs and this is due to the formation of one of two possible conformations at the dimer initiation site (see introduction figure 1.10). In the gel shown overleaf (fig 3.2) the bands corresponding to the dimeric RNA have dimerised during phenol extraction from the in-vitro transcription reaction, and probably represent the extended complex, whereby there is extended annealing between the stems of each SL1 stem loop structure (see fig 1.10). These RNAs are stable in the presence of TBE buffer and do not require magnesium for the maintenance of the dimer structure (Haddrick et al 1996).
Figure 3.2 An ethidium bromide stained agarose gel showing, Lane 1 = φX174 HaeIII marker, positions of the dimer and monomer RNA bands are shown, lanes 2+3 are RNA 2, lanes 4+5 are RNA 1. Mutated autocomplementary sequence RNAs can still dimerise see lane 2, although not as readily as wild type (lane 4).

The gel in figure 3.2 shows that RNA 1 dimerises readily during phenol extraction, presumably forming the extended linear RNA that is stable on a Tris-borate EDTA gel (figure 3.3). RNA 2 dimerised less readily under these conditions, the dimers formed being between the complementary stem regions (see fig 3.3). These RNAs with mutations in the autocomplementary sequence do not spontaneously dimerise, these dimers must be formed during phenol extraction, maybe due to increased localised concentration of RNA during phenol extraction. Higher oligomeric structures seen with RNA 2 (lane 2) must have formed by annealing between the stem regions in each RNA and would suggest that these phenol induced dimers do not initiate via a kissing loop interaction. These TBE gel stable dimers are referred to as phenol dimers and have previously been described (Haddrick et al 1996). These phenol dimers must have formed the stable dimeric structure, of the extended conformation at the dimer linkage site. The phenol induced dimer (or phenol dimer) is a model for the virion RNA DLS extended conformation.
Figure 3.3 RNA dimer structures Highly schematic diagram, showing two RNA SL1 monomers (1), the kissing loop interaction (2), the linear extended form (3), and the expected dimeric form between two RNAs with mutations in the autocomplementary sequence (4).

**Kissing Loop Mediated Dimerisation**

These RNAs can also dimerise by kissing loop mediated dimerisation depicted in figure 3.3.2. In this instance the palindromic nature of the stem loops causes annealing interactions between each RNA monomer. These structures are not stable on TBE gels and require the presence of magnesium in the gel for the dimer to stay together. This requirement for magnesium suggests that it is the metal ion that is responsible for maintenance of the secondary structure in the RNA molecules and / or magnesium ions are required at the dimer interaction interface. Figure 3.4A shows the effect of running a kissing loop dimer RNA on a Tris borate EDTA gel, and in figure 3.4B shows this same RNA run on a Tris -Borate 0.1mM MgCl2 agarose gel.
Figure 3.4

Figure 3.4 A shows RNA 1 run on an ethidium bromide stained agarose - Tris borate EDTA gel. Lane 1 = φX174 HaeIII marker, lane 2 Denatured RNA, lane 3 Native RNA (in water, not subjected to dimerisation in dimer buffer) and lane 4 kissing loop dimerised RNA. In 3.4B RNA 1 has been run on a Tris-borate gel containing 0.1mM MgCl₂, lane 1 = φX174 Hae III, Lane 2 Denatured RNA 1, lane 3 native RNA (in water) and in lane 4 kissing loop dimerised RNA.

Gel A shows that Tris borate EDTA gels are not suitable for the analysis of kissing loop dimers. It is clear (lane 4) that RNA that should have dimerised efficiently shows poor amounts of dimeric RNA on the gel. The bands at the monomer position are diffuse and smear up the gel, this is probably indicative of RNA dissociation in the gel dimer dissociation may be a reason for the apparent shift in monomer position of the RNA in lane 4.

Gel B clearly illustrates that the kissing loop dimers have to be analysed on gels containing magnesium, in order to prevent the dimer dissociation. The presence of magnesium in this agarose gel actually promoted the dimerisation of denatured RNA.

This gel associated dimerisation was reduced by lowering the gel magnesium concentration still further to 0.01mM. This is shown in figure 3.5.
This gel shows that an agarose gel containing 0.01mM MgCl₂ does not induce dimerisation of monomeric RNA (compare lane 2 above with lane 2 in figure 3.4B). RNA that is induced to dimerise by incubation in dimerisation buffer runs at the dimeric position, indicating that 0.01mM MgCl₂ is sufficient to maintain the dimeric structure.

3.3 Development of an RNA Dimerisation Competition Assay

From previously published work it is reasonable to suggest that the kissing loop dimer, in which ever conformation, either open or extended, contributes to dimer stability in the HIV-1 genome as kissing loop mutated virions have reduced amounts of dimeric RNA (Haddrick et al. 1996). The deletion or substitution of the residues at the tip of SL1 (the kissing contact region), causes a drastic reduction in dimeric RNA in virus produced from infectious molecular clones. The development of an in-vitro competition assay, that specifically disrupts pre-formed RNA dimers at the DIS would be useful in determining if this site of dimerisation contributes to the stability or maintenance of the mature virion dimeric genome, when used on purified virion genomic RNA. Competition experiments were carried out using various oligonucleotides that are antisense to the kissing loop, to analyse the effects these may have on the two versions of dimer that can form with SL1. The competition strategy assumes that during the course of the experiment the RNA dimers are in an equilibrium between association and dissociation, and that during
dissociation a competitive inhibitor of dimerisation may anneal to the autocomplementary sequence.

In the following experiments kissing loop dimer RNA-i was made by T7 run off transcription as described in materials and methods. RNA dimer concentration was kept at 4μM for each experiment. Table 3.1 shows the oligonucleotides used and gives their description, figure 3.6 shows the site of annealing on the SL1 sequence.

Table 3.1

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
<th>Site of annealing</th>
<th>Competitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>MH 14</td>
<td>GGTACATCAGCCCATATCACC</td>
<td>761-781</td>
<td>-</td>
</tr>
<tr>
<td>MH 20</td>
<td>GCTGTGCAGCCTCAAGCAAGCC</td>
<td>246-263</td>
<td>+</td>
</tr>
<tr>
<td>BM 4</td>
<td>TGGCGGCTT</td>
<td>255-263</td>
<td>?</td>
</tr>
<tr>
<td>BM 5</td>
<td>CTTGCCGTGCGCGCTTCAGCAAG</td>
<td>248-270</td>
<td>+</td>
</tr>
<tr>
<td>RNA-i</td>
<td>GCCGUUGCAGCGCUU</td>
<td>255-267</td>
<td>+</td>
</tr>
<tr>
<td>BM7</td>
<td>CTTGCCGTTAGAGAGTCAGCAAG</td>
<td>248-270</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 3.6

Figure 3.6. A diagrammatic representation of SL1 is drawn in A and B. Shown around SL1 are the sites the above oligonucleotides would be expected to anneal. ***** represents the nucleotides in the BM7 oligonucleotide that are not expected to interact with the RNA loop sequence, in SL1 (A), B shows SL1 with oligonucleotides expected to compete.
Kissing loop dimer competition

MH 14 is a gag derived oligonucleotide sequence (not expected to anneal to SL1) and was used to control for non specific competition. A kissing loop dimer of RNA 1 (figure 3.3.2) was incubated for 15 minutes at 37°C with an excess of each oligonucleotide in 1x cacodylate buffer.

This is a specific interaction, RNA dimers are not dissociated by the addition of an oligonucleotide of essentially irrelevant sequence (MH14), see fig 3.7A. Figure 3.7B shows that the dimer of RNA 1 was successfully competed out by the addition of oligonucleotide MH20 that contains an antisense sequence for the SL1 5' stem and the kissing loop region.

Figure 3.7

Figure 3.7 A Shows an ethidium bromide stained agarose gel made with TBM buffer loaded with, φX174 HaeIII marker (lane 1), then RNA dimer of RNA 3 incubated with an increasing molar ratio of competing oligonucleotide MH14 (a control oligonucleotide), as indicated.
To test other oligonucleotides against the preformed kissing loop dimer, these experiments were analysed on polyacrylamide gels, using RNA dimer that had been spiked with approximately 1% w/w 32P-end-labeled RNA of the same species. This allowed both autoradiography and phosphor image analysis. In these experiments a constant ratio of RNA dimer to competing oligonucleotide was incubated for increasing length of time at 37°C. The reactions were snap frozen on dry ice then run after the addition of glycerol to 30% v/v on a Tris-borate polyacrylamide gel containing 0.1mM magnesium chloride. Dried gels were then visualised by autoradiography. Phosphor image analysis was carried out using a Molecular Dynamics phosphor imager using Imagequant™ software. The relative percentage of dimer and monomer was calculated per lane. Figure 3.8 shows these results represented graphically.
Figure 3.8

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure3.8.png}
\caption{Competition for Kissing Loop Dimer by Competing Oligonucleotides}
\label{fig:figure3.8}
\end{figure}

This data shows that the control experiment using an oligonucleotide of gag derived sequence did not destabilise the preformed kissing loop dimer. In comparison those oligonucleotides that contained antisense SL1 sequences did cause dissociation of the preformed dimer. The longer of the oligonucleotides tested have a greater destabilising effect than those that are shorter (compare the data for the oligonucleotides BM4 and BM5). This is presumably because of the greater stability the longer oligonucleotide is able to confer when it is part of an RNA-DNA hybrid. This also would have the effect of destabilising the hairpin into a linear conformation. Destabilisation of this secondary structure would have the effect of removing the RNA from the monomer-dimer equilibrium altogether. Although not compared with a DNA equivalent, the small RNA oligonucleotide (RNA-i) was also able to interfere with the dimer structure. The BM5 oligonucleotide was tested at two concentrations in this experiment, both outcompete the kissing loop dimer by 5 hours, these results were reproduced on two occasions. The oligonucleotides broadly acted as to how they were predicted to behave based on their sequence (table 3.1)

**Phenol dimer competition**

As HIV-1 genomic RNA is TBE gel stable when analysed on agarose gels and visualised by northern blotting, one might regard the dimeric structure at the DLS to be of the
extended conformation, rather than just a loop-loop kissing interaction in the mature virion RNA. Experiments were therefore carried out to test the competition ability of the best competing oligonucleotide with the kissing loop dimer (BM5), on dimers of RNA 3, representing the extended conformation (phenol dimers). Figure 3.9 shows the result of a competition experiment using a phenol dimer of RNA 3, with oligonucleotide BM5 at a 5X molar excess to dimeric RNA. The experiments were performed at 37°C and 50°C (just below the melting temperature for extended complex), and reactions were again removed over time. Reactions were run on 1XTBE 6% polyacrylamide gels and the change in dimer to monomer monitored by phosphor image analysis of the dried gel.

**Figure 3.9**

![Competition with Phenol Induced Dimer at Elevated Temperatures](image)

*Figure 3.9* Shows the effect of an oligonucleotide competition experiment using BM5 as the competing oligo on a phenol induced dimer of RNA 3. This graph shows that even under elevated temperatures close to the Tm of the RNA dimer, the phenol RNA cannot be efficiently out competed by competitor oligonucleotides.
This shows that the extended phenol dimer cannot be efficiently competed by the BM5 oligonucleotide even under elevated temperature conditions, nearing that of the melting point of this phenol dimer. Conditions where the phenol dimer could be efficiently competed by the oligonucleotide were then investigated. The effect of formamide on the stability of the phenol dimer was investigated with a view to using this as a method of reducing the strength of the Watson-Crick base pairing between the RNA molecules in the dimer. Under conditions of stable but weakened dimers, the competition would again be tried. Figure 3.10 shows the effect of the formamide on the preformed dimer. In these experiments dimeric RNA was incubated for 15 minutes at 37°C, in 1X dimerisation buffer made up to the stated % v/v formamide.

**Figure 3.10**

![Figure 3.10](image)

*Figure 3.10* shows monomer and dimer phenol dimerised RNA at the stated concentration (v/v) of formamide.

Figure 3.10 shows that the phenol dimer is stable in formamide up to 40% v/v. This should have the effect of lowering the strength of the hydrogen bonds between the two RNAs.

At 40% formamide the competition was again tried using the BM5 oligonucleotide at 5X molar excess to dimeric RNA. Figure 3.11 shows that when the RNA dimer is incubated with BM5 under these conditions the dimer is melted during a 15 minute incubation when the competing oligonucleotide is at a 5X molar excess. The slight band shift of the monomer RNA when compared to control, presumably being due to the annealing of the DNA oligonucleotide.
Figure 3.11 shows an autoradiograph of phenol dimer RNA in 40% formamide / dimerisation buffer (lane 1), and when incubated with BM5 (lane 2). Evident is a slight band shift in the reaction + BM5, presumably due to annealing of the oligonucleotide.

Summary

Conditions were therefore optimised to target and interfere with each conformation of the RNA structure that could form the DLS at SL1. The kissing loop structure could be competed out, using oligonucleotides (e.g. BM5) in kissing loop dimer buffer at 37°C. The extended structure which may be expected in mature virion RNA can be outcompeted in 40% formamide using oligonucleotide BM5 at 37°C. If the dimeric structure at the SL1 DIS is important for the stability of the mature genomic RNA dimer, then the structure would be susceptible to competition by one of these two assays. Competition experiments were then carried out on extracted virion RNA to see if the oligo BM5 had any destabilising effect on the virion genome.

3.4 Competition Experiments with Extracted Virion Genomic RNA.

Preliminary experiments showed virion genomic RNA to be less stable than the phenol dimer RNA in the presence of formamide. Thus the stability of the genomic RNA was tested in formamide buffers of increasing concentration as had been done for the phenol dimer. The extracted virion RNA was precipitated and resuspended in 100mM NaCl, 50mM Tris 7.5 made up with increasing % v/v formamide. The RNA was incubated at 37°C for 1 hour, then run on 0.8% agarose 1X TBE gels. RNA was visualised by northern blot using a random primed 32P labeled DNA probe to the NL4.3 genomic sequence.
Figure 3.12 Shows denatured RNA (lane 1), native RNA (lane 2) Then increasing amounts of formamide in the reactions as indicated. The dimer and monomer positions of the genomic RNA are illustrated.

Figure 3.12 shows that the virion RNA dimer is stable up to 25% formamide (lane 7) Above this concentration and the amount of dimer shown on the Northern blot is rapidly decreased.

In the following experiments any kissing loop dimers of virion genomic RNA were competed by incubation in cacadylic acid buffer with oligonucleotide BM5 at a concentration of 5μM. This was carried out at 37°C for the time periods indicated on figure 3.13. The extended conformation at the DLS was competed for by incubation in 50mM Tris 7.5, 100mM NaCl plus deionised formamide at the concentration indicated and the BM5 oligo at 5μM. This concentration of oligonucleotide is estimated to be in extreme excess of the virion genomic RNA dimer present in these experiments.
Figure 3.13 shows that under conditions which result in dissociation of small, \textit{in-vitro} transcribed, kissing loop dimerised RNA have no effect on the virion RNA dimer, compare lanes 2 (native), with lane 3 after one hour. The competition in the presence of 20 and 25\% formamide also has no effect, compare the monomer + dimer bands of lanes 6 and 8 and those of 10 and 11. Noted in this experiment was a slight shift in mobility at the dimeric position of the RNA which was incubated with oligonucleotide, when compared to the dimeric RNA that has been incubated in the absence of competing oligonucleotide. The reasons for this slight increase in mobility are unclear.
These results show that genomic virion RNA cannot be dissociated by oligonucleotide directed to the SL1 kissing loop or extended structures after one hour. To investigate whether the competition needed a longer incubation time the experiment was carried out again increasing the competition time to up to 5 hours (figure 3.14).

Figure 3.14

Figure 3.14 Shows the competition experiment between BM5 and virion RNA as a northern blot. Denatured genomic HIV-1 RNA in lane 1, native RNA in lane 2, kissing loop competition lanes (3-7), the controls for the kissing loop competition (8-12), extended dimer competition (13-17), and the controls for the extended dimer competition (18-22).

This was a demanding experiment, during which crude RNA dimer was incubated for long time periods at 37°C. This has resulted in extended smearing of the loaded RNA. It is clear however that there is no decreased dimer or increased monomer over the time course of the experiment, as determined by eye. From these experiments it is clear that the dimeric RNA genome cannot be destabilised by oligonucleotides that are directed by antisense towards SL1. It is likely that the lane in the extreme far right of the gel (-BM5, 5 hours) has suffered from nuclease degradation during the 5 hour incubation at 37°C.
3.5 Summary

3.5.1 Kissing loop dimers of small dimeric RNAs can be dissociated by competing oligonucleotides

The preformed kissing loop dimer in small RNAs can be dissociated by using oligonucleotides targeted towards the kissing loop hairpin SL1. This interaction is sequence specific as oligonucleotides containing irrelevant sequence do not dissociate the dimer. Paillart et al (1996) showed that preformed dimeric RNA could be interfered with by a smaller sense RNA by a kissing loop interaction, which could be shown to be independent of stem sequence. In that study, smaller antisense oligonucleotides were used to out compete the kissing loop interaction. Skripkin et al (1996), published the results of an extensive study on the effects of small oligonucleotides targeted either by sense (small kissing loops) or antisense to a preformed in-vitro transcribed kissing loop dimer. This study found that in the case of sense oligonucleotides, an RNA sense loop is an absolute requirement, whereas the sequence and chemistry of the stem were unimportant, i.e. the kissing loop interaction was required. On the other hand, antisense oligonucleotides were found to be far more effective in their dimer destabilising properties and their non reversible effect thought to result from complete extension of the base pairing that had been initiated by the interactions at the loop interface. It is an intriguing possibility that this interaction could effectively be used as a target for therapy in HIV-1 infected patients. The loop sequence in SL1 is highly conserved, following only two consensus sequences. Virus lacking the ability to dimerise via SL1 are severely impaired in their ability to replicate in tissue culture, thus an antisense strategy may prove useful in preventing an RNA genome dimerising in the infected cell.

3.5.2 Dimers of extended conformation can be specifically dissociated in the presence of formamide

This work also demonstrated the ability of antisense oligonucleotides to compete with the extended phenol dimer RNA, under destabilising conditions. This dimer is a model for the SL1 dimer in its extended conformation. This structure may or may not be present in the mature virion RNA.

3.5.3 Genomic RNA is stable to conditions causing dissociation of small dimers

Neither of the conditions that interfered with the small transcripts had any destabilising effect on the extracted genomic RNA dimer. From this it can be reasoned that the genomic RNA dimer stability is not conferred by SL1 mediated dimerisation. Either this means that interactions at SL1 are not present, or that these are not required for genomic RNA stability. This would be due to the presence of further points of contact between each RNA molecule. To determine if interactions in SL1 are occurring in the mature
dimer a more direct chemical study is required, which could maybe involve chemical modification type experiments on extracted virion RNA.

### 3.5.4 Possible reasons for genomic RNA stability

The genomic RNA dimer stability was not sensitive to the competing oligonucleotide strategy, that dissociated smaller RNA transcripts. This may mean that further points of contact between the two RNAs keep the genomic dimer together. There are two lines of evidence for this, the first indirect evidence comes from what is seen when genomic RNA is analysed under both native and denatured conditions. There are many smaller subgenomic fragments that can be seen under denaturing conditions than RNA run native. These must be kept in the dimer by smaller, minor interactions between the RNA molecules such as those detected in Rous sarcoma virus (Lear et al. 1995). More direct evidence comes from the fact that kissing-loop mutants do actually contain some dimeric RNA (Haddrick et al. 1996) but at a reduced amount compared to wild type, suggesting dimerisation must be occurring at some other point in the genome. Hoglund et al. (1997), have further examined purified genomic RNA by electron microscopy and propose that the rU5 RNA hairpin may be interacting to form a dimeric structure (see figure 3.15).
Figure 3.15 Proposed points of contact between the two RNAs in the virion RNA dimer as proposed by Hoglund et al. (1997). Shown are the Tar hairpin, the primer binding site, and the stem loops SL3 and SL4.

At the moment there is still speculation as to the role of SL1 mediated RNA dimerisation in HIV-1 genomic RNA. It clearly has a major effect on dimerisation of the genomic RNA. Kissing loop mediated dimerisation may be the initial step in genome dimerisation which brings the two RNAs into such an orientation as to maximise the chances of further intramolecular interactions, such as those proposed for rU5. Maybe then, in the mature virus, p7 nucleocapsid flips the SL1 structure out into the extended conformation (Muriaux et al 1996), which is partly responsible for the increased Tm of the dimer. It is not known whether the extended structure plays any role in the genomic RNA dimer. From the point of view of the virus replication machinery, the presence of a very stable intermolecular structure at SL1 may dramatically reduce the efficiency of genome reverse transcription. In this context the kissing loop dimer would be the preferred structure at SL1 in the mature genomic RNA (Paillart et al 1996). However this may be overcome.
in-vivo by the action of p7 Nucleocapsid which is thought to work with the reverse transcriptase in destabilising RNA secondary structures (Klasens et al 1999). It is clear that these questions must be answered to gain a fuller understanding of the role of SL1 in genome dimerisation.
Chapter 4

In-Vitro Identification of RNA Binding sites for p55 Gag on the leader sequence of HIV-1 Genomic RNA

4.1 Introduction

The packaging of the HIV-1 RNA involves the recognition of the genomic RNA in a homogeneous mix of cellular RNAs by the structural polyprotein precursor, Gag. This interaction has to occur to initiate the packaging of an RNA genome into an assembling virion (see introduction). In the mature virion it is the p7 Nucleocapsid protein that associates with the genomic RNA. This protein has a non-specific RNA binding activity (Berkowitz and Goff 1994, Surovoy et al 1993) allowing the formation of the nucleocapsid core by the binding of protein to RNA throughout the genome. The fact that it is the p55 gag protein which is present in the immature virus particle suggests that it is this protein that must make initial contact with the viral RNA. Mutational analysis of the HIV-1 genome suggests that the RNA packaging signal, or psi (ψ) is located primarily in the 5' UTR of the viral RNA. This signal may include sequences 5' of the major splice donor, (Kaye and Lever 1996, Clever and Parslow 1997, Kim et al 1994, McBride and Panganiban 1997), and between this site and the gag start codon (Aldavoni et al 1996, Clavel et al 1990, Clever et al 1997, Lever et al 1989, Luban and Goff 1994). There is also some evidence that the packaging signal may also extend into the gag coding region (Luban and Goff 1994). Despite continued efforts a minimal binding site on the viral RNA for the p55 gag protein has not been described. Several groups have investigated the in-vitro binding properties of HIV-1 gag to RNA sequences from the 5' leader region. In each case specific binding to HIV-1 RNA has been observed in vitro. The following chapter details my studies on this interaction with the aim of finding a minimal RNA binding site on the HIV-1 RNA leader sequence for the p55 gag protein.

p55 gag was expressed as a fusion to Glutathione S-Transferase (GST) under conditions devised and optimised by my colleague Dr Pankaj Marya. This protein was shown to have affinity to the HIV-1 leader RNA, yet purified GST alone did not (figure 4.1A). This means that it is the p55 domain of the fusion protein that has affinity for RNA. The binding of GST-p55 was also shown to be specific for HIV-1 leader derived RNA as the protein showed no apparent binding to C1 RNA, a size matched control RNA derived from a human C1 sequence (figure 4.1B)
Another methodology for the analysis of RNA-protein interactions, the gel shift assay, was tried but was not routinely used due to its non-quantitative nature. Figure 4.1C shows the results of a gel shift experiment using GST-p55 and RNA H2 (HIV nucleotides 195-404) and GST-p55 with C2 RNA, (a size matched control RNA containing HIV-1 RRE sequence).

**Figure 4.1C**

This Gel shift experiment shows the binding characteristics of RNA H2 derived from the leader RNA sequence (see fig 4.2) and C2 (a size matched control), with GST-p55. HIV-1 leader derived RNA (H2) forms high molecular weight complexes that are retarded at the top of the gel, and some faster moving complexes (marked by arrows). C2 on the other hand has little interaction with GST-p55, and moves into the gel. (Figure taken from Maddison et al 1998).
The autoradiograph shows that high molecular weight complexes have formed which do not migrate into the gel in the presence of HIV-1 leader derived sequence. In the presence of C1 RNA there is no gel retardation and RNA migrates into the gel. This experiment suggests that either GST-p55 binds co-operatively to RNA or that free GST-p55 exists largely in an aggregated state, or maybe aggregated p55 has a preferential affinity for RNA.

4.2 GST-p55 Protein Purification

GST-p55 was expressed and purified from Topp II cells (Stratagene) as described in materials and methods. Figure 4.2 shows the protein at steps in the purification.

Figure 4.2

Expression in *E. coli* Topp II and a variety of protease deficient strains (not shown) results in substantial C-terminal truncation of the fusion protein, as evidenced by those smaller proteins that purify on the Glutathione Sepharose 4B (GS4B) matrix (lane 1). Immunoaffinity purification of the GS4B purified protein substantially purifies the GST-p55 protein as the major product (lane 2). However proteins just smaller than GST-p55 co-purify with the major product. As these proteins don't react with C terminal p6 antibody (lane 3), they must co-purify by protein-protein interactions with the full length fusion protein. These small amounts of contaminating protein were not expected to participate in the binding of RNA in these experiments as these proteins lack C-terminal regions including p7 Nucleocapsid which are important for gag binding. At the time, despite substantial effort, this was the purest protein that we could use in the following experiments. Only latterly in the course of my studies did I manage to purify native p55 (see chapter 5).
4.3 Deletion Analysis of the Nucleotides in the 5' Leader Sequence that are Required for Binding GST-p55

The full length RNA leader sequence, an RNA from +1 to 404 (viral RNA sequence), was progressively truncated from both the 5' and 3' ends to investigate GST-p55 binding in a deletion type analysis. This was carried out to investigate the nucleotides that may be involved in the binding of GST-p55. Figure 4.3 shows the mulfold prediction for the RNA secondary structure in the 5' leader sequence from nucleotides +1 to 406. In addition to computer analysis, this region has been subject to extensive secondary structure probing experiments (e.g. Clever et al 1995, Baudin et al 1993), and all substantially agree on the structure depicted in figure 4.3. Indicated in figure 4.3 are all those RNAs that bound GST-p55 in nitrocellulose filter binding assays, these are labelled (+). RNAs that bound GST-p55 showed equal or greater than 40% RNA bound at 136nM GST-p55. Those RNAs that showed no appreciable affinity are labelled (-), these typically showed less than 5% binding at 136nM in filter binding assays.
Figure 4.3 Shows the multifold predicted secondary structure for the leader sequence of HIV-1NL43 RNAs are shown diagrammatically with their 5' and 3' boundaries shown on the right hand side of the diagram. Indicated are the transactivation response hairpin (Tar), 1st polyadenylation signal (poly A), the primer binding site (PBS), then 4 smaller stem loops known as SL1, SL2, SL3 and SL4. SL4 may be predicted to form this extended structure as shown here or a truncated version starting from nucleotides 339-352 (H9). ++++ has been drawn to show the physical link between Tar and Poly A hairpins, which cannot be drawn easily due to the close proximity of the stem loop structures. Also labelled are the positions of the nucleotides that make up the 5' and 3' boundaries of the RNAs. + describes an RNA which binds to GST-p55, and - describes an RNA which does not bind GST-p55.
A number of RNAs were constructed and their length is indicated on figure 4.3. The construction of each RNA was based on the information gained from the binding experiments with preceding RNA. A nitrocellulose filter binding assay was employed to assess the binding capability of the RNA to GST-p55. As employed in these experiments this method is not suitable for determining the subtle differences in protein-RNA affinity. A positive control, generally H4 and a negative control, C1, were included in each filter binding set, to compare binding of new RNAs. Experiments for each RNA were carried out several times to confirm the binding characteristics of each RNA.

4.4 5' and 3' Deletions Centre Binding on SL4
RNAs H1(1-406) to H5(195-362) localised by deletion analysis a binding site for GST-p55 between nucleotides 278-362. This region contains the stem loops SL2, SL3 and SL4. The fact that an RNA starting at 278 and ending at 336 (RNA H6) didn't bind suggested that SL4 must be important for gag binding and that sequences within nucleotides 278-336 are possibly not sufficient for gag binding. The importance of SL4 in gag binding is further suggested by the fact an RNA from 317 to 404 (H7) containing SL4 and 3' sequence, also binds GST-p55.

4.5 RNA Sequences Flanking SL4 Don't Bind GST-p55
H8 was thus constructed representing SL4 truncated, starting at 338 and ending at 404. This RNA initially looked as if it bound GST-p55, and on this information RNAs H9 (nucleotides 338-352) to H12 (nucleotides 332-389) were made. As none of these RNAs actually bound, GST-p55 binding to H8 was repeated several times. H8 actually showed poor binding to GST-p55. H13 (353-404) and H14 (1-335) represent the 3' and 5' RNA sequences to SL4. As neither of these RNAs bind GST-p55, SL4 must constitute a minimal gag binding site.

4.6 Minimal Binding RNAs are centred on SL4
As an RNA representing truncated SL4, H9 (nucleotides 338-358) does not bind GST-p55 it was suggested that the interaction may need 3' sequence for binding. H16 was subsequently made by linearising the H9 plasmid (D1), at the Pvu II site in the pUC119 vector sequence. When transcribed this gave H9 with a 232 nucleotide tail. This RNA also failed to bind GST-p55, indicating other specific HIV-1 sequences are necessary, or that nucleotides 5' of SL4 were needed. H15 (nucleotides 317-362), H17 (nucleotides 317-379) and H18 (nucleotides 317-389) all bind to GST-p55 and tend to suggest that SL4 is important in RNA binding but small amounts of 5' and 3' sequence are needed for efficient binding.
4.7 Leader RNAs Deleted in SL4 bind GST-p55

As GST-p55 binding to HIV-1 transcripts seems to centre on SL4, the SL4 sequence was deleted from those longer RNAs H1 to H4, giving H19, H20, H21, H22. Surprisingly, all these RNAs bound to GST-p55 similarly to wild type in filter binding assays, see figure 4.4. These experiments were repeated three separate times generating a single data set for each RNA. Shown is a representative graph from one such experiment.

Figure 4.4

The deletion of SL4 (nucleotides 334-358) from these RNA transcripts has little effect on the binding capability of these RNAs to GST-p55. This suggests that although SL4 is important for the binding of RNAs such as in H15, H17 and H18, this interaction may not be essential for the binding of longer RNAs.

4.8 H4 and H22 Have Similar Affinity for GST-p55

As H4 and H22 both seemed to bind GST-p55 it was possible that GST-p55 bound H22 with less affinity than H4, but this assay was not able to detect this difference. Binding reactions were therefore set up with increasing amounts of excess tRNA to act as non-specific RNA competitor for the RNA-protein interaction. If GST-p55 bound H22 with less affinity than H4 then binding would probably be out competed by competitor, more easily than in a similar reaction with H4. Filter binding reactions were carried out using 136nM GST-p55 and 20000cpm RNA. Increasing amounts of non-specific competitor
tRNA was added to the reactions, binding was allowed to proceed for 15 minutes on ice, before application to the filter (see fig 4.5).

Figure 4.5

Filter binding in the presence of non-specific competitor RNA shows essentially similar binding curves for each RNA. On each occasion binding was carried out the binding for the deletion RNA was lower. This must mean that although SL4 constitutes a gag binding site, further gag binding sites must be present in the RNA leader.
4.9 A Minimal SL4 RNA Requires Small 5' or 3' Sequence.

Further RNAs were constructed to determine the minimal binding site centred on SL4. H15 (nucleotides 317-362) defined the 3' boundary of the minimal binding RNA to 362. H24 (nucleotides 333-389) defined the 5' boundary to nucleotide 333. However when these two minimal boundaries were put together in H26 (nucleotides 333-362), the RNA did not bind. Binding centred on SL4 must therefore require a short 5' or 3' sequence to facilitate binding of the protein to the stem loop. A diagrammatic representation of these structures is shown in figure 4.6

Figure 4.6

4.10 3' Boundary Determination for the H4/GST-p55 Interaction

The filter binding study showed that protein binding to RNA seemed dependent on an RNA tail either 5' or 3' of SL4. An experiment was devised to look at the 3' boundary, or the minimal 3' sequence on the H4 RNA that would still support binding. This experiment is described in chapter 2. Briefly, 5' kinased H4 RNA was hydrolysed to form a ladder. This ladder was used in a binding reaction then protein bound RNA applied to a nitrocellulose filter. RNA retained on the filter was eluted by boiling in formamide loading dye and the RNA analysed on 12% denaturing polyacrylamide gel, see figure 4.7. In this experiment the largest RNAs (which should all bind the protein) will run high up in the gel. As the RNAs become progressively shorter, from their 3' end, there should be a sudden and immediate break in the RNA ladder signifying the nucleotides that form the boundary of the RNA binding site.
Figure 4.7 Autoradiograph showing, purified RNA H4 used (Native), a T1 digest sizing ladder (T1 digest), the hydrolysed RNA used in the experiment (size ladder). Three binding reactions containing 0nM GST-p55, 68nM GST-p55 and 136nM GST-p55 are labelled. The position of the G residues sized from the T1 digests are shown, as are the positions of the predicted secondary structure in this RNA.
The RNA only bound to the filter in the presence of protein showing that the retention of RNA on the filter is due to RNA-protein binding (compare 0nM GST-p55, with 136nm GST-p55). Using this technique RNA corresponding to the 5' of SL2 has bound GST-p55. RNA composed of the complete SL2 with some 3' tail bound poorly. RNAs encompassing SL3 and SL4 also bound to GST-p55. RNA with 3' boundaries including SL3 or SL4 also bound GST-p55. Thus RNAs not containing SL4 also bound, as found during filter binding in the case of those RNAs containing a deletion of SL4.

It is unclear why these smaller RNAs, such as those forming the 5' stem of SL2 might apparently bind. As RNA corresponding to SL2 but including some 3' sequence binds GST-p55, it is likely that the 5' stem of SL2 may apparently bind to GST-p55, by association with other unlabelled RNAs carrying the 3' stem of SL2. Thus RNAs of this size may not actually bind on their own. Following this reasoning, further information cannot be gained on the 3' boundary of RNA binding sites using this experimental approach.

4.11 Role of Kissing Loop Dimer in RNA-gag Recognition

It is not known whether the virion genomic RNA dimerises before or after packaging. The two genomic RNA monomers may dimerise within the virion. It is likely that the RNA dimer matures within the virus particle as genomic RNA harvested from young virions is immature with respect to thermal stability (Fu and Rein 1993). The low packaging efficiency of kissing loop mutated virions (Laughrea et al 1997) may be due to the fact that p55 recognises a structural motif that is only present in the dimeric RNA, e.g. the tertiary structure formed by the interacting RNAs in the kissing loop region. It is possible then, that dimerisation occurs in the infected cell by kissing loop mediated dimerisation via SL1-SL1 interactions. Therefore does dimerisation of genomic RNA by SL1-SL1 interactions enhance binding to GST-p55? To investigate this idea, a kissing loop mutant and wild type RNA was analysed by filter binding for their comparative ability to bind GST-p55. In this experiment an RNA transcript from 195-404 (containing a non-complementary kissing loop sequence RNA, see chapter 3) was compared to its wild type, kissing loop competent RNA, H2 (RNA 3 in chapter 3), in filter binding reactions. In this experiment both the RNAs bound to GST-p55 as depicted by similar filter binding graphs (figure 4.8)
Figure 4.8

Filter Binding ACS Mutated RNA

Figure 4.8 Shows the filter binding curves for RNA H2 and the same RNA with substitutions in SL1 (a change from $\text{257GCGCGC}_{263}$ to $\text{257AAACGC}_{263}$). Negative control C1 again shows no binding.

As the mutated RNA is unable to form kissing loop dimers (Haddrick et al 1996) and wild type RNA readily dimerises, as confirmed by analysis on polyacrylamide gels, (not shown), this experiment suggests p55 doesn't recognise a kissing loop RNA any more efficiently than the monomeric RNA. This suggests that the primary RNA-protein interactions can take place between both monomeric and dimeric RNA.

4.12 Ribonuclease Footprinting Identifies further p55/RNA binding sites

Ribonuclease footprinting experiments were carried out to look at those RNA-protein interactions that occur in the RNAs that were identified as being able to bind GST-p55. As described in chapter 2, protein and RNA binding reactions were set up on ice. Ribonucleases T1 and V1 were added to the reaction after binding had occurred. After a short incubation the reactions were loaded onto denaturing polyacrylamide gels. RNase T1 was employed to cleave after unpaired G residues. RNase V1 was used to cleave after paired or stacked nucleotides. 5' end labelled RNAs H4(278-404) and H18(317-389) were used to identify protein-RNA interactions that were occurring in SL4 and SL2/SL3 respectively. In these experiments it was estimated by filter binding that about 75% of the available RNA was bound to the GST-p55 protein in the reactions containing highest concentrations of protein.

H18 was bound to increasing amounts of GST-p55. RNase T1 was added at a constant amount (0.1 unit) to each reaction. Shown in figure 4.9 is the autoradiograph of this experiment.

Figure 4.9

Figure 4.9 Autoradiograph of the H18-GST-p55 T1 Ribonuclease footprint. Alkaline hydrolysis ladder for sizing (RNA ladder). Reactions with an increasing amount of GST-p55 are shown. The positions of G residues in this RNA are shown down the right-hand side.
H18 shows greatest sensitivity to T1 in those areas predicted by Mulfold to be single stranded e.g. G328 between SL3 and SL4. There is no evidence of enhanced or suppressed T1 digestion by the presence of GST-p55. This suggests that any interactions that occur between the protein and SL4 may occur in the double stranded stem regions.


To investigate whether GST-p55 binding to SL4 may involve the stem motif, the digestion reactions on protein bound to H18 were carried out using ribonuclease V1 (see figure 4.10).

**Figure 4.10**

![Figure 4.10](image)

*Figure 4.10* V1 digests in the absence of protein are shown in A, using the increasing units of RNAse V1 as shown. Digests carried out on binding reactions containing protein are shown in B. Positions corresponding to nucleotides in the stem regions of SL1 that seem differentially sensitive to RNAse V1 in the presence of GST-p55 are shown.

In the absence of GST-p55 the H18 RNA shows 5 points of cleavage (G348 to C352), corresponding to the 3' half of the SL4 stem loop sequence. This suggests that SL4 exists in its smaller or truncated form as drawn in figure 4.17, rather than the extended form as
drawn in figure 4.3, in this RNA. Reactions between H18 and RNase V1 were modified in the presence of GST-p55 (4.10B).

The Nucleotides on the 5' side of the stem loop corresponding to nucleotides 342-343 show partial protection from V1 ribonuclease, as they are not cleaved to the same extent as RNA digested in the absence of protein. Nucleotides on the 3' side of the SL4 stem loop show an altered cleavage pattern to the digest that contains no protein as stronger bands are seen in the case of nucleotides G348 to C352. This altered cleavage must be enhanced cleavage due to the binding of GST-p55.

4.15 RNAse T1 Footprint of GST-p55 / H4 Reveals Sites of Protection In SL3
T1 digests were carried out on H4 / GST-p55 using increasing amounts of T1 ribonuclease. This longer RNA was used to investigate those interactions that must be occurring 5' to SL4 which are important for binding GST-p55 in those RNAs that have SL4 deleted (H19-H22). T1 Ribonuclease was added to binding reactions that had been set up in the presence or absence of protein. Figure 4.11A shows those reactions set up in the absence of GST-p55 and Figure 4.11B those reactions set up in the presence of GST-p55.
Figure 4.11 shows A T1 digests carried out in the absence of GST-p55 using increasing amounts of RNase T1 indicated. 4.11 B shows the reactions carried out in the presence of GST-p55, using increasing amounts of RNase T1 indicated. The positions of the nucleotides which show some protection are shown on the right-hand side.

T1 cleaves most strongly at those positions that are predicted by Mulfold to be single stranded e.g. G318 in the loop sequence of SL3. Residues that show protection from T1 cleavage, indicated by their lower band intensity in the presence of GST-p55 are G310, G315, G317, G318, and G325. These nucleotides are all contained in the SL3 stem loop, indicating the importance of SL3 stem and loop sequences in the binding of GST-p55.
4.16 RNAse V1 Footprint of GST-p55 / H4 Reveals sites of Protection In SL3 and SL2

V1 digests were carried out on H4 / GST-p55 using increasing amounts of V1 ribonuclease. This longer RNA was used to investigate those interactions that must be occurring 5' to SL4 in areas of double stranded RNA which cannot be detected by T1 digestion. V1 Ribonuclease was added to binding reactions that had been set up in the presence or absence of protein. Figure 4.12A shows those reactions set up in the absence of protein and Figure 4.12B those reactions set up in the presence of GST-p55.

Figure 4.12

Figure 4.12 Shows A V1 digests carried out in the absence of GST-p55 using increasing amounts of RNAse V1 indicated. B shows The reactions carried out in the presence of GST-p55, using increasing amounts of RNAse V1 indicated. The positions of the nucleotides which show some protection are shown on the right hand side U313 is not shown as it has merged with C312 in this scan.
V1 Ribonuclease shows cleavage at those regions predicted to be double stranded in SL3 and SL2. Residues that show protection from V1, indicated by their lower band intensity in the presence of GST-p55 are C312 and U313 in SL3 and nucleotides C297, G298, C299 and C300 in SL2. This shows that GST-p55 interacts with the stem sequence of SL2 and SL3 in addition to SL4, as shown by V1 digestion of H18 RNA.

4.17 Summary of Ribonuclease Footprinting Analysis

Figure 4.13A shows the points of RNAse sensitivity that were detectable in an RNA sequence from nucleotide 278 to 362 in the RNA leader of HIV-1.

**Figure 4.13 (A)**

![Diagram](image)

Figure 4.13A shows the points of sensitivity by RNases T1 and V1 superimposed on the secondary structure of this sequence as derived by Clever et al (1995). Cleavage by RNase T1 is scored on a scale of 1-3 triangles on the above diagram. This was as judged by eye from the footprinting gels. The points of most efficient cleavage occurred in those regions that have previously been predicted to be single stranded. There is some cleavage by RNase T1 in regions predicted to be double stranded, although not as strong as those predicted to be single stranded. V1 RNase only cleaved at regions that were expected to be double stranded. Cleavage by Ribonuclease V1 only gave a pattern at SL4 corresponding to 5 base paired nucleotides in the stem loop. This would be consistent with this structure adopting the shortened or truncated structure as drawn above rather than the structure drawn in figure 4.3.
Figure 4.13 (B) shows the points of cleavage by RNases T1 and V1 that are differentially sensitive to the presence of GST-p55. This was determined by eye using autoradiographs of the reaction and confirmed by phosphorimage analysis of the dried gel. This shows that SL4, which is essential for the binding of some RNAs shows protection from cleavage in residues 342 and 343, and enhanced cleavage of residues 348-351. This must mean that GST-p55 is interacting with the stem sequences in SL4, maybe covering nucleotides 342 and 343, whilst leaving exposed the nucleotides 348-351. The GST-p55 protein must also be interacting with the stem loop of SL3 and parts of SL2 as addition of the protein hinders the access of the nuclease to these sites. Presumably it is interaction with some or all of these domains that allows the SL4 deleted RNAs to bind GST-P55.
4.18 Northwestern Analysis of p55-gag / RNA Interactions

It is reasonable to assume, due to our initial control experiments, that the binding of p55-GST to RNA in the filter binding assays, and the interference of T1 and V1 digestion in the footprinting work is due to the Gag moiety of the fusion protein. However the possibility that the gag binding characteristics are modified by the 30kDa GST fusion cannot be ruled out. It was therefore desirable to repeat some of this work using authentic purified p55-gag protein. Chapter 5 describes a successful strategy to purify authentic recombinant p55 from E. coli, however this protein was only available at the end of my studies. Due to initial difficulties in protein purification, a membrane bound northwestern assay was used to try and investigate the p55 - RNA interaction, using partially purified or crude extracts. In this assay crude protein extracts are subject to SDS-PAGE, blotted onto nitrocellulose filters and the protein is then allowed to renature in buffer. The blot is then hybridised with radiolabeled RNA, and after washing subject to autoradiography. Specific protein-RNA interactions should be detected on film. Luban and Goff (1991) showed specific binding of GST tagged p55 in crude extracts to HIV-1 leader RNAs by a northwestern technique, and managed to locate at least two gag binding sites within the gag coding region. Using extracts of factor Xa cleaved GST-p55, northwestern blotting was tried with a view to testing some of the salient RNAs from H1-H26, for their ability to bind p55. This assay could also be used for directly comparing RNA binding by p55 and p55-GST, simultaneously on the same blot. The same conditions described by Luban and Goff were tried. p55-GST was used initially to try and optimise the conditions for hybridisation. Very high background binding of RNA to the gel was seen in these experiments, and was not improved by using alternative RNAs or extensive washing (data not shown). The type of membrane also did not appear to effect the non-specific binding. High background binding was seen throughout this investigation.
4.19 Northwestern Binding of HIV-1 RNA to immobilised GST-p55 is Dependent on KCl Concentration

Northwestern blots were carried out as described in chapter 2. This experiment shows the effect of KCl concentration on binding. Binding of HIV-1 RNA (H4) to immobilised p55-GST was seen when the potassium chloride concentration was adjusted. As Luban and Goff reported, binding of HIV-1 RNA to immobilised p55 seems to be dependent on KCl concentration, (figure 4.14). This experiment showed optimal binding at 90mM KCl.

Figure 4.14

The blot in figure 4.14 shows low overall binding. This autoradiograph was a 48 hour exposure, the GST-p55 bands are faint and there is high background. Binding is dependent on the concentration of KCl in the hybridisation, (fig 4.14 compare lane 0mM KCl with 90mM KCl.) This binding was difficult to reproduce.

4.20 RNA Binding to Immobilised p55 Protein is not Specific

Using large amounts of protein per lane (1-2μgs) it was possible to obtain reasonably reliable binding by HIV-1 RNA (H4) to the immobilised protein. However using a control RNA (C1), RNA-protein hybridisation can be seen to be non-specific as both RNAs bind to the protein with equal intensity in as judged by the intensity of the bands in the autoradiograph, see figure 4.15A and B.
HIV-1 RNA sequence (H4) and a control sequence (C1) both bind to the proteins on the protein blot. Both RNAs bind to the GST-p55 bands, the putative p55 band and the smaller truncates of gag that have purified on the GS4B matrix. This experiment shows that RNA binding to immobilised GST-p55 and p55 is non-specific under these conditions and therefore this method can't be used as a method for the analysis of RNA-gag interactions.

**4.21 GST-p55 Shows No Difference in Affinity Towards H4 or C1 When Immobilised to a Nitrocellulose Membrane**

In order to try and cut down the non-specific binding of C1 RNA to GST-p55 in a northwestern blot, increasing amounts of yeast tRNA was added to hybridisations containing $2 \times 10^5$ cpm/ml either RNA. The RNA bound to the GST-p55 protein was visualised by autoradiography, see figure 4.16.
Figure 4.16

Panel A shows those protein samples hybridised with the H4 RNA probe. Panel B shows those protein samples hybridised by the C1 RNA probe. The lanes on each blot are labelled with the amounts of tRNA per ml, added to each hybridisation. Shown is the putative band that corresponded to the size of GST-p55.

The addition of excess tRNA to each hybridisation did not out compete any of the non-specific binding of RNA to immobilised protein, RNA H4 and C1 both bind to the immobilised protein in the presence of 50μg/ml tRNA competitor. At higher concentrations of tRNA used in this experiment no binding is seen using either RNA, suggesting that both RNAs are binding with similar affinity to the immobilised protein.

4.22 Northwestern Blotting Extracts from Infected Cells Using H4 RNA Fails to Bind p55-gag

Infected cell extracts and non-infected cell extracts were compared in their ability to bind HIV-1 RNA by northwestern blotting. Northwestern analysis may require authentic protein to work, which may in part be due to any post-translational modification the protein undergoes in the eukaryotic cell. In this experiment extracts prepared by NP-40 lysis of 1x10^5 C8166 cells or C8166 cells infected with HIV-1NL-43 (7 day post infection) were subjected to SDS-PAGE then northwestern blotted with H4 RNA, the autoradiograph of this hybridisation is shown in figure 4.17.
Figure 4.17

Northwestern blot autoradiograph, probing uninfected cell extract (A) and infected cell extract (B) with RNA H4. C=cytoplasmic and N=nuclear fractions.

Figure 4.17 shows that the infected cell extract shows no obvious band at around 55kDa which may signify specific binding to p55. The RNA binds to various proteins in both the uninfected and infected cell extracts, these proteins may be cellular RNA binding proteins. Northwestern blotting may unsuitable for detecting RNA-protein interactions which are specific to infected cells.

4.23 Conclusions And Discussion

4.23.1 Filter Binding

p55 gag is the initial gag protein to make contact with the genomic RNA. Later on, after assembly and maturation, the cleavage products p7 Nucleocapsid and p15 bind non-specifically to RNA sequences. This work therefore utilised a p55-gag GST fusion protein to analyse the RNA protein interactions occurring during the initial stages of packaging. GST-p55 was shown to have affinity for leader derived RNA sequences and GST on its own showed no appreciable affinity. The RNA affinity was then shown to be specific for leader derived RNA, and that an RNA of the same size but of essentially irrelevant sequence did not bind the protein. These experiments showed that p55 has affinity for HIV-1 leader derived RNA when part of a fusion protein and that the 30kDa GST has no appreciable effect on RNA binding.
The filter binding data described in this chapter reveals that GST-p55 binds specifically to HIV-1 RNA transcripts derived from the 5' leader sequence. Transcripts which contain SL1-3 (H14), or SL2-3 (H6) do not bind gag. Transcripts that contain the secondary structures SL1-4 such as H4, and SL4 alone (H7, H15, H25) also bind gag. The minimal binding RNA is centred on SL4 and this RNA must include either 3' or 5' sequence for protein binding (as in H24 and H25). This suggested that SL4 was the primary gag binding site. The deletions of SL4 in RNAs H19-H22 were subsequently made. These all bound with the same apparent affinity as H1 even in the presence of 40 million-fold molar excess tRNA. The deletion of SL4 showed that gag binding sites are indeed present, 5' to SL4. As to a reason why these RNAs (H14 and H6) do not bind gag, it is possible that in the absence of 3' sequence these RNAs do not fold correctly to form the correct gag binding sites.

To further define the 3' boundary of the gag binding site, filter binding reactions were performed using a 32P end labelled ladder of RNA H4. RNA bound to the nitrocellulose filter was eluted in formamide loading dye and analysed. It would seem that SL2 RNA on its own has bound to p55-GST. As SL2 with a small 3' tail binds poorly in this experiment, it is most probable that these smallest RNAs interact with those unlabelled RNA sequences that represent the 3' of the SL2 stem sequence which would have been produced during the hydrolysis. If this is the case then the 3' boundary cannot be mapped utilising this technique, as other cold fragments of RNA may have been interacting with RNA during the protein-RNA binding step. This may have allowed these RNAs to apparently bind protein during the filter binding step. Boundary experiments like these were carried out by Allen et al (1996) and in their experiments using NCp7 protein on SELEX RNAs their experiments gave a far more definite cut off when analysing the protein binding to a stem-loop structure.

4.23.2 RNA Dimerisation is not Required for p55-RNA binding

It is of passing interest that the RNA H2(195-404) which is dimerisation competent as wild type, and H2 with the kissing loop mutated, which doesn't dimerise, both bind with similar binding profiles as determined by nitrocellulose filter binding. This may be evidence to show that a dimeric RNA is not essential for recognition by p55, and is not preferentially bound by p55. This may mean that in the infected cell, a dimeric RNA is not required or specifically recognised for packaging and that RNA dimerisation of the genome takes place after incorporation into the virus. It is also possible that if a dimeric RNA genome is required for packaging maybe the sequences recognised in this dimeric (packagable) RNA are not present in this 5' region of the viral genome.
4.23.3 Ribonuclease Footprinting Identifies RNA Binding sites 5' to SL4

The ribonuclease footprinting was carried out to identify at the nucleotide level the residues that are probably interacting with the p55-GST protein in SL4 and also in sequence 5' to SL4. The digests by the T1 and V1 enzymes were consistent with those that would be expected to occur from the Mulfold secondary structure prediction, and that of other groups who have probed the secondary structure of the RNA leader in HIV-1 (Clever et al 1995). The data derived from V1 digestion would tend to suggest that it is the smaller stem loop that may form in SL4 from 348-361, as in figure 4.13 rather than the extended structure prediction as shown in figure 4.3. This is because there seem to be a smaller number, (5), V1 cleavages at the SL4 than would be expected if SL4 existed in the extended structure (7). RNase footprinting using ribonuclease T1 on RNA H18 failed to show any points of differential nuclease sensitivity indicating that GST-p55 does not interact with single stranded sequence in this RNA. In the case of V1 ribonuclease however, the digest pattern was modified in the presence of gag. Two nucleotides G342 and 343 on the 5' side of SL4 stem loop were protected, whilst four nucleotides on the 3' of SL4 stem loop showed enhanced nuclease digestion. This is similar to what was found by Berglund et al (1997) who proposed binding of p7 Nucleocapsid to a SELEX enriched RNA involved the 5' of a stem loop leaving the 3' side available for RNase V1. To investigate gag binding upstream of SL4, H4 was used in further experiments. These show that residues in both SL2 and SL3 are also involved in gag binding. This confirms the filter binding data with those RNAs H19-22, proving that there are further binding sites 5' of SL4. So far only one other group has generated a footprint from p55-RNA in this region. Damgaard et al (1998) used GST-p55 to generate a footprint of the leader from +1 to 368. In this RNA two sites of ribonuclease sensitivity in the loop of SL4 showed 1.5-2 fold protection, whereas 3 sites in SL3 showed greater than 3 fold protection. This is broadly comparable to what was seen with my footprinting results. This group however noticed a 3 fold enhancement of cleavage in the loop region of SL2. While this was not seen in my experiments, this further suggests roles for SL2 in gag binding.

The footprinting results presented here are entirely consistent with those results that were obtained in the filter binding study. Both show that gag binds specifically to the 5' leader sequence of HIV-1 transcripts. Within this region gag may bind preferentially to SL2, SL3 and SL4. SELEX type experiments have shown that gag protein binds to stem loop structures, which do not obviously resemble those from HIV-1 (Lochrie et al 1997, Allen et al 1996 and Berglund et al 1997). The high frequency of G containing base pairs in this region (SL2-4) may indicate that HIV-1 gag binds to secondary structures with a high purine content. It may be that gag binds to a multipartite structure made up of SL2, SL3 and SL4, however from the results presented here, this is probably unlikely as the protein
tends to bind the different RNAs with equal affinity. It is more likely that gag binds to several elements independently. This maybe creates a large localised concentration of gag that is sufficient to package the RNA into an assembling virus.

The preferential packaging of the genomic RNA rather than subgenomic RNAs may be due to the fact that multiple gag binding sites occur in the full length RNA but, maybe only one or two in the spliced messages. This being the case, the efficiency of packaging in the virus by recognition of multiple binding elements by multiple gag proteins would be dependent on the number of binding elements that the RNA possessed. The major contributory factor to the selection of genomic RNA rather than spliced is the position of the binding sites, that were identified. The filter binding and RNase footprinting work showed that the binding elements identifiable in this study were in the region 3' of the first splice donor. The only RNAs carrying this sequence therefore are the full length, unspliced genomic messages. The other viral messages have this sequence spliced out as intron sequence.

The three dimensional structure of p7 Nucleocapsid bound to SL3 has recently been described, (DeGuzman et al 1998). This shows that the protein forms contacts with the major groove of the RNA stem and nucleotides 318, 319 and 320 of the loop, which is in good agreement with the footprinting data from this region. This p7 Nucleocapsid data may apply to p55 as the p7 Nucleocapsid domain is the region likely to be interacting with RNA in the context of p55 and it is conceivable that after the cleavage of the polyprotein there is no rearrangement of the p7-RNA interaction.

4.23.4 Northwestern Blotting

The Northwestern blotting of the GST-p55 - RNA interaction detected non-specific binding of the protein to RNA. The binding showed similar affinity to HIV-1 leader RNA and a size matched control RNA when hybridisations were carried out with an excess of tRNA. When GST-p55 is bound to an immobilised filter in this way the protein behaves like a generalised RNA binding protein lacking sequence specificity. No specific binding of RNA to protein was seen when infected cell extracts were used either. It is probable that the protein does'nt refold properly to form the correct conformation required for specific binding after denaturation on SDS PAGE. Also likely is that the protein can't bind RNA properly when it is immobilised to a nitrocellulose filter. These experiments showed that northwestern blotting was not a reliable enough technique to investigate RNA-protein interactions in this instance.
4.23.5 Future Experiments

All the work described in this chapter has used GST-p55 protein. While reproducible specific binding to RNA sequence has been presented, it cannot be ruled out that the GST domain must in some way alter the way in which this protein binds its RNA target. Glutathione-S-Transferase is able to dimerise, so it may be that in our experiments a dimer of GST-p55 is binding RNA, which is probably not representative of gag-RNA interactions in-vivo. If in vitro experiments are carried out any further, these must utilise intact p55 protein lacking any fusion tag (see chapter 5).

It would be desirable to carry out a detailed ribonuclease footprint of the entire leader sequence using native, intact p55 protein. Using the data derived from a study such as this predictions on RNA-protein binding could be made and incorporated into an in-vivo mutagenesis study. This might determine the exact roles of p55-RNA binding in terms of packaging in-vivo. It would be interesting to analyse the minimal core binding stochiometry of protein to the RNA in-vitro, i.e. how many protein monomers bind an RNA, maybe at multiple sites, or does a protein aggregate bind an RNA at a single site? It would also be interesting to investigate which regions direct protein binding preferentially, or initiate binding in the leader RNA. It would be useful to see which regions of p55 support the specific binding of the protein to RNA. If the p7 Nucleocapsid domain gives the protein its RNA binding capability, other regions in p55 must contribute to the selectivity of the protein-RNA interaction. This could be done by either protease treatment of the purified protein, followed by protein purification, or by protein mutagenesis studies. These type of experiments may yield useful information as to the nature of recognition of RNA by p55, and give insights into how specific binding is achieved.
Chapter 5

Purification of Recombinant p55 Gag from E.coli

5.1 Introduction
Most in-vitro HIV-1 RNA packaging studies carried out to date have used GST fusion proteins for investigations such as those described in chapter 4. A few groups have used authentic p55 gag made from in-vitro transcription translation systems (e.g. Geigenmuller and Linial 1996), or from baculovirus expression systems (Jowett et al 1992). However these interactions have been studied using northwestern type approaches in which the RNA-protein interaction probably behaves differently to that which occurs in solution. The advantages of using fusion proteins in studies like these is that these proteins are easily purified, give high yield and have good solubility. The GST tag has not been shown to have any appreciable effect on the RNA binding ability of the p55 domain of the fusion protein (see chapter 4). It would however be desirable to study this RNA-protein interaction using intact, native p55 protein, as it is possible that the GST tag may interfere with gag-gag interactions that may occur at the RNA binding site. GST protein actually dimerises itself (Tudyka and Skera 1997) and this may cause differences in the RNA binding ability of the p55 domain. To date, interactions between authentic p55 and RNA have not been studied in solution. These experiments need to be performed to gain a fuller understanding of the mechanisms involved in the recognition of RNA by p55. Described in this chapter are strategies used to express p55-gag in prokaryotic expression systems with a view to using the product in RNA-Protein interaction studies.

5.2 Cleavage of GST-p55 by FXa
The pGEX1-p55 construct had been designed such that there was a Factor Xa (FXa) cleavage site at the end of the GST domain that would precisely cleave the protein to give p55 starting with the N-terminal glycine as in virion expressed p55 (see fig 5.1).
Figure 5.1

Thus cleavage could potentially be carried out with thrombin or FXa, FXa giving the more authentic protein. GST-p55 showed no cleavage in the presence of thrombin (data not shown). To investigate the feasibility of GST-p55 cleavage by FXa an experiment was carried out in solution using purified GST-p55 protein. Figure 5.2 shows the cleavage of 1µg of GST-p55 with an increasing amount of FXa for 1 hour in FXa buffer. The figure shows a western blot of these reactions using the anti p6 polyclonal antibody, referred to in this chapter as p6 antibody.

Figure 5.2

GST-p55 is cleaved using the higher amounts of FXa. It is evident that the cleavage can go too far (see lane 1000ng FXa) as in this lane the band corresponding to a protein of 55kDa, assumed to be p55, seems to be degraded. FXa cleavage in solution is possible, but it was desirable to carry out the cleavage reaction on the Glutathione Sepharose 4B beads (GS4B). This would allow the chromatographic separation of the GST tag and any uncleaved protein from p55 in the reaction mix.
5.3 Cleavage of GST-p55 on GS4B Matrix

Experiments were carried out using GST-p55 bound to the GS4B matrix. PBS washed, GS4B bound GST-p55 was resuspended in FXa cleavage buffer and incubated with 1µg FXa per 40µl reaction volume. The reaction was incubated at 37°C and the samples were removed over time and analysed by 12% SDS-PAGE. Figure 5.3 shows a coomassie stained gel of this experiment. Under these conditions GST-p55 is not specifically cleaved. There is no protein of around 55kDa in the supernatant fractions, however total protein of high molecular weight seems to be reduced in those reactions incubated for longer periods of time, suggesting that there may be some non-specific protein degradation in these samples.

**Figure 5.3**

![Coomassie stained protein gel of a GST-p55 matrix bound FXa cleavage. M = marker proteins with sizes indicated, P= pellet fraction, S= supernatant fraction. t= Time of incubation indicated.](image)

This experiment showed that specific cleavage of the protein whilst bound to a solid support either does not occur at all or is very inefficient. This gel also highlights the problems associated with this method of expressing GST-p55, in that apparent C terminal truncation has occurred. When this protein is purified, these smaller truncates of the fusion protein co-purify with the intact protein. This is presumably due to protein-protein interactions between the gag domains in the fusion protein and has been noted by others (Saito *et al* 1995, Campbell and Rein 1999, Luban and Goff 1991). In the case of this GST fusion protein, GST-GST interactions may be occurring and allowing the co-purification of smaller contaminants.
5.4 Detergent Activated Cleavage of GST-p55

In an attempt to overcome the problem of poor cleavage a small amount of SDS was added to the reaction mix. This has been reported to be useful for the cleavage of some fusion proteins (Ellinger et al 1991). Essentially the experiment described in figure 5.3 was carried out again but this time 0.5% w/v SDS was included in the digestion buffer. Protein was analysed by running samples on a 12% PAGE followed by coomassie staining (figure 5.4a).

Figure 5.4a

![Figure 5.4a](image)

Figure 5.4a Coomassie stained protein gel of cleavage reactions carried out in the presence of 0.5% SDS. M = marker proteins with sizes indicated, P= pellet fraction, S= supernatant fraction. t= Time of incubation indicated (minutes). There is lot of GST-p55 protein in the supernatant lanes (e.g. t=0 S) which is probably indicative of GS4B bead carry over in this experiment.

GST-p55 could therefore be cleaved by FXa in the presence of SDS giving a presumptive p55 band at around 55kDa. As the presence of SDS was probably undesirable for the future uses of the purified protein parallel cleavages were carried out in 0% and 0.1% w/v SDS. The reactions were performed at room temperature for between 0-60 minutes. Figure 5.4b shows that a concentration of 0.1% SDS is sufficient to facilitate GST-p55 cleavage.
The desired cleavage is signified by a protein band in the supernatant lane at a size of around 55kDa. This presumably meant that the inclusion of 0.1% w/v SDS in the cleavage reaction was sufficient to activate cleavage, by whatever means.

Cleavage reactions were set up to compare the use of alternative detergents. SDS, Triton X-100, and Nonidet P-40 were all tested at a concentration of 0.1%, and compared to a reaction set up with no detergent. This experiment is shown in figure 5.4c.
In these experiments, when SDS is included in the buffer at a concentration of 0.1%w/v, cleavage of the fusion protein is seen. Protein of about 55kDa (presumed to be p55) occurs in the SDS supernatant lane. Cleavage occurred in the other reactions although this was minor when compared to the SDS sample (not enough to see on the previous scanned image). These experiments were repeated with differing amounts of these other detergents but no further cleavage was seen (data not shown).

5.5 Cleavage and Immunoaffinity purification of p55 from p55-GST.

The FXa cleavage of the fusion protein in the presence of SDS was not easy to control as over-digestion regularly occurred and the extent of this seemed to vary from experiment to experiment. Although cleavage should have been arrested by the addition of 2mM PMSF, the cleavage reaction still occurred in the crude reaction. Cleavage would also occur in the reaction mix if left at 4°C overnight in the presence of protease inhibitors, showing that the reaction was difficult to stop. As the cleaved protein in the reaction supernatant was present with many smaller contaminating proteins this meant that the protein had to be further purified. The following gel (figure 5.5) show an example of the unpredictable nature of this reaction (in that p55 is the minor product), and an attempt to immunoaffinity purify the p55 protein. These results are examples of many attempts at p55 purification from GST-p55.

**Figure 5.5**

[Image: Coomassie stained gel of FXa cut experiment. FXa cut is the fraction loaded onto the antibody column, 5,6,7 and 8 are those fractions that were eluted from the immunoaffinity column. The position of the presumed p55 band is shown.]
In conclusion I found that FXa cleavage of GST-p55 was unworkable due to uncontrollable cleavage of the fusion protein, which resulted in poor yield of partially purified product. Purification of the protein that could be cleaved by immunoaffinity chromatography resulted in the co-purification of smaller proteins. The presence of SDS in the reaction mix, essential for digestion was undesirable for the subsequent uses of this protein.

5.6 Expression of p55 from pET 16b-p55 construct.
The construction of pET16b-p55 is described in the materials and methods. Essentially, the p55 gag sequence was excised from the p55-pGEX-1 construct and ligated into the pET16B vector at the Bam H1 and Nco 1 restriction sites. Expression of p55 from this construct in a DE3 lysogen should have allowed the synthesis of native p55 with a short extension to the N-terminus (see materials and methods).

The pET16b-p55 construct was initially transformed into BL21 cells and the production of p55 monitored by western blotting using the p6 antibody, after the cells were induced from 1-5 hours. Also analysed was the solubility of the protein over time. The culture was induced at an OD595 of 0.75, by the addition of 0.5mM IPTG. Samples were taken every hour for five hours and the samples sonicated to obtain soluble and insoluble fractions. Figure 5.6 shows the result of this induction as an anti-p6 western blot.

Figure 5.6

![Western blot of a p55 induction from pET16B-p55 in BL21 cells, using p6 antibody. Time in hours is indicated, S= soluble fraction, I= soluble fraction, U= Uninduced, M= marker. The position of the band corresponding to p55 is marked.]

When p55 was expressed from the pET16B construct in E.coli BL21 cells there was extensive N terminal truncation of the p55 protein as evidenced by the smaller bands that
react with the p6 antibody in western blotting. The intact protein also seemed to accumulate in insoluble fractions over time in this system (compare lanes at one hour with those at 5 hours). On a coomassie stained gel (not shown) these proteins did not appear as dominant bands as might be expected in a high level expression system such as this, this may suggest p55 could be toxic to prokaryotic cells.

Expression from this construct was again tried in another DE3 lysogen E.coli B834. This was directly compared to BL21 inductions (figure 5.7). This induction was carried out at 37°C and also at 30°C, as it has been reported that proteins are more soluble at lower temperatures and less susceptible to proteolysis.

**Figure 5.7**

![Western blot using the p6 antibody](image)

**Figure 5.7** Western blot using the p6 antibody. S= soluble, I= soluble, U= Uninduced, M= marker. The temperature of each induction is indicated, 30=30°C, 37=37°C.

p55 expressed in E. coli B834 cells showed greater solubility at 30°C than 37°C, whilst from E. coli BL21 cells the protein solubility was similar at each temperature. Both cell types show extensive N terminal proteolysis as shown in the above blot (which was probed using the p6 antibody). This extensive proteolysis would have necessitated the production of an N-terminal antibody for protein purification. Even if this could have been achieved there would have been problems due to gag-gag interactions with the smaller fragments interfering in the purification, as was seen in the purification's from GST-tagged protein. It is not understood why this protein also suffered N-terminal cleavages, this has not been previously described when expressing this protein in E.coli.

As the protein was truncated both N and C terminally it is most likely that the truncations seen are due proteolysis rather than an mRNA stability problem within the cell.
5.7 Expression and purification of p55 from pTYB4-p55 construct

The construction of pTYB4-p55 is described in materials and methods. This vector was used with the intention of expressing p55 as a fusion to an intein-chitin binding domain fusion tag. Under appropriate conditions the intein domain can be self cleaved from the fusion protein and p55 would be released. However, as shown in figure 5.8B this construct expresses a 55kDa protein. A recheck of the MH1 primer used to make this construct revealed a TGA stop codon immediately before the EcoR1 restriction site. Therefore this protein was essentially translated from a p55 message with a 2kb 3' tail (the intein fusion message) on the 3' end of the mRNA (figure 5.8A).

Figure 5.8A

![Diagram of p55 expression](image)

Figure 5.8A, p55 was expressed from the pTYB4 vector from RNA transcribed as an intein fusion mRNA but translated as p55 protein with the absence of intein fusion due to an in frame UGA stop codon.

5.8 Induction of p55 from pTYB4-p55 in E.coli

BL21 and B834 E. coli cells were transformed with the construct and the expression pattern of p55 from this construct was analysed. Inductions were carried out for 5 hours at 37°C or 16 hours at 16°C. Figure 5.8B shows the anti p6 western blot results of this induction.
Figure 5.8B

A protein of approximately 55kDa shows up in a western blot using the p6 antibody after induction and then cell lysis in SDS buffer. The protein expresses in both BL21 and B834 E.coli cells. Induction at 37°C was actually better than overnight induction at 16°C resulting in more product and less contaminating truncates, (which were very low in abundance and are not visible on this scanned image). Using the p6 antibody in the western blot suggested that the protein essentially existed as a single species. This meant that purification could be achieved using this polyclonal antibody in an antibody column, with little problem from the effects of contaminating cleavage products. The fact that the protein runs as a single band suggests that the 3' message sequence that encodes the intein CBD protein must protect the p55 mRNA sequence from 3' degradation.

Sonicates were made of small aliquots of induced culture that had been frozen down, the soluble and insoluble fractions were then analysed by western blotting, again using the p6 antibody. Figure 5.9 shows a western blot of the protein in these soluble and insoluble fractions.
p55 was expressed predominantly as a soluble protein in both B834 and BL21 E. coli cells. There are smaller amounts of truncated protein evident in those induction's that were carried out for 16 hours at 16°C, but none at 37°C, (not detectable on the above scanned image). For expression of protein, BL21 cells were induced at 37°C for 5 hours using 0.5mM IPTG at OD595= 1.0.

5.9 Immunoaffinity Purification

The following shows the results of an immunoaffinity purification of the protein using the p6 antibody column as described in chapter 2. Figure 5.10A shows the results of the antibody elution as a coomassie stained polyacrylamide gel. 5.10B shows the protein samples before purification, protein flow through, and approximately 100ng of the purified product.
p55 purifies from the sonicate with little of the contamination that was seen previously with the FXa cleaved protein. The small amounts of contaminating protein that can be seen on these gels could easily be removed on a gel filtration column. From a 200ml induction the estimated total protein recovery was 1.6 milligrams as worked out from A260 and A280 readings of each fraction. To confirm the identity of the protein that was purified, the N-terminus of the protein was sequenced at the University of Leicester Protein and Nucleic Acid Chemistry Laboratory. The first 7 amino acids of the protein were:

Gly-Ala-Arg-Ala-Ser-Val-Leu

This is 100% homologous to the N-terminus of the matrix domain of the HIV-1 p55 gag protein confirming that this method for expression and purification of p55gag, gave the desired protein, and protein prepared by this method are not N-terminally truncated.

5.10 Summary

5.10.1 p55 Purification from FXa cleaved GST-p55
Cleavage of GST-p55 to give p55 by FXa was unreliable and needed the presence of SDS in the cleavage reaction. Problems encountered using this method were due to the uncontrollable cleavage of the protein on the GS4B matrix, which gave low yields and...
contaminating smaller fragments. The non-specific cleavage may (in part) be due to the purity of the preparation of FXa used, (an in house prepared batch of protein), a commercial batch was not tried. Problems may have been compounded by using large amounts of FXa. The cleavage of p55 from the GST tag might have been made easier by engineering a run of glycines known as a glycine linker adjacent to the cleavage site (Rodriguez and Carrasco 1995). This reportedly allows more specific, efficient cleavage, and this may also have cut down on the non-specific cleavages that were seen, by reducing overall cleavage time. It would be interesting to test if protein expressed from a construct that had been prepared this way cleaved any more easily than the fusion protein used in this study. Preparations of protein prepared in this way would not have ideally been suited to the downstream applications of this protein. It would be quite likely the SDS used in the cleavage of the protein could carry through in the purification and be deleterious to RNA binding studies.

5.10.2 Expression of p55 from pET16B-p55
Expression of protein from the pET16B construct resulted in serious N and C terminal truncations of the expressed protein when analysed by western blotting. This occurred under all conditions tested. Cleavages and low yield of p55 gag have been seen before in *E.coli* expression systems (Saito *et al.* 1995, Campbell and Rein 1999, Berkowitz *et al.* 1993, Ehrlich *et al.* 1996 and Luban and Goff 1991). Due to these problems, efforts to express p55 from this construct were halted. Expression from this construct was further complicated by the apparent N terminal cleavages as well. Simple purification from these inductions would have necessitated an N terminal antibody to gag which was not available and would have required recovery of protein from insoluble cell fractions.

5.10.3 Expression of p55 from pTYB4-p55
p55 was expressed from the pTYB4 based construct, essentially allowing transcription of the p55 mRNA fused to an Intein-CBD RNA. When induced in *E.coli* BL21 cells, p55 is expressed as an intact, soluble protein that is easily purified using a C-terminal antibody column. This protein has a mature N terminus lacking the N-terminal methionine and should have an authentic C terminus (if expressed correctly from construct). Due to expression in a prokaryotic cell the N-terminal glycine would not be myristylated as mature HIV-1 p55 is. The observed C terminal truncation seen with the pGEX-1 vector was not seen in the case of this construct. It is most likely that the observed C-terminal proteolysis is due to RNA instability at the 3' end of the message in the pGEX and pET systems. In these cases the truncated proteins were presumably translated from shortened RNAs. In the pTYB4 construct the RNA from which gag has been translated is much bigger than those RNAs transcribed from the pET16B and pGEX vectors. Any 3' RNA
degradation would have shortened intein-chitin binding domain mRNA, which due to the stop codon at the end of gag would not have been translated. Addition of a 3’ RNA tail to constructs expressing eukaryotic proteins in prokaryotic expression systems may be of use if their expression is typified by C terminal truncation. It is likely that the addition of a 3’ RNA tail to an RNA message protects the mRNA from 3’ RNAses digestion, as the only prokaryotic ribonucleases characterised to date cleave 3’ to 5’ (Makrides 1996). As it stands, conditions have been optimised whereby soluble, authentic p55 gag can be expressed and purified with ease from *E. coli* cultures, even though it is not quite in the manner I anticipated.

### 5.10.4 Future Experiments

This protein has not been used in any further RNA-protein binding experiments. It would be desirable to carry out a detailed study of the interaction of this protein with the 5’ leader sequence as this has never been studied using the native protein. Further questions that may be answered using the purified protein could include, a further detailed ribonuclease footprint of the entire HIV-1 leader RNA, as all those done to date have utilised GST fusions proteins. It is entirely conceivable that the GST tag may obstruct the RNA cleavage by the ribonuclease used in these reactions, or dimeric GST domains in the protein may alter the binding characteristics of the protein to the RNA. The protein would be useful in further RNA binding experiments to obtain detailed binding affinity data for each of the RNA templates constructed for the filter binding work. The protein could be used to work out the minimal binding stoichiometry of those protein RNA interactions that assemble in vitro in the filter binding reaction, this may enhance our understanding of viral core formation. The purified protein could also be useful in studying the protein-protein interactions that occur in the *in-vitro* assembly of immature core-like particles (Campbell and Rein 1999), and the specificity of RNA packaging in these systems. It may be possible to use this protein in ultrastructural studies of p55 alone and of p55 bound to viral RNA sequences. The p55 protein expressed in the prokaryotic cell is not modified at its N-terminus by the addition of a myristate group. For studies where a mature protein might be desirable myristylation might be achieved *in-vitro* after purification either chemically, or as Hermamida-Matsumoto and Resh (1999) describe, biologically using reticulocyte extracts.
Chapter 6

In-vivo Studies on RNA packaging in HIV-1

6.1 Introduction

The site of RNA packaging, or RNA selection in infected cells has not yet been determined. It is possible that p55 which has localised to the nucleus of the infected cell may associate with nascent RNA transcripts, in this compartment. In this case some unspliced viral RNA transported by the rev dependant pathway to the cytoplasm may have small amounts of p55 bound to it. This might determine that this RNA acts as a genomic RNA to be packaged rather than translated. Alternatively selection may occur in the cytoplasm, in which case the selection and binding of RNA by gag may be entirely random. It may be envisaged that there is a translation 'pool' of RNA that only serves as mRNA and another pool which can only serve as genome. As HIV-1 viral cores have not been seen in the cell cytoplasm, or for that matter in any other type C retrovirus infection, gag bound RNA must somehow associate with the plasma membrane where budding occurs. This suggests that a p55-RNA complex either associates with the membrane or associates with membrane bound gag. It is also possible that free RNA is captured by a developing virus particle at the inner surface of the plasma membrane. It is therefore of interest to analyse extracts prepared from infected cells that would show where gag and RNA interactions were taking place. This would give new insights into any regulation of genome packaging or selection, and might further define the steps leading to the incorporation of genome into a virus.

The role of p55 with respect to RNA binding was initially investigated in the nuclei of infected cells. Evidence for RNA binding in the nuclear compartment would have interesting implications for the virus lifecycle, as it may mean that RNA was selected straight after transcription in the nucleus, and thus would not have to compete with ribosomes in the cytoplasm for translation. Initially western blotting was carried out to confirm that p55 can be found in the nucleus. RNA-p55 interactions were also analysed in soluble cytoplasmic and cytoplasmic membrane fractions, including the plasma membrane from infected cells.
6.2 Infected Cells can be Fractionated into Cytoplasmic and Nuclear Fractions

To investigate the nuclear localisation of gag proteins in the infected cell, a control western blot was performed to make sure that nuclear extracts were not contaminated with cytoplasmic proteins. Two control western blots were carried out on nuclear and cytoplasmic extracts prepared by NP-40 lysis. HIV-1 gp120, is a cytoplasmic membrane bound glycoprotein which is not localised to the nucleus, and was used to assess contamination of nuclei with cytoplasmic proteins. Similarly the cytoskeletal protein vinculin interacts with proteins at the inner face of the plasma membrane in the cross linking of actin molecules. Vinculin should also show no nuclear involvement (Professor D. Crichley, Department of Biochemistry, University of Leicester -personal communication).

**Figure 6.1**

![Western blot showing equivalent cell loadings of nuclei and cytoplasms. C= cytoplasmic (extranuclear) fraction, N= Nuclear fraction. Blot for vinculin (panel A), blot for gp120 (panel B). In the vinculin blot two different cell lines have been used C8166 and CEM 4 as shown. For gp120 only C8166 cells have been used.](image)

The western blots shown in figure 6.1 show that there was no cytoplasmic contamination of the nuclear fractions. Uncontaminated nuclear fractions are thus easily purified. Vinculin is an abundant cellular protein, which was easily detected by western blotting therefore this procedure was used to validate all subsequent experiments.
6.3 Gag Proteins Localise to the Nucleus of the Infected Cell

Crude nuclear extracts prepared by NP-40 lysis of C8166 cells infected with HIV-1 NL4-3, 60 hours post infection, were subject to SDS-PAGE, electroblotted onto a nitrocellulose membrane and probed using antibodies to p24, p17 and p6. Figure 6.2A shows gag proteins which co-localise to the nuclei of the infected cell. Equal amounts of cytoplasm and nuclei (in cell equivalents) were loaded in each lane of the gel.

Figure 6.2A

![Western blot showing the nuclear localisation of HIV-1 gag proteins in the infected cell.](image)

- Panel A western using p24 mAb (ARP313), Panel B western using p17 mAb (ARP 315), Panel C western using p6 polyclonal antibody. Indicated are the size of the gag proteins detected.

p55 gag and the proteins representing cleavage intermediates of p55 localise to the nucleus of infected cells. These gag proteins represent cleavage intermediates of p55, their sizes corresponding to those expected by specific cleavage due to the viral protease. The intermediate cleavage products seen in the above gels and in other studies tend to suggest that the viral protease may be functional in the infected cell cytoplasm or maybe at the cytoplasmic side of the plasma membrane. It is interesting to note that p17 matrix cannot be detected in the nuclei. This suggests that cleavage of the p55 protein is not occurring in the nuclei (or we would see this protein), and that p17 is not translocating to the nucleus after cleavage in the cytosol or at the membrane. All other gag protein intermediates in the nucleus contain the p17 domain of the gag protein maybe suggesting that this domain is important for nuclear targeting (see figure 6.12B for possible cleavage products). The matrix domain has a well defined nuclear localisation signal which may be involved in transfer of freshly reverse transcribed cDNA to the nucleus after virus infection. This activity must act in competition with a membrane targeting sequence motif and the presence
of an N-myristylated N-terminus, both contributing to the targeting and anchoring of the p55 gag protein to the plasma membrane (see discussion).

**Figure 6.12 B**

This diagram shows predicted cleavage events that have formed those protein cleavage intermediates shown in figure 6.12A. p49 could be formed from the cleavage of p6 from p55 as in 2. P15 may be produced from the cleavage of p55 between p2 and p7 as in 3. P43 may be produced from the cleavage of p55 or p49 between p2 and p7 domains 3. P41 may be produced by cleavage of p55/p49/p43 between p17 and p24 and p2 and p7 5. The mature proteins are illustrated 6.
6.4 Gag appears in the nucleus 24 hours post infection

The time course of nuclear p55 gag accumulation was then determined. This was to investigate the point at which gag protein can be detected in the nucleus. If it were the case that gag only appeared in the nucleus at a late time point in a chronic infection, this would suggest that nuclear p55 does not play a vital role in viral replication. Figure 6.3 shows nuclear p55 accumulation over time as western blots from this experiment.

**Figure 6.3**

![Western blot showing the time course of an infection.](image)

Size markers indicated C= cytoplasmic (extranuclear) fraction, N= Nuclear fraction. Times of each fraction are indicated (hours).

The blot showed that p55 gag could first be detected in the nuclei of the infected cell at around 21-27 hours post infection. This result was obtained on two separate occasions. As this blot shows equivalent cell loading for each lane it is apparent (as measured by eye) that p55 seems to be accumulating in the nuclei with time. Figure 6.4 shows the p24 culture supernatant data for this experiment.
The assay shows p24 titres of released virus, rising after 24-27 hours post infection. At this point p55 exists in both the cytoplasmic and nuclear locations. It is possible therefore that nuclear p55 may play a role in a productive infection of HIV-1.

6.5 p55 Gag Builds up in the Nuclei of the Infected Cell.

For nuclear p55 to play an active role in the packaging of HIV-1 genomic RNA, the protein must be transported from the cytoplasm to the nucleus and then back out again. p55 might travel to the nucleus by virtue of the nuclear localisation signal (NLS) in the matrix domain of the protein. p55 may be then be transported out bound to the RNA by a rev dependent, or some other nuclear to cytoplasmic RNA transport pathway. Alternatively, it is possible that p55 gag may exit from the nucleus during cell division after the breakdown of the nuclear membrane. To investigate whether p55 reaches a steady state distribution between the nucleus and the cytoplasm or whether the protein merely accumulates in the nucleus, cells infected with HIV-1 were grown in a long term infection and the location of p55 determined by western blotting extracts taken over the time period of the infection. In this experiment, equal loading of cytoplasmic and nuclear extracts, as cell equivalents have been loaded in each half of the gels. The blots were probed with both vinculin and p24
monoclonal antibodies simultaneously to show the presence of any contamination of the nucleus with cytoplasmic extract. Performing the double antibody probe proved useful as a control for cytoplasmic contamination of nuclei. Vinculin did not cross react with any gag protein and p24mAb did not cross react with vinculin (data not shown). Figure 6.5 shows a panel of the western blots from this experiment.

**Figure 6.5**

![Western blots from infected cell extracts taken at the times indicated. Extranuclear and nuclear extracts are shown, increasing amounts of extract are loaded for each fraction. The positions of the p55 gag protein and vinculin bands are labelled as determined by double antibody probing.](image)

This experiment shows that p55 accumulates over time in the nuclei of the infected cell, and the relative ratio between cytoplasmic to nuclear protein changes, as judged by eye. At about day 21 the majority of the p55 gag protein is nuclear and there is hardly any cytoplasmic p55. At this point in the infection high numbers of cells (typically 80%) are still viable as evidenced by trypan blue staining but they do not appear to be growing in number. By the end of this experiment the cells had recovered with respect to viability and growth rate. At this point the distribution of p55 resembled that of the productive infection again, with cells remaining in a chronically infected state shedding low amounts of virus.
The titres of p24 released into the culture supernatant are shown in figure 6.6, these effectively mirror what is seen with the distribution of p55 in the infected cell. Essentially the titre of virus being shed into the culture supernatant, rises during the days of the efficient productive infection during the first week of the experiment. When the p55 distribution becomes biased towards the nucleus, the culture sheds low amounts of virus (p24).

**Figure 6.6**

![Extracellular p24 Titres during long term infection](image)

**Figure 6.6** p24 titres from the tissue culture supernatant, from the long term infection described above.

The data presented here, would tend to suggest that the p55 gag protein that is nuclear localised does not move back to the cytoplasm. It is possible that p55 is targeted to the nucleus due to the presence of functional NLS domains rather than p55 targeting to the membrane which is what normally occurs to this protein. In this situation the nucleus is likely to be a dead end for the protein in terms of its role as a virus structural protein. The protein may have an as yet unassigned role in the nucleus, but this was not investigated in this work.

**6.6 Myristylation State of Nuclear p55**

The addition of a myristic acid residue to the N-terminus of HIV-1 p55 gag is a well characterised post-translational modification to the protein which, is proposed to be
important in the targeting and anchoring of the protein to the plasma membrane of the infected cell. If this is the case, then is myristylated gag exclusively cytoplasmic? To further investigate the role of myristylation on nuclear targeting of p55 gag, infected cells were metabolically labelled with $^3$H-myristic acid, fractionated into nuclear and extranuclear fractions and then the gag proteins analysed. Figure 6.7 shows a western blot for the nuclear and cytoplasmic extracts made after a 16 hour label using antibodies to p24, p17 and vinculin. Also shown is the fluorograph of this filter, showing those proteins that are labelled with $^3$H. The labelling protocol used in this experiment is detailed in the materials and methods. To summarise this procedure, 24 hours post infection, the time at which newly synthesised gag proteins are generally detected, $5 \times 10^6$ C8166 cells were grown in media containing 50μCi/ml $^3$H 9,10-myristic acid. After a further 16 hour incubation cells were harvested and fractionated.

**Figure 6.7**

Panel A, shows western blot of the extracts, Panel B shows the fluorograph of the same gel, illustrating those proteins that are tritium labelled. M= marker, C= cytoplasmic fraction (extranuclear), N= Nuclear fraction. The approximate sizes of the various proteins are shown.
Tritiated and hence myristylated nuclear gag proteins include p17, p55, p41 and p49. There is also another N-myristylated protein p28 which shows up on the western blot and fluorograph. This protein has been described before (Furuishi et al. 1996) and is thought to be derived from p17 (with a myristylated N-terminus), but must also contain part of p24, based on its size and reactivity with p24 mAb. All of the p17 containing gag proteins that colocalise to the nucleus, can be seen on the fluorograph, indicating that at least a proportion of the total nuclear gag is myristylated like that of the extranuclear proteins. Myristylated gag is present in both the nuclear and extranuclear fractions.

To determine whether myristylated gag represents a minor part of the total gag in the nucleus an infection was carried out whereby the cells were labelled at 20 hours post infection for a further 24 hours. 20 hours post infection is the approximate start time of gag production. Extracts were made from the infected cells and increasing amounts of each fraction were run on an SDS PAGE gel and then western blotted (figure 6.8A). The bands corresponding to p55 gag were then cut out and radioactivity was quantitated by liquid scintillation counting (see figure 6.8C). In figure 6.8A the average pixel density for the first three bands was plotted for the scanned gel as determined using NIH image software. This detection method did not allow accurate quantification of the bands corresponding to the higher amounts of protein in the other lanes.

**Figure 6.8**
**Figure 6.8** Blot A shows the western blot for p55 on the cytoplasmic and nuclear extracts. Graph B shows the results of densitometrically scanning the western blot A. Graph C shows the relative amounts of tritiated p55 in Blot A.
The plotted data derived from densitometric scanning of the western blot autoradiograph shows that at about 44 hours post infection there appeared to be about the same amount of p55 in each fraction. The scintillation counting of p55 in these fractions shows that the myristylated component of p55 in the nucleus represents about 20-30% of the amount in the cytoplasm. As the total amounts of p55 in each fraction are approximately the same this must mean that the nuclear p55 is underrepresented in myristylated gag. Both forms of p55 gag, myristylated and non-myristylated, are present in the nucleus, although the amount of myristylated gag in the nucleus is reduced compared to that from the extranuclear fraction.

### 6.7 Nuclear p55 does not bind RNA Extranuclear p55 does.

The obvious question as to a role for nuclear p55 was, is this protein binding RNA in the nucleus? To investigate whether p55 was interacting with RNA in the nuclei of infected cells, photoaffinity crosslinking was carried out. Infected cell nuclear and cytoplasmic extracts were UV-irradiated to crosslink RNA-protein interactions. Total RNA was then purified using RNeasy spin columns (Qiagen). This step essentially purified total RNA, and consequently any protein that had been crosslinked to the RNA. The purified RNA was then subjected to extensive RNAse A digestion to effectively release any bound protein. The digest was then run on a PAGE gel and western blotted for gag. This protocol typically gave RNA yields of around 100μg RNA from the extranuclear fraction and around 15-20μg, from the nuclear fraction, from 1x10^7 infected cells. Figure 6.9A shows a representative gel of total RNA run on 0.8% agarose gel to check the integrity of the purified RNA.

**Figure 6.9 A**

![0.8% Agarose TBE gel. M= Marker, C= Cytoplasmic (extranuclear) RNA, N= Nuclear RNA, + = +UV irradiation, - = -UV irradiation. The Ribosomal RNA bands that are visible are labelled.](image)
Figure 6.9B shows the nuclear and cytoplasmic distribution of gag in the cells used in this experiment confirming that at the time this experiment was carried out p55 gag was in the cell nuclei. To carry out the analysis of the cross-linked RNA-protein isolated, the whole sample representing all the RNA isolated from each fraction was used as described in section 2.2.12 of materials and methods. The photoaffinity crosslinking results are shown in figure 6.9C, essentially these show that p55 co-purifies with total RNA isolated from the extranuclear fraction, and only those fractions that were subjected to UV-irradiation.

**Figure 6.9B**

![Western blot of increasing amounts of the Nuclear (N) and cytoplasmic extracts (C) used in this experiment to show nuclear and cytoplasmic distribution of gag.](image)

**Figure 6.9C**

![Western blot of the UV cross linked extracts following the described preparation. M= molecular weight markers, C= cytoplasmic (extranuclear), N= Nuclear extract (-) denotes no UV cross linking, (+) denotes UV cross linking.](image)
The cross linking of p55 to RNA only occurs in those UV irradiated fractions from the cytoplasm and not the nucleus. This would suggest that the nuclear p55 does not bind RNA in the nucleus and thus probably doesn't play an active part in RNA packaging. This assay does not tell us whether p55 is bound to HIV genomic RNA, merely that p55 is able to cross link to RNA during UV irradiation. From the previous in-vitro studies by ourselves and others it is assumed that p55 is co-purifying with genomic RNA. If p55 formed complexes with other RNAs then there would be no enrichment of the RNA genome in the virion as theoretically any RNA could be packaged.

6.8 Membrane and Cytoplasmic Fractionation

The UV cross linking experiment described (6.7) does not distinguish between those RNA-p55 complexes that are occurring in the cytoplasm and those which may be occurring at the cell membrane as the extranuclear homogenate contained both. It is a matter for debate whether RNA is targeted for packaging by binding p55 in the cytoplasm, or whether the RNA is 'captured' into an assembling particle by p55 that is bound at the plasma membrane.

To address this problem with a view to determining where p55-RNA binding may be occurring the 'cytoplasmic' preparation was further fractionated into cytosol and membrane fractions. Initial attempts at fractionating cytoplasm and membrane was based on high speed (100000g) centrifugations of the crude extranuclear lysate (figure 6.10). In this procedure the membrane should pellet and the cytoplasmic proteins should stay in the supernatant (Spearman et al 1994). However during this procedure the RNA in the sample also pellets (determined by A260 analysis of RNA extracted from each fraction) and was thus unsuitable for the analysis of RNA-p55 interactions.
6.9 Extraction and Analysis of RNA-Protein Interactions Isolated from Membrane and Cytoplasmic Fractions

Fractions were therefore separated using step sucrose density gradient centrifugation. This has been used in many studies looking at various membrane associated viral proteins, such as the membrane association of influenza matrix protein (Zhang and Lamb 1996), and the gag proteins of HIV-1 (Spearman et al 1997, Ono and Freed 1999 and Paillart and Gottlinger 1999). During this procedure cellular membranes (including the plasma membrane) in the homogenate float to form a visible band at the interface between the 65% and 10% w/v sucrose layers. As in the analysis of the nuclear and cytoplasmic fractions, the extranuclear fraction was UV irradiated and loaded at the bottom of a step 80/65/10% w/v sucrose density gradient as detailed in materials and methods. After an 18 hour spin at 100000g 1ml fractions were removed from the top of the gradient and a 100μl aliquot of this removed for analysis by SDS-PAGE and western blotting for gag proteins and to measure the density of each fraction. The fractions were processed as described in materials and methods, essentially RNA was purified from the fractions using RNeasy spin columns, (Qiagen), followed by RNAse A digestion and western blotting as shown in figure 6.11A. Figure 6.11B shows the density profile of the sucrose gradient with the 10%-65% sucrose interface where the membranes accumulate highlighted. Figure 6.11C shows the gag protein western blot of the extracts in each fraction.

Figure 6.10

Western blot showing membrane fraction (M) and cytoplasmic fraction (C) proteins from high speed centrifugation of the extranuclear fraction. This was blotted and probed with p24, p17 and vinculin mAbs.
Summary of method used to purify cytoplasmic and membrane bound gag-RNA complexes

C8166 Cells infected with HIV-1

Homogenisation

UV irradiate extract

Step Sucrose Density Gradient

Membrane fraction

Cytoplasmic fraction

10% Sucrose

65% sucrose

75% sucrose

100000g 18 hours

Total RNA extraction on RNeasy membrane based spin columns

RNase A Digestion

RNase A digested RNA sample loaded onto SDS PAGE and western blotted for gag
Figure 6.11A shows the method used to purify and analyse the RNA-protein interactions in the membrane and cytoplasmic fractions. B shows the density profile of the sucrose gradient, the position of the 10/65% interface was at fraction 6-7. C shows the Western blot for gag and vinculin with the various proteins highlighted. Indicated are those fractions pooled to make the membrane and cytoplasmic fractions.

Figure 6.11C shows the relative amounts of gag protein in each fraction. There are small amounts of vinculin in the plasma membrane, fractions 13-17, as would be expected (vinculin interacts with plasma membrane bound proteins). Of the total p55 gag about 20% seems to be membrane localised, fractions 6,7 and 8. p55 gag was assumed to be plasma membrane localised. The gag protein p49 only appears to be membrane associated, and is not seen in the cytoplasmic fractions. Some p24 is membrane associated, this may
represent p24 that is part of a membrane associated virus, as p24 has no membrane targeting and binding sequence of its own. Cleaved gag proteins also exist in the cytoplasmic fractions. These must be due to cleavages that have occurred in the cytoplasm or represent protein that has dissociated from the membrane. Fractions 6, 7 and 8 were combined to make the membrane fraction, whilst 13, 14, 15 and 16 were pooled to make the cytoplasmic fraction.

The p17 matrix protein on this fraction profile (figure 6.12) shows that, p17 exists in both the membrane and the cytoplasmic fractions as for the other gag proteins, similarly suggesting that this protein is not primarily membrane associated (see discussion). This blot is a reprobe of the that in figure 6.11C and shows a cleaner blot in this experiment.

**Figure 6.12**

![Western blot for p17 MA reprobe of the blot in figure 6.11C, showing the location of p17MA in the infected cell.](image)

**6.10 Confirmation of Membrane Fractionation by gp120/gp160 Distribution**

To show that these fractions are true representations of cytoplasmic membrane and cytoplasm, a small amount of each fraction making up each of the total cytoplasmic and membrane fractions were combined and western blotted for HIV-1 gp120/gp160. gp120 is a cytoplasmic membrane glycoprotein of which the majority, if not all should reside in the cytoplasmic membrane fraction. Ono and Freed (1999) used gp41 as a plasma membrane control in western blotting when they were looking at gag proteins in plasma membrane fractions that had been isolated on sucrose density gradients. Figure 6.13 shows the western blot for gp120/gp160, showing that in these experiments the putative membrane fraction contains these membrane glycoproteins whereas the cytoplasm does not.
In this experiment there is slight distortion in the gp120/160 bands on the above gel due to the way the SDS-PAGE was run. This experiment however shows that cytoplasmic membrane can be purified from cytoplasm using sucrose density gradient centrifugation based on the distribution of the infected cell surface markers gp120/gp160.

6.11 p55 UV Cross Links With RNA From The Cytoplasmic Fraction

After confirmation that the homogenate had separated into the expected fractions by western blotting, the RNA purification procedure and RNAse A digestion was carried out on each pooled fraction. Western blotting for p55 using a p24 mAb was carried out and these results are shown in figure 6.14.

Figure 6.14 Shows the western blot of the RNAse A digested cross linked RNA from the cytoplasm and membrane fractions. C= Cytoplasmic, M= Membrane. (+), UV irradiated, (-), no UV irradiation. Marker sizes are illustrated.
Figure 6.24 shows that p55 only cross links with purified RNA in the cytoplasmic fractions that have been UV irradiated. This has been demonstrated in three separate experiments with the same results. There is no evidence of a p55 specific band in the non-UV treated lanes or the UV irradiated membrane fraction. A further control that shows the specificity of the interaction, is that none of the other gag proteins that can be detected with a p24 mAb are picked up, proteins such as p24, p41 and p49.

This experiment provides the first evidence that selection of packagable genomic RNA by the p55 gag protein occurs at the cytoplasmic level, and may not involve membrane bound p55 interacting directly with the RNA in the cytoplasm.

### 6.12 p55 Gag Binds Genomic RNA and Not Subgenomic RNA

All the work so far described has shown interactions between p55 and RNA. It has given no indication as to what RNA sequence the protein is primarily bound to, as these experiments could have detected binding to cellular RNAs, HIV genomic RNAs or spliced HIV RNAs. To investigate this question, RT-PCR was utilised to analyse the RNA that would co-purify with immunoprecipitated p55 from a cytoplasmic extract. The primer pair AL17 and BM15 were used to amplify singly spliced RNA derived from mature vif message, while the primer pair AL17 and AL18 were used to amplify unspliced RNA representing genomic RNA (see figure 6.15). It might be expected that p55 would bind to genomic RNA but not the spliced message.

**Figure 6.15**

Amplifications using the AL17/AL18 primer pair only amplify genomic RNA sequence giving an expected PCR product size of 437bp. The primer pair AL17 and BM15 will only
amplify vif, spliced message, giving an expected PCR product size of 390bp. Random hexamer primed cDNA was thus PCR amplified to detect each of these messages in the immunoprecipitate (p55 bound) or the supernatant, (unbound cytoplasmic RNA). Figure 6.16 shows the preliminary results of these PCRs.

**Figure 6.16**

![Figure 6.16](image)

No PCR product was amplified in each of the negative controls by either reaction, thus showing no contamination of these reaction sets by contaminating laboratory sequence (lanes 2 and 7) or DNA from the infected cell (lanes 3 and 8). Both of the positive controls amplify cDNA derived from the immunoprecipitate supernatant, which shows that RNA remains intact during the IP step. These PCR products are of the expected size and are assumed to represent the amplified sequences depicted in figure 6.15. PCR of cDNA derived from the immunoprecipitate RNA shows no or extremely low amounts of spliced message sequence (lane 4) while there is a strong band from the PCR which is amplifying cDNA which represents the genomic sequence.

The RT-PCR results suggest that cytoplasmic p55 forms complexes with genomic RNA and by inference this is what is probably seen in the UV cross linking experiments. So gag
binds to the genomic RNA in the cytoplasm as the RT-PCR only detects genomic and not spliced RNA. Low levels of the spliced message were picked up by RT-PCR (extremely faint band in lane 4). These may represent extremely low amounts of spliced RNA specifically bound to p55, but it is also possible that small amounts of non-specific RNA may come through the washing step as contaminants in an immune complex. It is conceivable that small amounts of RNA other than genomic bind p55, as abundant cellular messages such as β-actin have been detected in virus particles. These experiments would ideally lend themselves to Ribonuclease protection analysis which would allow quantification of each RNA species in each sample.

6.13 Summary and Conclusions

6.13.1 HIV-1 p55 Co-purifies With Nuclear and Cytoplasmic Fractions

The nuclear localisation of HIV-1 gag proteins, previously visualised by EM (Royer et al 1991), was confirmed by western blotting crude nuclear and cytoplasmic fractions. This may be a specific process as some proteins appear to be localised only to one compartment, e.g. p49. p24 shows low amounts of nuclear localisation, there are reports that p24 does contain a putative NLS but its role and significance has not been studied. p17 MA is only found in the extranuclear fraction suggesting that matrix alone cannot translocate to the nucleus in productively infected cells, or maybe p17 is degraded in the nucleus. All the other p17 containing gag proteins could be found in the nuclear fractions. This has been seen before in previous studies investigating the nuclear localisation of MoMuLV gag proteins in infected cells (Nash et al 1993). Studies investigating HIV-1 gag and its nuclear involvement have utilised baculovirus or yeast systems expressing p55 gag (Royer et al. 1991, Biemans et al 1992) and these have concentrated on visualisation of gag protein by immunoelectron microscopy. These papers therefore do not distinguish between the different gag proteins that associate with the nuclei.

This investigation has not been carried out before in the case of HIV-1 and it was a surprise that cellular p17 was only found in the extranuclear fraction when, all the other proteins containing the p17 domain can co-fractionate with the nuclei. This must mean that cleaved, mature p17 is somehow restricted from the nucleus while all the other p17 containing proteins are able to translocate there. The proteolysis of p55 gag is thought to take place within the immature virus by the dimerisation and activation of the viral protease, concomitant with budding from the plasma membrane (Kaplan et al 1994). It is entirely feasible that gag could be cleaved by cellular protease’s that recognise the cleavage sites in p55. The intermediate cleavage products seen in these gels (eg figure 6.2A) and in other studies tend to suggest that the viral protease may be functional in the infected cell cytoplasm or maybe at the cytoplasmic side of the plasma membrane. This being the case,
these proteins must be in loose association with the membrane as they are also found in the nuclei. The absence of p17 in the nucleus suggests that cleavage of the p55 protein is not occurring in the nuclei (or we would see this protein), and that p17 is not translocating to the nucleus after cleavage in the cytosol or at the membrane. All other partially cleaved gag proteins in the nucleus contain the p17 domain of the gag protein suggesting that this domain may be important for nuclear targeting. This is in contrast then to what happens in the newly infected cell, Sharova and Bukrinskaya (1991), showed that in the recently infected cell p17 matrix and p17 matrix containing proteins are both transported to the nuclei within 2 hours of the infection. This suggests that there was some change in the targeting of these proteins away from the membrane and into the nucleus.

It is now established that in the newly infected cell, matrix may have a role in the preintegration complex targeting newly reverse transcribed cDNA to the cell nucleus (Bukrinsky et al. 1992, Bukrinsky et al. 1993a, Bukrinsky et al. 1993b, Bukrinskaya et al. 1996). Thus in this case the matrix NLS is active at this point in the virus lifecycle. The viral growth experiments looking at the accumulation of p55 in the nucleus over time failed to pick up any nuclear gag protein that localised there just after infection. This may have been due to insufficient sensitivity of the assay, and the fact that samples were only taken after 3 hours, which was probably too long after complex translocation to the nucleus. This growth experiment showed that p55 localises to the nucleus at the same time as p24 release, and must mean that this nuclear accumulation coincides with efficient virus production, signifying nuclear p55 may have a role in the infected cell.

6.13.2 p55 Accumulates With The Nuclear Fraction Over Time

It was apparent that in long term cultures p55 builds up in the nucleus to levels that eventually exceed that of the extranuclear fraction. During this extended tissue culture infection the cells appeared to stop increasing in numbers, until around day 21 when the cells settle into a chronic infection. At this point the cells show a cytoplasmic to nuclear distribution of gag that is comparable to the early infection. Although this is probably not relevant to an infection in vivo it is plausible that in the chronically infected state the cells are dividing at a normal rate and gag is being released from the nuclei, after the breakdown of the nuclear membrane at cell division. This could be a reason why high amounts of nuclear p55 were not seen during the chronic infection.

6.13.3 Myristylated p55 Can Be Detected In The Nucleus

To understand why gag proteins translocate to the nucleus, it was thought that myristylation may play a role in the nuclear rather than membrane targeting of the gag proteins. As discussed in the introduction, the N terminus of p17/p55 contains a conserved basic region,
and an N-terminal myristate group. This is thought to contribute to the targeting of p55 gag to the plasma membrane, by forming electrostatic contacts with the membrane acidic phospholipids (Zhou et al. 1994). As mutations that block the addition of this moiety to the N-terminus of the polyprotein have the effect of inhibiting particle release (Freed et al. 1994, Gottlinger et al. 1989) by allowing the retention of p55 in the cytoplasm, it is reasonable to suggest that an absence of protein myristylation may allow an active NLS to localise p55 to the nuclei.

In these labelling experiments all the nuclear localising pl7 MA containing gag proteins can still be found in the nuclei as labelled, and by inference, myristylated protein. This is interesting as it means some myristylated p55 can localise to the nucleus possibly due an active NLS which is functional in the presence of the myristate group. This was investigated further by semi quantitative western blots which showed that myristylated p55 actually only accounts for around 20-30% of that of cytoplasmic p55, when the total p55 protein in each fraction is roughly equivalent. Thus lack of myristylation may account for some of the nuclear p55. Recently (see introduction) a myristyl switch mechanism has been proposed, that in addition to myristylation further regulates membrane association. This myristyl switch mechanism proposes that the conformation of the protein at the N-terminus of pl7/p55 acts to sequester the myristate group (Zhou and Resh 1996, Spearman et al. 1997). This is thought to allow the controlled exposure of the myristyl group for insertion into the plasma membrane and consequently prevent the targeting of the protein to any other cellular membrane (Paillart and Gottlinger 1999, Sandefur et al. 1998).

The trigger for the conformational change, or the 'conformational myristyl switch' that allows the exposure of the myristate group may be either phosphorylation (Spearman et al. 1997) or proteolytic cleavage by the viral protease (Hermida-Matsumoto and Resh 1999). It is possible therefore that myristylated nuclear gag protein has been targeted to the nucleus due to the sequestration, by whatever means, of the myristate group and hence has functional nuclear targeting due to the NLS in matrix. Further experiments should be carried out using matrix nuclear localisation signal mutants to confirm that this is the reason for p55 nuclear targeting.

6.13.4 Photoaffinity Cross Linking Detects p55-RNA Complexes in the Cytoplasm but not the Nucleus
Using photoaffinity cross linking to detect general gag-RNA interactions in cellular extracts, this work has shown that p55 is the only gag protein detectable (using p24 and pl7 monoclonal antibodies) which specifically co-purifies with RNA under cross linking conditions. There is no evidence that the nuclear gag complexes with RNA. However the
p55 in the extranuclear fraction does co-purify with RNA after UV cross linking. The cross linking with extranuclear RNA was expected and was the positive control for this experiment.

It is surprising that this protein should not cross link with RNA in the nucleus, as one might expect p55 to bind its cognate RNA in any environment. It was not investigated, but it may be that this subset of protein co-localises with some other cellular protein, preventing RNA binding. This could occur, or it may be that in-vivo co-factors are required for RNA binding that are not present in the nucleus. It is also conceivable that this protein binds RNA very weakly in the buffered environment of the nucleus which is different from that of the cytoplasm. Finally it is also possible that this protein may be anchored in nuclear membranes, and maybe this aggravates RNA binding. This last point is probably unlikely as EM studies in cells expressing gag have shown that gag proteins exist in the intranuclear space in addition to some being present on nuclear membranes (Royer et al 1991).

6.13.5 Analysis of Membrane Bound and Cytoplasmic p55-RNA Interactions

Since the p55-RNA cross linked interactions were detected in the extranuclear fractions which are essentially made up of both cytoplasm and plasma membrane it was desirable to further fractionate this mixed fraction and again look at cross linked RNA-protein interactions. It might be assumed that RNA would cross link with p55 at the plasma membrane, as this is the place where C-type retroviral assembly takes place. However nobody has detected RNA-p55 interactions at this site and it is not known whether the protein interacts with RNA in the cytoplasm, before membrane association.

To fractionate membrane, and hence membrane proteins from cytoplasm, step sucrose density gradients were used as per published protocols. A simple membrane pelleting protocol, was found to pellet all the RNA as well which was not useful for this study. These gradients showed that membrane could be successfully fractionated on the basis of HIV-1 gp120/160 and cytoplasmic vinculin distribution. Using a UV cross linking approach as had been applied to the nuclear and extranuclear fractions, this methodology repeatedly found that p55 only bound to RNA in the cytoplasmic fractions that had been UV irradiated. No evidence for RNA-p55 cross linking was found in the membrane fractions. This last point is interesting, it may mean that membrane bound gag takes no part in the primary RNA-p55 interactions that take place, and that cytoplasmic p55 does. These results perhaps suggest a model whereby some p55 in the cytoplasm binds a packagable RNA, then these proteins associate with further p55 proteins that are already membrane associated.
Recently Lee et al. (1999) have described the formation of detergent resistant gag complexes in the cytoplasm of the infected cell. These are thought to then associate with the plasma membrane and form budding virus structures. The detergent resistant complexes described are less dense than the immature capsid cores and this is possibly a reason why they have not been detected by electron microscopy. It is possible then, that it is these complexes which associate with genomic RNA in the cytoplasm and this is what has been detected in these cross linking experiments.

It is unclear why p55 at the membrane should not co-purify with RNA on cross-linking. It is plausible that on RNA binding at the membrane, virus assembly kinetics are sufficiently fast enough to mean that p55-RNA interactions here are not detectable. This may be because they are either enveloped into a particle that buds from the surface of the cell, or on RNA binding the p55 polyprotein is rapidly cleaved by the viral protease. Due to the presence of a large amount of different gag proteins including matrix in the cytoplasm it is not discounted that membrane bound proteins may dissociate from the membrane over time although this has not been recognised as a cause of high amounts of cleaved gag by those other groups using this methodology on wild type HIV-1 (Spearman et al. 1997, Ono and Freed 1999, Paillart and Gottlinger 1999).

The presence of cytoplasmic p55 in association with RNA may be evidence that the initial point of RNA selection occurs at the level of the cytoplasm and that this complex is then recruited to the site of virus assembly by the membrane bound gag protein.

Indirect evidence supporting this finding has shown that in order for RNA to be efficiently packaged, a cytoplasmic location for the RNA, mediated by rev dependent transport, is needed in cell based packaging assays where rev is provided in trans, (Helg-Maria et al 1999). This is consistent with nuclear p55 not having a direct role in RNA packaging.

### 6.13.6 RT-PCR Analysis Confirms p55-RNA Association

The UV cross linking data gives no idea as to which species of RNA is bound to p55 in these experiments. RT-PCR was used to show that the genomic unspliced message was immunoprecipitated with monoclonal antibodies to p24/p55, while a spliced HIV-1 vif message was not. This is reasonable evidence to show that in these experiments p55 has been associated with the genomic RNA of HIV-1.

These studies show the first biochemical evidence as to how the initial stages of RNA packaging may be taking place in HIV-1 or maybe in any type C retrovirus.
6.13.7 Future In-Vivo Experiments

Future studies in the area of localising the selection and trafficking of packagable RNA, into an assembling particle should utilise fluorescence laser microscopy. It would be interesting to track p55 gag by either antibody localisation, or by using a fusion to Green Fluorescent Protein, in a living cell system, over time. This may give an insight into the targeting of p55 to both the nucleus and to the membrane. Studies such as this could then be complemented by looking at RNA localisation in these cell systems over time with the possibility that RNA and protein interactions could be identified at the same time in the living cell.
Chapter 7

Conclusions

7.1 RNA Dimerisation

The retroviral genome contained in the virion particle is a dimeric structure consisting of two identical RNA molecules. The requirement for two copies of genomic RNA and its dimeric state have been implicated in many important aspects of virion function. These include allowing two extensively nicked RNAs to be held together, allowing a certain degree of genetic recombination, and important roles in reverse transcription and maybe also in RNA packaging. The kissing loop model, describing a method of RNA dimerisation in HIV has been extensively studied both directly in-vitro and indirectly in-vivo. The kissing loop interaction mediated by the autocomplementary sequence in SL1 is clearly important for viral replication, but maybe not necessarily so for dimerisation as dimeric RNA can still be extracted from mutated virion, albeit at a reduced amount to wild type. Another type of approach to see if the kissing loop interaction is important in the mature virion dimer, would be to see if this contact does exist and if so does it confer stability to the virion RNA dimer. In this thesis I have described in-vitro competition assays that were designed to interfere with preformed dimer structures that are thought to form at the dimer initiation site between SL1 domains. Conditions were optimised to allow the dissociation of preformed, in-vitro transcribed, RNA dimers in-vitro. These same conditions were then tested on extracted virion RNA.

The conditions that cause the dissociation of small RNA dimers in-vitro did not cause the dissociation of extracted genomic RNA. This infers that either the dimer linkage structure is not present in the mature RNA, or that competition at this site does not destabilise the virion dimer maybe due to the presence of further points of contact throughout the genomic RNA. In-vivo mutagenesis studies suggest the kissing loop sequence is important for dimerisation but direct proof of their involvement in virion RNA dimers is still lacking. To determine whether dimeric structures formed by SL1-SL1 interactions are important to the virion RNA dimer it would be useful to study this interaction by a more direct biochemical means. The kissing loop structure needs to be analysed to investigate what, (if any) structure exists between SL1 domains in the mature RNA. This could probably be carried out using a chemical modification interference type approach on extracted virion RNA, where analysis of chemically modified virion RNA could be carried out by primer extension. This may yield useful information as to whether structures predicted to form in virion RNA, based on in-vitro
models actually occur in the context of virion RNA and the virus particle. In this respect it would be interesting to see if those sequences predicted to contribute to a guanine tetrad structure \textit{in-vitro} (Sundquist and Heaphy 1993) are present in virion virion dimer and contribute to the number of RNA contacts throughout the genome.

Further points of contact undoubtedly exist throughout the retroviral genome, it would be interesting to see if those interactions proposed between the poly A hairpin (Hoglund \textit{et al} 1997) have any effect on virion RNA dimer stability by carrying out mutagenesis studies on this region. A full deletion type approach of the virion genome \textit{in-vitro} may yield additional regions which contribute to the stability of the genomic RNA dimer \textit{in-vivo}.

\textbf{7.2 RNA–Protein interactions \textit{in-vitro}}

The specific event allowing the packaging of a genomic RNA into a virus particle is mediated by the interaction between the p55 gag protein and sequences in the genomic RNA. This interaction between the virion RNA and the p55-gag was studied \textit{in-vitro} using GST-p55 and HIV-1 leader RNA progressively truncated by deletion analysis. The aim was to define a minimal binding site for the gag protein. This study found that multiple sites probably bind gag individually and, in the context of the leader act to bind a high localised concentration of p55 gag which maybe precipitates virus particle assembly. One of these binding signals, SL4, which comprised the minimal recognisable binding site, acts to bind p55 in a dominant way in some RNAs, but is not the only element binding gag in longer RNA, as was confirmed by ribonuclease footprinting. This must mean that the RNA binding signal in the 5' leader is multipartite, consisting of several elements that each individually bind p55. This has been shown in previous studies using virion packaging systems. Virion approaches have identified these sequences and others, which were not identified in this study e.g. the tar hairpin. The \textit{in-vitro} binding approach should yield specific sites which can be targeted by mutagenesis of viral clones. This approach would allow the specific nucleotide residues which may interact with gag to be targeted specifically rather than using the unsubtle approaches carried out by other groups which have used larger deletions in virus based packaging systems. Larger deletions in virions probably have indirect effects on packaging by virtue of the multiple roles RNA structures from the 5'UTR probably have. Differences between \textit{in-vitro} biochemical studies and \textit{in-vivo} mutagenesis studies could be due in part to other, as yet unidentified cellular co-factors which may contribute to RNA-protein interactions and RNA packaging \textit{in-vivo}. 

165
This work found no evidence for the preferential binding of protein to RNA kissing-loop dimer, which agrees with in-vivo evidence showing that monomeric RNA can be packaged into viral particles.

The analysis of p55-RNA interactions in-vitro needs to be expanded by investigations on RNA binding using p55 gag, with preparations such as I have been able to obtain from p55 expression in E. coli cells. It would be interesting to look at the effect of cellular extracts on the in-vitro binding activity of the protein as maybe unidentified cellular factors are involved in the interaction. This preparation of p55-gag would be useful in an extensive footprinting analysis of the viral leader sequence. This would provide information on protein binding sites in the RNA leader that could be used in an in-vivo mutagenesis study.

7.3 RNA-Protein interactions in-vivo

Before these experiments were carried out the cellular location of the protein-RNA interactions involved in genomic RNA packaging were not known. As p55 is synthesised in the cytoplasm on free ribosomes and then translocates to the plasma membrane it was possible that RNA binding could occur at either of these locations. During this study I have shown that p55 and some of the processing intermediates colocalise to the nucleus of infected cells. Analysis of these gag proteins by UV photoaffinity cross linking suggests that nuclear p55 probably doesn't take part in RNA binding and by implication, packaging. The localisation of p55 to the nuclei of infected cells may be due to the nuclear localisation signal which is present in the matrix domain of the protein, this suggestion is purely speculative and deserves further study. Both myristylated and non-myristylated forms of p55 were found in the nuclear fractions, suggesting that lack of myristylation does not govern nuclear targeting.

Using UV-cross linking and step sucrose gradient centrifugation, cytoplasmic (including plasma) membranes were purified from cytoplasm, and the p55-RNA interactions were analysed. These experiments showed that p55 interacts with RNA in the cytoplasm and this interaction could not be detected in these experiments in the membrane fraction. The fact that no interaction could be detected in the membrane fraction may be due to the virus assembly kinetics, as it is possible that on binding to the membrane an RNA-protein complex rapidly buds to form a virus particle.

The finding that p55 interacts with RNA in the cytoplasm is novel and perhaps surprising, presumably meaning that a p55-RNA complex is then recruited to the site of virus assembly at the plasma membrane. The translocation of this complex may be
brought about by interactions between gag and cytoskeletal proteins such as actin. RT-PCR was used to show that the interaction between p55 and RNA detected in these experiments was specific as a spliced viral message, vif, was not detected in a gag immunoprecipitate, whereas genomic RNA sequence was. These in-vivo interactions are summarised in figure 7.1.

**Figure 7.1 In-Vivo RNA packaging summary**

1) Gag proteins including p55, proteolytic processing intermediates and mature gag proteins can be detected in the cytoplasm, nucleus and membrane fractions of the infected cell. It was interesting to find cleaved gag proteins in the cellular fractions as these are generally associated with mature virions.

2) Substantial amounts of gag proteins, including p55 accumulate in the nucleus of the infected cell. These proteins accumulate over time suggesting that they may not be involved in RNA transport. No evidence could be found to suggest that p55 bound RNA in the nuclear fractions, suggesting no role in packaging or RNA selection. Both myristylated and non-myristylated forms of p55 gag were found in nuclear fractions, although the myristylated form was underrepresented.
3) p55 gag was found to bind RNA in the cytoplasmic fractions, suggesting that this binding intermediate then goes to associate with p55 gag residing at the membrane (4). As no p55-RNA interactions were detected in membrane fractions, this maybe suggests that the assembly kinetics are extremely rapid on the envelopment of a genome and therefore are not detected.

The logical progression of this work, would be to analyse the RNA-p55 interaction in-vivo using microscopic techniques. It might be possible to look at p55 and RNA using confocal laser microscopy of infected cells, probing p55 with fluorescent monoclonal antibody and HIV-1 genomic RNA with a labelled oligonucleotide probe. Co-localisation of the two probes should occur in the cytoplasm of the cell confirming my biochemical results. It would also be useful to use in-vitro models to look at p55-RNA interactions in cell extracts, to try and identify the type of complexes that exist in the cytosol before particle budding. Experiments such as this could be developed into model systems where reconstitution experiments could look at the effect of specific components on in-vitro particle formation and packaging, for example, presence or absence of protein myristylation.

7.4 Genome Packaging in other Viruses

HIV-1 genomic RNA packaging is perhaps one of the most extensively studied mammalian virus packaging system to date. This is mainly due to the interest RNA packaging has as a retrovirus antiviral target, and the fact that lentiviruses could potentially be useful as vectors for gene transfer. Packaging in other virus systems has also been investigated and shows that viruses can share similar or have very diverse packaging strategies. The following virus examples are described for illustrative purposes.

The Alphavirus family including Sindbis virus, Ross-river virus and Semliki forest virus, all possess a positive sense RNA viral genome. Like HIV, subgenomic message has to be excluded from the packaging process. These viruses package their genome via interactions between the virus capsid protein and sequences in the 5' region of the genomic RNA. Again like in retroviruses this interaction is thought to involve RNA secondary structure recognition. The location for this interaction is thought to take place in the cytoplasm of the cell, forming a capsid structure. This then migrates to the cell membrane and exits the cell by budding. This specific RNA-protein recognition event therefore contrasts to another RNA virus family, the Picornaviruses, which include polio virus. The polio virus genomic RNA is thought to be packaged into a preassembled virus core either by threading through a pore structure in the virus core, or by binding to channels on the outer surface of the structure. It is thought that
binding to an outer surface channel may trigger internalisation of the RNA and then subsequent maturation of the capsid. Sequences important for RNA packaging in picornaviruses are at present not well defined.

DNA viruses such as Adenovirus also package their genome into a pre-formed or partially formed capsid in the nuclei of the infected cell. It is thought that the DNA genome of Adenoviruses is packaged by recognition of AT-rich repeats approximately 250bp into the viral chromosome. These sequences are thought to create DNA bending which may be recognised by packaging proteins or the partially assembled capsid. This theme is broadly similar in Herpes viruses which also package their DNA genome into a nuclear assembled viral core. With this virus two specific recognition elements are recognised within the virion genome by viral and, or cellular proteins, one of which appears to be a nuclease. DNA corresponding to one genome size is packaged by recognition of a similar sequence the equivalent of one genome away. Genomic DNA is cleaved from concatomeric genome and the genome is packaged.

These strategies have been briefly described to show the diversity of mechanisms in genome selection and packaging within mammalian viruses. While a packaging strategy may be broadly conserved within a virus family there seem to be large differences between the different virus families.
References


Bukrinsky, MI., Sharova, N., McDonald, TL., Pushkarskya, T., Tarpley, WG and Stevenson, M. (1993b) Association of integrase, matrix and reverse transcriptase
antigens of human immunodeficiencyvirus type 1 with viral nucleic acids following acute infection. Proceedings of the National Academy of Science USA. vol. 90: p6125-6129.


Liang, C., Rong, L., Quan, Y., Laughrea, M., Kleiman, L. and Wainberg, M.A. (1999) Mutations within four distinct gag proteins are required to restore replication of human immunodeficiency virus type 1 after deletion mutagenesis within the dimerisation initiation site. Journal of Virology. vol.73: p7014-7020.

Linial, M., Medeiros, E. and Hayward, W.S. (1978) An avian oncorenretovirus mutant


