

Characterisation of RNA aptamers that bind to HIV-1 gp120



Carla Cohen

Sir William Dunn School of Pathology
and
Wolfson College

A thesis submitted for the degree of Doctor of Philosophy
at the University of Oxford, Trinity Term 2006

Characterisation of RNA aptamers that bind to HIV-1 gp120

Carla Cohen

Sir William Dunn School of Pathology and Wolfson College
A thesis submitted for the degree of Doctor of Philosophy at the
University of Oxford, Trinity Term 2006

Abstract

RNA aptamers with 2'-fluoro-pyrimidine chemistry were previously selected by *in vitro* evolution to bind to monomeric HIV-1 gp120 from the R5 strain BaL. A group of 36 novel aptamers were cloned and sequenced from the heterogeneous pool and were tested for their ability to bind to gp120. The diversity of the RNA secondary structure of these, and 27 aptamers isolated previously, was analysed using a bioinformatics approach. This showed that eight aptamers contain a common branched motif, and RNA mutagenesis indicated that this structure is probably required for gp120 binding. Chemically synthesised derivatives of one such aptamer, B40, were designed and tested for binding to gp120. Truncation was found to decrease their binding, but the introduction of point mutations to stabilise the branched conformation and 2'-O-dimethylallyl-modified residues to stabilise helices increased binding to levels greater than that of the parental aptamer. The aptamer epitope on gp120 was mapped by testing aptamer binding to alanine-scanning mutants and deletion mutants of gp120 using a novel plate-based assay. This study showed that the aptamer binding site overlaps with the CCR5 epitope and is confined to four key residues at the base of the V3 loop, one of which is highly conserved. This finding may account for the observation that a number of aptamers were shown previously to neutralise a range of HIV-1 R5 clinical isolates in PBMC cultures. Interestingly however, the aptamer was unable to neutralise HIV-1 pseudovirus in a cell line, which is most likely due to the increased levels of cell-surface CCR5 in

cell lines compared to PBMC. Future work should focus on identifying the structure and epitopes of other anti-gp120 aptamers as well as testing neutralisation of HIV-1, HIV-2 and SIV by the B40-derived aptamers. These aptamers can be used as tools to investigate the HIV-1 entry pathway and also have the potential to be developed as anti-HIV-1 microbicides.

Acknowledgements

I would like to thank my supervisor, William James, for all of his help, direction and support throughout my PhD. I have been fortunate to work with over 20 post-docs, graduate students and undergraduates in the James lab and I thank them all for the time spent discussing everything from the details of a PCR to the meaning of research. I have learnt a lot during my time here from their diverse personalities and wide-ranging interests. Most importantly I would like to thank Shooz Khati and Antu Dey for all of their help and instruction when I joined the lab and from whose work and understanding I have been able to develop my own DPhil project.

I am indebted to a number of other people in the Dunn School and elsewhere for providing reagents and giving advice: Mike Puklavec for assistance with tissue culture; Alice Kearney for help with BIAcore; Nigel Rust for help with flow cytometry; Simon McGowan, Rune Lyngsø and Jotun Hein for help with bioinformatics; Debbie Hatherley and Nick Clarkson for providing plasmids and help with cloning; Ralph Pantophlet for providing pseudovirus plasmids and many emails troubleshooting techniques; Ally Evans and Lucinda Risius for sorting out all my admin; and everyone else who works in the CIU for making it an friendly place to work. I would also like to thank Herman Waldmann for giving me the opportunity to work in such a great department, in an environment that has enabled me to develop personally and professionally as well as scientifically.

In joining Wolfson College I have got to know many friends who have been there to help when things were tough and to celebrate when things were going well. Thanks especially to Emma, James and Rhian.

Thanks very much to my family, Mum, Dad and Victoria who have continued to give me advice, support (and money!) throughout my long education, and who have learnt to ask the right questions at the right time. I promise that one day I will sit down and explain it all properly.

Finally and most importantly I would like to thank my husband, Richard, who gallantly came to live in Oxford so that I could do my PhD in the first place, and has been amazing at looking after me and making sure I got to the end. Richard, I couldn't have done it without you.

Freddie Mercury died in November 1991, and this was the first time I really understood what HIV and AIDS were. This DPhil is dedicated to his memory.

“... I wish to confirm that I have been tested HIV positive and have AIDS ... the time has come now for my friends and fans around the world to know the truth and I hope that everyone will join with my doctors and all those worldwide in the fight against this terrible disease.”

Freddie Mercury, November 23rd 1991

Last public statement, made one day before his death

Abbreviations

Ab	Antibody
AIDS	Acquired immunodeficiency syndrome
AZT	Azidothymidine
Biotin-16-ddUTP	biotin- ϵ -aminocaproyl- γ -aminobutyryl-[5-(3-aminoallyl)-2', 3'-dideoxy-uridine-5'-triphosphate] tetralithium salt
CA	Capsid
CD4i	CD4-induced (site)
CD4L	CD4 leader
CM5	Carboxymethylated 5 (chip)
CHO	Chinese hamster ovary
CO	Codon optimised
CTL	Cytotoxic lymphocyte
DC	Dendritic cell
DC-SIGN	Dendritic cell-specific ICAM-grabbing nonintegrin
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
dNTP	Deoxynucleotide 5'-triphosphate
EDC	N-ethyl-N'-(3-dimethylaminopropyl) carbodiimide
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
FACS	Fluorescence-activated cell sorting
FCS	Fetal calf serum
FITC	Fluorescein isothiocyanate
FRET	Fluorescence resonance energy transfer
GAG	Glycosaminoglycan
GALT	Gut-associated lymphoid tissue
gp	Glycoprotein
gp120-BH	gp120-biotinylation site-His tag fusion protein
gp120-CB	gp120-rat CD4 domains 3 & 4- biotinylation site
HAART	Highly active anti-retroviral therapy
hCD4	Human CD4
HEK 293T	Human embryonic kidney 293T cells
HEPES	N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
HIV-1	Human immunodeficiency virus type 1
HIV-2	Human immunodeficiency virus type 2
HRP	Horseradish peroxidase
HSPG	Heparan sulfate proteoglycans
hCD4	Human CD4
Ig	Immunoglobulin
IL-2	Interleukin-2

IN	Integrase
LTR	Long terminal repeat
MA	Matrix
MDS	Multidimensional scaling
MFE	Minimum free energy
MOI	Multiplicity of infection
MSX	Methioninesulfoxamine
NC	Nucleocapsid
NHS	N-hydroxysuccinimide
NIH ARRRP	National Institute of Health AIDS Research and Reagent Reference Program
ntd	Nucleotide
PAGE	Polyacrylamide gel electrophoresis
PBMC	Peripheral blood mononuclear cell
PBL	Peripheral blood lymphocyte
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PE	Phycoerythrin
PEG	Polyethylene glycol
PEI	Poly(ethylenimine)
PHA	Phytohemagglutinin
PIC	Pre-integration complex
PR	Protease
PVDF	Polyvinylidene fluoride
qPCR	Quantitative PCR
Rantes	Regulated-upon activation, normal T0cell expressed and secreted
rCD4d3&4	Rat CD4 domains 3 and 4
RRE	Rev response element
RNA	Ribonucleic Acid
RPMI-1640	Roswell Park Memorial Institute 1640 medium
RT	Reverse transcriptase
SA	Streptavidin
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SIV	Simian immunodeficiency virus
SPR	Surface plasmon resonance
TAR	Tat activation region
TBS	Tris buffered saline
TCS	Tissue culture supernatant
TEMED	N, N, N, N'-tetramethylethylenediamine
TMB	3,3',5,5'-Tetramethylbenzidine
WHO	World Health Organisation
X-gal	5-bromo-4-chloro-3-indoyl- β -D-galactopyranoside

Table of Contents

CHAPTER 1: INTRODUCTION	5
1.1 OVERVIEW OF HIV-1 BIOLOGY	6
1.1.1 <i>Epidemiology</i>	6
1.1.2 <i>HIV-1 genome organisation</i>	7
1.1.3 <i>HIV-1 structure</i>	10
1.1.4 <i>HIV-1 life cycle</i>	10
1.1.5 <i>Gp120 structure</i>	14
1.1.6 <i>Transmission and disease progression</i>	18
1.1.7 <i>Immune evasion</i>	21
1.1.8 <i>Current strategies for HIV-1 prevention and treatment</i>	22
1.2 APTAMERS	29
1.2.1 <i>Aptamer selection</i>	29
1.2.2 <i>Applications of aptamers</i>	33
1.2.3 <i>Anti-gp120 aptamers</i>	36
1.2.4 <i>Surface Plasmon Resonance</i>	37
1.3 RNA SEQUENCE AND STRUCTURE.....	39
1.3.1 <i>RNA sequence alignment</i>	40
1.3.2 <i>RNA structure prediction</i>	41
1.3.3 <i>Comparison of RNA structures</i>	43
1.4 AIMS AND SCOPE OF THE THESIS	45
CHAPTER 2: MATERIALS & METHODS	48
2.1 MATERIALS	48
2.1.1 <i>General suppliers</i>	48
2.1.2 <i>General materials & suppliers</i>	49
2.1.3 <i>Monoclonal antibodies</i>	52
2.1.4 <i>Polyclonal antibodies</i>	52
2.1.5 <i>Buffers and solutions</i>	53
2.2 MANIPULATION OF NUCLEIC ACIDS.....	54
2.2.1 <i>Polymerase Chain Reaction</i>	54
2.2.2 <i>2'-fluoro in vitro transcription</i>	56
2.2.3 <i>Agarose gel electrophoresis</i>	56
2.2.4 <i>Denaturing polyacrylamide gel electrophoresis</i>	58
2.2.5 <i>Native polyacrylamide gel electrophoresis</i>	58
2.2.6 <i>Phenol:Chloroform extraction of nucleic acid</i>	59
2.2.7 <i>Ethanol precipitation of nucleic acid</i>	59
2.2.8 <i>Removal of free nucleotides</i>	60
2.2.9 <i>Spectrophotometry</i>	60
2.2.10 <i>Gel purification of RNA</i>	60
2.2.11 <i>5' end-labelling of RNA with [γ-^{32}P] ATP</i>	61
2.2.12 <i>3' end-labelling of RNA with biotin-16-ddUTP</i>	61
2.2.13 <i>Aptamer refolding</i>	62
2.3 CLONING TECHNIQUES.....	62
2.3.1 <i>Restriction digests</i>	62
2.3.2 <i>Alkaline phosphatase treatment</i>	63
2.3.3 <i>Gel purification of DNA fragments</i>	63
2.3.4 <i>Ligation of DNA fragments</i>	63
2.3.5 <i>Transformation of competent cells</i>	64
2.3.6 <i>DNA sequencing</i>	64
2.3.7 <i>Amplification of plasmid DNA</i>	65
2.4 TISSUE CULTURE TECHNIQUES	65

2.4.1	Maintenance of cell lines.....	65
2.4.2	Preparation and culture of PBL.....	66
2.4.3	Flow cytometry.....	66
2.4.4	Transient transfections.....	67
2.4.5	Desulphation and removal of cell-surface heparan sulphate proteoglycans ..	68
2.5	EXPRESSION OF GP120.....	69
2.5.1	Expression of gp120 fusion proteins in 293T cells.....	69
2.5.2	Concentration of supernatant.....	70
2.5.3	Biotinylation of gp120-CB	70
2.5.4	Other protein sources.....	70
2.6	USE OF HIV-1 PSEUDOVIRUS	71
2.6.1	Generation of HIV-1 pseudovirus	71
2.6.2	Transduction of target cells.....	71
2.6.3	Neutralisation of HIV-1 pseudovirus	72
2.6.4	Detection of luciferase activity.....	72
2.6.5	Detection of HIV-1 DNA by quantitative PCR.....	73
2.7	PROTEIN TECHNIQUES.....	75
2.7.1	SDS polyacrylamide gel electrophoresis.....	75
2.7.2	Western blotting	75
2.7.3	Quantitation of bands using ImageQuant	76
2.7.4	Gp120 ELISA	76
2.7.5	Inhibition ELISA.....	77
2.7.6	Competition assays.....	78
2.7.7	Slot-blot.....	79
2.7.8	Plate-based assay for aptamer-gp120 interaction	79
2.7.9	Immunoaffinity chromatography.....	80
2.8	STATISTICAL TECHNIQUES	80
2.8.1	Nonlinear regression.....	81
2.8.2	Derivation of sample concentration.....	81
2.8.3	T-tests	82
2.9	COMPUTATIONAL METHODS	83
2.9.1	Modelling aptamer structures with mfold	83
2.9.2	RSmatch implementation.....	83
2.9.3	ClustalW alignment and phylogenetic trees	85
2.9.4	Sfold implementation.....	85
2.10	SURFACE PLASMON RESONANCE	86
2.10.1	Immobilisation.....	86
2.10.2	Binding analysis	88
2.10.3	Chip regeneration	89
CHAPTER 3: GP120 EXPRESSION.....		91
3.1	INTRODUCTION	91
3.2	EXPRESSION OF GP120 FUSION PROTEINS FROM pEF-BOS VECTORS	92
3.2.1	Cloning the gp120 gene into pEF-BOS expression vectors.....	94
3.2.2	Transient transfection of 293T cells.....	97
3.3	MAKING A STABLE CELL LINE EXPRESSING GP120-CB	101
3.3.1	Cloning the gp120-CB construct into pEE14.....	102
3.3.2	Selection of positive clones	105
3.4	CODON OPTIMISATION	105
3.4.1	Cloning of codon-optimised gp120 into pEF-BOS-CB	107
3.4.2	Comparison of expression.....	107
3.5	OPTIMISATION OF TRANSFECTION METHOD	113
3.6	COMPARISON OF EXPRESSION METHODS	115
3.6.1	Comparison by Western blot.....	116
3.6.2	Comparison of antibody binding using SPR	116

3.6.3	Comparison of hCD4, aptamer 299.5 and CCR5 peptide binding by SPR ...	123
3.7	DISCUSSION	128
CHAPTER 4: APTAMER BINDING AND STRUCTURE		131
4.1	INTRODUCTION	131
4.2	CLONING AND SEQUENCING OF NOVEL ANTI-GP120 APTAMERS	133
4.3	APTAMER BINDING TO GP120.....	135
4.4	COMPARISON OF RNA STRUCTURES.....	141
4.4.1	Experimental design.....	141
4.4.2	Optimisation of RSmatch parameters.....	144
4.4.3	Data analysis.....	146
4.4.4	Network analysis	154
4.4.5	Analysis of RNA structures within a population.....	154
4.4.6	Alignment of loop sequences	162
4.5	DISCUSSION	165
CHAPTER 5: APTAMER MINIMISATION.....		170
5.1	INTRODUCTION	170
5.2	METHODS	171
5.2.1	Design of synthetic aptamers	171
5.2.2	Aptamer binding assay.....	171
5.3	STABILISATION OF THE BRANCHED FORM OF B40t77	172
5.4	SHORTENING THE STEM	175
5.5	REMOVAL OF AA MISMATCH.....	175
5.6	SHORTENING STEM WITH REMOVED AA MISMATCH	178
5.7	STABILISATION OF BRANCHED FORM OF TRUNCATED APTAMERS.....	178
5.8	INTRODUCTION OF 2'-O-DIMETHYLALLYL GROUPS	180
5.9	COMBINATION OF 2'-O-DIMETHYLALLYL MODIFICATIONS AND STABILISATION OF THE BRANCHED FORM	183
5.10	DISCUSSION	185
CHAPTER 6: DEFINITION OF THE APTAMER BINDING SITE ON GP120.....		190
6.1	INTRODUCTION	190
6.2	METHODS	192
6.2.1	Design of a synthetic biotinylated aptamer B40 derivative.....	192
6.2.2	Production of gp120 mutants	193
6.3	ANALYSIS OF GP120-APTAMER INTERACTION BY SPR.....	194
6.4	ANALYSIS OF APTAMER-GP120 INTERACTION BY AN ELISA-TYPE METHOD.....	196
6.4.1	Optimisation of an aptamer-gp120 binding assay	196
6.4.2	Estimation of gp120 concentration by ELISA	199
6.4.3	Estimation of gp120 concentration by Western blot	201
6.4.4	Interaction of gp120 mutants with aptamer 299.5	203
6.5	DISCUSSION	205
6.5.1	Summary of results.....	205
6.5.2	Choice of methods.....	205
6.5.3	Involvement of gp120 variable loops	206
6.5.4	Specific gp120 residues involved in aptamer-gp120 interaction	207
6.5.5	Model of aptamer binding to gp120.....	209
6.5.6	CD4 independence of aptamer binding.....	211
6.5.7	Sequence conservation in gp120	212
6.5.8	Future work.....	213
6.5.9	Conclusions	214

CHAPTER 7: ATTEMPTS TO GROUP APTAMERS BY COMPETITION	
BINDING ASSAYS.....	216
7.1 INTRODUCTION	216
7.2 OPTIMISATION OF A COMPETITION ASSAY	218
7.2.1 <i>Competition assay with radiolabelled aptamer</i>	218
7.2.2 <i>Competition with biotinylated aptamer</i>	220
7.2.3 <i>Competition assay using immobilised aptamer</i>	222
7.2.4 <i>Competition with antibodies</i>	224
7.3 INTERACTION OF OTHER APTAMERS WITH GP120 MUTANTS.....	226
7.3.1 <i>Slot-blot method</i>	230
7.3.2 <i>Plate-based method</i>	230
7.4 DISCUSSION	232
CHAPTER 8: USE OF PSEUDOVIRUSES TO EVALUATE APTAMER	
NEUTRALISATION	235
8.1 INTRODUCTION	235
8.2 APTAMER NEUTRALISATION OF PSEUDOVIRUSES	236
8.2.1 <i>Optimisation of the neutralisation assay</i>	236
8.2.2 <i>Desulphation and removal of cell-surface heparan sulphate proteoglycans</i>	238
8.2.3 <i>Neutralisation of pseudovirus in PBLs</i>	241
8.2.4 <i>Detection of HIV-1 pseudovirus DNA by qPCR</i>	243
8.3 DISCUSSION	246
CHAPTER 9: GENERAL DISCUSSION.....	250
9.1 SUMMARY OF RESULTS	250
9.2 ASPECTS OF GP120 STRUCTURE	250
9.3 STRUCTURAL ANALYSIS OF ANTI-GP120 APTAMERS	251
9.4 EVIDENCE FOR THE APTAMER BINDING SITE ON GP120	253
9.5 APTAMERS AS DIAGNOSTIC TOOLS.....	255
9.6 THERAPEUTIC APPLICATIONS OF APTAMERS	256
9.6.1 <i>Screening for small molecule inhibitors</i>	256
9.6.2 <i>Structure-based drug design</i>	256
9.6.3 <i>Aptamers as microbicides</i>	257
9.7 CONCLUDING REMARKS.....	257
BIBLIOGRAPHY	259
APPENDIX 1: APTAMER BINDING TO THE GP120-CB FUSION PROTEIN.....	277
APPENDIX 2: EXTENSIVE BIOINFORMATICS RESULTS.....	278

Chapter 1: Introduction

- Overview of HIV-1 biology
- Introduction to aptamers
- RNA sequence and structure
- Aims and scope of thesis

Chapter 1: Introduction

This thesis describes the characterisation of aptamers that bind to gp120, their uses as tools to investigate the HIV-1 entry pathway and their potential as therapeutic agents. By way of introduction I will describe relevant features of HIV-1 biology, outline current aptamer technology and applications and give details of techniques pertinent to the analysis of RNA structure.

1.1 Overview of HIV-1 biology

1.1.1 Epidemiology

We are in the midst of an HIV-1 pandemic: statistics from the World Health Organisation (WHO) show that at the end of 2005, 40.3 million people were infected with HIV-1. In 2005 alone there were 4.9 million new infections and 3.1 million AIDS-related deaths (UNAIDS/WHO 2005). Since the first identification of an unusual incidence of pneumonia in homosexual men in 1981 (Centers for Disease Control 1981) and the isolation of HIV-1 in 1983 (Barre-Sinoussi *et al* 1983) our understanding of HIV-1 biology has developed extensively. However, despite the discovery of a number of anti-retroviral drugs, a vaccine has proved elusive and the pandemic continues.

Retroviruses are characterised by their RNA genome, which is reverse transcribed to DNA and integrated into the host genome to provide a template for the production of new virus particles. HIV-1 is one member of the lentivirus subfamily of retroviruses, which also includes HIV-2, simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), equine infectious anaemia virus (EIAV) and caprine arthritis encephalitis virus (CAEV). These viruses are characterised by their rod- or cone-shaped core and their ability to infect non-dividing cells.

Phylogenetic trees based on *env* gene sequences show that HIV-1 is more closely related to SIV from chimpanzees (SIV_{cpz}) than to HIV-2 or SIV from macaques (SIV_{sm}). There are three groups of HIV-1, M (main), N (not M, not O) and O (outlier), each of which is thought to represent an independent crossover event from chimpanzees to humans (Gao *et al* 1999). Most circulating HIV-1 populations are group M, and this group is subdivided into clades A, B, C, D, F, G, H and J. The worldwide distribution of these clades varies, with clades C and B predominating in Africa, and clade B predominating in Europe and the USA (reviewed in Thomson *et al* 2002). In addition, a number of circulating recombinant subtypes have been generated by the infection of one individual with more than one clade of virus, and multiple quasispecies (related viruses that have sequence variation) exist within each patient. The huge genetic variation of HIV-1 contributes to its evasion of the immune response and has implications for the development of drugs and vaccines.

1.1.2 HIV-1 genome organisation

The organisation of the HIV-1 genome is shown in Figure 1.1a. The genome is approximately ten kilobases in length and consists of nine genes that encode fifteen proteins. All retroviruses contain the genes *gag*, *pol* and *env*, but HIV-1 also has genes for a number of auxiliary factors (Li *et al* 2005). The genome is flanked by 5' and 3' long terminal repeats (LTR) that promote and regulate HIV-1 transcription and are required for integration of the HIV-1 cDNA into the host genome. Each viral protein has unique functions that are summarised in Table 1.1. The *pol* and *gag* genes are expressed as polyproteins that are subsequently cleaved in the virion by the viral protease. The *gag* gene is most commonly expressed as a 55 kDa polyprotein, Pr55^{Gag}, but in 5-10 % of cases a frameshift mutation occurs resulting in the expression of the 160 kDa Pr160^{GagPol} polyprotein. This mechanism regulates the proportions of PR, RT and IN proteins that are produced. The *env* gene is

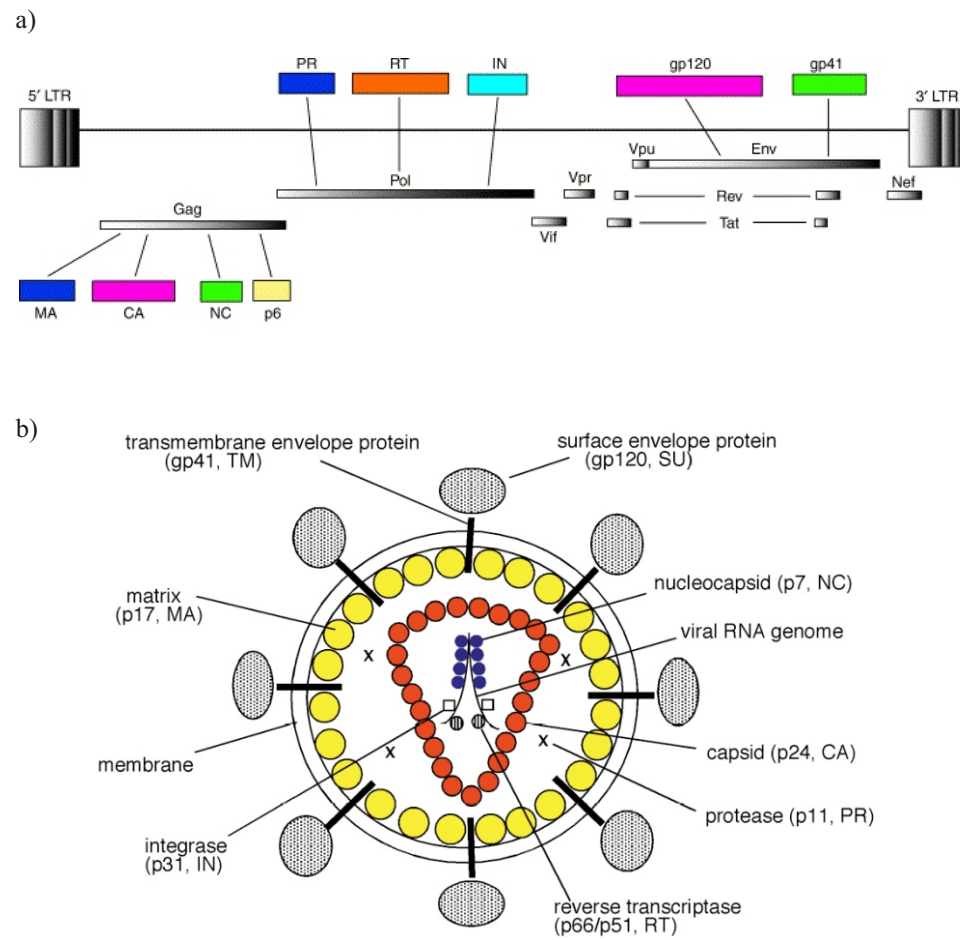


Figure 1.1:

- a) Structure of the HIV-1 genome.
 b) Schematic representation of an HIV-1 virion.
 Adapted from Freed, 1998 and 2004.

Gene	Encoded protein(s)	Function
<i>gag</i>	Matrix (MA)	Targeting of Pr55 ^{gag} and Pr160 ^{GagPol} to plasma membrane Nuclear import of the PIC Lining of inner surface of virus membrane Infection of non-dividing cells
	Capsid (CA)	Structural role in virion Viral assembly Role in uncoating through binding cyclophilin A
	Nucleocapsid (NC)	Packaging of RNA into virion Increasing stability of RNA within viral core Enhancement of other nucleic-acid dependent steps
	p6	Incorporation of Vpr into virion
<i>pol</i>	Protease (PR)	Cleavage of polyproteins Virion maturation
	Reverse transcriptase (RT)	Reverse transcription
	Integrase (IN)	Integration
<i>vif</i>	Viral infectivity factor (Vif)	Counteraction of cellular restriction factors Stimulation of reverse transcription
<i>vpr</i>	Viral protein r (Vpr)	Weak stimulation of viral gene expression Cell cycle arrest in G ₂ Nuclear import of the viral PIC
<i>vpu</i>	Viral protein u (Vpu)	Enhancement of viral particle release Degradation of CD4
<i>rev</i>	Regulator of Expression of the Virion (Rev)	Nuclear export of viral mRNA Increased stability of viral mRNA
<i>env</i>	Env	Downregulation of CD4
	gp120 (SU)	Attachment of virion to the host cell
	gp41 (TM)	Fusion of viral and cellular membranes
<i>tat</i>	Transactivator of Transcription (Tat)	Promoter of viral transcriptional elongation Induction of apoptosis
<i>nef</i>	Negative factor (Nef)	Suppression of viral gene expression Downregulation of cell-surface CD4 and MHC class I Enhancement of viral infectivity Interference with host-cell signal transduction

Table 1.1: Functions of viral gene products.

also expressed as a polyprotein (Env or gp160) and this is cleaved to gp120 and gp41 (also called SU and TM) by a cellular protease during its trafficking to the surface.

1.1.3 HIV-1 structure

A schematic depiction of the structure of an HIV-1 particle is shown in Figure 1.1b. The core of the virus is made from the capsid protein (CA), inside which is the viral RNA, integrase (IN), protease (PR) and reverse transcriptase (RT). Two copies of the viral RNA genome are stabilised by the interactions with the nucleocapsid (NC) protein. The capsid is surrounded by a phospholipid membrane that is formed when the virus buds from the host cell. The membrane contains the viral proteins gp120 and gp41 as well as a number of host cell proteins that are taken up in the budding process. The matrix protein (MA) associates with the inner surface of the viral membrane through its myristic acid modification.

1.1.4 HIV-1 life cycle

Over the last 25 years a detailed understanding of the HIV-1 life cycle has been established, and this knowledge has been crucial in the development of antiviral drugs. The HIV-1 life cycle has been reviewed in detail (for example in Frankel and Young 1998, Freed 2001, Greene and Peterlin 2002), and here I will provide a summary of the main events as outlined in Figure 1.2.

The discoveries of CD4 (Dalglish *et al* 1984), CXCR4 (Feng *et al* 1996) and CCR5 (Deng *et al* 1996) as cellular receptors of HIV-1 represented significant milestones in understanding the viral life cycle. HIV-1 entry to the cell first requires binding of the gp120 surface glycoprotein to cell-surface CD4 (described in more detail in Section 1.1.5). This interaction causes conformational changes in gp120 that allow it to bind concurrently to the coreceptor (Figure 1.3). *In vitro* studies have shown that HIV-1 is able to use

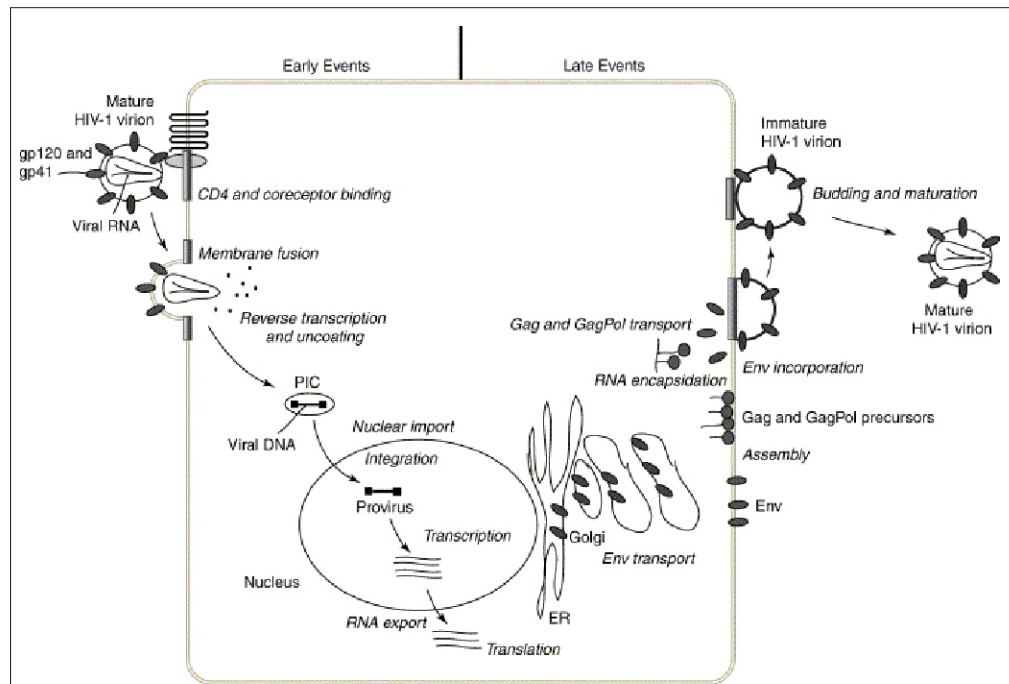


Figure 1.2: Events in the HIV-1 life cycle.
Adapted from Freed, 2004. See text for details.

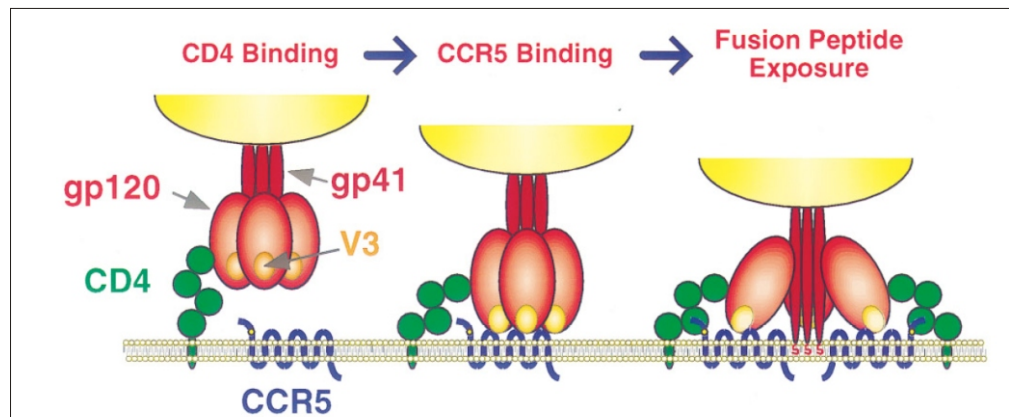


Figure 1.3: HIV-1 entry mechanism.
Adapted from Doms and Pieper, 1997. See text for details.

a multitude of coreceptors but *in vivo* its usage is limited to CXCR4 or CCR5 (Doms and Trono 2000). Originally, viruses were defined as M-tropic or T-tropic according to their ability to infect macrophages or T cell lines. Since it has been determined that viral tropism is predominantly due to the expression of CXCR4 or CCR5 on the cell surface of T cells and macrophages respectively, viruses are now classified as X4, R5 or dual-tropic according to their ability to use CXCR4 or CCR5 (Berger *et al* 1998).

Following binding of gp120 to CD4 and the coreceptor, drastic conformational changes occur in gp41 that ultimately lead to fusion of the viral and cellular membranes and release of the viral core into the cytoplasm (Chan and Kim 1998). The gp41 ectodomain is triggered from the native state to an intermediate formation in which the hydrophobic “fusion peptide” is exposed and inserted into the target cell membrane (Figure 1.3). Further conformational changes result in the formation of the fusion-active state, or hairpin, in which the N-helix and C-helix form a six-helix bundle (Chan *et al* 1997; Weissenhorn *et al* 1997). This “spring-loaded” mechanism is similar to the mechanism of influenza virus entry using HA₂ (Carr and Kim 1993).

Once inside the cell, the viral particle is uncoated and the reverse transcription complex is generated and released from the plasma membrane. This complex docks onto actin microfilaments and reverse transcription takes place, producing double-stranded viral cDNA (Telesnitsky and Goff 1997). Vif is required to overcome interference by host restriction factors such as APOBEC3G and TRIM5 α (Kremer and Schnierle 2005). Once reverse transcription is complete the viral pre-integration complex (PIC) is formed. The PIC is transported on microtubules to the nucleus, which it enters through nuclear pores by exploiting the Importin system. At this point viral cDNA is integrated into the host genome by the viral integrase and other host proteins such as HMGI(Y) and BAF (Brown 1997). A state of either latent or productive infection is then entered, depending on the cellular environment,

the state of the chromatin in which the provirus has integrated, and the availability of transcription factors.

Viral transcription by RNA polymerase II is a highly regulated process that involves the viral transcription factor, Tat, and many host transcription factors. As mentioned above, differential splicing of viral mRNA is required in order to regulate the production of different viral proteins. The binding of Rev to the Rev response element (RRE), a stem-loop structure present within the viral mRNA, allows export of mRNA to the cytoplasm. The proteins are expressed using the host-cell translational machinery and viral assembly primarily occurs at the plasma membrane and is directed by products of the *gag* gene.

Env is inserted cotranslationally into the lumen of the endoplasmic reticulum (ER) through its N-terminal signal peptide, and is stopped by a signal in gp41. In the ER and Golgi, gp120 becomes heavily glycosylated, intra-domain disulphide bonds are formed, and oligomerisation occurs (Freed and Martin 1995). Env is cleaved into gp120 and gp41 at the C-terminus of a Lys/Arg-X-Lys/Arg-Arg motif by a host furin or furin-like enzyme, but the subunits remain associated through non-covalent interactions. The gp41 is required to anchor the complex to the plasma membrane and, as the gp41-gp120 interactions are weak, gp120 is shed from the cell surface, although this is more common in some viral strains than others. This shedding may account for the fact that the predominant antibody response is raised against monomeric gp120. It is uncertain how gp120 and gp41 are incorporated into virions but they are thought to interact with the MA domain of Pr55^{Gag}.

Virus particles bud from lipid rafts in the plasma membrane, and because of this the viral membrane incorporates high levels of cholesterol that increase its stability and its fusion potential. Host cell proteins are also incorporated into the new viral membrane and the virion core. Viruses either bud into the

extracellular space or across a viral synapse into an uninfected cell (Piguet and Sattentau 2004). Nascent viral particles are not infectious until a process of maturation has been completed. The Pr55^{Gag} and Pr160^{GagPol} polyproteins are cleaved by the viral protease and this causes other structural rearrangements that are required for virus maturation. Immature and mature virions can be distinguished by electron microscopy: the “doughnut-shaped” core of the immature virus is seen to condense upon maturation (Briggs *et al* 2004).

1.1.5 Gp120 structure

As described above, gp120 on the virion surface binds to CD4 and the coreceptor to enable viral entry. The amino acid sequence of gp120 consists of five constant regions interspersed with five variable regions, which are so named due to their comparative sequence conservation or variability across HIV-1, HIV-2 and SIV (Starcich *et al* 1986). In the 1990s, various mutagenesis and antibody-binding studies provided the first insights into gp120 structure and function, and this was largely confirmed when the crystal structure of gp120 was solved (Kwong *et al* 1998).

Three gp120 structures have now been solved and are shown in Figure 1.4. The first gp120 structure (Kwong *et al* 1998) was deleted in variable regions one, two and three (Δ V1/V2/V3), deglycosylated and complexed with two domains of CD4 and the Fab fragment of antibody 17b (Figure 1.4 a and b). This was followed by crystallisation of glycosylated gp120 that was only Δ V1/V2-deleted, and was also complexed with two domains of CD4 and the Fab fragment of another CD4i antibody, X5 (Figure 1.4c, Huang *et al* 2005). Most recently the structure of an unliganded gp120 from SIV was solved that is Δ V1/V2/V3-deleted (Figure 1.4d, Chen *et al* 2005). The main caveat with the comparison of these structures is that the unliganded gp120 is from SIV. The particular strain of SIV used in the study, SIVmac32H, is able to infect cells in a CD4-independent manner with moderate efficiency (Reeves *et al*

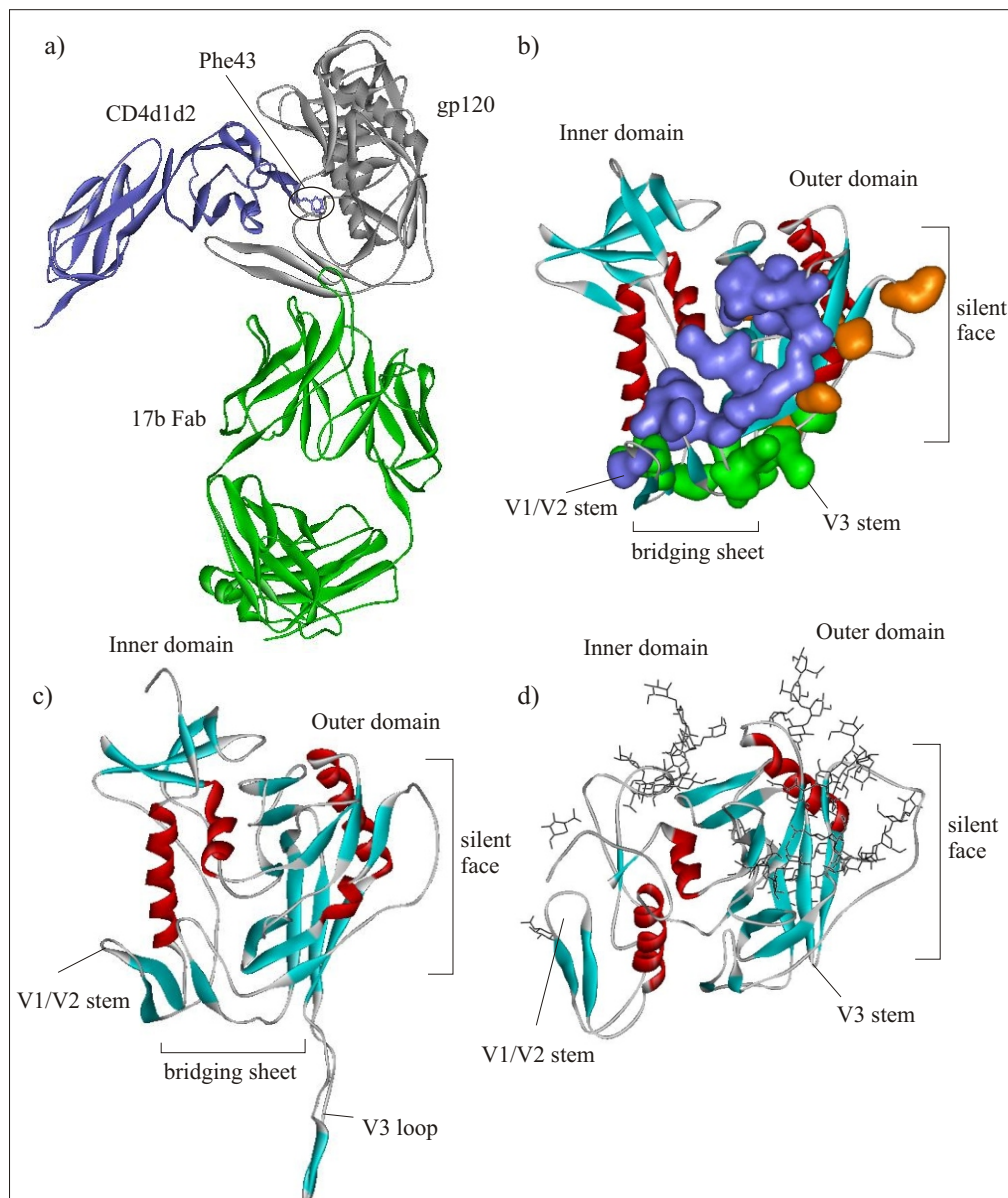


Figure 1.4: Structures of gp120.

- Structure of HIV-1 gp120 (grey) in complex with two domains of CD4 (blue) and the Fab fragment of antibody 17b (green). The atoms of residue Phe43 of CD4 are drawn. Adapted from Kwong *et al* 1998.
- Structure of HIV-1 gp120 showing the CD4 binding site (blue), CCR5 binding site (green) and 2G12 binding site (orange). Adapted from Kwong *et al* 1998, Rizzuto *et al* 1998, Rizzuto and Sodroski 2000, Cormier *et al* 2001, Cormier and Dragic 2002, Trkola *et al* 1996.
- Structure of HIV-1 gp120 including the V3 loop. Adapted from Huang *et al* 2005.
- Structure of unliganded SIV gp120. Glycosylation is shown in black. Adapted from Chen *et al* 2005.

Structures (b-d) are shown in the standard orientation in which the viral membrane would be orientated above and the cell membrane below the figure. Structure (a) is shown in a conformation that is rotated 90° to the left relative to the other structures. In (b) and (c) the CD4 and antibody domains are omitted for clarity. Secondary structures are depicted schematically.

1999), so there may be some structural differences between this molecule and unliganded gp120 from HIV-1. Hopefully the structure of unliganded HIV-1 gp120 will be solved in the future.

Electron microscopy studies have shown that it is likely that the functional gp120 spike is a heterotrimeric gp120/gp41 complex, held together by the trimerisation domain of gp41 and by non-covalent interactions between gp41 and gp120 (Zhu *et al* 2003). Other studies indicate that dimers and tetramers of gp120 also exist, probably due to shedding of gp120 from the virion surface and inefficient trimerisation in the ER (Kuznetsov *et al* 2003). The structure of the gp120/gp41 heterotrimer has been modelled based on the crystal structure of monomeric gp120 (Kwong *et al* 2000).

The monomeric, liganded structure of gp120 (Figure 1.4 a and b) consists of an inner domain, outer domain and bridging sheet (Kwong *et al* 1998). The inner domain consists of constant regions one and five and forms the major contact with gp41, and the outer domain is highly glycosylated. The CD4 binding site is at the interface between the inner domain, bridging sheet and outer domain, and is reliant on a hydrophobic cavity in which residue Phe43 of CD4 lies (Figure 1.4a and b). CD4 binding causes conformational changes in gp120 that exposes the CCR5 binding site and drastically enhances binding of CCR5 (Wu *et al* 1996). These structural changes can be seen by comparing the unliganded and liganded forms of gp120 (Figure 1.4b-d; Kwong *et al* 1998; Chen *et al* 2005). The outer domain conformation is similar in all three structures, whereas substructures of the inner domain are repositioned and the bridging sheet is disrupted when the gp120 is not bound to CD4 (discussed further in Chapter 6). It has been suggested that monomeric gp120 is highly flexible and its conformation becomes restricted on CD4 binding (Pantophlet and Burton 2006), but it is unknown how the flexibility is affected in the trimeric form.

The CCR5 binding site was described from mapping studies (Rizzuto *et al* 1998; Rizzuto and Sodroski 2000) and is thought to involve the V3 loop and core regions of gp120 (Figure 1.4b). The antibodies 17b and X5 that were co-crystallised with gp120 are thought to bind to similar epitopes as CCR5 because their binding is also CD4-induced and overlapping gp120 residues are required (Figure 1.4a; Thali *et al* 1993; Moulard *et al* 2002).

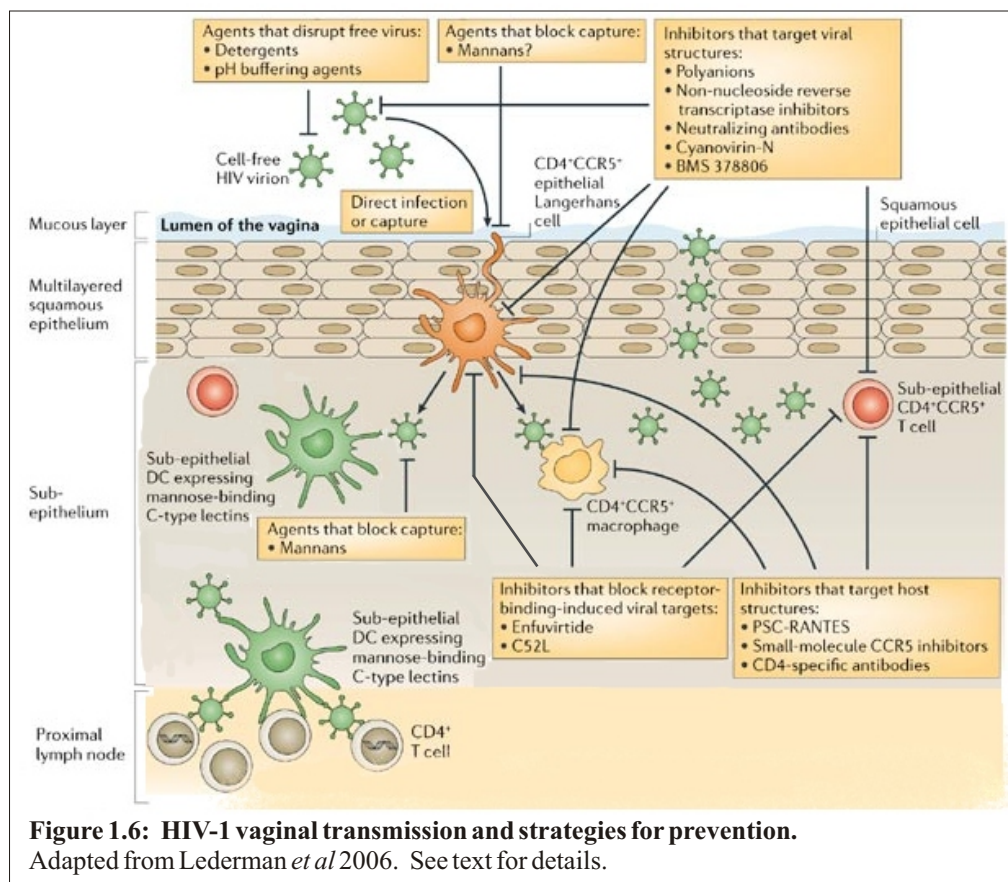
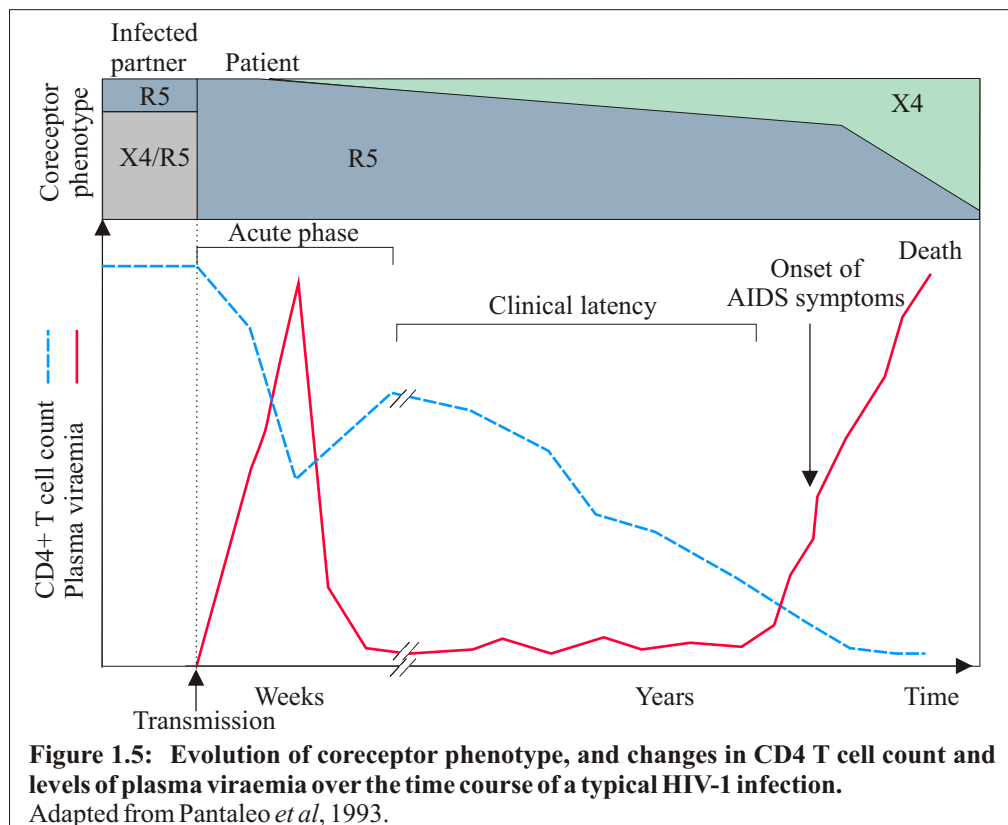
Antibody responses are predominantly directed against gp120 but they are ineffective in controlling viral replication because of a number of structural features of gp120. Although the functional form of gp120 is the trimer, monomeric gp120 is shed from the virus surface so antibodies are raised against epitopes on gp120 that are occluded in the trimeric form, and are non-neutralising (Kwong *et al* 2000; Kwong *et al* 2002; Yang *et al* 2002). In the gp120 trimer the CD4 binding site is masked by the V3 loop (Chen *et al* 2005) and the CCR5 binding site is blocked by the V1 and V2 loops of adjacent gp120 monomers (Kwong *et al* 2000) until gp120 binds CD4. This occlusion and the fact that the sites are only briefly exposed during viral entry reduces the opportunity for antibodies to be raised against these conserved sites. Antibody b12 is a rare example of an antibody whose epitope overlaps with the CD4 binding site. It was isolated from an HIV-1-infected patient and was found to neutralise a wide range of isolates of HIV-1 due to its unusually long CDR3 loop (Saphire *et al* 2001; Pantophlet *et al* 2003). Antibodies that bind to the coreceptor binding site in a CD4-induced manner (CD4i antibodies) have also been isolated from patients on rare occasions, such as antibodies 17b and 48d (Thali *et al* 1993), or made by immunising animals with the gp120-CD4-CCR5 complex (Moulard *et al* 2002). Another study showed that CD4i-antibody Fab fragments bind to gp120 more efficiently than the whole molecule (Labrijn *et al* 2003). This indicates that binding of the whole antibody is sterically hindered by the gp120 oligomerisation and variable loops, whereas the smaller Fab fragments are able to penetrate the complex and access the binding site without inhibition.

50 % of the surface of gp120 is covered in carbohydrates on the region known as the “silent face” (Figure 1.4b-d; Wyatt *et al* 1998). This glycosylation makes the exposed part of gp120 poorly immunogenic: only one antibody (2G12) that binds to this epitope has been isolated from a patient (Scanlan *et al* 2002). Not only does this glycan shield cover the surface of the gp120 trimer (Kwong *et al* 2000) but the pattern of glycosylation can evolve throughout the course of an infection (Wei *et al* 2003). Gp120 residues that are required for antibody 2G12 binding are shown in Figure 1.4b.

1.1.6 Transmission and disease progression

Transmission of HIV-1 between individuals may occur in a number of ways. Infectious virus particles are present in the semen, blood and breast milk leading to transmission through homosexual and heterosexual intercourse, blood transfusions, sharing of needles by intravenous drug users and from mother to child. HIV-1 infection originally predominated in homosexual men in the Western world, but heterosexual transmission is now the most prevalent mode of infection in all parts of the world. Women constitute 60 % of the HIV-1-positive population and in the developing world they may have little opportunity to protect themselves if they are unable to insist on preventative methods such as use of condoms (UNAIDS/WHO 2005). Therefore the development of antiviral microbicides that can be used by women may help to curb the spread of HIV-1 in developing countries. There are also many political and economic factors that need to be addressed, such as availability and cost of anti-viral drugs, microbicides and vaccines, education of susceptible populations, and reduction of the social stigma associated with the disease.

A typical course of an HIV-1 infection is shown in Figure 1.5. Although HIV-1 variants that utilise both CCR5 and CXCR4 for cell entry exist within an infected individual, CCR5-using (R5) viruses are predominantly



transmitted. Individuals with a homozygous $\Delta 32$ CCR5 mutation, who have no cell-surface CCR5, are highly resistant to HIV-1 infection (Liu *et al* 1996), which highlights the importance of CCR5 for transmission. This means that targeting the gp120-CCR5 interaction is a rational method of preventing HIV-1 infection. The probability of a sexual encounter with an HIV-1 infected individual resulting in established infection is thought to be related to the viral load (Quinn *et al* 2000), so most transmission occurs in the acute stages of infection when the viral load peaks and few clinical symptoms occur so the patient is unaware that they are infected.

Figure 1.6 outlines the steps in heterosexual transmission of HIV-1 across the vaginal mucosa and ways that this process can be intervened (see Section 1.1.8). The virus crosses the mucosal barrier and infects Langerhans cells, a type of CD4⁺ CCR5⁺ dendritic cell (Hu *et al* 2000). In addition, cell-free virus may access the submucosa through tears in the epithelium and infect T cells, macrophages and dendritic cells (DCs). Next the virus is disseminated to lymph nodes through a mechanism that is still somewhat unclear, but probably involves DCs, CD4⁺ T cells and macrophages. A number of studies have indicated that HIV-1 is able to bind to DC-SIGN on the surface of DCs, resulting in transfer of the virus to T cells without actual infection of the DC (Geijtenbeek *et al* 2000; Kwon *et al* 2002). It is clear that HIV-1 gp120 is able to bind to DC-SIGN *in vitro* but there is still debate over whether or not this is a physiological mechanism of viral dissemination.

The virus is transmitted to the lymph nodes where infection of resting CD4⁺ T cells and macrophages establishes latency while infection of activated CD4⁺ T cells causes virus production. By day seven the virus has spread to the gut-associated lymphoid tissue (GALT) and plasma viraemia peaks at day 21. Immune responses to the virus do occur at this stage, but they are ineffective due to their insufficient magnitude, breadth and rapidity. Cytotoxic lymphocyte (CTL) responses are directed predominantly against

epitopes in Gag and Tat (Mothe *et al* 2002). These responses may contribute to CD4⁺ T cell death but they fail to control the virus as the number of productively infected cells is too high and escape mutants emerge rapidly (Price *et al* 1997). Similarly, antibody responses tend to be non-neutralising so have little effect on viral replication (Safrit *et al* 1994). Later in the course of infection, anti-gp120 neutralising antibodies may be produced but the virus is able to mutate rapidly and evade the antibody response. In a small number of cases neutralising antibodies are raised against conserved epitopes of gp120 and viral escape is less marked, but by this stage a latent infection has been established so, while the antibody response may delay the course of disease, the virus will not be cleared.

There is a significant decline in viral load after the acute stage of infection because of these immune responses, and due to the fact that the population of susceptible memory T cells has been saturated by HIV-1 infection (Borrow *et al* 1994; Safrit *et al* 1994; Phillips 1996). At this point the clinically latent phase is entered (Figure 1.5), which may last for several years. Latent infection of resting T cells, as well as of other cells in immune-privileged sites such as glial cells of the central nervous system, allows the viral infection to persist for extended periods, even in patients on anti-retroviral therapy (Finzi *et al* 1999). There are low levels of viral replication, the immunodeficiency slowly progresses and the patient's CD4⁺ T cell count declines, until eventually the viral load increases and clinical symptoms occur, leading to AIDS and finally death. In 51 % of cases, particularly in clade B, the viral tropism evolves from R5 to X4 (Figure 1.5), and this evolution tends to correlate with the onset of disease (Schuitemaker *et al* 1992; Scarlatti *et al* 1997).

1.1.7 Immune evasion

HIV-1 is able to evolve rapidly according to selective pressures from CTL and antibody responses due to the nature of its replication (McMichael 1998;

Parren *et al* 1999). There is a high rate of error in reverse transcription of approximately one nucleotide per replication cycle (Mansky and Temin 1995) and the high turnover of virus particles means that the viral pool is renewed every day. The variable regions of gp120 are particularly adapted to a high level of mutation and can adapt to evade antibody responses without a loss of viral fitness (Burton *et al* 2004).

Anti-HIV-1 CTL responses are impaired because HIV-1 infection alters the expression of a number of cell-surface molecules. Gp120 binds to CD4 in the ER and slows its export, and Vpu and Nef also contribute to CD4 down-regulation, resulting in decreased T cell activation (Chen *et al* 1996). Nef is also responsible for the down-regulation of MHCI, leading to reduced antigen presentation (Collins *et al* 1998). In the late stages of disease Nef also up-regulates Fas ligand (FasL), which induces apoptosis of CTLs via the Fas-FasL pathway (Xu *et al* 1999).

1.1.8 Current strategies for HIV-1 prevention and treatment

The development of a number of antiviral drugs has meant that many patients in the developed world are able to live for several years after being diagnosed as HIV-1-positive. Current treatments are unable to clear HIV-1 infection because latent viral reservoirs are inaccessible to replication inhibitors, and antiviral treatments are yet to become commonplace in the developing world because of their expense. Therefore new strategies to prevent the transmission of HIV-1 are needed, and existing treatments should be improved to lower their cost and reduce their side-effects. A huge global effort has gone into the development of a vaccine to prevent HIV-1 infection but this has so far proved unsuccessful. A recent review describes the reasons that current efforts to produce an HIV-1 vaccine have failed and outlines future strategies (McMichael 2006).

1.1.8.1 *Antiviral drugs*

Most existing anti-HIV-1 drugs target the viral reverse transcriptase or protease (reviewed in Kaufmann and Cooper 2000). Soon after the discovery of the first reverse transcriptase inhibitor, AZT (Ezzell 1987), it was realised that viral resistance developed very rapidly. The implementation of HAART (highly active antiretroviral therapy), in which reverse transcriptase and protease inhibitors are used in combination, can prevent the emergence of drug-resistant strains for more than four years. New drugs such as darunavir (or TMC-114), a novel protease inhibitor (Kovalevsky *et al* 2006), are being developed to specifically target drug-resistant strains. Enfuvirtide (T-20) is a fusion inhibitor that binds to gp41 and prevents the conformational changes that are required for virus entry. This drug is currently used as a last resort for patients who no longer respond to HAART, because it requires twice-daily injections (Lalezari *et al* 2003; Lazzarin *et al* 2003). Other inhibitors that target viral entry, integration and virion maturation are being developed, and it would be possible to inhibit accessory factors such as Nef and Vif. Although Nef is not absolutely required for HIV-1 infection, its deletion decreases virulence (Deacon *et al* 1995) so inhibition of Nef may increase the patient's life-span. Inhibition of Vif would enable cellular restriction factors such as APOBEC3G to inhibit viral replication (Kremer and Schnierle 2005). A new strategy is to try and stimulate the latent reservoirs of HIV-1 in order to completely eradicate the virus (Pomerantz 2002). Gene therapy approaches are also being investigated, for example introduction of siRNA by lentiviral vectors (Morris and Rossi 2006).

1.1.8.2 *Prevention of transmission*

The predominant mechanism of HIV-1 transmission is through heterosexual sex, and there are a number of ways in which this can be prevented. Educating people about HIV-1 prevention and promoting testing are key factors. Abstinence and monogamy are promoted by conservative and

religious groups but for many women this is not a practical option. Similarly, condoms have proved to be a highly successful method of preventing HIV-1 transmission (De Vincenzi 1994) but reluctance to use them and the misconstrued belief that they do not prevent the spread of HIV-1 means that they can have limited success in some societies. There is growing evidence that male circumcision can reduce the probability of HIV-1 infection occurring from sexual intercourse (Auvert *et al* 2005; Siegfried 2005), but again some cultures are not open to this. Prevention of other sexually transmitted diseases such as HSV-2 also reduces the chance of HIV-1 infection, probably because other infections damage the vaginal epithelium (Barton *et al* 2005).

HIV-1 is also transmitted from mother to child and the chance of this occurring can be reduced by administering a single dose of nevirapine (a non-nucleoside reverse transcriptase inhibitor) to them both, and avoiding breastfeeding when antiviral therapies are unavailable (Farley *et al* 2002). In the developed world, post-exposure prophylaxis is available following occupational exposure, such as a needle-stick injury, or non-occupational exposure, such as rape.

1.1.8.3 Microbicides

Antiviral microbicides are being developed as a novel way to prevent HIV-1 transmission across the vaginal epithelium (Figure 1.6; Lederman *et al* 2006). I will discuss the types of inhibitor and delivery methods as well as safety, acceptability, efficacy and cost.

The first microbicides to be investigated for clinical use were detergents that disrupt the viral membrane, or agents that modify pH and inactivate the virus. However one promising detergent, Nonoxynol-9, was found to increase the rate of HIV-1 transmission because it damages the vaginal mucosa (Hillier *et*

al 2005). Another compound, BufferGel, lowers the vaginal pH to 4, which disrupts the virion structure, but there is uncertainty over whether or not it will work in the presence of alkaline semen (Mayer *et al* 2001). Given these caveats, I predict that more specific viral inhibitors are likely to be more successful. Existing antiviral drugs that target viral replication could also be used as microbicides; for example the non-nucleoside reverse transcriptase inhibitor TMC120 is being tested in a vaginal ring (Malcolm *et al* 2005). However problems regarding viral resistance may be encountered if the same inhibitor is used as a microbicide and a drug treatment.

A number of candidate microbicides target viral proteins involved in transmission (Figure 1.6). Polyanions such as PRO2000 and heparan sulphate inhibit charge interactions between the V3 loop of R5 gp120 and the coreceptor in a CD4-dependent manner (Vives *et al* 2005), and X4 strains in a CD4-independent manner (Moulard *et al* 2000). Lectins such as Cyanovirin-N, an 11kDa protein expressed from *Nostoc ellipsosporum*, inhibit R5 and X4 HIV-1 and SIV by binding to the glycans on gp120 (Boyd *et al* 1997). The monoclonal antibody 2G12 also binds to these glycans, and has been used in combination with other antibodies for systemic infusion to prevent HIV-1 infection (Mascola *et al* 2000), but this has led to the production of resistant strains. A number of compounds that inhibit the interaction of gp120 with CD4 have been tested in macaque models for their ability to prevent HIV-1 infection. These include monoclonal antibody b12, which binds the CD4 epitope of gp120 (Veazey *et al* 2003b), and BMS 378806, which prevents the conformational changes in gp120 that are required for virus entry (Si *et al* 2004). The inhibitors Enfuvirtide and C52L prevent conformational changes in gp41, and have been used to prevent SHIV infection of macaques in vaginal challenge experiments (Veazey *et al* 2005). The advantage of targeting the virus is that there should be fewer adverse effects on the vaginal epithelium, and current studies have shown that viral inhibitors tend to work at lower concentrations than cellular inhibitors, and

are therefore cheaper (Klasse *et al* 2006). However, there may be problems with viral resistance developing, particularly with prolonged use. As with vaccines, it may be difficult to find a microbicide that can target a wide range of HIV-1 strains, and it is uncertain whether solely inhibiting R5 virus may promote the transmission of X4 strains.

A number of groups are developing microbicides that target host-cell molecules involved in the virus entry pathway (Figure 1.6). Mannan was tested as a microbicide because it blocks binding of the virus to DC-SIGN on dendritic cells *in vitro* (Geijtenbeek *et al* 2000), but it did not prevent infection of macaques with SHIV in a vaginal challenge model (Veazey *et al* 2005). A number of synthetic analogues of chemokines have been shown to inhibit viral replication and inhibit HIV-1 infection of macaques in a vaginal challenge model, for example AOP-RANTES (Mack *et al* 1998) and PSC-RANTES (Lederman *et al* 2004). A single dose of these inhibitors may have a longer-lasting effect than an inhibitor of a viral protein because they sequester the chemokine receptor from the cell surface. Other small molecule inhibitors of CCR5 prevent gp120 binding by inducing a prohibitive conformation in the coreceptor, for example TAK-779 (Baba *et al* 1999) and CMPD167 (Veazey *et al* 2005). These molecules have the potential to be used either as topically-applied microbicides or as drug treatments. An advantage of targeting the cell-surface receptors rather than the virus is that viral resistance cannot develop, though it is uncertain how these molecules may affect host cell function. Direct inhibitors of CD4 cannot be used as they would disrupt T cell signalling. Individuals with a homozygous $\Delta 32$ CCR5 mutation seem to have an intact immune system so many groups feel that CCR5 is an acceptable receptor to target. Trials must confirm that CCR5 inhibitors do not have adverse effects, for example chemokine analogues could activate downstream signalling pathways. Also it is questionable whether blocking the coreceptor is sufficient to prevent HIV-1 infection. It is possible that the virus could be disseminated by dendritic cells to other sites

where CCR5 is not blocked, and from there establish an infection. However a number of studies have shown that CCR5 inhibitors can prevent HIV-1 infection of macaques in a vaginal challenge model (Veazey *et al* 2003a; Lederman *et al* 2004; Veazey *et al* 2005). In all of these *in vivo* studies the inhibitor had to be applied at a concentration three logs higher than in the *in vitro* experiments, which may reflect the need for the compound to penetrate deeper into the vaginal epithelium in order to be effective.

Most candidate microbicides are being developed with a view to being used in a single-dose gel-based formulation, but other delivery methods are possible. Vaginal rings, pessaries or suppositories may be used to release the compound over a prolonged period, for example the slow release of the microbicide TMC120 (Malcolm *et al* 2005). Systems like these may have increased compliance, acceptability and efficacy than a gel-based microbicide. Another approach is rectal or vaginal colonisation with recombinant bacteria that secrete peptide inhibitors, such as the *Escherichia coli* strain Nissle 1917 that secretes a fusion inhibitor (Rao *et al* 2005). For this method to be successful the protein must be secreted at high enough concentrations over a long term period. There is also a risk that the genetic modification could affect the bacterial pathogenicity or immunogenicity. Systemic infusion of antibodies has been used successfully to prevent HIV-1 infection, but not only is this a very costly method that is reliant on frequent injections, it has also produced resistant strains of HIV-1 (Mascola *et al* 2000).

There are a number of safety issues that need to be addressed when developing candidate microbicides. Clinical trials should check that the compound does not irritate the epithelium and cause inflammation, like Nonoxynol-9 (Hillier *et al* 2005). Similarly, the use of chemokine analogues must be monitored carefully as chemokines recruit HIV-1-susceptible cells to the epithelium. Proteins need to be checked for carcinogenic effects and immunogenicity, particularly in long-term use. Microbicides should be tested

on both healthy and HIV-1 positive patients since their use in infected patients may increase the risk of resistant strains emerging.

One key issue in microbicide development is their acceptability. They must be easy to use, unobtrusive and not unpleasant, otherwise people will simply avoid them, and ideally the compound should remain active for several hours after application. Several CCR5 inhibitors were shown to be active in a macaque model when the challenge occurred six hours after the application of the inhibitor (Veazey *et al* 2005). Public perception may be a problem for implementation of bacterial microbicides. Many people avoided “GM crops” and the same could happen with “GM bacteria”. Similarly, siRNA therapies could be labelled “gene-therapy” and may be shunned by the public.

The efficacy of microbicides may be hampered by inconsistent use, unacceptability or cost. Microbicides must be cheap enough to be widely available in developing countries, ideally costing less than \$1 per application and this is yet to be achieved. However, the cost of a vaginal ring that releases an RT inhibitor is estimated at \$5-10 per dose, which may be acceptable given that one ring could last for several months (Malcolm *et al* 2005). If an effective microbicide is found, cheaper production methods can be developed later to enable its widespread distribution. Microbicides need to be active against a panel of viruses, and it would be possible to use the same panel as has been recommended for testing candidate vaccines (Mascola *et al* 2005). One of the main drawbacks with primate models of HIV-1 challenge is the limited number of HIV, SHIV and SIV strains available, so these studies will have to be combined. Clinical trials should be carefully designed so that time and money is not wasted testing compounds whose efficacy is not sufficient, but at the same time potentially useful compounds must not slip through the net.

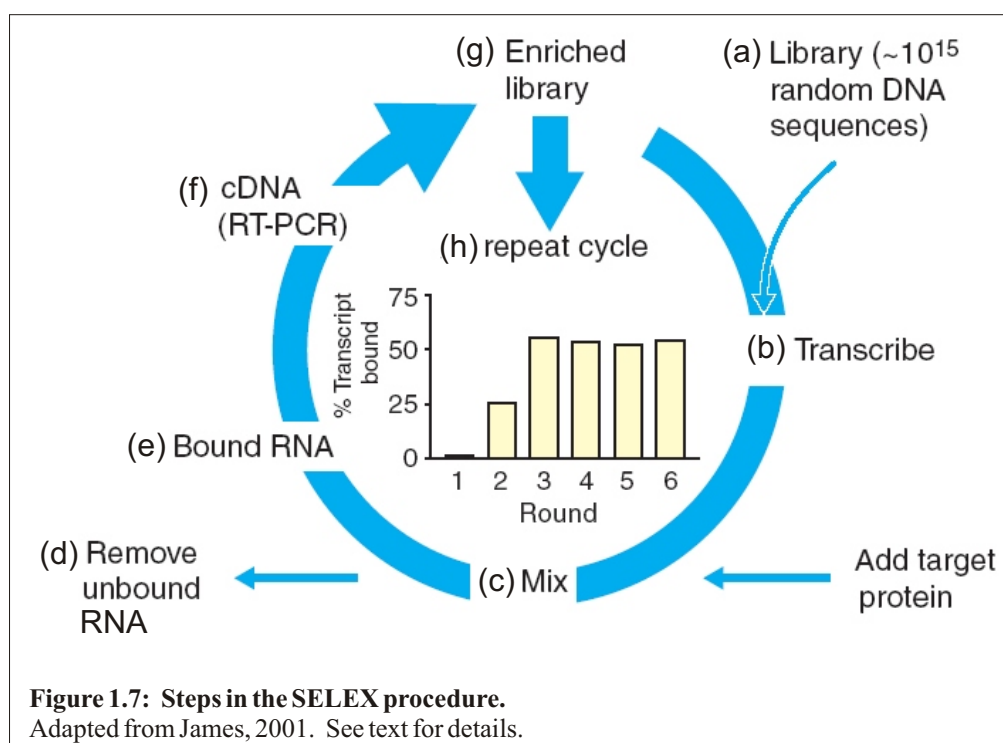
Current progress indicates that anti-HIV-1 microbicides have potential to reduce the transmission of the virus and could be used in combination with antiretroviral therapy and social factors to help stem the epidemic

1.2 Aptamers

Aptamers are synthetic nucleic acid ligands that are selected by *in vitro* evolution from a randomised pool to bind to a target of choice. They are characterised by their high specificity and affinity for the target, which gives them a wide range of applications as detection reagents, in elucidating protein function, in drug discovery, and as therapeutic agents. Previously in the James lab RNA aptamers with 2'-fluoro-pyrimidine chemistry were selected against HIV-1 gp120 from the strain BaL (Khatri *et al* 2003). In this section I will discuss RNA aptamer selection and potential applications, as well as previous work on anti-gp120 aptamers.

1.2.1 Aptamer selection

The process of *in vitro* selection or SELEX (selection of ligands by exponential enrichment) was described by Tuerk and Gold in 1990, the same year that Ellington and Szostak coined the term “aptamer” (Ellington and Szostak 1990; Tuerk and Gold 1990). The SELEX process to obtain RNA aptamers is outlined in Figure 1.7. The starting pool (a) consists of approximately 10^{15} DNA oligonucleotides with a random sequence of defined length flanked by 5' and 3' constant regions. The 5' constant region includes the binding site for T7 RNA polymerase, and the DNA pool is used as a template for *in vitro* transcription to produce an RNA pool (b). Due to the nature of RNA folding, each random sequence will fold into a unique tertiary structure. These sequences are then mixed with the target protein (c), for example on a nitrocellulose membrane, magnetic beads, or a BIAcore chip. By chance, a small number of RNA molecules will fold into structures that



bind to the target. Unbound RNA is removed and discarded (d), after which the bound fraction is eluted (e). This selected pool is then used as a template for RT-PCR (f) to produce an enriched DNA library (g) and the next round of selection is begun (h). In the first round of selection the proportion of RNA that binds to the target is very low, usually less than 0.1 %, and this proportion increases after each round of selection until the proportion of molecules that bind reaches a plateau. Multiple rounds result in the selection a pool of RNA sequences that bind to the target with high specificity and affinity. Many steps in the SELEX process have to be optimised to maintain the maximum diversity, particularly in the first round of selection. Full details of the experimental procedures have been described elsewhere (Fitzwater and Polisky 1996). Recent efforts have focussed on the automation of the SELEX procedure (Cox *et al* 2002).

Negative selection of aptamers can be performed to increase their specificity. For example aptamers raised against ATP were also found to bind ADP and AMP because the common adenosine moiety is the preferred recognition target (Sassanfar and Szostak 1993). Through a process of negative selection, the Szostak group were able to select aptamers that bound to ATP with an affinity three logs higher than to AMP (Sazani *et al* 2004).

Instead of a simple binding assay, aptamers can be selected for catalytic or inhibitory functions. For example Koizumi and colleagues were able to select a catalytic aptamer, a ribozyme, that undergoes self-cleavage only in the presence of cGMP and cAMP (Koizumi *et al* 1999).

The aptamer RNA chemistry can be modified pre- or post-selection in a number of ways to increase the stability or bioavailability of the RNA, or to enable its detection. The introduction of 2'-fluoro-pyrimidines, 2'-amino-pyrimidines (Pieken *et al* 1991) or 2'-O-methyl-purines (Rhodes *et al* 2000) increases the RNA stability by decreasing susceptibility to digestion by

nucleases. However this advantage is offset by a decrease in transcription efficiency by wild-type T7 RNA polymerase. Until recently, 2'-O-methyl modifications had to be introduced after the selection procedure, but a mutant T7 RNA polymerase that incorporates modified nucleotides has now been described enabling SELEX with 2'-O-methyl-pyrimidines to be performed (Chelliserrykattil and Ellington 2004). 5'-iodo- or 5'-bromo-uridines can be incorporated so that the RNA aptamer can be chemically crosslinked to a protein target (Bock *et al* 2004). Aptamers can be detected in assays by covalently linking markers such as biotin or FITC to their termini or to an internal site. In some cases the addition of a marker may affect the function of the aptamer, so its position should be chosen carefully (Sayer *et al* 2004), unless it is incorporated in the SELEX procedure (Blank *et al* 2001). When aptamers are used therapeutically they can be modified by the addition of polyethylene glycol (PEG) to increase their bioavailability (Zhou and Wang 2006). Once the nucleotides involved in target binding have been identified, for example through nuclease digestion and footprinting assays, the RNA can often be truncated without loss of function (Sayer *et al* 2004). This requires that the minimal binding region retains its secondary structure in the absence of the rest of the RNA sequence.

In the 16 years since aptamer technology was developed, a huge number of molecules have been used as targets, including small molecules such as dopamine (Mannironi *et al* 1997), peptides such as vasopressin (Porschke *et al* 2006), amino acids (Geiger *et al* 1996), proteins such as the hepatitis C NS3 protease (Fukuda *et al* 2000), nucleic acids such as the TAR RNA element of HIV-1 (Duconge and Toulme 1999), and whole tumour cells (Blank *et al* 2001).

The aptamers raised against vasopressin are known as spiegelmers and were developed by exploiting the phenomenon of chirality (Porschke *et al* 2006). The RNA aptamers were initially raised against a chemically-synthesised

form of the target made with unnatural D-amino acids. Having selected one sequence, this individual aptamer was then synthesised with mirror-image bases. The mirror-image aptamer, or spiegelmer, binds to the natural L-amino acid target molecule and has much higher biostability than normal RNA.

Nucleic acids may seem an unusual target for aptamers since the reverse-complement of a sequence should be a ligand. However viral RNA often has complex secondary structures that prevent complementary sequences from binding. Consequently aptamers that were raised against the HIV-1 TAR sequence (Duconge and Toulme 1999) bound to the target more strongly than the antisense RNA (Beaurain *et al* 2003).

1.2.2 Applications of aptamers

Aptamers tend to have affinities in the low nanomolar range, and are highly specific for their target. These properties make them suitable for a large range of applications; as detection reagents, in elucidating protein function, in drug discovery and as therapeutic agents. In a number of instances aptamers are able to be used instead of, and have advantages over, monoclonal antibodies. Recent reviews assess the potential applications of aptamers (Nimjee *et al* 2005; Proske *et al* 2005; James 2006).

A number of aptamers have been used as detection reagents where monoclonal antibodies are either unavailable or not specific enough. Aptamers raised against prion fibrils from infected mouse brain, PrP^{Sc}, were able to detect this diseased form with higher affinity than the cellular form, PrP^C (Rhie *et al* 2003). At the time, there was no antibody available that would bind to PrP^{Sc}, so the biotinylated aptamer was employed in binding assays and histology (Sayer 2004). Novel assays have been developed in which fluorogenic probes are conjugated to aptamers (Nutiu and Li 2005). One such mechanism employs fluorescence resonance energy transfer (FRET): in the unbound aptamer structure the probes are not near enough

each other to interact, but upon binding to the target the aptamer structure changes so that the probes become adjacent and FRET occurs. ELISA-type assays have been developed to detect the protein target with biotinylated aptamers and have proved to be 1000-fold more sensitive than detection with the equivalent antibody (Baldrich *et al* 2004; Baldrich *et al* 2005).

Aptamers have had many roles in the investigation of protein function. For example, aptamers that can differentiate between the phosphorylated and unphosphorylated forms of the MAP kinase Erk-2 have helped identify the functions of this protein in the MAPK signalling pathway (Seiwert *et al* 2000). Similarly, aptamers that bind to both human and mouse CXCL10 can be used as tools to investigate cross-talk between the chemokine and its receptor (Marro *et al* 2005).

A recent review outlines the potential for aptamers to be used in high throughput screening for drug discovery (Green *et al* 2001). Libraries of “drug-like” molecules can be screened in competition assays to find compounds that block the aptamer-target interaction. This is particularly advantageous because the aptamer can be labelled (see above) and this avoids labelling the potential drug or target, which may affect their binding properties. In addition the crystal structures of aptamers in complex with their targets have the potential to be used in structure-based drug design. Many such complexes have now been solved (Patel 1997), but this information is yet to be used in this way.

The potential of aptamers to be used therapeutically has been realised in the development and FDA approval of Macugen (Pegaptanib), an aptamer against VEGF, as a treatment for age-related macular degeneration (Ng *et al* 2006). Nucleic acids have a number of properties that make them poor therapeutic agents but these problems can be overcome. The nuclease resistance and shelf-life of RNA can be increased by the introduction of modified

nucleotides and 3' capping (see above). RNA is quickly cleared by the kidney so the anti-VEGF aptamer was conjugated to PEG to increase its bioavailability. Encouragingly the aptamers are non-toxic and poorly immunogenic due to their similarity to host molecules. An aptamer raised against factor IXa is in clinical trials to replace heparin as an anticoagulant for use in patients undergoing cardiopulmonary bypass surgery (Nimjee *et al* 2006). This aptamer is made with 2'-fluoro-pyrimidines to increase its stability and is conjugated to cholesterol to reduce renal clearance (Rusconi *et al* 2004). Interestingly, the complementary sequence acts as an antidote, enabling the anticoagulant activity of the aptamer to be reversed rapidly (Rusconi *et al* 2002).

In order for aptamers to be used therapeutically they need to be delivered locally to target specific sites, or by injection to act systemically for a limited period (as in the examples above). With respect to HIV-1 therapy this would make aptamers more suited to development as topically-applied prophylactic microbicides than antiviral drugs. Intracellular viral targets could be inhibited by the use of lentiviral vectors to deliver the gene encoding the aptamer into the cell. However this field is surrounded by problems including poor transduction efficiency and the risks associated with genetic recombination. It is uncertain whether or not the aptamer would fold into an active conformation in the intracellular environment and they would be degraded rapidly because only 2'-hydroxy-nucleotides would be incorporated. Therefore extracellular molecules, such as the HIV-1 surface glycoproteins, are more likely to be successful therapeutic aptamer targets.

Aptamers may prove to be more expensive than more typical “drug-like” molecules, but their cost of production will depend on the aptamer length, sequence composition and modifications. However it will be possible to produce therapeutic aptamers in areas where there is sufficient demand due to

a lack of acceptable alternative therapies, for example in the prevention and treatment of HIV-1.

1.2.3 Anti-gp120 aptamers

Work in the James lab was originally directed at selecting aptamers with 2'-fluoro-pyrimidines to bind to gp120 from the HIV-1 X4 strain, HXB2 (Sayer *et al* 2002). After nine rounds of selection, 33 aptamers were cloned and sequenced and 16 bound to monomeric HXB2 gp120. The aptamers bound to HXB2 gp120 with affinities in the order of 10^{-7} M but did not bind to gp120 from the R5 strain of HIV-1, BaL, and none of the aptamers were able to neutralise HIV-1_{HXB2} in the T cell line H9. Following this result, a new SELEX procedure was performed to raise aptamers with 2'-fluoro-pyrimidines against gp120 from the R5 HIV-1 strain BaL (Khatri *et al* 2003). All of the 27 aptamers that were cloned and sequenced from the round 5 pool bound to monomeric BaL gp120 with K_d values in the range of 5-100 nM. 19 of the 27 aptamers tested neutralised HIV-1 strain BaL in PBMCs by more than three logs of infectivity. Most interestingly, 11 aptamers also neutralised five HIV-1 R5 primary isolates from group M and one isolate from group O (Khatri *et al* 2003).

One anti-gp120 aptamer from the R5 selection, aptamer B40, was chosen for detailed structural analysis. The minimal binding region of the aptamer was identified through structural modelling, nuclease digestion, footprinting and mutagenesis (Dey *et al* 2005a). A truncated aptamer, B40t77, based on the minimal binding region, was shown to bind to gp120 and neutralise HIV-1_{BaL} in PBMCs comparably to the parental aptamer, B40. Additionally, the study demonstrated that substitution of the 2'-fluoro-uracil residues with 2'-hydroxy-uracil abrogated B40 binding to gp120, but that substitution of the 2'-fluoro-cytosine with 2'-hydroxyl-cytosine did not affect binding, indicating that some of the 2'-fluoro-uracil residues contact gp120 directly (Dey *et al* 2005a).

Further experiments began to investigate the binding site of aptamer B40 gp120. Deglycosylation of gp120 did not affect aptamer binding, so it is unlikely that the aptamer epitope is on the carbohydrate face (Dey *et al* 2005b). While aptamer binding to wild-type gp120 is not CD4-dependent, binding of the aptamer to $\Delta V1/V2$ or $\Delta V1/V2/V3$ gp120 is CD4-induced, indicating that the variable regions play a role in aptamer binding (Dey *et al* 2005b). Finally, competition experiments were performed in which the aptamer B40 was shown to compete with a sulphated peptide constituting the 22 N-terminal amino acids of CCR5, indicating that the aptamer and CCR5 binding sites on gp120 may overlap (Dey *et al* 2005b).

These experiments have provided some interesting observations about the anti-gp120 aptamers and have identified many questions that still need to be answered. A more detailed understanding of the aptamer-gp120 interaction and the mechanism of neutralisation may lead to the use of aptamers in investigating the HIV-1 entry pathway and as therapeutic agents.

1.2.4 Surface Plasmon Resonance

The interaction of aptamers and gp120 can be analysed conveniently by surface plasmon resonance (SPR) using a BIAcore 2000. This technique enables the interaction to be studied in real time and does not require labelling of either the aptamer or the gp120. In the SPR experimental set-up one molecule (the ligand) is immobilised on a sensor chip surface, and the second molecule (the analyte) is passed over the surface in an appropriate buffer.

Figure 1.8a outlines the principles behind SPR. A gold layer lies in between the glass sensor surface and the buffer, and this is struck by polarised light that is directed over a range of incidence angles. The light is internally reflected because the surface and buffer have different refractive indices, and an evanescent wave is generated when the light hits the glass surface. This produces electron charge density waves (plasmon waves) in the gold layer,

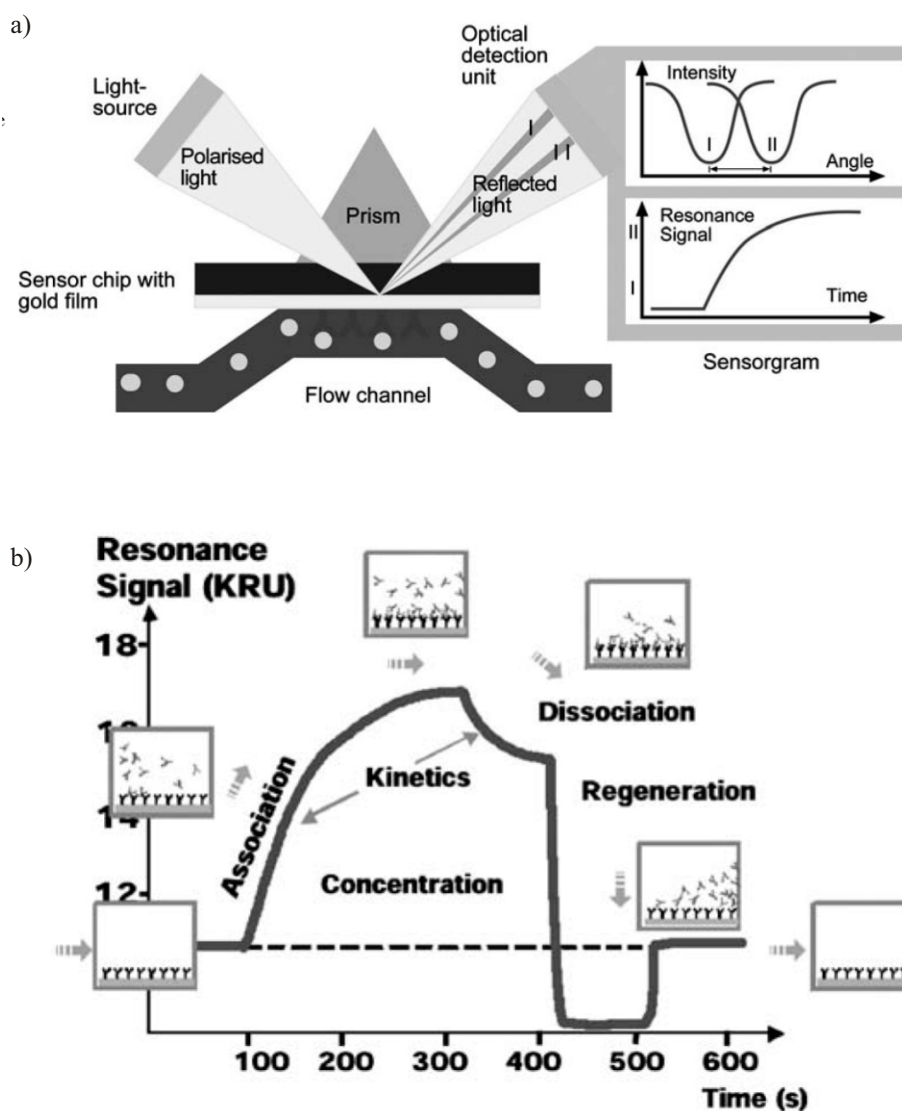


Figure 1.8: Principles of Surface Plasmon Resonance.

a) Schematic depiction of the process involved in ligand detection by SPR.

b) Typical real-time detection of ligand-analyte interaction.

Adapted from Cooper, 2003. See text for a detailed description.

decreasing the intensity of the reflected light. The incident “SPR angle” is the angle at which the minimum intensity of reflected light is produced, and this angle is determined by the refractive index of the buffer. When molecules interact with the sensor surface, i.e. the ligand is immobilised or the analyte binds to the ligand, the SPR angle changes. This is represented as a change in the SPR response and is proportional to the mass of the bound material.

The interaction between the analyte and the ligand is observed in real time, and a typical binding curve is shown in Figure 1.8b. As the analyte is injected the analyte and the ligand associate until equilibrium binding is reached, generating a binding curve that reaches a plateau. Once the injection is stopped the analyte dissociates and the response decreases. The analyte can be completely removed from the sensor surface by a regeneration step, after which the response returns to the baseline.

SPR technology has been applied to a wide range of applications to detect protein-protein, protein-DNA and protein-RNA interactions. The technique is less suited to the detection of small molecules because the response is proportional to the mass of the analyte. It is possible to use SPR to analyse binding in a quantitative manner to calculate kinetic parameters and binding affinities (Van der Merwe 2001).

1.3 RNA sequence and structure

Single-stranded RNA forms secondary and tertiary structures in solution due to internal base-pairing. In nature such structures often confer functional properties to the RNA, such as the TAR and RRE elements in HIV-1 RNA. Common structural motifs often exist in related sequences, for example in tRNA structures. These secondary and tertiary structures are often required to

promote or inhibit RNA-RNA, RNA-DNA or RNA-protein interactions or, for aptamers, the ability to bind to the target molecule.

There are a number of methods available to analyse the sequence and structure of RNA. The secondary structure of RNA molecules can be analysed by laboratory-based experiments methods. Such techniques are very rigorous but are time-consuming to perform. The structure of many RNA molecules can be analysed by computational methods where the use of laboratory methods would be impractical.

1.3.1 RNA sequence alignment

Single-stranded RNA molecules are produced in the laboratory by *in vitro* transcription from a cDNA template using an RNA polymerase such as T7 polymerase (Losick 1972). Alternatively, short RNA molecules of up to approximately 60 nucleotides can be produced by solid-phase chemical synthesis (Sproat 1993). Primary RNA sequences are generally inferred from their cDNA template, and a number of RNA sequences can be compared using the bioinformatics program, ClustalW (Thompson *et al* 1994).

ClustalW is used to detect homology between related DNA, RNA or protein sequences. It is based on a progressive algorithm developed nearly 20 years ago (Feng and Doolittle 1987), which uses dynamic programming (Eddy 2004b) to produce a mathematically optimal alignment. The program uses a table of scores for matches and mismatches, and penalties for insertions and deletions. In order to analyse eight or more sequences within a reasonable amount of time and range of computing power, heuristic approaches are used to simplify the algorithm. ClustalW represents an optimised version of the original program, and is freely available on the web at www.ebi.ac.uk/clustalw.

1.3.2 RNA structure prediction

The secondary structure of an RNA molecule, either alone or in complex with a protein, can be predicted using either experimental or computational methods. Enzymatic probing and primer extension are both robust experimental methods that have been employed successfully to identify RNA secondary structures (Ehresmann *et al* 1987). More recently, RNA structures have been solved by X-ray crystallography (Golden and Kundrot 2003) and NMR (Latham *et al* 2005). All of these methods are difficult to perform and very time consuming, and the development of computational approaches has enabled the prediction of many more RNA secondary structures.

The most commonly used RNA folding algorithms are based on principles of free energy minimisation, for example Mfold (Zuker and Stiegler 1981; Zuker 2003) and RNAfold, from the Vienna RNA server (Hofacker *et al* 1994; Hofacker 2003). The minimum free energy (MFE) structure is that which has the lowest Gibbs free energy in its predicted conformation, usually expressed in kcal/mol. Most RNA folding algorithms compare the energy of a given secondary structure with other structures for the same sequence in order to find the MFE structure. This approach uses the Turner free energy parameters of Watson-Crick base-pairs that were calculated by optical melting experiments (Tinoco *et al* 1973), subsequently refined (Freier *et al* 1986; Xia *et al* 1998), and modified to include GU base pairs and loops (Mathews *et al* 1999). Since the number of possible structures for one RNA sequence grows exponentially with sequence length, it is impossible to generate all of them (a sequence of 100 nucleotides has 10^{25} possible structures that would take 10^{13} years to compute!), so dynamic programming algorithms are used to calculate the energy of all possible structures without implicitly generating them, and the MFE structure is calculated as follows. The first step, called “fill”, determines the lowest conformational free energy for each possible sequence fragment, starting with the shortest, and then for fragments of increasing length until the energy of the full-length sequence has been calculated. The

energy of the longer fragments is calculated by the sum of the energies of the short fragments of which the longer fragment is composed. Finally a procedure known as “traceback” is performed in order to determine the structure that corresponds to this minimal energy. A more extensive description of this recursive algorithm can be found in a review by Eddy in 2004 (Eddy 2004a).

A number of groups have modified the original Mfold algorithm (Zuker and Stiegler 1981) to increase its speed and accuracy, and to make the analysis more biologically relevant. One such program, PKNOTS, is able to predict the presence of pseudoknot RNA structures that are not detected by Mfold or RNAfold (Rivas and Eddy 1999). For many RNA molecules the MFE structure does not correspond to the functional RNA structure, and some RNA molecules, known as riboswitches, exist in more than one conformation according to their environment (Mandal and Breaker 2004). For this reason the original algorithms have been modified to enable calculation of suboptimal energy structures (Zuker 1989; Wuchty *et al* 1999). When using the default Mfold parameters, structures with energies within 5 % of the energy of the MFE structure are included in the output.

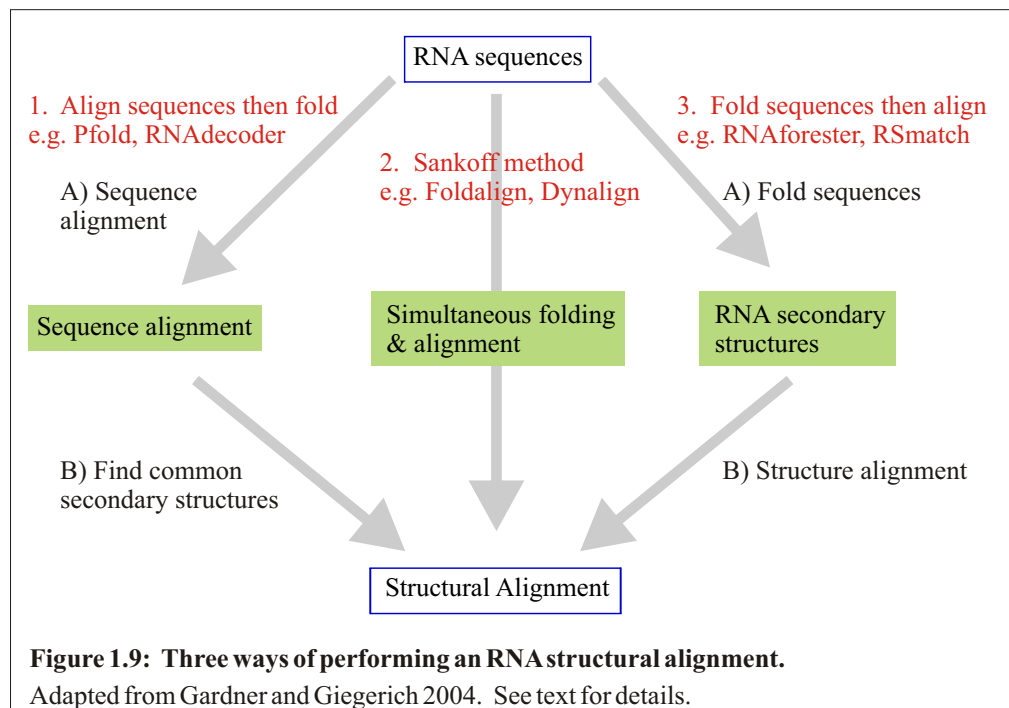
A new approach to RNA structure prediction was developed based on the principle of statistical sampling of possible RNA structures (Ding and Lawrence 1999; 2001; 2003). This program, Sfold, still uses the Turner energy parameters described above, but does not use a recursive algorithm as employed by Mfold and RNAfold. Instead it is designed to choose base pairs probabilistically in accordance with the partition functions for all possible sequence fragments. The probability of sampling a given structure is the same as the probability of it occurring in the thermodynamic ensemble. The Sfold web server not only has options for RNA structure prediction but also for other applications such as prediction of the accessibility of regions that are targets for antisense DNA (Ding *et al* 2004).

1.3.3 Comparison of RNA structures

When a functional RNA secondary structure motif has been identified it can be useful to find out whether or not a similar motif also exists in other RNA sequences. As with RNA structure prediction, programs are available to enable the comparison of multiple structures that are too numerous or too complex to be analysed by eye. These programs fall into three distinct groups and are used in different situations depending on the type and number of sequences being analysed (Figure 1.9).

The first approach is useful for identifying homologous structural motifs in closely related RNA sequences such as tRNAs. The program first performs a primary sequence alignment using a program such as ClustalW (described above) before looking for homologous structural features by identifying structure-neutral mutations. A program based on these principles, RNAdedecoder, was used to identify regions of structural homology within the HCV and polio virus genomes (Pedersen *et al* 2004).

The second method is known as the Sankoff algorithm and involves simultaneous folding and alignment of a set of homologous RNA sequences from which a common fold is inferred (Sankoff 1985). This algorithm requires such huge computing power and time that it has never been implemented in its original form. However, there are a number of programs that use restricted versions of the algorithm, such as Foldalign (Havgaard *et al* 2005a; Havgaard *et al* 2005b) and Dynalign (Mathews and Turner 2002; Mathews 2005). Although the Sankoff algorithm itself does not require the sequences to be related, programs using modified versions tend to involve an evolutionary component, so again this method is most effective for analysing related sequences.



The third method of RNA structural alignment does not require the sequences to be related because there is no primary sequence alignment step. Instead the RNA structures of each sequence are predicted using an MFE method such as Mfold or RNAfold (see above). These secondary structures are then compared in order to find a consensus structure. This type of approach is therefore most suited to the comparison of aptamer structures because they are not related by divergent evolution. Fewer programs exist that use this approach than for the first two methods; two examples are RSmatch (Liu *et al* 2005) and RNAforester (Hochsmann *et al* 2003; Hochsmann *et al* 2004).

A wide range of bioinformatics programs are available to analyse various aspects of RNA sequence sets, and the data obtained from these analyses can be used to direct further laboratory-based experiments.

1.4 Aims and scope of the thesis

The aims of this thesis are:

1. To map the binding site of aptamer B40 on gp120.
2. To design truncated synthetic derivatives of aptamer B40 that have the potential to be developed as anti-HIV-1 microbicides.
3. To investigate the stability of the B40 minimal binding motif and determine the extent of structural variation within the family of anti-gp120 aptamers.

Chapter 2 describes the common materials and methods used throughout this investigation and Chapter 3 gives details of the production of soluble gp120 for use in biochemical experiments. I cloned and sequenced a new set of aptamers from the SELEX round 5 pool of anti-gp120 aptamers produced by M. Khati. The ability of these aptamers to bind to gp120 and their structural

diversity are investigated in Chapter 4. In Chapter 5 I describe the design of B40 derivatives in which the minimal binding motif is stabilised and assess their suitability for microbicide development using a gp120-binding assay. Chapter 6 describes the mapping of the aptamer binding site on gp120 using mutagenesis. I tried to use a competition assay to find out whether or not other aptamers bind to the same site on gp120 and this is described in Chapter 7. In Chapter 8 I attempt to use a pseudovirus assay to investigate neutralisation of the B40 derivatives. The data is summarised and its wider implications are discussed in Chapter 9.

Chapter 2:

Materials and Methods

- Materials
- Manipulation of nucleic acids
- Cloning techniques
- Tissue culture techniques
- Expression of gp120
- Use of HIV-1 pseudovirus
- Protein techniques
- Statistical techniques
- Computational methods
- Surface plasmon resonance

Chapter 2: Materials & Methods

2.1 *Materials*

2.1.1 General suppliers

Aalto Bioreagents, Dublin, Eire	National Blood Service, Bristol, UK
Advanced Biomedical, Oldham, UK	New England Biolabs, Beverly, MA, USA
Avidity, Denver, CO, USA	NIH AIDS Research & Reagent Reference Program, (NIH ARRRP), Germantown, MD, USA
BD Biosciences, Oxford, UK	Pierce, Rockford, IL, USA
BIAcore, Uppsala, Sweden	Pharmacia LKB, Uppsala, Sweden
BDH Ltd, Poole, UK	Promega Corporation, Madison, WI, USA
Biogenesis, Poole, UK	QIAgen Ltd, Crawley, UK
Bio-Rad Laboratories Inc, UK	RNA-Tec, Leuven, Belgium
Bio-Whittaker, Poole, UK	Roche Diagnostics Ltd, Lewes, UK
Boehringer Mannheim, Indianapolis, IN, USA	Seikagaku Corporation, Tokyo, Japan
Calbiochem, San Diego, CA, USA	Sigma Chemical Company, St Louis, MO, USA
Cambrex, Nottingham, UK	Spectrum Laboratories, Breda, The Netherlands
Eurogentec, Seraing, Belgium	Stratagene, La Jolla, CA, USA
GE Healthcare, Chalfont St Giles, UK	TriLink BioTechnologies Inc, San Diego, CA, USA
Gibco BRL Life Technology, Paisely, UK	VWR International, West Chester, PA, US
Invitrogen, Groningen, The Netherlands	
Merck & Co Inc, Whitehouse Station, NJ, USA	
Millipore, Billerica, MA, USA	
MP Biomedicals Inc, Cambridge, UK	

2.1.2 General materials & suppliers

2'-fluorocytidine-5'-triphosphate	TriLink BioTechnologies
2'-fluorouridine-5'-triphosphate	TriLink BioTechnologies
4 % NuSieve pre-cast agarose gels	BioWhittaker
4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid (HEPES)	Sigma
40 % acrylamide/bis solution 19:1	Bio-Rad
5-bromo-4-chloro-3-indoyl- β -D-galactopyranoside (X-gal)	Sigma
Adenosine 5'-triphosphate	GE Healthcare
Agarose	Flowgen
Ammonium acetate	Sigma
Ammonium persulphate	Bio-Rad
Ampicillin sodium salt	Sigma
Azidothymidine	Sigma
Biotin-16-ddUTP	Roche
Boric acid	Merck
Bovine serum albumin	Sigma
Bromophenol blue	BDH
Calcium chloride	BDH
CCR5 sulphated N-terminal peptide	Advanced Biomedical
Chloroform	Sigma
Deoxyribonuclease I	Sigma
Deoxynucleotide triphosphates	GE Healthcare
Dimethyl sulphoxide	Sigma
Dithiothreitol	BDH
DNA molecular weight markers	Invitrogen
Dulbecco's Modified Eagle Medium	Gibco BRL
Empigen	Calbiochem
Ethanol	VWR International

Ethidium bromide	Sigma
Ethylenediaminetetraacetic acid	BDH
Foetal calf serum	Sigma
Ficoll-paque	GE Healthcare
Formaldehyde	Sigma
Formamide	Sigma
FuGENE 6 transfection reagent	Roche
G418 disulphate salt	Sigma
Glycerol	VWR International
Glycine	Sigma
Guanosine 5'-triphosphate	GE Healthcare
HIV-1 _{BaL} gp120, recombinant	NIH ARRRP
Hydrochloric acid	BDH
Hygromycin B	Sigma
ImmunoPure TMB substrate kit	Pierce
L-Glutamine	Sigma
Magnesium chloride	Sigma
Methanol	BDH
N,N,N',N'-tetramethylethylenediamine (TEMED)	Sigma
Oligonucleotides	Sigma
Penicillin/Streptomycin	Gibco BRL
Peroxidase-conjugated streptavidin	Pierce
Phenol:chloroform:isoamyl alcohol 125:24:1	Sigma
Phenol:chloroform:isoamyl alcohol 25:24:1	Sigma
Phosphate buffered saline solution	SWDSOP
Phosphate buffered saline tablets	Sigma
Potassium chloride	BDH
Protease inhibitor cocktail tablets	Roche
Puromycin	Sigma
Rantes, human recombinant	Sigma

RPMI-1640	Gibco BRL
Sephadex TM G50 fine	Pharmacia
Spectra/Por molecular porous membrane tubing	Spectrum laboratories
Sodium acetate	BDH
Sodium chlorate	Sigma
Sodium chloride	BDH
Sodium dodecyl sulphate	BDH
Sodium hydroxide	BDH
Spermidine	Sigma
Synthetic RNA oligonucleotides	RNA-Tec
T4 DNA ligase	New England Biolabs
T7 RNA polymerase	New England Biolabs
<i>Taq</i> DNA Polymerase	Promega
Tris base	MP Biomedicals Inc
Trypsin	Gibco BRL
Tween 20	Sigma
Ultra pure water	Sigma
Urea	BDH

2.1.3 Monoclonal antibodies

Name	Isotype	Antigen	Host	Conjugate	Supplier
10E4	IgM	Heparan sulphate	Mouse	None	Seikagaku
17b	IgG1	gp120	Human	None	NIH ARRRP
2D7	IgG2a	Human CCR5	Mouse	None	NIH ARRRP
2G12	IgG1	gp120	Human	None	NIH ARRRP
553046	IgG2a	Human CD4	Rat	FITC	BD Biosciences
555993	IgG2a	Human CCR5	Mouse	PE	BD Biosciences
b12	IgG1	gp120	Human	None	NIH ARRRP
B4	IgG2a	Human CD4	Mouse	None	NIH ARRRP
OX68	IgG1	Rat CD4d3&4	Mouse	None	Neil Barclay

2.1.4 Polyclonal antibodies

Antigen	Host organism	Conjugate	Supplier
gp120 C-terminal peptide	Sheep	None	Aalto Bioreagents Ltd
gp120	Goat	Biotin	Biogenesis
HIV	Human	None	NIH ARRRP
Human IgG	Goat	HRP	Pierce
Mouse IgG	Goat	HRP	Sigma
Mouse IgG	Mouse	PE	BD Biosciences
Control IgG1	Mouse	FITC	BD Biosciences
Control IgG2a	Mouse	PE	BD Biosciences

2.1.5 Buffers and solutions

2 × sample buffer for SDS-PAGE: 125 mM Tris-HCl pH 6.8; 20% (v/v) glycerol; 20% (w/v) SDS; 0.1 M DTT; 0.25% (w/v) bromophenol blue

5 × CHBS buffer: 50 mM HEPES; 752 mM NaCl; 5 mM MgCl₂; 5 mM CaCl₂; 13.5 mM KCl, pH 7.4

5 × fluoro-transcription buffer: 200 mM Tris-HCl ; 30 mM MgCl₂; 5 mM spermidine, pH 8.1

5 × SDS PAGE running buffer: 124 mM Tris; 450 mM glycine; 0.5% SDS

50 × TAE electrophoresis buffer: 1 M Tris; 50 mM EDTA, pH 8.5

20 × TBE electrophoresis buffer: 1.78 M Tris; 0.5 M boric acid; 50 mM EDTA, pH 8.0

Agarose DNA loading dye: 0.125% (w/v) bromophenol blue; 50% (v/v) glycerol; 0.8 mM Tris; 1.8 mM EDTA pH 8.5; 1 mM boric acid

FACS blocking buffer: 5 % rabbit serum in FACS washing buffer

FACS washing buffer: 0.5 % BSA; 2 mM NaN₃; 5 mM EDTA in PBS

Formamide loading dye: 95% (v/v) formamide; 0.25% (w/v) bromophenol blue; 0.25 % (w/v) xylene cyanol; 1 mM EDTA

Low salt buffer: 25 mM Tris-HCl pH 7.5; 140 mM NaCl; 0.02 % azide

High salt buffer: 25 mM Tris-HCl pH 7.5; 500 mM NaCl; 0.02 % azide

Phosphate buffered saline (PBS): 137 mM NaCl; 2.7 mM KCl; 4.3 mM Na₂HPO₄·7H₂O; 1.4 mM KH₂PO₄

PBS-T: 0.05 % Tween in PBS

PBS-BT: 0.05 % Tween; 1 % BSA in PBS

RNA elution buffer: 0.5 M ammonium acetate; 1 mM EDTA, pH 6.5

TBS: 20 mM Tris-HCl pH7.5; 500 mM NaCl

Transfer buffer: 25 mM Tris; 90 mM glycine; 20% methanol

Western washing buffer: 50 mM Tris-HCl pH 7.4; 150 mM NaCl; 0.01 % Tween 20

Western stripping buffer: 100 mM β-mercaptoethanol; 62.5 mM Tris-HCl pH 6.8; 2 % SDS

2.2 Manipulation of nucleic acids

2.2.1 Polymerase Chain Reaction

The polymerase chain reaction (PCR) is used to amplify DNA. Primers are designed that anneal to the 5' and 3' ends of the sequence following denaturation of the double-stranded DNA template. The primer is then elongated by the addition of complementary deoxynucleotide triphosphates (dNTPs) by a polymerase enzyme, such as *Taq* DNA polymerase. This process is repeated several times, with the PCR product from one round of amplification acting as a template for the next round, thus exponentially amplifying the sequence of interest. The reagents for a PCR are mixed with care to avoid contamination with foreign DNA, using barrier tips and positive displacement pipettes where necessary.

PCRs were performed using a PTC-200 Peltier Thermal Cycler DNA Engine (MJ Research). To amplify aptamer DNA, 0.2 µg DNA template was amplified by PCR in 1 × reaction buffer (Promega), 25 mM magnesium chloride, 0.25 mM dNTPs, with 15 µM primers and 0.1 U/µL *Taq* DNA polymerase (Promega) in a 500 µL reaction. After an initial denaturation step of 3 minutes at 95 °C, the cycling begins with a short denaturation step for 30 seconds at 95 °C. The primers are annealed for 1 minute at an appropriate temperature (according to the melting temperature of the primers and the template, see Table 2.2) followed by an extension time of 1 minute at 72 °C. This cycle is repeated 20-40 times and is followed by a final elongation step of 8 minutes at 72 °C. If required, a gradient PCR was performed in order to optimise the annealing temperature. In this case the annealing temperature was raised incrementally across a PCR block between two defined temperatures, e.g. from 55-75 °C. Table 2.1 outlines the primers and conditions used in each PCR and Table 2.2 gives the primer sequences. PCR products were analysed by agarose gel electrophoresis (Figure 2.1a) to assess

PCR	Primers	[Mg ²⁺] (mM)	Annealing temperature (°C)	Cycles
SELEX round 5 PCR	T3 SELEX T7 SELEX	2.5	72	20
TA colony PCR	T3 SELEX T7 SELEX	2.5	72	20
Aptamer PCR	T3 SELEX T7 SELEX	2.5	72	25
Aptamer PCR for B40t77	B40t_a B40t_b	2.5	51	25
PCR of gp120 DNA from baculovirus	Fp2Bac Rp2Bac	1.5	55	35
PCR of gp120 DNA to insert restriction sites	Gp120 XbaI fwd 3' gp120 Sal	2.5	72	35
pEF-BOS colony PCR	1391 1373	2.5	55	25
Quantitative PCR	LTR-P MH531 MH532	5	58	40

Table 2.1: Specific primers and conditions used in polymerase chain reactions.

Primer	Sequence (5' to 3')	T_m^a (°C)
T3 SELEX	AATTA ACCCT CACTA AAGGG AACTG TTGTG AGTCT CATGT CGAA	61
T7 SELEX	TAATA CGACT CACTA TAGGG AGACA AGACT AGACG CTCAA	60
B40t_a	TAATA CGACT CACTA TAGGG AGACA AGACT AGACG C	60
B40t_b	GGGAA ACAA CCAAT CGCG	58
Fp2Bac	CCGGA TTATT CATAC CGTCC CACC	70
Rp2Bac	GGGTC CAAGT TTCCC TGTA G AACTC	76
Gp120 XbaI fwd	ACCAT CGGGC GTTCT AGAGG ATCCA TGAGA	60
3' gp120 Sal	GTCGA CCCGC CTCCT CTTTT TTCTC TCTGC ACCAC	104
LTR-P ^b	CAGTG GCGCC GAACA GGGA	68
MH531	TGTGT GCCCG TCTGT TGTGT	62
MH532	GAGTC CTGCG TCGAG AGAGC	66
1391	GGGGG GAGGG GTTTT ATGCG ATGG	78
1373	CAGGG ATTTC TTGTC TCCCA CG	68
M13 fwd	GTAAA ACGAC GGCCA G	50
M13 rev	CAGGA AACAG CTATG AC	50
Env seq 5'	CCCCA TAATA GACAG TGA	52
Env seq 3'	ATAGC AGTAG CTGAG GGG	52
1711	CTACA AATGT GGTAT GGCTG	58
1712	TGACA CGAAG CTTGG GC	54

Table 2.2: Primer sequences^a Primer melting temperatures were calculated using the Wallace formula:

$$T_m = 2(A+T) + 4(C+G)$$

^b LTR-P has 5' 6-FAM and 3' TAMRA modifications

their purity and length, and to check for contamination in the negative control which is set up with no DNA template.

2.2.2 2'-fluoro *in vitro* transcription

Aptamer DNA that had been amplified by PCR was used as a template for 2'-fluoro *in vitro* transcription. Typically, 20 µg DNA template were incubated overnight at 37 °C in 1 × fluoro-transcription buffer (see Section 2.1.5), with 1 mM ATP, 1 mM GTP, 1 mM 2'-fluoro-CTP, 1 mM 2'-fluoro-UTP and 3.63 U/µL T7 RNA polymerase in a 250 µL reaction volume. The resulting transcripts were treated with DNase I for 30 minutes 37 °C in the above reaction conditions with the addition of 6 mM magnesium chloride and 100 mM sodium acetate pH 5.2. The enzymes were removed by phenol:chloroform extraction (see Section 2.2.6) and the RNA was ethanol precipitated (see Section 2.2.7) and resuspended in 50 µL water. Unincorporated nucleotides were removed by passing the sample through a Sephadex column (see Section 2.2.8). RNA transcripts were analysed by denaturing gel electrophoresis (Section 2.2.4) to assess the length of the transcript and the extent of DNase I treatment (Figure 2.1b).

2.2.3 Agarose gel electrophoresis

Pre-cast 4 % NuSieve agarose gels made with 1 × TBE buffer (see Section 2.1.5) containing ethidium bromide were used to analyse DNA fragments of less than 200 bp. 1 % agarose gels made in 1 × TAE buffer (see Section 2.1.5) were used to analyse DNA fragments and plasmids from 500 bp to several kbp in size. DNA ladders were used as markers against which the length of the DNA samples could be measured. Samples were mixed with agarose DNA loading dye (see Section 2.1.5) and were electrophoresed in TBE or TAE buffer respectively, containing 0.001 % ethidium bromide for approximately one hour at 100 V. The samples were visualised under UV light. When the samples were used for gel purification the gels were stained with ethidium bromide after electrophoresis.

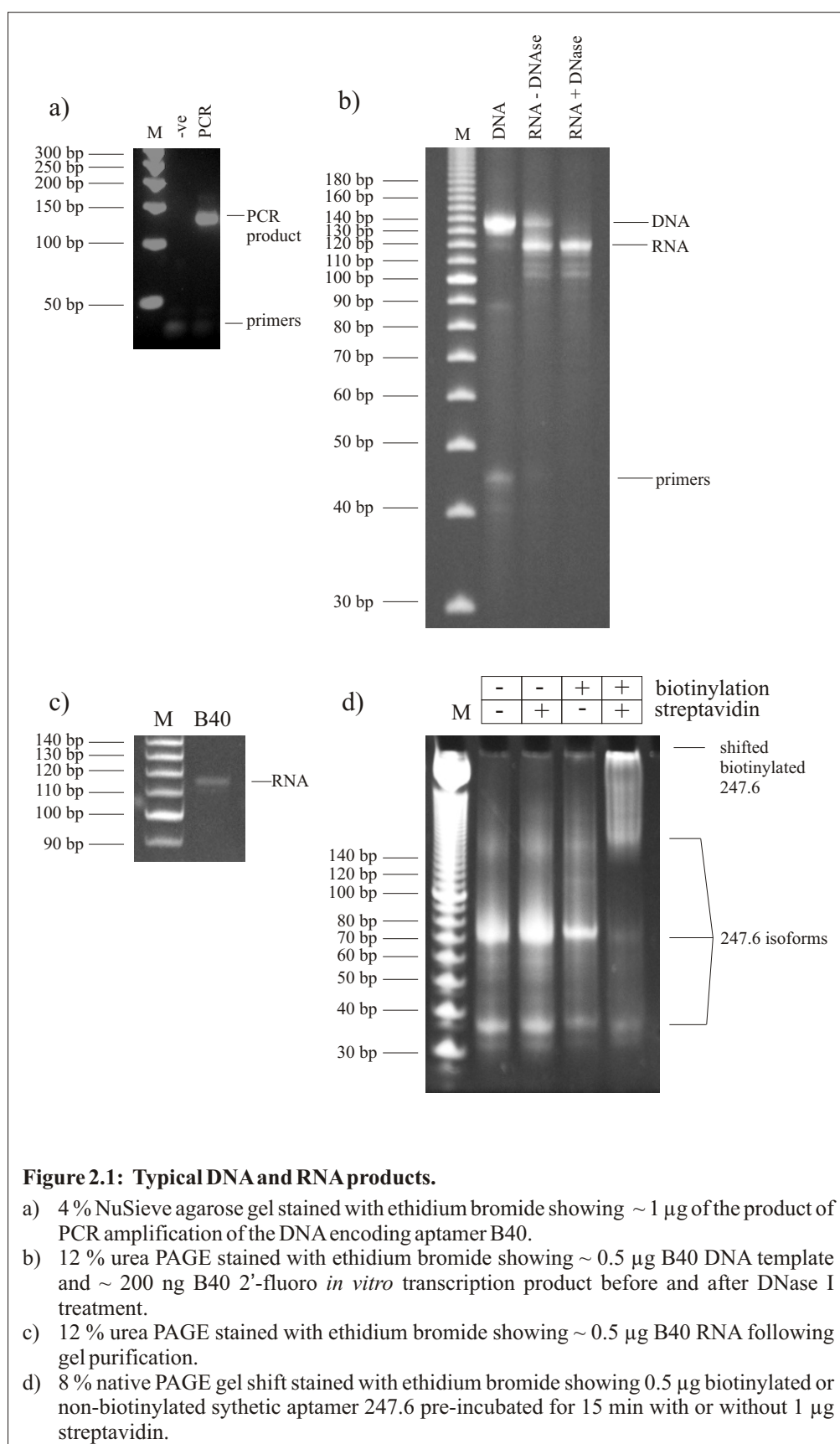


Figure 2.1: Typical DNA and RNA products.

- 4 % NuSieve agarose gel stained with ethidium bromide showing ~ 1 μ g of the product of PCR amplification of the DNA encoding aptamer B40.
- 12 % urea PAGE gel stained with ethidium bromide showing ~ 0.5 μ g B40 DNA template and ~ 200 ng B40 2'-fluoro *in vitro* transcription product before and after DNase I treatment.
- 12 % urea PAGE gel stained with ethidium bromide showing ~ 0.5 μ g B40 RNA following gel purification.
- 8 % native PAGE gel shift assay showing 0.5 μ g biotinylated or non-biotinylated synthetic aptamer 247.6 pre-incubated for 15 min with or without 1 μ g streptavidin.

2.2.4 Denaturing polyacrylamide gel electrophoresis

Denaturing polyacrylamide gel electrophoresis (PAGE) was used to analyse RNA fragments of less than 200 bp. For convenience, gels were prepared from a pre-made 25 % acrylamide-8M Urea mixture to give the appropriate acrylamide percentage as indicated in the Table 2.3 below, with a final concentration of 8 M Urea, $1 \times$ TBE, 0.05 % APS and 0.1 % TEMED in 50 mL.

Gel percentage	8 %	10 %	12 %	15 %
25 % acrylamide-8M Urea	16	20	24	30
8M Urea	31.3	27.2	23.3	17.2
$20 \times$ TBE	2.5	2.5	2.5	2.5
10 % APS	0.25	0.25	0.25	0.25
TEMED	0.05	0.05	0.05	0.05

Table 2.3: Composition of denaturing PAGE gels of varying percentages. All volumes are given in mL.

Gels were pre-run in $1 \times$ TBE buffer for 15 minutes at 30 W to ensure samples remained denatured once loaded. Samples (including DNA ladders) were mixed 1:1 with formamide buffer (see Section 2.1.5) and denatured for 3 minutes at 95 °C, then electrophoresed for one hour at 30 W. Gels were stained for 15 minutes in a 0.002 % solution of ethidium bromide in water and then visualised under UV light.

2.2.5 Native polyacrylamide gel electrophoresis

Native (non-denaturing) PAGE was used to analyse the extent of RNA biotinylation in a gel-shift assay (Figure 2.1d). Gels were made with 8 % polyacrylamide, 0.1 % TEMED and 0.06 % ammonium persulphate in $1 \times$ TBE buffer using the Bio-Rad mini-protean II system. Biotinylated RNA samples, which were pre-incubated with or without streptavidin, were mixed

with agarose gel DNA loading dye then electrophoresed in $1 \times$ TBE buffer for approximately one hour at 100 V. Gels were stained for 15 minutes in a 0.002 % solution of ethidium bromide in water and visualised under UV light.

2.2.6 Phenol:Chloroform extraction of nucleic acid

Proteins were removed from DNA or RNA samples by phenol:chloroform extraction. Phenol:chloroform:isoamyl alcohol 25:24:1 (DNA) or phenol:chloroform:isoamyl alcohol 124:24:1 (RNA) was added to the sample to give a volume ratio of 1:1 (typically 250 μ L of each). The emulsion was mixed by vigorously agitating the tube using a Vortex Mixer (Jencons) for two minutes, then the layers were separated by centrifugation for two minutes at 12 000 rpm in a benchtop MSE Micro Centaur centrifuge. The aqueous phase was harvested into a new tube, and the phenol phase was re-extracted with an equal volume of water. The emulsion was mixed and centrifuged again, then the aqueous phase was harvested and pooled with that from the first round of extraction. An equal volume of chloroform was added to the aqueous phases and the mixture was again mixed and centrifuged, and then the supernatant was recovered. Typically the sample was then ethanol precipitated immediately.

2.2.7 Ethanol precipitation of nucleic acid

DNA or RNA was precipitated by the addition of $1/10^{\text{th}}$ volume of 3 M sodium acetate pH 5.2 and 2.5 volumes of 100 % ethanol. This was incubated on dry ice for 20-30 minutes then centrifuged at 4 °C for 30-40 minutes at 17 000 rpm in a Heraeus Biofuge 22R. The supernatant was decanted and the pellet was resuspended in 200 μ L of 70 % ethanol then centrifuged again for 15 minutes at 4 °C at 17 000 rpm. The supernatant was decanted and the pellet was dried by centrifuging in a vacuum for 1-2 minutes. Finally the dried pellet was resuspended in an appropriate volume of water and stored at -20 °C.

2.2.8 Removal of free nucleotides

Free nucleotides or nucleosides were removed from PCR or *in vitro* transcription products by passing the sample through a Sephadex column. 2 mL Sephadex (G50 fine, DNA Grade) was loaded into a 5 mL spin column and washed with three volumes of water. The column was centrifuged in an MSE Mistral 1 000 centrifuge for 5 minutes at $500 \times g$ to remove excess water, then the sample was loaded onto the column and the centrifugation was repeated. The sample was collected in a small Eppendorf tube situated below the column.

2.2.9 Spectrophotometry

The concentration of DNA and RNA was estimated by reading the absorbance of the nucleic acid at a wavelength of 260 nm using a spectrophotometer (Nanodrop technologies). The purity of the nucleic acid was estimated by dividing the absorbance at a wavelength of 260 nm by the absorbance at a wavelength of 280 nm. The expected ratios for pure nucleic acid were 1.0 and 1.5 for DNA and RNA respectively.

2.2.10 Gel purification of RNA

Approximately 10 μ g of transcribed RNA aptamer was separated by 12 % denaturing PAGE (see Section 2.2.4) and was visualised by brief staining in ethidium bromide. The RNA band was cut out and crushed with a pipette tip, and the crushed gel was resuspended in 1.5 mL RNA elution buffer (see Section 2.1.5) and mixed with 1.5 mL phenol:chloroform:isoamyl alcohol 125:24:1 by vigorously agitating the tube using a Vortex mixer (Jencons) for 5 minutes then rolling for 30-60 minutes. The emulsion layers were then separated by centrifugation in a Heraeus Biofuge 22R for 10 minutes at 11 800 rpm at 4 °C and the aqueous phase was harvested and kept on ice. The phenol phase was re-extracted with 1.5 mL RNA elution buffer as before and the second aqueous phase was harvested and pooled with the first. The RNA

was then ethanol precipitated twice (Section 2.2.7) and resuspended in 20 μL water. The purity of the RNA was checked by gel electrophoresis (Figure 2.1c), and the concentration was estimated by spectrophotometry (Section 2.2.9). Normally about 4 μg RNA were recovered.

2.2.11 5' end-labelling of RNA with [γ - ^{32}P] ATP

4 μg gel-purified RNA were dephosphorylated with 0.2 units/ μL alkaline phosphatase (Amersham Biosciences) in $1 \times$ alkaline phosphatase buffer (Amersham Biosciences) in a 40 μL reaction volume for 1 hour at 37 $^{\circ}\text{C}$. The RNA was extracted with phenol:chloroform to remove the enzyme (see Section 2.2.6), then ethanol precipitated (see Section 2.2.7) and resuspended in 5 μL water.

The 4 μg dephosphorylated RNA was treated with 0.5 U/ μL T4 polynucleotide kinase (PNK, Roche) in $1 \times$ T4 PNK reaction buffer (Roche) in the presence of 35 μCi adenosine 5'-[γ - ^{32}P] triphosphate (Amersham) for 1-2 hours at 37 $^{\circ}\text{C}$. The resulting RNA molecule is 5'-end-labeled with ^{32}P . This RNA was subsequently purified from 12 % denaturing PAGE (see Section 2.2.10). The activity of the RNA was determined by Cherenkov counting, thus enabling the concentration of the labelled RNA to be estimated. Samples were used within two weeks of labelling while they retained sufficient activity (the half-life of ^{32}P is 14.3 days).

2.2.12 3' end-labelling of RNA with biotin-16-ddUTP

RNA aptamers were end-labelled with biotin- ϵ -aminocaproyl- γ -aminobutyryl-[5-(3-aminoallyl)-2', 3'-dideoxy-uridine-5'-triphosphate] tetralithium salt (biotin-16-ddUTP) by incubation with terminal transferase. Various reaction conditions were tested (see Section 7.3) and it was found that maximum biotinylation occurred when 3.5 μg aptamer were incubated with 100 μM biotin-16-ddUTP (Roche) and 2 U/ μL terminal transferase (New England Biolabs) in the presence of 2.5 mM cobalt (II) chloride and $1 \times$ reaction buffer

(New England Biolabs) in a 50 μ L reaction, incubated for 30 minutes at 37 °C. The reaction was stopped by the addition of EDTA at a final concentration of 40 mM and unincorporated biotin-16-ddUTP was removed by passing the sample through a Sephadex column (see Section 2.2.8). In order to assess the extent of biotinylation, 0.1 μ g biotinylated RNA was pre-incubated with 0.5 μ g streptavidin for 15 minutes at room temperature then analysed by native PAGE in a gel-shift assay (Figure 2.1d).

2.2.13 Aptamer refolding

Before the RNA aptamers were used in any assay they were refolded to maximise structural homogeneity in the solution. The RNA, diluted to an appropriate concentration in water, was denatured for 3 minutes at 95 °C, cooled on ice for one minute, then incubated for 10 minutes at room temperature, followed by the addition of 1/5th volume of 5 $\times$ CHBS buffer (see Section 2.1.5). Following a further 10 minute incubation at room temperature the RNA aptamer was incubated on ice for a period of at least one hour. If convenient, the RNA was refolded at a concentration higher than required in the assay and subsequently diluted in 1 $\times$ CHBS buffer or tissue culture medium as appropriate.

2.3 Cloning techniques

2.3.1 Restriction digests

Typically, 1 μ g plasmid DNA was incubated for 1 hour at 37 °C with 10 units of a restriction endonuclease in the presence of 1 $\times$ reaction buffer (supplied with the enzyme) and BSA if specified by the manufacturer. Where digestion with more than one enzyme was required, the plasmid was incubated with both enzymes concurrently if the buffer conditions were compatible. If not, the plasmid was digested with the first enzyme, the buffer conditions were

adjusted accordingly and then the plasmid was incubated with the second enzyme.

2.3.2 Alkaline phosphatase treatment

To prevent self-ligation, digested plasmids were treated with alkaline phosphatase to remove the 5'-phosphates on the sticky ends of the DNA. Therefore ligation could only occur between a plasmid and a DNA fragment (i.e. the inserted gene of interest) that had been digested with the compatible restriction enzymes but not treated with alkaline phosphatase. 1 µg plasmid were dephosphorylated with 0.2 U/µL alkaline phosphatase (Amersham Biosciences) in 1 × alkaline phosphatase buffer (Amersham Biosciences) in a 30 µL reaction for 1 hour at 37 °C. The DNA was extracted with phenol:chloroform (see Section 2.2.6) to remove the enzyme, then ethanol precipitated (see Section 2.2.7) and resuspended in 20 µL water.

2.3.3 Gel purification of DNA fragments

Approximately 1 µg of digested DNA fragments were separated by electrophoresis on 1 % agarose gels. The bands were visualised briefly using ethidium bromide and excised using a blade. The DNA was then extracted from the gel piece using the QIAquick Gel Extraction Kit (QIAGEN) and was eluted in 30 µL water. The concentration was estimated by spectrophotometry and the purity and size of the fragments were checked by agarose gel electrophoresis.

2.3.4 Ligation of DNA fragments

DNA fragments that had been digested with appropriate restriction enzymes, alkaline phosphatase treated and gel purified were ligated using T4 DNA ligase. Plasmid and insert were mixed together at molar ratios of 1:0, 1:1 and 1:3 and incubated overnight at 16 °C in the presence of 400 units T4 DNA ligase and 1 × reaction buffer (both New England Biolabs).

2.3.5 Transformation of competent cells

Bacteria were cultured using aseptic technique in LB medium (tryptone 10 g/L, yeast extract 5 g/L, sodium chloride 10 g/L) containing 100 µg/mL ampicillin or on LB agar plates (LB medium with 0.7 % agar) containing 100 µg/mL ampicillin unless otherwise specified. Aliquots of *Escherichia coli* Top10F' competent cells (Invitrogen) were thawed on ice. 2 µL DNA from each ligation reaction were mixed gently with the cells and incubated on ice for 30 minutes. The cells were heat-shocked for one minute at 42 °C, then 250 µL LB medium was added and the cells were incubated at 37 °C for 1 hour, shaking at 200 rpm. These cells were used to streak agar plates, which were incubated overnight at 37 °C. Colonies from the 1:1 or 3:1 plates were picked and used to inoculate 5 mL cultures, which were incubated overnight at 37 °C shaking at 200 rpm. The cells were then centrifuged at $500 \times g$ for 10 minutes and the supernatant was discarded. Plasmid DNA was extracted from the cell pellet using the QIAprep Spin Miniprep kit (QIAGEN) and was eluted in 50 µL water. Presence of the insert was confirmed by digestion and DNA sequencing. Glycerol stocks of transformed bacteria were made by mixing 0.7 mL bacterial culture with 0.3 mL 100 % sterile glycerol and were stored at -80 °C.

2.3.6 DNA sequencing

100 ng plasmid DNA and 3.2 pmol primer were mixed in a 6 µL reaction and the DNA was sequenced using dideoxy terminator sequencing by Julian Robinson through the SWDSOP in-house service. DNA sequences were analysed using BioEdit Sequence Alignment Editor software version 7.0.4.1 and plasmid maps were drawn using VectorNTI.

2.3.7 Amplification of plasmid DNA

DNA maxipreps were performed when large amounts of plasmid DNA were required, for example for transfection. Bacteria from a glycerol stock were streaked onto agar plates and incubated overnight at 37 °C. Colonies were picked and used to inoculate 5 mL 8 hour cultures, and these cultures in turn were used to inoculate 250 mL overnight cultures. The bacteria were centrifuged for 20 minutes at 6 000 rpm at 4 °C in a Beckman J2-21 centrifuge with a F10B rotor and the supernatant was discarded. DNA was extracted from the cell pellet using a HISpeed Plasmid Maxi Kit (QIAGEN) and was eluted in 1 mL sterile water. The DNA concentration was estimated using spectrophotometry (Section 2.2.9) and agarose gel electrophoresis (Section 2.2.3).

2.4 Tissue Culture techniques

2.4.1 Maintenance of cell lines

Human embryonic kidney cell line 293T was cultured in Dulbecco's modified eagle medium (DMEM) with the addition of 10 % Fœtal calf serum, 50 µg/mL penicillin, 50 µg/mL streptomycin and 2 mM L-Glutamine at 37 °C in 5 % CO₂. U87.CD4.CCR5 cells (from the NIH ARRRP) (Deng *et al* 1996) were cultured in DMEM with the addition of 10 % FCS, 50 µg/mL penicillin 50 µg/mL streptomycin, 2 mM L-Glutamine, 300 µg/mL G418 and 1 µg/mL puromycin at 37 °C in 5 % CO₂. Chinese Hamster Ovary cell line K1 (CHO-K1 cells) were cultured in CB2 medium (a variant of Glasgow Minimal Essential Medium) with the addition of 10 % FCS, 50 µg/mL penicillin, 50 µg/mL streptomycin and 2 mM L-Glutamine, unless under MSX selection in which case L-Glutamine was not added.

Cells were detached from tissue-culture treated plastic with a solution of 0.25 % trypsin and 1 mM EDTA in PBS, and were routinely passaged by

between 1:3 and 1:10 dilution (depending on confluence) two to three times per week. Stocks of all cell lines were stored under liquid nitrogen in 90 % dimethylsulphoxide (DMSO) and 10 % FCS. FCS was heat-inactivated for 30 minutes at 56 °C and filtered through 0.2 µm before use.

2.4.2 Preparation and culture of PBL

Buffy coats were acquired from the National Blood Service in Bristol. 50 mL buffy coat was diluted 1:1 with 50 mL PBS then added slowly to three tubes each containing 15 mL Ficoll-paque. A Ficoll gradient was produced by centrifugation of the tubes in a Beckman GPR centrifuge for 20 minutes at 2 000 rpm with no brake. The serum layer was discarded and the lymphocytes were harvested and resuspended in 50 mL ice-cold PBS. After a further centrifugation for 10 minutes at 1 500 rpm with half brake, the cells were washed 3-5 times by re-suspending in 50 mL ice-cold PBS and centrifugation for 5 minutes at 1 000 rpm with full brake. After the final wash the cells were resuspended in RPMI-1640 with 10 % FCS and 1 µg/mL phytohemagglutinin (PHA, Sigma) and were distributed in flasks at 1×10^6 cells/mL. After three days of incubation at 37 °C in 10 % CO₂ the peripheral blood lymphocytes (PBLs) were stimulated with 70 U/mL interleukin-2 (IL-2, Pharmacia). Cells were counted every few days and diluted in medium to maintain a concentration of 1×10^6 cells/mL.

2.4.3 Flow cytometry

The surface molecules of various cell types were analysed by flow cytometry. Cells were detached from tissue culture-treated plastic with 1 mM EDTA if required, and were centrifuged for 10 min at 1 000 rpm then resuspended in FACS blocking buffer (see Section 2.1.5) and distributed in 96-well plates at 5×10^5 cells/well in 50 µL/well. After a 30 minute incubation on ice, 50 µL/well of primary antibody diluted in FACS blocking buffer was added at a concentration that had previously been determined by titration. The cells were incubated on ice with primary antibody for one hour, then washed three

times in FACS washing buffer (see Section 2.1.5) and finally resuspended in 50 μ L of secondary antibody that had been diluted in FACS blocking buffer to a suitable concentration. The cells were incubated with the secondary antibody for one hour on ice, and then washed four times as before. Finally the cells were resuspended in 1 mL FACS washing buffer and analysed using the FACSsort flow cytometer (Becton Dickinson). Typically 50 000 events were counted per sample, and the data were analysed using CellQuest version 3.3 (BD Biosciences) and WinMDI version 2.8 software.

2.4.4 Transient transfections

2.4.4.1 *Calcium phosphate transfection*

Cells to be transfected were seeded in T75 flasks so that they would be at 35-50 % confluence the following day. The tissue culture medium was replaced two hours prior to transfection. A 1 mL solution of 20 μ g sterile plasmid DNA in 0.25 M calcium chloride was added slowly (over a period of 1 minute) to 1 mL $2 \times$ HBS buffer (see Section 2.1.5), while the tube was vigorously mixed using a Vortex mixer. This caused a cloudy precipitate to form, which was slowly added to the medium on the cells. The cells were then returned to the incubator, and the medium was replaced the following day to remove the precipitate. The cells were then incubated for several days until required.

2.4.4.2 *FuGENE 6 transfection*

Cells to be transfected were seeded in T75 flasks so that they would be at 50 % confluence the following day, and were incubated overnight. The tissue culture medium was replaced two hours prior to transfection. Typically, 100 μ L serum-free DMEM was added to an Eppendorf tube, and 30 μ L FuGENE 6 (Roche) was added slowly. Care was taken not to let the FuGENE 6

directly touch the plastic tube. 10 µg sterile plasmid DNA in water was then added slowly to the mixture of DMEM and FuGENE, and the mixture was incubated at room temperature for 15 minutes. The amounts of DMEM, FuGENE 6 and DNA added were varied according to the needs of each individual experiment. The mixture was added slowly to the tissue culture medium, during which the flask was agitated and care was taken to ensure the mixture did not contact the cells directly. The cells were then returned to the incubator, and the medium was replaced the following day. The cells were then incubated for several days until required.

2.4.4.3 *PEI transfection*

Cells to be transfected were seeded in T75 flasks so that they would be at 50 % confluence the following day, and were incubated overnight. The tissue culture medium was replaced two hours prior to transfection. Typically, 100 µL serum-free DMEM was added to an Eppendorf tube, and 10 µg sterile plasmid DNA in water was added slowly. 50 µL poly(ethylenimine) (PEI) was then added slowly to the mixture of DMEM and DNA, and the mixture was incubated at room temperature for 15 minutes. The amounts of DMEM, DNA and PEI added were varied according to the needs of each individual experiment, and in particular the ratio of DNA:PEI was optimised for each plasmid. The mixture was slowly added to the tissue culture medium, during which the flask was agitated and care was taken to ensure the mixture did not contact the cells directly. The cells were then returned to the incubator, and the medium was replaced the following day. The cells were then incubated for several days until required.

2.4.5 Desulphation and removal of cell-surface heparan sulphate proteoglycans

U87.CD4.CCR5 cells were cultured for one week in the presence of 20 mM sodium chlorate to inhibit the sulphation of heparan sulphate proteoglycans

(HSPGs). Sodium chlorate inhibits the enzyme sulphate adenylyltransferase, which is part of the pathway that results in HSPG sulphation, without reducing cell viability. During the one week period the cells were treated with trypsin three times in order to remove residual sulphated HSPGs.

To remove cell surface HSPGs, U87.CD4.CCR5 cells were distributed in flat-bottom 96-well plates at 1.5×10^4 cells/well. After an overnight incubation the medium was removed from several wells and replaced with 20 μ L of heparitinase (Seikagaku) at 0.1 U/ml for 3 hours at 37 °C.

2.5 Expression of gp120

2.5.1 Expression of gp120 fusion proteins in 293T cells

293T cells were transfected with plasmid constructs expressing gp120 fusion proteins (see Section 3.2) by one of the three methods described in Section 2.4.4. Typically, cells in a T75 flask were transfected with 10 μ g plasmid. The following day, the medium was replaced and the cells were incubated at 37 °C for two to four days. When the medium was later going to be concentrated (see Section 2.5.2), serum-free medium such as X-vivo 10 (Invitrogen) was used. Optimisation showed that maximum protein expression occurred four days post-transfection (see Section 3.2.2). At this point, the tissue culture supernatant was harvested, centrifuged for 10 minutes at 2 000 rpm to remove cell debris, then filter sterilised through 0.2 μ m. The cell supernatant was either stored at 4 °C for short periods, or was stored in aliquots at -20 °C. Alternatively protein was harvested from the cell lysate, in which case the cells were harvested by trypsinisation and centrifuged for 10 minutes at 1 000 rpm. The cell pellet was then resuspended in 2 mL PBS and was subjected to three freeze-thaw cycles to lyse the cells. Then the sample

was centrifuged for 10 minutes at 1 000 rpm to remove cell debris and the supernatant was stored at 4 °C.

2.5.2 Concentration of supernatant

15 mL tissue culture supernatant containing the gp120-CB fusion protein (see Section 3.2) were concentrated in a Centriprep YM-10 spin column (Millipore) according to the manufacturers instructions. This resulted in a 23-fold concentration of the protein (0.65 mL final volume). The protein could only be concentrated in this way when it was expressed in serum-free medium such as X-vivo 10 (Invitrogen).

2.5.3 Biotinylation of gp120-CB

The gp120-CB fusion protein was biotinylated with BirA biotin ligase (see Section 3.2). 0.65 mL concentrated tissue culture supernatant was diluted to 15 mL in 10 mM Tris-HCl pH 8, then re-concentrated to 0.65 mL (Section 2.5.2). 0.65 mL gp120-CB was mixed with 81 µL Biomix A and 81 µL Biomix B and 1 µL BirA (all Avidity). If required, reaction volumes were scaled up and the gp120-CB:Biomix A:Biomix B ratio was retained at 8:1:1. The reaction was incubated overnight at room temperature. The protein was then dialysed in 1 L PBS with three changes of buffer over two days to remove any unbound biotin. The protein was then stored in aliquots at -20 °C and repeated freezing and thawing was avoided.

2.5.4 Other protein sources

HIV-1 gp120 from strain BaL that was expressed in Sf9 cells was produced by A. Dey. HIV-1 gp120 from strain IIIB and soluble human CD4 were both expressed in CHO-K1 cells and were produced by M. Khati. HIV-1 gp120 from strain BaL that was expressed in 293T cells was obtained from the NIH ARRRP.

2.6 Use of HIV-1 pseudovirus

2.6.1 Generation of HIV-1 pseudovirus

293T cells were distributed in 6-well plates at 3×10^5 cells/well and incubated overnight. The following day the supernatant was replaced with 2 ml fresh medium 2 hr prior to transfection. 4 µg sterile pNL4.3-Luc.R-E- DNA and 2 µg sterile pSVIII-Env DNA were transfected into 293T cells using 18 µL FuGENE 6 (Roche) according to the manufacturer's instructions. Briefly, the DNA was added dropwise to an Eppendorf tube already containing 100 µL DMEM and the FuGENE 6. After a 15 minute incubation at room temperature the mixture was added dropwise to the cells, whilst the 6-well plate was agitated. Alternatively the transfection was performed in a T25 or T75 flask, in which case the number of cells and amounts of DNA and FuGENE 6 were scaled up accordingly. The medium was replaced 24 hours post-transfection, and the cell supernatant containing the pseudovirus was harvested after a further 48 hours. The supernatant was centrifuged for 10 minutes at 2000 rpm to remove cell debris, then filtered through 0.2 µm to remove contaminating DNA. When the sample would be used subsequently for qPCR analysis, 200 µL supernatant was treated with DNase I (700 U) for 1 hr at 37 °C in the presence of 10 µM sodium acetate pH 7 and 5.6 µM magnesium chloride. The pseudovirus was then either used to transduce target cells (see below) or was stored in aliquots at -80 °C. Alternatively, 1 % Empigen was added to the sample and the pseudovirions were lysed by agitating briefly using a Vortex mixer in order to release gp120 that could later be used in a biochemical assay.

2.6.2 Transduction of target cells

Pseudovirus was used to transduce target cells that express CD4 and CCR5; either the cell line U87.CD4.CCR5 or PBLs. U87.CD4.CCR5 cells were seeded in flat-bottom 96-well plates at a density of 1.5×10^4 cells/well in

200 μL medium and incubated overnight in order for the cells to form an adherent monolayer. The medium was removed from the cells and replaced with 100 μL of a 5-fold serial dilution of pseudovirus. PBLs that had been PHA- and IL-2-stimulated for four days (see Section 2.4.2) were seeded in a round-bottom 96-well plate at a density of 1×10^5 cell/well in 50 μL /well, and 50 μL of a 5-fold serial dilution of pseudovirus was added. Alternatively a 6-well plate was used for the assay (so that the DNA could be extracted from the cells for a qPCR) in which case the cell number and volumes were scaled up accordingly. The cells were incubated with the pseudovirus for three days to allow transduction to occur, at which point transduction efficiency was measured either by luciferase activity (Section 2.6.4) or quantitative PCR (Section 2.6.5).

2.6.3 Neutralisation of HIV-1 pseudovirus

U87.CD4.CCR5 target cells were seeded as above in preparation for transduction. The pseudovirus was diluted to an appropriate concentration (one that would give several thousand luciferase response units when transduction was uninhibited) and incubated for 1 hr at 37 °C in the presence of a serial dilution of inhibitor (either antibody or aptamer). The medium was removed from the target cells and replaced with 100 μL of pseudovirus and inhibitor and incubated at 37 °C. After 24 hours the pseudovirus and inhibitor were removed from the cells and replaced with 100 μL fresh medium. The cells were incubated for a further three days and then transduction efficiency was measured by luciferase activity (Section 2.6.4).

2.6.4 Detection of luciferase activity

To detect luciferase activity in the transduced cells, the cell supernatant was removed (following a 10 min centrifugation step at 1000 rpm for PBLs) and 100 μL 1 x lysis buffer (Promega) was added. The cells were dislodged from the surface of the wells and mixed vigorously by pipetting up and down. The cell lysates were then stored at -20 °C for at least two hours, then thawed and

centrifuged at 2000 rpm for 10 minutes. 50 μ L of the resulting supernatant was then transferred into a white, flat-bottom 96-well plate and 50 μ L Bright-Glo luciferase substrate (Promega) was added. The reaction of the substrate catalysed by the luciferase in the supernatant from the cell lysate produces a chemiluminescent signal which was detected using a Lucy 3 luminometer (Anthos).

2.6.5 Detection of HIV-1 DNA by quantitative PCR

The rationale behind quantitative PCR (qPCR) is to detect the amount of DNA product produced after each PCR cycle and thus extrapolate the number of copies of the detected sequence that were in the input of the reaction. This method was first developed by (Heid *et al* 1996). The sequence is detected by a pair of primers that are designed to anneal at the 5' and 3' end of the required sequence. In addition, a third probe primer is designed, which anneals in between the 5' and 3' primer sequences once the PCR product has been synthesised. The probe DNA has a fluorescent reporter dye and a quencher at the 5' and 3' ends respectively. When the dye and the quencher are in this close proximity there is no signal from the dye. Once the probe has annealed to the PCR product the 5' to 3' exonuclease activity of the DNA polymerase cleaves the 5' end of the probe releasing the fluorophore. At this point the fluorophore is released into the solution so it is no longer quenched and it produces a fluorescent signal which is directly proportional to the number of copies of PCR product. This fluorescence is detected by a real-time PCR machine. It is possible for multiple reactions to occur using the same DNA sample if different primers and probes are used that have fluorophores that emit with different spectra.

The qPCR was set up such that there were two reactions occurring in the same tube. The first primer pair (from Butler *et al* 2001) amplify a region of HIV-1 cDNA between the 5' LTR sequence and the 5' end of the *gag* gene, which will be detected once reverse transcription is completed. The second primer

pair amplifies a 92 bp region of the human β -actin gene to normalise for the number of cells per sample (supplied as a kit from Eurogentec). Transduced cells were harvested from the 6-well plate and a DNA extraction was performed using the DNEasy tissue kit (QIAGEN), resulting in each DNA sample being eluted in 200 μ L water then concentrated to 30 μ L using a vacuum. Approximately 15 ng DNA was used per sample in the qPCR. The pNL4.3Luc.R⁻E⁻ plasmid and human genomic DNA, serially diluted to known copy numbers, were used as positive controls for the detection of HIV-1 DNA and the β -actin gene respectively. The HIV-1 primer sequences are given in Table 2.2, and the probe has a conjugated 5' 6-FAM fluorophore and 3' TAMRA quencher. The β -actin primers and probe were supplied with the β -actin control kit (Eurogentec), and the probe has a 5' Yakima Yellow fluorophore and 3' Eclipse DarkQuencher. The qPCR MasterMix (Eurogentec) contains the dNTPs, reaction buffer, magnesium chloride (5 mM final concentration) and HotGoldStar DNA polymerase. The addition of Tween 20 was found to increase the resolution of detection. The reaction components were mixed as follows: 1 $\times$ qPCR MasterMix; 0.45 μ M HIV-1 primers; 0.45 μ M HIV-1 probe; 0.8 μ M β -actin primers; 0.25 μ M β -actin probe; 15 % Tween 20. 20 μ L of the PCR mixture was added to 10 μ L of each sample or water for a blank control. After an initial denaturing step of 50 °C for 2 minutes then 95 °C for 10 minutes, the DNA was amplified by 40 cycles of 15 seconds denaturing at 95 °C followed by 90 seconds of elongation at 58 °C. The qPCR was performed in the Chromo 4 four-colour real-time PCR detector (Bio-Rad) and analysed with Opticon Monitor analysis software.

2.7 Protein techniques

2.7.1 SDS polyacrylamide gel electrophoresis

SDS PAGE was used to separate proteins according to molecular weight. Gels were poured using the Bio-Rad Mini Protean II system. Firstly, a 10 % separating gel was prepared with 0.1 % SDS, 10 % acrylamide/Bis, 0.05 % APS, 0.05 % TEMED in 1M Tris-HCl pH 8.8. 70 % isopropanol was layered over the gel while setting to ensure an even finish. Once the gel had set the isopropanol was removed and a 4 % stacking gel was prepared with 0.1 % SDS, 4 % acrylamide/Bis, 0.05 % APS, 0.05 TEMED in 0.125 M Tris-HCl pH 6.8. Samples were mixed with SDS PAGE loading dye (see Section 2.1.5) then heat-denatured for 3 minutes at 95 °C prior to loading. Pre-stained protein standards (Bio-Rad) were loaded concurrently. The samples were electrophoresed at 100 V for 1 hour in SDS PAGE running buffer (see Section 2.1.5).

2.7.2 Western blotting

Samples separated by SDS PAGE were transferred to Immun-Blot PVDF (polyvinylidene fluoride) membrane (Bio-Rad) by applying a current either for 2 hours at 100 V or overnight at 20 V in transfer buffer (see Section 2.1.5). The PVDF membrane was pre-wetted in 100 % methanol, rinsed with distilled water, then equilibrated in transfer buffer before use. All subsequent steps were performed in a small tray on a rocking apparatus at room temperature unless otherwise indicated. The membrane was blocked for two hours or overnight in 10 mL Western blocking buffer (see Section 2.1.5) then rinsed briefly with Western washing buffer (see Section 2.1.5). The primary antibody was diluted to an appropriate concentration (usually 1 µg/mL) in 5 mL Western blocking buffer and the membrane was incubated in primary antibody for one hour. The membrane was then washed with Western washing buffer by three 15 minute incubations, then incubated for one hour in

the secondary peroxidase-conjugated antibody, which again had been diluted to an appropriate concentration in 5 mL Western blocking buffer. The membrane was then washed four times with 15 minute incubations in Western washing buffer. Finally, ECL detection reagents (GE Healthcare) were mixed 1:1 and applied to the membrane at 0.125 mL/cm², which was then exposed to Kodak MS film for one minute in the dark. If required, the antibodies were stripped from the membrane by incubation at 50 °C for 30 minutes in Western stripping buffer (see Section 2.1.5) and then the membrane was blocked and re-probed with different antibodies using the same method as above.

2.7.3 Quantitation of bands using ImageQuant

ImageQuant software version 5.2 (Molecular Dynamics) was used to quantify bands on Western blots. The film image of the blot was captured digitally by an HP ScanJet 2400 using HP Photo & Imaging Gallery software version 1.1. The resulting Tagged Image File (TIF) was cropped to the region of interest using Picture Manager (Microsoft). The file was then opened in ImageQuant Tools version 2.2 (Molecular Dynamics) and converted to “.gel” format so that it could be opened in the ImageQuant program. Here the contrast of the image was adjusted to use the optimal spectrum (without affecting the underlying data). Bands were outlined with the rectangle tool, and the background was corrected with the Local Area setting so that the outline of the box around each band was used to set the background. The volume (in arbitrary units) is the integrated intensity of all the pixels in the rectangle, excluding the background. A volume report was then created to analyse the density of each band, and these values were compared to a band of known concentration on the same gel.

2.7.4 Gp120 ELISA

Gp120 concentrations were analysed using a standard ELISA (enzyme-linked immunosorbent assay) procedure developed by the Burton group (Pantophlet *et al* 2003). All volumes were 50 µL/well and all steps were performed at

room temperature on rocking apparatus unless otherwise stated. Antibody D7324 (Aalto Bioreagents) was diluted to 5 µg/mL in PBS and adsorbed to 96-well plates overnight at 4 °C. The plates were washed twice with PBS-T (see Section 2.1.5), then blocked for one hour with 3 % BSA in PBS. The BSA was removed and replaced by a serial dilution of gp120 made in PBS-BT (see Section 2.1.5) in triplicate, and incubated for 2-4 hours. The plates were washed three times with PBS-T. HIV polyclonal serum (HIV-Ig, NIH ARRRP) was diluted to 1 µg/mL in PBS-BT and added to the wells for one hour, then the plates were washed five times with PBS-T. Peroxidase-conjugated anti-human IgG raised in goat was diluted 1 000-fold and added to the wells for one hour. After five more washes with PBS-T, ImmunoPure TMB substrates (Pierce) were mixed 1:1 and added to the wells. After 10-30 minutes a blue colour developed, and at this point the reaction was stopped by the addition of 2 M sulphuric acid to the wells. The absorbance at a wavelength of 450 nm was read using an Emax ELISA plate reader (Molecular Devices).

2.7.5 Inhibition ELISA

An inhibition ELISA was used to estimate the concentration of rat CD4 domains 3 and 4 (rCD4d3&4) fusion proteins by comparison to soluble rat CD4 as a control. This technique was developed by the Barclay group (Brown and Barclay 1994). All volumes were 50 µL/well and all steps were performed at room temperature on rocking apparatus unless otherwise stated. A target plate was prepared by adsorption of rat CD4 at 5 µg/mL in PBS for one hour. The plate was washed once with PBS-T, then blocked with a solution of 0.5 % BSA in PBS for 30 minutes. Meanwhile, an inhibitor plate was prepared in which a 2-fold serial dilution of rCD4d3&4-fusion protein and a 5-fold serial dilution of soluble rat CD4 (starting from 10 µg/mL) was made in PBS. Tissue culture supernatant containing monoclonal antibody OX68 (from Neil Barclay), which binds to rCD4d3&4, was added to the inhibitor plate at an appropriate dilution (made in 0.5 % BSA in PBS) that had

previously been determined by titration (courtesy of Steve Simmonds), and was incubated for one hour. After these incubations the blocking solution was removed from the target plate and was replaced with the serial dilutions from the inhibitor plate, in duplicate, and incubated for one hour. The target plate was then washed three times with PBS-T. Peroxidase-conjugated anti-mouse IgG was diluted 5 000-fold in PBS-BT and added to the target plate for one hour. After five more washes with PBS-T, ImmunoPure TMB substrates (Pierce) were mixed 1:1 and added to the wells. After 10-30 minutes a blue colour developed, and at this point the reaction was stopped by the addition of 2 M sulphuric acid to the wells. The absorbance at a wavelength of 450 nm was read using an Emax ELISA plate reader. Unlike a standard ELISA in which the absorbance is directly proportional to the concentration of antigen in the sample, in an inhibition ELISA the absorbance is inversely proportional to the concentration of antigen in the sample.

2.7.6 Competition assays

Competition assays were performed in 96-well plates. Assays using streptavidin-coated plates were performed with 100 μ L reagent volumes, and all other assays were performed with 50 μ L reagent volumes. HRP-conjugated streptavidin, HRP-conjugated anti-human IgG, HIV-Ig and 2G12 were all used at 1 μ g/mL. Washes were performed with PBS-T. Aptamers were refolded in CHBS on the same day as the assay was performed (see Section 2.2.13), and samples were analysed in triplicate. If TMB (Pierce) was used as the detection reagent the reaction was stopped by the addition of an equal volume of 2M sulphuric acid and the absorbance was read at a wavelength of 450 nm using a spectrophotometer. If SuperSignal (Pierce) was used as the detection reagent the luminescence was measured using a luminometer immediately after the addition of the substrate. Data are presented as the percent of binding achieved in the absence of competitor, and then are transformed to show the percent inhibition. More specific details are given in the description of each particular assay in Chapter 7.

2.7.7 Slot-blot

A PVDF membrane was assembled in a slot-blot apparatus (Bio-Rad) and washed with 100 μ L TBS-T (see Section 2.1.5) per slot, then serial 5-fold dilutions of samples in TBS-T were added to the slot-blot at 100 μ L per slot. Gp120-CB dilutions were made starting from 1 μ M and pseudovirus-derived gp120 samples were made starting from neat. The membrane was washed with 200 μ L TBS-T per slot then removed from the apparatus. The membrane was blocked for 1 hour with 5 % milk powder in TBS-T followed by two 5-minute washes with TBS-T. For antibody labelling the membrane was stained with 5 mL 2G12 at 1 μ g/mL in TBS-T for 1 hour then washed twice for 5 minutes with TBS-T. The membrane was then stained with HRP-conjugated anti-human IgG diluted 1000-fold in TBS-T for 1 hour, then washed three times for 5 minutes in TBS-T. ECL substrate was used to develop the blot, which was exposed to a STORM phosphorimage screen overnight and the results were analysed using ImageQuant software. The same blot was then washed twice for 5 minutes with TBS-T and re-probed with aptamer B40. Refolded, radiolabelled aptamer B40 was added to the membrane at 0.33 nM in 5 mL TBS-T, then washed once with TBS-T. The membrane was then exposed to a phosphorimage screen overnight as above.

2.7.8 Plate-based assay for aptamer-gp120 interaction

All volumes are 50 μ L and steps were performed at room temperature unless otherwise stated. Antibody D7324 (5 μ g/mL in PBS) was adsorbed to 96-well ELISA plates overnight at 4 °C. The plate was then washed twice with PBS-T and blocked for 1 hour with 3 % BSA in PBS. Wild-type or mutant gp120 from lysed pseudovirus supernatants were diluted to 100 ng/mL in PBS-BT and were added to the plates for 3 hours. Meanwhile, radiolabelled aptamers B40, B4, B156 and B197 were refolded and 5-fold serial dilutions

were made in CHBS. The plates were washed three times with PBS-T and the aptamers were added to the plates for one hour. Samples were added in triplicate and their order on the plates was randomised. The plate was washed once with PBS-T, wrapped on paper towels to dry the wells and then exposed to a phosphorscreen overnight. The resulting image was analysed using ImageQuant software. The volume of each spot was determined and was normalised for the concentration and activity of each sample, and the amount of binding seen in the absence of any gp120 was subtracted.

2.7.9 Immunoaffinity chromatography

A 7 mL column was prepared in which purified OX68 antibody was covalently immobilised on a sepharose 4B matrix. 500 mL tissue culture supernatant from CHO-K1 cells containing the gp120-CB fusion protein was applied to the column and allowed to run overnight. The column was washed with 20 column volumes of low salt buffer (see Section 2.1.5) then pulsed with 5 column volumes of high salt buffer (see Section 2.1.5) then 5 column volumes of low salt buffer. 1 mL fractions were collected and immediately neutralised with 50 μ L 1 M Tris pH8 until the pH of the eluant reached 2.5. Each fraction was analysed by spectrophotometry and the absorbance at a wavelength of 280 nm was measured to ascertain if there was any protein in the fraction. The exact concentration of the gp120-CB in a number of fractions was determined using the bicinchoninic acid (BCA) method (Pierce) and by SDS PAGE (see Section 2.7.1.).

2.8 Statistical techniques

Statistical analysis and curve-fitting was performed using GraphPad Prism software version 3.02.

2.8.1 Nonlinear regression

Binding curves, for example in ELISA (Section 2.7.4), inhibition ELISA (Section 2.7.5) and aptamer-gp120 binding assays (Section 2.7.8) were analysed by nonlinear regression. A graph was created in which sample concentration was plotted on the x axis and the observed response (i.e. absorbance at a wavelength of 450 nm or luciferase response) was plotted on the y axis. Replicates were plotted and error bars representing the standard error were used. Any background values (such as the binding of antibody with no gp120) were subtracted from the binding curve to set the baseline at zero, then the sample concentration values were converted to logarithmic values using the transform function. The sigmoidal dose-response curve was modelled using the in-built equation:

$$y = bottom + \frac{top - bottom}{1 + 10^{(\log EC_{50} - x)Hillslope}}$$

where x is the logarithm of concentration, y is the response and y starts at *bottom* and goes to *top* with a sigmoid shape.

The value for *bottom* was set to zero because of the previous baseline subtraction. If required, *top* was also set to a fixed value (for example the *top* value generated for wild-type gp120 was the set as the *top* for analysis of mutants of gp120). Various parameters such as the $\log EC_{50}$ and *Hillslope* were calculated automatically by the program.

2.8.2 Derivation of sample concentration

In an inhibition ELISA, the standard curve of known rat CD4 concentration was modelled by nonlinear regression as above and the parameters for *top*, *bottom*, $\log EC_{50}$ and *Hillslope* were noted. A graph of the binding curve of a sample to OX68 was created, and the baseline was removed and the data were

transformed to the logarithmic value on the x axis as above. The data were then transformed by the user-defined equation:

$$x = \log EC_{50} - \frac{\log_{10}(top - bottom)}{\left(\frac{(y - bottom) - 1}{Hillslope} \right)}$$

where x is the logarithm of concentration, y is the response and the values for $\log EC_{50}$ and *Hillslope* are those defined by the standard curve.

This transformation creates a new graph in which the x axis is the logarithm of sample concentration and the y axis is the logarithm of rat CD4 concentration. The new curve is a straight line which was then analysed by nonlinear regression using the in-built function. The y -intercept gives the logarithm of rat CD4 concentration where the logarithm of the sample concentration is zero, i.e. where the concentration of the sample is 1 which corresponds to the undiluted sample. Therefore the concentration of the sample can be derived by solving the equation:

$$x = 10^y$$

where x is the logarithm of sample concentration and y is the logarithm of rat CD4 concentration.

2.8.3 T-tests

T-tests were used to compare aptamer binding to wild-type gp120 versus mutant gp120. A data table was created for the $\log EC_{50}$ values of aptamer binding from at least two independent experiments. The T-test function was used to perform a two-tailed unpaired T-test and P values of less than 0.05 were considered significant.

2.9 Computational methods

2.9.1 Modelling aptamer structures with mfold

Aptamer structures were modelled using mfold via the web server at <http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi> (Zuker 2003). Two-dimensional structures were then drawn using RNAviz version 2.0. Constrained folding was performed by stipulating where base-pairs must and must not occur using the available options.

2.9.2 RSmatch implementation

Experiments using RSmatch (Liu *et al* 2005) were carried out on a unix system hosted at orac.molbiol.ox.ac.uk. The sequences of the 27 aptamer sequences from (Khati *et al* 2003) plus the 34 novel sequences (see Section 4.2) were used to create aptamer sequences libraries in fasta format. The full-length sequence library was `all_aptamers.fasta` and the variable only library was `all_aptamers_variable.fasta`. Additionally, individual sequence files of the full-length and variable sequence of all aptamers were created in fasta format, for example `B40.fasta` and `B40_variable.fasta`, and these were used as the query sequences.

RSmatch requires the input of various options and parameters. A typical unix instruction appears as:

```
rsmatch -p dsearch -D all_aptamers.fasta -Q B40.fasta -o
B40_full.result.out -r 1 -W 120
```

1. Type of search (-p)

This represents the type of query that will be used. The options are a general database search (`dsearch`), an iterative database search (`isearch`), a multiple RNA structural alignment (`mrna`) or a pairwise RNA structural alignment

(prsa). Here I used dsearch that involves finding patterns that match locally or globally to a given query structure. mrsa is used when a small number of RNA structures are known to contain a shared motif. isearch updates the scoring matrix with the latest result (I took a similar approach to this but instead performed the iterations by eye). prsa is used to compare just two RNA sequences.

2. Input sequence library (-D or -d)

The input sequence library in fasta format (-D) or structure format (-d).

3. Gap penalty (-g)

The penalty for mismatches in the alignment. The default value of -6 was used.

4. Output (-o)

Defines the location of the output file, which takes the suffix .result.out.

5. Range of folding energy (-r).

The optimisation of this parameter is discussed in Chapter 4 (Section 4.4.2).

6. Window length (-W)

W is the length of sequence windows (default length 100 ntd) within a long sequence that are folded independently by RSmatch. In this study W was set to 120 so that each aptamer consisted of one window.

7. Sliding step length (-S)

This is the value, expressed as a ratio of W, at which each new sliding window starts. The default value of 0.5 was used as this was an irrelevant parameter, given that each aptamer sequence consists of only one window.

8. Global alignment (-G)

A global (T) or local (F) alignment can be used, and in this study a local alignment was chosen as I was interested in looking at small loops.

2.9.3 ClustalW alignment and phylogenetic trees

RNA sequence alignments were performed with ClustalW (Thompson *et al* 1994) through the web server by Des Higgins at <http://cbi.labri.u-bordeaux.fr/outils/Pise/clustalw.html>. This interface was chosen because it produces an output file that can be used to draw phylogenetic trees that include bootstrap values. The sequences corresponding to aptamer loops 1 and 2 (see Section 4.1) were assembled into fasta files and were submitted using the default alignment parameters. The resulting alignment (.aln) file was opened with BioEdit and the phylogenetic tree (.dnd) file was used to calculate a phylogenetic tree displaying bootstrap values using the same web interface. Phylogenetic trees were drawn using Phylodendron via the CBRG website. Random sequences were generated using the Bioinfo server at <http://users.ugent.be/~mdgroeve/bioinfo/>.

2.9.4 Sfold implementation

Aptamer structures were analysed using Sfold through the Sfold web server (Ding *et al* 2004) at <http://sfold.wadsworth.org/index.pl> using default parameters.

Invaluable help with computational methods was received from Simon McGowan from the Computational Biology Research Group (CBRG) in SWDSOP, Oxford who installed RSmatch on the unix system and wrote Perl scripts to automatically submit query sequences to RSmatch as well as giving advice on ClustalW and phylogenetics.

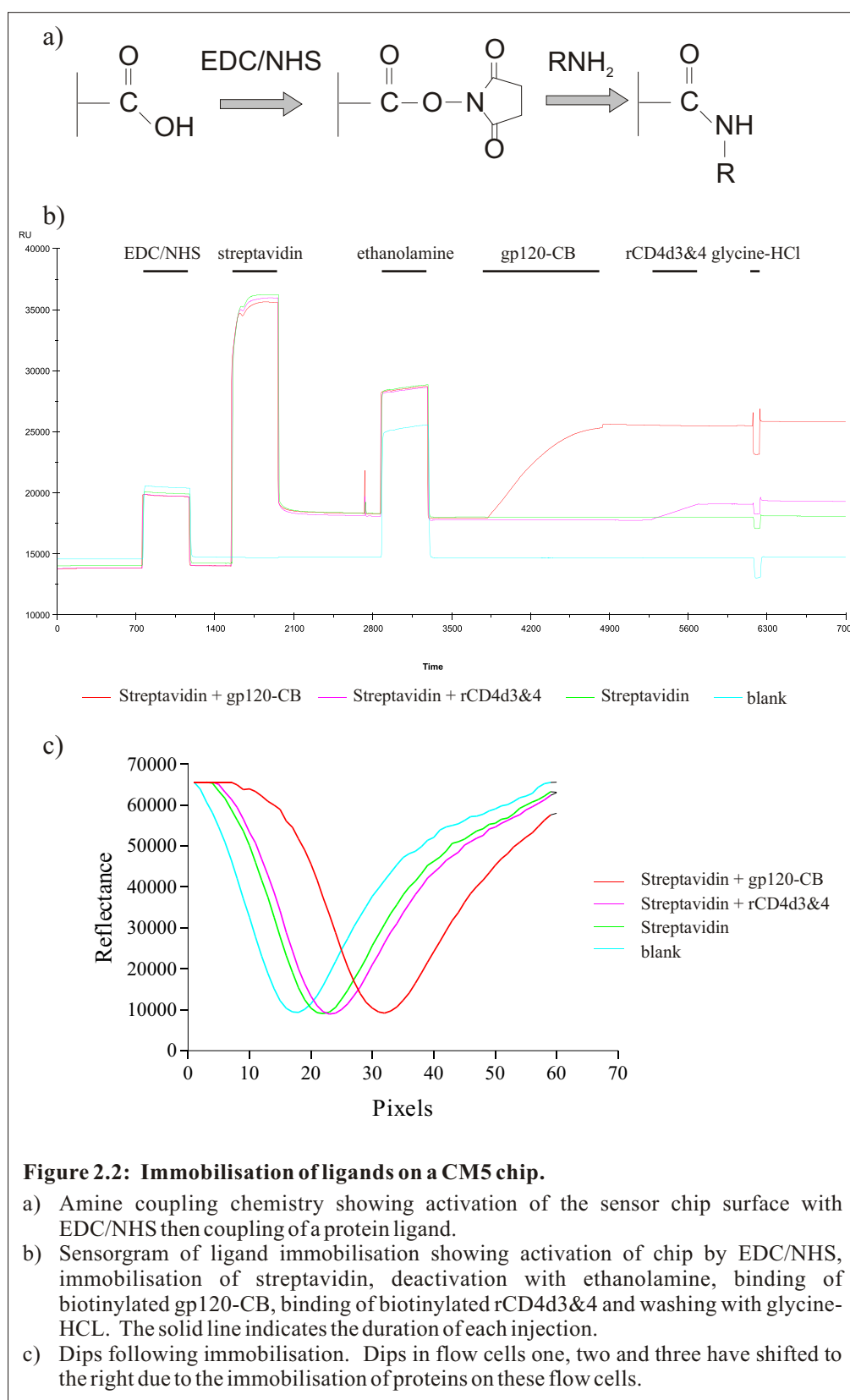
2.10 Surface plasmon resonance

Protein-protein and protein-RNA interactions were analysed using a BIAcore 2000 instrument and data were analysed using BIAevaluation software version 3.2. The running buffer used in all experiments was $1 \times$ CHBS buffer (see Section 2.1.5) which had been filtered, then degassed for 30 minutes using a vacuum. In order to avoid the introduction of air bubbles into the system, all samples were centrifuged at 12 000 rpm in a MSE Micro Centaur centrifuge for 30-60 seconds before injection. The flow rate was 5 μ L/min unless otherwise stated and experiments were performed at 25 °C.

2.10.1 Immobilisation

Proteins were immobilised on a BIAcore CM5 sensor chip using amine coupling chemistry (see Figure 2.2a). Equal volumes of N-ethyl-N'-(3-dimethylaminopropyl) carbodiimide (EDC) and N-hydroxysuccinimide (NHS, both BIAcore) were mixed and 35 μ L was injected over two or four flow cells to activate the sensor surface. The protein to be immobilised was diluted to between 50 and 100 μ g/mL in a solution of 10 mM sodium acetate at a pH in the range 4-6 (the optimal pH for immobilisation of each individual protein was determined experimentally). 35 μ L of this protein was injected over the appropriate flow cells. The sensor surface was then deactivated by an injection of 35 μ L 1 M ethanolamine-HCl pH 8.5 (BIAcore) over all flow cells. To remove any residual unbound protein, 5 μ L glycine-HCl pH 2.5 (BIAcore) was injected over all the flow cells.

To immobilise a biotinylated protein, streptavidin was first immobilised by the amine coupling method described above, using streptavidin at 100 μ g/mL in 10 mM sodium acetate pH5. After deactivation, a suitable volume of the biotinylated protein was injected over the required flow cell, then the chip was washed with glycine-HCl pH 2.5. Finally all streptavidin sites were blocked



with a 10 μ L injection of 160 μ g/mL biotin in $1 \times$ CHBS. Figure 2.2b shows a sensorgram of a typical immobilisation protocol in which the biotinylated gp120-CB fusion protein and a control of biotinylated rCD4d3&4 are immobilised via streptavidin on flow cells one and two, respectively, of a CM5 chip. In such instances streptavidin alone was immobilised on flow cell three as a control for non-specific binding to streptavidin and flow cell four was left blank.

In order to check the flow cell surface after immobilisation, dips were analysed. When a ligand is immobilised on the flow cell surface, the dips are shifted to the right in proportion with the amount of ligand immobilised. The dips in the flow cell should be all of the same depth with a reflection minimum at approximately 10,000 reflectance units. Figure 2.2c shows typical dips following the standard immobilisation protocol described above.

2.10.2 Binding analysis

Once the ligand(s) had been immobilised on the sensor chip surface, various analytes could be passed over the sensor chip surface to see if they bound to the ligand. Samples were diluted to an appropriate concentration, usually 0.1-1 μ M, in $1 \times$ CHBS buffer and 10-35 μ L were injected over all flow cells used in the experiment.

In SPR the response generated when an analyte binds to a ligand is dependent upon a number of variables; the affinity of the analyte for the ligand, the amount of ligand immobilised, the molecular mass of the analyte and the concentration of the analyte injected. Therefore binding responses were normalised for these variables so that different analytes could be compared.

BIAevaluation software was used to find the response at maximum binding or at a given time point after the start of injection. The baseline was normalised by subtracting the response generated in the blank flow cell. This was

necessary because injection of a sample in a buffer other than $1 \times$ CHBS, for example when tissue culture supernatant was injected, results in a bulk effect response in all flow cells.

2.10.3 Chip regeneration

After binding of an analyte it was possible (in most cases) to regenerate the sensor chip surface by injection of either glycine-HCl at low pH, or sodium hydroxide at high pH. The change in pH was sufficient to disrupt bonds between the ligand and analyte and remove the analyte from the flow cell. Typically 10 μ L of regeneration agent was injected, after which the flow rate was increased to 100 μ L/min for 10 minutes to wash the flow cells with buffer and ensure removal of the regenerating agent. All regeneration techniques were tested to ensure that they did not denature the protein immobilised on the sensor chip surface, and that subsequent injections of analyte were not affected. This method was used successfully to analyse binding of several analytes to one ligand in a single experiment, for example binding of different aptamers to immobilised gp120.

Chapter 3: Gp120 expression

- Expression of gp120 fusion proteins from pEF-BOS vectors
- Making a stable cell line expressing gp120-CB
- Codon optimisation
- Optimisation of transfection method
- Comparison of expression methods

Chapter 3: Gp120 expression

3.1 Introduction

My first aim was to develop an expression system for HIV-1_{BaL} gp120. Previously in the James lab, gp120 was expressed as a FLAG-tagged protein from a baculovirus vector in Sf9 insect cells (Lin *et al* 2000). However this expression system proved time-consuming, expensive and gave low yields. Also, because Sf9 cells have somewhat limited post-translational glycosylation machinery compared to mammalian cells (Yeh *et al* 1993), the resulting gp120 was not fully glycosylated giving it a reduced molecular mass of 100 kDa.

I chose to develop a mammalian expression system so that the recombinant protein would be fully glycosylated and would be secreted into the tissue culture supernatant (TCS). The TCS containing gp120 could be used directly in biochemical assays or the gp120 could be purified by affinity chromatography using the fusion domain, either domains three and four of rat CD4 or a His tag, if required.

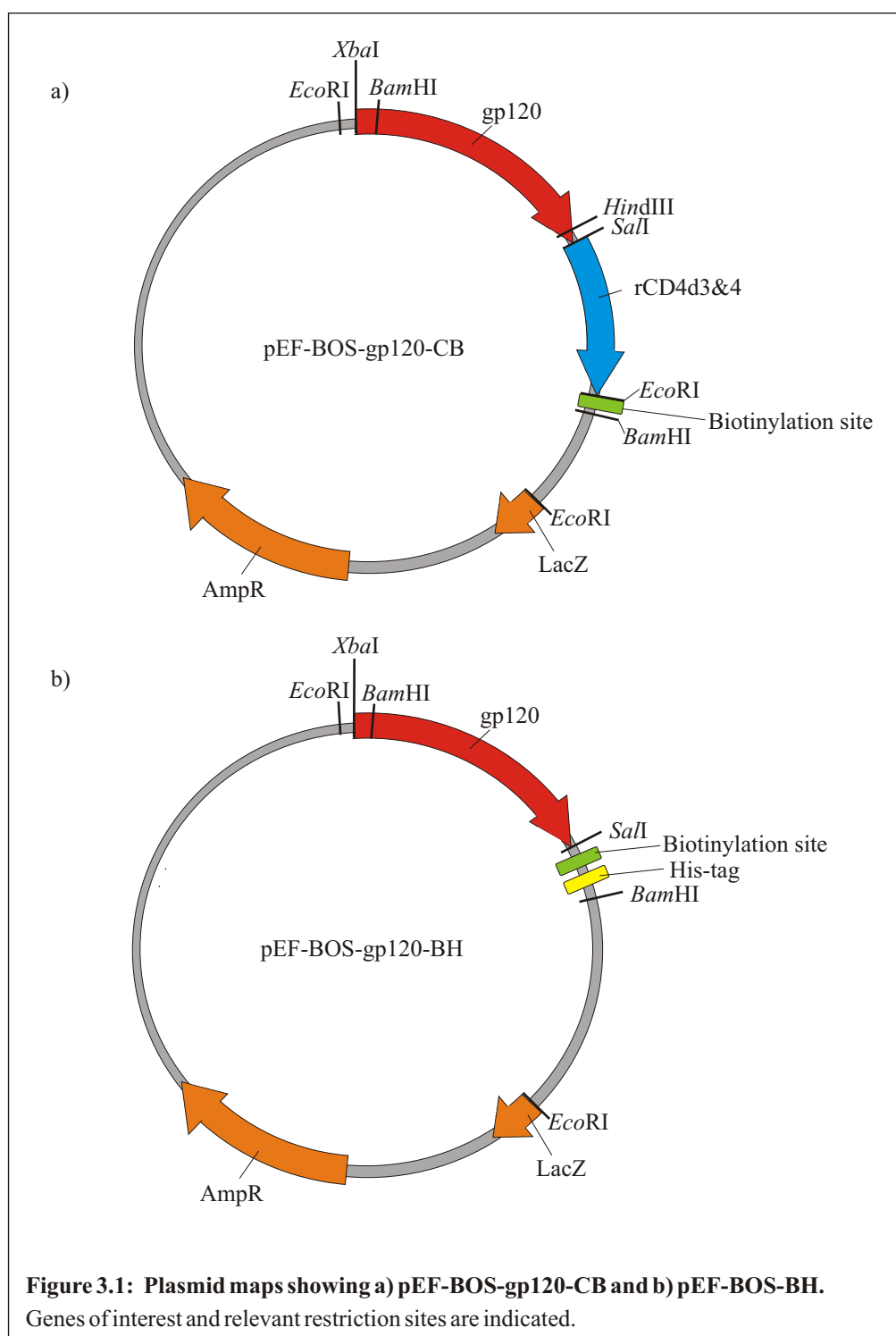
This chapter describes the expression of gp120 as a fusion protein with domains three and four of rat CD4 in 293T cells and CHO-K1 cells. I optimised the expression using a time-course assay, investigated the use of alternative transfection reagents and found that codon-optimisation of the gp120 gene increased expression. I compared the binding of this protein to several antibodies, human CD4 (hCD4), aptamer and an N-terminal peptide of CCR5 with that of monomeric gp120 in order to verify the conformation of the gp120 in the fusion protein. I was able to produce sufficient quantities of gp120-CB to use in the experiments in the rest of the thesis.

3.2 Expression of gp120 fusion proteins from pEF-BOS vectors

Two variants of the pEF-BOS mammalian expression vector (Mizushima and Nagata 1990) described below are routinely used by the Barclay group in Oxford for the expression of fusion proteins. The pEF-BOS vector was derived from pUC119 with the addition of the powerful human EF-1 α promoter, the SV40 origin of replication and a polyadenylation signal. pEF-BOS vectors can be introduced into mammalian cells such as HEK 293T cells by transient transfection, and the proteins are harvested from the cell supernatant.

The pEF-BOS-rCD4L-rCD4d3&4-biotinylation site (pEF-BOS-CB) vector includes the rat CD4 leader sequence flanked by *Xba*I and *Sal*I restriction sites, immunoglobulin domains three and four of rat CD4 (rCD4d3&4) flanked by *Sal*I and *Eco*RI restriction sites and a biotinylation site plus stop codon flanked by *Eco*RI and *Bam*HI restriction sites (Brown *et al* 1998). The biotinylation site encodes the amino acid sequence NSGSLHHILDAQKMVWNHR, which is a target for the *Escherichia coli* biotin holoenzyme synthetase, BirA (Schatz 1993). I cloned the gp120 leader sequence and gp120 gene into the vector to replace the rat CD4 leader sequence to produce a fusion construct of gp120-rCD4d3&4-biotinylation site. This construct will be referred to as pEF-BOS-gp120-CB and the expressed protein as gp120-CB. A plasmid map of pEF-BOS-gp120-CB is shown in Figure 3.1a. Since neither the gp120 nor the CD4 domains contain a transmembrane domain, it was expected that the protein would be secreted.

The pEF-BOS-rCD4L-rCD4d3&4-biotinylation site-His tag (pEF-BOS-BH) vector was derived from pEF-BOS-CB by removal of the stop codon and



addition of a His-tag (peptide sequence TLARSTHHHHHH) plus stop codon after the biotinylation site (Brown *et al* 1998). I cloned the gp120 leader sequence and gp120 gene into the vector to replace the rat CD4 leader sequence to produce a fusion construct of gp120-biotinylation site-His tag. This plasmid will be referred to as pEF-BOS-gp120-BH and the expressed protein as gp120-BH. A plasmid map of pEF-BOS-gp120-CB is shown in Figure 3.1b. Since neither the gp120 nor the His tag contain a transmembrane domain, it was expected that the protein would be secreted.

3.2.1 Cloning the gp120 gene into pEF-BOS expression vectors

Previously in the James lab, gp120 was expressed with a FLAG tag from a recombinant baculovirus vector in Sf9 cells (Lin *et al* 2000). Baculovirus DNA was extracted by SDS precipitation so that the gp120 gene could be used for cloning. 1 % SDS was added to the baculovirus supernatant and heated to 65 °C for 10 min, following which a phenol:chloroform extraction (see Section 2.2.6) and ethanol precipitation (see Section 2.2.7) were performed and the DNA pellet was resuspended in 1/10th the original volume in water. 0.25 µL of this DNA template (approximately 75 ng) was amplified by PCR (see Section 2.2.1) with the primers Fp2Bac and Rp2Bac, using Easy-A polymerase (Stratagene), which is a combination of *Taq* and *Pfu* DNA polymerases to give both high efficiency amplification of long PCR products as well as proof-reading activity to avoid the introduction of mutations. A gradient PCR was performed to optimise the annealing temperature (Figure 3.2a). The 1536 bp PCR product was formed in most reaction conditions so the PCR products were pooled and purified from an agarose gel (Figure 3.2b, see Section 2.2.3).

Approximately 80 ng of this gp120 PCR product was used as a template for a new PCR. In this reaction, primers were designed that would insert *Xba*I and *Sal*I restriction sites at the 5' and 3' ends of the gene respectively, remove the

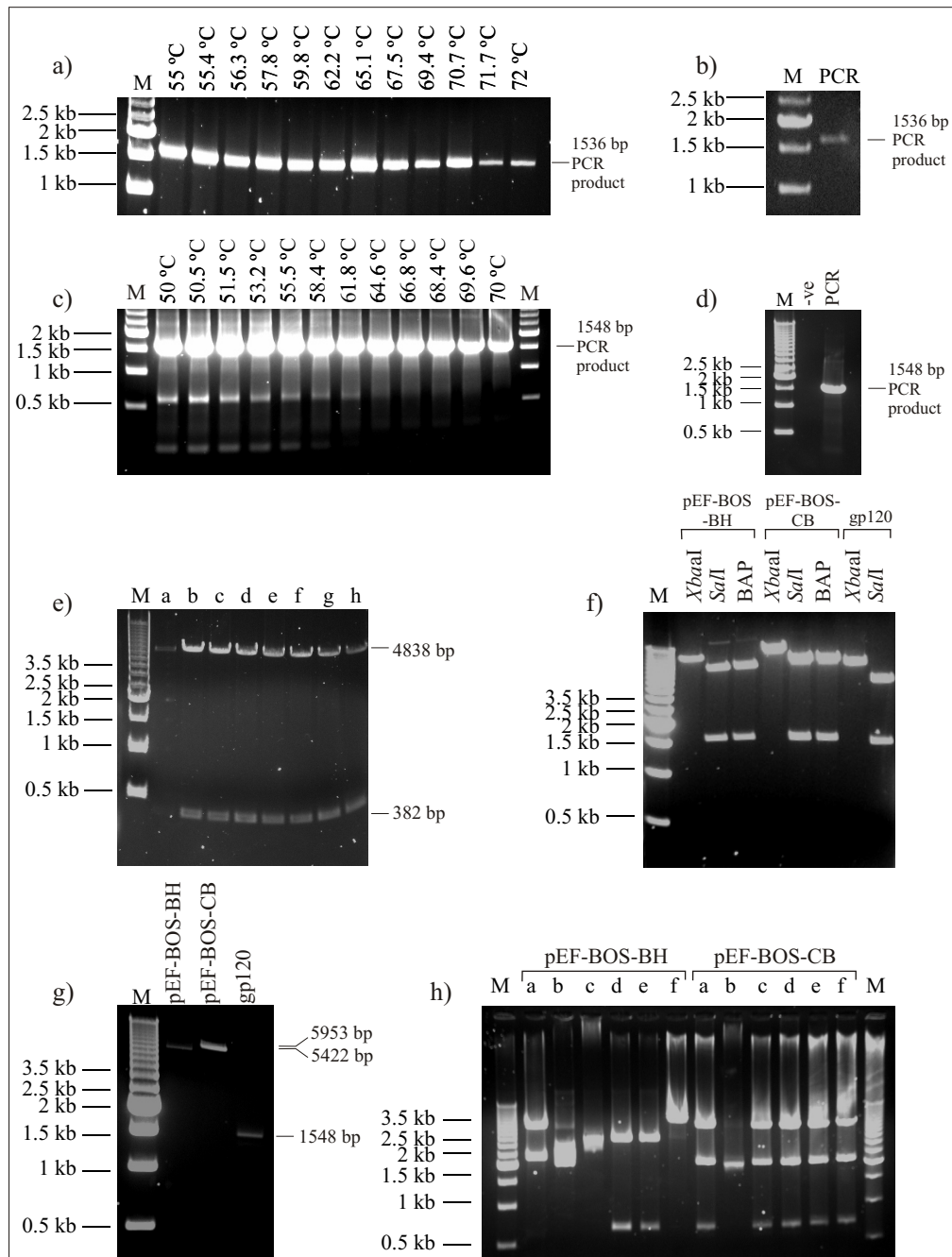


Figure 3.2: 1 % agarose gels stained with ethidium bromide showing DNA products involved in cloning the gp120 gene into pEF-BOS vectors.

- Gradient PCR amplification of the gp120 gene from baculovirus DNA.
- Gel-purified pooled PCR products from the 55 °C to 70.7 °C reactions in (a).
- Gradient PCR amplification of the gp120 gene to insert *Xba*I and *Sal*I restriction sites.
- PCR amplification as in (c) at the optimal annealing temperature of 65 °C.
- Nsi*I digestion products of the TOPO TA-gp120 plasmid from eight colonies.
- Products from treatment of pEF-BOS-alcam-BH, pEF-BOS-alcam-CB and TOPO TA-gp120 with *Xba*I, *Sal*I and bacterial alkaline phosphatase (BAP).
- Gel purification of the fragments produced in (f).
- Eco*RI digestion products of pEF-BOS-gp120-CB and pEF-BOS-gp120-BH from six colonies of each plasmid.

FLAG tag and add a linker of three glycine residues that would later join the gp120 to the rCD4d3&4. A gradient PCR was performed with the primers gp120 Xba fwd and 3' gp120 Sal using Easy-A polymerase (Figure 3.2c). The PCR was then repeated using the optimal annealing temperature of 65 °C (Figure 3.2d). The resulting ~1548 bp PCR product was immediately ligated into the TOPO TA cloning vector (Invitrogen) and then transformed into competent cells (see Sections 2.3.4 and 2.3.5). Colonies were picked and used to inoculate 5 mL cultures from which DNA was extracted) and the plasmid DNA from eight colonies was digested with *NsiI* restriction endonuclease and analysed by gel electrophoresis (Figure 3.2e). The presence of 4838 bp and 382 bp restriction digest products indicated that the gp120 gene was inserted into the TOPO TA vector, and this was confirmed by DNA sequencing (see Section 2.3.6). While some clones did contain mutations in the gp120 gene (due to polymerase errors in PCR amplification), clones B and H did not contain mutations, so clone B was chosen for further study.

The cloning step to insert the gp120 gene into the TOPO TA vector was included so that the gp120 gene with inserted *XbaI* and *SalI* restriction sites did not have to be digested as a linear DNA fragment. Instead, ligation of a fresh PCR product into the TOPO TA cloning vector allows digestion of the same DNA from a plasmid. This is a more favourable reaction than digestion of linear DNA and the digestion products can easily be seen by agarose gel electrophoresis.

N. Barclay kindly provided me with aliquots of the pEF-BOS-CB and pEF-BOS-BH plasmids each containing a ~1500 bp *alcam* gene insert (plasmids pEF-BOS-*alcam*-CB and pEF-BOS-*alcam*-BH respectively). These plasmids and the TOPO TA vector clone B containing the gp120 gene were digested with *XbaI* and *SalI* restriction endonucleases (Section 2.3.1) and the pEF-BOS vectors were treated with alkaline phosphatase (Section 2.3.2) to prevent self-ligation (Figure 3.2f). The 5953 bp and 5422 bp fragments, from

pEF-BOS-CB and pEF-BOS-BH vectors respectively, and the 1548 bp fragment of the gp120 gene were purified from an agarose gel (see Section 2.2.3, Figure 3.2g), the fragments were ligated (see Section 2.3.4) and transformed into competent cells, and plasmid DNA was extracted from six clones of each construct. The plasmids were digested with *EcoRI* restriction endonuclease and the digestion products were analysed by agarose gel electrophoresis (Figure 3.2h). The expected sizes of the digestion products were 4620 bp, 2103 bp and 776 bp, for the pEF-BOS-gp120-CB construct, and 4620 bp and 2348 bp for the pEF-BOS-gp120-BH construct. Digestion of pEF-BOS-gp120-CB clones a, c, d, e and f and pEF-BOS-gp120-BH clone a produced bands of the expected size and DNA sequencing was performed to check the sequence of the constructs (see Section 2.3.6). The protein sequences of gp120-CB and gp120-BH are given in Figure 3.3 and 3.4.

3.2.2 Transient transfection of 293T cells

293T cells were transiently transfected with pEF-BOS-gp120-CB and pEF-BOS-gp120-BH using the calcium phosphate method (see Section 2.4.4). The medium was changed 24 hours post transfection and TCS was harvested and cell lysates were prepared four days post transfection. A gp120 ELISA was performed (see Section 2.7.4) to analyse gp120 expression from the two constructs in the TCS and cell lysate (Figure 3.5). The gp120-BH fusion was only expressed at low levels in the cell lysate and supernatant (0.016 µg/mL and 0.041 µg/mL respectively). The gp120-CB fusion protein was expressed well and the majority was secreted in the supernatant (0.49 µg/mL in supernatant, 0.036 µg/mL in cell lysate). The gp120-CB fusion protein was therefore chosen for further study. In order to increase the protein concentration the TCS was concentrated using Centricon filter (see Section 2.5.2). Biotinylation with BirA was performed as described in Section 2.5.3.

A time-course assay was performed in order to optimise the time point at which the TCS should be harvested. 293T cells were transfected with the

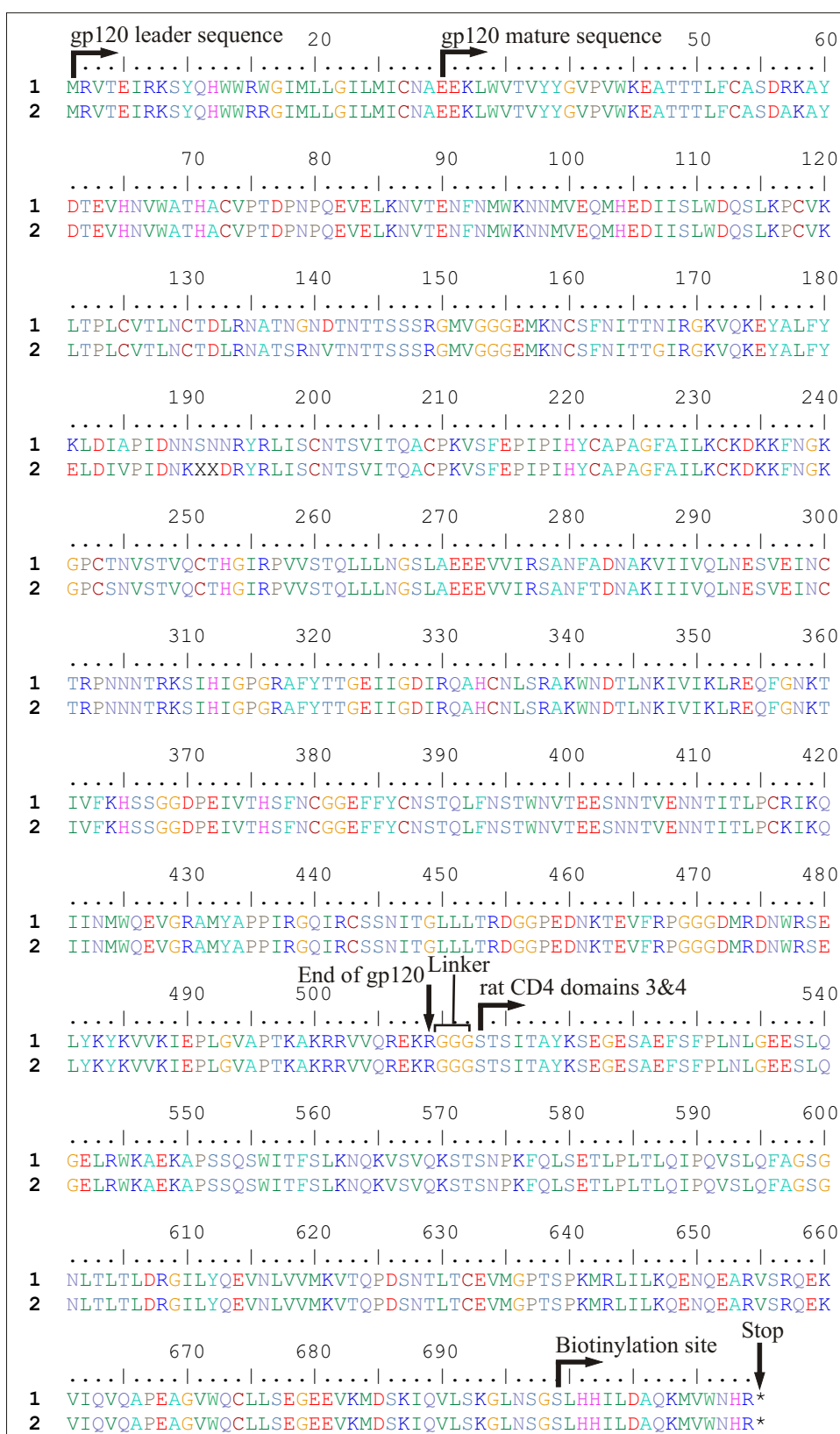
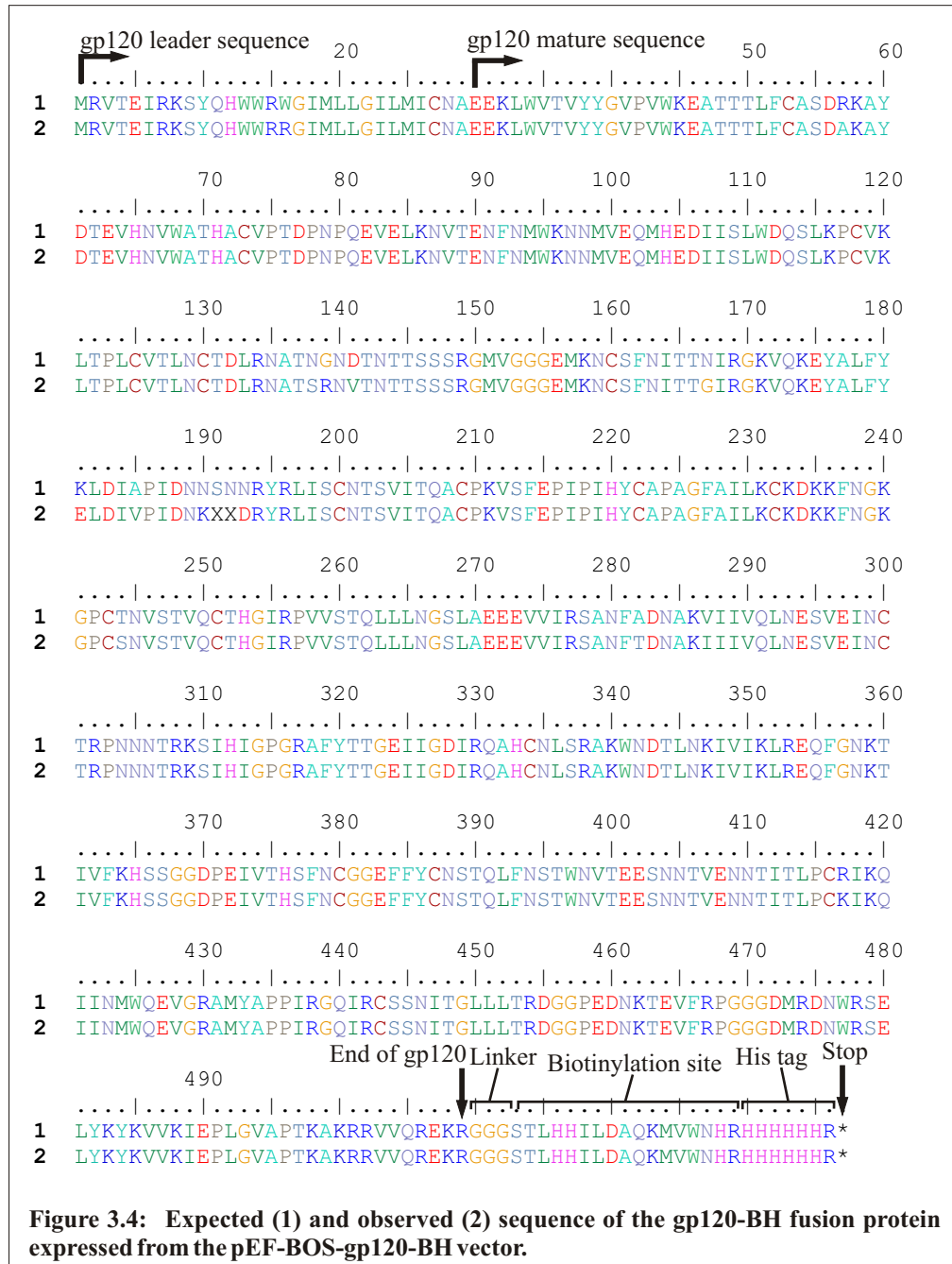
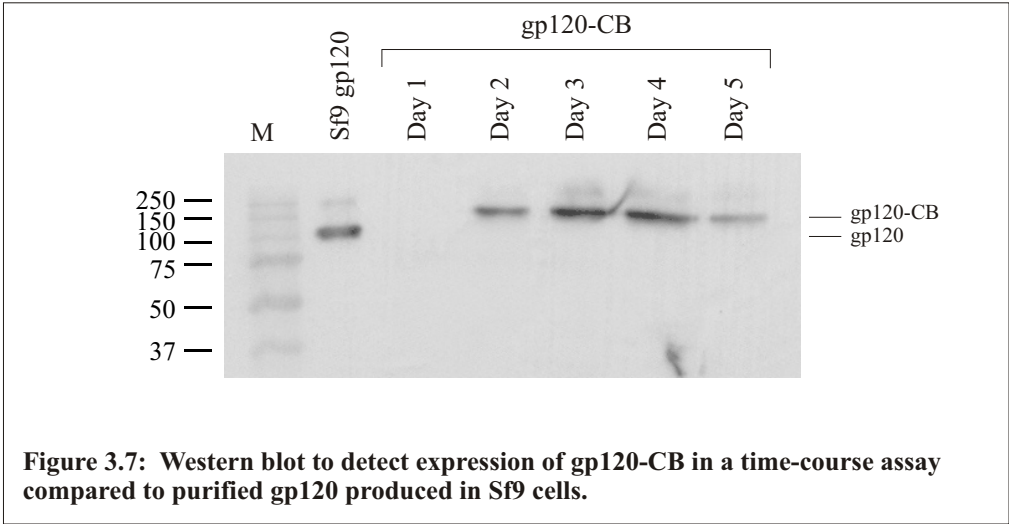
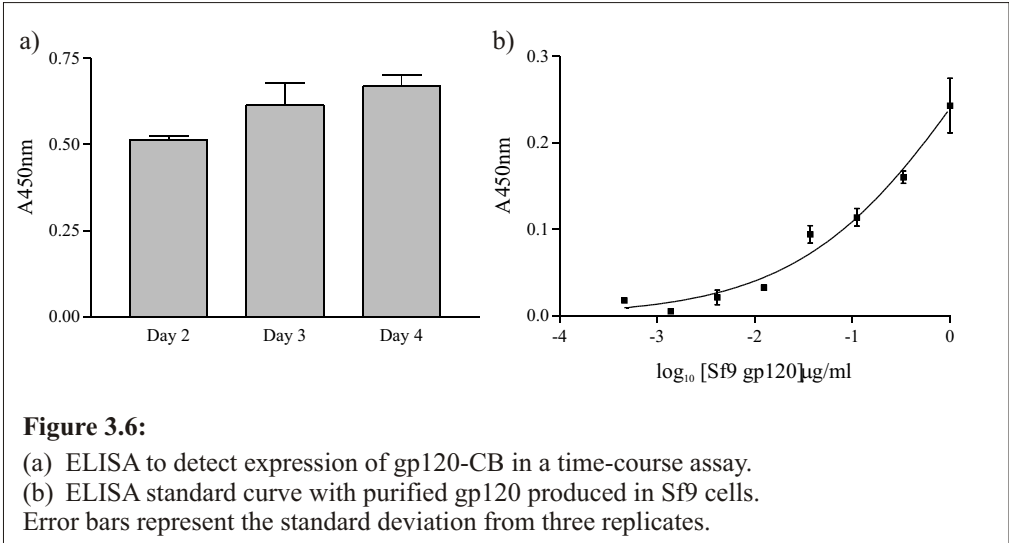
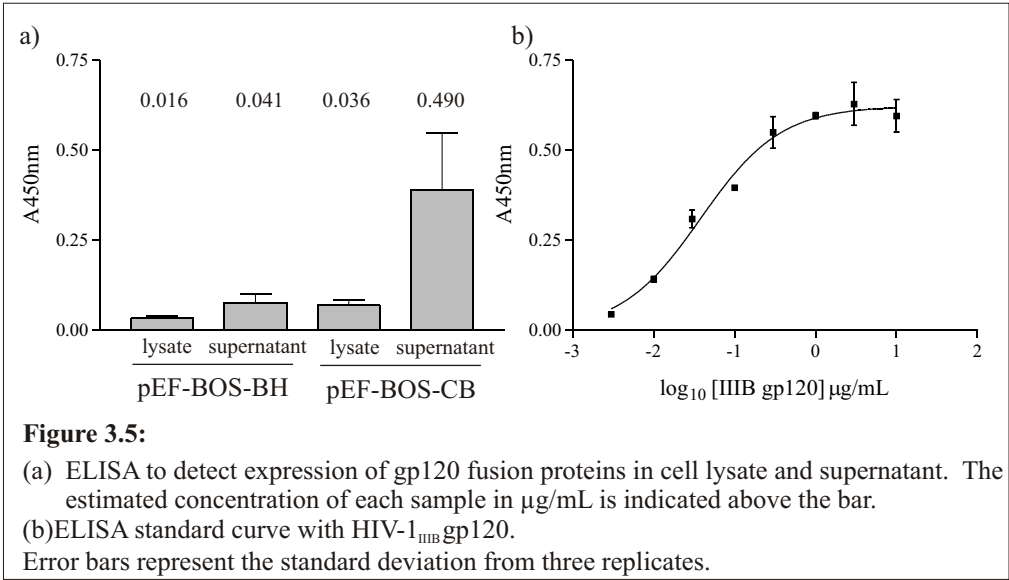


Figure 3.3: Expected (1) and observed (2) sequence of the gp120-CB fusion protein expressed from the pEF-BOS-gp120-CB vector.





pEF-BOS-gp120-CB plasmid using the calcium phosphate method (Section 2.4.4) the medium was changed 24 hours post transfection and then aliquots of TCS were taken two, three and four days post transfection. The expression of gp120-CB in the supernatant was analysed by ELISA (Section 2.7.4, Figure 3.6). Gp120-CB expression increased over the time period analysed and maximum expression was achieved at day four. It was not possible to estimate protein concentrations accurately using this ELISA because the standard curve was not performed with gp120 at high enough concentrations. However the concentration of gp120-CB at each time point was at least 1 µg/mL.

A second time-course assay was performed using the same method as above, but continued until day five post transfection. This time the samples were analysed by Western blot (see Section 2.7.2) and a sample of gp120 expressed in Sf9 cells was included as a control (Figure 3.7). The Western blot was stained with antibody 2G12 at 1 µg/mL and peroxidase-conjugated anti-human IgG at 1/1000 dilution. This analysis shows that peak gp120-CB expression occurs at day four post-transfection and that after this time point the protein begins to degrade. Therefore, further gp120-CB preparations were harvested at four days post-transfection. As expected, the apparent molecular mass of the gp120-CB fusion protein from 293T was 145 kDa (120 kDa gp120 plus 25 kDa rCD4d3&4). The apparent molecular mass of gp120 produced in Sf9 cells is 100 kDa, owing to reduced glycosylation of the gp120.

3.3 *Making a stable cell line expressing gp120-CB*

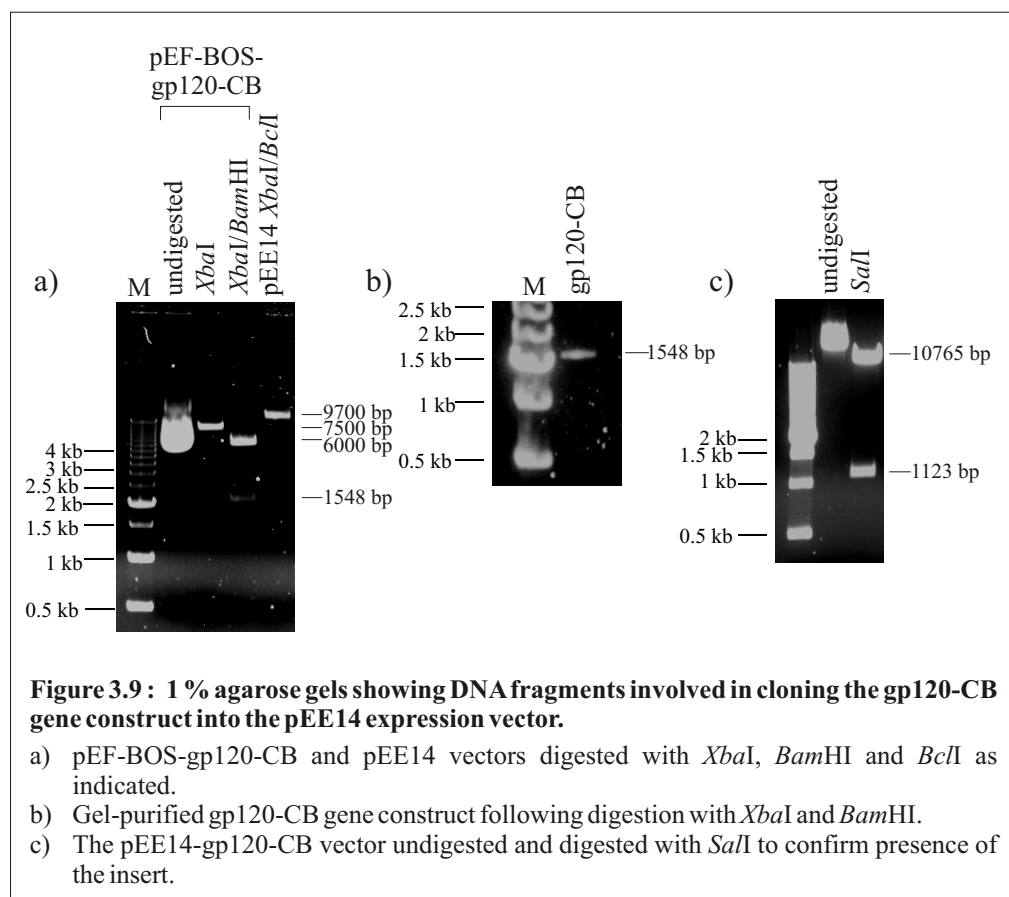
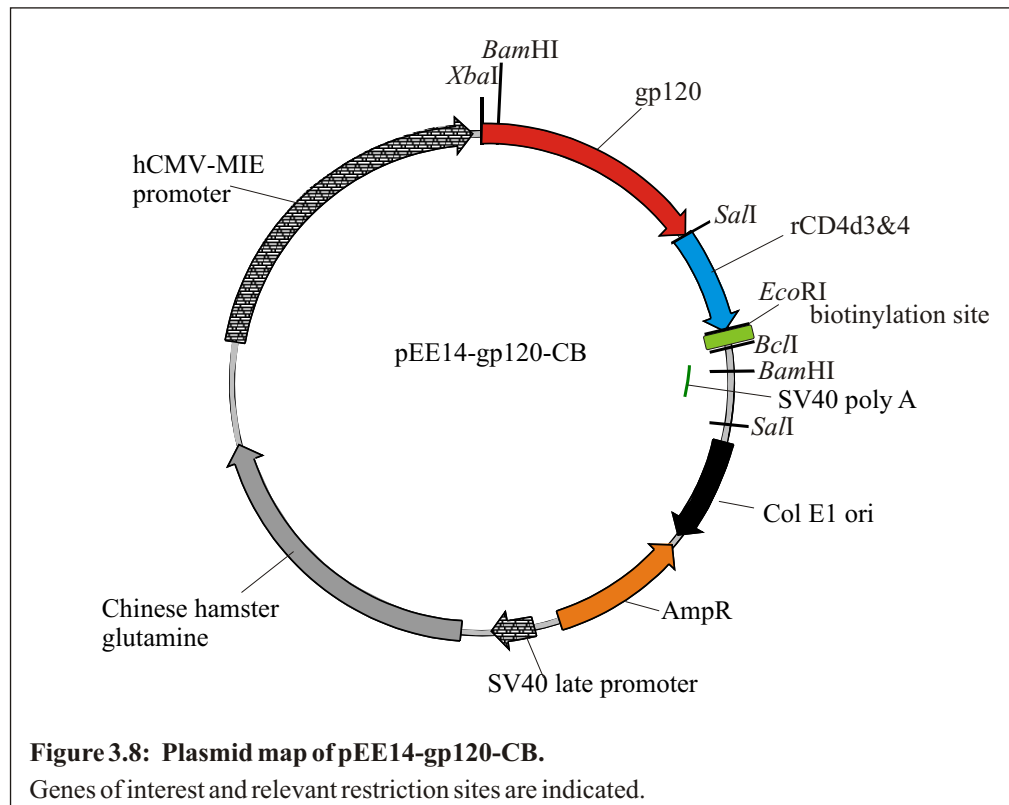
Following the success of the expression of gp120 as a rCD4d3&4-fusion protein in mammalian cells, I decided to make a stable cell line in order to

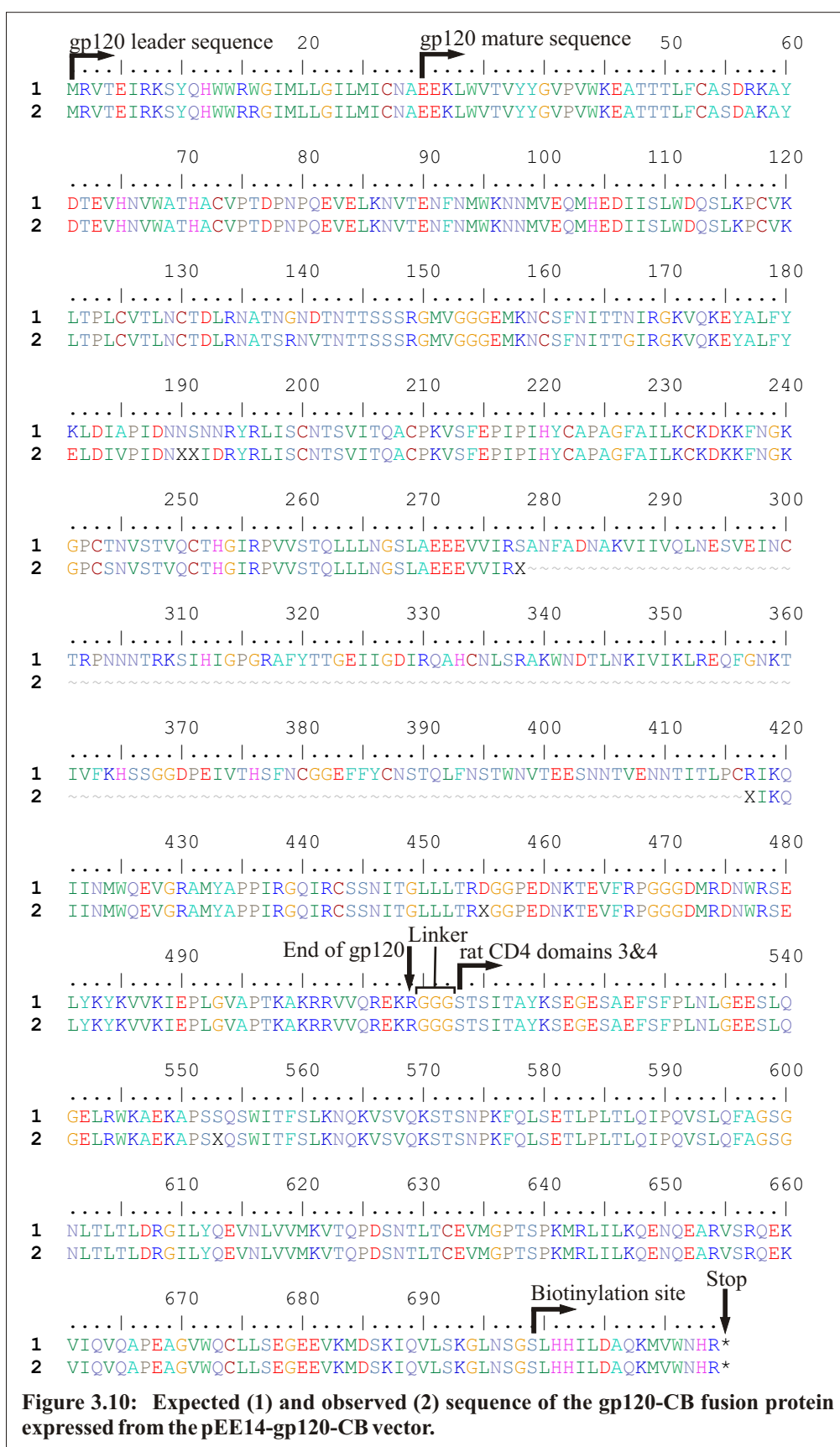
produce large quantities of gp120-CB that could be purified and used in biochemical assays.

The pEE14 cloning vector (Cockett *et al* 1990) is derived from pEE6 (Stephens and Cockett 1989). pEE14 contains the human cytomegalovirus (HCMV) promoter, an origin of replication from pBR328, a polyadenylation site and the glutamine synthetase minigene. The expression of glutamine synthetase allows cells transfected with pEE14 to grow in the presence of the inhibitor methioninesulfoxamine (MSX), allowing selection of clones that retain the plasmid. This vector has been used previously to express other proteins as fusion proteins with rCD4d3&4 in CHO-K1 cells (Brown and Barclay 1994). The gp120-CB construct was subcloned into pEE14 and used to transfect CHO-K1 cells. Positive clones were selected in the presence of MSX to obtain a stable cell line. A plasmid map of the pEE14-gp120-CB construct is shown in Figure 3.8.

3.3.1 Cloning the gp120-CB construct into pEE14

N. Barclay kindly provided the pEE14 vector that had already been digested with *Xba*I and *Bcl*II restriction endonucleases. The pEF-BOS-gp120-CB vector was digested with *Xba*I and *Bam*HI restriction endonucleases (Section 2.3.1, Figure 3.9a) as the *Bcl*II and *Bam*HI recognition sequences are compatible. Note that the restriction sites are lost upon ligation of the insert and plasmid because the two enzymes are not isoschizomers. The digested 1548 bp gp120-CB construct was gel purified (Figure 3.9b) and ligated into the digested pEE14 vector. The resulting plasmid was transformed into competent cells and plasmid DNA was purified from a 5 mL bacterial culture. The pEE14-gp120-CB plasmid was digested with *Sal*I which resulted in two bands of 1123 bp and 10 765 bp as expected (Figure 3.9c), and the presence of the gp120-CB construct was confirmed by DNA sequencing (Figure 3.10).



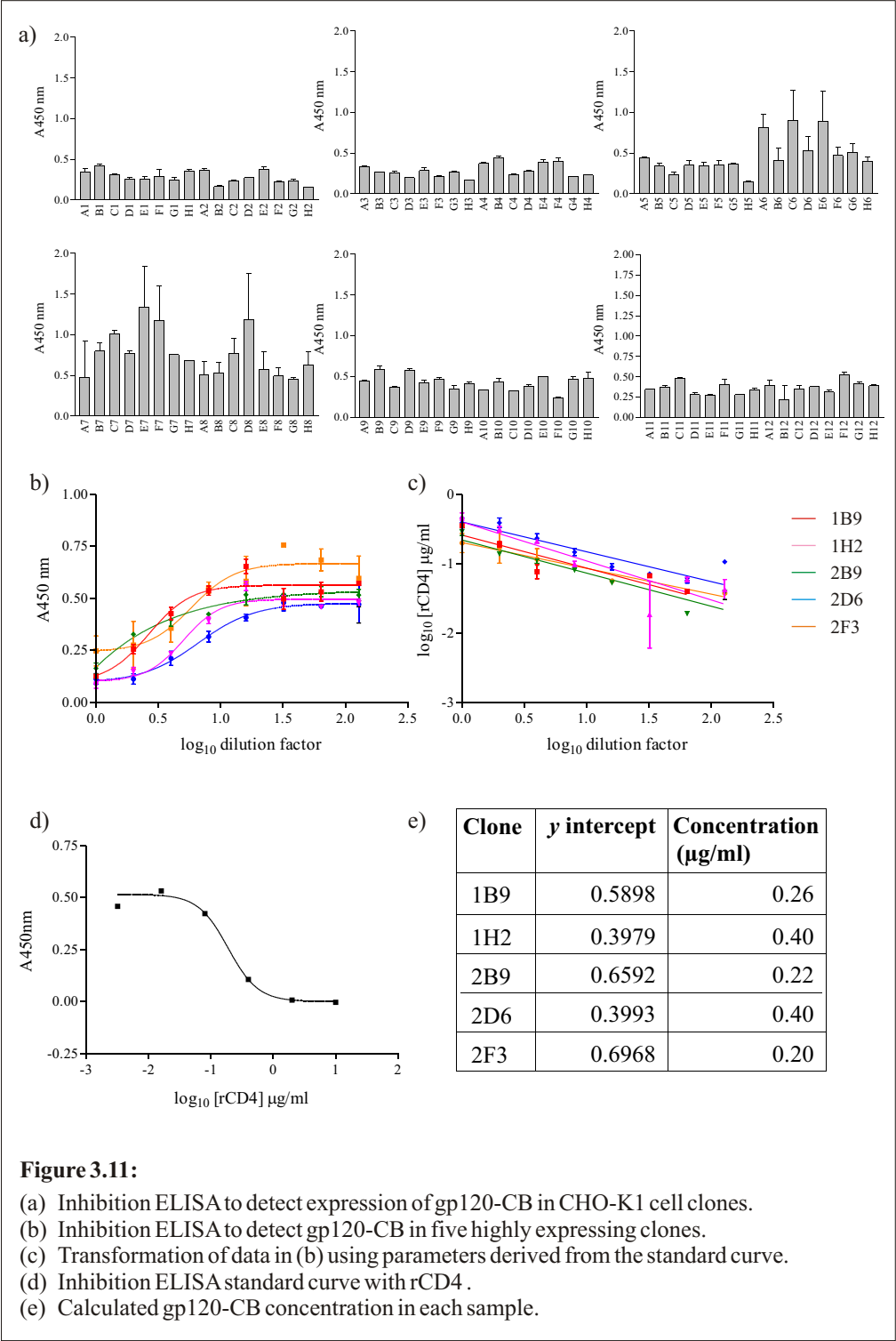


3.3.2 Selection of positive clones

The pEE14-gp120-CB plasmid was transfected into 10^6 CHO-K1 cells in a T75 flask using FuGENE 6 (Section 2.4.4). 24 hours post transfection the cells were trypsinised and redistributed in 96-well plates at 10^4 cells per well in the presence of 25 μ M or 30 μ M MSX. Cells that retained the plasmid, which expressed the glutamine synthetase minigene, were able to grow in the presence of the MSX inhibitor. After 14 days the TCS of each clone was assayed for gp120-CB expression by inhibition ELISA (see Section 2.7.5), representative examples of which are shown in Figure 3.11a. Five clones expressed high levels of gp120-CB and were expanded into T25 and then T75 tissue culture flasks. Tissue culture supernatant from these clones (1B9, 1H2, 2B9, DD6 and 2F3) was analysed by a more quantitative inhibition ELISA using a serial dilution of each sample. The results of this assay and the estimated protein concentrations from each clone are shown in Figure 3.11b-e. Clone 1H2 was chosen for further study as it expressed the highest levels of gp120-CB (0.4 μ g/mL). Clone 1H2 was expanded into a 1 L cell factory and the gp120-CB was purified from the TCS after two weeks by immunoaffinity chromatography using an OX68 column (see Section 2.7.9). Further analysis of this protein is described later in the Chapter.

3.4 Codon optimisation

It is well documented that the genetic code is degenerate and that organisms from different phylogenetic groups use different codons to encode their proteins. It has been shown that altering viral DNA codons to those used by highly-expressed genes from mammalian cells (codon optimisation) can increase expression levels of viral genes in mammalian systems (Nguyen *et al* 2004). The reason for this is two-fold: firstly, that the translational machinery in mammalian cells is better adapted to use mammalian codons and secondly, that the viral RNA may form secondary and tertiary structural interactions that



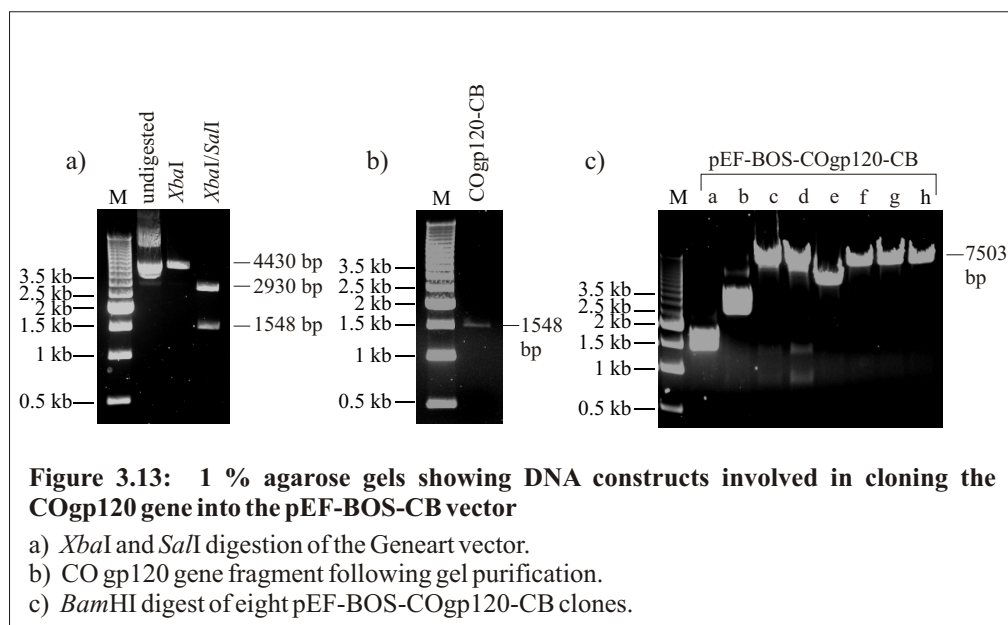
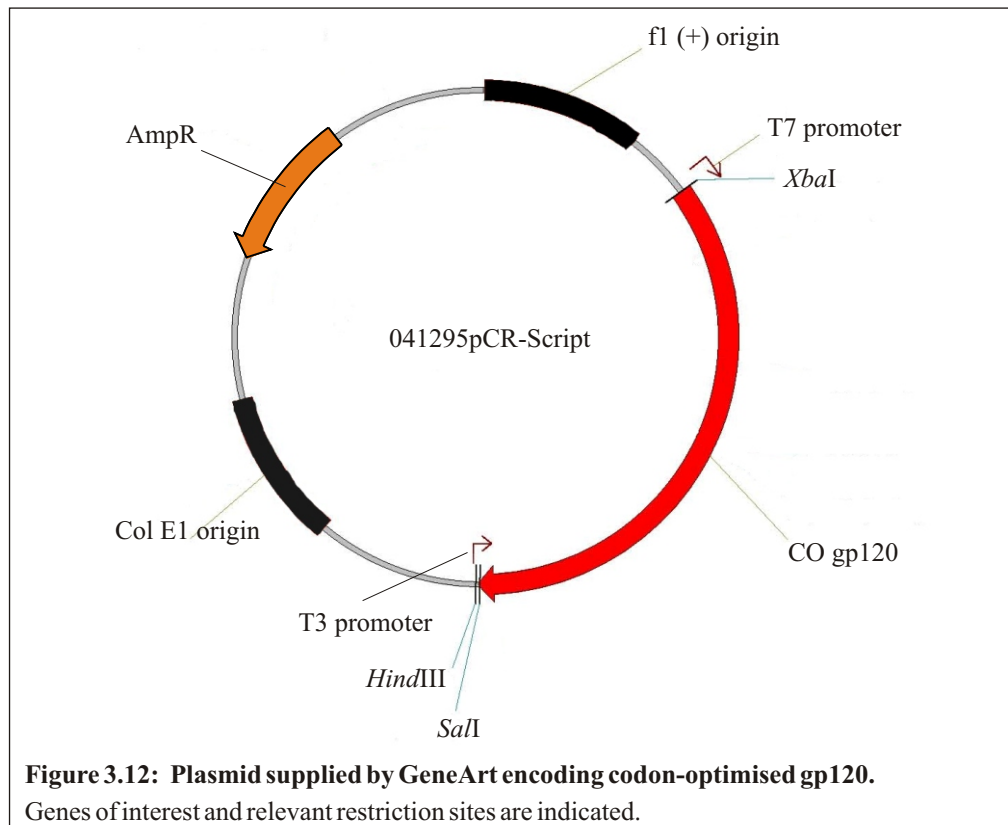
may impede the progression of mammalian enzymes. Haas *et al* showed that codon optimisation of the HIV-1 *env* gene from the MN strain dramatically increased gp120 expression from 293T cells (Haas *et al* 1996). Therefore I designed a codon-optimised gp120 gene from the HIV-1 strain BaL and this was synthesised in a vector by the company Genart (Figure 3.12). The codon-optimised gp120 gene (COgp120) was constructed with flanking *XbaI* and *SalI* sites, allowing it to be cloned into the pEF-BOS-CB expression vector and expressed in 293T cells.

3.4.1 Cloning of codon-optimised gp120 into pEF-BOS-CB

The Genart plasmid was digested with *XbaI* and *SalI* restriction endonucleases (see Section 2.3.1, Figure 3.13a) and the resulting 1548 bp fragment was purified from an agarose gel (Section 2.3.3, Figure 3.13b). This insert was ligated into pEF-BOS-CB that had been digested with the same restriction endonucleases, and the ligated construct was transformed into competent cells (Section 2.3.5). Plasmid DNA was purified from 5 mL cultures of eight clones and was analysed by *BamHI* digestion to look for the presence of the insert (Figure 3.13c). Digestion of five clones produced the single 7503 bp fragment expected, and the presence of the insert was confirmed by DNA sequencing (Figure 3.14).

3.4.2 Comparison of expression

293T cells were transiently transfected with pEF-BOS-gp120-CB or pEF-BOS-COgp120-CB using FuGENE (see Section 2.4.4). TCS was harvested after four days and was analysed by inhibition ELISA to determine the concentration of protein in the supernatant (Figure 3.15a-d). Codon optimisation of the gp120 gene resulted in a 3.5-fold increase in expression from 4.89 µg/mL to 16.9 µg/mL. The TCS was also analysed by Western blot (Figure 3.15e) to confirm the size of the protein expressed and the relative



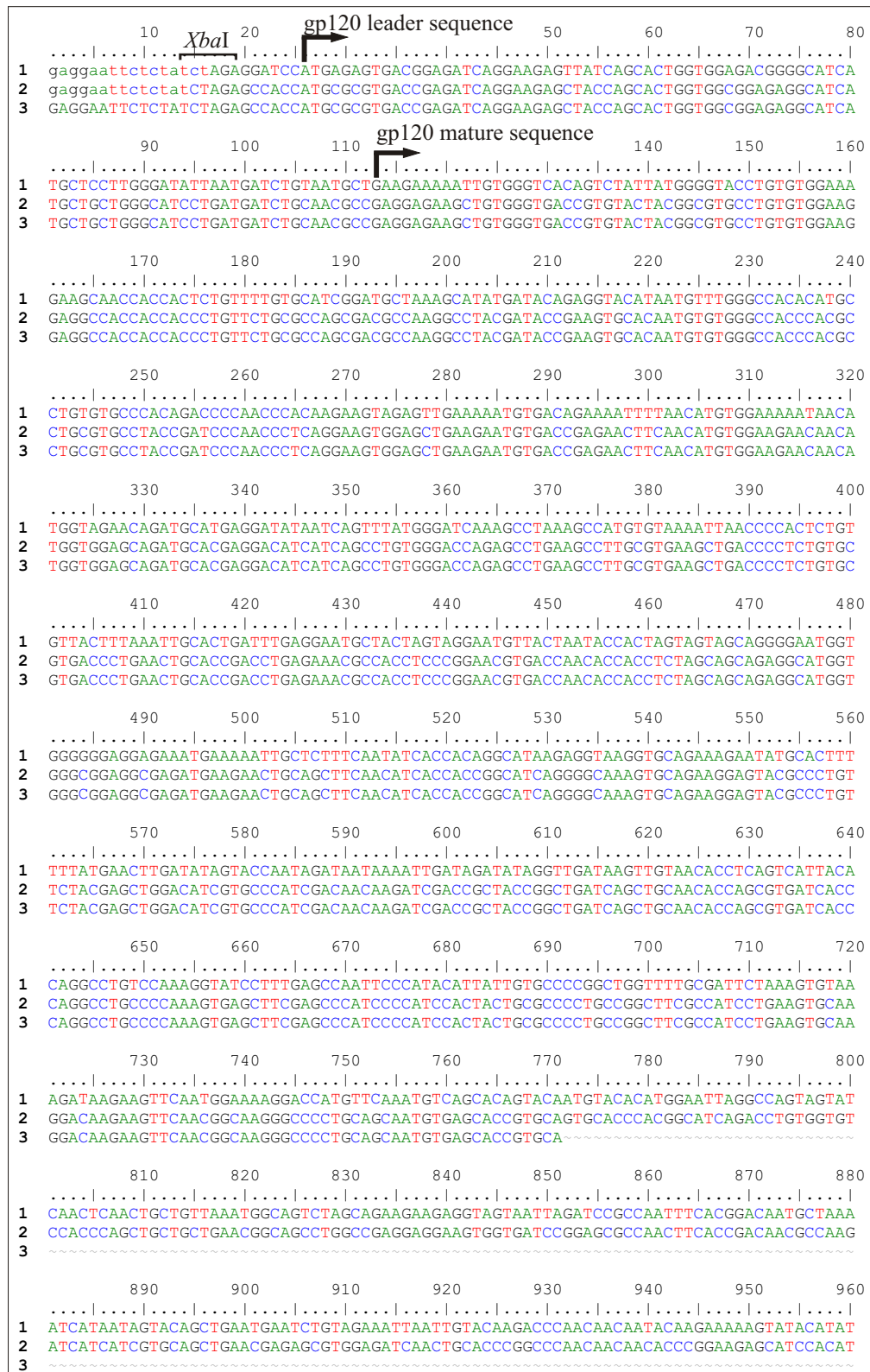
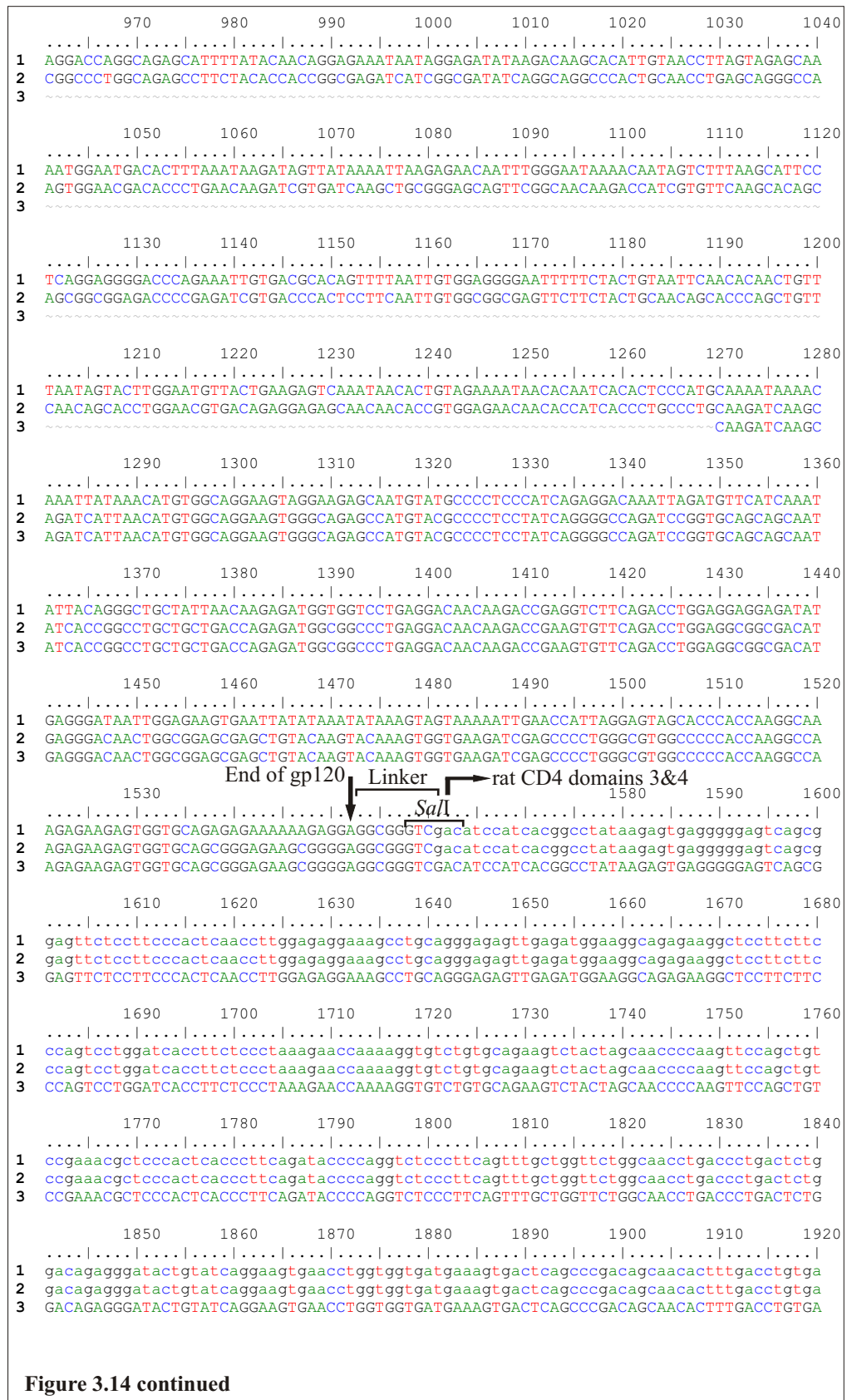
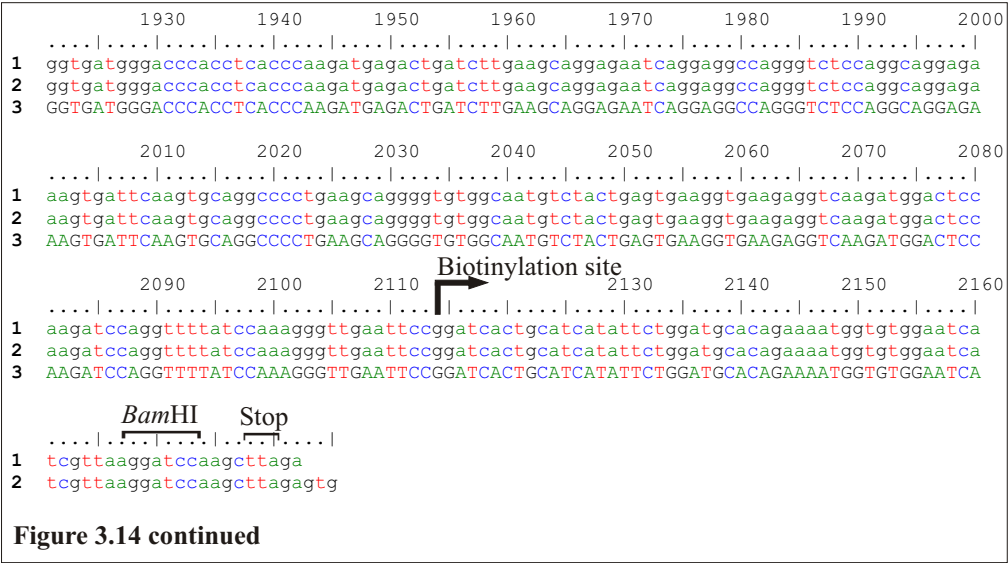
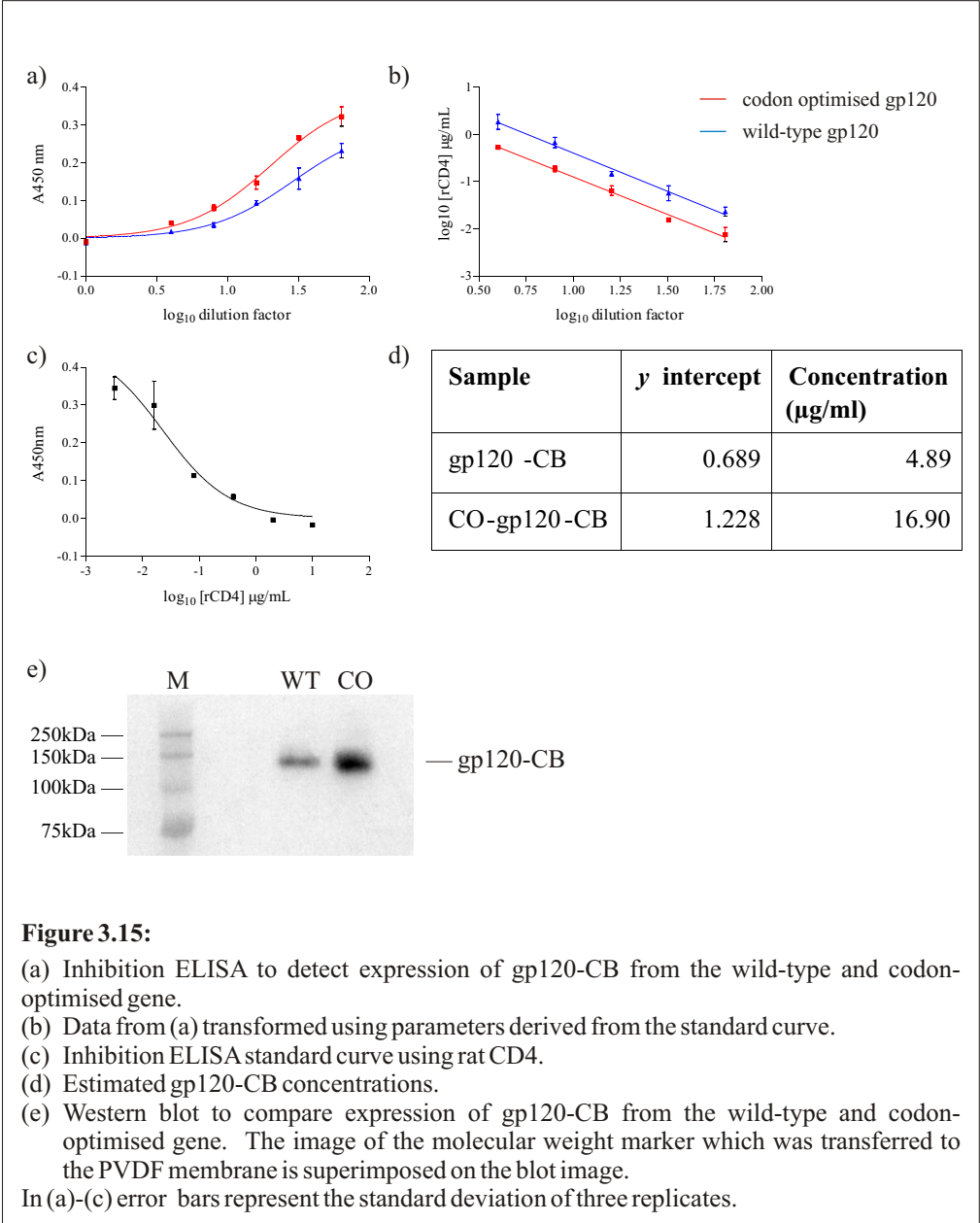


Figure 3.14: DNA sequence of the gp120-CB gene construct (1) compared to expected (2) and observed (3) DNA sequence of the codon-optimised gp120-CB gene construct.







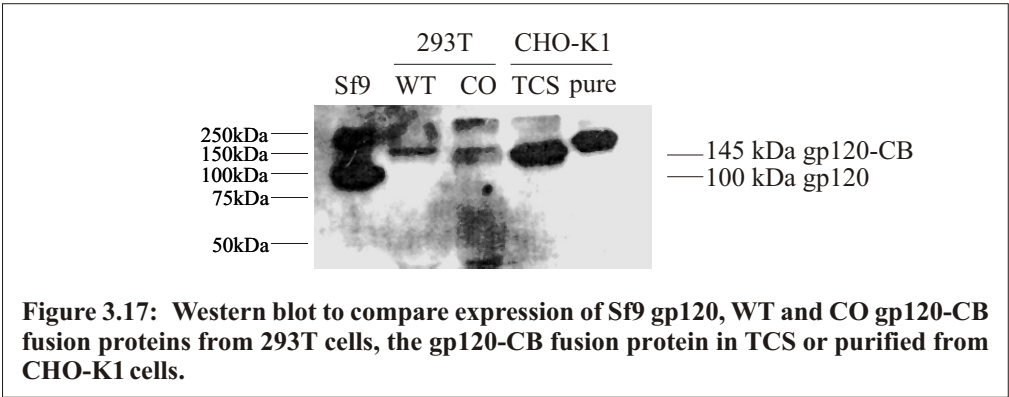
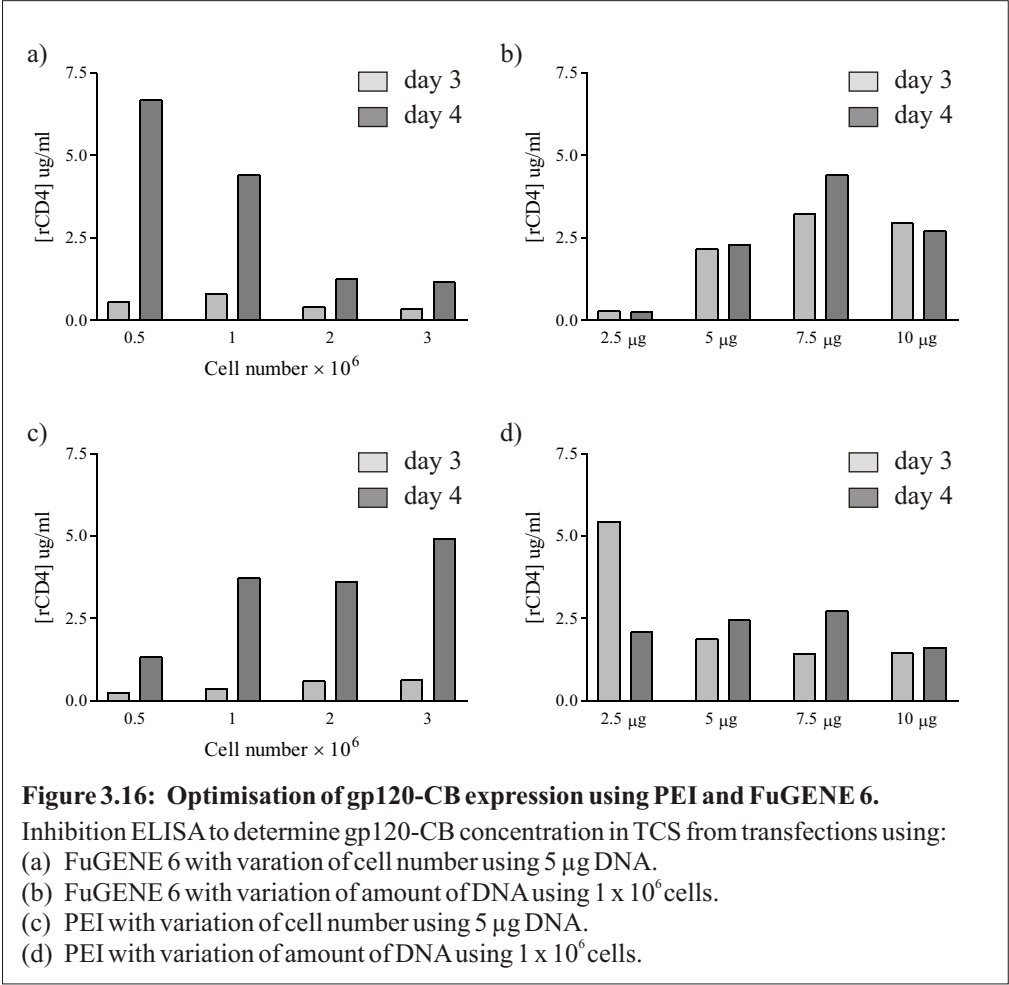
expression levels. Although the increase in expression due to codon optimisation was more modest than has been seen for other proteins (Haas *et al* 1996) the increase was significant enough that the construct encoding the codon-optimised gene was used for subsequent protein production.

3.5 Optimisation of transfection method

Transient transfection of 293T cells with plasmid DNA was traditionally performed by the calcium phosphate method (Gorman 1985) but more recently trends have been to use liposome-based methods of transfection, for example with FuGENE 6 (Roche). While transfection with FuGENE is very efficient, the reagent is very costly, so I took this opportunity to investigate the potential of poly(ethylenimine) (PEI) as a transfection reagent (Godbey *et al* 1999; Durocher *et al* 2002). PEI is a polyvalent cation that will intercalate with DNA and neutralise its charge, thus enabling its uptake by cells.

I optimised gp120-CB expression from the pEF-BOS-COgp120-CB plasmid using the FuGENE 6 and PEI transfection methods (see Section 2.4.4). 5 µg DNA was transfected into 10^6 cells in a T75 flask, and the ratio of DNA:FuGENE 6 and DNA:PEI were maintained at 1:3 and 1:5 respectively. The amount of DNA transfected and the number of cells per flask were then varied. The TCS was harvested and analysed by inhibition ELISA three and four days post transfection to determine the concentration of protein expressed in each set of conditions (Figure 3.16). In almost all cases the expression level was greater when the protein was harvested four days post-transfection.

When FuGENE was used as the transfection reagent and 5 µg DNA were transfected, increasing the number of cells decreased gp120-CB expression. When a constant number of cells were used, increasing the amount of plasmid



DNA transfected increased protein expression up to 7.5 μg . Transfection of larger amounts of DNA may be toxic to the cells so the levels of protein expression were lower when 10 μg DNA were transfected.

When PEI was used as the transfection reagent and 5 μg DNA were transfected, increasing the number of cells increased the amount of protein expression. When a constant number of cells were used, increasing the amount of plasmid transfected did not seem to have a very great effect, suggesting that the plasmid was in excess.

Maximum protein expression was achieved when 5×10^5 cells were transfected with 5 μg plasmid using FuGENE 6 (6.5 $\mu\text{g/mL}$). However, similar levels of expression (5 $\mu\text{g/mL}$) were obtained when 3×10^6 cells were transfected with 5 μg DNA using PEI. Given that FuGENE 6 is much more expensive than PEI, I decided to use PEI in subsequent experiments, transfecting 3×10^6 cells with 5 μg DNA. Notably, the PEI was very toxic to the cells and cell viability was increased if the medium was changed to X-vivo 10 the day after transfection.

3.6 Comparison of expression methods

I have now described the expression of gp120-CB from 293T cells and CHO-K1 cells and using a codon-optimised gene. In order to verify that the gp120-CB was in the correct conformation when expressed in the different cell types I performed comparisons with monomeric gp120 expressed from Sf9 cells (made by A. Dey) and monomeric gp120 expressed from 293T cells (from the NIH ARRRP).

3.6.1 Comparison by Western blot

Purified gp120 expressed from Sf9 cells, gp120-CB TCS from 293T cells, gp120-CB TCS from CHO-K1 cells and purified gp120-CB from CHO-K1 cells were compared by Western blot (see Section 2.7.2). The blot was probed with antibody 2G12 at 1 µg/mL and with peroxidase-conjugated anti-human IgG at 1/1000 dilution. The results of the Western blot are shown in Figure 3.17.

The predominant band from Sf9 gp120 had an apparent molecular mass of 100 kDa as expected. Gp120-CB produced in 293T cells had an apparent molecular weight of 145 kDa, and expression was greater from the codon-optimised gene, as seen before (Figure 3.17). A second, undefined band of higher molecular mass was also present. Gp120-CB expressed in the TCS of CHO-K1 cells produced a band at an apparent molecular weight of 145 kDa as expected, but again a second undefined band of higher molecular mass was present. A sample of the same gp120-CB that had been purified by immunoaffinity chromatography (see Section 2.7.9) had a higher apparent molecular mass of 200 kDa and this was the only band. From these results I concluded that gp120-CB in the TCS of 293T and CHO-K1 cells was expressed as expected but that purification of gp120-CB by immunoaffinity chromatography altered the protein such that its mobility on SDS PAGE was altered.

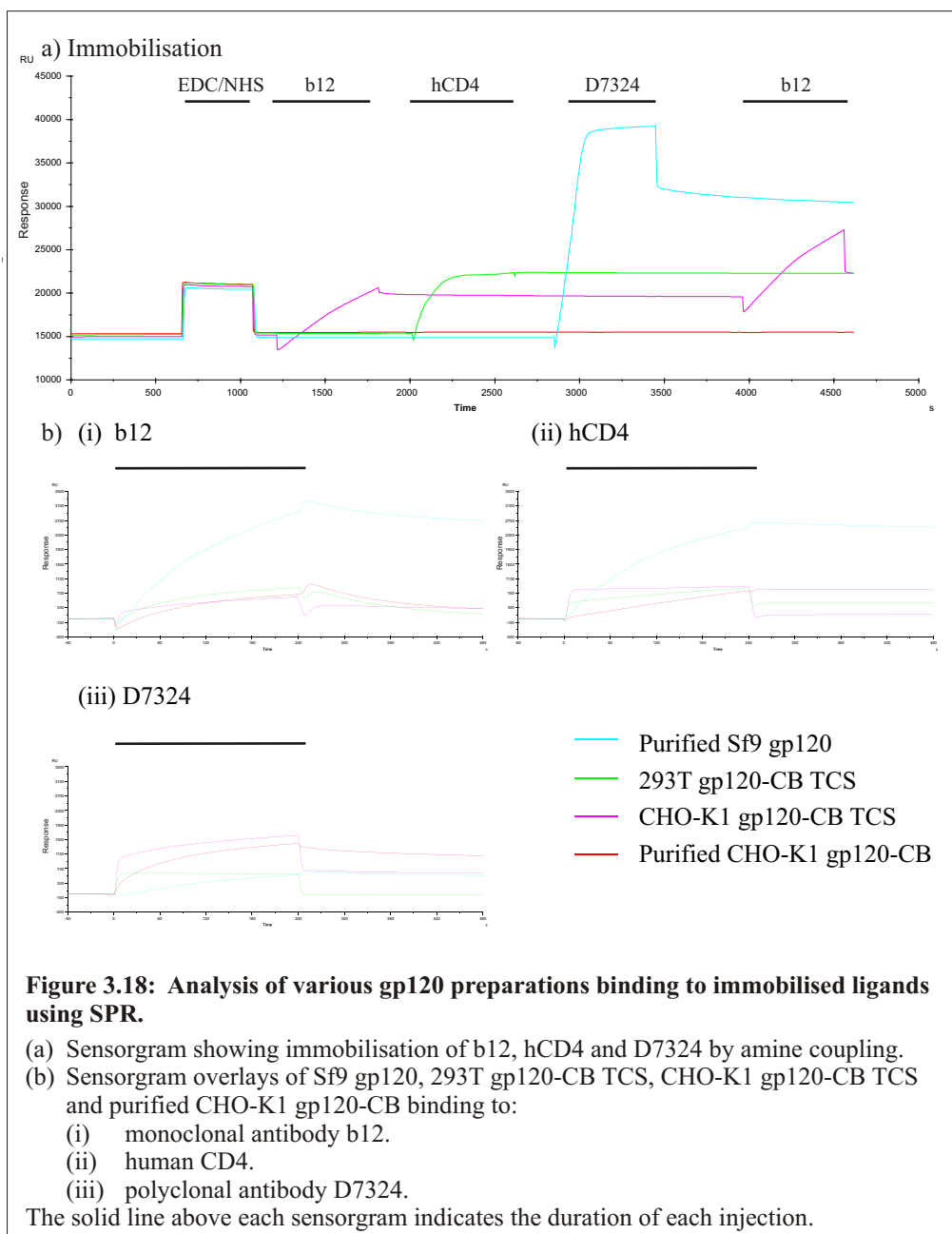
3.6.2 Comparison of antibody binding using SPR

A large number of antibodies that bind to various epitopes of gp120 have been described (Gorny and Zolla-Pazner 2003). I chose to analyse the interaction of gp120 and gp120-CB with CD4 and a number of anti-gp120 antibodies that bind to distinct epitopes on gp120, as described below. This would enable me to verify the presence of these epitopes on the fusion protein which would confirm that the gp120 was expressed in the same conformation as monomeric gp120. I also analysed OX68 antibody binding to check the

conformation of rCD4d3&4. I used two SPR methods to analyse these interactions, one in which hCD4 and antibodies were immobilised on a sensor chip surface and the various gp120 preparations were injected as analytes, and the second in which different gp120 preparations were immobilised and various antibodies and hCD4 were injected as analytes.

The neutralising monoclonal antibody b12 (Burton *et al* 1994; Roben *et al* 1994) was originally isolated from an HIV-1-infected individual by the Burton group, and functional data (Zwick *et al* 2003) and the crystal structure of b12 (Saphire *et al* 2001) indicate that the antibody binds in the CD4-binding site of gp120 due to an unusually long CDR H3 loop. Polyclonal antibody D7324 was raised against a peptide from the C5 region of gp120 at the C-terminus of the protein (Aalto bioreagents). The neutralising monoclonal antibody 447-52D was originally isolated from an HIV-1-infected individual by the Zolla-Pazner group, and reacts with a synthetic peptide from the V3 loop of the strain MN (Gorny *et al* 1992). The crystal structure of 447-52D in complex with the V3 loop peptide has been determined (Stanfield *et al* 2004). The neutralising monoclonal antibody 2G12 (Kunert *et al* 1998) was originally isolated from an HIV-1-infected individual by the Burton group. Epitope mapping studies have shown that 2G12 binds to a carbohydrate-dependent epitope on gp120 that predominantly binds to N-linked glycans attached to gp120 residues N295 and N332 (Scanlan *et al* 2002). The OX68 monoclonal antibody was raised in mice against rat CD4 by Don Mason and is directed against domains three and four (Brown and Barclay 1994).

In the first method, monoclonal antibody b12, hCD4 and polyclonal antibody D7324 were immobilised on a BIAcore CM5 chip by amine coupling (see Section 2.10.1, Figure 3.18a). Two injections of b12 were required to achieve the desired level of immobilisation. Purified gp120 from Sf9 cells (100 µg/mL in PBS), gp120-CB TCS from 293T cells (12 µg/mL), gp120-CB TCS



from CHO-K1 cells (5 µg/mL) and purified gp120-CB from CHO-K1 cells (100 µg/mL in PBS) were injected over the immobilised ligands. 35 µL of each gp120 preparation was injected after which the chip was regenerated with 5 µL glycine-HCl pH 2.5. Figure 3.18b shows sensorgram overlays of the binding of each preparation of gp120 to b12, hCD4 and D7324. The binding response of each gp120 preparation 600 seconds after the start of the injection, normalised for the molecular mass and concentration of the analyte, is summarised in Figure 3.20a.

Purified Sf9-expressed gp120 bound very strongly to immobilised antibody b12 and human CD4 as expected. This gp120 bound less well to antibody D7324 which may be because D7324 is a polyclonal antibody so some antibodies in the mixture may not bind to this strain of gp120. Gp120-CB TCS from 293T cells bound moderately to b12 and human CD4, but did not bind to D7324. The overall reduced binding compared to that of Sf9 gp120 is due to the lower concentration of the sample, but it was surprising that there was no binding at all to D7324. Gp120-CB TCS from CHO-K1 cells bound moderately to b12 and D7324, again at reduced levels due to the low concentration of gp120-CB in the sample, but there was no binding to hCD4. This caused me to question the conformation of the protein produced in this way. Purified gp120-CB from CHO-K1 cells bound moderately to b12, human CD4, and D7324. The fact that the gp120-CB from CHO-K1 cells shows different binding properties once it has been purified implies that the purification process caused structural changes in the protein.

In order to confirm these results I designed a second experiment in which the various gp120 preparations were immobilised and antibodies and hCD4 were injected over the sensor surface. Purified Sf9 gp120 and streptavidin were immobilised on the CM5 chip using amine coupling (see Section 2.10.1). Biotinylated gp120-CB from 293T cells TCS or purified biotinylated gp120-CB from CHO-K1 cells were then immobilised via streptavidin. The

immobilisation was performed so that equimolar amounts of each gp120 were immobilised and a sensorgram of the immobilisation procedure is shown in Figure 3.19a. I had to inject 25 μL purified biotinylated gp120-CB from CHO-K1 cells in order to achieve the required level of immobilisation. Given that the protein in this preparation was fairly concentrated I expected to need to inject about 5 μL . This finding implied that the gp120-CB was not fully biotinylated.

Human CD4 (26 $\mu\text{g/mL}$ in CHBS buffer), antibody 2G12 (100 $\mu\text{g/mL}$ in CHBS buffer), antibody b12 (100 $\mu\text{g/mL}$ in CHBS buffer), antibody OX68 (TCS), antibody 447-52D (TCS) and antibody D7324 (100 $\mu\text{g/mL}$ in CHBS) were injected over the sensor chip surface. 20 μL of each analyte was injected, after which the chip was regenerated with 5 μL of glycine-HCl pH 2.5 followed by 2 μL 50 mM sodium hydroxide. Figure 3.19b shows sensorgram overlays of each ligand binding to the immobilised gp120s. A summary of binding responses as 300 seconds post injection (i.e. just after the end of the injection) is given in Figure 3.20b.

Human CD4 bound to 293T gp120-CB TCS and Sf9 gp120 as expected but only weakly to purified gp120-CB from CHO-K1 cells. The same binding pattern was also seen with binding of b12 and 447-52D, indicating that the purified gp120-CB from CHO-K1 cells is not folded in the expected conformation. Antibody 2G12 bound more strongly to the Sf9 gp120 than to the gp120-CB constructs. This may be due to the difference in glycosylation patterns between the different cell types, and it is also possible that the rCD4d3&4 in the fusion protein obscures the 2G12 epitope. OX68 bound to both fusion proteins but not to the Sf9 gp120 as expected, indicating that the rCD4d3&4 are correctly folded. D7324 only bound significantly to the gp120-CB from 293T cells, which is contradictory to the data in the first experiment. It is possible that the polyclonal nature of the D7324 antibody

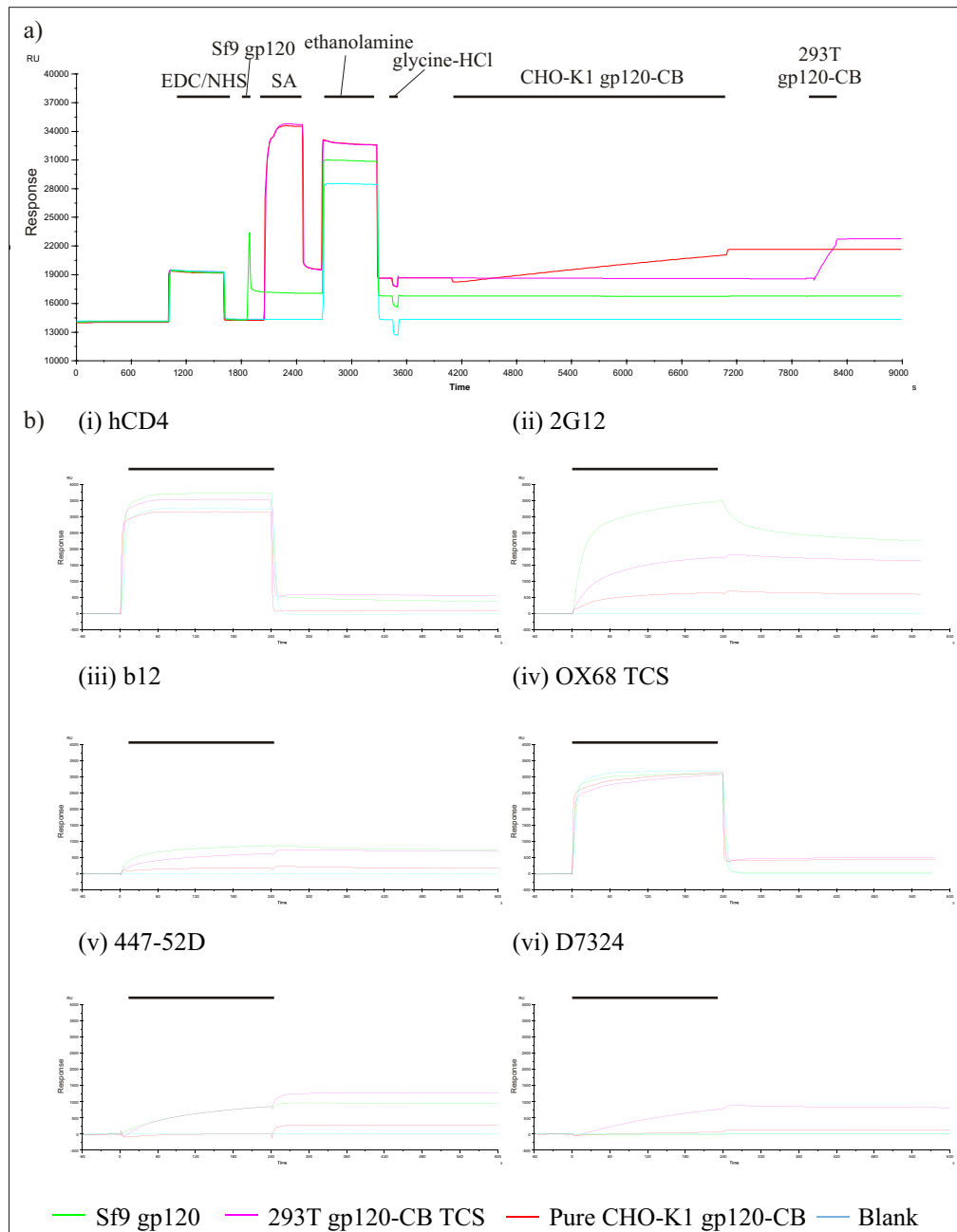


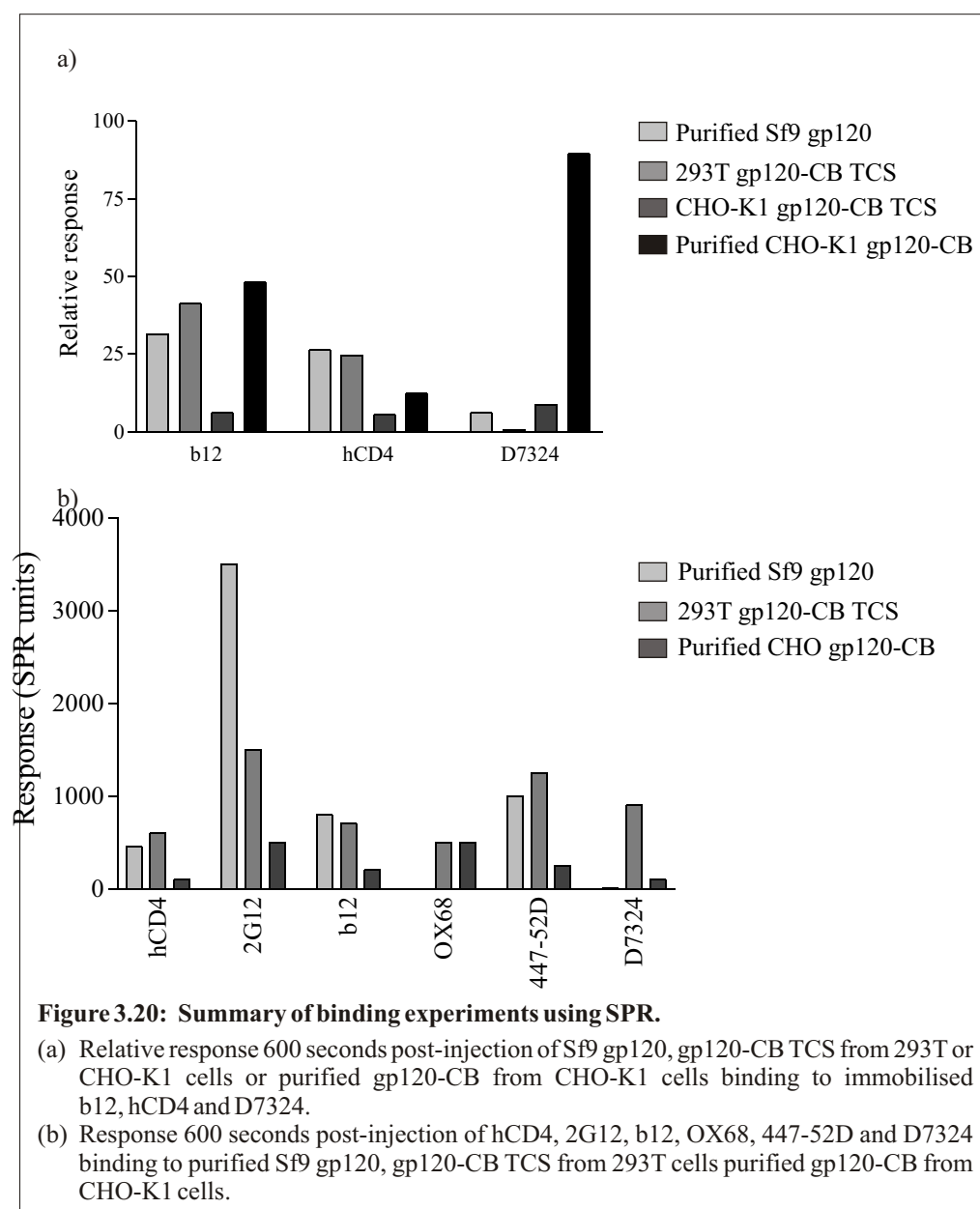
Figure 3.19: Analysis of various ligands binding to immobilised gp120s using SPR.

(a) Sensorgram showing immobilisation of streptavidin and Sf9 gp120 by amine coupling, followed by capture of biotinylated gp120-CB from CHO-K1 cells or 293T cells.

(b) Sensorgram overlays of:

- (i) hCD4
- (ii) 2G12
- (iii) b12
- (iv) OX68
- (v) 447-52D
- (vi) D7324

binding to the immobilised gp120s. The solid line above each sensorgram indicates the duration of each injection.



causes it to behave differently when it is immobilised on a sensor chip surface compared to when it is in solution.

From these experiments I concluded that the gp120-CB fusion protein expressed in 293T cells has a similar conformation to that of the monomeric gp120 produced from Sf9 cells. Differences in binding to D7324 can be explained by the fact that this antibody is polyclonal, and differences in 2G12 binding are probably due to differences in glycosylation patterns. In contrast the gp120-CB expressed in CHO-K1 cells showed marked differences in antibody binding compared to the Sf9 gp120, particularly in the second experiment. Given that this protein also showed an unexpectedly high molecular mass in the Western blot (Figure 3.17) I cannot be sure that the conformation of the gp120-CB fusion protein expressed in CHO-K1 cells is similar enough to that of the Sf9 gp120 to be used in biochemical experiments. In subsequent experiments I used gp120-CB TCS from 293T cells or monomeric gp120 from the NIH ARRRP.

3.6.3 Comparison of hCD4, aptamer 299.5 and CCR5 peptide binding by SPR

The preceding experiments, using antibodies and CD4, revealed subtle differences in the conformation of gp120 prepared under alternative expression conditions. I reasoned that such differences might also affect the binding site on gp120 for other ligands, such as CCR5 or aptamer. To test this I decided to study the interaction of these gp120 preparations with an N-terminal peptide of CCR5, aptamer 299.5 and human CD4.

A peptide comprising the 22 most amino-terminal residues of CCR5 binds to a CD4-inducible site of gp120, and can neutralise HIV-1 infection (Farzan *et al* 2000). The peptide is sulphated at two of its four tyrosines, and this sulphation is required in full-length CCR5 for the binding of gp120 and indeed for entry of HIV-1 (Farzan *et al* 1998). The synthesised peptide is also

biotinylated at the 5' end. Recently, the James group showed that aptamer B40 competes with the CCR5 peptide for gp120 binding (Dey *et al* 2005b).

Aptamer 299.5 (described in Chapter 6) is derived from the neutralising aptamer B40. It is likely that aptamer 299.5 binds to a region of gp120 that overlaps with the binding site of CCR5 and other CD4i antibodies such as 17b and 48d. Aptamer 299.5 was chemically synthesised and includes a biotinylated UTP at the 5' end of the RNA molecule.

Human CD4 and streptavidin were immobilised on a CM5 sensor chip by amine coupling (see Section 2.10.1). The CCR5 peptide and aptamer 299.5 were then immobilised on the streptavidin flow cells via their conjugated biotin tags. Any remaining sites were blocked by the injection of free biotin. A sensorgram showing the immobilisation procedure is shown in Figure 3.21a.

As controls, anti-hCD4 monoclonal antibody B4 (Wang *et al* 1999) (100 µg/mL in CHBS buffer), the CCR5-binding chemokine RANTES (5 µg/mL in CHBS) and purified rCD4d3&4 (100 nM in CHBS) were injected over the sensor surface. 35 µL of each ligand were injected and the chip was regenerated with 5 µL glycine-HCl pH2.5 if required (Figure 3.21b). As expected, monoclonal antibody B4 bound to CD4 only, and rCD4d3&4 did not bind to any of the immobilised ligands. RANTES did bind to the CCR5 peptide and rather interestingly also bound to aptamer 299.5 (discussed further in Section 9.4).

I then investigated binding of monomeric gp120 from Sf9 cells, monomeric gp120 from 293T cells (from the NIH ARRRP) and gp120-CB from 293T cells to the immobilised ligands. All gp120s were injected at 100 nM and were pre-incubated for one hour at 37 °C in the presence or absence of 200 nM soluble human CD4. 35 µL of each gp120 or gp120-hCD4 complex

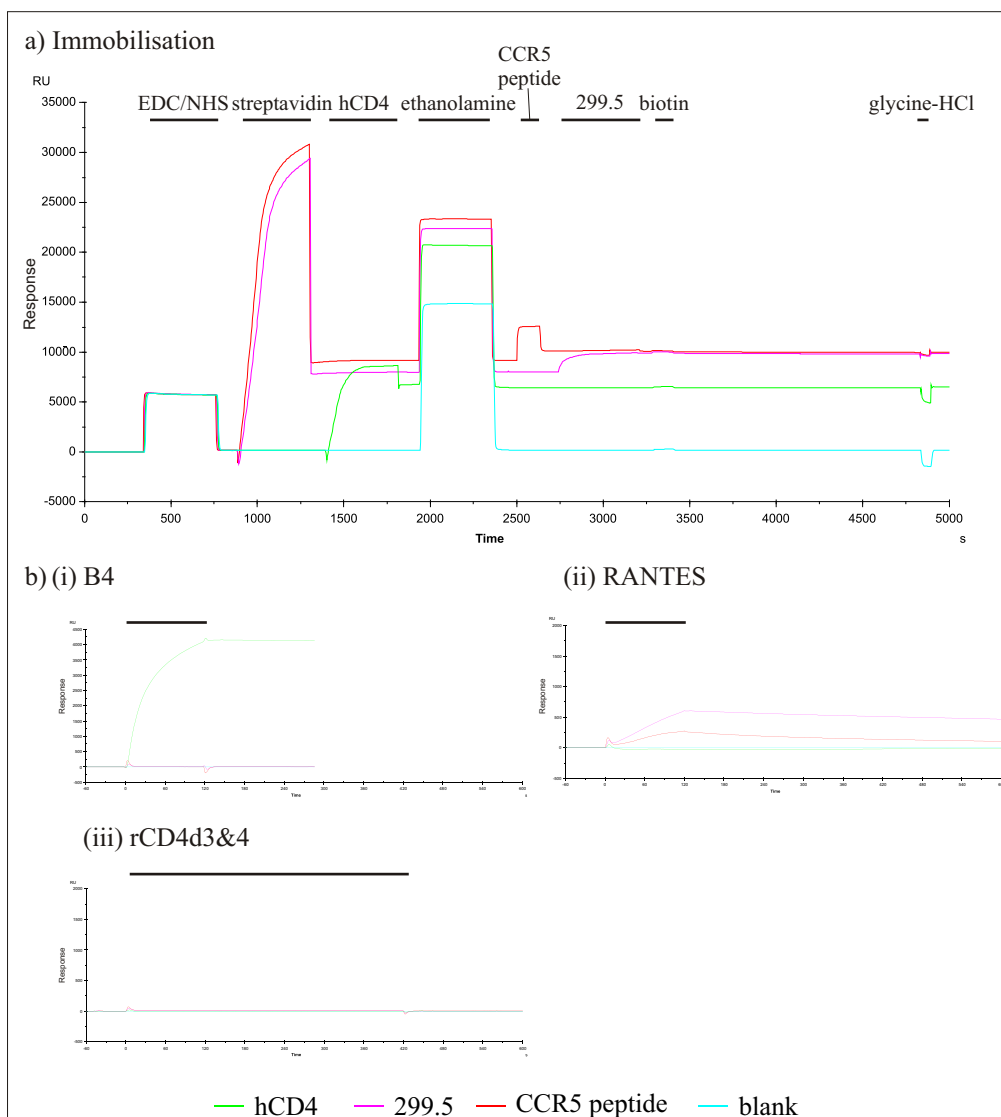


Figure 3.21: Analysis of various control molecules binding to hCD4, aptamer 299.5 and CCR5 peptide by SPR.

(a) Sensorgram showing immobilisation of hCD4 and streptavidin chip by amine coupling, capture of biotinylated CCR5 N-terminal peptide and biotinylated aptamer 299.5 blocking with free biotin and washing with glycine-HCl.

(b) Sensorgram overlays of control molecules binding to the immobilised ligands:

- (i) monoclonal anti-hCD4 antibody B4.
- (ii) Rantes.
- (iii) rCD4d3&4.

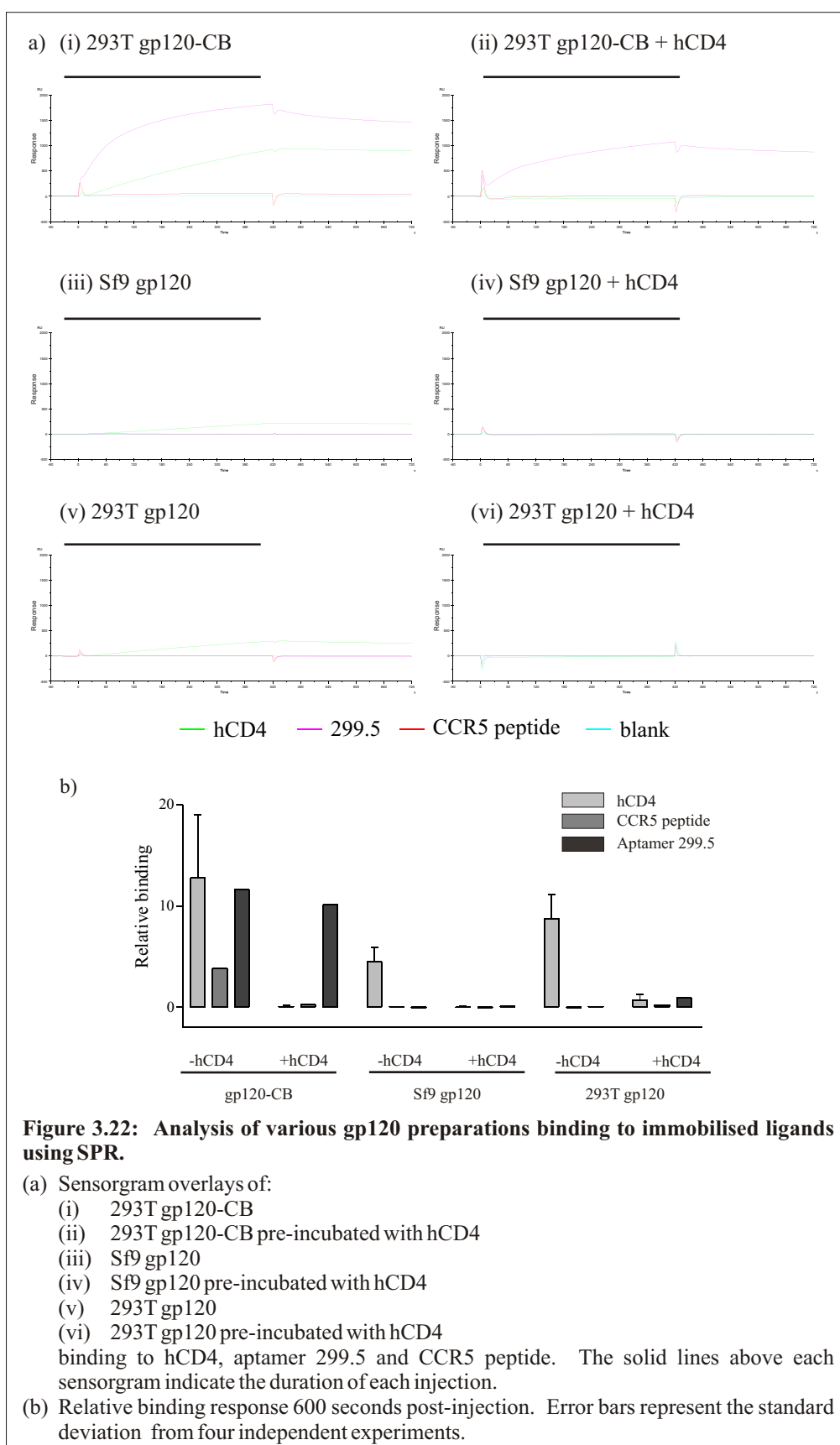
The solid line above each sensorgram indicates the duration of each injection.

were injected and the chip was regenerated with 5 μ L glycine-HCl pH 2.5 on all four flow cells and 5 μ L 7M urea on the aptamer flow cell if required. Representative binding curves are shown in Figure 3.22a, and Figure 3.22b summarises binding data from four independent experiments. Responses are shown at 600 second post-injection and are normalised for the molecular mass of the analyte and different immobilisation levels of the ligands.

Gp120-CB from 293T cells bound strongly to immobilised hCD4 and this binding was inhibited by pre-incubation with soluble hCD4 as expected. Gp120-CB also bound strongly to aptamer 299.5, and this binding was not affected by the presence or absence of soluble hCD4. Gp120-CB only showed minimal binding to the CCR5 peptide in the absence of soluble hCD4 as expected, but surprisingly no CD4-induced binding to the CCR5 peptide was observed. These results contradict previous work (Dey *et al* 2005b) that showed CD4-induced binding of Sf9 gp120 to the CCR5 peptide.

Monomeric gp120 from Sf9 cells bound to immobilised hCD4 and this binding was inhibited by pre-incubation with soluble hCD4 as expected. Sf9 gp120 did not bind to aptamer 299.5 in the presence or absence of soluble hCD4. Sf9 gp120 did not bind to the CCR5 peptide in the absence of soluble hCD4 as expected, but no CD4 induced binding to the CCR5 peptide was observed.

Gp120 from 293T cells bound strongly to immobilised hCD4 and this binding was inhibited by pre-incubation with soluble hCD4 as expected. Surprisingly, 293T gp120 did not bind to aptamer 299.5, but binding was slightly increased by the presence of soluble hCD4. 293T gp120 only showed minimal binding to the CCR5 peptide in the absence of soluble hCD4 as expected, and showed a very slight induction of CCR5 peptide binding when pre-incubated with soluble hCD4.



These results indicate that there are differences in the conformations of the different gp120s. While there are no significant differences between the gp120 expressed from 293T cells or Sf9 cells, many differences are seen when the gp120 is expressed as a fusion protein with rCD4d3&4. In particular the gp120-CB binds much more strongly to aptamer 299.5 than monomeric gp120 both in the presence and absence of soluble hCD4.

3.7 Discussion

In this chapter I have outlined a new expression system for gp120 in mammalian cells, in which the gp120 is expressed as a fusion protein with rCD4d3&4 and with a biotinylation site that is a target for BirA biotin ligase. The expression of this protein was increased when the gp120 gene was codon-optimised. I derived a CHO-K1 cell line that stably expresses the gp120-CB construct but this system produced protein that had a different conformation from that produced in 293T cells. Therefore codon-optimised gp120-CB from 293T cells was used in subsequent experiments.

I have explored different options for the transient transfection of 293T cells and have found that the use of PEI as a transfection reagent provides a cheap, reliable method of transfection. Optimisation of gp120-CB expression using PEI transfection resulted in the expression of gp120-CB at approximately 5 µg/mL at four days post transfection. The protein concentration could be increased 20- to 30-fold by concentration of the TCS using a Centricon spin column and the gp120-CB could be biotinylated by the BirA biotin ligase if required.

Various gp120 preparations were compared to verify the conformation of gp120 when expressed as a fusion protein. Firstly, I compared binding of gp120 from Sf9 cells and gp120-CB from 293T cells and CHO-K1 cells to

hCD4 and several antibodies. This showed that the gp120-CB fusion protein expressed in 293T cells had similar properties to the gp120 produced in Sf9 cells. Secondly I investigated the binding of these proteins, and gp120 from 293T cells, to immobilised hCD4, a CCR5 peptide and aptamer 299.5 in the presence or absence of soluble hCD4. This analysis showed that there are subtle differences between the gp120-CB fusion protein and monomeric gp120, particularly involving the binding to aptamers.

The fact that monomeric gp120 did not bind to immobilised aptamer contrasts with later experiments (see Chapters 4 and 5) in which aptamers bind to immobilised gp120. The most likely explanation is that the proteins have different conformations depending on whether they are immobilised on a surface or are in solution. I propose that immobilisation of gp120 opens up its conformation so that the aptamer binding site is more exposed. When gp120 is expressed as a fusion protein, I believe the presence of rCD4d3&4 means that the gp120 is in a more open conformation in solution as well as when it is immobilised. These properties mean that the gp120-CB is able to bind to immobilised aptamer more readily than soluble monomeric gp120. When the gp120-CB is used in future experiments one must take into consideration the fact that this protein may have a slightly different conformation from monomeric gp120.

Chapter 4:

Aptamer binding and structure

- Cloning and sequencing of novel anti-gp120 aptamers
- Aptamer binding to gp120
- Comparison of RNA structures

Chapter 4: Aptamer binding and structure

4.1 Introduction

In Chapter 1 I described previous work in the James lab in which RNA aptamers were selected against gp120 from the HIV-1 strain BaL using SPR (Khatri *et al* 2003). These aptamers consist of a 50-nucleotide variable region, which is flanked by 5' and 3' constant regions (22 and 42 nucleotides in length, respectively) and are made with 2'-fluoro-pyrimidines. The constant regions allow amplification of the aptamer DNA by PCR. The SELEX process results in a pool of RNA aptamers that bind to the target molecule, each round producing aptamers with higher specificity and affinity. M. Khatri cloned and sequenced 27 aptamers from the round 5 SELEX pool and these aptamers were numbered and given the prefix B, to stand for BaL (Figure 4.2).

Once aptamers have been selected they are often found to contain regions of primary sequence homology that form part of the region required for target binding. For example, 30 out of 45 aptamers cloned following selection against the hepatitis C NS3 protease contained a 9-base conserved sequence (Fukuda *et al* 2000). Unexpectedly there was no primary sequence homology in the variable region of the anti-gp120 aptamers. All of the anti-gp120 aptamers bound to gp120, 19 neutralised HIV-1 BaL in PBMC and several showed considerable neutralisation of a range of HIV-1 primary isolates (Khatri *et al* 2003).

The structure and binding mechanism of one aptamer, B40, has been studied in detail (Dey *et al* 2005a; Dey *et al* 2005b) and the delineation of its binding

site on gp120 is described in Chapter 6. I wanted to investigate whether or not other anti-gp120 aptamers bind to the same site on gp120. Often when aptamers are selected they all bind to the same part of the target molecule (or “aptatope”), for example anti-rat CD4 aptamers all bind to the CD42-like region in domain 1 of CD4 (Kraus *et al* 1998). If all the aptamers bind to the same site on gp120 then further investigations can focus on aptamer B40 alone, but if some aptamers bind to distinct epitopes on gp120 then these should be also be studied further. Aptamers that bind to different parts of gp120 could potentially be used in combination with one another to study the HIV-1 entry mechanism and as therapeutic agents.

My first aim was to clone and sequence more anti-gp120 aptamers from the round 5 SELEX pool. If duplicate sequences were cloned, this would enable me to estimate the extent of sequence variation within the pool of aptamers. Having verified that the novel aptamers bound to gp120 I then planned to use all the anti-gp120 aptamers to investigate whether or not any of them bind to the same gp120 aptatope as B40. Given that not all of the aptamers tested were able to neutralise HIV-1, this led me to believe that different aptatopes may be present in the aptamer pool. I first approached this problem using a competition assay to see if other aptamers could bind to gp120 at the same time as B40. If both aptamers could bind concurrently this would indicate they had distinct gp120 aptatopes, whereas if they competed with one another for binding this would imply they share the same aptatope. Unfortunately this method (described in Chapter 7) was unsuccessful because of non-specific competition from any RNA, so next I used a computational method to investigate the structural diversity of the family of anti-gp120 aptamers.

A. Dey studied the secondary structure of four anti-gp120 aptamers, B4, B40, B28 and B116, using folding simulations, energy landscape mapping and endonuclease digestion (unpublished work, Dey 2004). On comparing these structures by eye, he noticed that three of the aptamers contained a

homologous structure in which two small stem loops branch from a central loop (Figure 4.6a). Loops 1 and 2 seemed to contain small regions of primary sequence homology; a CAXC motif in loop 1 and an ACXXG motif in loop 2. These homologous sequences, that are apparent when considered in the context of the RNA secondary structure, are not detected when a primary sequence alignment is performed because the sequences are too short and occur in completely different parts of the aptamer sequences. Subsequent mutational analysis, footprinting studies and truncation of B40 (see Chapter 5 and Dey *et al* 2005a) showed that it is likely that these loops are in the footprinted region of B40 that binds to gp120 (Figure 4.8a). I decided to find out whether this motif was present in any other anti-gp120 aptamers. I hypothesised that other aptamers containing this branched structure may also bind to the same aptatope as B40 on gp120.

In order to do this I used the bioinformatics program RSmatch (Liu *et al* 2005) to perform an RNA structural alignment of all the anti-gp120 aptamers isolated to date. I was able to categorise the aptamers into three structural groups based on homologous motifs, namely those that contained the B40-like branched motif, those that contained a mirror image of the branched motif and those that were structurally unrelated. I used the program Sfold (Ding *et al* 2004) to estimate the probability that B40-like structures exist in solution. The structural relationships between the aptamers was explored using a network diagram and the sequences of homologous loops were analysed using ClustalW (Thompson *et al* 1994) to find consensus sequences.

4.2 Cloning and sequencing of novel anti-gp120 aptamers

RNA from the round 5 SELEX pool of anti-gp120 aptamers was reverse-transcribed to produce DNA, which was used as a template for PCR (Section

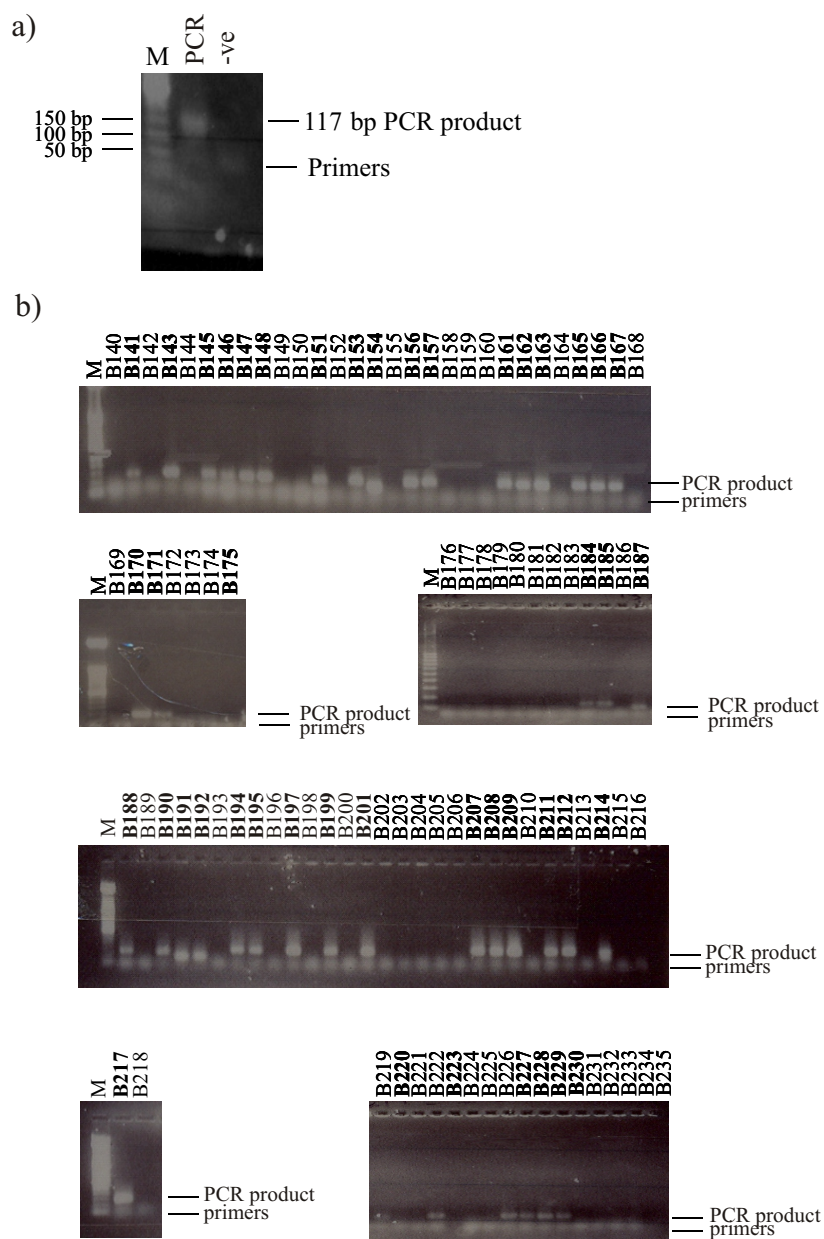


Figure 4.1: 1 % agarose gels stained with ethidium bromide showing the products of PCR amplification of aptamer DNA.

- (a) PCR amplification of aptamer DNA from the round 5 SELEX pool.
 (b) Colony PCR to check for inserts following shotgun cloning of aptamer DNA into the pCR3.1 TA cloning vector. Positive clones are indicated in **bold** type.

2.2.1, Figure 4.1a). The PCR products were immediately ligated into the pCR3.1 TA cloning vector (Invitrogen) with T4 DNA ligase (see Section 2.3.4), so that each plasmid would contain a unique insert. The pool of plasmids was transformed into competent cells (Section 2.3.5) and a colony PCR was performed on the resulting clones in order to detect the presence of inserts in the cloning vector. 45 of the 95 colonies tested contained an insert (indicated in bold type in Figure 4.1b), and these colonies were used to seed 5 mL cultures from which plasmid DNA was extracted and sequenced (Sections 2.3.6). Figure 4.2 shows the sequences of anti-gp120 aptamers isolated previously (Khati *et al* 2003) and Figure 4.3 shows the sequences of the novel anti-gp120 aptamers. The variable sequences of all the aptamers were all unique except for B165 and B166. I assumed that these were identical because the colonies were adjacent and the samples had become contaminated, so B166 was excluded from further study. A further eight clones were discarded because DNA sequencing showed that the plasmid preparation was not monoclonal. None of the new sequences shared any primary sequence homology, nor were any identical, providing no indication as to the extent of sequence variation in the round 5 SELEX pool. Most aptamers were the expected length of 117 nucleotides, except aptamers B154 and B190 that have considerable deletions in the variable region.

4.3 Aptamer binding to gp120

The binding of the novel anti-gp120 aptamers was analysed by SPR using a BIAcore 2000. The plasmid DNA of each monoclonal aptamer was used as a template for individual PCRs (see Section 2.2.1). The PCR products were ethanol-precipitated and used as templates for *in vitro* transcription reactions with 2'-fluoro-pyrimidines (see Section 2.2.2). The resulting RNA aptamers were phenol:chloroform extracted, ethanol-precipitated and refolded in CHBS buffer at 1 μ M before use (see Sections 2.2.6, 2.2.7 and 2.2.13).

GGGAGACAAAGACUAGACGCUCAAACCUAUAAGACACCUUUUUUUGGCC...AAUUCGACAUGAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAGACCGGUGUCCUGAUCCCAAUGCCACAAGUACCAUUGCAGUGAGUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAGACGCGUUAAGGAGAUUAGGCACAGUUGGACAGUGAUCGSCUGAGUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAAGGCGCUAAUAGUGCCUGUAGCCUCCUCAAUCCGACUCAUUAUUAAGUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAACCAAGGSCUAAAGUCCUUAAGAGGUCACAA...CUAAAGUUCGCAUGAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAACCAAGGSCUAAAGUCCUUAAGAGGUCACAA...ACUGCGUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAAGACCUUAACCGUGAGAUACAGCUUCUUCGAGC...ACGCGAC...UUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAACCGAUUUUCUCUGUAGUGGGAUUUUCCCGCUGAACUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAACCUACCUAAUUAUAAACUUUGGGCAGUAUCCCGUUUGCUUUUUCUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAAGUUAUUAUACACAGGUUAAGCGUUAACUUCGUGGACAGCAAGAAUCCUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAACAUUGCCCAAUUCCUUGAAUUCUGACUGCUGGUAUAGACUUAUAGACUUAUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAAGGAGAAUUAUGAGCGGGAACAACUUUGUCCUGUUCGUGUCCGUAUGAGCGUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAUCCUUGAGGGCCCCAUGACUGAUAGUAUUCUGGCC...UCGAGUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUAAUUGUGGCCACGCCGUAUUUACGUUUUACCGCACGGAUUGGUUUUGUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAAGUGUUAUAGCUGGCCACAACUUCGUGUUCUUAACGGCCUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAGUCAUGGACACCCUUUCUGUUUCUGAUUGCCACAACCGAUUCCAAUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAUCUGCGUAGACCAUUUUCCAAUACAAGUACCGCGUCUACAAGUUAUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAACCGGACCAACCCUAGACUUAUGCACAUCUGUUUUUUGCUCUUUCCCCGAUUUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAUUGGUG...CCGAGUGG...UUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAAAGUUUACUGUCCGAGCAAGAUCCUUAUUGGCGCCGCUUUAUGACUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAACCCCAUGGCACGCCGAUUCGUUUGCUGCCGCGUCCAUAAAUUUCUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAA...AUGACUAUCCGCAACAGCCACAAGUCUUAUUCGCGCCAACCCUUGCUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAACCGUGCUCUAUCUUUUAUUCUGUGGGCUCUGCGGUAAGCCUUAAGCUCUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAUAUACAGCGAAGUUAACAGCCCAUACAGGUACAACUUGCGCCGACAAGUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAAGCGCAACCCGUUAUACCCUCCAUCUGGUUAUCCCGCUUAAAGCUUCCCUUAUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAACCUUACCUUACCUACCUACCUUACCUU...GAGCCUAACGCCGCCGACACCCUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAACGUAUUCUACGUAAGCGUAGACUCCGUGGCGGGCUGUCC...AUUUUAUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU
GGGAGACAAAGACUAGACGCUCAAACGCCAGGUUAUACGGCCGGAUUCGACUAGACUCACAAAGUCCCUUUAGUGAGGGUAAAUUU

Figure 4.2: Sequences of anti-gp120 aptamers cloned and sequenced by M. Khati (Khati *et al* 2003). Note the 5' and 3' constant regions flanking the 50-nucleotide variable regions. Dots represent deletions.

[illegible]

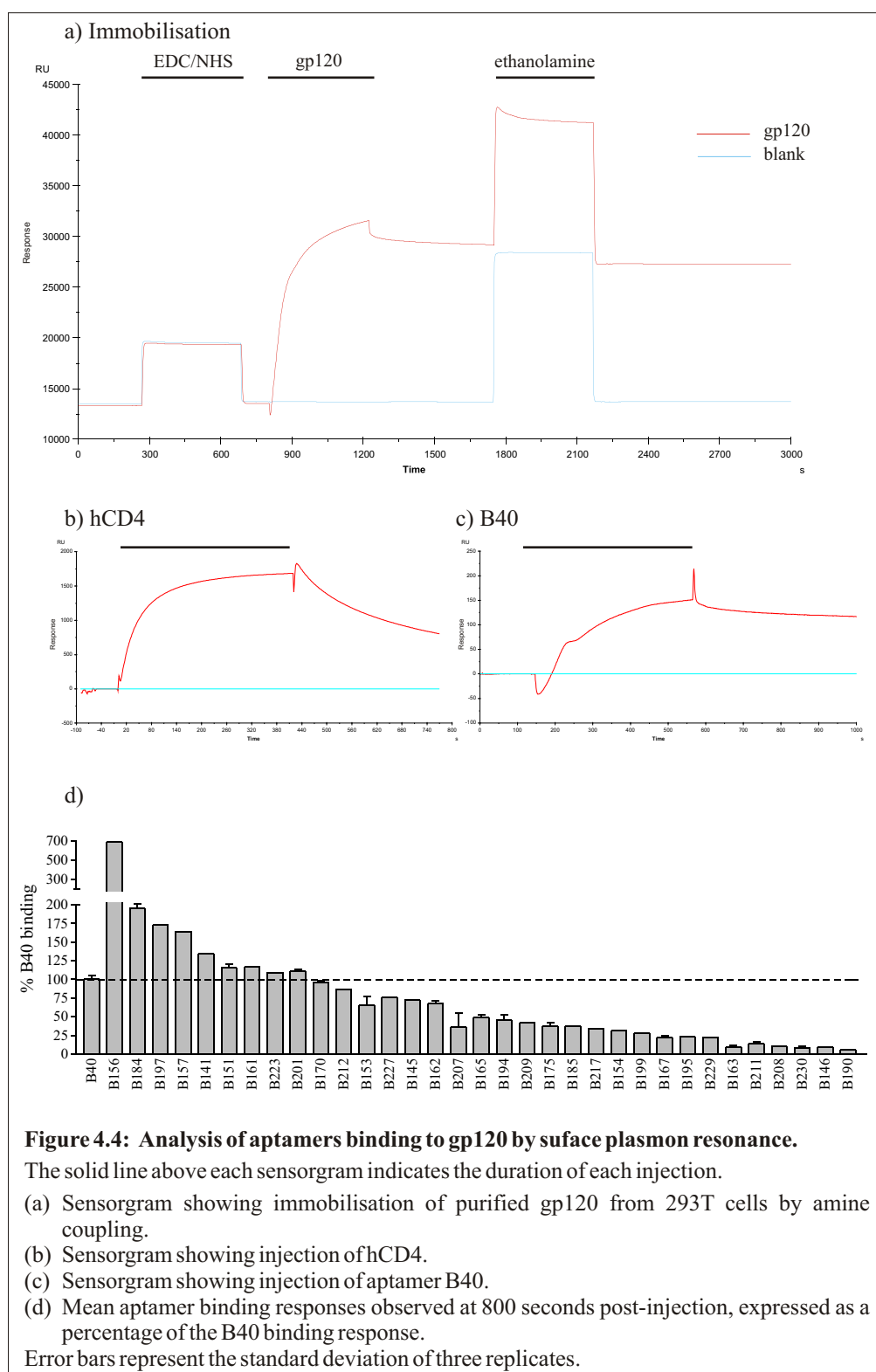
Figure 4.3: Sequences of novel anti-gp120 aptamers cloned and sequenced in this project.

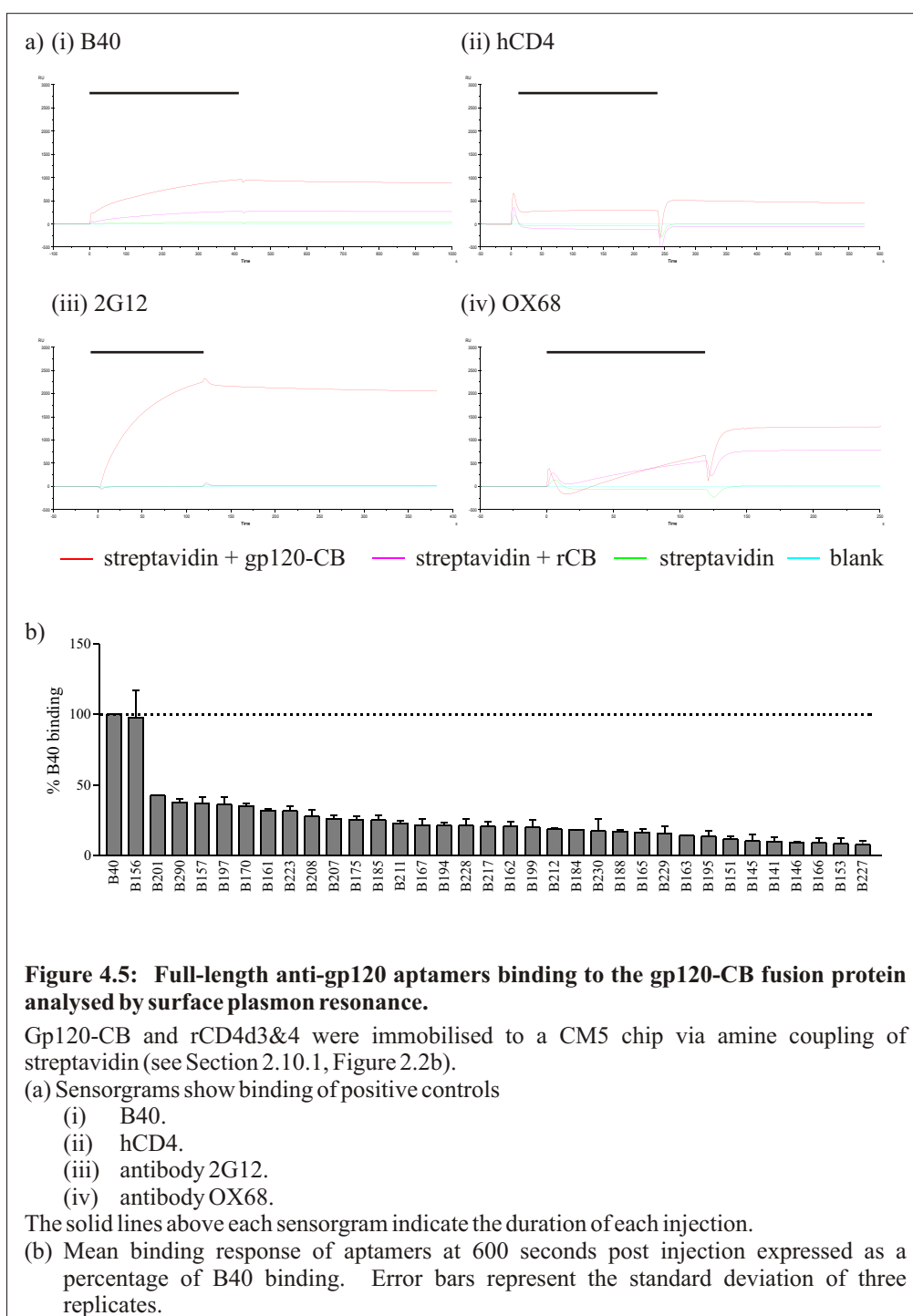
Note the 5' and 3' constant regions flanking the 50-nucleotide variable regions. Dots represent deletions.

Soluble gp120 produced in 293T cells was obtained from the NIH ARRRP. The gp120 was immobilised on flow cell 1 of a CM5 sensor chip using amine coupling (see Section 2.10.1, Figure 4.4a) and flow cell 2 was left blank as a control. Soluble hCD4 and aptamer B40 were injected as positive controls over flow cells 1 and 2 (Figure 4.4b and c). 35 μ L soluble hCD4 (100 μ g/mL in CHBS) were injected and the chip was regenerated with a 5 μ L injection of glycine-HCl pH 2.5. 35 μ L aptamer B40 (1 μ M in CHBS) were injected and the chip was regenerated with a 5 μ L injection of 10 mM NaOH. As expected, both control molecules bound to gp120 but not to the blank flow cell.

Each novel anti-gp120 aptamer was injected in duplicate (35 μ L injection at 1 μ M in CHBS) over flow cells 1 and 2, and the order of sample injection was randomised. The chip was regenerated after each aptamer injection with 5 μ L of 10 mM NaOH, after which the flow rate was increased to 100 μ L/min for 15 min in order to remove any residual NaOH. The binding response of each aptamer at 800 seconds post-injection was noted and is expressed as a percentage of B40 binding (Figure 4.4d). At this time-point the injection has finished and the aptamer is dissociating at a constant rate. As all the aptamers had the same molecular mass and concentration, the binding response is proportional to the affinity of the aptamer for the gp120.

The aptamer binding affinities for gp120 varied considerably compared to aptamer B40. Notably one aptamer, B156, bound to gp120 with much higher affinity than B40. Several aptamers bound with comparable affinities, and the remainder bound with lower affinities, some only reaching 5 % of the B40 binding response. The experiment was also repeated to investigate aptamer binding to the gp120-CB fusion protein. The results were very similar to the binding of aptamers to monomeric gp120, and are shown in Figure 4.5.





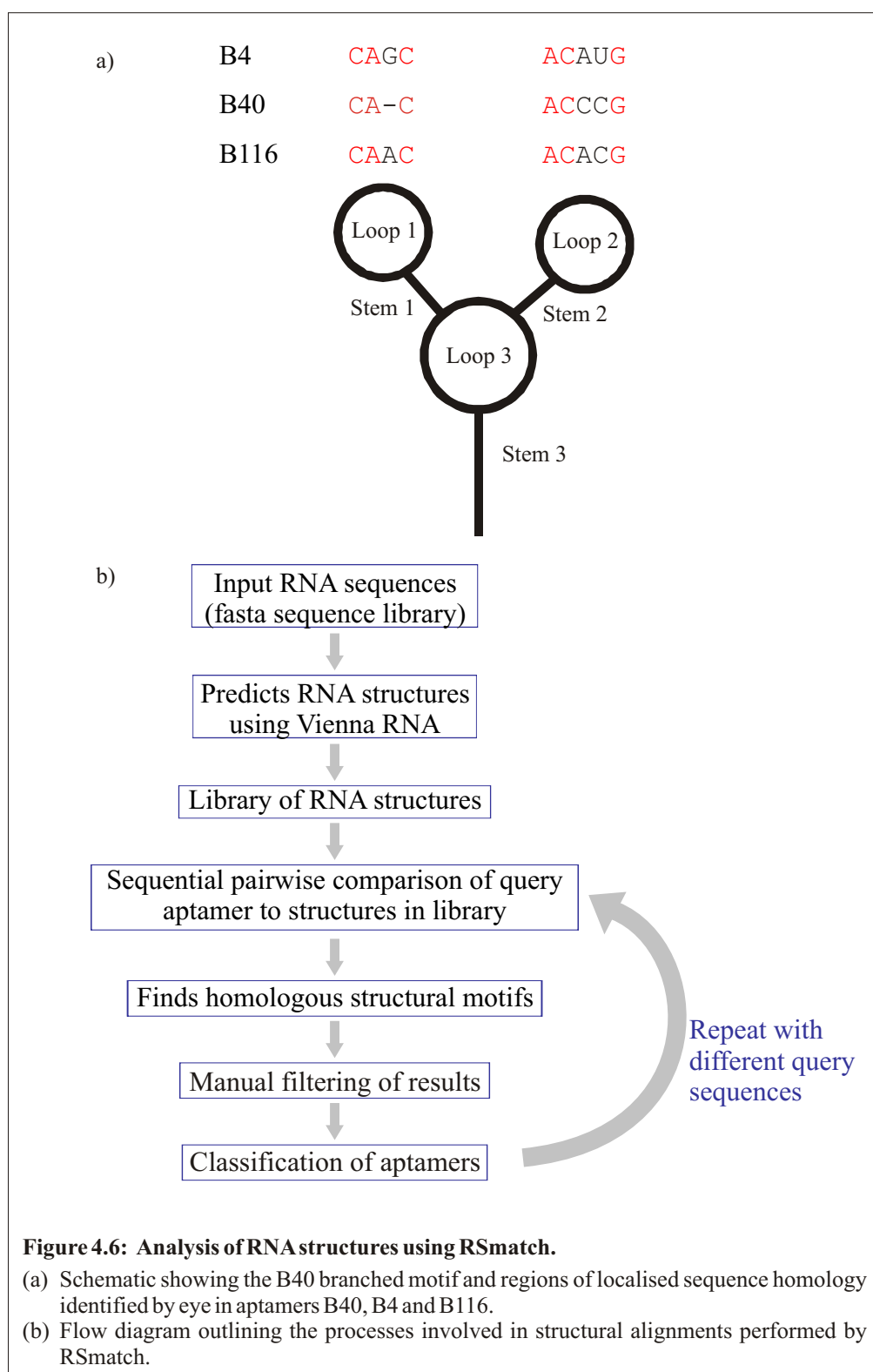
4.4 Comparison of RNA structures

4.4.1 Experimental design

The program RSmatch (Liu *et al* 2005) was used to investigate whether or not aptamers other than B40, B4 and B116 contained the branched motif (Figure 4.6a). A number of programs are available to perform RNA structural alignments and the rationale behind the use of RSmatch has been explored in Chapter 1 (Section 1.3.3). RSmatch initially uses the RNAfold algorithm from the Vienna RNA package (Hofacker 2003) to predict RNA structures that exist within a certain distance from the minimum free energy (MFE) structure, creating a library of possible RNA structures for all the aptamer sequences. The program then performs pairwise structural alignments between a query structure and the database of library structures and homologous motifs or “hits” are identified. B40 was used as the query structure to find other aptamers that shared the B40 branched structure, and then the structural alignment was repeated using other “B40-like” aptamers as query structures. A flow diagram outlining the processes performed by RSmatch is shown in Figure 4.6b. This iterative method identified homologous motifs in other aptamers, and using this information I was able to classify aptamers into three structural groups.

Definitions of the aptamer libraries are given in Chapter 2 (Section 2.9.2). The RSmatch output is given in Stockholm format in which unpaired bases are represented by dots and base-pairs are represented by parentheses.

Initially the full-length aptamer library was used as the input file for RSmatch and this was probed with the full-length aptamer B40 sequence. This resulted in most of the hits coming from the 5' and 3' constant regions, with only a small number of hits (which were in fact more relevant) coming from the variable regions. I have defined the hits obtained with this method as follows:



Type 1: The hit is in the variable region of both the query and library aptamer (this is the most interesting type of hit).

Type 2: The hit is in the variable region of the query aptamer and the constant region of the library aptamer.

Type 3: The hit is in the constant region of the query aptamer and the variable region of the database aptamer.

Type 4: The sequences are identical because the hit is in the constant region of both aptamers.

Later, whenever the full-length library was probed with a full-length query, I filtered the results manually so that only type 1 hits were included.

RSmatch was also implemented using the variable-sequence aptamer library as the input file, and this library was probed with both full-length and variable query sequences. This approach meant that the aptamer folding simulation had to be interpreted with caution, as the structure of the variable region alone is sometimes different from the structure of the same bases when folded in the complete aptamer. Therefore I manually screened the results in order to see if a motif identified as a hit in the variable sequence library was present in the context of the whole aptamer structure. I used Mfold to predict the aptamer structures, and motifs that did not occur in these predicted structures, even when specified using the constraining options, were excluded from the results.

It is possible to predetermine the structure of the query aptamer by using a structure file rather than a sequence file as the input. This approach was used when the structure of the query sequence produced by RNAfold was not satisfactory. For example, Mfold predicts that aptamer B116 folds into three structures, only two of which contain the B40-like loops of interest. However the MFE structure produced by RNAfold, which would be used as the query structure if the sequence file was used as the input, does not contain the B40-like motif. Therefore, in this case, a user-defined B116 structure file, whose

structure does contain the B40-like motif, was used as the query. Similarly, when a variable-sequence query was used, a user-defined structure file was made if the structure of the variable region generated by RNAfold was different from the structure of the same bases when modelled in the full-length aptamer.

4.4.2 Optimisation of RSmatch parameters

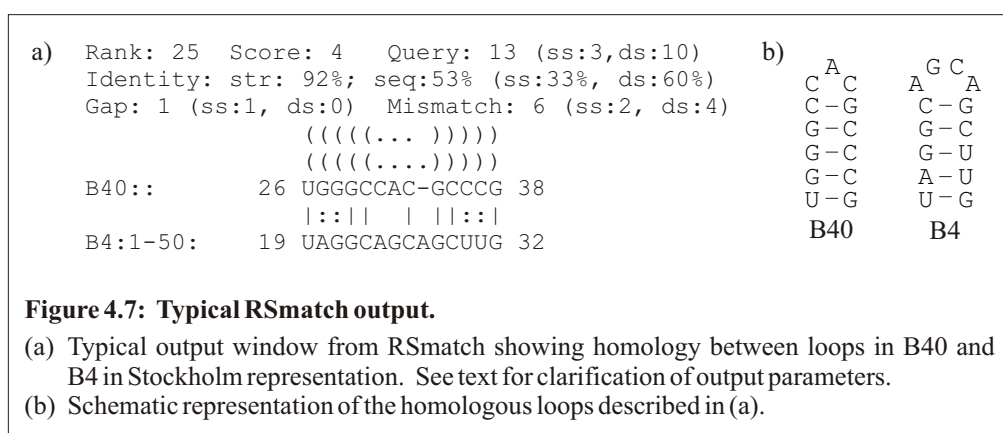
Options for a number of parameters need to be specified when running RSmatch (see Section 2.9.2). In this study the parameter *r* was optimised. This value defines the number of alternative structures that are formed when the folding algorithm is implemented. Alternative structures with an energy difference from the MFE structure of up to the value of *r* (in kcal/mol) are included. The default value for this parameter is 0, i.e. only the MFE structure for each RNA sequence is considered.

The full-length aptamer library was analysed using RSmatch with full-length B40 as the query sequence, and the *r* value was varied from 0 to 3 (Table 4.1). As expected, the total number of structures in the library rises with the “*r*” value, but interestingly this did not increase the total number of hits. In fact, increasing the *r* value above 1 decreased the number of type 1, 2 and 3 hits obtained (see Section 4.4.1 for definitions of type 1, 2 and 3 hits). This may be because the program cannot find any matches other than in the homologous regions when there are so many structures to compare. An *r* value of 1 was used in further experiments because this value produced the greatest number of type 1 hits.

r value	Structures	Total hits	Type 1 hits	Type 2 hits	Type 3 hits
0	97	61	4	2	2
0.5	588	66	6	6	2
1	2082	66	5	5	1
1.5	6551	66	3	4	0
2	18252	57	4	4	0
3	112000	53	0	0	0

Table 4.1: Results of optimisation of the parameter “r” in RSmatch.

See text for definitions of type 1, 2 and 3 hits.



4.4.3 Data analysis

Figure 4.7 shows a typical RSmatch result with the alignment of a homologous loop found in aptamers B40 and B4, and a schematic depiction of these loops. Identical bases are indicated by a dash and mutations that conserve the secondary structure are indicated by a colon. The complete RSmatch results detailing the query structure and all hits except type 4 hits are given in Appendix 2.

The analysis was first performed with B40 as the query, after which the aptamer that seemed to contain the most strongly B40-like or Mirror structure was used as the query. Table 4.2 summarises the RSmatch results, indicating which queries produced which hits. Based on these hits, and the position of the structural motif in the context of the whole aptamer structure (as predicted by Mfold), the aptamers were classified into three structural groups; those that are similar to B40 (B40-like), those that share homologous loops but in a mirror orientation (Mirror), and those that are structurally unrelated to B40 (Other). 19 aptamers were found to be B40-like, 9 were found to have Mirror structures and 18 were classified as Other. Figure 4.8a-d shows the predicted structures of aptamers that are representative of each of these groups. Figure 4.8e shows the structure of aptamer B197, which is classed as Other, but can be constrained to include B40-like motifs. This constrained structure has an energy of -12.12 kcal/mol, compared to the MFE structure energy of -19.3 kcal/mol, making it unlikely that the induced B40-like structure would ever spontaneously form in solution. Similarly, the structures of B156 that are Mirror have much less negative free energies than the linear structures that are also predicted to occur. This led me to use the program Sfold to investigate the range of structural conformations that exist for one aptamer sequence, and the probability that a B40-like or Mirror conformation actually exists (see Section 4.4.5).

Table 4.2: Assignment of aptamers to structural groups based on structural homology.

The Loop 1 and Loop 2 columns indicate which aptamer was the query structure when homologous motifs were found in that loop. Asterisks indicate where more than one loop 1 or loop 2 match was found when different query structures were used.

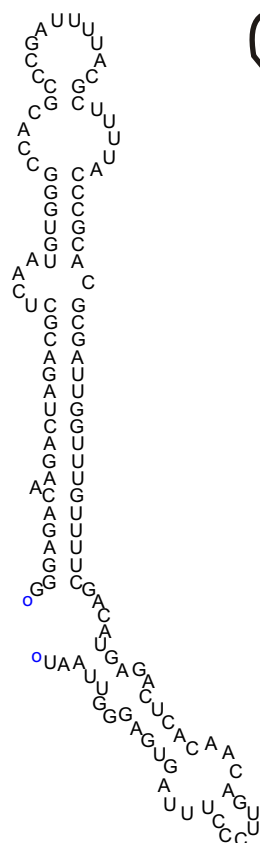
- B11 has one motif homologous to B40 loop 1 and a second motif homologous to B116, B156 & B163 loop 1.
- B19 has one motif homologous to B40 loop 1 and a second motif homologous to B116 loop 1.
- B28 has one motif homologous to B156 loop 1 and a second motif homologous to B40 loop 1.
- B31 has one motif homologous to B40 loop 2 and a second motif homologous to B141 & B229 loop 2.
- B33 has one motif homologous to B62 loop 1 and a second motif homologous to B229 loop 1.
- B86 has one motif homologous to B163 loop 1 and a second motif homologous to B116 loop 1.
- B137 has one motif homologous to B40 loop 2 and a second motif homologous to B229 loop 2.
- B141 has one motif homologous to B116 & B156 loop 1 and a second motif homologous to B163 & B229 loop 1.
- B145 has one motif homologous to B4 loop 1, a second motif homologous to B4 & B116 loop 1 and a third motif homologous to B156 loop 1.
- B165 has one motif homologous to B4 & B40 loop 1, and a second motif homologous to B116 loop 1.
- B167 has one motif homologous to B4 & B163 loop 1 and a second motif homologous to B116, B141 & B84 loop 1.
- B197 has one motif homologous to B4 loop 1 and a second motif homologous to B116 & B141 loop 1.
- B207 has one motif homologous to B4 & B229 loop 1 and a second motif homologous to B116, B141, B62 & B163 loop 1.

Aptamers in which no homologous motifs were found are not included in the table. These aptamers are B44, B55, B65, B81, B132, B136, B151, B153, B154, B161, B162, B170, B190, B208 and B209.

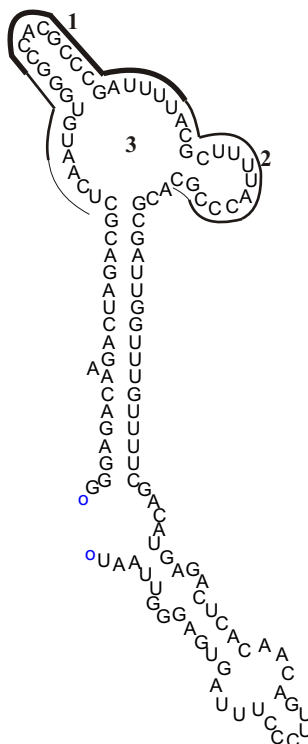
Aptamer	Loop 1	Loop 2	Class
B1	B4		Mirror
B3		B141, B62	Other
B4	B40, B116, B156, B84, B163		B40-like
B5	B4		Other
B9	B40		B40-like
B11	B40, B116, B156, B163*		B40-like
B19	B40, B116*		B40-like
B28	B40, B156*		B40-like
B30		B163	Other
B31		B40, B141, B229*	B40-like
B33		B62, B229*	Other
B36		B229	Mirror
B38	B141, B156		Other
B40	B4, B116, B229	B156	B40-like
B45		B141, B163, B229	B40-like
B62	B141		Mirror
B82	B116		Mirror
B84	B4, B156	B116	Mirror
B86	B116, B163*	B40	Mirror
B114		B141	Other
B116	B40, B4, B141, B163		B40-like
B137		B40, B229*	Mirror
B141	B116, B156, B163, B229*	B40	B40-like
B145	B4, B116, B156*		B40-like
B146	B40		B40-like
B156	B4, B141, B156, B163		Mirror
B157		B116	Other
B163	B4, B156, B229		B40-like
B165	B40, B4, B116*		B40-like
B167	B4, B116, B141, B84, B163*		B40-like
B175	B4		Other
B184		B116	Other
B185	B229		B40-like
B188	B4		Other
B194		B40	Other
B195	B4		Other
B197	B40, B116, B141*	B62, B229	Other
B199	B116, B163, B229		Other
B201		B4	Other
B207	B4, B116, B141, B163,	B62	B40-like
B211		B141, B62	Other
B212	B116, B229		Other
B217	B116, B163, B229		B40-like
B223	B4, B163		Other
B229	B116, B163	B141, B62	B40-like
B230		B4	Mirror

a) B40

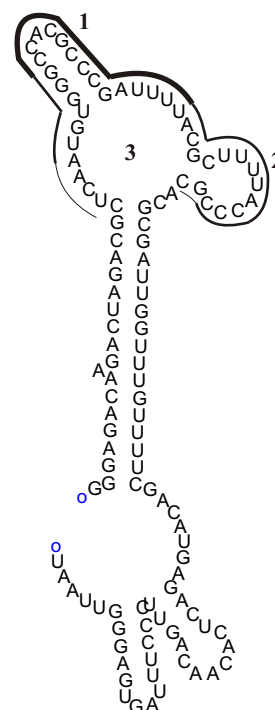
-30 kcal/mol



-27.3 kcal/mol



-26.7 kcal/mol

**Figure 4.8: Alternative structures of anti-gp120 RNA aptamers.**

- (a) Aptamer B40.
- (b) Aptamer B4 (B40-like).
- (c) Aptamer B156 (Mirror).
- (d) Aptamer B161 (Other).
- (e) Aptamer B197 (Other) including an induced B40-like structure.

The structures and free energies of were predicted by Mfold. Motifs that are homologous to B40 loops 1, 2 and, 3 are indicated. The outlined region of B40 indicates nucleotides that are required for binding to gp120 and the thickness of the line represents the certainty that those residues are required (Dey *et al* 2005a).

b) B4

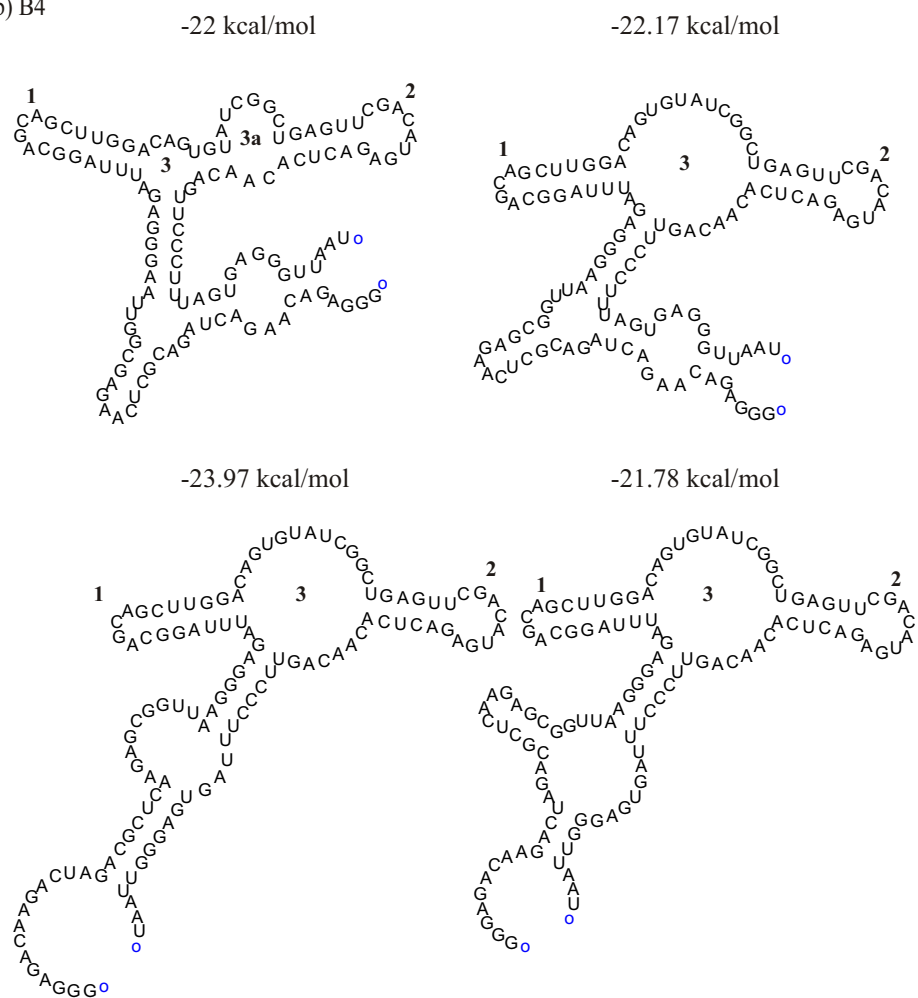
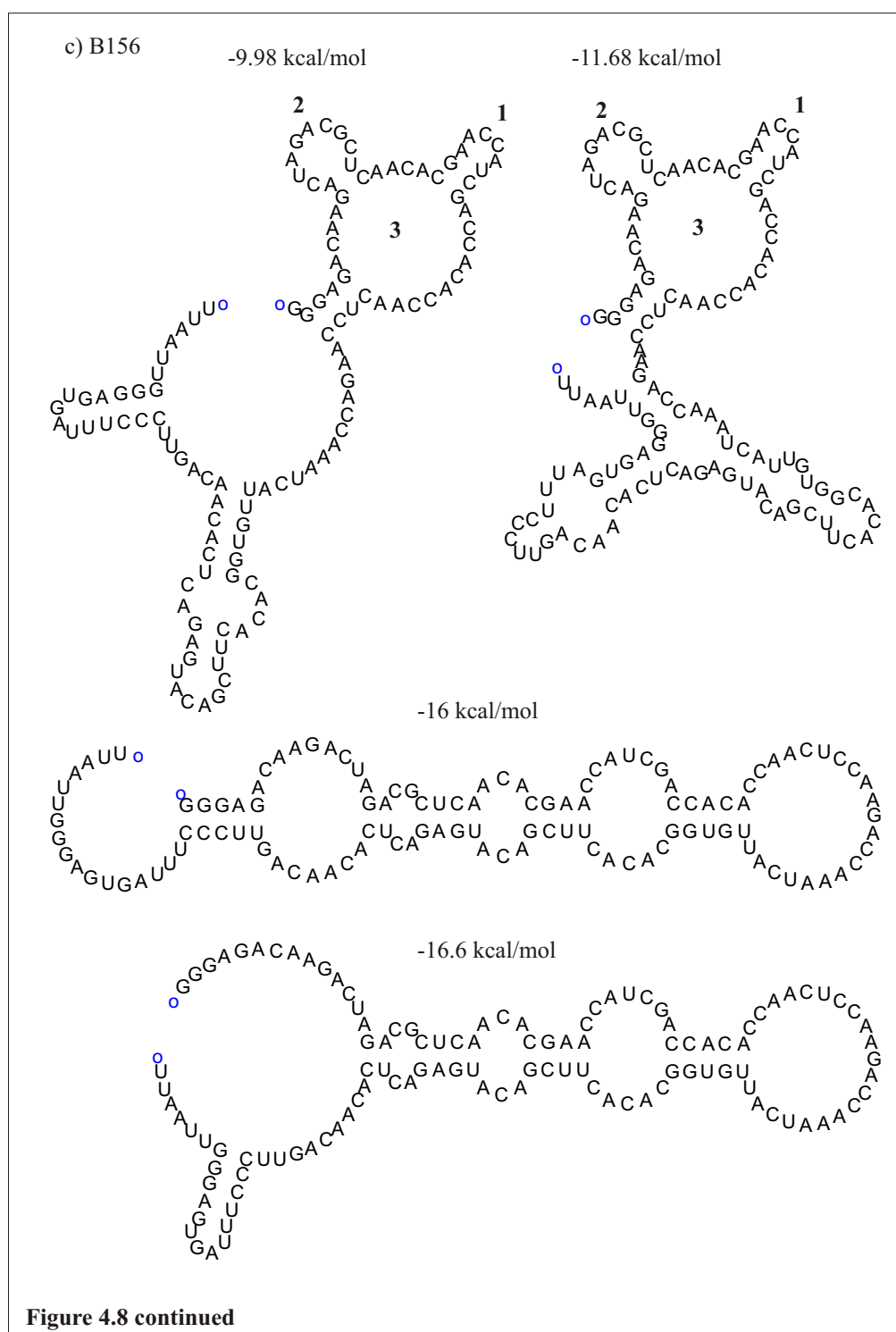
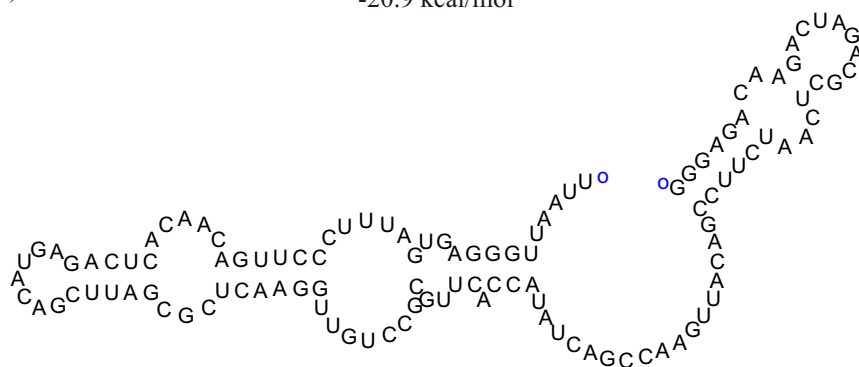


Figure 4.8 continued

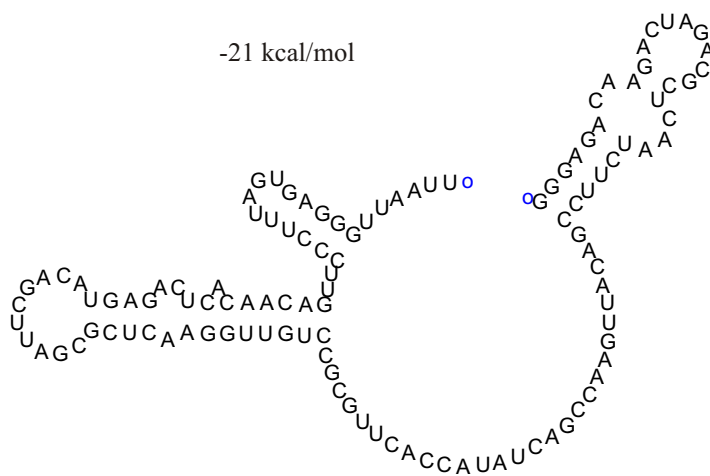


d) B161

-20.9 kcal/mol



-21 kcal/mol



-19.26 kcal/mol

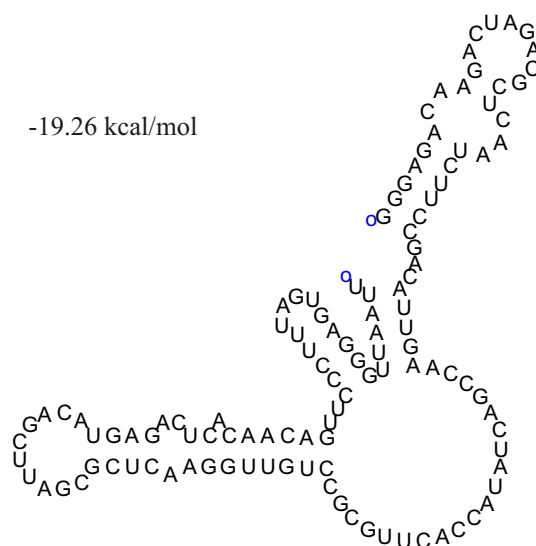
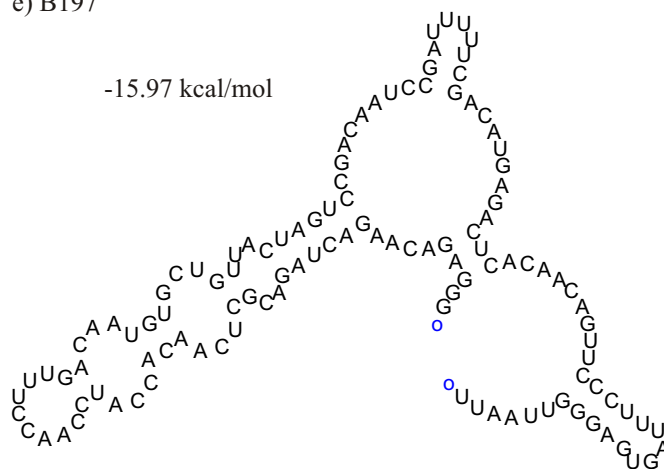


Figure 4.8 continued

e) B197

-15.97 kcal/mol



-19.3 kcal/mol

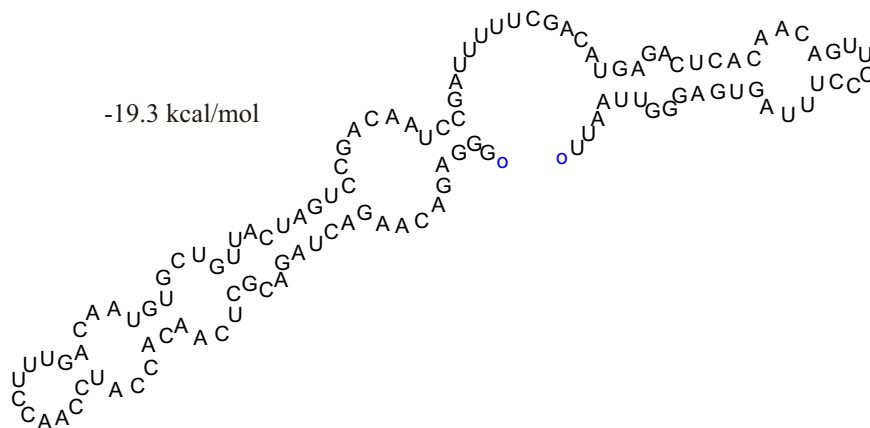
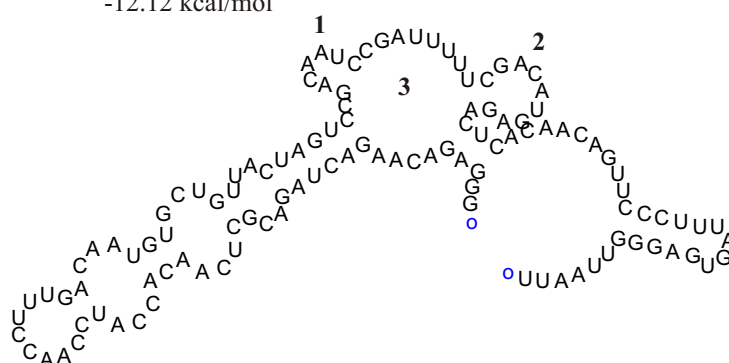
Constrained B40-like structure
-12.12 kcal/mol

Figure 4.8 continued

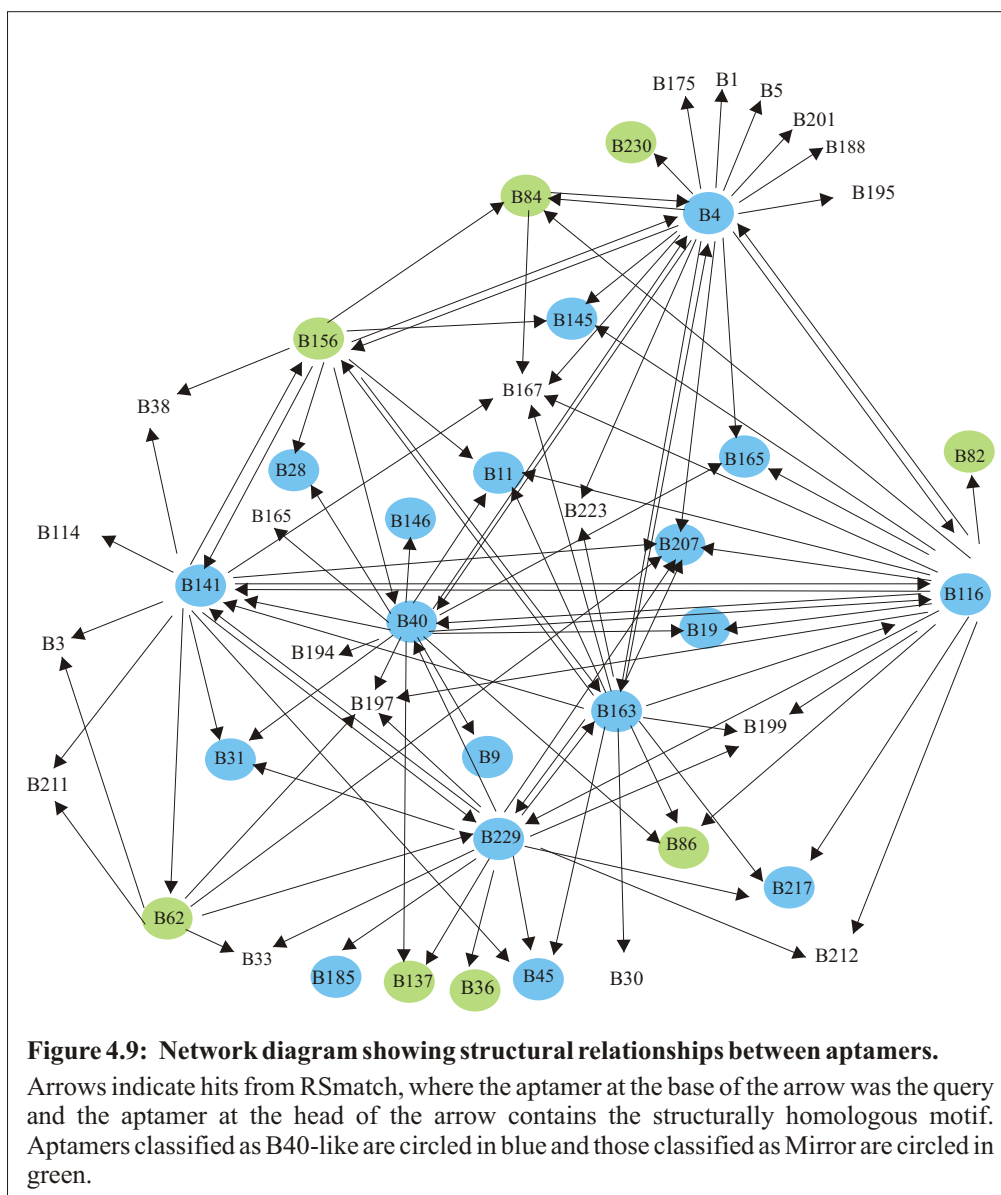
4.4.4 Network analysis

When analysing the RSmatch data I noticed that some aptamers seemed to be more structurally related than others. For example when B40 is used as the query structure, a homologous region in B4 is found, and the reciprocal structure in B40 is found when B4 is used as the query structure. However, in some cases this is not true: the B40 query does not find any structural homologues in B156 but when B156 is the query a homologous structure is found in B40. I chose to represent this information as a network diagram in order to find out if some groups of aptamers are more related to one another, and whether or not there are any distinct boundaries within the data set.

Figure 4.9 shows the network diagram in which each aptamer is represented as a node and hits are represented by arrows. B40-like and Mirror aptamers are shown in blue and green respectively. The network is biased towards B40-like aptamers as these were used most often as the query structures so will have more connections, and because aptamers that did not have any homologous motifs are not included (18 aptamers). This network provides support for the hypothesis that a number of the aptamers in this group are structurally related, because there are multiple links between multiple nodes. It also highlights the fact that when B4 was used as a probe many hits occurred that were not significant when considered in the structural context of the aptamer, i.e. there are many false positives. A group of aptamers that are in the Other category, aptamers B211, B3, B114 and B38, share homologous motifs with a limited number of probe aptamers, so it is possible that members of this group are structurally related through a non-B40-like motif.

4.4.5 Analysis of RNA structures within a population

RNA molecules are inherently flexible, and any given RNA sequence may exist in multiple structures in solution (as will be shown for aptamer B40 in



Chapter 5). Therefore I wanted to explore the likelihood of any particular B40-like or Mirror structure (the structures of interest) actually existing in solution. A crude method is to compare the energy of the structure of interest to the energy of the MFE structure. If this energy difference is large, then it is possible that the structure of interest may never form because it is not thermodynamically favourable.

Table 4.3 shows the energy difference between the B40-like or Mirror structure with that of the MFE structure for each aptamer, as predicted by Mfold. In seven cases, the B40-like structure is the same as the MFE structure, and therefore it is highly likely that this structure exists in solution as it is the most thermodynamically favourable structure. Other aptamers differ in the energies between these two structures by up to 8.82 kcal/mol. The greater the energy difference, the less likely it is that the B40-like or Mirror structure will exist. From these data I would predict that the B40-like or Mirror structures in B4, B31, B116, B137, B146 and B165 are present in these aptamers' structural populations.

A more robust approach to this problem was taken by analysing each aptamer sequence using Sfold (Ding *et al* 2004). Sfold is an RNA structure prediction program that identifies clusters of structures by sampling the Boltzmann distribution of the population of possible structures. The program calculates the probability of each cluster and finds structures that represent each cluster centroid and the ensemble centroid, and the MFE structure.

Sfold also produces a three-dimensional energy landscape representation of the differences between clusters. The x and y axes are measures of multidimensional scaling (MDS) (Kruskal and Wish 1977) and the z axis is the free energy of each structure. MDS is used to represent the distance between three-dimensional structures in two dimensions. The energy landscape provides a visualisation of how similar the clusters are to one

Aptamer	Shape	Energy of B40-like or Mirror structure (kcal/mol)	Energy of MFE structure (kcal/mol)	Energy difference (kcal/mol)
B1	Mirror	-13.54	-17.88	4.34
B4	B40-like	-23.97	-23.97	0
B9	B40-like	-13.57	-16.97	3.4
B11	B40-like	-19.49	-26.5	7.01
B19	B40-like	-16.9	-18.7	1.8
B28	B40-like	-23.31	-26.3	2.99
B31	B40-like	-18.33	-18.33	0
B33	B40-like	-11.89	-19.7	7.81
B40	B40-like	-27.3	-30	2.7
B45	B40-like	-16.06	-17.5	1.44
B62	Mirror	-15.8	-22.2	6.4
B82	Mirror	-15.4	-19.2	3.8
B84	Mirror	-18.1	-20.2	2.1
B86	Mirror	-20.08	-28.9	8.82
B116	B40-like	-19.2	-19.2	0
B137	Mirror	-22.77	-22.77	0
B141	B40-like	-19.16	-22.86	3.7
B145	B40-like	-22.57	-27.6	5.03
B146	B40-like	-22.04	-22.47	0.43
B156	Mirror	-12.78	-16.6	3.82
B163	B40-like	-19.27	-21.1	1.83
B165	B40-like	-26.28	-26.28	0
B167	B40-like	-24.41	-26.6	2.19
B170	B40-like	-12.03	-18.2	6.17
B185	B40-like	-21.23	-24.4	3.17
B217	B40-like	-23.62	-27.03	3.41
B229	B40-like	-21.28	-25	3.72
B230	Mirror	-17.33	-22.4	5.07

Table 4.3: Relative stability of each aptamer's B40-like or Mirror structure compared to the MFE structure.

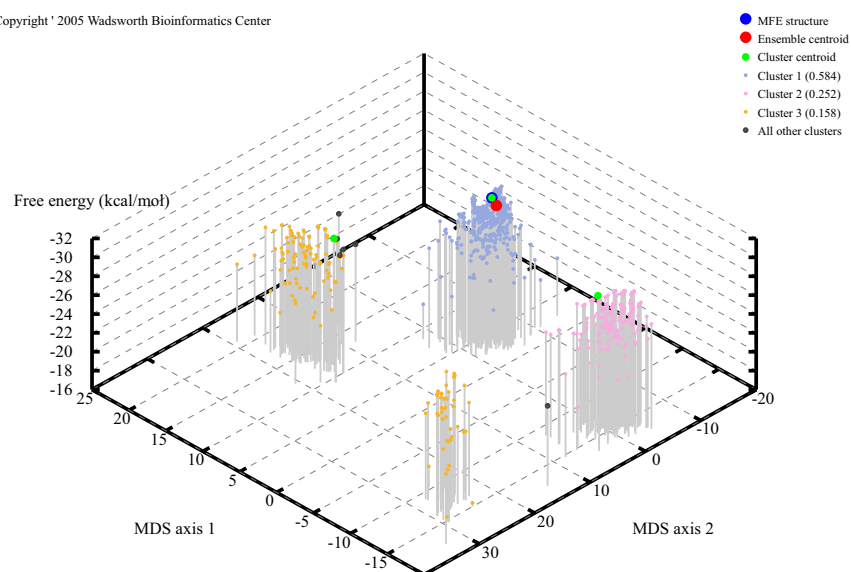
another, and how much energy would be required for an RNA structure to change between conformations in different clusters. To demonstrate this, the energy landscapes of B40 and B145 are given in Figure 4.10. B40 has three clearly delineated structural clusters whereas the three structural clusters of aptamer B145 are distributed over the free energy landscape. This correlates with the fact that Mfold, identifying structures whose energies are within 5 % of that of the MFE structure, predicts only three structures for B40 but nine structures for B145. This analysis predicts that while the structure of B40 that occurs in solution is fairly rigid once formed, B145 RNA molecules are more likely to flip-flop between different conformations.

Aptamer sequences were analysed using Sfold, and Table 4.4 summarises the number of clusters, the probability of the occurrence of each cluster, which clusters contain the MFE sequences, which contain the homologous motifs from RSmatch, and the energy of the MFE structure for each aptamer. In some cases, for example aptamers B1 and B86, the structure of interest only occurs in clusters that have a very low probability of occurring so it is unlikely that the structure of interest will exist in these cases. In contrast the opposite is true for several aptamers so it is likely that their structures of interest predominate in solution. From these results I would predict that the B40-like or Mirror structures in B4, B45, B84, B137, B165, B188, B207, B229 and B230 are significant in the structural populations of these aptamers.

In order to compare these two methods of analysis (energy comparison and Sfold), I plotted the energy difference between the MFE and the structure of interest versus the cumulative probability of the cluster(s) in which the structure of interest occurs (Figure 4.11). The significant negative correlation ($r = -0.3798$) indicates that these two methods are complementary. On the graph I have highlighted aptamers whose structure of interest have a particularly low or high probability of existing in solution, as well as those

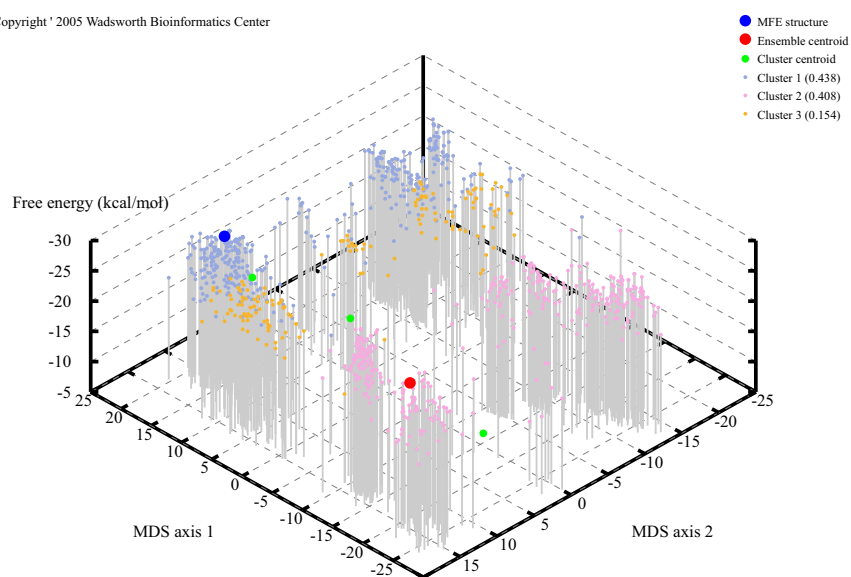
a) B40

Copyright © 2005 Wadsworth Bioinformatics Center



b) B145

Copyright © 2005 Wadsworth Bioinformatics Center

**Figure 4.10: Energy landscape maps generated by Sfold.**

a) Aptamer B40.

b) Aptamer B145.

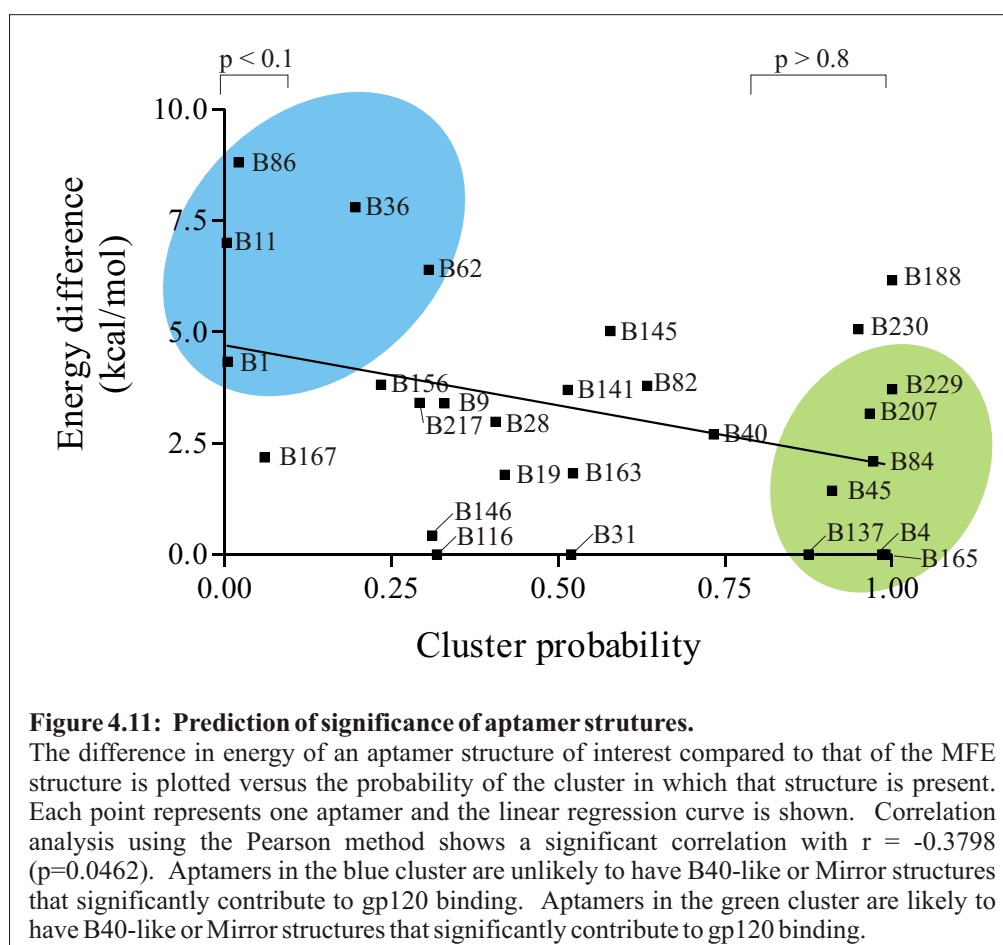
Structural clusters are represented by different colours (see individual legends) and the MFE structure, ensemble centroid and cluster centroids are indicated.

Aptamer	Number of clusters	Cluster probabilities	Free energy of MFE structure (kcal/mol)
B1	18	0.362* 0.356 0.189 0.026 0.020 0.018 0.008 0.007 0.003 [¶] 0.002 0.002 0.001 0.001 0.001 0.001 0.001 [¶] 0.001 0.001	-20.4
B4	6	0.776* [¶] 0.195 [¶] 0.015 [¶] 0.008 0.005 0.001	-26.1
B9	3	0.615* 0.329 0.056	-18.6
B11	6	0.454* 0.352 0.097 0.089 0.005 0.003 [¶]	-27.5
B19	2	0.580* 0.420 [¶]	-21.7
B28	2	0.594* 0.406 [¶]	-26.3
B31	3	0.519* [¶] 0.445 0.036	-21.1
B36	8	0.421* 0.297 0.106 [¶] 0.090 [¶] 0.065 0.015 0.003 0.003	
B40	3	0.593* [¶] 0.267 0.140 [¶]	-30.8
B45	2	0.910* [¶] 0.090	-20.0
B62	6	0.653* 0.289 [¶] 0.038 0.016 [¶] 0.003 0.001 [¶]	-22.2
B82	3	0.633 [¶] 0.219* 0.148	-19.2
B84	7	0.748* [¶] 0.181 [¶] 0.037 [¶] 0.025 0.005 [¶] 0.003 0.001 [¶]	-21.6
B86	4	0.835* 0.143 0.021 [¶] 0.001	-29.4
B116	2	0.682* 0.318 [¶]	-21.9
B137	2	0.875* [¶] 0.125	-24.2
B141	2	0.514 [¶] 0.486*	-25.6
B145	4	0.467* [¶] 0.392 0.111 [¶] 0.030	-28.2
B146	4	0.687 0.311* [¶] 0.001 0.001	-25.0
B156	2	0.766* 0.234 [¶]	-16.6
B163	2	0.522 [¶] 0.478*	-21.1
B165	6	0.840* [¶] 0.088 [¶] 0.030 [¶] 0.028 [¶] 0.010 0.004 [¶]	-27.1
B167	3	0.594* 0.346 0.060 [¶]	-26.6
B188	2	0.824* [¶] 0.176 [¶]	-31.1
B207	2	0.967* [¶] 0.033	-20.9
B217	4	0.613* 0.292 [¶] 0.092 0.003	-28.9
B229	2	0.948* [¶] 0.052 [¶]	-27.4
B230	2	0.950* [¶] 0.050	-23.4

Table 4.4: Number of structural clusters for B40-like and Mirror aptamers predicted by Sfold, the probability of the existence of each structural cluster and the energy of the MFE structure.

* marks the cluster in which the MFE structure is located.

[¶] marks clusters in which B40-like or Mirror structures are located.



that have both high probabilities and low energy differences (green), or low probabilities and high energy differences (blue).

Taking these data into account I would predict that structures of interest in B4, B45, B84, B137, B165, B207 and B229 make up a significant proportion of the structures within the population whereas the structure of interest of B1, B11, B36, B62 and B86 are unlikely to exist in solution and therefore are probably not relevant to gp120 binding.

4.4.6 Alignment of loop sequences

The sequences of the homologous loops found by RSmatch were aligned using ClustalW (Thompson *et al* 1994) in order to address whether there is localised sequence homology between structurally related aptamers. The loop sequence was defined as the sequence of the unpaired bases at the end of a stem loop, plus the last base pair in the stem. Where one aptamer contained more than one homologous loop these were included separately (loop 1a, loop 1b etc) in the alignment.

The ClustalW alignment is shown in Figure 4.12a. Some regions of local sequence homology can be identified by eye, for example several loops contain the CAC motif from the original hypothesis. However there is no clear delineation of these groupings, for example the CAC motif then becomes a CAU motif, then a CACC motif down the alignment. To represent this more clearly a phylogenetic tree was drawn (Figure 4.12b). There are no significantly distinct groups of aptamer loops present within the phylogenetic tree. Bootstrap values represent the probability (from 1-1000) that a branch occurs when the phylogenetic tree is drawn multiple times, and values greater than 700 tend to be considered significant. This tree has very few Bootstrap values greater than 700 implying that there are no significant sequence relationships in the loop sequences.

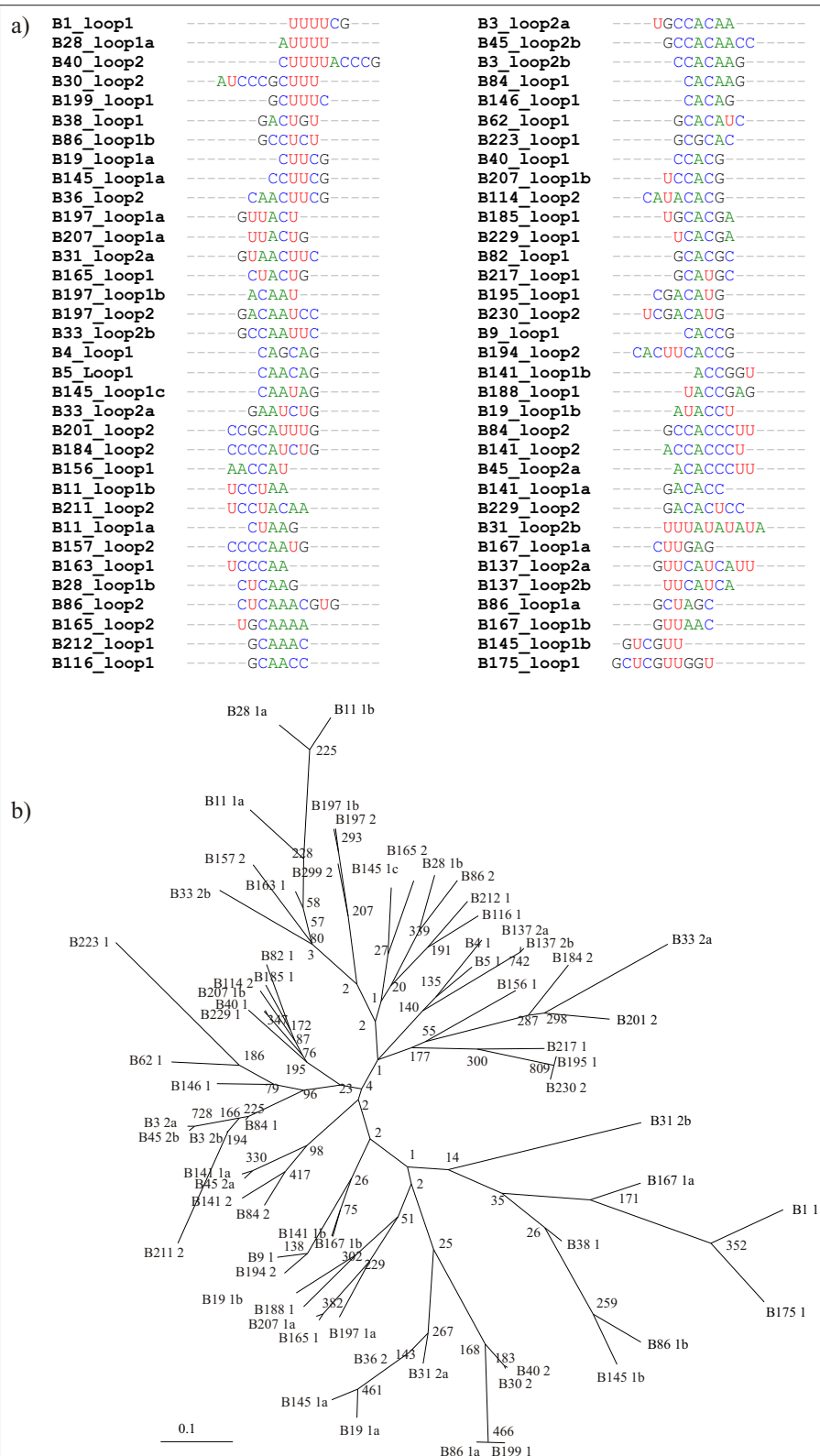
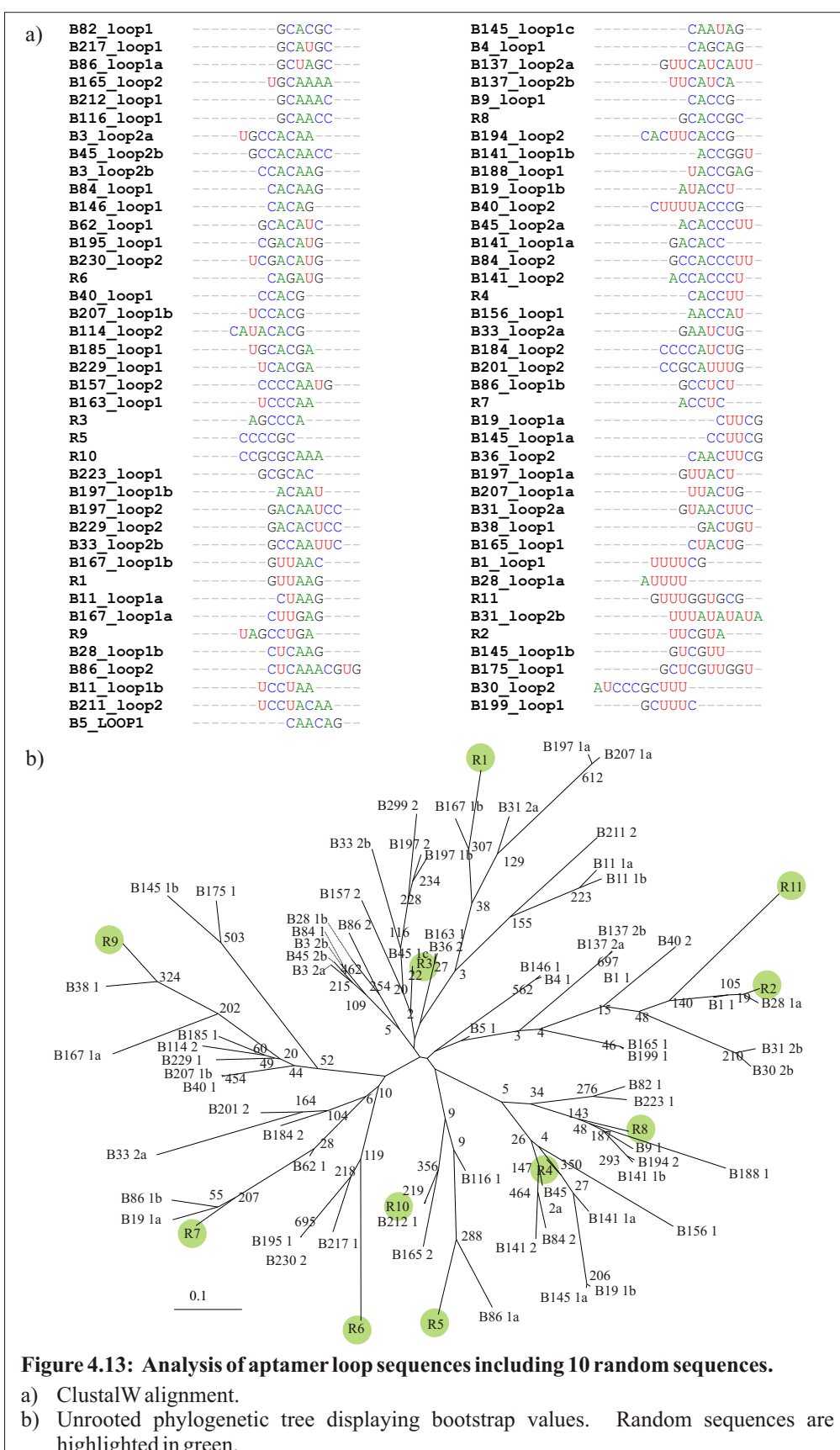


Figure 4.12: Analysis of aptamer loop sequences.

- a) ClustalW alignment.
b) Unrooted phylogenetic tree displaying bootstrap values.



There was no clustering of loop 1 and loop 2 homologues, leading me to the conclusion that it is unlikely that two distinct types of loops are present in the B40-like structures. This is further supported by the fact that there is a continuous range of loop lengths from 5-10 bases, rather than a group of loops with 3 bases (as seen for B40 loop 1) and a group of loops with 10 bases (like B40 loop 2). The predominant loop length was 4 bases.

In order to confirm this the ClustalW alignment was repeated and a new phylogenetic tree was drawn, this time including 11 random RNA sequences from 5-10 bases in length (sequences R1-11, Figure 4.13a). If the aptamer loop sequences were related to one other then they would cluster together and the random sequences would be phylogenetically distinct. However the ClustalW alignment and phylogenetic tree (Figure 4.13) show that the random sequences are inserted between aptamer loop sequences, indicating that in fact there are no significant regions of sequence homology within the aptamer loops.

4.5 Discussion

I have described the cloning and sequencing of a set of new aptamers from the SELEX pool produced by Makobetsa Khati. The binding of all these aptamers to gp120 was tested using SPR, and all of the aptamers bound as expected.

Computational methods were chosen to look for structural similarities between the aptamers. Several aptamers contain motifs that are homologous to the loops in the branched motif of aptamer B40, supporting the original hypothesis that there are common motifs in different aptamers. The use of a network diagram has provided further support for the hypothesis that a number of these aptamers are structurally related. The probability of B40-like

or Mirror structures existing in solution was estimated using Sfold. From this analysis I predict that aptamers B4, B45, B84, B137, B165, B207 and B299 form B40-like structures that significantly contribute to gp120 binding. Therefore it is likely that these aptamers also bind to the same site on gp120 as aptamer B40.

ClustalW and phylogenetic analysis showed that there was no support for the hypothesis that there are small regions of sequence similarity between the homologous loops. This could be further explored by mapping the nucleotides required for gp120 binding of putative B40-like aptamers, and by mapping their gp120 epitopes (see Chapter 7).

There are a number of caveats to be considered when employing computational methods, both in general and specifically to these experiments. It has been suggested that RNA structure prediction programs based on thermodynamic principles are not always accurate, and that their results should be interpreted with caution. For example, it has been estimated that Mfold correctly predicts 73 % of base pairs within an RNA structure (Mathews *et al* 1999). A further complication in this specific case is that the aptamers are synthesised with 2'-fluoro-pyrimidines, which may affect the aptamer structure. A fluorine atom is a much stronger hydrogen-acceptor than the wild-type hydroxyl group so this modification may affect the bond-forming and bond-breaking energies which are used by Mfold to predict the RNA structure. However it has been shown that Mfold correctly predicts the structure of aptamer B40 (Dey *et al* 2005a) so it is clear that the 2'-fluoro modification does not render the RNA structure completely different from the equivalent 2'-hydroxyl RNA structure. Other experiments in this Chapter used the RNA structure prediction program Sfold, and although this uses statistical methods rather than thermodynamic principles to predict RNA structures, the results from the two methods were generally similar.

All bioinformatics programs to date have been designed with the intention of analysing genes. There are key differences between genes and aptamer sequences that leave most bioinformatics programs ill-equipped to analyse aptamer sequences satisfactorily. Many programs work on the assumption that the sequences being compared are related through a common ancestor, having evolved divergently. In the case of aptamers there is no common ancestor, and the SELEX method is a process of convergent rather than divergent evolution. Not only that, but the aptamer sequences also contain constant regions of 100 % sequence homology. RSmatch was chosen because it puts very little emphasis on evolutionary components (unlike other similar alignment programs that perform an initial sequence alignment and then a subsequent structural alignment). I found that the best way to overcome the shortfalls was to analyse both the full-length and the variable-only aptamer sequences with RSmatch.

These conditions meant that instead of the bioinformatics experiments being performed in an unbiased statistical manner, a lot of human intervention was required in order to filter and rationalise the results. In some cases this may help to strengthen the data, for example by eye it is possible to see where a particular structural motif occurs within the whole aptamer structure. However it may also lead to bias and the inclusion of more false-positives and false-negatives than when a bioinformatics program is run unsupervised using genetic data as intended.

Having highlighted these caveats, it would be an interesting exercise in bioinformatics to design a structural alignment program that was suited to aptamer sequences. It would also be interesting to modify RSmatch to use Sfold for the RNA folding component, to see whether or not differences would then occur in the structural alignment.

The results from any bioinformatics analysis should be followed up with wet-lab experiments to complement the *in silico* data. I had planned to use the RSmatch results to design competition assays to ascertain whether or not aptamers that seemed to have B40-like structures also compete with one another for binding to gp120, and whether there may be other groups of aptamers that bind to different regions of gp120. Unfortunately, despite several optimisation steps, I was unable to reach such a conclusion. The development and implementation of the competition assay is described in Chapter 7.

Chapter 5:

Aptamer minimisation

- Stabilisation of the branched form of B40t77
- Shortening the stem
- Removal of AA mismatch
- Stabilisation of the branched form of truncated aptamers
- Introduction of 2'-o-dimethylallyl groups

Chapter 5: Aptamer minimisation

5.1 Introduction

Anti-gp120 aptamers have the potential to be used as prophylactic microbicides to prevent HIV-1 infection. For this to be realised, the cost of aptamer production needs to be reduced, and one way in which this can be done is by shortening the aptamer. Full-length 117-nucleotide aptamers are synthesised by *in vitro* transcription with 2'-fluoro-pyrimidines, which is time-consuming, expensive and impractical on a large scale. Reduction of the aptamer length to fewer than 60 nucleotides would enable their chemical synthesis with the same 2'-fluoro-pyrimidine chemistry (Sproat 1993). This method would be cheaper, could be scaled up to produce the milligram quantities required for clinical use, and would also allow the inclusion of additional RNA modifications that are incompatible with *in vitro* transcription using T7 polymerase.

The secondary structure of aptamer B40 has been studied extensively using theoretical modelling, enzymatic digestion, footprinting and mutagenesis to delineate which nucleotides are required for binding to gp120 (Dey *et al* 2005a). It is likely that two alternative conformations of B40, a branched structure and a linear structure, exist in equilibrium in solution, and that the branched structure contains the functional motif that binds to gp120. Figure 4.8 shows the alternative structures and also highlights the footprinted region that is required for gp120 binding. A truncated aptamer, B40t77, which constitutes only nucleotides 1-77 of B40, was shown to have comparable gp120 binding and HIV-1 neutralisation activity as full-length B40 (Dey *et al* 2005a).

The initial aims of this study were two-fold; firstly to characterise mutations in B40t77 that stabilise the branched form, and secondly to design a truncated aptamer that could be chemically synthesised and that would retain gp120-binding activity. I then went on to combine these two aspects and produce synthetic aptamers of fewer than 60 nucleotides in length that are stabilised to induce the branched conformation. Finally I investigated the introduction of 2'-O-dimethylallyl modifications on bases within helices that would further stabilise that aptamer structure.

5.2 Methods

5.2.1 Design of synthetic aptamers

Brian Sproat of RNA-Tec, Belgium, provided the chemically synthesised aptamers described below. Mfold was used to predict the structures and free energies of the aptamers. It is not known exactly how the introduction of 2'-O-dimethylallyl-modified bases would affect the energy of an RNA structure so the energies of these aptamers were not estimated.

5.2.2 Aptamer binding assay

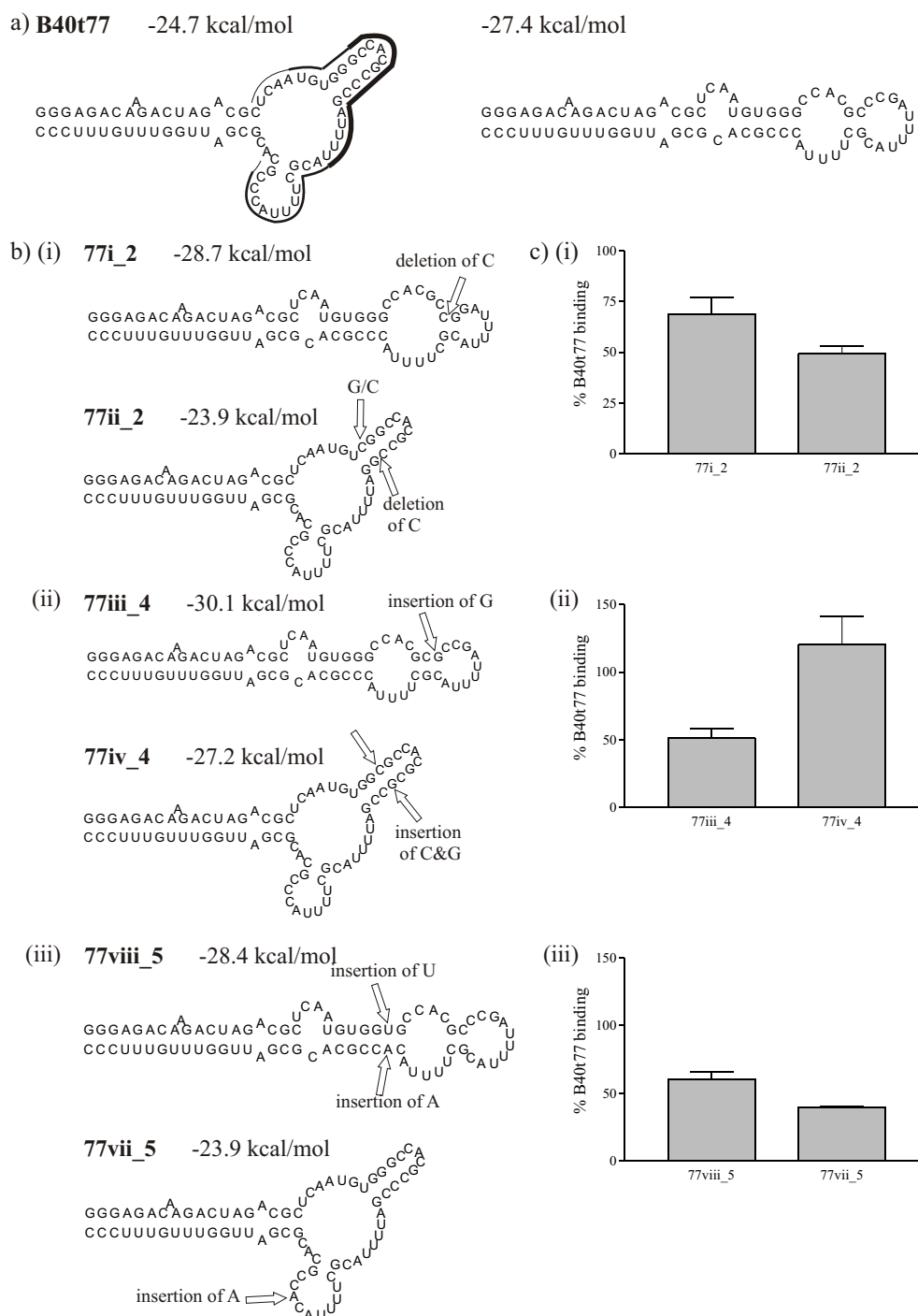
SPR was used to analyse the interaction of aptamers with gp120 using the same method as described in Chapter 4 (Section 4.3). Briefly, gp120 was immobilised on flow cell 1 of a CM5 chip and flow cell 2 was left blank. Aptamers were injected and the binding response at 600 seconds post injection was recorded, at which point the aptamer was dissociating at a constant rate. Each aptamer was injected in triplicate and the sequence of injections was randomised. Binding responses were normalised for the molecular mass of the aptamer and are expressed as a percentage of B40t77 binding.

5.3 Stabilisation of the branched form of B40t77

Three pairs of aptamers based on B40t77 were designed to investigate stabilisation of the branched and linear conformations. The first aptamer of each pair contained one or more mutations that are predicted to stabilise the linear form, which was expected to decrease aptamer binding to gp120. The second aptamer contained a compensating mutation that was expected to stabilise the branched form and restore gp120 binding.

Mfold was used to predict which mutations would stabilise the linear or branched conformations and to design the aptamers. The corresponding DNA oligomers were obtained from Sigma and amplified by PCR (see Section 2.2.1). The PCR products were ethanol precipitated and used as templates for *in vitro* transcription reactions with 2'-fluoro-pyrimidines (Section 2.2.2). The transcription products were phenol:chloroform extracted, ethanol precipitated and refolded in CHBS buffer at 1 μ M (see Sections 2.2.6, 2.2.7 and 2.2.13) then analysed for gp120 binding using SPR (see Section 5.2.2). Figure 5.1 shows the structures of the three pairs of aptamer mutants and the relative responses observed on gp120 binding expressed as a percentage of B40t77 binding. The two alternative structures of B40t77 and the footprinted region, that constitutes the minimal region required for gp120 binding, are also shown for reference.

The structures and gp120 binding of the first pair, aptamers 77i_2 and 77ii_2, is shown in Figure 5.1(i). Aptamer 77i_2 contains a deletion of residue C37, which is predicted to stabilise the linear conformation and this aptamer showed a decrease in gp120 binding as expected. Mutation of residue 27 from G to C in aptamer 77ii_2 was predicted to restore the branched structure, but this aptamer showed even less binding to gp120 than 77i_2. This may be

**Figure 5.1:**

- (a) Alternative structures and free energies of B40t77 as predicted by Mfold. The footprinted region is indicated by a solid line.
- (b) Structures of three pairs of aptamers (i, ii and iii) based on B40t77 designed with mutations that stabilise either the branched or linear conformation. Structures and their free energies are shown as predicted by Mfold.
- (c) Aptamer gp120-binding responses of aptamer pairs (i, ii and iii) observed 600 seconds after the start of injection expressed as a percentage of B40t77 binding and normalised for aptamer size. Error bars represent the standard deviation of three replicates.

because residue G27 is in the footprinted region, so this mutation may disrupt the interaction with gp120 and eliminate any increase in binding that would be achieved due to stabilisation of the branched form.

The structures and gp120 binding of the second pair, aptamers 77iii_4 and 77iv_4 is shown in Figure 5.1(ii). Aptamer 77iii_4 contains an inserted G after residue C35 that is predicted to stabilise the linear conformation, and this aptamer did show decreased binding to gp120 binding. Insertion of a compensating C residue after G28 in aptamer 77iv_4 was predicted to stabilise the branched conformation, and 77iv_4 binding was restored to wild-type levels, as expected. The insertion of this extra GC base-pair in stem 3 does not affect the binding to gp120 even though it is in the footprinted region.

The structures and gp120 binding of the third pair, aptamers 77viii_5 and 77vii_5, is shown in Figure 5.1(iii). Insertion of a U after nucleotide G27 and an A after nucleotide C53 in aptamer 77viii_5 are predicted to stabilise the linear conformation. A decrease in binding to gp120 was observed for this aptamer as expected. Aptamer 77vii_5 only has the inserted A, and this was predicted to stabilise the branched conformation and consequently increase binding. However 77vii_5 showed even less binding to gp120 than 77viii_5, which is probably due to the modification of nucleotides within the footprinted region.

While any number of mutant pairs in which binding was not restored by the introduction of a compensating mutation may be explained by disruption within the footprinted region, one successful reversion supports the hypothesis that the branched structure preferentially binds to gp120.

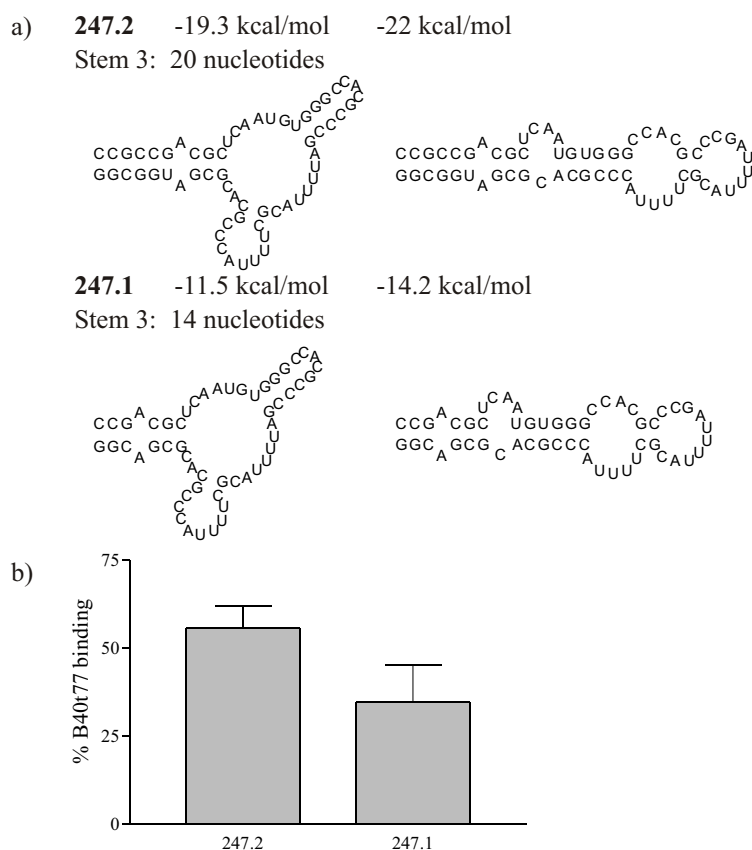
5.4 *Shortening the stem*

I designed synthetic aptamers to investigate how much the long stem of B40t77 (stem 3, Figure 4.5) could be truncated without decreasing its ability to bind to gp120. I hypothesised that removal of nucleotides in stem 3 would not affect binding of the aptamers to gp120 as it is not within the footprinted region. Synthetic aptamers 247.2 and 247.1 were designed with shortened stems of 20 and 14 nucleotides respectively, compared to B40t77 which has 37 nucleotides in stem 3. These aptamers were predicted to form both the linear and branched structures. The aptamer structures and binding to gp120 are shown in Figure 5.2.

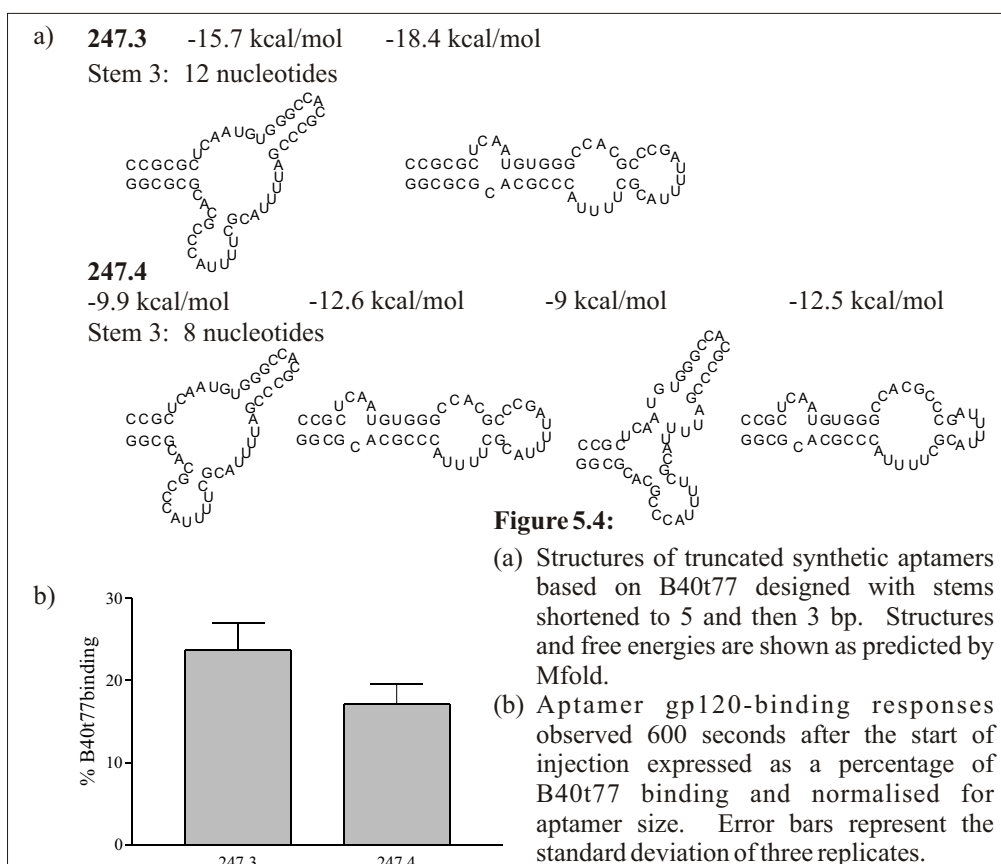
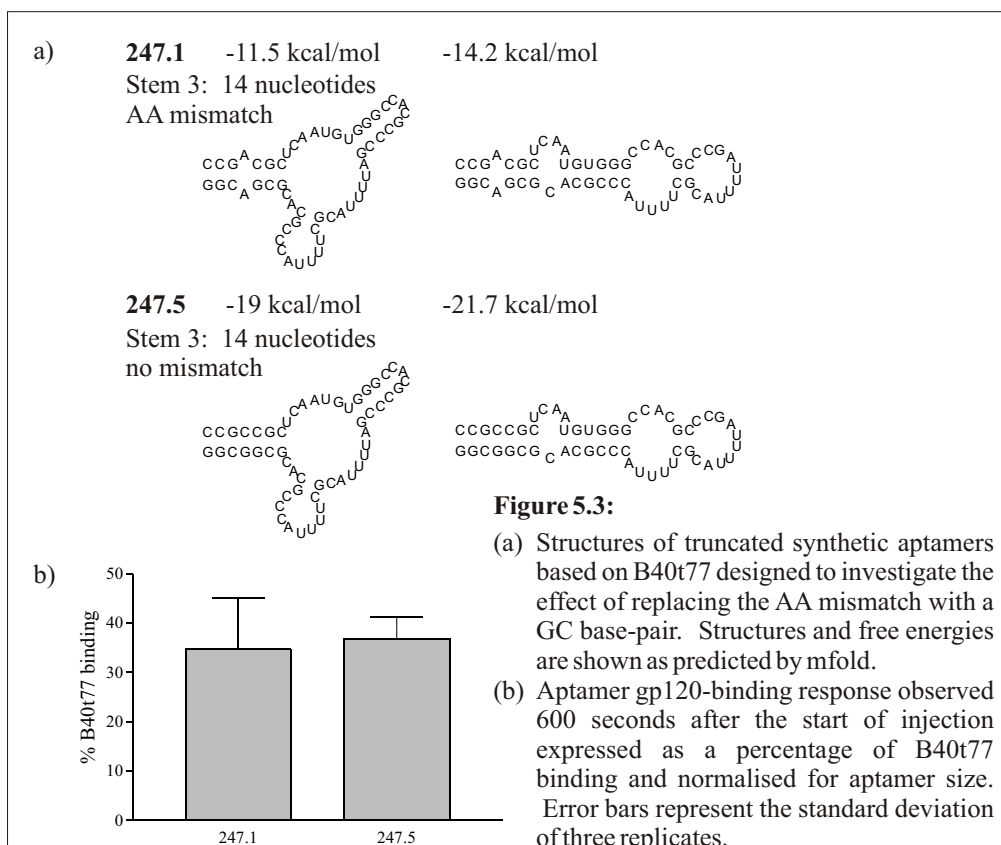
This experiment showed that decreasing the length of stem 3 also decreased the gp120 binding activity of each aptamer. It is possible that the longer stem 3 in B40t77 stabilises the overall aptamer structure, and that shortening it destabilises the overall aptamer structure and decreases binding, despite the fact that nucleotides in the footprinted region are not directly affected.

5.5 *Removal of AA mismatch*

Stem 3 of B40t77 contains an AA mismatch. I hypothesised that replacement of the AA mismatch with a GC base-pair would stabilise the stem and therefore increase the overall aptamer stability. This aptamer was predicted to fold into both the linear and branched conformations. Figure 5.3 compares the binding of aptamers 247.1, that has 14 nucleotides in stem 3 and an AA mismatch, to aptamer 247.5, that has 14 nucleotides in stem 3 but a GC base-pair in place of the AA mismatch. Surprisingly these aptamers bound comparably to gp120, indicating either that the stability of stem 3 was not significantly improved by removal of the AA mismatch, or that increased stability of stem 3 did not affect the stability of the whole aptamer structure.

**Figure 5.2:**

- (a) Structures of truncated synthetic aptamers based on B40t77 designed with truncations in stem 3. Structures and their free energies are shown as predicted by Mfold.
- (b) Aptamer gp120-binding responses observed 600 seconds after the start of injection expressed as a percentage of B40t77 binding and normalised for aptamer size. Error bars represent the standard deviation of three replicates.



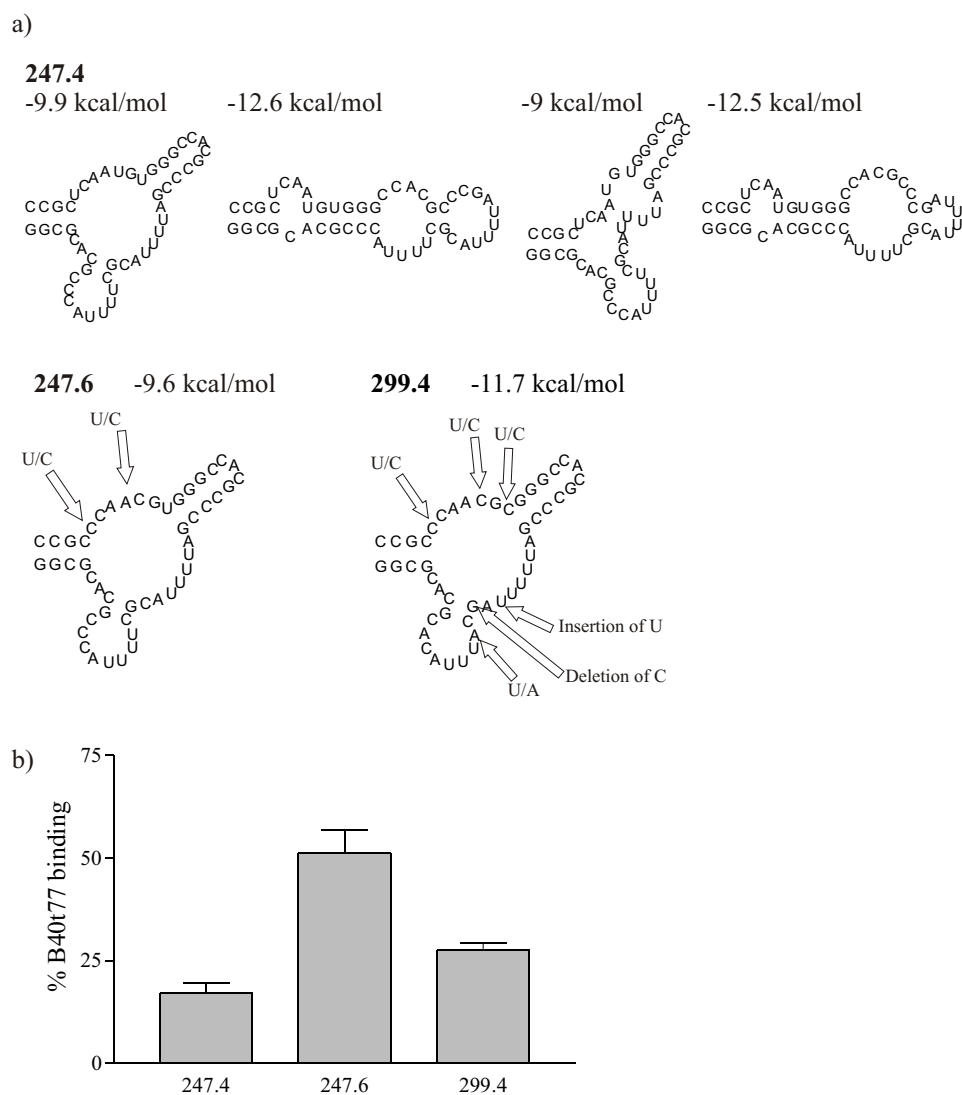
5.6 *Shortening stem with removed AA mismatch*

Despite the previous result two more synthetic aptamers were designed that had 12 and 8 nucleotides in stem 3 (aptamers 247.3 and 247.4, respectively) without the AA mismatch. As expected these aptamers showed an even greater decrease in binding to gp120 (Figure 5.4). Mfold predicts that aptamer 247.4 will fold into four different conformations, only one of which is expected to bind to gp120. It is probable that the short stems in these two aptamers are not sufficient to stabilise the branched conformation that is required for gp120 binding.

5.7 *Stabilisation of branched form of truncated aptamers*

Following these disappointing results using aptamers with shortened stems, I decided to investigate whether or not mutations that stabilise the branched form could be used to increase binding of truncated aptamers to gp120. Aptamer 247.6 was designed with 8 nucleotides in stem 3 and mutations that were predicted to increase the probability of the formation of the branched structure. Aptamer 299.4 was designed with a CG base pair at the base of stem 1 in place of the UG base pair to stabilise the helix. As this U/C substitution was predicted to disrupt the branched structure, several other mutations were also introduced to stabilise the branched form: insertion of a U after U28, deletion of C30 and substitution of U 33 to A, and the two U to A substitutions from 247.6. Aptamer structures and gp120-binding responses are shown in Figure 5.5.

Encouragingly, aptamer 247.6 showed more binding to gp120 than the equivalent aptamer 247.4 that has 8 nucleotides in stem 3 but no stabilising

**Figure 5.5:**

- (a) Structures of truncated synthetic aptamers based on B40t77 with 3 bp in stem 3 and mutations that stabilise the branched form (247.6) or to allow substitution of a UA base-pair with a more stable GC base-pair, compared to the equivalent aptamer 247.4 that has no mutations. Structures and free energies are shown as predicted by Mfold.
- (b) Aptamer gp120-binding responses observed 600 seconds after the start of injection expressed as a percentage of B40t77 binding and normalised for aptamer size. Error bars represent the standard deviation of three replicates.

mutations. Aptamer 299.4 showed a smaller increase in binding which is probably because this aptamer contained six mutations, all of which were in the footprinted region. These individual mutations disrupted gp120 binding despite the fact that the branched form was stabilised.

5.8 Introduction of 2'-O-dimethylallyl groups

RNA helices can be stabilised by the inclusion of 2'-O-dimethylallyl modifications in tandem internal base pairs (Iribarren *et al* 1990; Lamond and Sproat 1993; B. Sproat, personal communication). The structure of 2'-O-dimethylallyl-cytosine triphosphate is shown in Figure 5.6a. Aptamer 265.1 was designed with 12 nucleotides in stem 3, which includes three 2'-O-dimethylallyl-modified tandem base pairs (in the case of cytosine triphosphate this modification replaces the previous 2'-fluoro modification). Compared to the equivalent unmodified aptamer (aptamer 247.3), aptamer 265.1 showed a very large increase in binding to gp120. In fact aptamer 265.1 showed a greater than 3-fold increase in binding to gp120 compared to unmodified B40t77 (Figure 5.6).

Next I chose to investigate the effect of introducing tandem pairs of 2'-O-dimethylallyl-modified bases in three different positions in stem 1 (aptamers 265.2, 265.3 and 265.4). The 2'-O-dimethylallyl-modified base pairs are predicted to stabilise stem 1 resulting in the formation of the branched structure alone. These aptamers all bound to gp120 to approximately the same degree as B40t77 (Figure 5.7), which is a considerable increase in binding compared to the equivalent unmodified aptamer, 247.3. However binding levels were still lower than that of 265.1, which is probably because only one 2'-O-dimethylallyl-modified base pair was introduced to each of these aptamers compared to the introduction of three pairs of modified bases in aptamer 265.1. The modified bases are in the footprinted region but apparently do not disrupt the gp120-aptamer interaction or such an increase in

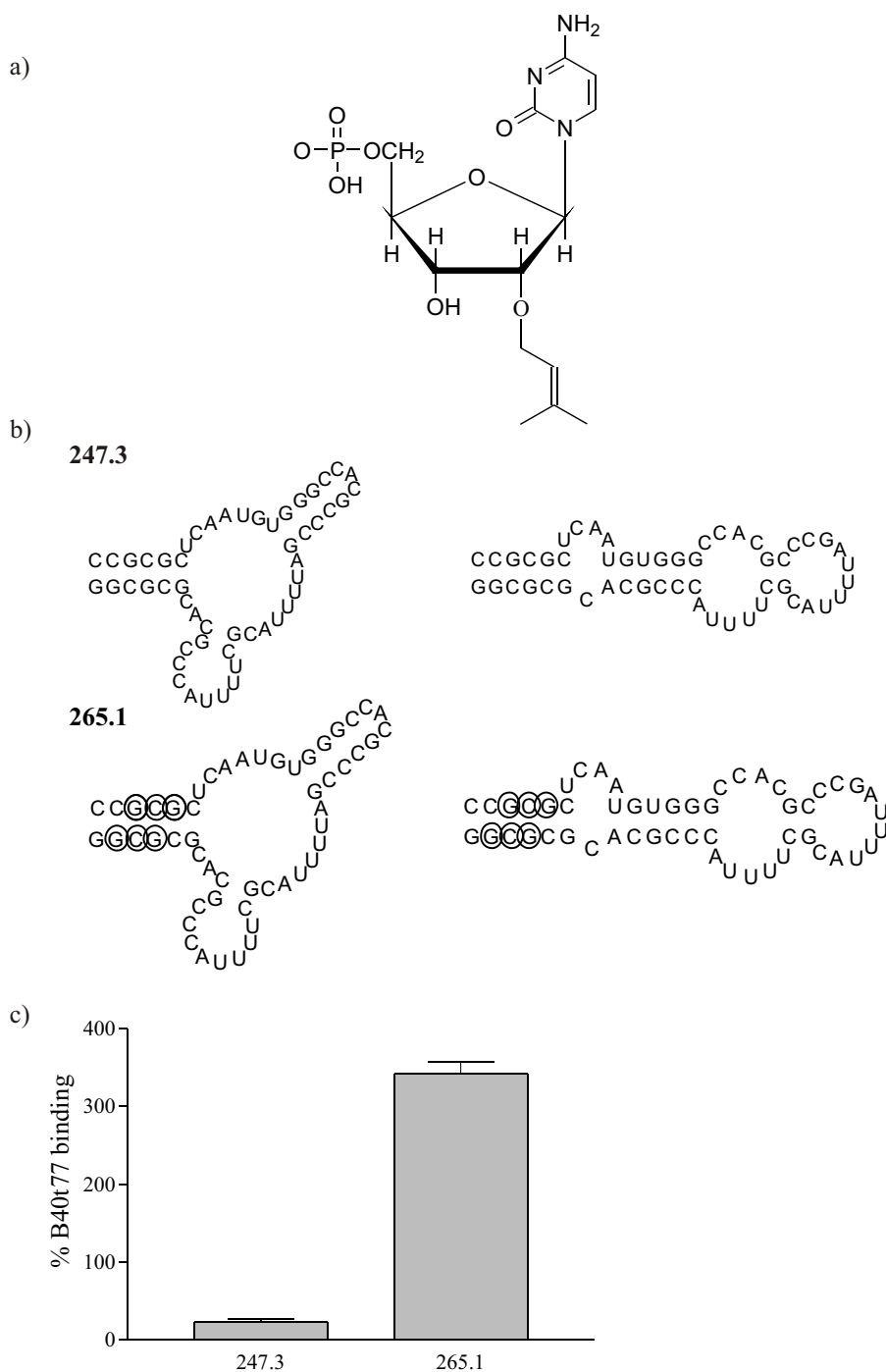
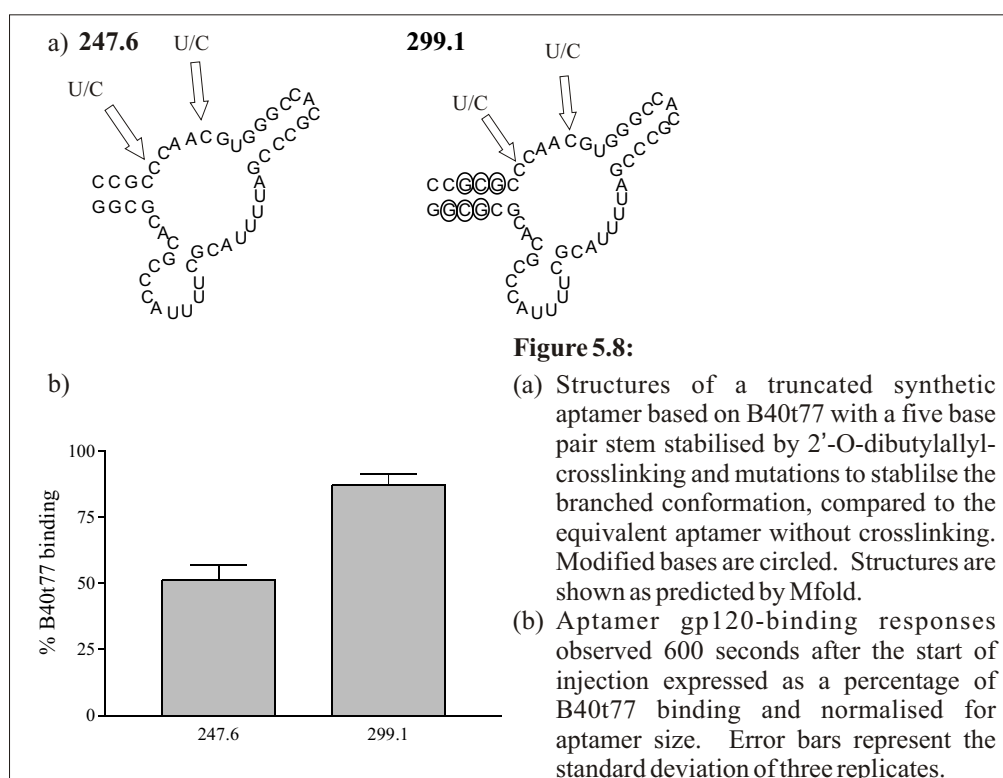
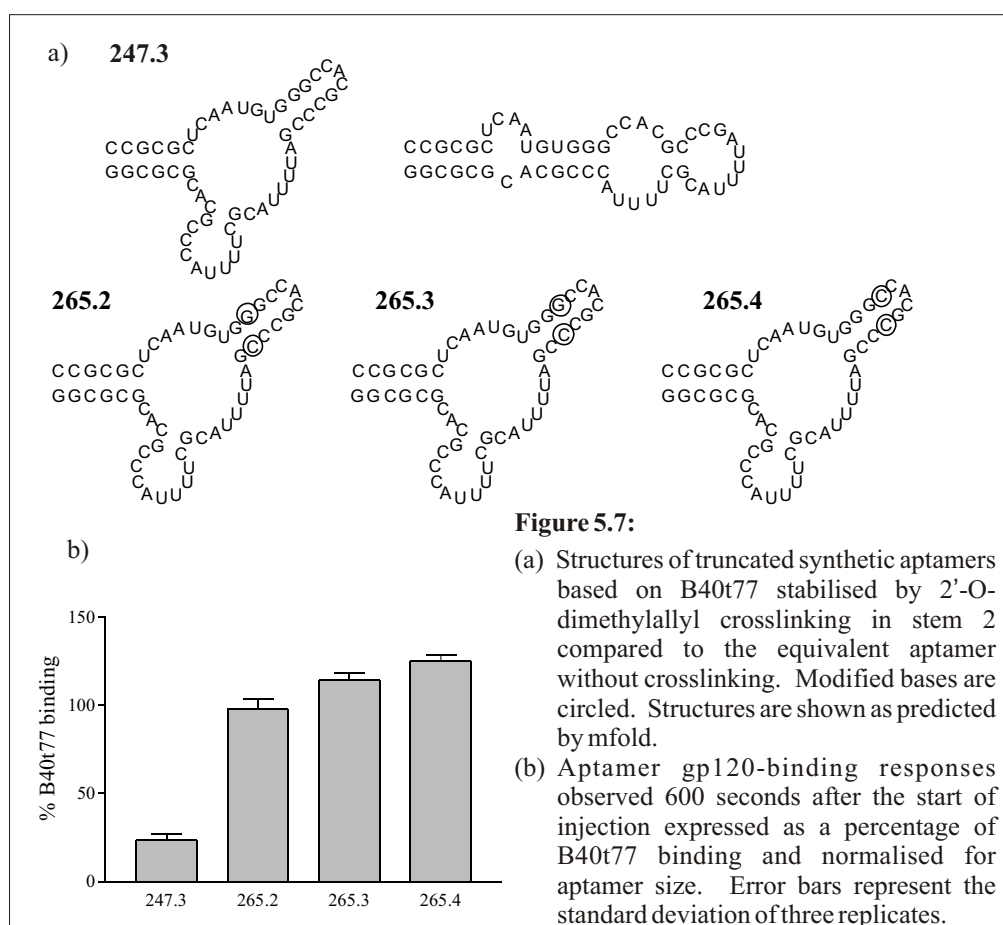


Figure 5.6:

- Chemical structure of 2'-O-dimethylallyl-cytosine triphosphate.
- Structures of a truncated synthetic aptamer based on B40t77 with a 5 bp in stem 3 that is stabilised by dimethylallyl crosslinking, compared to the equivalent aptamer without crosslinking. Modified bases are circled. Structures are shown as predicted by Mfold.
- Aptamer gp120-binding responses observed 600 seconds after the start of injection expressed as a percentage of B40t77 binding and normalised for aptamer size. Error bars represent the standard deviation of three replicates.



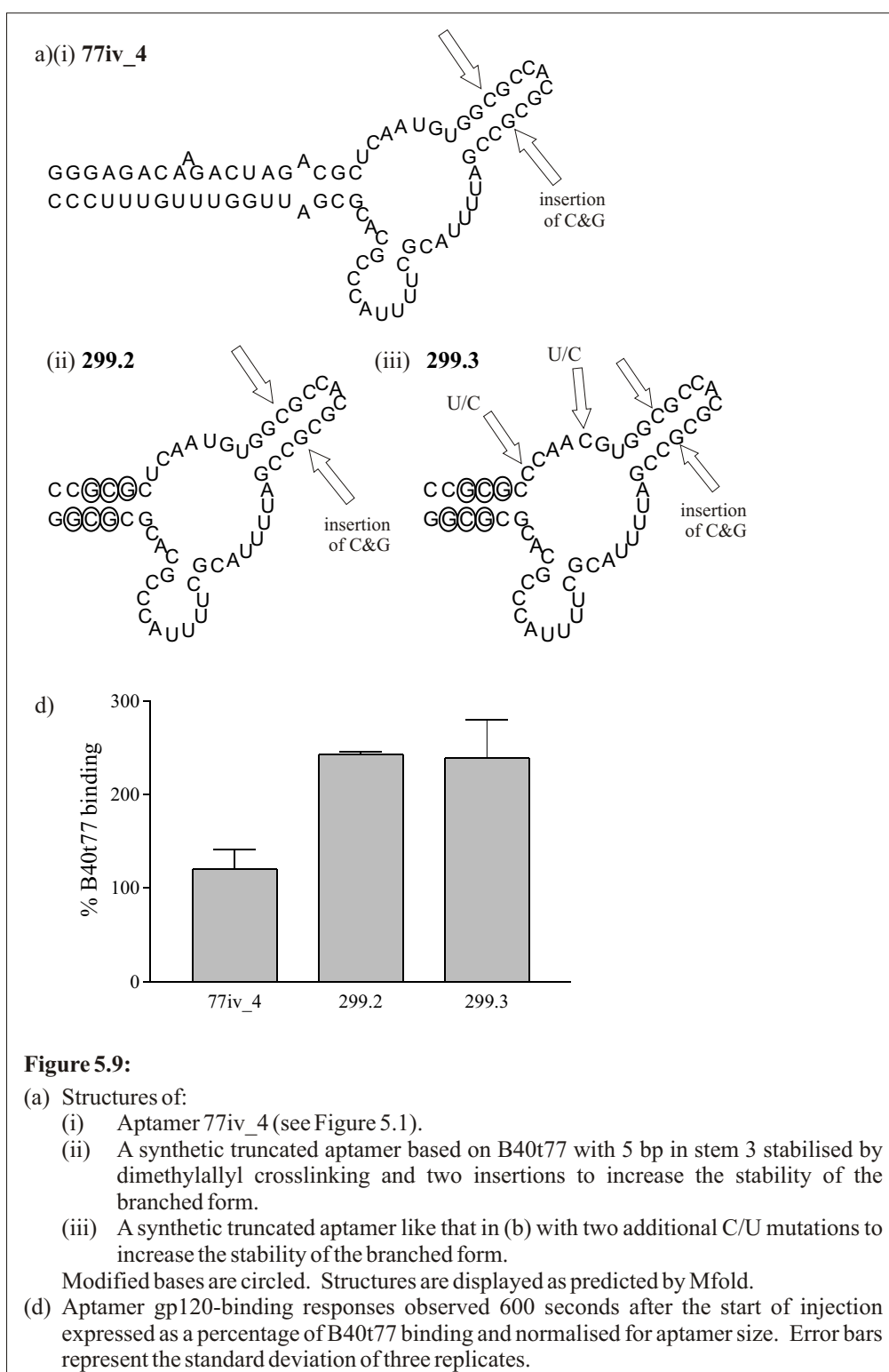
binding compared to the unmodified aptamer would not have been seen. This is probably because the 2'-O-dimethylallyl groups lie within the helix, so do not interact directly with gp120.

5.9 Combination of 2'-O-dimethylallyl modifications and stabilisation of the branched form

Having found that the introduction of 2'-O-dimethylallyl modified bases successfully stabilised stem 3 and correspondingly increased gp120 binding, I went on to introduce a similar modification into an aptamer that also contained mutations to increase the stability of the branched form (aptamer 299.1, based on aptamer 247.6). At this time 2'-O-dimethylallyl-CTP and 2'-O-dimethylallyl-GTP were unavailable through RNA-Tec so aptamer 299.1 was synthesised with 2'-O-dibutylallyl-modified bases instead. Introduction of 2'-O-dibutylallyl-modified bases should increase helix stability in the same way as 2'-O-dimethylallyl-modified bases (Sproat and Lamond 1993).

Aptamer 299.1 showed increased binding compared to aptamer 247.6 as expected (Figure 5.8), but the effect of introducing the 2'-O-dibutylallyl-modified bases was not as substantial as the effect of introducing a similar modification to an aptamer with no other stabilising mutations (compare with aptamers 247.3 and 265.1, Figure 5.6). Whether this difference is due to the aptamer mutations or the difference in 2'-O-dimethylallyl and 2'-O-dibutylallyl modifications is uncertain.

I designed aptamer 299.2 with three pairs of 2'-O-dimethylallyl-modified bases in the truncated stem 3 and the C and G insertions from aptamer 77iv_4 that are predicted to increase the stability of the branched form. I also designed a derivative of aptamer 299.2, aptamer 299.3, which additionally



contained the two U/C substitutions from aptamer 247.6 to further increase stability of the branched form. Both of these aptamers showed improved binding to gp120 compared to B40t77 and aptamer 77iv_4 although there was little difference between aptamers 299.2 and 299.3 (Figure 5.9).

5.10 Discussion

I have designed a panel of aptamers derived from B40t77 that contain various modifications to increase the binding to gp120 while decreasing the aptamer size. The aptamer modifications, consequences for aptamer structure and relative gp120 binding are summarised in Table 5.1. Using mutagenesis I confirmed that the branched form of B40t77 binds preferentially to gp120. Shortening stem 3 had detrimental effects on aptamer binding to gp120, and while this was not affected by the removal of the AA mismatch, binding was restored when stem 3 was stabilised with three pairs of 2'-O-dimethylallyl-modified bases. I investigated various combinations of truncations, mutations and modifications and found that aptamer 265.1 showed the greatest binding to gp120. Aptamer 265.1 is 52 bases long with 12 nucleotides in stem 3, which is stabilised by three pairs of 2'-O-dimethylallyl-modified bases and there are no other mutations in the aptamer.

I observed that there was a more modest increase in gp120 binding when single 2'-O-dimethylallyl-modified base pairs were introduced in stem 1 (aptamers 265.2, 265.3 and 265.4) than when three 2'-O-dimethylallyl-modified base pairs were introduced in stem 3 (aptamer 265.1). I believe that this is due to the reduced number of cross links rather than disruption of the gp120-binding by altering the footprinted region. Given that both mutations and 2'-O-dimethylallyl modifications in stem 1 had less of a detrimental effect on aptamer binding to gp120 than comparable modifications in other parts of the footprinted region, it is possible that stem 1 is not as crucial for aptamer

Aptamer	Length (ntds)	Features	Outcome	%B40t77 binding
77i_2	77	Deletion of C	Stabilise linear form	68
77ii_2	77	Deletion of C, G/C substitution	Stabilise branched form	49
77iii_4	78	Insertion of G	Stabilise linear form	51
77iv_4	79	Insertion of G and C	Stabilise branched form	120
77viii_5	78	Insertion of U and A	Stabilise linear form	61
77vii_5	79	Insertion of A	Stabilise branched form	40
247.2	60	Stem 3: 20 nucleotides	Truncated	56
247.1	54	Stem 3: 14 nucleotides	Truncated	35
247.5	54	Stem 3: 14 nucleotides Remove AA mismatch	Truncated Increased stability	37
247.3	52	Stem 3: 12 nucleotides Remove AA mismatch	Truncated Increased stability	23
247.4	48	Stem 3: 8 nucleotides Remove AA mismatch	Truncated Increased stability	17
247.6	48	Stem 3: 8 nucleotides Remove AA mismatch 2 x U/C substitution	Truncated Increased stability Stabilise branched form	51
299.4	48	Stem 3: 8 nucleotides Remove AA mismatch Substitute CG bp for UA bp	Truncated Increased stability	28
265.1	52	Stem 3: 12 nucleotides dma stabilisation in stem 3	Truncated Increased stability	342
265.2	52	Stem 3: 12 nucleotides dma stabilisation in stem 1	Truncated Increased stability	98
265.3	52	Stem 3: 12 nucleotides dma stabilisation in stem 1	Truncated Increased stability	114
265.4	52	Stem 3: 12 nucleotides dma stabilisation in stem 1	Truncated Increased stability	125
299.1	52	Stem 3: 12 nucleotides dma stabilisation in stem 3 2 x U/C substitutions	Truncated Increased stability Stabilise branched form	87
299.2	54	Stem 3: 12 nucleotides dma stabilisation in stem 3 Insertion of G and C	Truncated Increased stability Stabilise branched form	243
299.3	54	Stem 3: 12 nucleotides dma stabilisation in stem 3 Insertion of G and C 2 x U/C substations	Truncated Increased stability Stabilise branched form	239

Table 5.1: Summary of data relating to B40t77 aptamer derivatives.

binding to gp120 as was previously thought (Dey *et al* 2005a). In collaboration with Peter Kwong, we are currently attempting to crystallise the gp120-aptamer complex, which will confirm which bases in the aptamer directly contact gp120.

Having seen that the introduction of 2'-O-dimethylallyl-modified base pairs increases aptamer stability and binding to gp120 I would like to design aptamers that contain these modifications in both stem 1 and stem 3, as I believe these modified aptamers would show even greater binding to gp120.

It was interesting to find that aptamer 299.1, which has 2'-O-dimethylallyl-modified base pairs in stem 3 and U/C substitutions to stabilise the branched form, did not bind to gp120 as well as aptamer 265.1, that contained the 2'-O-dimethylallyl-modified base pairs in stem 3 alone (compare Figures 5.6 and 5.8). I believe that this indicates that the U/C substitutions may affect bases that directly contact gp120. On their own, these substitutions stabilise the branched form of gp120 and increase aptamer binding (compare aptamers 247.4 and 247.6, Figure 5.5). However, the presence of three 2'-O-dimethylallyl-modified base pairs in stem 3 overrides any stabilisation from the U/C substitutions, so these mutations instead become detrimental to aptamer binding.

The results presented here describe the binding of aptamers to monomeric HIV-1 gp120 from strain BaL. I also tested these aptamers for binding to the gp120-CB fusion protein that was expressed in 293T cells (see Chapter 3). Overall there were no significant differences compared to the binding of these aptamers to gp120, and the results are shown in Appendix 1. Gp120 on the surface of HIV-1 exists as a trimeric spike so it would be interesting to see if similar levels of aptamer binding are seen to recombinant trimeric gp120.

The strongest gp120-binding aptamer B40 derivative, aptamer 265.1, will be taken forward for development as a candidate microbicide. The next key experiment is to verify that this aptamer will neutralise HIV-1 in PBMC to the same extent as B40 and B40t77. Experiments are also currently underway to investigate the stability of this aptamer over a 24-hour period in conditions such as human plasma, and porcine vaginal and rectal swabs. In the future we plan to collaborate with groups at UCLA to test aptamer neutralisation of HIV-1 in human tissue explant models and, if these are successful, the aptamer may be tested in animal models and later in clinical trials.

Chapter 6: Definition of the aptamer binding site on gp120

- Optimisation of methods
- Interaction of aptamer with gp120 Mutants
- Model of the aptamer binding site

Chapter 6: Definition of the aptamer binding site on gp120

6.1 Introduction

A number of previous studies have used panels of gp120 alanine-scanning mutants to map the epitopes of antibodies on gp120 (Scanlan *et al* 2002; Pantophlet *et al* 2003; Darbha *et al* 2004). I decided to use a similar approach to determine which gp120 residues are important for the binding of aptamer B40 to gp120. I was able to map the aptamer binding site on gp120 to four key gp120 residues and show that aptamer binding is dependent on the gp120 variable loops. These results indicated that it is likely that the binding site of B40 overlaps with the CCR5 binding site, which is consistent with previous work (Dey *et al* 2005b).

I chose to analyse the interaction of an aptamer derived from B40 (see Section 6.2.1) with a number of gp120 mutants in order characterise the aptamer binding site on gp120. I used the results of previous work (Dey *et al* 2005b) to design the mutagenesis experiment. The mutants fell into two categories: deletion mutants and alanine-scanning mutants. Mutants in which variable loops V1, V1 and V2 or V3 were deleted from gp120 ($\Delta V1$, $\Delta V1/V2$ and $\Delta V3$ respectively) were chosen because these loops have key roles in the binding of gp120 to the CCR5 coreceptor and various CD4i antibodies (Thali *et al* 1993; Wyatt *et al* 1995; Kwong *et al* 1998; Rizzuto *et al* 1998; Cormier and Dragic 2002). Alanine mutants were chosen that are important for the binding of gp120 to CCR5 (Rizzuto *et al* 1998; Rizzuto and Sodroski 2000; Cormier *et al* 2001; Cormier and Dragic 2002) and to the CD4i antibodies 17b (Thali *et al* 1993; Rizzuto *et al* 1998) and X5 (Darbha *et al* 2004). Table 6.1

Mutant ^a	Domain ^b	Epitope ^c	Conservation ^d
K121A	C1	CCR5 bs, 17b bs	B
K207A	C2	CCR5 bs, 17b bs, 48d bs X5 bs	B
N295A	C2	Glycosylation	
T297S	V3		
R298A	V3	CCR5 bs, X5 bs	
P299A	V3		
N301	V3	CCR5 bs	
N302	V3	CCR5 bs	
P313A	V3	CCR5 bs	
A329K	V3		
N339A	C3	Glycosylation	
D368A	C3	CD4 bs	A
R419A	C4	CCR5 bs, 17b bs	A
I420A	C4	CCR5 bs, 17b bs, 48d bs, X5 bs	B
K421A	C4	CCR5 bs, 17b bs, 48d bs	B
Q422A	C4	CCR5 bs, 17b bs, X5 bs	A
I423A	C4	17b bs, X5 bs	
W427A	C4	CD4 bs	B
K432A	C4	X5 bs	
Y435A	C4	17b bs, 48d bs, X5 bs	B
ΔV1	V1		
ΔV1/V2	V1/V2		
ΔV3	V3		

Table 6.1: Alanine scanning and deletion mutants used to map the aptamer binding site on gp120 (adapted from Pantophlet *et al* 2003).

^a Mutations in HIV-1 JR-CSF gp120. Amino acid numbering is relative to that of the HIV-1 isolate HXB2 where 1 is the initial methionine (26). Mutants in **bold** are those which result in a significant loss of binding to aptamer 299.5.

^b Domains in gp120: V, variable loop; C, constant region.

^c Epitopes affected by amino acid mutations; bs, binding site.

^d Conservation is defined as in Kwong *et al* 1998: A, conserved amongst all primary HIV-1 isolates; B, conserved amongst all primate immunodeficiency viruses.

summarises the gp120 mutants used in this study, indicating the location of each mutant within gp120, importance in binding to CCR5, 17b or X5, and conservation amongst HIV-1 and other primate immunodeficiency viruses. Residues A329, T297 and P299 have not previously been suggested to be important for CCR5 or CD4i antibody binding but were included because of their location in the V3 loop. Residues W427 and D368 are in the CD4 binding site (Kwong *et al* 1998), and residues N295 and N339 are important glycosylation sites (Scanlan *et al* 2002). Mutants of these four residues were included as controls that were not expected to affect aptamer binding.

I initially used an SPR method to investigate whether or not aptamer 299.5 would bind to the gp120 mutants. This method proved unsuccessful, so I developed an ELISA-type assay that is similar to the methods employed by other groups to investigate antibody binding sites on gp120. After some optimisation, this technique allowed me to determine the EC₅₀ values of aptamer 299.5 binding to each gp120 mutant, and with this information I was able to map the binding site of aptamer 299.5 on gp120.

6.2 Methods

6.2.1 Design of a synthetic biotinylated aptamer B40 derivative

The methods in this Chapter require the aptamer to be biotinylated so that it can be detected with HRP-conjugated streptavidin or immobilised on a streptavidin-coated plate or chip. Rather than use full-length aptamer B40, in this study I chose to use a synthetic aptamer based on aptamer 247.6 (Figure 5.5). Use of aptamer B40 would have required biotinylation of the aptamer with terminal transferase to conjugate biotinylated ddUTP to the 5' end, which would have involved an impractical amount of work considering the quantity of material required and the inefficiency and cost of the biotinylation reaction

(see Chapter 7). Chapter 5 describes the design of a number of synthetic aptamers derived from B40 that contain the branched motif encompassing nucleotides thought to be important for the binding of B40 to gp120. Aptamer 247.6 is 48 nucleotides in length, has 8 nucleotides in stem 3 with no AA mismatch, and contains two U/C substitutions that are predicted to stabilise the branched conformation. I designed a variant of aptamer 247.6, aptamer 299.5, which has an additional biotinylated uridine residue at the 5' end of the RNA molecule. This modification does not affect binding of the aptamer to gp120 (data not shown). Use of aptamer 299.5 relies on the assumption that the binding sites of B40 and 299.5 on gp120 are identical, but I believe this is the case because 299.5 is based on the nucleotides of B40 that constitute the minimal binding region.

6.2.2 Production of gp120 mutants

HIV-1 pseudoviruses are generated by the transfection of 293T cells with two plasmids, pNL4.3Luc.R⁻E⁻ and pSVIII-Env (Connor *et al* 1995, see Section 2.6). pNL4.3Luc.R⁻E⁻ encodes the genome of HIV-1 strain NL4.3, but has inactivating frameshift mutations in *rev* and *env* and a gene for luciferase in place of *nef* (Connor *et al* 1995). The pSVIII-Env plasmid (Helseth *et al* 1990) encodes the *env* and *rev* genes. The resulting pseudovirions are capable of transduction of susceptible cell lines (see Chapter 8), but here the pseudovirions were lysed with detergent to release gp120 (see Section 2.6.1). Mutants of gp120 were produced by performing the transfection with pSVIII-Env plasmids encoding alanine-scanning or deletion mutants of JR-CSF gp120, kindly provided by R. Pantophlet (Zwick *et al* 2001; Scanlan *et al* 2002; Pantophlet *et al* 2003). I anticipated that the analysis of aptamer 299.5 binding to gp120 from the HIV-1 R5 strain JR-CSF would be representative of binding to gp120 from other R5 strains, because aptamer B40 was shown to neutralise diverse strains of HIV-1 (Khati *et al* 2003).

6.3 Analysis of gp120-aptamer interaction by SPR

I designed an SPR assay to investigate aptamer binding to gp120 mutants. Human CD4 and aptamer 299.5 were immobilised on a sensor chip and samples containing either wild-type or mutant gp120 were injected over the surface. I anticipated that most gp120 mutations would not affect binding to CD4 (except for the control mutations D368A and W427A), so the amount of binding to CD4 would be proportional to the gp120 concentration in the sample. The response upon binding to aptamer 299.5 could therefore be normalised relative to the CD4-binding response.

Sensorgrams of the immobilisation procedure are shown in Figure 6.1a and b. Approximately 7 000 RU hCD4 (flow cell 3) and 8 000 RU streptavidin (flow cells 1 and 2) were immobilised on a CM5 chip using amine coupling (Section 2.10.1). Two injections of streptavidin were necessary to achieve the required level of binding. 35 μ L biotinylated aptamer 299.5 (at 2 μ M following refolding in CHBS) were injected on flow cell 1 to immobilise 1 000 RU. 35 μ L biotinylated polyclonal anti-gp120 antibody (40 μ g/mL in CHBS) were injected on flow cell 2 but no binding was seen. The injection was repeated using a 10-fold higher concentration of antibody but there was still no binding, so the flow cell was disregarded in the remainder of the experiment. Flow cells 1 and 2 were blocked with a 5 μ L injection of biotin (180 μ g/mL in CHBS) and flow cell 4 was left blank.

As a positive control, 35 μ L of the gp120-CB fusion protein (36 μ g/mL in CHBS) were injected over all four flow cells (Figure 6.1c). As expected, the gp120-CB bound to hCD4 and aptamer 299.5 but not to the blank flow cell. The surface of the chip was regenerated with a 5 μ L injection of glycine-HCl pH 2.5 followed by a 5 μ L injection of 7 M Urea. The flow rate was then

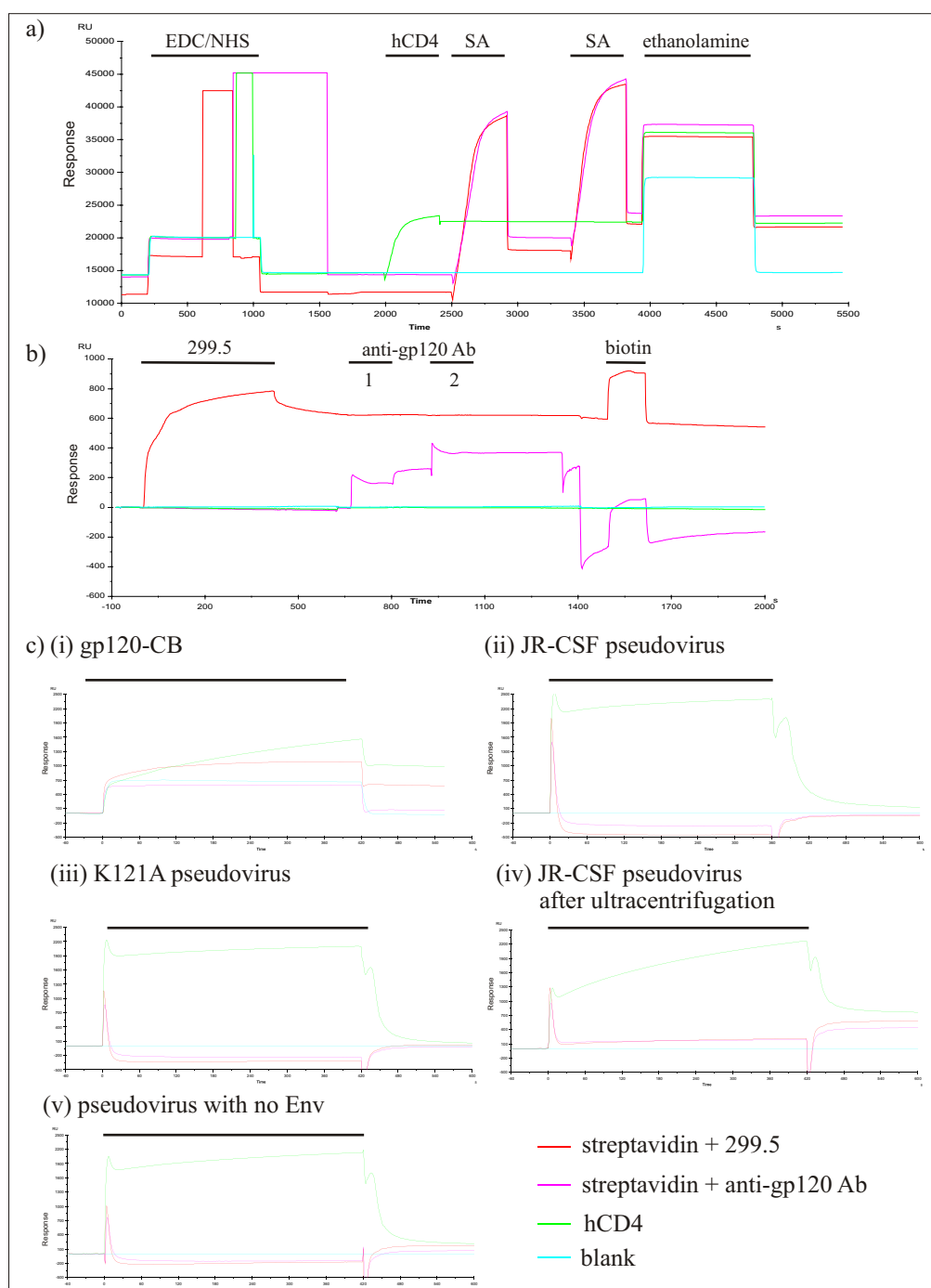


Figure 6.1: Use of SPR to analyse gp120-aptamer binding.

- Sensorgram showing immobilisation of streptavidin and hCD4 by amine coupling.
- Sensorgram showing immobilisation of biotinylated aptamer 299.5 and biotinylated polyclonal anti-gp120 antibody to streptavidin, followed by blocking with biotin.
- Sensorgrams showing injections of:
 - gp120-CB.
 - JR-CSF gp120 from pseudovirus supernatant.
 - K121A gp120 from pseudovirus supernatant.
 - JR-CSF gp120 from pseudovirus supernatant following ultracentrifugation.
 - Supernatant from pseudovirus containing no Env.

The solid lines above each sensorgram indicate the duration of each injection.

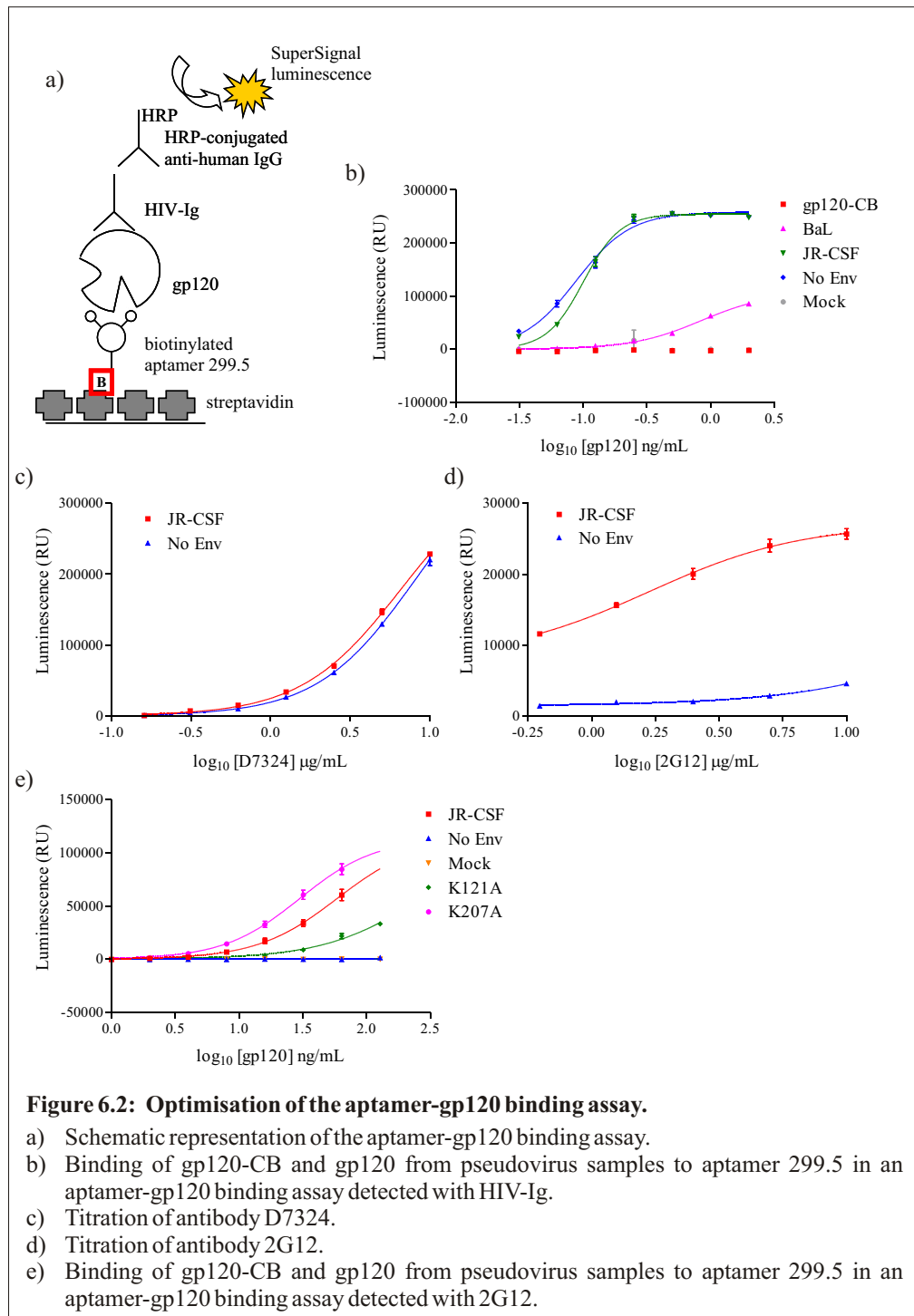
increased to 100 $\mu\text{L}/\text{min}$ for 10 min in order to ensure that both regeneration agents had been completely removed.

Pseudovirus made with wild-type JR-CSF Env, K121A JR-CSF Env or no Env was produced by transfection of 293T cells (Section 2.6.1). Supernatant was harvested, pseudovirions were lysed by the addition of 1 % Empigen, and 35 μL were injected over all four flow cells (Figure 6.1c). In one instance the wild-type JR-CSF sample was subjected to ultracentrifugation at $100\,000 \times g$ for 1 hour at 4 °C and the pellet was resuspended in 50 μL PBS prior to injection. In all cases injection of the sample resulted in a large response in flow cell 3 (hCD4) and no response in the other flow cells. The response in flow cell 3 rapidly deteriorated at the end of the injection indicating that it did not represent true binding. At this stage it became clear that this method was not suitable for analysing aptamer interactions with pseudovirus-derived gp120 mutants so I decided to investigate a different approach.

6.4 Analysis of aptamer-gp120 interaction by an ELISA-type method

6.4.1 Optimisation of an aptamer-gp120 binding assay

I designed an ELISA-type assay to investigate aptamer-gp120 interactions, which is depicted schematically in Figure 6.2a. Streptavidin-coated 96-well plates were washed three times with PBS-T (PBS containing 0.05 % Tween). Aptamer 299.5 was refolded at a final concentration of 100 nM in CHBS and 100 $\mu\text{L}/\text{well}$ was added for 30 minutes. The aptamer was removed, the plate was washed twice with PBS-T, and a blocking solution of 0.8 μM biotin and 1 % BSA in PBS was added at 200 $\mu\text{L}/\text{well}$ for 30 minutes. The plate was then washed three times with PBS-T, then 100 μL of serially-diluted tissue



culture supernatants containing wild-type or mutant JR-CSF Env were added to the plates for at least two hours. All samples were analysed in triplicate and the order of samples on the plates was randomised. The plates were washed three times with PBS-T and HIV-Ig at 1 µg/mL in CHBS was added at 100 µL/well for 1 hour. The plates were washed five times with PBS-T and HRP-conjugated anti-human IgG diluted 1/1000 in CHBS was added at 100 µL/well for 1 hour. After five final washes with PBS-T, SuperSignal substrate was added at 100 µL/well and the luminescence in each well was recorded using a luminometer.

Figure 6.2b shows the results of the ELISA-based aptamer binding assay. Pseudovirus supernatants containing gp120 from JR-CSF or BaL, with no Env, or that were mock-transfected were tested. The gp120-CB fusion protein was included as a positive control. Gp120 concentrations in each sample had been estimated by gp120 ELISA (see Section 2.7.4 and below) and two-fold serial dilutions were made starting from 2 ng/mL. The gp120-CB did not bind to the aptamer, BaL gp120 bound moderately to the aptamer, and no binding was seen with the mock-transfected sample. However JR-CSF gp120 showed a very high binding response, and this was also observed with the sample in which no Env was present. I therefore concluded that there must be a contaminating factor in the JR-CSF and no Env pseudovirus samples that was causing this response. I suspected that the polyclonal HIV-Ig may be detecting other HIV-1 proteins in the sample, so I decided to investigate the use of alternative primary antibodies.

A titration experiment was performed using a similar set up as above. Pseudovirus supernatants containing JR-CSF Env or no Env were diluted 1:3 in PBS and were detected with a serial dilution of the anti-gp120 polyclonal antibody D7324. There was no difference in detection of the two samples (Figure 6.2c) so I concluded that D7324 was not a suitable antibody to use. The same titration was performed using the monoclonal antibody 2G12,

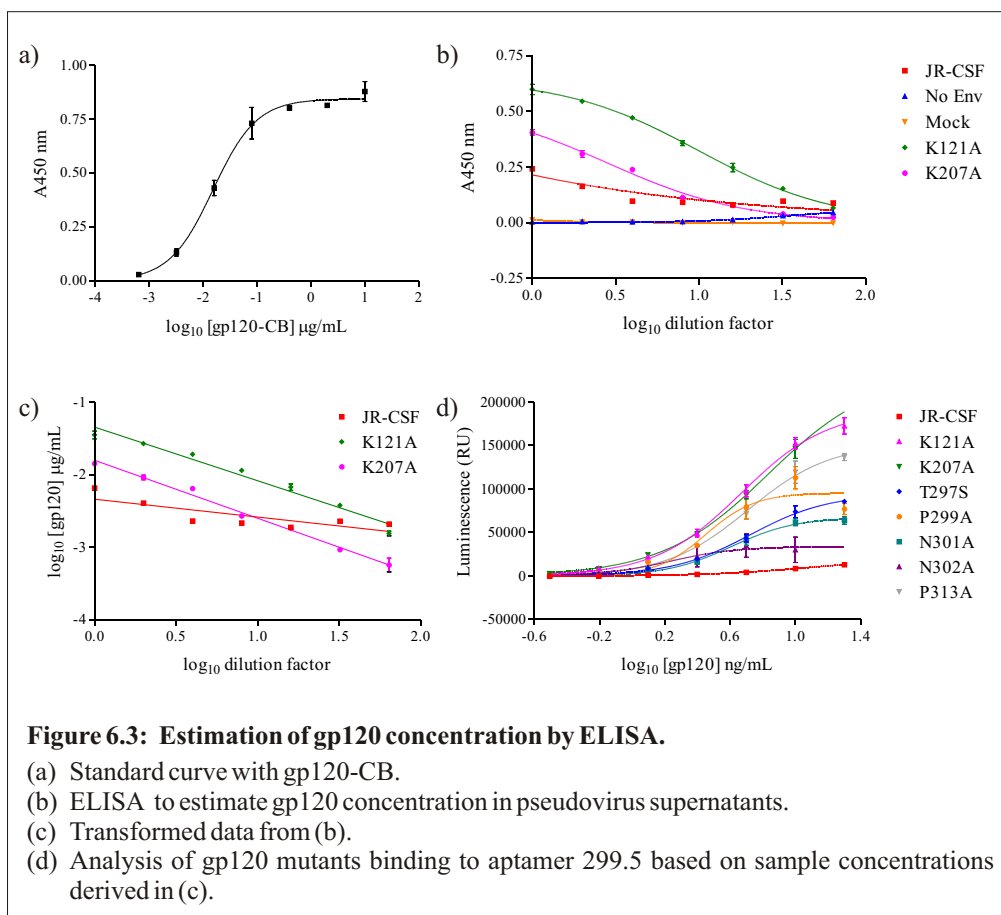
which is directed against the carbohydrate epitope of gp120. This time a specific binding response to JR-CSF gp120 was observed (Figure 6.2d), so antibody 2G12 was used in further experiments.

Serial two-fold dilutions of pseudovirus supernatants containing wild-type JR-CSF Env, no Env, K121A or K207A Env, or from mock-transfection, were made in CHBS starting from 100 ng/mL. Gp120 concentrations were estimated by gp120 ELISA (see Section 6.4.2 below). These samples were analysed in an aptamer-gp120 binding assay that used 2G12 at 1 µg/mL as the primary antibody. Figure 6.2e shows the binding curves generated in this assay. This time the JR-CSF Env sample bound to aptamer 299.5 and no binding was seen to the mock-transfected sample or the sample with no Env, indicating that the gp120-aptamer interaction was specific. The sample containing K121A gp120 showed reduced binding and the sample containing K207A gp120 showed slightly increased binding compared to wild-type gp120. These results showed that a robust, specific assay had been developed.

6.4.2 Estimation of gp120 concentration by ELISA

The concentration of mutant gp120s from pseudovirus supernatants was estimated by gp120 ELISA (Figure 6.3a-c). A standard curve was generated using known concentrations of the gp120-CB fusion protein. Serial two-fold dilutions of supernatants were made in PBS, and the titration curves were transformed using parameters from the standard curve in order to estimate their concentrations (Section 2.8.2).

These supernatants were then used in the aptamer-gp120 binding assay. Serial two-fold dilutions were made of each gp120 sample, starting from 100 ng/mL, and representative data showing binding to aptamer 299.5 are given in Figure 6.3d. Unexpectedly, all of the mutants bound more strongly to aptamer 299.5 than the wild-type JR-CSF gp120. This led me to believe



Sample	ELISA concentration (ng/mL)	Dilution factor in assay	Western concentration (ng/mL)
JR-CSF WT	85.11	4.26	648.5
K121A	45.19	2.26	326.39
K207A	15.81	1.58	208.86
T297S	47.86	2.39	618.54
P299A	26.36	1.32	42.89
N301A	33.27	1.66	57.96
N302A	15.56	1.56	62.46
P313A	25.23	1.26	258.58
D368A	39.72	1.99	181.65

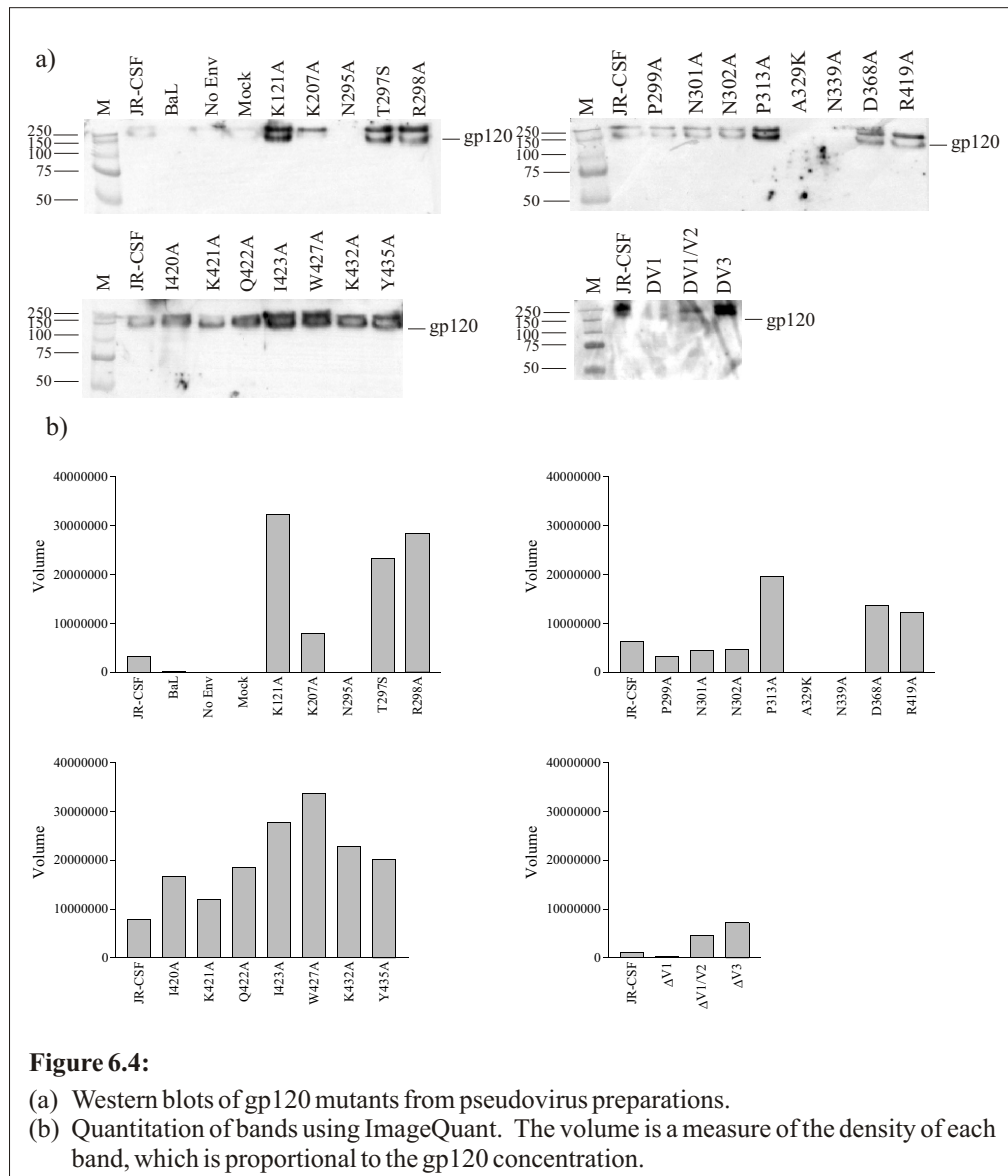
Table 6.2: Concentration of gp120 mutants expressed from pseudovirus as determined by ELISA and quantified Western blot, compared to the dilution factor used in the initial aptamer binding assay.

that there may be a problem with the method used to estimate protein concentration. Table 6.2 shows the concentration of gp120 determined from the gp120 ELISA, and the dilution factor that was used in the aptamer-gp120 binding assay. Most of the mutants were diluted between one- and two-fold to achieve the required starting concentration of 100 ng/mL whereas the wild-type JR-CSF gp120 was diluted more than four-fold. If, instead of using the concentration data from the ELISA, it is assumed that all of the gp120s are expressed at roughly equal concentrations, then the wild-type JR-CSF sample has been added to the assay at a four-fold lower concentration than the mutant gp120 samples. This could account for the unexpectedly high binding of the mutant gp120s to aptamer 299.5 compared to the wild-type.

6.4.3 Estimation of gp120 concentration by Western blot

I decided to use a second method to estimate the gp120 concentration of the same samples in order to find out if there was a problem with the gp120 ELISA. Each mutant gp120 sample was analysed by Western blot detected with 2G12 and HRP-conjugated anti-human IgG (Section 2.7.2). The concentration of each gp120 was estimated by comparing the density of the bands to that from a sample of known concentration (Section 2.7.3, Figure 6.4). Two bands can be seen in most samples, which may correspond to uncleaved gp160 and gp120, and both were included in the quantification. Mutants N295A and N339A were not detected in the Western blot because these mutations abrogate the binding of 2G12 (Scanlan *et al* 2002), so these were excluded from the remainder of the study.

Table 6.2 shows the concentration estimates of a selection of gp120 samples determined by the two different methods. It is evident that the estimates are different, confirming my previous doubts over the accuracy of the ELISA-based method. I believe the concentration estimates from the Western blot are more reliable because it is a visual method that can be verified by eye.



Therefore gp120 concentrations were estimated by Western blot in subsequent experiments. Notably, use of HIV-Ig produced unsatisfactory results in both the gp120 ELISA and the aptamer-gp120 binding assay, so it should not be used in future assays that use pseudovirus samples.

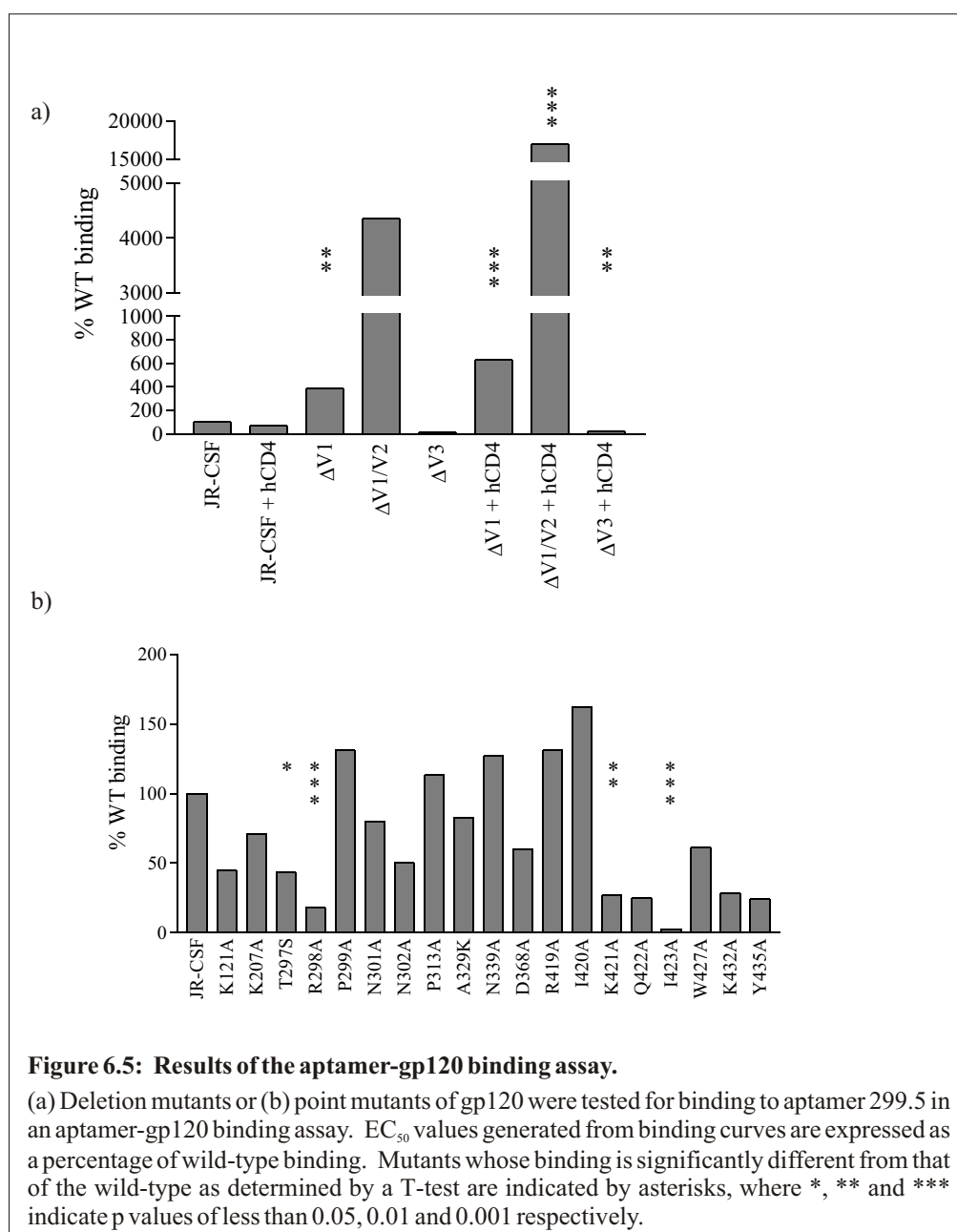
6.4.4 Interaction of gp120 mutants with aptamer 299.5

Following optimisation of the aptamer-gp120 binding assay and satisfactory estimation of gp120 concentration in each sample, I was able to analyse the interaction of each gp120 mutant with aptamer 299.5. Each mutant was analysed in at least two independent assays. Aptamer binding to deletion mutants was tested in the presence and absence of hCD4; in these experiments the gp120 was pre-incubated with or without 100 nM hCD4 for 1 hour at 37 °C. The resulting binding curves were modelled using GraphPad Prism software to generate EC_{50} values (Section 2.8.1). The mean EC_{50} value for each mutant is plotted as a percentage of wild-type binding in Figure 6.5, which was calculated using the equation:

$$\% WT \text{ binding} = \left(\frac{EC_{50}WT}{EC_{50}mut} \right) \times 100$$

where $EC_{50}WT$ is the mean EC_{50} value of JR-CSF gp120, $EC_{50}mut$ is the mean EC_{50} value of the mutant gp120, and WT stands for wild-type. EC_{50} values were compared using a T-test.

The binding of gp120 deletion mutants ($\Delta V1$, $\Delta V1/V2$ and $\Delta V3$) to aptamer 299.5 was tested in the presence or absence of hCD4 (Figure 6.5a). The $\Delta V1$ and $\Delta V1/V2$ mutants showed a significant increase in aptamer binding compared to the wild-type, whereas the $\Delta V3$ mutant showed a significant



decrease in aptamer binding. Binding of the aptamer to these deletion mutants and wild-type gp120 was unaffected by the addition of hCD4.

Of the 19 point mutants of gp120 tested, four showed a significant decrease in binding to aptamer 299.5; mutants T297S, R298A, K421A and I423A (Figure 6.5b).

6.5 Discussion

6.5.1 Summary of results

A panel of deletion and alanine-scanning mutants of gp120 were expressed using a pseudovirus system. Following an unsuccessful attempt using SPR, the interaction of these mutant gp120s with a truncated B40-derived aptamer was analysed using an aptamer-gp120 binding assay in an ELISA-type format, which was optimised to maximise binding specificity. Estimation of gp120 sample concentrations by ELISA test was unsatisfactory, so this method was replaced by a quantitative Western blot.

6.5.2 Choice of methods

It was surprising that the gp120 ELISA method used to estimate protein concentration was found to be inaccurate, given that this method is used by many groups. In experiments to determine the epitope of anti-gp120 antibodies, titration curves of gp120 binding using the antibody of interest are typically compared to the amount of binding seen using polyclonal HIV-Ig at a single concentration to normalise for gp120 concentration (Scanlan *et al* 2002; Pantophlet *et al* 2003; Darbha *et al* 2004). In this study, use of a titration curve of HIV-Ig, rather than a single concentration, has highlighted the fact that non-specific interactions may occur between HIV-Ig and other HIV-1 proteins in pseudovirus-derived gp120 samples.

Use of a Western blot to estimate gp120 concentration, while more robust than the ELISA method, is still not without fault. High sample concentrations could have saturated the autoradiograph leading to an underestimation of concentration. The accuracy could have been increased by repeating the Western blots with a range of standard and sample concentrations and by using a detection system with a greater dynamic range, but this was impractical due to limited sample availability, the large number of samples and a lack of time. Given that the calculated sample concentrations obtained using this method were comparable to those estimated by eye, the method was considered satisfactory for the purpose of these experiments.

It was disappointing that the SPR method to investigate aptamer-gp120 interactions was unsuccessful, as the inclusion of hCD4 as a control against which aptamer binding could be normalised would have negated the need for a separate estimation of gp120 concentration. Further investigation into why this method failed could prove useful in the future.

6.5.3 Involvement of gp120 variable loops

The variable loops play important roles in the binding of CCR5 and CD4i antibodies to gp120. The gp120-CCR5 interaction has been studied in detail, showing that the V3 loop is critical for CCR5 binding, as even single mutations in V3 can abrogate CCR5 binding (Wu *et al* 1996; Cormier and Dragic 2002). $\Delta V1/V2$ gp120 binds to CCR5 with the same pattern as wild-type gp120, i.e. binding is CD4-dependent (Wu *et al* 1996). Antibodies 17b and 48d both show reduced binding to $\Delta V3$ gp120, but binding is restored if CD4 is present. Deletion of the V1 and V2 loops also reduces 17b and 48d binding, but this is not restored in the presence of CD4. The deletion of V1 alone, however, has no effect on the binding of these antibodies to gp120 (Thali *et al* 1993; Wyatt *et al* 1995). The binding of antibody X5 to gp120 has a slightly different dependence on variable loops; it does show CD4-

dependent binding to $\Delta V3$ gp120, but unlike 17b and 48d it is able to bind to $\Delta V1/V2$ and $\Delta V1/V2/V3$ gp120 (Moulard *et al* 2002).

In this study I have shown that the binding of aptamer 299.5 to gp120 is dependent on different loop requirements compared to the binding of gp120 to CCR5 and CD4i antibodies. The deletion of the V1 loop results in an increase of gp120 binding to aptamer 299.5, regardless of the presence or absence of CD4, a feature that is unique to the aptamer. Deletion of the V1 and V2 loops results in increased aptamer binding, regardless of the presence or absence of CD4, which is the same pattern of binding seen with antibody X5 but not the other CD4i antibodies or CD4. In contrast to antibody binding, aptamer binding to V3-deleted gp120 is abrogated regardless of the presence or absence of human CD4. These findings suggest that while the variable loops of gp120 are obviously important for aptamer 299.5 binding, the mechanism is somewhat different from the binding of CD4i antibodies and CCR5. One possibility is that removal of the V3 loop causes the gp120 structure to collapse inwards, restricting the exposure of the aptamer binding site.

6.5.4 Specific gp120 residues involved in aptamer-gp120 interaction

Four point mutations in gp120 result in a significant loss of binding to aptamer 299.5: T297S, R298A, K421A and I423A. Two of these residues, R298 and K421 are positively charged and it is probable that, on mutation to alanine, the loss of positive charge decreases the binding of the negatively charged RNA aptamer. However, other more specific interactions may also occur between the aptamer and these four residues.

Residue T297 in gp120 has not previously been implicated in CCR5 or CD4i antibody binding because it was not tested in previous studies that used alanine scanning mutants of gp120 to map the CCR5, X5, 17b or 48d binding site (Thali *et al* 1993; Rizzuto *et al* 1998; Cormier *et al* 2000; Rizzuto and

Sodroski 2000; Cormier *et al* 2001; Darbha *et al* 2004). Mutation of T297 to serine significantly decreases aptamer 299.5 binding to gp120 and it is possible that this mutation would also affect CCR5 or CD4i antibody binding.

R298 is a critical residue involved in the gp120-CCR5 interaction that binds to the sulphated tyrosines in the N-terminus of CCR5, a property that is mimicked in the binding of 6-O-sulphated syndecans to gp120 (de Parseval *et al* 2005). Mutation of R298 to alanine results in a loss of CCR5 and syndecan binding (Cormier *et al* 2001; de Parseval *et al* 2005). This mutation also reduces binding of antibody X5 (Darbha *et al* 2004) but not 17b or 48d (Thali *et al* 1993) to gp120. The fact that this mutation also reduces aptamer 299.5 binding lends further support to the hypothesis that the aptamer binding site overlaps with the CCR5 binding site on gp120.

Mutation of K421 to alanine does not affect the binding of antibodies X5 or 17b, but it does decrease binding of 48d and CCR5 to gp120 (Thali *et al* 1993; Rizzuto *et al* 1998; Rizzuto and Sodroski 2000; Cormier *et al* 2001; Darbha *et al* 2004). I423 is an important determinant of antibody X5 binding to gp120, and is also thought to be involved in CCR5 binding, but was not included in the study testing alanine mutants of gp120 for binding to antibodies 17b and 48d (Thali *et al* 1993; Rizzuto *et al* 1998; Rizzuto and Sodroski 2000; Cormier *et al* 2001; Darbha *et al* 2004). A number of similar studies indicate that contiguous residues R419, I420, K421 and Q422 are all important for CCR5 binding to gp120 (Rizzuto *et al* 1998; Rizzuto and Sodroski 2000; Cormier *et al* 2001). The fact that mutation of K421 and I423 also affect binding of aptamer 299.5 to gp120 again indicates that aptamer 299.5 shares parts of the CCR5 and antibody binding sites.

This mutagenesis study has mapped the aptamer binding site on gp120 to four key residues. From this information and the analysis of aptamer binding to gp120 deletion mutants I believe that the aptamer binding site shares some

similarity with the CCR5 and CD4i antibody binding sites on gp120, but that there are also some differences particularly regarding the CD4-dependence of binding.

6.5.5 Model of aptamer binding to gp120

The first structure of gp120 was solved by Kwong *et al*, and shows a V1/V2/V3-deleted, deglycosylated gp120 from the HIV-1 X4 strain HXB2 in complex with two domains of CD4 and the Fab fragment of antibody 17b (Kwong *et al* 1998). More recently, a structure of V3-containing gp120 from the R5 strain JR-FL in complex with two domains of CD4 and the Fab fragment of antibody X5 was solved (Huang *et al* 2005). Figure 6.6 a and b shows this second structure in which the gp120 is shown in the “heart-shaped” orientation such that the viral membrane would be above and the target membrane would be below each Figure. In this model the position from where the V1 and V2 loops would protrude (the V1/V2 stem), the bridging sheet, the β 20- β 21 sheet and the inner and outer domains are indicated. In Figure 6.6a residues that are important for CD4 and CCR5 binding are shown in blue and green, respectively, and in Figure 6.6b residues that were found to be important for aptamer binding are indicated in pink, and those that were tested but that did not significantly reduce aptamer binding are shown in grey. From this model one can see that the aptamer binding site predominantly involves core gp120 residues at the base of the V3 loop, and it is possible to visualise how removal of the V3 loop would change the conformation of gp120 in such a way that binding would be abrogated.

The structure of gp120 that contains the V3 loop (Huang *et al* 2005) has shed further light on the domains of gp120 that are required for binding to CCR5 and CD4i antibodies. From this work, Hwang *et al* suggested that the N-terminus of CCR5 binds to the gp120 core and base of the V3 loop, while the tip of the V3 loop interacts with the second extracellular loop of CCR5. I propose that the aptamer binding site mimics that of the binding of the CCR5

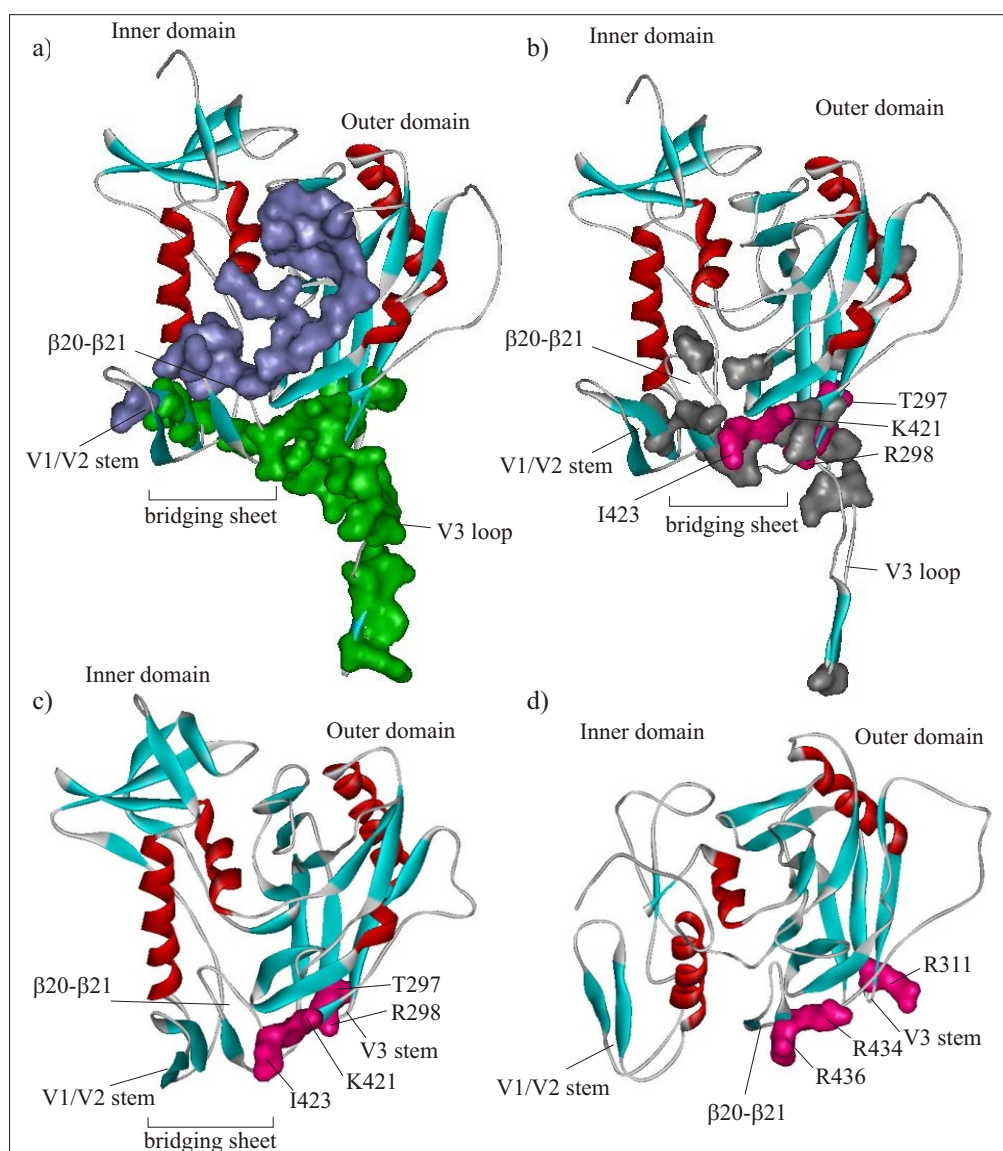


Figure 6.6: Comparison of gp120 structures.

(a) and (b) Structure of HIV-1 V3-containing gp120 in complex with two domains of hCD4 and the Fab fragment of antibody X5 (adapted from Huang *et al* 2005).

(c) Structure of HIV-1 gp120 in complex with two domains of human CD4 and the Fab fragment of antibody 17b (from Kwong *et al* 1998).

(d) Structure of unliganded SIV gp120 (from Chen *et al* 2005).

Secondary structures are indicated schematically. Carbohydrate chains, CD4 domains and antibody Fab fragments have been omitted for clarity. Residues involved in CD4 binding (blue), CCR5 binding (green) and aptamer binding (pink) and residues whose mutation did not affect aptamer binding (grey) are highlighted, and other relevant structural features are indicated.

N-terminus as it involves binding of core residues of gp120. However the aptamer binding site involves fewer residues because it is a small soluble molecule unlike the membrane-bound CCR5. Similarly, the small size of the aptamer (17 kDa) compared to an antibody (150 kDa) means that its binding site is more precise.

6.5.6 CD4 independence of aptamer binding

CCR5 and the CD4i antibodies 17b, 48d, and X5 all show a great increase in binding to gp120 when CD4 is present (Thali *et al* 1993; Wu *et al* 1996; Moulard *et al* 2002). This is because the interaction of CD4 with gp120 causes a conformational shift in gp120 in which the variable loops move and the residues required for CCR5 or CD4i antibody binding are exposed (Wu *et al* 1996). In contrast, the presence or absence of CD4 does not affect binding of aptamer 299.5 to gp120. This is somewhat surprising, given the indication from previous work that aptamer B40 binds in a site that overlaps with the CCR5 binding site (Dey *et al* 2005b). The simplest explanation for this is that the anti-gp120 aptamers were selected against gp120 in the absence of CD4, rather than a gp120-CD4 complex, so the aptamer binding site must be exposed in unliganded gp120. A second explanation is that the small size of aptamer 299.5 compared to antibodies means that it is not sterically hindered by the variable loops of gp120 and can bind to gp120 in the absence of CD4. Evidence that the Fv and Fab portions of antibody X5 neutralise HIV-1 more potently than the whole IgG lend support to this explanation (Labrijn *et al* 2003).

The structure of gp120 from SIV in an unliganded state was published in 2005 (Chen *et al* 2005), providing the first insight into differences in the conformation of gp120 in the CD4-bound and unliganded states. Figure 6.6b-d compares the two CD4-bound HIV-1 gp120 structures (Kwong *et al* 1998; Huang *et al* 2005) with the unliganded SIV gp120 structure (Chen *et al* 2005). In the paper describing the SIV gp120 structure, it was noted that in

the unliganded form the V1/V2 stem of gp120 is displaced outwards, disrupting the region known as the bridging sheet so that the V1/V2 stem is no longer adjacent to the β 20- β 21 sheet. This conformational change is thought to expose the CCR5 epitope. In Figure 6.6b-d residues that are important for aptamer binding to HIV-1 gp120, and the equivalent residues in SIV gp120, are indicated (SIV gp120 uses a different numbering system from HIV-1 gp120). This comparison highlights the fact that all of the aptamer binding site residues are in a domain of gp120 that is not affected by CD4 binding, providing a further explanation for the observation that aptamer binding to gp120 is CD4-independent.

6.5.7 Sequence conservation in gp120

Sequence variation, particularly in the variable loops of gp120, enables HIV-1 to evade the host immune system. Many residues can be changed without detriment to the virus, but there are a few key residues that are highly conserved throughout HIV-1 and other primate immunodeficiency viruses. It is believed that the highly conserved residues are particularly important for the function of gp120 including binding to CD4 and the coreceptor. Three of the residues in the aptamer binding site, T297, R298 and I423, are not conserved across HIV-1 strains. This implies that aptamer 299.5 may be unable to bind to gp120s from HIV-1 isolates that have different residues in these positions. However aptamer binding also involves residue K421, which is highly conserved across all primate immunodeficiency viruses. This may explain why aptamer B40 was shown to neutralise HIV-1 from a number of diverse strains (Khati *et al* 2003). These observations raise some interesting questions; namely whether or not aptamer 299.5 would be able to bind to gp120 from X4 HIV-1 isolates, HIV-2 and SIV, and also whether or not it is able to neutralise these viruses.

Several other experiments also indicate that B40-derived aptamers may be able to neutralise other immunodeficiency viruses. M. Khati showed that two

other aptamers raised against BaL gp120 from the same library as B40, aptamers B19 and B36, bind to gp120 from the X4 HIV-1 strain IIIB ((Khati 2002). He also performed a basic experiment that indicated aptamer B36 also neutralised the virus, but further experiments are needed to verify these data. It has already been shown that aptamer B40 is able to neutralise a diverse range of R5 HIV-1 isolates (Khati *et al* 2003). Should aptamer 299.5 be found to neutralise X4 HIV-1 strains, HIV-2 or SIV, this would not only provide even more clues as to the mechanism of action but it would also have important implications for the use of such an aptamer as an anti-HIV-1 microbicide.

6.5.8 Future work

It would be possible to perform a more extensive mutagenesis of gp120 and test aptamer binding. From the crystal structure of gp120 I have identified residues that appear to be in adjacent positions to the aptamer binding site at the base of the V3 loop. The most interesting residues to investigate in a further study would be those that have a positive charge and those that are particularly conserved across HIV-1. Positively charged residues that are adjacent to the aptamer binding site include N300, R327, Q328, H330, N332, N377, N425, R440, Q442 and R444 but of these only N377 is conserved amongst HIV-1. Other conserved residues include cysteines that are important for formation of disulphide bonds, uncharged residues that are less likely to affect aptamer binding and prolines that may be particularly important in the formation of secondary structures within gp120, so these are unsuitable candidates for mutagenesis. Therefore it seems that additional gp120 mutagenesis studies would not add much to the current understanding of the aptamer binding site. An alternative approach to delineate the aptamer binding site even more precisely is to co-crystallize the aptamer and gp120, and this project is being attempted in collaboration with Peter Kwong.

As mentioned above, investigation into the interaction of aptamers with X4 strains of HIV-1, HIV-2 and SIV would provide further insight into aptamer function. Experiments are planned to investigate whether or not aptamers can bind to gp120 from these viruses and neutralise them in cell culture.

6.5.9 Conclusions

This study has supported previous evidence that the aptamer binding site on gp120 overlaps with that of the CCR5 N-terminus and antibody 17b. The aptamer binding site has been more specifically delineated and was shown to involve binding to four key gp120 residues, one of which is highly conserved. Further studies including the crystallisation of a B40-derived aptamer in complex with gp120 and investigation of the interaction of aptamers with other HIV-1, HIV-2 and SIV isolates may serve to add even more depth to our understanding of the aptamer-gp120 interaction.

Chapter 7:

Attempts to group aptamers by competition binding assays

- Optimisation of a competition assay
- Interaction of other aptamers
with gp120 mutants

Chapter 7: Attempts to group aptamers by competition binding assays

7.1 Introduction

Having characterised the binding site of a B40-derived aptamer on gp120, I decided to investigate whether or not other aptamers bind to the same site. If all other aptamers share the B40 binding site and mechanism of action, then future work can focus on B40 alone. However, should it become evident that some aptamers have different gp120 aptatopes, they need to be characterised further, using gp120 mutagenesis. Aptamers with different gp120 binding sites could potentially be used together as anti-HIV-1 microbicides, and targeting the HIV-1 gp120 at multiple sites may increase the efficacy of such a treatment and reduce the emergence of escape mutants.

I set out to establish whether or not the anti-gp120 aptamers bind to the same site on gp120 using a competition assay. In theory, two aptamers with different gp120 aptatopes will be able to bind concurrently to gp120, whereas two aptamers with the same gp120 aptatope will compete with one another for binding. A competition assay was designed in which the binding to gp120 of one labelled aptamer to gp120 is detected in the presence of increasing concentrations of unlabelled competitor. Development of such a competition assay required many stages of optimisation and, while I was able to increase its signal-to-noise ratio and reproducibility sufficiently, non-specific competition from irrelevant aptamers was a recurring problem. Because of this I was unable to determine whether or not the aptamers share the same binding mechanism using this method.

I then investigated whether or not aptamers would be able to compete with anti-gp120 antibodies, which would enable me to map their binding site on gp120. However I saw no competition between aptamer B40 and antibody 17b, so I did not take this study further.

Next I decided to investigate whether binding of other aptamers to gp120 would be affected by the same gp120 point mutations that were shown to constitute the binding site of aptamer 299.5 in Chapter 6. Aptamers of interest were chosen based on the data from the computational analysis of aptamer structure in Chapter 4. This approach failed because it was not possible to produce the microgram quantities of biotinylated RNA that were required for the aptamer-gp120 binding assay.

Therefore I developed a new assay to analyse aptamer binding to the gp120 mutants. The first method used a slot-blot apparatus to bind the mutant gp120s to a membrane, which was then probed with radiolabelled aptamer. This was unsuccessful because the pseudovirus-derived gp120 samples were not concentrated enough to be detected. Next, I designed a plate-based assay in which immobilised mutant gp120 proteins were detected with radiolabelled aptamers, and although some data was obtained it was not reproducible enough to be interpreted. At this point it was clear that optimisation of this new assay was going to take a considerable amount of work, and my own time constraints meant that I was unable to complete this process.

Even with the employment and optimisation of several different methods I was therefore unable to determine whether or not all the anti-gp120 aptamers share the same gp120 binding site.

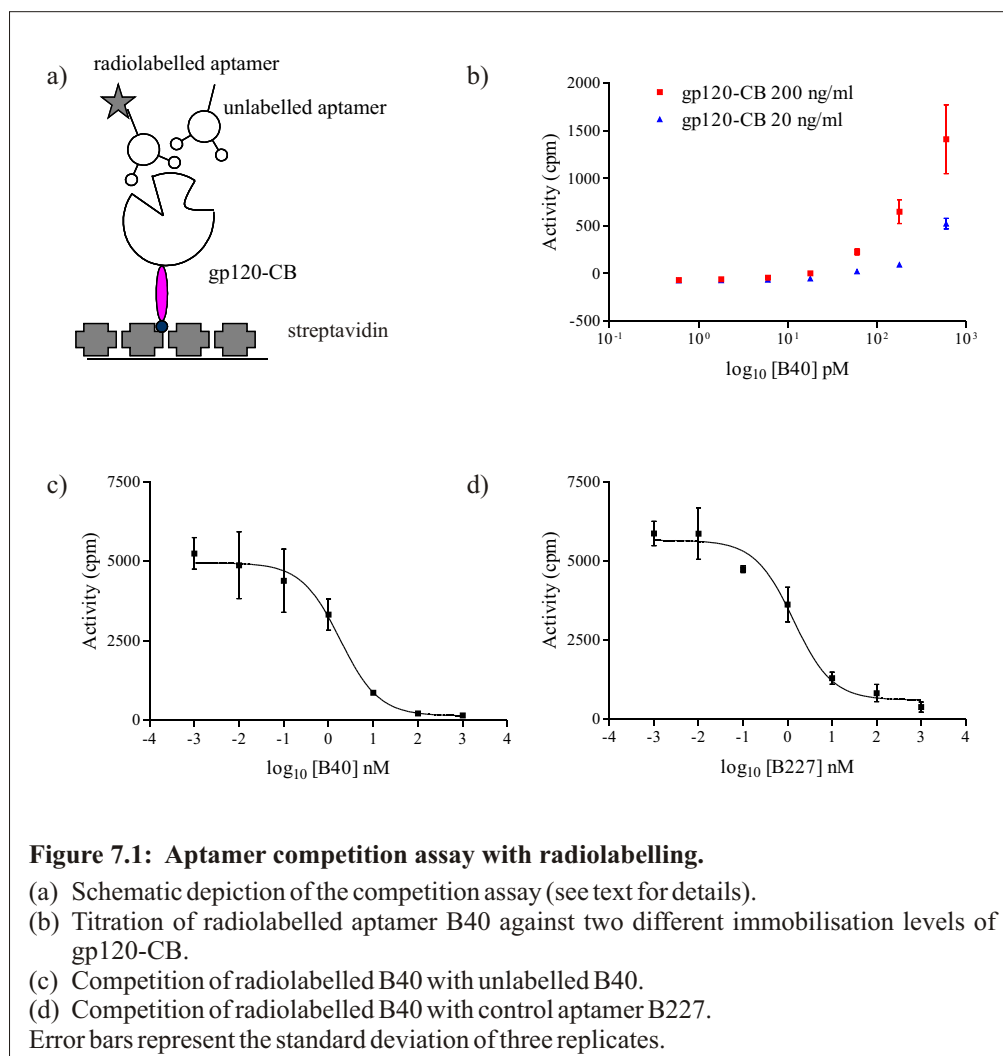
7.2 Optimisation of a competition assay

Details of steps common to all competition assays are given in Chapter 2 (Section 2.7.6), so the following descriptions only highlight features unique to each particular experiment.

7.2.1 Competition assay with radiolabelled aptamer

The first competition assay is shown schematically in Figure 7.1a. The biotinylated gp120-CB fusion protein was immobilised on a streptavidin-coated 96-well plate, which was then blocked for one hour with PBS containing 1 % BSA and 0.8 μ M biotin. A fixed concentration of radiolabelled aptamer B40 was mixed with increased concentrations of unlabelled competitor, and the aptamers were then added to the wells containing gp120-CB. Unbound aptamer was removed by washing and the activity in each well was measured by Cherenkov counting.

Initially, radiolabelled aptamer B40 was titrated against two different immobilisation levels of gp120-CB (Figure 7.1b) in order to determine the concentrations of gp120-CB and radiolabelled B40 that should be used in order to achieve a suitable signal. A competition assay was then performed in which 20 ng/mL gp120-CB was immobilised on the plate and 180 pM radiolabelled B40 was pre-incubated with a serial dilution of either unlabelled aptamer B40, or B227. Self-competition of B40 provided a positive control and aptamer B227 was chosen as a negative control as it showed very little binding to gp120 (Figure 4.4). Figure 7.1 c and d shows that increasing concentrations of unlabelled competitor resulted in reduced binding of labelled B40 to gp120-CB in both cases. I concluded that this was an unsuitable experimental set-up in which to perform aptamer competition due to the non-specific competition seen with the negative control aptamer.



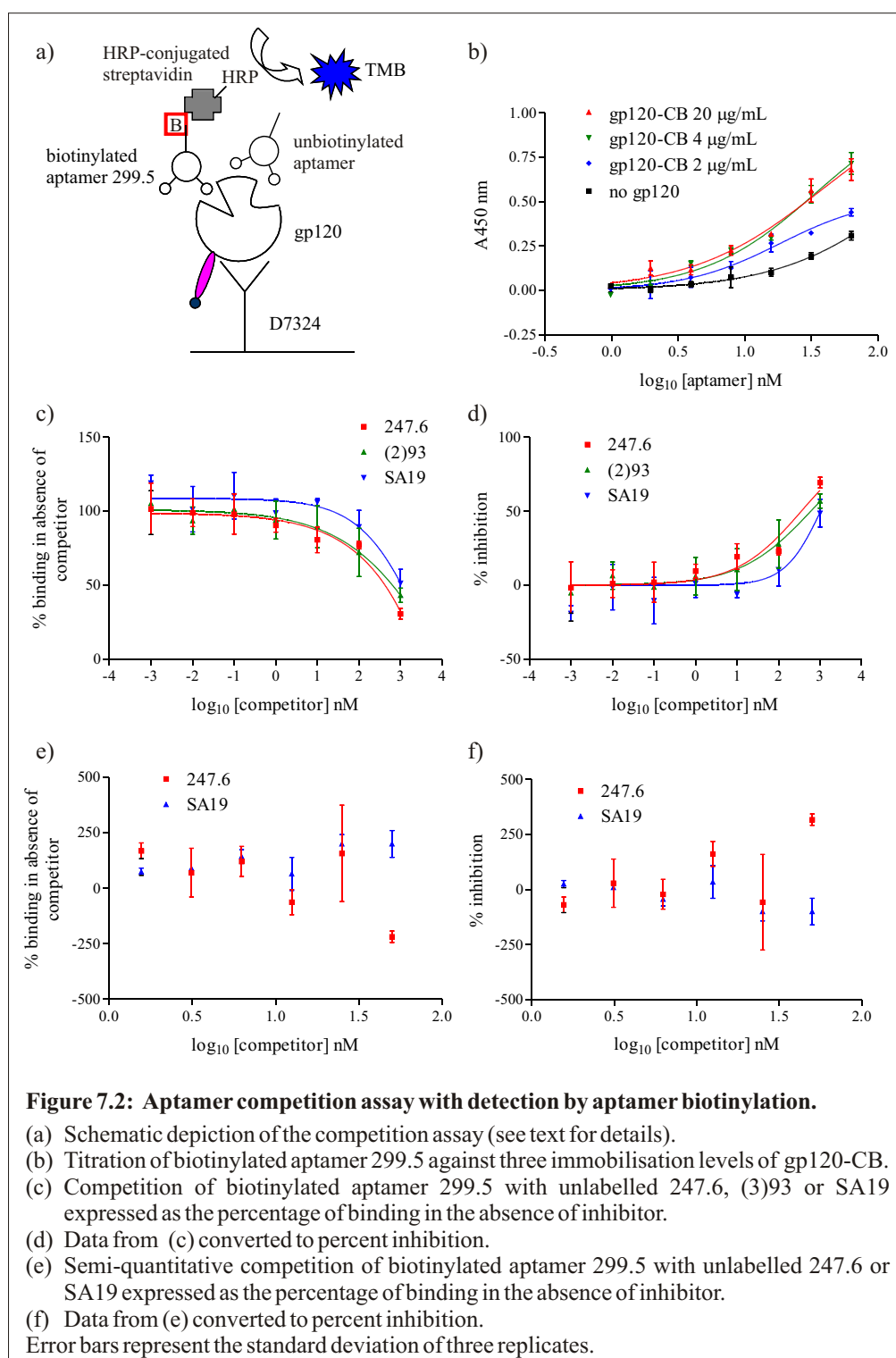
7.2.2 Competition with biotinylated aptamer

Next I designed a competition assay in which biotinylated aptamer 299.5 (a B40 derivative described in Chapter 6) was detected with HRP-conjugated streptavidin. Aptamer 247.6 has the same sequence as aptamer 299.5 but is unbiotinylated, so was used as a positive control (see Chapter 5). Aptamers that had been selected from the same SELEX library but against different target molecules were used as negative controls. Aptamer (2)93 is a truncated aptamer selected against prion protein (Rhie *et al* 2003) and aptamer SA19 was selected against streptavidin (Tahiri-Alaoui *et al* 2002).

The competition assay is depicted schematically in Figure 7.2a. Gp120-CB was immobilised on a 96-well plate via antibody D7324 in the same manner as the gp120 ELISA (Section 2.7.4). Labelled and unlabelled aptamers were pre-mixed, then added to the wells containing gp120-CB. After washing, bound biotinylated aptamer was detected with HRP-conjugated streptavidin and TMB substrate. A titration of aptamer 299.5 was performed against three different immobilisation levels of gp120-CB in order to determine the concentrations of aptamer 299.5 and gp120-CB that should be used (Figure 7.2b). Further experiments were performed using 4 µg/mL gp120-CB and 100 nM aptamer 299.5.

A competition assay was performed in which aptamer 299.5 was pre-incubated with a serial dilution of aptamer 247.6, aptamer (2)93 or aptamer SA19 (Figure 7.2 c and d). Increasing concentrations of competitor reduced binding of aptamer 299.5 to gp120-CB in all three cases, demonstrating again that non-specific competition was occurring.

In an attempt to reduce non-specific competition, the assay was performed in a less quantitative manner. Gp120-CB was immobilised on a 96-well plate via antibody D7324 as above. This time the serial dilution of unlabelled competitor aptamer was first allowed to bind to the gp120-CB for one hour.



After one wash the biotinylated aptamer 299.5 was added and allowed to bind for one hour. One further wash was performed and then the presence of aptamer 299.5 was again detected using HRP-conjugated streptavidin and TMB. The results in Figure 7.2 e and f show that there was no clear difference between competition of aptamer 299.5 with either the positive (247.6) or negative (SA19) control aptamers, so this modified method did not overcome the problem of non-specific competition.

7.2.3 Competition assay using immobilised aptamer

The Baldrich group have optimised an assay to investigate the interaction of an anti-thrombin aptamer with its ligand, and it was found that immobilisation of the aptamer rather than the thrombin reduced non-specific binding (Baldrich *et al* 2004; Baldrich *et al* 2005). Therefore I took a similar approach (depicted schematically in Figure 7.3a) and immobilised biotinylated aptamer 299.5 on streptavidin-coated plates. Gp120-CB was pre-incubated with the competitor aptamer prior to adding to wells containing aptamer 299.5. After washing, the presence of gp120-CB was detected with HIV-Ig and HRP-conjugated anti-human IgG, and the assay was developed with TMB.

A titration of gp120-CB binding to immobilised biotinylated aptamer 299.5 was performed in order to ascertain the appropriate gp120-CB concentration to use (Figure 7.3b). A concentration of 1.25 µg/mL was used in further experiments.

A competition assay was performed in which gp120-CB was pre-incubated with serial dilutions of aptamer 247.6 or SA19 prior to binding to immobilised aptamer 299.5 (Figure 7.3 c and d). Increasing concentrations of competitor reduced the binding of aptamer 299.5 to gp120-CB in both cases, indicating that non-specific competition was still a problem. Additionally the observed

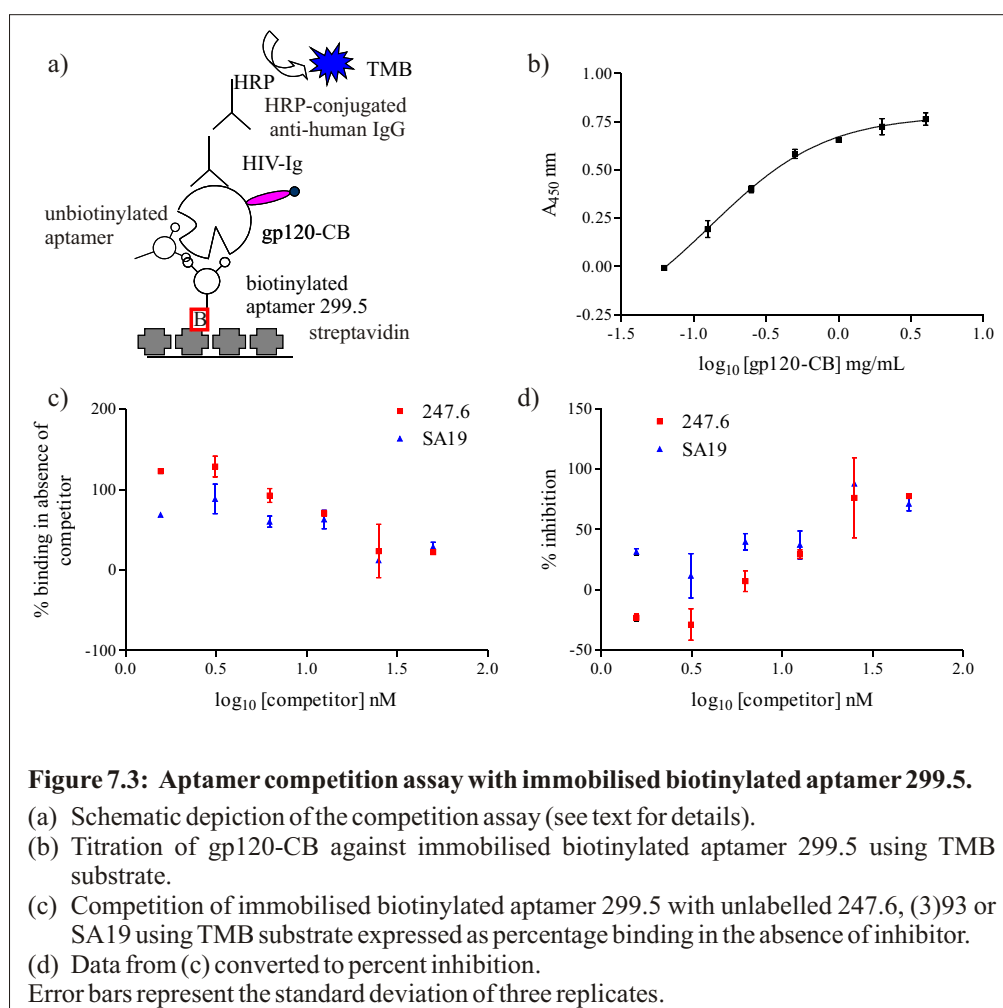


Figure 7.3: Aptamer competition assay with immobilised biotinylated aptamer 299.5.

- (a) Schematic depiction of the competition assay (see text for details).
 (b) Titration of gp120-CB against immobilised biotinylated aptamer 299.5 using TMB substrate.
 (c) Competition of immobilised biotinylated aptamer 299.5 with unlabelled 247.6, (3)93 or SA19 using TMB substrate expressed as percentage binding in the absence of inhibitor.
 (d) Data from (c) converted to percent inhibition.
 Error bars represent the standard deviation of three replicates.

responses had a small dynamic range due to the limitations of the TMB detection system.

To this end, the TMB was replaced with the luminescent detection reagent SuperSignal in order to improve the dynamic range of the response. A titration of gp120-CB on immobilised aptamer 299.5 was performed using SuperSignal as the detection reagent (Figure 7.4b) and a concentration of 16 µg/mL was used in further experiments. The competition assay was performed as in the previous experiment (depicted schematically in Figure 7.4a) and again non-specific competition with the negative control aptamer was seen (Figure 7.4 c and d).

In order to increase reproducibility and decrease the standard variation, the order of samples on the plate was randomised, and additional negative and positive controls were included; J42, an aptamer selected against HXB2 gp120, full-length B40, and B40t77. The results in Figure 7.4 e and f show that there was still no difference in competition between any of the samples. Given that antibody HIV-Ig has been shown to cause problems in other gp120-binding assays I repeated the assay using antibody 2G12 to detect gp120-CB but the same result was observed (data not shown). At this point I concluded that non-specific competition from negative control aptamers was an insurmountable problem and it would not be possible to determine whether or not all the aptamers bind to the same site on gp120 using a competition assay.

7.2.4 Competition with antibodies

I next chose to investigate competition of aptamers with antibodies. From previous work (Dey *et al* 2005b) and the delineation of the aptamer binding site in Chapter 6, aptamer B40 and its derivatives are expected to compete with CD4i antibodies for binding to gp120. I designed an assay to investigate the competition of aptamer B40 and other aptamers antibodies 17b and 2G12,

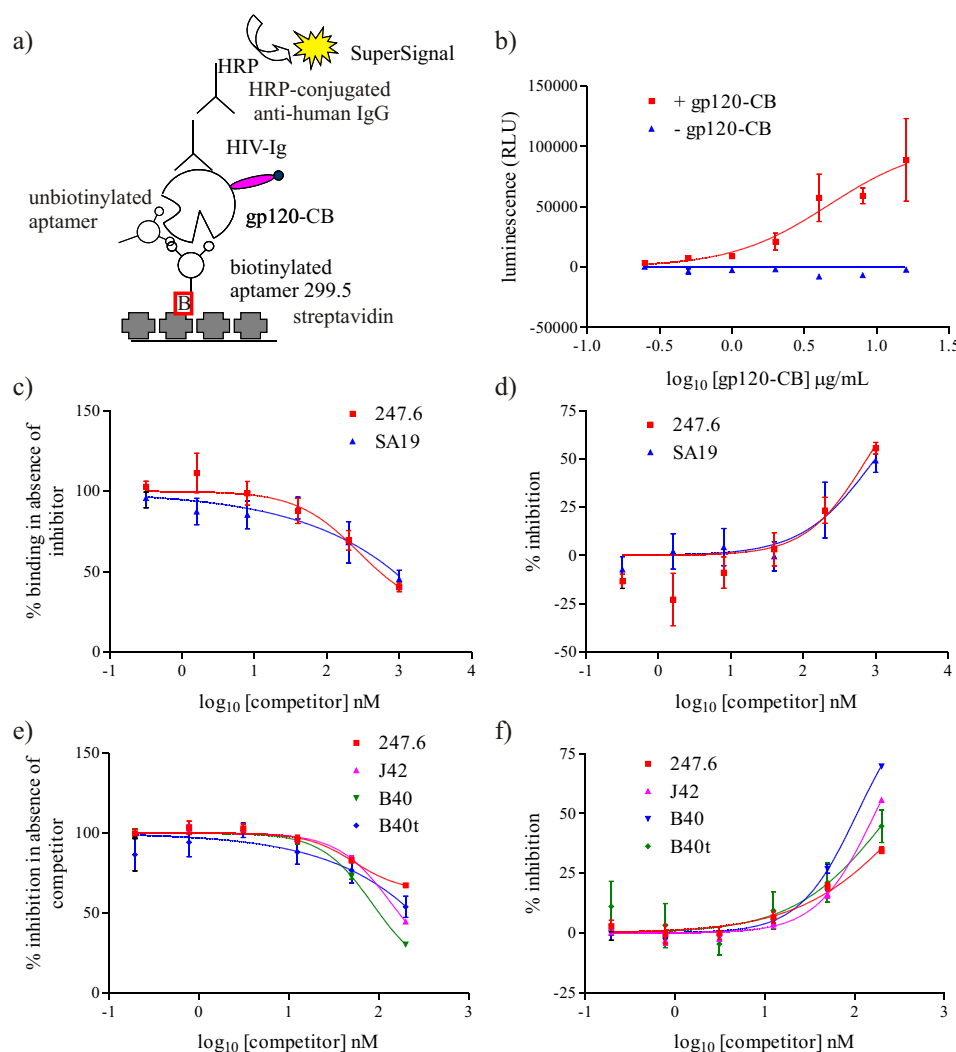


Figure 7.4: Aptamer competition assay with immobilised biotinylated aptamer 299.5 using SuperSignal detection reagent.

- (a) Schematic depiction of the competition assay (see text for details).
 - (b) Titration of gp120-CB against immobilised biotinylated aptamer 299.5 using SuperSignal as the detection reagent.
 - (c) Competition of immobilised biotinylated aptamer 299.5 with unlabelled 247.6, (3)93 or SA19 using SuperSignal detection reagent, expressed as percentage binding in the absence of inhibitor.
 - (d) Data from (f) converted to percent inhibition.
 - (e) Competition of immobilised biotinylated aptamer 299.5 with unlabelled 247.6, (3)93 or SA19, using SuperSignal substrate, following optimisation of tRNA concentration, expressed as percentage binding in the absence inhibitor.
 - (f) Data from (h) converted to percent inhibition.
- Error bars represent the standard deviation of three replicates.

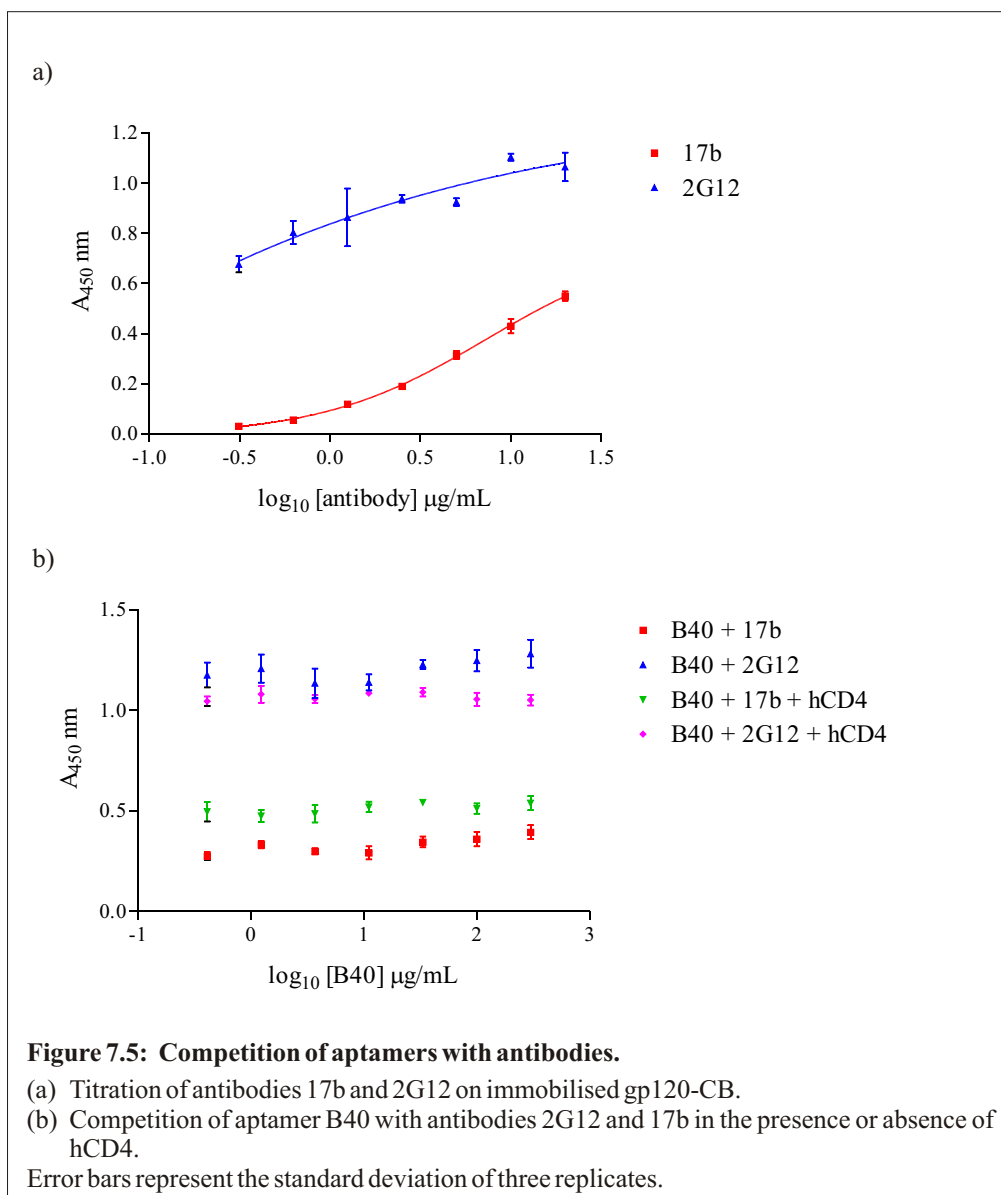
in the presence or absence of human CD4. I would expect aptamer B40 to compete with antibody 17b, which is a CD4i-antibody, but not 2G12, which binds to the carbohydrate face of gp120.

Gp120-CB was immobilised on a 96-well plate via D7324 in the same manner as the gp120 ELISA (Section 2.7.4) and antibodies 2G12 and 17b were titrated in order to determine the optimal concentrations to use (Figure 7.5a). As expected, 17b did not bind very well to gp120-CB in the absence of CD4. Antibody 2G12 was used at 1 µg/mL and antibody 17b was used at 10 µg/mL in further experiments.

For the competition assay, gp120-CB was immobilised as above and, if required, hCD4 was added to the plates at 1 µg/mL for 1 hour. After three washes, a serial dilution of aptamer B40 was added to the wells for one hour. After three more washes, antibody 17b or 2G12 were added for one hour. The antibodies were detected with HRP-conjugated anti-human IgG and TMB substrate. The results in Figure 7.5 show that aptamer B40 did not compete with either of the antibodies regardless of the presence of CD4. This was an unexpected result given the previous data that B40 competes with 17b (Dey *et al* 2005b), and because of this, the assay was not pursued as a method to investigate the competition of other aptamers with antibodies.

7.3 Interaction of other aptamers with gp120 mutants

Chapter 6 describes the delineation of the aptamer 299.5 binding site on gp120 by investigating which point mutations in gp120 reduce aptamer binding. I decided to see whether or not these same mutations also reduce the binding of other aptamers to gp120, as this would allow me to conclude



whether or not other aptamers share the same binding site on gp120 as aptamer 299.5.

It was apparent that too many experiments would be required to test the binding of all the anti-gp120 aptamers to all the gp120 mutants (1260 reactions). I selected three aptamers to study that represented the different structural groups identified from the computational analysis of aptamer secondary structure in Chapter 4: aptamer B4 (B40-like), B156 (Mirror) and B197 (Other). I chose to investigate their binding to four mutants of gp120; two that decreased aptamer 299.5 binding (I420A and K421A), one that increased aptamer 299.5 binding (P299A) and one that had no effect on aptamer 299.5 binding because it is in the CD4 binding site on gp120 (D368A).

The aptamer-gp120 binding assay employed in Chapter 6 requires the aptamer to be biotinylated. Whereas I previously used a chemically-synthesised biotinylated aptamer, here it was necessary to biotinylate the full-length aptamers using terminal transferase. This enzyme catalyses the conjugation of biotinylated-ddUTP on to the 3' end of an RNA molecule. The reaction was optimised by varying the amount of biotin-16-ddUTP and terminal transferase added to the reaction. The products of the reaction were pre-incubated in the presence or absence of streptavidin then analysed by a gel-shift assay (Section 2.2.12, Figure 7.6). Biotinylated aptamer 299.5 was included as a positive control. Figure 7.6 shows that the most optimal conditions were when 5 μ L biotin-16-ddUTP and 5 μ L terminal transferase were used. However even in this reaction only 60 % of the aptamer was biotinylated, which shows that the reaction was still inefficient. Unfortunately these reaction conditions were unsuitable to be scaled up to biotinylate microgram quantities of all three aptamers due to the large cost involved. Therefore I was unable to investigate the binding of the aptamers to gp120 mutants using the same method as in Chapter 6.

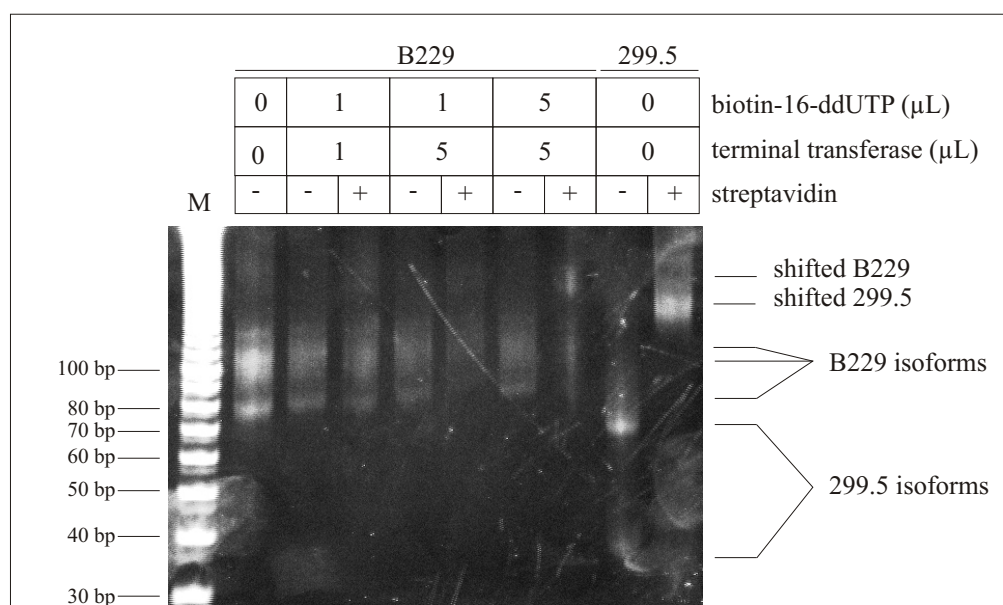


Figure 7.6: Gel-shift assay.

8 % native PAGE gel-shift assay stained with ethidium bromide to show the extent of aptamer biotinylation under different reaction conditions.

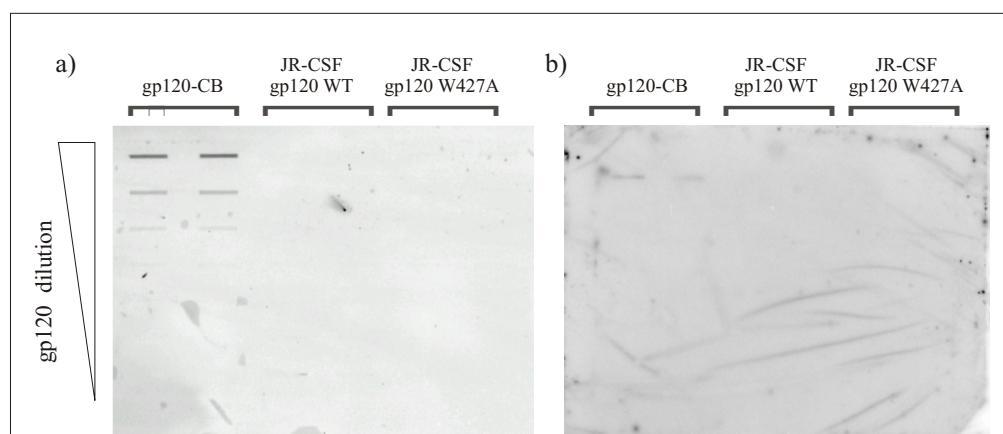


Figure 7.7: Slot-blot.

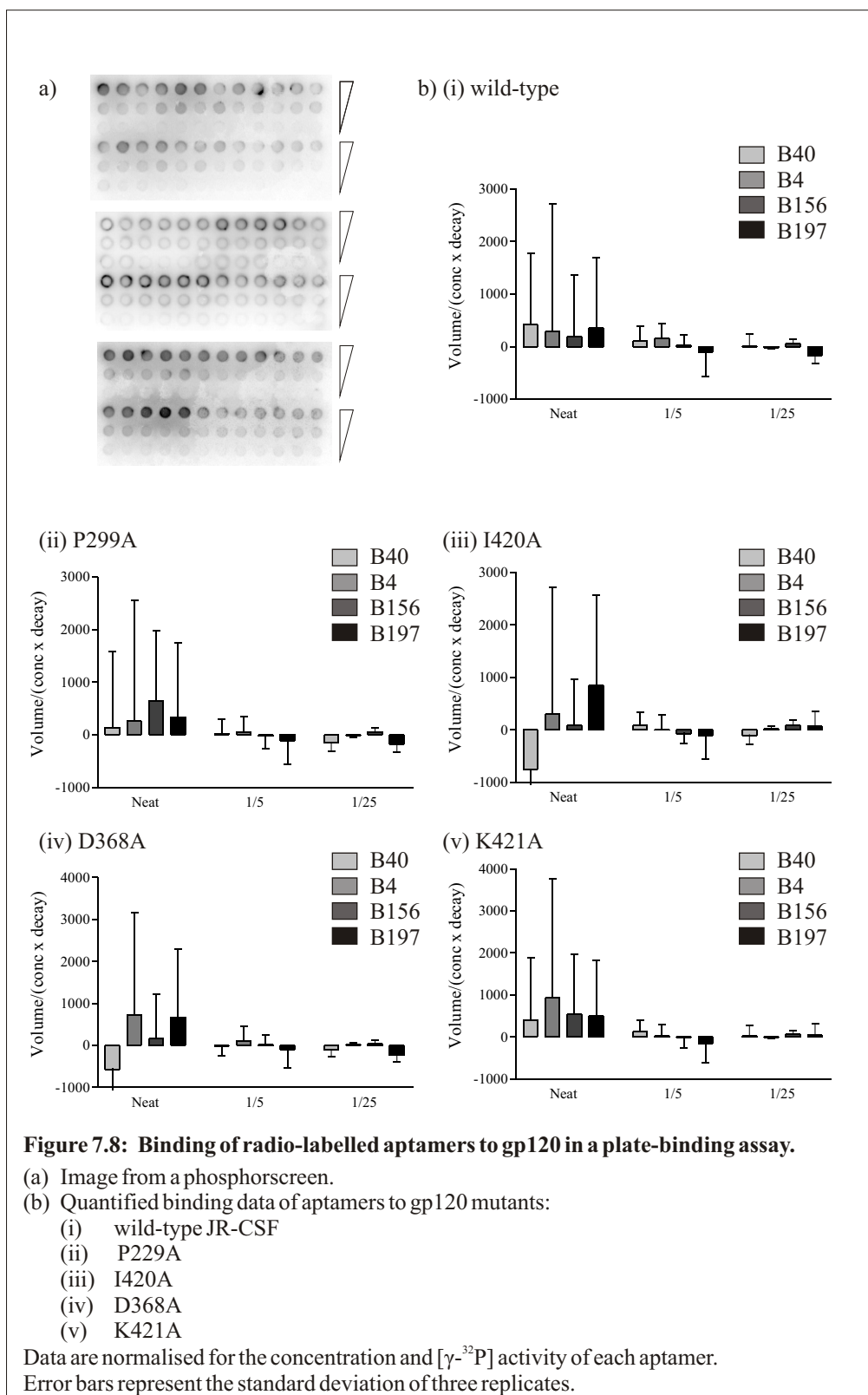
Images from a phosphorimage screen of a slot-blot showing detection of serial dilutions of gp120-CB, wild-type gp120 or W427A gp120, with (a) antibody 2G12 and peroxidase-conjugated anti-human IgG, or (b) or radio-labelled aptamer B40.

7.3.1 Slot-blot method

A new assay was designed in which serial dilutions of the pseudovirus-derived gp120 mutants were immobilised on a membrane using a slot-blot apparatus. The membrane was then probed with antibody 2G12 and HRP-conjugated anti-human IgG to normalise for protein concentration, then with radiolabelled aptamer. Full details of the method are given in Chapter 2 (Section 2.7.7) and Figure 7.7 shows the result of the slot-blot. Three dilutions of the gp120-CB sample were successfully detected with 2G12 and the highest concentration was also detected with aptamer B40. None of the pseudovirus-derived gp120 samples were detected, presumably because the sample concentration was too low. Therefore this slot-blot method was deemed unsuitable for use as a tool to investigate aptamer-gp120 interactions.

7.3.2 Plate-based method

In a final attempt to analyse the interaction of aptamers with gp120 mutants a plate-based assay was designed. Pseudovirus-derived gp120 samples were captured on 96-well plates via D7324. Serial dilutions of radiolabelled aptamers were added to the wells and bound aptamer was detected by exposing the plate to a phosphorscreen. The detailed method is given in Chapter 2 (Section 2.7.8) and the results of the assay are shown in Figure 7.8. While some aptamer binding can be seen the data is not clear enough to be interpreted. The assay was obviously unsatisfactory in a number of respects and it would take a great deal of time to optimise. At this point I decided not to take this study further, so I was unable to determine whether or not other aptamers are able to bind to point mutants of gp120.



7.4 Discussion

In this chapter I have investigated a number of methods to analyse the interaction of aptamers with gp120 in order to determine whether or not they have the same gp120 aptatope as B40. I found that it was not possible to use a competition assay in which labelled and unlabelled aptamers competed for gp120-binding, because there was always too much non-specific competition from unrelated aptamers. This may be because aptamer 299.5 binds to gp120 via positively charged residues that also bind non-specifically to any negatively charged RNA molecule. I also found that B40 did not compete with antibody 17b in an ELISA-type format, so I did not investigate the competition of other aptamers with antibodies. In contrast, previous work using an SPR assay showed that aptamer B40 does compete with antibody 17b (Dey *et al* 2005b). These differences were probably seen because gp120 was immobilised in the ELISA but the aptamer-gp120 interaction occurred in solution in the SPR assay. Immobilisation of gp120 may cause subtle changes in its structure that affect antibody and aptamer binding.

I decided not to repeat the SPR assay of aptamer-antibody competition (Dey *et al* 2005b) because I could not repeat a similar assay showing CD4-induced binding of gp120 to an immobilised CCR5-peptide (see Chapter 3). I could have performed the aptamer-antibody competition using a similar set up to the aptamer-gp120 binding assay in Chapter 6, and perhaps this is a method that could be investigated in the future. I was also concerned that the reason I could not replicate the aptamer-antibody competition assay was because I used the gp120-CB fusion protein rather than monomeric gp120. Therefore I repeated the competition assay using monomeric gp120 but there was still no competition between antibodies and the aptamer (data not shown).

I attempted to analyse the interaction of a number of aptamers with mutants of gp120 that were previously found to affect the binding of aptamer 299.5. It was not possible to use the aptamer-gp120 binding assay described in Chapter 6 because sufficient quantities of biotinylated full-length aptamers could not be produced. Two alternative methods were investigated, the slot-blot and plate-based assay, but both methods proved unsuccessful due to low sample concentration and poor signal-to-noise ratios. I was surprised that insufficient pseudovirus-derived gp120 was transferred onto the slot-blot to be detected by antibody 2G12 and aptamer 299.5 because this method is similar to the Western blot that was used to estimate gp120 concentrations. Perhaps optimisation of factors such as buffer conditions may have increased gp120 binding to the membrane and enabled use of this assay to investigate aptamer binding to gp120 mutants.

It is disappointing that I have been unable to identify the binding sites of other anti-gp120 aptamers on gp120. There are two possible ways in which this could be achieved but both require a substantial amount of work so fall outside the timescale of this thesis. It would be possible to optimise the plate-binding or slot-blot assay to determine whether or not other aptamers are able to bind to gp120 point mutants. Alternatively, the minimal binding region of other aptamers could be identified, leading to the design of truncated aptamers that could be chemically synthesised and biotinylated. These could then be tested for binding to gp120 mutants using the method from Chapter 6.

Chapter 8: Use of pseudoviruses to evaluate aptamer neutralisation

- Aptamer neutralisation of pseudoviruses
- Desulphation and removal of cell-surface HSPGs
- Neutralisation of pseudovirus in PBLs
- Detection of HIV-1 pseudovirus DNA by quantitative PCR

Chapter 8: Use of pseudoviruses to evaluate aptamer neutralisation

8.1 Introduction

This chapter describes the use of a pseudovirus system as a model to investigate HIV-1 neutralisation by aptamers. In previous work the neutralisation of infectious HIV-1 in PBMCs by full-length B40 and B40t77 was performed using a TCID50 assay system and p24 ELISA (Khati *et al* 2003; Dey *et al* 2005a). These are robust systems but are very time-consuming to perform. I chose to investigate the use of a pseudovirus system as an alternative model that would have several advantages; it would have a simple luciferase read-out, would use cell lines rather than primary cells, provide the ability to use multiple Envs including mutants, and circumvent manipulation of infectious HIV-1. Such systems are used routinely to assay the neutralisation potential of anti-HIV-1 antibodies, for example to assess their potential as candidate vaccines (Binley *et al* 2004), and it is widely assumed that they provide a representative model of HIV-1 entry. I planned to use a pseudovirus system to assay the neutralisation potential of the synthetic B40-derived aptamers described in Chapter 5.

HIV-1 pseudoviruses are produced by the co-transfection of 293T cells with two plasmids, pNL4.3Luc.R⁻E⁻ and pSVIII-Env (Section 2.6.1). Plasmid pNL4.3Luc.R⁻E⁻ encodes the backbone of HIV-1 from strain pNL4.3 but is deleted in *rev* and *env* and contains the *luc* gene that encodes luciferase, and plasmid pSVIII-Env encodes HIV-1 *rev* and *env*. Env from the strain JR-FL was used in all the experiments in this Chapter because it was found to produce pseudovirus with the highest titres. These pseudoviruses are able to

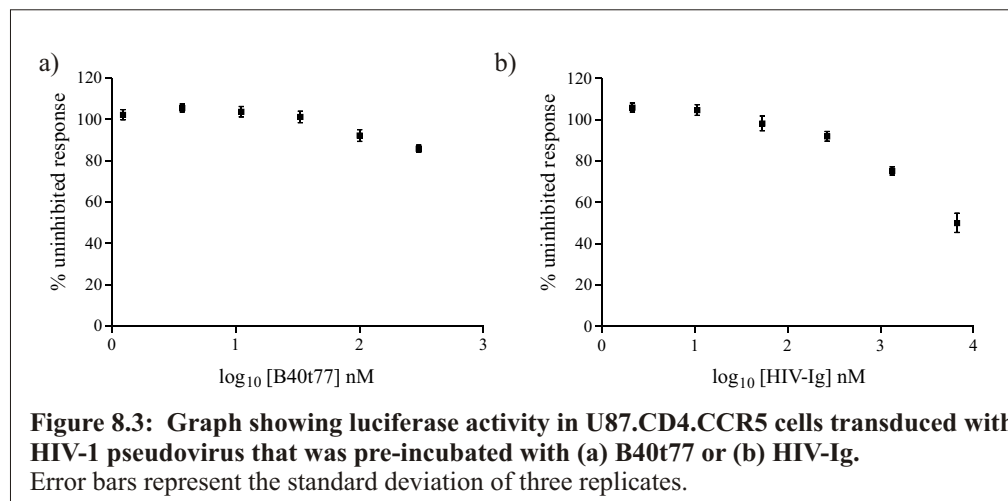
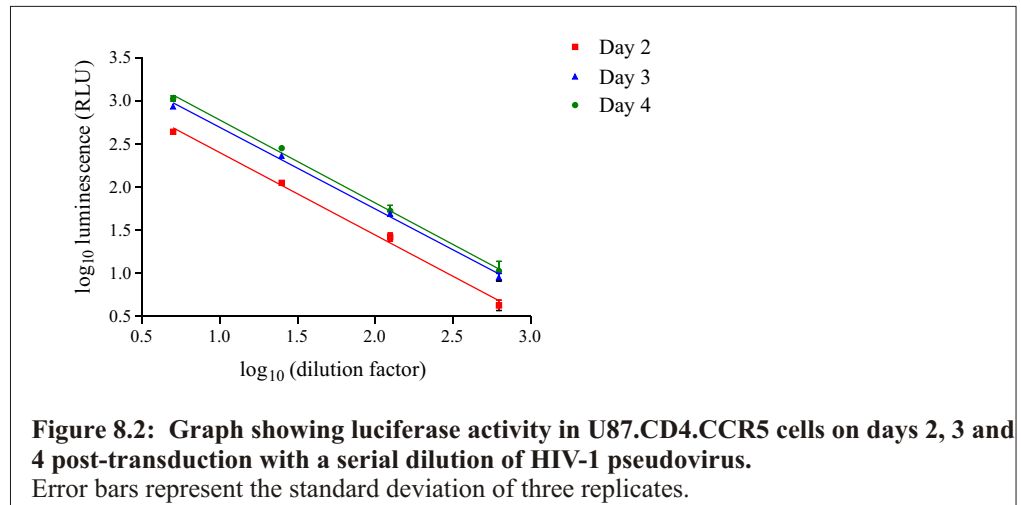
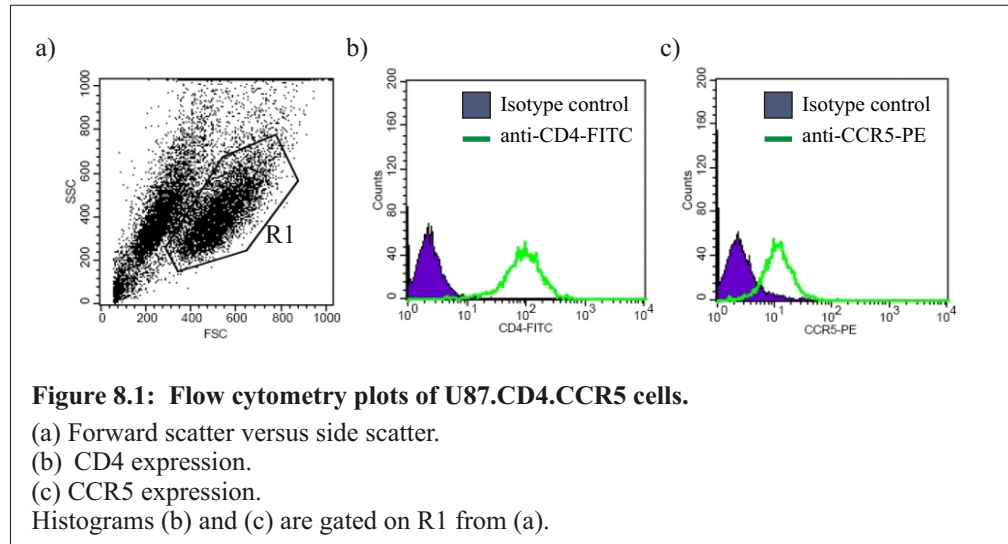
transduce target cells that express CD4 and an appropriate coreceptor but will not produce further progeny as only the pNL4.3Luc.R⁻.E⁻ plasmid, which does not encode the *env* gene, is packaged into the virion. Upon transduction the luciferase gene is expressed in the target cell and its activity is detected by the addition of a luminescent substrate that is added to the cell lysate. For neutralisation assays the pseudovirus is incubated with serial dilutions of inhibitor prior to transduction (Section 2.6.3).

In contrast to the neutralisation of infectious HIV-1 in PBMCs, aptamer B40t77 showed very poor neutralisation of HIV-1 pseudovirus in cell lines. Following this unexpected result the reasons for the difference were explored. Removal of cell-surface heparan sulphate proteoglycans (HSPGs) from the target cells was investigated, but this had no effect on the neutralisation potency of aptamer B40t77. Next the transduction of primary lymphocytes was investigated but insufficient luciferase activity was observed to be able to perform a neutralisation assay. Finally a quantitative PCR (qPCR) was employed to try and detect HIV-1 DNA in the transduced PBL but this was also unsuccessful. While I was unable to test the neutralisation potency of the B40-derived synthetic aptamers using this pseudovirus system, these experiments highlighted previously unexplored differences between pseudovirus model systems and real HIV-1 infection.

8.2 Aptamer neutralisation of pseudoviruses

8.2.1 Optimisation of the neutralisation assay

U87.CD4.CCR5 cells (Princen *et al* 2004) were chosen as target cells for pseudovirus transduction because they express CD4 and CCR5. Satisfactory expression levels of these cell surface molecules were seen by flow cytometry (Figure 8.1).



HIV-1 pseudovirus was produced by transfection of 293T cells and serial dilutions were used to transduce U87.CD4.CCR5 cells (see Section 2.6.2). The luciferase activity in the target cells was assayed two, three and four days post-transduction in order to ascertain the best time-point to use in future experiments. Figure 8.2 shows the result of the optimisation and subsequent experiments were assayed on day three post-transduction.

A neutralisation assay was performed in which aptamer B40t77 or polyclonal antibody HIV-Ig was used to inhibit pseudovirus transduction of U87.CD4.CCR5 cells (Section 2.6.2). Aptamer B40t77 was refolded to 300 nM in CHBS buffer, a concentration that had previously been shown to neutralise infectious HIV-1 100-fold, and a three-fold serial dilution of aptamer was pre-incubated with pseudovirus for one hour prior to transduction. A three-fold serial dilution of HIV-Ig from 6.6 μ M was used because a concentration of 6.25 μ M is required to neutralise infectious HIV-1 according to the product literature. Figure 8.3 shows that concentration-dependent neutralisation of pseudovirus by HIV-Ig was observed as expected, but that even the highest concentration of B40t77 (300 nM) only reduced the pseudovirus transduction by 10 %.

This was a very surprising result that led me to investigate why there was such a strong difference in aptamer B40t77 neutralisation of HIV-1 pseudovirus compared to infectious HIV-1.

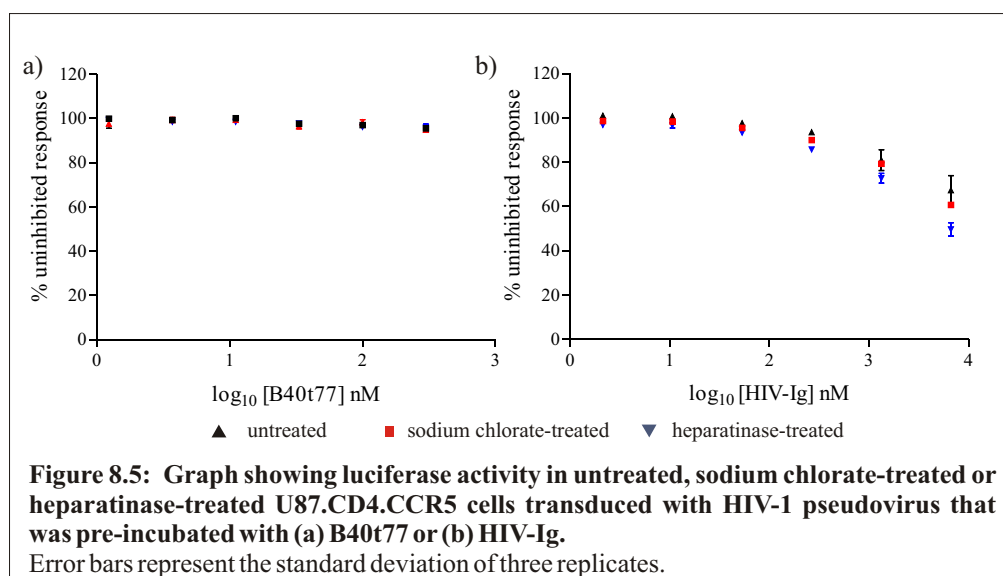
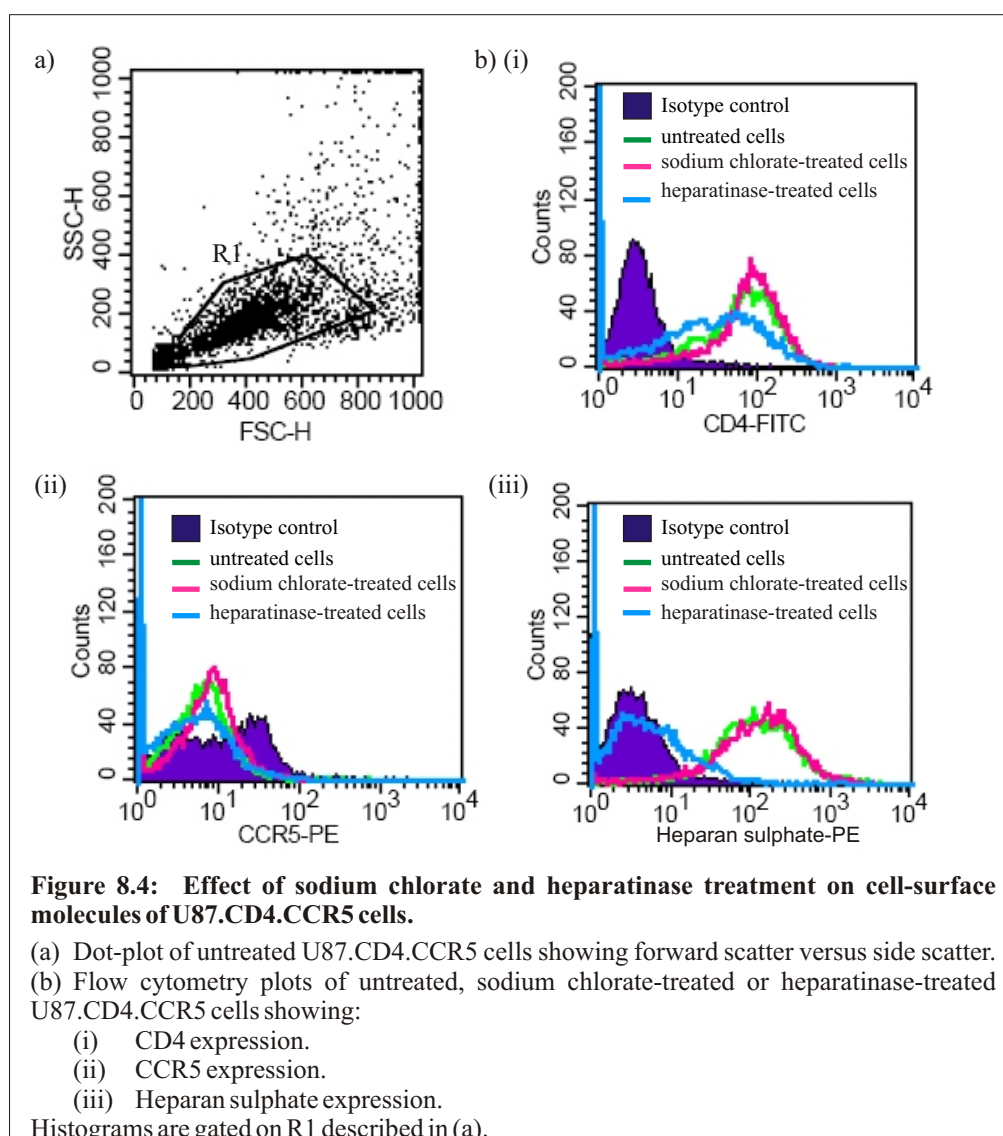
8.2.2 Desulphation and removal of cell-surface heparan sulphate proteoglycans

Heparan sulphate proteoglycans (HSPGs) are a group of glycosaminoglycans (GAGs) that are highly abundant on the surface of some cell lines but not primary lymphocytes (Ibrahim *et al* 1999). I hypothesised that the failure of aptamers to neutralise HIV-1 pseudovirus compared to the neutralisation of infectious HIV-1 may be due to the presence of cell-surface HSPGs on

U87.CD4.CCR5 cells. Residue R298 is a key determinant of gp120 binding to syndecans (a related group of GAGs), CCR5 (de Parseval *et al* 2005) and aptamer 299.5 (see Chapter 6). It is thought that the syndecan binding to gp120 mimics that of the sulphated N-terminus of CCR5. Since aptamer B40 is also thought to bind to gp120 in a similar manner to the N-terminus of CCR5 it is possible that HSPGs on the cell surface may compete with the aptamer for binding to gp120, thus reducing its neutralisation potency.

U87.CD4.CCR5 cells were treated with either sodium chlorate, to prevent sulphation of HSPGs (Humphries and Silbert 1988), or heparitinase, to remove all cell-surface HSPGs (Section 2.4.5). These cells were analysed by flow cytometry for expression of CD4, CCR5 and heparan sulphate (Figure 8.4). The sodium chlorate-treated cells showed the same levels of CD4, CCR5 and heparan sulphate expression as the untreated cells. This was expected because the anti-heparan sulphate antibody binds to desulphated heparan sulphate. The heparitinase-treated cells showed a broader distribution of CD4 expression with a lower mean fluorescent intensity, and the CCR5 expression was comparable to that on the untreated cells. As expected, the levels of HSPGs were reduced considerably compared to the untreated cells. All three cell types showed lower CCR5 expression than in the previous experiment but this did not affect the ability of the pseudovirus to transduce the cells.

I tested B40t77 and HIV-Ig neutralisation of pseudovirus using untreated, sodium chlorate-treated, or heparitinase-treated U87.CD4.CCR5 cells (Figure 8.5). These treatments did not have any effect on the ability of aptamer B40t77 to neutralise the pseudovirus. The neutralisation by HIV-Ig was marginally more potent in the heparitinase-treated cells. From these results it was clear that cell-surface HSPGs on the target cell did not specifically reduce the ability of B40t77 to neutralise HIV-1 pseudovirus compared to infectious HIV-1.

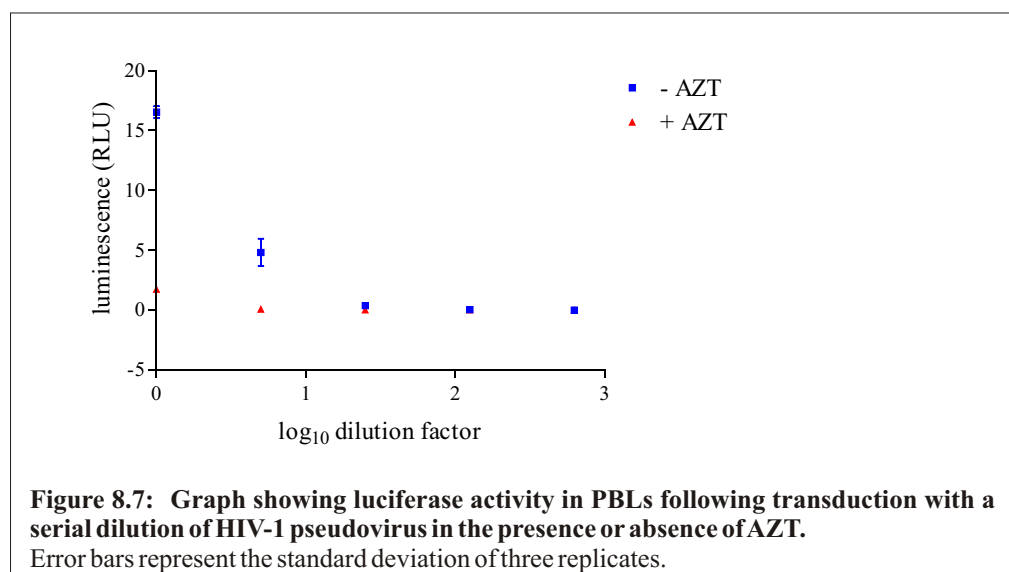
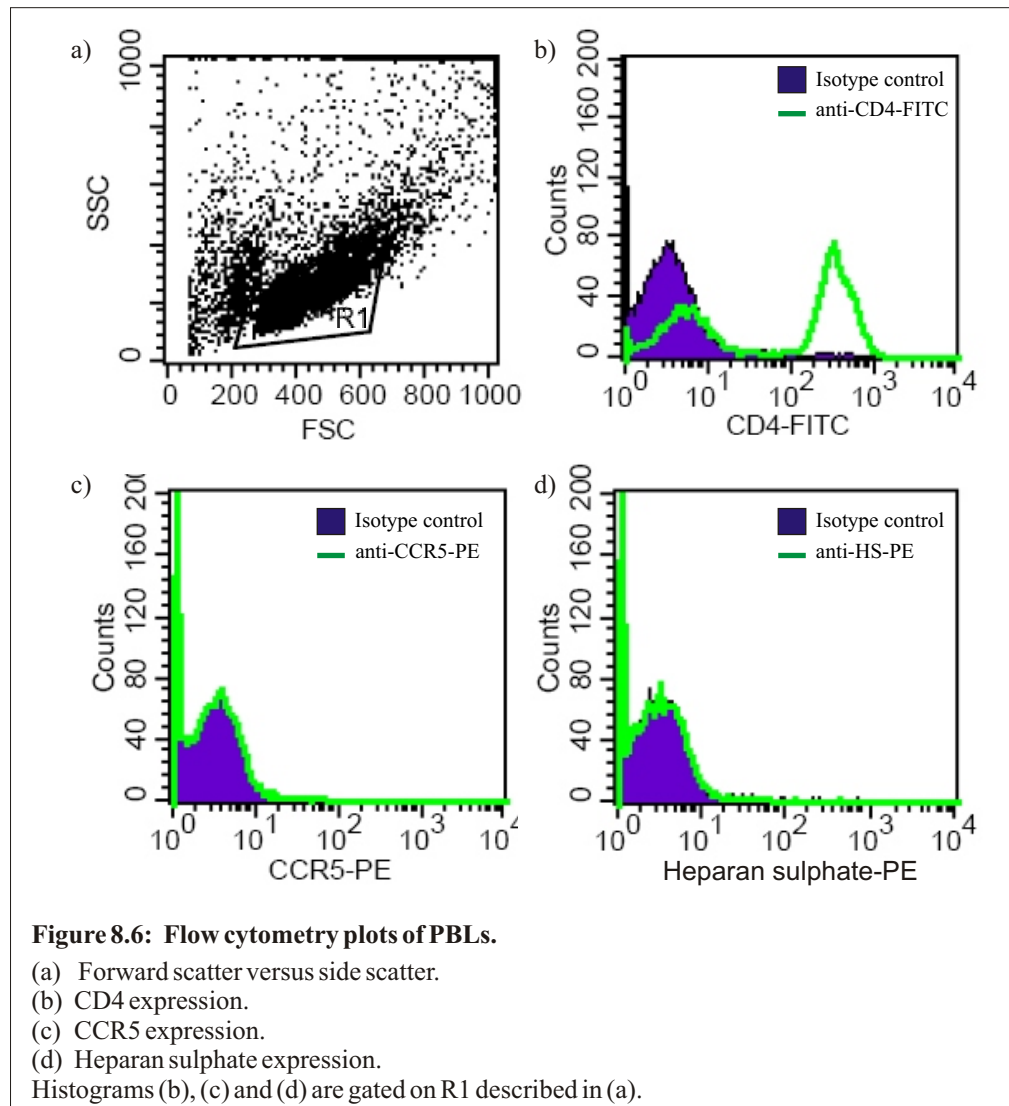


8.2.3 Neutralisation of pseudovirus in PBLs

An important difference between neutralisation assays involving pseudovirus compared to those using infectious HIV-1 is the type of target cell. Previous experiments demonstrating neutralisation of infectious HIV-1 by aptamer B40t77 were performed in primary cells, whereas pseudovirus experiments are usually performed using cell lines. I next chose to investigate whether or not aptamer B40t77 would be able to neutralise HIV-1 pseudovirus when primary peripheral blood lymphocytes (PBLs) were used as the target cell.

PBLs were isolated from Buffy Coats (see Section 2.4.2) and were analysed by flow cytometry to look for cell-surface expression of CD4, CCR5 and heparan sulphate (Figure 8.6). The cells expressed high levels of CD4 and low levels of CCR5 and heparan sulphate as expected.

Pseudovirus transduction of primary cells tends to be far less efficient than transduction of transfected cell lines whereas, paradoxically, infection of primary lymphocytes by HIV-1 is more efficient than infection of cell lines. In order to ascertain how much pseudovirus to use in the neutralisation assay, PBLs were transduced with serial dilutions of HIV-1 pseudovirus (Figure 8.7). This titration was performed in the presence or absence of AZT to prove that any luciferase activity detected was due to transduction of the cells and not from pseudovirus particles that had attached to the cell surface. The pseudovirus did transduce the PBLs, but the levels of luciferase activity observed were 100-fold lower than when U87.CD4.CCR5 cells were used (compare to Figure 8.2). It was uncertain whether this was due to low efficiency of pseudovirus transduction or low expression of luciferase in the PBLs. From these data it was clear that a neutralisation assay could not be performed because the signal-to-noise ratio was not high enough for modest levels of neutralisation to be detected.



8.2.4 Detection of HIV-1 pseudovirus DNA by qPCR

A qPCR was investigated as an alternative and potentially more sensitive method than the luciferase assay to measure transduction of PBLs by HIV-1 pseudovirus. Serial dilutions of DNase-1-treated HIV-1 pseudovirus were used to transduce PBLs in the presence or absence of AZT, and after three days the cellular DNA was harvested and used as a template for qPCR (Section 2.6.5). Serial dilutions of the pNL4.3Luc.R⁻E⁻ plasmid and human genomic DNA were used as standards. Two pairs of primers were used, one to detect HIV-1 DNA, and one to detect the β -actin gene as a control for cell number.

Figure 8.8 shows the graphical qPCR data and Table 8.1 shows the C(T) values (the number of cycles at which the curve crosses the threshold). Satisfactory standard curves were generated (Figure 8.8 a-c). The β -actin gene was detected in the pNL4.3Luc.R⁻E⁻ control because HeLa DNA was added to the sample to increase the stability of DNA at low copy numbers. The β -actin gene was detected in all the samples of transduced cells, indicating that a similar amount of DNA was extracted in each sample (Table 8.1). HIV-1 DNA was detected in the samples of cells transduced with neat pseudovirus in both the presence and absence of AZT but no HIV-1 DNA was detected in any other samples. Table 8.1 shows that at maximum a single copy of HIV-1 DNA was present in any sample of transduced cells. The fact that the same result was seen for samples in the presence and absence of AZT implies that this was due to contamination with plasmid DNA rather than being a genuine result. From this I concluded that the qPCR was not a suitable assay to detect HIV-1 DNA in transduced cells unless time was taken to optimise the method, so the neutralisation of pseudoviruses by aptamer B40t77 using PBLs could not be tested.

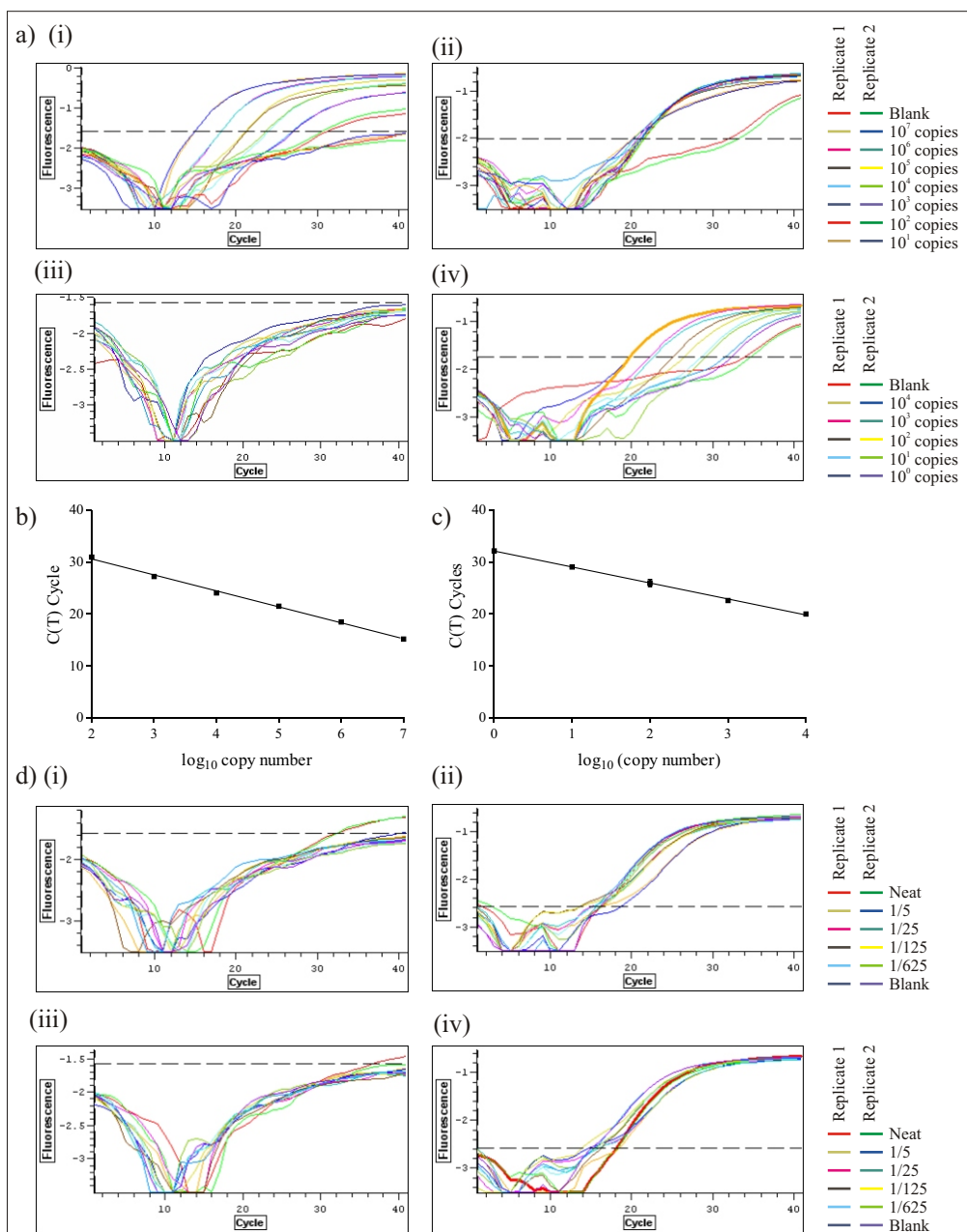


Figure 8.8: qPCR to detect HIV-1 DNA in PBLs transduced with HIV-1 pseudovirus.

- (a) qPCR amplification of
 - (i) pNL4.3-Luc DNA using HIV-1 primers.
 - (ii) pNL4.3-Luc DNA using β -actin primers.
 - (iii) human genomic DNA using HIV-1 primers.
 - (iv) human genomic DNA using β -actin primers.
- (b) HIV-1 standard curve.
- (c) β -actin standard curve.
- (d) qPCR amplification of DNA extracted from PBLs transduced with JR-FL pseudovirus
 - (i) in the absence of AZT using HIV-1 primers.
 - (ii) in the absence of AZT using β -actin primers.
 - (iii) in the presence of AZT using HIV-1 primers.
 - (iv) in the presence of AZT using β -actin primers.

DNA	Sample	C(T)	molecules
HIV-1	Neat -AZT	32.04	1
HIV-1	Neat -AZT	32.24	1
HIV-1	Neat +AZT	36.47	1
HIV-1	Neat +AZT	None	N/A
HIV-1	1/5 -AZT	None	N/A
HIV-1	1/5 -AZT	39.98	1
HIV-1	1/5 +AZT	None	N/A
HIV-1	1/5 +AZT	None	N/A
β -actin	Neat -AZT	20.77	4601.04
β -actin	Neat -AZT	20.59	5233.69
β -actin	Neat +AZT	21.73	2244.72
β -actin	Neat +AZT	21.91	1959.94
β -actin	1/5 -AZT	24.18	362.91
β -actin	1/5 -AZT	24.30	331.938
β -actin	1/5 +AZT	22.78	1029.68
β -actin	1/5 +AZT	22.63	1146.77
β -actin	1/25 -AZT	21.74	2222.16
β -actin	1/25 -AZT	21.76	2190.09
β -actin	1/25 +AZT	22.68	1109.04
β -actin	1/25 +AZT	22.63	1152.87
β -actin	1/125 -AZT	21.78	2165.53
β -actin	1/125 -AZT	22.02	1804.36
β -actin	1/125 +AZT	21.16	3418.08
β -actin	1/125 +AZT	20.74	4677.44
β -actin	1/625 -AZT	20.89	4187.13
β -actin	1/625 -AZT	21.07	3661.84
β -actin	1/625 +AZT	21.04	3754.88
β -actin	1/625 +AZT	20.94	4047.51
β -actin	Mock -AZT	20.28	6598.86
β -actin	Mock -AZT	20.42	5952.25
β -actin	Mock +AZT	19.90	8753.87
β -actin	Mock +AZT	20.01	8073.96

Table 8.1: Quantification of number of HIV-1 or β -actin DNA molecules per sample in PBLs transduced with HIV-1 pseudovirus in the presence or absence of AZT, calculated in comparison to standards of known copy number.

8.3 Discussion

HIV-1 pseudoviruses were used to transduce U87.CD4.CCR5 cells using methods adapted from the literature. Aptamer B40t77 did not neutralise pseudovirus transduction of U87.CD4.CCR5 cells, in stark contrast to the potent neutralisation of infectious HIV-1 in lymphocytes by the same aptamer (Dey *et al* 2005a). I hypothesised that HSPGs on the target cell surface may compete with B40t77 for gp120 binding and thus inhibit neutralisation, but this proved not to be the case because removal of cell-surface HSPGs from the target cell did not affect the ability of aptamer B40t77 to neutralise HIV-1 pseudovirus. It was not possible to find out whether or not B40t77 neutralised pseudovirus transduction of PBLs because transduction was too inefficient. The luciferase assay was replaced by a qPCR in an attempt to increase the signal-to-noise ratio of the output but frustratingly HIV-1 DNA could not be detected in this way. The qPCR system is currently being optimised by other members of the James lab as it would be a useful tool in the analysis of neutralisation of both infectious HIV-1 and pseudovirus by aptamers.

In these experiments pseudoviruses were produced with Env from the HIV-1 strain JR-FL. It is possible that the poor neutralisation of these pseudoviruses by aptamer B40t77 was due to this choice of Env plasmid as it has not been shown that B40t77 neutralises infectious HIV-1 of strain JR-FL. However aptamer B40 does neutralise a wide range of R5 clinical isolates of HIV-1 (Khati *et al* 2003) and therefore it is unlikely that B40 would not neutralise JR-FL HIV-1 at all. Nevertheless future experiments are planned to investigate the neutralisation of this and other strains of HIV-1, including those that use the CXCR4 coreceptor, as well as HIV-2 and SIV by B40-derived aptamers (see Chapter 6).

The attempt to neutralise HIV-1 pseudovirus by aptamer B40t77 has raised some interesting questions. It is generally assumed that the HIV-1 pseudovirus system provides a representative model of HIV-1 entry but there are clearly some differences between this model system and real HIV-1 infection. A study by the Burton group (Binley *et al* 2004) compared neutralisation of 25 primary HIV-1 isolates by five antibodies, either with infectious virus in PBMC or pseudovirus in U87.CD4.CCR5 cells. Their data showed 73 % concordance overall between the two methods, with the discrepancies being primarily attributed to the coreceptor density on the target cell surface and variation in Env sequence of infectious HIV-1 due to passaging. Like aptamer B40t77, antibody X5 was more potently neutralising in the PBMC assay than the pseudovirus assay although the difference was less substantial. The mechanism of neutralisation by X5 (by blocking the CCR5 binding site) is similar that of aptamer neutralisation, so it is interesting that both these molecules showed weaker neutralisation in the pseudovirus assay. It is likely that these differences are due to higher levels of CCR5 expression on U87.CD4.CCR5 cells than PBMCs. I believe that CCR5 competes with antibody X5 and aptamer B40t77 binding for gp120, so in both cases so more inhibitor is required to achieve the same level of neutralisation, however further investigation is required to verify this. Overall the study by Binley *et al* concluded that the pseudovirus system provided a representative model of HIV-1 neutralisation by antibodies, whereas I have shown that this is not the case for neutralisation by aptamers.

HIV-1 infects primary cells much more efficiently than cell lines, but the opposite is true for pseudovirus transduction. Further investigation is underway in the James lab to quantify this observation and to assess more precisely the neutralisation potency of aptamer B40t77 cell lines and primary cells using pseudovirus and infectious HIV-1.

Since aptamer B40t77 did not neutralise HIV-1 pseudovirus I was unable to use this system to assay neutralisation by other B40-derived aptamers. At the moment only the gp120-binding capacity of these aptamers has been investigated; a property that does not necessarily correlate with neutralisation potency. Analysis of aptamer neutralisation of infectious HIV-1 is now a priority, so that we can ascertain whether or not these are suitable molecules to take forward as candidate microbicides.

Chapter 9: General Discussion

- Summary of results
- Aspects of gp120 structure
- Structural analysis of anti-gp120 aptamers
- Evidence for the aptamer binding site on gp120
- Aptamers as diagnostic tools
- Therapeutic applications of aptamers

Chapter 9: General Discussion

9.1 *Summary of results*

I cloned and expressed the gp120-CB fusion protein as a source of soluble gp120 to be used in subsequent experiments, and I cloned and sequenced a number of novel anti-gp120 aptamers from the round 5 SELEX pool (Khati *et al* 2003). I used a bioinformatics approach to compare the structures of all the anti-gp120 aptamers isolated to date, finding that a number of the aptamers share a similar structural motif but that novel gp120-binding structures are also likely to be present. The structure of aptamer B40 has previously been studied in detail (Dey *et al* 2005a), and I built on this work in order to design truncated, chemically-synthesised aptamers whose stability was increased by the introduction of mutations and crosslinking. These aptamers were screened for binding to gp120 and the strongest binders will be taken forward as candidate microbicides. One of these B40-derivatives was used to map the aptamer binding site on gp120, which I achieved by developing a new assay to analyse aptamer binding to deletion and point mutants of gp120. This work showed that the aptamer binding site on gp120 overlaps with the CCR5 binding site, and that aptamer binding involves specific interactions with four gp120 residues, one of which is highly conserved. I was unable to determine whether or not other aptamers bind to the same site on gp120 as aptamer B40, and I found that aptamer B40t77 is unable to neutralise HIV-1 pseudovirus.

9.2 *Aspects of gp120 structure*

The gp120-CB fusion protein was expressed from 293T cells using a transient transfection system and from CHO-K1 cells by making a stable cell line. In order to verify that the gp120 was folded into the correct conformation I

tested binding of various gp120 ligands (antibodies, aptamer, CD4 and an N-terminal peptide of CCR5) to these gp120-CB preparations and compared it to the ligand binding to monomeric gp120 produced in Sf9 or 293T cells. From this comprehensive analysis I ascertained that there are in fact subtle differences in the gp120 structure between the monomeric and fusion proteins. Most interestingly I found that expression of gp120 as a fusion protein increased its binding to aptamer and CCR5 in a CD4-independent manner but only when it was not immobilised on a sensor chip surface. It was concerning that I was unable to repeat previous work that showed CD4-induced binding of gp120 to CCR5 using an SPR assay (Dey *et al* 2005b), and this meant that later I did not pursue a similar method to study aptamer and antibody competition for gp120 binding. These experiments have highlighted the flexibility of the gp120 structure that subtle structural changes occur both upon its binding to ligands and when it is expressed as a fusion protein. Retrospectively it may have been preferable to persevere with expression of monomeric gp120 from Sf9 insect cells, as I was unable to purify the gp120-CB without altering the binding pattern of various antibodies.

9.3 Structural analysis of anti-gp120 aptamers

Having cloned and sequenced a number of novel anti-gp120 aptamers my aim was to analyse their structures in order to ascertain whether homologous structural motifs were present in the aptamer population. I chose a bioinformatics approach to identify motifs homologous to the branched structure previously identified by eye in aptamers B40, B4 and B116. This analysis showed that indeed there are other aptamers that contain the branched structural motif as well as others that contain the motif in a mirror orientation and some that are structurally unrelated to B40. I also used the program Sfold to estimate the proportion of RNA molecules within the thermodynamic ensemble that are likely to contain the B40-like motif. I concluded that aptamers B4, B40, B45, B84, B137, B165, B207 and B299 are structurally

related to B40 and that for these aptamers the branched motif is likely to contribute significantly to gp120 binding. I used ClustalW to analyse the loop sequences of the homologous branched structures but was unable to find evidence to support the hypothesis that there are regions of local sequence homology within these loops. Therefore it is probable that it is the overall branched motif rather than individual sequence components that are required for gp120 binding. Given that some of the anti-gp120 aptamers did not seem to contain the branched structure, future work should focus on exploring other structural motifs in aptamers that are required for gp120 binding and mapping alternative aptamer binding sites on gp120. These aptamers could be used not only in further elucidation of the HIV-1 entry pathway but also in combination with B40-derived aptamers as anti-HIV-1 microbicides.

One may question the need to investigate other anti-gp120 aptamers given that B40 has been shown to neutralise HIV-1 in PBMC and that synthetic derivatives of B40 have been designed with the aim of developing them as anti-HIV-1 microbicides. I have previously noted that other aptamers found to bind to different sites on gp120 could be used in combination with B40-derived aptamers as microbicides in order to increase their effectiveness. It is also possible that other aptamers may neutralise HIV-1 more potently than aptamer B40. B40 was originally chosen for detailed analysis because energy landscape mapping analysis (Dey 2004) predicted it to be the most likely aptamer to crystallise; it was not the best gp120-binding or HIV-1-neutralising aptamer. I would particularly like to investigate the neutralisation of HIV-1 by aptamer B156, because this aptamer was shown to have a higher affinity for gp120 than aptamer B40, and because it contains a structural motif that may be related to the branched structure.

I was unable to substantiate the findings from the computational experiments with *in vitro* studies because both the competition assay and an attempt to analyse the interaction of full-length aptamers with gp120 mutants failed.

This was disappointing because computational approaches rely heavily on assumptions regarding the thermodynamics of RNA folding, as well as the evolutionary relationship of the sequences being analysed. This may mean that the results of the computational analysis do not provide a reasonable representation of the RNA structures that actually exist. Interestingly, as I mentioned in the discussion of Chapter 4, this work has highlighted caveats in the analysis of aptamer RNA sequences due to the fact that bioinformatics programs are designed with the intention of analysing RNA sequences that are related by evolution. Optimisation of bioinformatics programs for the analysis of aptamer sequences is a stimulating challenge for the future.

9.4 Evidence for the aptamer binding site on gp120

A number of lines of evidence support the hypothesis that the aptamer binding site on gp120 lies in the core region at the base of the V3 loop, which overlaps with the CCR5 binding site. This idea was first suggested by the study showing that aptamer B40t77 competes with an N-terminal peptide of CCR5 for binding to gp120 (Dey *et al* 2005b). In Chapter 6 I used gp120 mutagenesis to show that an aptamer derived from B40 requires four gp120 residues for binding, three of which have previously been shown to be important for gp120 binding to CCR5 (Rizzuto *et al* 1998; Rizzuto and Sodroski 2000; Cormier *et al* 2001; Cormier and Dragic 2002). Using deletion mutants of gp120 I showed that while aptamer binding is dependent on variable loops one, two and three, the pattern of binding is different from that of CCR5 and CD4i antibodies. This implies that there may be some differences in the mechanism by which aptamers bind to gp120. Further evidence comes from the observation that the chemokine Rantes, the natural ligand for CCR5, binds to aptamer 299.5 (Section 3.6.3, Figure 3.21). This was an unexpected finding; indeed Rantes was merely injected as a positive control for the immobilisation of the CCR5 N-terminal peptide. This result

indicates that the structure of aptamer 299.5 mimics that of the N-terminus of CCR5 and again strengthens the argument that they both bind to the same site on gp120.

All the aptamer binding studies were performed using monomeric gp120. The surface of an HIV-1 virion is covered with spikes consisting of gp120 trimers that are non-covalently linked to trimers of gp41 (reviewed in Pantophlet and Burton 2006). It is not known precisely how the structure of gp120 changes between the monomeric and trimeric states, but it is believed that trimerisation masks many antibody epitopes (Kwong *et al* 2000). Evidence suggests that antibodies that bind to the gp120 trimer are able to neutralise HIV-1 more potently than those that bind preferentially to monomeric gp120, see for example (Yang *et al* 2002). Additionally, Kwong and colleagues demonstrated that the binding of non-neutralising antibodies to gp120 involves an increase in entropy that corresponds to conformational restriction of gp120 flexibility (Kwong *et al* 2002). This entropic penalty, incurred due to occlusion from glycosylation and oligomerisation, and also due to the large antibody footprint, means that antibody binding to the gp120 trimer is hindered. Antibody 17b, whose binding site overlaps with that of CCR5 and thus the aptamer, is able to bind to the gp120 trimer (Thali *et al* 1993; Kwong *et al* 2000). It would be interesting to see if, like neutralising antibodies, the aptamer binds to trimeric gp120. From the parallels with neutralising antibodies outlined above I would expect the aptamer binding site to be exposed on the gp120 trimer. If this is found to be the case this may initiate investigations to find out, for example, the number of aptamer molecules required to neutralise one virus particle and how many spikes must be bound to achieve neutralisation.

One study has shown that antibody binding affinities for gp120 point mutants do not necessarily correlate with neutralisation potency of pseudoviruses containing those point mutations, indicating that the antibody affinity for the

monomeric and trimeric forms of gp120 are different (Pantophlet *et al* 2003). I found that aptamer B40t77 was unable to neutralise wild-type pseudoviruses (Chapter 8), so unfortunately I was unable to test whether or not its neutralisation potency was affected by point mutations in the gp120 of pseudoviruses.

If, as described above, the aptamer binding site does involve conserved residues that are also required for CCR5 binding, this would explain why several aptamers were able to neutralise of a panel of HIV-1 isolates potentially (Khati *et al* 2003). A project shortly to be undertaken in the James lab is the investigation of whether or not aptamers will also neutralise HIV-2 and SIV in PBMCs.

9.5 Aptamers as diagnostic tools

One of the unexpected findings in this thesis was the poor neutralisation of pseudoviruses in U87.CD4.CCR5 cells by aptamer B40t77, particularly in comparison to aptamer neutralisation of infectious HIV-1 in PBMC (Dey *et al* 2005a). The aptamers that were selected against X4 HIV-1 strain HXB2 did not neutralise the virus in a cell line, H9 (Sayer *et al* 2002). Therefore it would be interesting to investigate whether or not these aptamers are able to neutralise X4 and perhaps R5 HIV-1 in PBMCs. Previous unpublished work has also indicated that neutralisation of R5 HIV-1 by the aptamers raised against BaL gp120 was more potent in PBMC cultures than in purified macrophage cultures (Khati 2002), although this observation is yet to be repeated by others. Cell-surface expression of CCR5 varies considerably among different cell types and I have noticed that increasing CCR5 expression (PBMC < macrophage < U87.CD4.CCR5) correlates with decreasing aptamer neutralisation potency. Considering that the CD4i-antibody X5 was also found to neutralise infectious HIV-1 in PBMC more potently than pseudovirus in U87.CD4.CCR5 cells (Binley *et al* 2004), this

supports the hypothesis that increased levels of cell-surface CCR5 compete with the inhibitor for gp120 binding, reducing its neutralisation potency. In the future aptamers could be used as diagnostic tools to expose other detailed aspects of the HIV-1 entry pathway, such as this, that cannot be explored using antibodies.

9.6 *Therapeutic applications of aptamers*

9.6.1 Screening for small molecule inhibitors

Current HIV-1 treatments have several limitations including unpleasant side-effects, problems with patient adherence, and emergence of resistant strains, so development of new drugs that overcome these drawbacks is a continual challenge in HIV-1 therapeutics. It would be possible to screen libraries of “drug-like” molecules in a high-throughput assay to find inhibitors of the aptamer-gp120 interaction. Such molecules are expected to bind to gp120 in the same conserved binding site as the aptamer and may be found to inhibit HIV-1 infection. Inhibitors that bind to conserved regions of gp120 are less likely to promote the emergence of resistant strains because conserved residues cannot be mutated without a significant reduction of viral fitness (Pastore *et al* 2004). It would be possible to perform such a high-throughput screen based on the aptamer-gp120 binding assay developed in Chapter 6.

9.6.2 Structure-based drug design

Aptamer B40t77 could not be crystallised (Dey 2004), probably because the unliganded RNA structure is too flexible. An attempt to crystallise the structure of aptamer 247.6 in complex with gp120 is in progress, in collaboration with P. Kwong. The aptamer-gp120 complex is probably locked in a more rigid state than either the aptamer or gp120 alone, so is more likely to be crystallised successfully. Information from this study will prove invaluable in confirming the previous studies that have mapped the aptamer

binding site, and would also enable structure-based drug design of anti-HIV-1 therapeutic agents.

9.6.3 Aptamers as microbicides

Microbicides have several advantages as anti-HIV-1 agents, as outlined in Chapter 1 (Section 1.1.8). I have designed a number of synthetic aptamers that have the potential to be developed as microbicides. The aptamers have been tested for gp120 binding and, while it is assumed that they will also neutralise HIV-1 in PBMC culture, verification of this is of utmost importance. Following this experiment the synthetic aptamers can be taken forward as candidate microbicides, a process that will involve many further steps of optimisation. Already the labile RNA molecules have been stabilised by the use of 2'-fluoro-pyrimidines, mutagenesis and 2'-O-dimethylallyl crosslinking, as well as being truncated to reduce the cost of production. Further modifications such as 5' and 3' capping, 2'-O-methyl-purine substitution, polyethylene glycol-conjugation or packaging in microspheres may be required, as were necessary for the clinical application of the anti-VEGF aptamer “Pentagnib” (Ng *et al* 2006). Current investigations indicate that the aptamers are stable in samples from porcine vaginal and rectal swabs for at least 24 hours (unpublished work) and further collaborations are underway to investigate aptamer neutralisation of HIV-1 in human vaginal explant models.

9.7 Concluding remarks

In this thesis I have described the structural analysis of several novel anti-gp120 aptamers, the design of truncated, stabilised synthetic aptamers, and the delineation of the aptamer binding site of one such aptamer. These experiments have also highlighted aspects of the conformational flexibility of gp120, uncovered issues relating to the bioinformatic analysis of aptamer sequences, questioned the reliability of some previously well-used methods

and exposed flaws in the use of pseudoviruses as a model system. In the future it is likely that aptamers will have significant roles in both basic science and in therapeutic applications.

Bibliography

- Auvert, B., D. Taljaard, E. Lagarde, J. Sobngwi-Tambekou, R. Sitta and A. Puren (2005). "Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial." *PLoS Med* **2**(11): e298.
- Baba, M., O. Nishimura, N. Kanzaki, M. Okamoto, H. Sawada, Y. Iizawa, M. Shiraishi, Y. Aramaki, K. Okonogi, Y. Ogawa, K. Meguro and M. Fujino (1999). "A small-molecule, nonpeptide CCR5 antagonist with highly potent and selective anti-HIV-1 activity." *PNAS* **96**(10): 5698-5703.
- Baldrich, E., J. L. Acero, G. Reekmans, W. Laureyn and C. K. O'Sullivan (2005). "Displacement enzyme linked aptamer assay." *Anal Chem* **77**(15): 4774-84.
- Baldrich, E., A. Restrepo and C. K. O'Sullivan (2004). "Aptasensor development: elucidation of critical parameters for optimal aptamer performance." *Anal Chem* **76**(23): 7053-63.
- Barre-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dautet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum and L. Montagnier (1983). "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)." *Science* **220**(4599): 868-71.
- Barton, S., C. Celum and T. W. Shacker (2005). "The role of anti-HSV therapeutics in the HIV-infected host and in controlling the HIV epidemic." *Herpes* **12**(1): 15-22.
- Beaurain, F., C. Di Primo, J. J. Toulme and M. Laguerre (2003). "Molecular dynamics reveals the stabilizing role of loop closing residues in kissing interactions: comparison between TAR-TAR* and TAR-aptamer." *Nucl. Acids Res.* **31**(14): 4275-4284.
- Berger, E. A., R. W. Doms, E. M. Fenyo, B. T. Korber, D. R. Littman, J. P. Moore, Q. J. Sattentau, H. Schuitemaker, J. Sodroski and R. A. Weiss (1998). "A new classification for HIV-1." *Nature* **391**(6664): 240.
- Binley, J. M., T. Wrin, B. Korber, M. B. Zwick, M. Wang, C. Chappey, G. Stiegler, R. Kunert, S. Zolla-Pazner, H. Katinger, C. J. Petropoulos and D. R. Burton (2004). "Comprehensive Cross-Clade Neutralization Analysis of a Panel of Anti-Human Immunodeficiency Virus Type 1 Monoclonal Antibodies." *J. Virol.* **78**(23): 13232-13252.
- Blank, M., T. Weinschenk, M. Priemer and H. Schluesener (2001). "Systematic Evolution of a DNA Aptamer Binding to Rat Brain Tumor Microvessels. SELECTIVE TARGETING OF ENDOTHELIAL REGULATORY PROTEIN PIGPEN." *J. Biol. Chem.* **276**(19): 16464-16468.
- Bock, C., M. Coleman, B. Collins, J. Davis, G. Foulds, L. Gold, C. Greef, J. Heil, J. S. Heilig, B. Hicke, M. N. Hurst, G. M. Husar, D. Miller, R. Ostroff, H. Petach, D. Schneider, B. Vant-Hull, S. Waugh, A. Weiss,

- S. K. Wilcox and D. Zichi (2004). "Photoaptamer arrays applied to multiplexed proteomic analysis." *Proteomics* **4**(3): 609-18.
- Borrow, P., H. Lewicki, B. H. Hahn, G. M. Shaw and M. B. Oldstone (1994). "Virus-specific CD8⁺ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection." *J. Virol.* **68**(9): 6103-6110.
- Boyd, M. R., K. R. Gustafson, J. B. McMahon, R. H. Shoemaker, B. R. O'Keefe, T. Mori, R. J. Gulakowski, L. Wu, M. I. Rivera, C. M. Laurencot, M. J. Currens, J. H. Cardellina, 2nd, R. W. Buckheit, Jr., P. L. Nara, L. K. Pannell, R. C. Sowder, 2nd and L. E. Henderson (1997). "Discovery of cyanovirin-N, a novel human immunodeficiency virus- inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development." *Antimicrob. Agents Chemother.* **41**(7): 1521-1530.
- Briggs, J. A. G., M. N. Simon, I. Gross, H.-G. Krausslich, S. D. Fuller, V. M. Vogt and M. C. Johnson (2004). "The stoichiometry of Gag protein in HIV-1." *Nat Struct Mol Biol* **11**(7): 672.
- Brown, M. H. and A. N. Barclay (1994). "Expression of immunoglobulin and scavenger receptor superfamily domains as chimeric proteins with domains 3 and 4 of CD4 for ligand analysis." *Protein Eng.* **7**(4): 515-521.
- Brown, M. H., K. Boles, P. A. van der Merwe, V. Kumar, P. A. Mathew and A. N. Barclay (1998). "2B4, the natural killer and T cell immunoglobulin superfamily surface protein, is a ligand for CD48." *J Exp Med* **188**(11): 2083-90.
- Brown, P. O. (1997). Integration. *Retroviruses*. J. M. Coffin and V. H. E. Cold Spring Harbor, Cold Spring Harbor Laboratory Press: 161-203.
- Burton, D. R., R. C. Desrosiers, R. W. Doms, W. C. Koff, P. D. Kwong, J. P. Moore, G. J. Nabel, J. Sodroski, I. A. Wilson and R. T. Wyatt (2004). "HIV vaccine design and the neutralizing antibody problem." **5**(3): 233-236.
- Burton, D. R., J. Pyati, R. Koduri, S. J. Sharp, G. B. Thornton, P. W. Parren, L. S. Sawyer, R. M. Hendry, N. Dunlop, P. L. Nara and et al. (1994). "Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody." *Science* **266**(5187): 1024-7.
- Butler, S. L., M. S. Hansen and F. D. Bushman (2001). "A quantitative assay for HIV DNA integration in vivo." *Nat Med* **7**(5): 631-4.
- Carr, C. M. and P. S. Kim (1993). "A spring-loaded mechanism for the conformational change of influenza hemagglutinin." *Cell* **73**(4): 823.
- Centers for Disease Control, C. (1981). "Pneumocystis pneumonia--Los Angeles." *MMWR Morb Mortal Wkly Rep* **30**(21): 250-2.
- Chan, D. C., D. Fass, J. M. Berger and P. S. Kim (1997). "Core Structure of gp41 from the HIV Envelope Glycoprotein." *Cell* **89**(2): 263.
- Chan, D. C. and P. S. Kim (1998). "HIV Entry and Its Inhibition." *Cell* **93**(5): 681.

- Chelliserrykattil, J. and A. D. Ellington (2004). "Evolution of a T7 RNA polymerase variant that transcribes 2[prime]-O-methyl RNA." Nat Biotech **22**(9): 1155.
- Chen, B., E. M. Vogan, H. Gong, J. J. Skehel, D. C. Wiley and S. C. Harrison (2005). "Structure of an unliganded simian immunodeficiency virus gp120 core." Nature **433**(7028): 834-841.
- Chen, B. K., R. T. Gandhi and D. Baltimore (1996). "CD4 down-modulation during infection of human T cells with human immunodeficiency virus type 1 involves independent activities of vpu, env, and nef." J. Virol. **70**(9): 6044-6053.
- Cockett, M. I., C. R. Bebbington and G. T. Yarranton (1990). "High level expression of tissue inhibitor of metalloproteinases in Chinese hamster ovary cells using glutamine synthetase gene amplification." Biotechnology (N Y) **8**(7): 662-7.
- Collins, K. L., B. K. Chen, S. A. Kalams, B. D. Walker and D. Baltimore (1998). "HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes." Nature **391**(6665): 397.
- Connor, R. I., B. K. Chen, S. Choe and N. R. Landau (1995). "Vpr Is Required for Efficient Replication of Human Immunodeficiency Virus Type-1 in Mononuclear Phagocytes." Virology **206**(2): 935.
- Cooper, M. A. (2003). "Label-free screening of bio-molecular interactions." Anal Bioanal Chem **377**(5): 834-42.
- Cormier, E. G. and T. Dragic (2002). "The Crown and Stem of the V3 Loop Play Distinct Roles in Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Interactions with the CCR5 Coreceptor." J. Virol. **76**(17): 8953-8957.
- Cormier, E. G., M. Persuh, D. A. D. Thompson, S. W. Lin, T. P. Sakmar, W. C. Olson and T. Dragic (2000). "Specific interaction of CCR5 amino-terminal domain peptides containing sulfotyrosines with HIV-1 envelope glycoprotein gp120." PNAS **97**(11): 5762-5767.
- Cormier, E. G., D. N. H. Tran, L. Yukhayeva, W. C. Olson and T. Dragic (2001). "Mapping the Determinants of the CCR5 Amino-Terminal Sulfopeptide Interaction with Soluble Human Immunodeficiency Virus Type 1 gp120-CD4 Complexes." J. Virol. **75**(12): 5541-5549.
- Cox, J. C., A. Hayhurst, J. Hesselberth, T. S. Bayer, G. Georgiou and A. D. Ellington (2002). "Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer." Nucleic Acids Res **30**(20): e108.
- Dalglish, A. G., P. C. Beverley, P. R. Clapham, D. H. Crawford, M. F. Greaves and R. A. Weiss (1984). "The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus." Nature **312**(5996): 763-7.
- Darbha, R., S. Phogat, A. F. Labrijn, Y. Shu, Y. Gu, M. Andrykovitch, M. Y. Zhang, R. Pantophlet, L. Martin, C. Vita, D. R. Burton, D. S. Dimitrov and X. Ji (2004). "Crystal structure of the broadly cross-reactive HIV-

- 1-neutralizing Fab X5 and fine mapping of its epitope." Biochemistry **43**(6): 1410-7.
- de Parseval, A., M. D. Bobardt, A. Chatterji, U. Chatterji, J. H. Elder, G. David, S. Zolla-Pazner, M. Farzan, T.-H. Lee and P. A. Gally (2005). "A Highly Conserved Arginine in gp120 Governs HIV-1 Binding to Both Syndecans and CCR5 via Sulfated Motifs." J. Biol. Chem. **280**(47): 39493-39504.
- De Vincenzi, I. (1994). "A Longitudinal Study of Human Immunodeficiency Virus Transmission by Heterosexual Partners." N Engl J Med **331**(6): 341-346.
- Deacon, N. J., A. Tsykin, A. Solomon, K. Smith, M. Ludford-Menting, D. J. Hooker, D. A. McPhee, A. L. Greenway, A. Ellett, C. Chatfield, V. A. Lawson, S. Crowe, A. Maerz, S. Sonza, J. Learmont, J. S. Sullivan, A. Cunningham, D. Dwyer, D. Dowton and J. Mills (1995). "Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients." Science **270**(5238): 988-91.
- Deng, H., R. Liu, W. Ellmeier, S. Choe, D. Unutmaz, M. Burkhart, P. Di Marzio, S. Marmon, R. E. Sutton, C. M. Hill, C. B. Davis, S. C. Peiper, T. J. Schall, D. R. Littman and N. R. Landau (1996). "Identification of a major co-receptor for primary isolates of HIV-1." Nature **381**(6584): 661-6.
- Dey, A. K. (2004). Structural characterisation of RNA aptamers against the gp120 surface envelope glycoprotein of a macrophage tropic strain of HIV-1. Department of Biochemistry. Oxford, University of Oxford. **DPhil**.
- Dey, A. K., C. Griffiths, S. M. Lea and W. James (2005a). "Structural characterization of an anti-gp120 RNA aptamer that neutralizes R5 strains of HIV-1." Rna **11**(6): 873-84.
- Dey, A. K., M. Khati, M. Tang, R. Wyatt, S. M. Lea and W. James (2005b). "An Aptamer That Neutralizes R5 Strains of Human Immunodeficiency Virus Type 1 Blocks gp120-CCR5 Interaction." J. Virol. **79**(21): 13806-13810.
- Ding, Y., C. Y. Chan and C. E. Lawrence (2004). "Sfold web server for statistical folding and rational design of nucleic acids." Nucl. Acids Res. **32**(suppl_2): W135-141.
- Ding, Y. and C. E. Lawrence (1999). "A Bayesian statistical algorithm for RNA secondary structure prediction." Computers & Chemistry **23**(3-4): 387.
- Ding, Y. and C. E. Lawrence (2001). "Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond." Nucl. Acids Res. **29**(5): 1034-1046.
- Ding, Y. and C. E. Lawrence (2003). "A statistical sampling algorithm for RNA secondary structure prediction." Nucl. Acids Res. **31**(24): 7280-7301.

- Doms, R. W. and S. C. Peiper (1997). "Unwelcomed guests with master keys: how HIV uses chemokine receptors for cellular entry." Virology **235**(2): 179-90.
- Doms, R. W. and D. Trono (2000). "The plasma membrane as a combat zone in the HIV battlefield." Genes Dev **14**(21): 2677-88.
- Duconge, F. and J. J. Toulme (1999). "In vitro selection identifies key determinants for loop-loop interactions: RNA aptamers selective for the TAR RNA element of HIV-1." RNA **5**(12): 1605-1614.
- Durocher, Y., S. Perret and A. Kamen (2002). "High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells." Nucleic Acids Res **30**(2): E9.
- Eddy, S. R. (2004a). "How do RNA folding algorithms work?" Nat Biotechnol **22**(11): 1457-8.
- Eddy, S. R. (2004b). "What is dynamic programming?" **22**(7): 909-910.
- Ehresmann, C., F. Baudin, M. Mougél, P. Romby, J. P. Ebel and B. Ehresmann (1987). "Probing the structure of RNAs in solution." Nucleic Acids Res **15**(22): 9109-28.
- Ellington, A. D. and J. W. Szostak (1990). "In vitro selection of RNA molecules that bind specific ligands." Nature **346**(6287): 818-22.
- Ezzell, C. (1987). "AIDS drug gets green light." Nature **329**(6142): 751.
- Farley, T., P. Gaillard, I. de Vincenzi, C. Osborne and I. de Zoysa (2002). "Mother-to-child transmission of HIV-1 infection." The Lancet **360**(9349): 1974.
- Farzan, M., H. Choe, L. Vaca, K. Martin, Y. Sun, E. Desjardins, N. Ruffing, L. Wu, R. Wyatt, N. Gerard, C. Gerard and J. Sodroski (1998). "A tyrosine-rich region in the N terminus of CCR5 is important for human immunodeficiency virus type 1 entry and mediates an association between gp120 and CCR5." J Virol **72**(2): 1160-4.
- Farzan, M., N. Vasilieva, C. E. Schnitzler, S. Chung, J. Robinson, N. P. Gerard, C. Gerard, H. Choe and J. Sodroski (2000). "A Tyrosine-sulfated Peptide Based on the N Terminus of CCR5 Interacts with a CD4-enhanced Epitope of the HIV-1 gp120 Envelope Glycoprotein and Inhibits HIV-1 Entry." J. Biol. Chem. **275**(43): 33516-33521.
- Feng, D. F. and R. F. Doolittle (1987). "Progressive sequence alignment as a prerequisite to correct phylogenetic trees." J Mol Evol **25**(4): 351-60.
- Feng, Y., C. C. Broder, P. E. Kennedy and E. A. Berger (1996). "HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor." Science **272**(5263): 872-7.
- Finzi, D., J. Blankson, J. D. Siliciano, J. B. Margolick, K. Chadwick, T. Pierson, K. Smith, J. Lisziewicz, F. Lori, C. Flexner, T. C. Quinn, R. E. Chaisson, E. Rosenberg, B. Walker, S. Gange, J. Gallant and R. F. Siliciano (1999). "Latent infection of CD4 T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy." Nature Medicine **5**(5): 512.

- Fitzwater, T. and B. Polisky (1996). "A SELEX primer." Methods Enzymol **267**: 275-301.
- Freed, E. O. (1998). "HIV-1 Gag Proteins: Diverse Functions in the Virus Life Cycle." Virology **251**(1): 1.
- Freed, E. O. (2001). "HIV-1 replication." Somat Cell Mol Genet **26**(1-6): 13-33.
- Freed, E. O. (2004). "HIV-1 and the host cell: an intimate association." Trends in Microbiology **12**(4): 170-177.
- Freed, E. O. and M. A. Martin (1995). "The Role of Human Immunodeficiency Virus Type 1 Envelope Glycoproteins in Virus Infection." J. Biol. Chem. **270**(41): 23883-23886.
- Freier, S. M., R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson and D. H. Turner (1986). "Improved Free-Energy Parameters for Predictions of RNA Duplex Stability." PNAS **83**(24): 9373-9377.
- Fukuda, K., D. Vishnuvardhan, S. Sekiya, J. Hwang, N. Kakiuchi, K. Taira, K. Shimotohno, P. K. Kumar and S. Nishikawa (2000). "Isolation and characterization of RNA aptamers specific for the hepatitis C virus nonstructural protein 3 protease." Eur J Biochem **267**(12): 3685-94.
- Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp and B. H. Hahn (1999). "Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes." Nature **397**(6718): 436.
- Gardner, P. P. and R. Giegerich (2004). "A comprehensive comparison of comparative RNA structure prediction approaches." BMC Bioinformatics **5**: 140.
- Geiger, A., P. Burgstaller, H. von der Eltz, A. Roeder and M. Famulok (1996). "RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity." Nucl. Acids Res. **24**(6): 1029-1036.
- Geijtenbeek, T. B. H., D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. F. van Duijnhoven, J. Middel, I. L. M. H. A. Cornelissen, H. S. L. M. Nottet, V. N. KewalRamani and D. R. Littman (2000). "DC-SIGN, a Dendritic Cell-Specific HIV-1-Binding Protein that Enhances trans-Infection of T Cells." Cell **100**(5): 587.
- Godbey, W. T., K. K. Wu and A. G. Mikos (1999). "Poly(ethylenimine) and its role in gene delivery." J Control Release **60**(2-3): 149-60.
- Golden, B. L. and C. E. Kundrot (2003). "RNA crystallization." Journal of Structural Biology **142**(1): 98.
- Gorman, C. (1985). High efficiency gene transfer into mammalian cells. DNA Cloning: A practical approach. D. M. Glover. Oxford, IRL Press. **II**: 143-190.
- Gorny, M. K., A. J. Conley, S. Karwowska, A. Buchbinder, J. Y. Xu, E. A. Emini, S. Koenig and S. Zolla-Pazner (1992). "Neutralization of diverse human immunodeficiency virus type 1 variants by an anti-V3 human monoclonal antibody." J Virol **66**(12): 7538-42.

- Gorny, M. K. and S. Zolla-Pazner (2003). Human Monoclonal Antibodies that Neutralize HIV-1. HIV Immunology and HIV/SIV vaccine database: 37-51.
- Green, L. S., C. Bell and N. Janjic (2001). "Aptamers as reagents for high-throughput screening." Biotechniques **30**(5): 1094-6, 1098, 1100 passim.
- Greene, W. C. and B. M. Peterlin (2002). "Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy." Nat Med **8**(7): 673-80.
- Haas, J., E. C. Park and B. Seed (1996). "Codon usage limitation in the expression of HIV-1 envelope glycoprotein." Curr Biol **6**(3): 315-24.
- Havgaard, J. H., R. Lyngso, G. D. Stormo and J. Gorodkin (2005a). "Pairwise local structural alignment of RNA sequences with sequence similarity less than 40%." Bioinformatics: bti279.
- Havgaard, J. H., R. B. Lyngso and J. Gorodkin (2005b). "The FOLDALIGN web server for pairwise structural RNA alignment and mutual motif search." Nucl. Acids Res. **33**(suppl_2): W650-653.
- Heid, C. A., J. Stevens, K. J. Livak and P. M. Williams (1996). "Real time quantitative PCR." Genome Res **6**(10): 986-94.
- Helseth, E., M. Kowalski, D. Gabuzda, U. Olshevsky, W. Haseltine and J. Sodroski (1990). "Rapid complementation assays measuring replicative potential of human immunodeficiency virus type 1 envelope glycoprotein mutants." J Virol **64**(5): 2416-20.
- Hillier, S. L., T. Moench, R. Shattock, R. Black, P. Reichelderfer and F. Veronese (2005). "In vitro and in vivo: the story of nonoxynol 9." J Acquir Immune Defic Syndr **39**(1): 1-8.
- Hochsmann, M., T. Toller, R. Giegerich and S. Kurtz (2003). "Local similarity in RNA secondary structures." Proc IEEE Comput Soc Bioinform Conf **2**: 159-68.
- Hochsmann, M., B. Voss and R. Giegerich (2004). "Pure multiple RNA secondary structure alignments: A progressive profile approach." IEEE/ACM Transactions on Computational Biology and Bioinformatics **1**(1): 53-62.
- Hofacker, I. L. (2003). "Vienna RNA secondary structure server." Nucl. Acids Res. **31**(13): 3429-3431.
- Hofacker, I. L., W. Fontana, P. F. Stadler and P. Schuster (1994). "Fast Folding and Comparison of RNA Secondary Structures." Monatshefte fur Chemie **125**: 167-188.
- Hu, J., M. B. Gardner and C. J. Miller (2000). "Simian Immunodeficiency Virus Rapidly Penetrates the Cervicovaginal Mucosa after Intravaginal Inoculation and Infects Intraepithelial Dendritic Cells." J. Virol. **74**(13): 6087-6095.
- Huang, C.-c., M. Tang, M.-Y. Zhang, S. Majeed, E. Montabana, R. L. Stanfield, D. S. Dimitrov, B. Korber, J. Sodroski, I. A. Wilson, R. Wyatt and P. D. Kwong (2005). "Structure of a V3-Containing HIV-1 gp120 Core." Science **310**(5750): 1025-1028.

- Humphries, D. E. and J. E. Silbert (1988). "Chlorate: a reversible inhibitor of proteoglycan sulfation." Biochem Biophys Res Commun **154**(1): 365-71.
- Ibrahim, J., P. Griffin, D. R. Coombe, C. C. Rider and W. James (1999). "Cell-surface heparan sulfate facilitates human immunodeficiency virus Type 1 entry into some cell lines but not primary lymphocytes." Virus Res **60**(2): 159-69.
- Iribarren, A. M., B. S. Sproat, P. Neuner, I. Sulston, U. Ryder and A. I. Lamond (1990). "2'-O-Alkyl Oligoribonucleotides as Antisense Probes." PNAS **87**(19): 7747-7751.
- James, W. (2001). "Nucleic acid and polypeptide aptamers: a powerful approach to ligand discovery." Curr Opin Pharmacol **1**(5): 540-6.
- James, W. (2006). "Aptamers in the virologists' tool kit." Journal of General Virology **In press**.
- Kaufmann, G. R. and D. A. Cooper (2000). "Antiretroviral therapy of HIV-1 infection: established treatment strategies and new therapeutic options." Curr Opin Microbiol **3**(5): 508-14.
- Khati, M. (2002). Macrophage-HIV interactions: Aptamers against the gp120 surface envelope glycoprotein of the macrophage tropic strains of HIV-1. Sir William Dunn School of Pathology. Oxford, University of Oxford. **DPhil**.
- Khati, M., M. Schuman, J. Ibrahim, Q. Sattentau, S. Gordon and W. James (2003). "Neutralization of infectivity of diverse R5 clinical isolates of human immunodeficiency virus type 1 by gp120-binding 2'F-RNA aptamers." J Virol **77**(23): 12692-8.
- Klasse, P. J., R. Shattock and J. Moore (2006). "Which Topical Microbicides for Blocking HIV-1 Transmission Will Work in the Real World?" PLOS Medicine **3**(9): e351.
- Koizumi, M., G. A. Soukup, J. N. Q. Kerr and R. R. Breaker (1999). "Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP." Nat Struct Mol Biol **6**(11): 1062.
- Kovalevsky, A. Y., Y. Tie, F. Liu, P. I. Boross, Y. F. Wang, S. Leshchenko, A. K. Ghosh, R. W. Harrison and I. T. Weber (2006). "Effectiveness of nonpeptide clinical inhibitor TMC-114 on HIV-1 protease with highly drug resistant mutations D30N, I50V, and L90M." J Med Chem **49**(4): 1379-87.
- Kraus, E., W. James and A. N. Barclay (1998). "Cutting edge: novel RNA ligands able to bind CD4 antigen and inhibit CD4+ T lymphocyte function." J Immunol **160**(11): 5209-12.
- Kremer, M. and B. S. Schnierle (2005). "HIV-1 Vif: HIV's weapon against the cellular defense factor APOBEC3G." Curr HIV Res **3**(4): 339-44.
- Kruskal, J. B. and M. Wish (1977). Multidimensional scaling.
- Kunert, R., F. Ruker and H. Katinger (1998). "Molecular characterization of five neutralizing anti-HIV type 1 antibodies: identification of nonconventional D segments in the human monoclonal antibodies 2G12 and 2F5." AIDS Res Hum Retroviruses **14**(13): 1115-28.

- Kuznetsov, Y. G., J. G. Victoria, W. E. Robinson, Jr. and A. McPherson (2003). "Atomic Force Microscopy Investigation of Human Immunodeficiency Virus (HIV) and HIV-Infected Lymphocytes." J. Virol. **77**(22): 11896-11909.
- Kwon, D. S., G. Gregorio, N. Bitton, W. A. Hendrickson and D. R. Littman (2002). "DC-SIGN-Mediated Internalization of HIV Is Required for Trans-Enhancement of T Cell Infection." Immunity **16**(1): 135.
- Kwong, P. D., M. L. Doyle, D. J. Casper, C. Cicala, S. A. Leavitt, S. Majeed, T. D. Steenbeke, M. Venturi, I. Chaiken, M. Fung, H. Katinger, P. W. Parren, J. Robinson, D. Van Ryk, L. Wang, D. R. Burton, E. Freire, R. Wyatt, J. Sodroski, W. A. Hendrickson and J. Arthos (2002). "HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites." Nature **420**(6916): 678-82.
- Kwong, P. D., R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski and W. A. Hendrickson (1998). "Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody." Nature **393**(6686): 648-59.
- Kwong, P. D., R. Wyatt, Q. J. Sattentau, J. Sodroski and W. A. Hendrickson (2000). "Oligomeric Modeling and Electrostatic Analysis of the gp120 Envelope Glycoprotein of Human Immunodeficiency Virus." J. Virol. **74**(4): 1961-1972.
- Labrijn, A. F., P. Poignard, A. Raja, M. B. Zwick, K. Delgado, M. Franti, J. Binley, V. Vivona, C. Grundner, C. C. Huang, M. Venturi, C. J. Petropoulos, T. Wrin, D. S. Dimitrov, J. Robinson, P. D. Kwong, R. T. Wyatt, J. Sodroski and D. R. Burton (2003). "Access of antibody molecules to the conserved coreceptor binding site on glycoprotein gp120 is sterically restricted on primary human immunodeficiency virus type 1." J Virol **77**(19): 10557-65.
- Lalezari, J. P., K. Henry, M. O'Hearn, J. S. G. Montaner, P. J. Piliero, B. Trottier, S. Walmsley, C. Cohen, D. R. Kuritzkes, J. J. Eron, Jr., J. Chung, R. DeMasi, L. Donatucci, C. Drobnes, J. Delehanty, M. Salgo and T. S. G. the (2003). "Enfuvirtide, an HIV-1 Fusion Inhibitor, for Drug-Resistant HIV Infection in North and South America." N Engl J Med **348**(22): 2175-2185.
- Lamond, A. I. and B. S. Sproat (1993). "Antisense oligonucleotides made of 2'-O-alkylRNA: their properties and applications in RNA biochemistry." FEBS Letters **325**(1-2): 123.
- Latham, M. P., D. J. Brown, S. A. McCallum and A. Pardi (2005). "NMR methods for studying the structure and dynamics of RNA." Chembiochem **6**(9): 1492-505.
- Lazzarin, A., B. Clotet, D. Cooper, J. Reynes, K. Arasteh, M. Nelson, C. Katlama, H.-J. Stellbrink, J.-F. Delfraissy, J. Lange, L. Huson, R. DeMasi, C. Wat, J. Delehanty, C. Drobnes, M. Salgo and T. S. G. the (2003). "Efficacy of Enfuvirtide in Patients Infected with Drug-Resistant HIV-1 in Europe and Australia." N Engl J Med **348**(22): 2186-2195.

- Lederman, M. M., R. E. Offord and O. Hartley (2006). "Microbicides and other topical strategies to prevent vaginal transmission of HIV." Nat Rev Immunol **6**(5): 371.
- Lederman, M. M., R. S. Veazey, R. Offord, D. E. Mosier, J. Dufour, M. Mefford, M. Piatak, Jr., J. D. Lifson, J. R. Salkowitz, B. Rodriguez, A. Blauvelt and O. Hartley (2004). "Prevention of Vaginal SHIV Transmission in Rhesus Macaques Through Inhibition of CCR5." Science **306**(5695): 485-487.
- Li, L., H. S. Li, C. D. Pauza, M. Bukrinsky and R. Y. Zhao (2005). "Roles of HIV-1 auxiliary proteins in viral pathogenesis and host-pathogen interactions." Cell Res **15**(11-12): 923.
- Lin, C. L., A. K. Sewell, G. F. Gao, K. T. Whelan, R. E. Phillips and J. M. Austyn (2000). "Macrophage-tropic HIV induces and exploits dendritic cell chemotaxis." J Exp Med **192**(4): 587-94.
- Liu, J., J. T. Wang, J. Hu and B. Tian (2005). "A method for aligning RNA secondary structures and its application to RNA motif detection." BMC Bioinformatics **6**(1): 89.
- Liu, R., W. A. Paxton, S. Choe, D. Ceradini, S. R. Martin, R. Horuk, M. E. MacDonald, H. Stuhlmann, R. A. Koup and N. R. Landau (1996). "Homozygous Defect in HIV-1 Coreceptor Accounts for Resistance of Some Multiply-Exposed Individuals to HIV-1 Infection." Cell **86**(3): 367.
- Losick, R. (1972). "In vitro transcription." Annu Rev Biochem **41**: 409-46.
- Mack, M., B. Luckow, P. J. Nelson, J. Cihak, G. Simmons, P. R. Clapham, N. Signoret, M. Marsh, M. Stangassinger, F. Borlat, T. N. C. Wells, D. Schlondorff and A. E. I. Proudfoot (1998). "Aminooxypentane-RANTES Induces CCR5 Internalization but Inhibits Recycling: A Novel Inhibitory Mechanism of HIV Infectivity." J. Exp. Med. **187**(8): 1215-1224.
- Malcolm, R. K., A. D. Woolfson, C. F. Toner, R. J. Morrow and S. D. McCullagh (2005). "Long-term, controlled release of the HIV microbicide TMC120 from silicone elastomer vaginal rings." J. Antimicrob. Chemother. **56**(5): 954-956.
- Mandal, M. and R. R. Breaker (2004). "Adenine riboswitches and gene activation by disruption of a transcription terminator." Nat Struct Mol Biol **11**(1): 29-35.
- Mannironi, C., A. Di Nardo, P. Fruscoloni and G. P. Tocchini-Valentini (1997). "In vitro selection of dopamine RNA ligands." Biochemistry **36**(32): 9726-34.
- Mansky, L. M. and H. M. Temin (1995). "Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase." J. Virol. **69**(8): 5087-5094.
- Marro, M. L., D. A. Daniels, A. McNamee, D. P. Andrew, T. D. Chapman, M. S. Jiang, Z. Wu, J. L. Smith, K. K. Patel and K. L. Gearing (2005). "Identification of potent and selective RNA antagonists of the IFN-

- gamma-inducible CXCL10 chemokine." Biochemistry **44**(23): 8449-60.
- Mascola, J. R., P. D'Souza, P. Gilbert, B. H. Hahn, N. L. Haigwood, L. Morris, C. J. Petropoulos, V. R. Polonis, M. Sarzotti and D. C. Montefiori (2005). "Recommendations for the Design and Use of Standard Virus Panels To Assess Neutralizing Antibody Responses Elicited by Candidate Human Immunodeficiency Virus Type 1 Vaccines." J. Virol. **79**(16): 10103-10107.
- Mascola, J. R., G. Stiegler, T. C. VanCott, H. Katinger, C. B. Carpenter, C. E. Hanson, H. Beary, D. Hayes, S. S. Frankel, D. L. Birx and M. G. Lewis (2000). "Protection of macaques against vaginal transmission of a pathogenic HIV-1/SIV chimeric virus by passive infusion of neutralizing antibodies." Nat Med **6**(2): 207.
- Mathews, D. H. (2005). "Predicting a set of minimal free energy RNA secondary structures common to two sequences." Bioinformatics **21**(10): 2246-2253.
- Mathews, D. H., J. Sabina, M. Zuker and D. H. Turner (1999). "Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure." Journal of Molecular Biology **288**(5): 911-940.
- Mathews, D. H. and D. H. Turner (2002). "Dyalign: an algorithm for finding the secondary structure common to two RNA sequences." Journal of Molecular Biology **317**(2): 191.
- Mayer, K. H., J. Peipert, T. Fleming, A. Fullem, T. Moench, S. Cu-Uvin, M. Bentley, M. Chesney and Z. Rosenberg (2001). "Safety and tolerability of BufferGel, a novel vaginal microbicide, in women in the United States." Clin Infect Dis **32**(3): 476-82.
- McMichael, A. (1998). "T Cell Responses and Viral Escape." Cell **93**(5): 673.
- McMichael, A. J. (2006). "HIV vaccines." Annu Rev Immunol **24**: 227-55.
- Mizushima, S. and S. Nagata (1990). "pEF-BOS, a powerful mammalian expression vector." Nucl. Acids Res. **18**(17): 5322-.
- Morris, K. V. and J. J. Rossi (2006). "Lentivirus-mediated RNA interference therapy for human immunodeficiency virus type 1 infection." Hum Gene Ther **17**(5): 479-86.
- Mothe, B. R., H. Horton, D. K. Carter, T. M. Allen, M. E. Liebl, P. Skinner, T. U. Vogel, S. Fuenger, K. Vielhuber, W. Rehrauer, N. Wilson, G. Franchini, J. D. Altman, A. Haase, L. J. Picker, D. B. Allison and D. I. Watkins (2002). "Dominance of CD8 Responses Specific for Epitopes Bound by a Single Major Histocompatibility Complex Class I Molecule during the Acute Phase of Viral Infection." J. Virol. **76**(2): 875-884.
- Moulard, M., H. Lortat-Jacob, I. Mondor, G. Roca, R. Wyatt, J. Sodroski, L. Zhao, W. Olson, P. D. Kwong and Q. J. Sattentau (2000). "Selective Interactions of Polyanions with Basic Surfaces on Human Immunodeficiency Virus Type 1 gp120." J. Virol. **74**(4): 1948-1960.

- Moulard, M., S. K. Phogat, Y. Shu, A. F. Labrijn, X. Xiao, J. M. Binley, M. Y. Zhang, I. A. Sidorov, C. C. Broder, J. Robinson, P. W. Parren, D. R. Burton and D. S. Dimitrov (2002). "Broadly cross-reactive HIV-1-neutralizing human monoclonal Fab selected for binding to gp120-CD4-CCR5 complexes." Proc Natl Acad Sci U S A **99**(10): 6913-8.
- Ng, E. W., D. T. Shima, P. Calias, E. T. Cunningham, Jr., D. R. Guyer and A. P. Adamis (2006). "Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease." Nat Rev Drug Discov **5**(2): 123-32.
- Nguyen, K. L., M. Ilano, H. Akari, E. Miyagi, E. M. Poeschla, K. Strebel and S. Bour (2004). "Codon optimization of the HIV-1 vpu and vif genes stabilizes their mRNA and allows for highly efficient Rev-independent expression." Virology **319**(2): 163-75.
- Nimjee, S. M., J. R. Keys, G. A. Pitoc, G. Quick, C. P. Rusconi and B. A. Sullenger (2006). "A Novel Antidote-Controlled Anticoagulant Reduces Thrombin Generation and Inflammation and Improves Cardiac Function in Cardiopulmonary Bypass Surgery." Molecular Therapy **14**(3): 408.
- Nimjee, S. M., C. P. Rusconi and B. A. Sullenger (2005). "Aptamers: an emerging class of therapeutics." Annu Rev Med **56**: 555-83.
- Nutiu, R. and Y. Li (2005). "Aptamers with fluorescence-signaling properties." Methods **37**(1): 16.
- Pantaleo, G., C. Graziosi and A. S. Fauci (1993). "The Immunopathogenesis of Human Immunodeficiency Virus Infection." N Engl J Med **328**(5): 327-335.
- Pantophlet, R. and D. R. Burton (2006). "GP120: target for neutralizing HIV-1 antibodies." Annu Rev Immunol **24**: 739-69.
- Pantophlet, R., E. Ollmann Saphire, P. Poignard, P. W. Parren, I. A. Wilson and D. R. Burton (2003). "Fine mapping of the interaction of neutralizing and nonneutralizing monoclonal antibodies with the CD4 binding site of human immunodeficiency virus type 1 gp120." J Virol **77**(1): 642-58.
- Parren, P. W., J. P. Moore, D. R. Burton and Q. J. Sattentau (1999). "The neutralizing antibody response to HIV-1: viral evasion and escape from humoral immunity." Aids **13 Suppl A**: S137-62.
- Pastore, C., A. Ramos and D. E. Mosier (2004). "Intrinsic Obstacles to Human Immunodeficiency Virus Type 1 Coreceptor Switching." J. Virol. **78**(14): 7565-7574.
- Patel, D. J. (1997). "Structural analysis of nucleic acid aptamers." Current Opinion in Chemical Biology **1**(1): 32.
- Pedersen, J. S., I. M. Meyer, R. Forsberg, P. Simmonds and J. Hein (2004). "A comparative method for finding and folding RNA secondary structures within protein-coding regions." Nucl. Acids Res. **32**(16): 4925-4936.
- Phillips, A. N. (1996). "Reduction of HIV concentration during acute infection: independence from a specific immune response." Science **271**(5248): 497-9.

- Pieken, W. A., D. B. Olsen, H. Aurup, D. M. Williams, O. Heidenreich, F. Benseler and F. Eckstein (1991). "Structure-function relationship of hammerhead ribozymes as probed by 2'-modifications." Nucleic Acids Symp Ser(24): 51-3.
- Piguet, V. and Q. Sattentau (2004). "Dangerous liaisons at the virological synapse." J. Clin. Invest. **114**(5): 605-610.
- Pomerantz, R. J. (2002). "Reservoirs of human immunodeficiency virus type 1: the main obstacles to viral eradication." Clin Infect Dis **34**(1): 91-7.
- Price, D. A., P. J. R. Goulder, P. Klenerman, A. K. Sewell, P. J. Easterbrook, M. Troop, C. R. M. Bangham and R. E. Phillips (1997). "Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection." PNAS **94**(5): 1890-1895.
- Princen, K., S. Hatse, K. Vermeire, E. De Clercq and D. Schols (2004). "Establishment of a novel CCR5 and CXCR4 expressing CD4+ cell line which is highly sensitive to HIV and suitable for high-throughput evaluation of CCR5 and CXCR4 antagonists." Retrovirology **1**(1): 2.
- Proske, D., M. Blank, R. Buhmann and A. Resch (2005). "Aptamers--basic research, drug development, and clinical applications." Appl Microbiol Biotechnol **69**(4): 367-74.
- Purschke, W. G., D. Eulberg, K. Buchner, S. Vonhoff and S. Klussmann (2006). "An L-RNA-based aquaretic agent that inhibits vasopressin in vivo." PNAS **103**(13): 5173-5178.
- Quinn, T. C., M. J. Wawer, N. Sewankambo, D. Serwadda, C. Li, F. Wabwire-Mangen, M. O. Meehan, T. Lutalo, R. H. Gray and G. The Rakai Project Study (2000). "Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1." N Engl J Med **342**(13): 921-929.
- Rao, S., S. Hu, L. McHugh, K. Lueders, K. Henry, Q. Zhao, R. A. Fekete, S. Kar, S. Adhya and D. H. Hamer (2005). "From the Cover: Toward a live microbial microbicide for HIV: Commensal bacteria secreting an HIV fusion inhibitor peptide." PNAS **102**(34): 11993-11998.
- Reeves, J. D., S. Hibbitts, G. Simmons, A. McKnight, J. M. Azevedo-Pereira, J. Moniz-Pereira and P. R. Clapham (1999). "Primary Human Immunodeficiency Virus Type 2 (HIV-2) Isolates Infect CD4-Negative Cells via CCR5 and CXCR4: Comparison with HIV-1 and Simian Immunodeficiency Virus and Relevance to Cell Tropism In Vivo." J. Virol. **73**(9): 7795-7804.
- Rhie, A., L. Kirby, N. Sayer, R. Wellesley, P. Disterer, I. Sylvester, A. Gill, J. Hope, W. James and A. Tahiri-Alaoui (2003). "Characterization of 2'-Fluoro-RNA Aptamers That Bind Preferentially to Disease-associated Conformations of Prion Protein and Inhibit Conversion." J. Biol. Chem. **278**(41): 39697-39705.
- Rhodes, A., A. Deakin, J. Spaul, B. Coomber, A. Aitken, P. Life and S. Rees (2000). "The Generation and Characterization of Antagonist RNA Aptamers to Human Oncostatin M." J. Biol. Chem. **275**(37): 28555-28561.

- Rivas, E. and S. R. Eddy (1999). "A dynamic programming algorithm for RNA structure prediction including pseudoknots." Journal of Molecular Biology **285**(5): 2053-2068.
- Rizzuto, C. and J. Sodroski (2000). "Fine definition of a conserved CCR5-binding region on the human immunodeficiency virus type 1 glycoprotein 120." AIDS Res Hum Retroviruses **16**(8): 741-9.
- Rizzuto, C. D., R. Wyatt, N. Hernandez-Ramos, Y. Sun, P. D. Kwong, W. A. Hendrickson and J. Sodroski (1998). "A Conserved HIV gp120 Glycoprotein Structure Involved in Chemokine Receptor Binding." Science **280**(5371): 1949-1953.
- Roben, P., J. P. Moore, M. Thali, J. Sodroski, C. F. Barbas, 3rd and D. R. Burton (1994). "Recognition properties of a panel of human recombinant Fab fragments to the CD4 binding site of gp120 that show differing abilities to neutralize human immunodeficiency virus type 1." J. Virol. **68**(8): 4821-4828.
- Rusconi, C. P., J. D. Roberts, G. A. Pitoc, S. M. Nimjee, R. R. White, G. Quick, E. Scardino, W. P. Fay and B. A. Sullenger (2004). "Antidote-mediated control of an anticoagulant aptamer in vivo." Nat Biotechnol **22**(11): 1423-1428.
- Rusconi, C. P., E. Scardino, J. Layzer, G. A. Pitoc, T. L. Ortel, D. Monroe and B. A. Sullenger (2002). "RNA aptamers as reversible antagonists of coagulation factor IXa." Nature **419**(6902): 90.
- Safrit, J. T., C. A. Andrews, T. Zhu, D. D. Ho and R. A. Koup (1994). "Characterization of human immunodeficiency virus type 1-specific cytotoxic T lymphocyte clones isolated during acute seroconversion: recognition of autologous virus sequences within a conserved immunodominant epitope." J. Exp. Med. **179**(2): 463-472.
- Sankoff, D. (1985). "Simultaneous Solution of the RNA Folding, Alignment and Protosequence Problems." SIAM Journal on Applied Mathematics **45**(5): 810-825.
- Saphire, E. O., P. W. Parren, R. Pantophlet, M. B. Zwick, G. M. Morris, P. M. Rudd, R. A. Dwek, R. L. Stanfield, D. R. Burton and I. A. Wilson (2001). "Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design." Science **293**(5532): 1155-9.
- Sassanfar, M. and J. W. Szostak (1993). "An RNA motif that binds ATP." Nature **364**(6437): 550.
- Sayer, N. (2004). The characterisation of aptamers that bind to prion protein or glycoprotein 120. Sir William Dunn School of Pathology. Oxford, University of Oxford. **DPhil**.
- Sayer, N., J. Ibrahim, K. Turner, A. Tahiri-Alaoui and W. James (2002). "Structural characterization of a 2'F-RNA aptamer that binds a HIV-1 SU glycoprotein, gp120." Biochem Biophys Res Commun **293**(3): 924-31.
- Sayer, N. M., M. Cubin, A. Rhie, M. Bullock, A. Tahiri-Alaoui and W. James (2004). "Structural determinants of conformationally selective, prion-binding aptamers." J Biol Chem **279**(13): 13102-9.

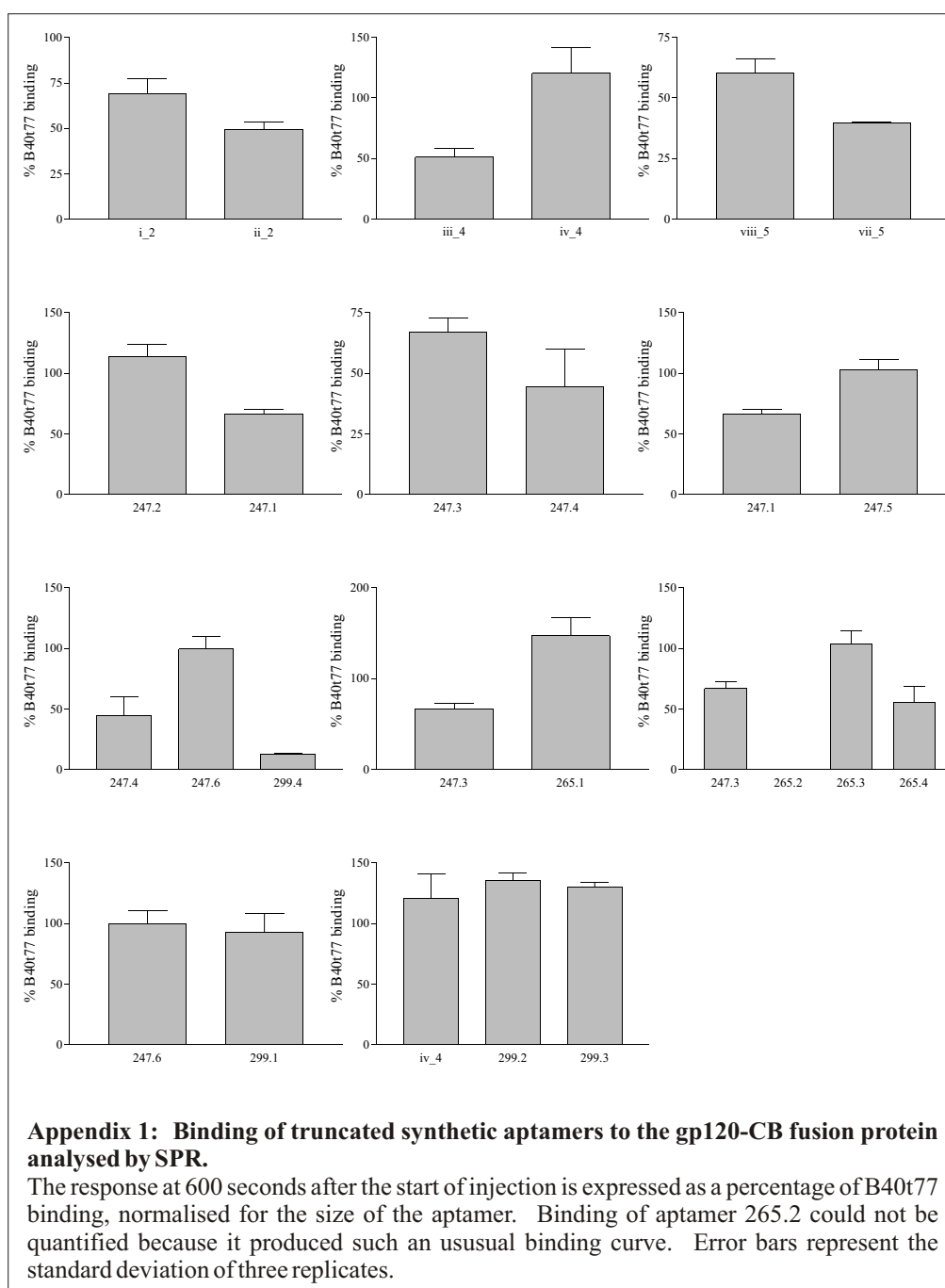
- Sazani, P. L., R. Larralde and J. W. Szostak (2004). "A small aptamer with strong and specific recognition of the triphosphate of ATP." J Am Chem Soc **126**(27): 8370-1.
- Scanlan, C. N., R. Pantophlet, M. R. Wormald, E. Ollmann Saphire, R. Stanfield, I. A. Wilson, H. Katinger, R. A. Dwek, P. M. Rudd and D. R. Burton (2002). "The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of alpha1-->2 mannose residues on the outer face of gp120." J Virol **76**(14): 7306-21.
- Scarlatti, G., E. Tresoldi, A. Bjorndal, R. Fredriksson, C. Colognesi, H. K. Deng, M. S. Malnati, A. Plebani, A. G. Siccardi, D. R. Littman, E. M. Fenyo and P. Lusso (1997). "In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression." Nat Med **3**(11): 1259-65.
- Schatz, P. J. (1993). "Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli." Biotechnology (N Y) **11**(10): 1138-43.
- Schuitmaker, H., M. Koot, N. A. Kootstra, M. W. Dercksen, R. E. de Goede, R. P. van Steenwijk, J. M. Lange, J. K. Schattenkerk, F. Miedema and M. Tersmette (1992). "Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population." J. Virol. **66**(3): 1354-1360.
- Seiwert, S. D., T. Stines Nahreini, S. Aigner, N. G. Ahn and O. C. Uhlenbeck (2000). "RNA aptamers as pathway-specific MAP kinase inhibitors." Chemistry & Biology **7**(11): 833.
- Si, Z., N. Madani, J. M. Cox, J. J. Chruma, J. C. Klein, A. Schon, N. Phan, L. Wang, A. C. Biorn, S. Cocklin, I. Chaiken, E. Freire, A. B. Smith, III and J. G. Sodroski (2004). "Small-molecule inhibitors of HIV-1 entry block receptor-induced conformational changes in the viral envelope glycoproteins." PNAS **101**(14): 5036-5041.
- Siegfried, N. (2005). "Does male circumcision prevent HIV infection?" PLoS Med **2**(11): e393.
- Sproat, B. S. (1993). "Chemical nucleic acid synthesis, modification and labelling." Current Opinion in Biotechnology **4**(1): 20.
- Sproat, B. S. and A. I. Lamond (1993). 2'-O-alkyloligoribonucleotides. Antisense Research and Applications. S. T. Crooke and B. Lebleu, CRC Press: 351-362.
- Stanfield, R., M. Gorny, C. Williams, S. Zolla-Pazner and I. A. Wilson (2004). "Structural Rationale for the Broad Neutralization of HIV-1 by Human Monoclonal Antibody 447-52D." Structure **12**: 193-204.
- Starcich, B. R., B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. Wolf, E. S. Parks, W. P. Parks, S. F. Josephs, R. C. Gallo and et al. (1986). "Identification and characterization of conserved and variable

- regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS." Cell **45**(5): 637-48.
- Stephens, P. E. and M. I. Cockett (1989). "The construction of a highly efficient and versatile set of mammalian expression vectors." Nucleic Acids Res **17**(17): 7110.
- Tahiri-Alaoui, A., L. Frigotto, N. Manville, J. Ibrahim, P. Romby and W. James (2002). "High affinity nucleic acid aptamers for streptavidin incorporated into bi-specific capture ligands." Nucl. Acids Res. **30**(10): e45-.
- Telesnitsky, A. and S. P. Goff (1997). Reverse transcription and the generation of retroviral DNA. Retroviruses. J. M. Coffin, H. S. H. and V. H. E. Cold Spring Harbor, Cold Spring Harbor Laboratory Press: 121-160.
- Thali, M., J. P. Moore, C. Furman, M. Charles, D. D. Ho, J. Robinson and J. Sodroski (1993). "Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding." J Virol **67**(7): 3978-88.
- Thompson, J., D. Higgins and T. Gibson (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucl. Acids Res. **22**(22): 4673-4680.
- Thomson, M. M., L. Perez-Alvarez and R. Najera (2002). "Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy." The Lancet Infectious Diseases **2**(8): 461.
- Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levin, O. C. Uhlenbeck, D. M. Crothers and J. Bralla (1973). "Improved estimation of secondary structure in ribonucleic acids." Nat New Biol **246**(150): 40-1.
- Trkola, A., M. Purtscher, T. Muster, C. Ballaun, A. Buchacher, N. Sullivan, K. Srinivasan, J. Sodroski, J. P. Moore and H. Katinger (1996). "Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1." J. Virol. **70**(2): 1100-1108.
- Tuerk, C. and L. Gold (1990). "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science **249**(4968): 505-10.
- UNAIDS/WHO (2005). "UNAIDS/WHO AIDS epidemic update 2005."
- Van der Merwe, P. A. (2001). Surface Plasmon Resonance. Protein-Ligand interactions: hydrodynamics and calorimetry. S. Harding and P. Z. Chowdhry. Oxford, Oxford University Press: 137-170.
- Veazey, R. S., P. J. Klasse, T. J. Ketas, J. D. Reeves, M. Piatak, Jr., K. Kunstman, S. E. Kuhmann, P. A. Marx, J. D. Lifson, J. Dufour, M. Mefford, I. Pandrea, S. M. Wolinsky, R. W. Doms, J. A. DeMartino, S. J. Siciliano, K. Lyons, M. S. Springer and J. P. Moore (2003a). "Use of a Small Molecule CCR5 Inhibitor in Macaques to Treat Simian Immunodeficiency Virus Infection or Prevent Simian-Human Immunodeficiency Virus Infection." J. Exp. Med. **198**(10): 1551-1562.

- Veazey, R. S., P. J. Klasse, S. M. Schader, Q. Hu, T. J. Ketas, M. Lu, P. A. Marx, J. Dufour, R. J. Colonna, R. J. Shattock, M. S. Springer and J. P. Moore (2005). "Protection of macaques from vaginal SHIV challenge by vaginally delivered inhibitors of virus-cell fusion." *Nature* **438**(7064): 99.
- Veazey, R. S., R. J. Shattock, M. Pope, J. C. Kirijan, J. Jones, Q. Hu, T. Ketas, P. A. Marx, P. J. Klasse, D. R. Burton and J. P. Moore (2003b). "Prevention of virus transmission to macaque monkeys by a vaginally applied monoclonal antibody to HIV-1 gp120." *Nat Med* **9**(3): 343.
- Vives, R. R., A. Imberty, Q. J. Sattentau and H. Lortat-Jacob (2005). "Heparan Sulfate Targets the HIV-1 Envelope Glycoprotein gp120 Coreceptor Binding Site." *J. Biol. Chem.* **280**(22): 21353-21357.
- Wang, C. Y., L. S. W. Sawyer, K. K. Murthy, X. Fang, A. M. Walfield, J. Ye, J. J. G. Wang, P. D. Chen, M. L. Li, M. T. Salas, M. Shen, M.-C. Gauduin, R. W. Boyle, R. A. Koup, D. C. Montefiori, J. R. Mascola, W. C. Koff and C. V. Hanson (1999). "Postexposure immunoprophylaxis of primary isolates by an antibody to HIV receptor complex." *PNAS* **96**(18): 10367-10372.
- Wei, X., J. M. Decker, S. Wang, H. Hui, J. C. Kappes, X. Wu, J. F. Salazar-Gonzalez, M. G. Salazar, J. M. Kilby, M. S. Saag, N. L. Komarova, M. A. Nowak, B. H. Hahn, P. D. Kwong and G. M. Shaw (2003). "Antibody neutralization and escape by HIV-1." *Nature* **422**(6929): 307-12.
- Weissenhorn, W., A. Dessen, S. C. Harrison, J. J. Skehel and D. C. Wiley (1997). "Atomic structure of the ectodomain from HIV-1 gp41." *Nature* **387**(6631): 426.
- Wu, L., N. P. Gerard, R. Wyatt, H. Choe, C. Parolin, N. Ruffing, A. Borsetti, A. A. Cardoso, E. Desjardin, W. Newman, C. Gerard and J. Sodroski (1996). "CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5." *Nature* **384**(6605): 179.
- Wuchty, S., W. Fontana, I. L. Hofacker and P. Schuster (1999). "Complete suboptimal folding of RNA and the stability of secondary structures." *Biopolymers* **49**(2): 145-65.
- Wyatt, R., P. D. Kwong, E. Desjardins, R. W. Sweet, J. Robinson, W. A. Hendrickson and J. G. Sodroski (1998). "The antigenic structure of the HIV gp120 envelope glycoprotein." *Nature* **393**(6686): 705-11.
- Wyatt, R., J. Moore, M. Accola, E. Desjardin, J. Robinson and J. Sodroski (1995). "Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding." *J. Virol.* **69**(9): 5723-5733.
- Xia, T., J. SantaLucia, Jr., M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox and D. H. Turner (1998). "Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs." *Biochemistry* **37**(42): 14719-35.

- Xu, X.-N., B. Laffert, G. R. Screaton, M. Kraft, D. Wolf, W. Kolanus, J. Mongkolsapay, A. J. McMichael and A. S. Baur (1999). "Induction of Fas Ligand Expression by HIV Involves the Interaction of Nef with the T Cell Receptor zeta Chain." *J. Exp. Med.* **189**(9): 1489-1496.
- Yang, X., J. Lee, E. M. Mahony, P. D. Kwong, R. Wyatt and J. Sodroski (2002). "Highly Stable Trimers Formed by Human Immunodeficiency Virus Type 1 Envelope Glycoproteins Fused with the Trimeric Motif of T4 Bacteriophage Fibrin." *J. Virol.* **76**(9): 4634-4642.
- Yeh, J. C., J. R. Seals, C. I. Murphy, H. van Halbeek and R. D. Cummings (1993). "Site-specific N-glycosylation and oligosaccharide structures of recombinant HIV-1 gp120 derived from a baculovirus expression system." *Biochemistry* **32**(41): 11087-99.
- Zhou, B. and B. Wang (2006). "Pegaptanib for the treatment of age-related macular degeneration." *Experimental Eye Research* **83**(3): 615.
- Zhu, P., E. Chertova, J. Bess, Jr., J. D. Lifson, L. O. Arthur, J. Liu, K. A. Taylor and K. H. Roux (2003). "Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions." *PNAS* **100**(26): 15812-15817.
- Zuker, M. (1989). "On finding all suboptimal foldings of an RNA molecule." *Science* **244**(4900): 48-52.
- Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." *Nucl. Acids Res.* **31**(13): 3406-3415.
- Zuker, M. and P. Stiegler (1981). "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information." *Nucleic Acids Res* **9**(1): 133-48.
- Zwick, M. B., R. Kelleher, R. Jensen, A. F. Labrijn, M. Wang, G. V. Quinnan, Jr., P. W. H. I. Parren and D. R. Burton (2003). "A Novel Human Antibody against Human Immunodeficiency Virus Type 1 gp120 Is V1, V2, and V3 Loop Dependent and Helps Delimit the Epitope of the Broadly Neutralizing Antibody Immunoglobulin G1 b12." *J. Virol.* **77**(12): 6965-6978.
- Zwick, M. B., M. Wang, P. Poignard, G. Stiegler, H. Katinger, D. R. Burton and P. W. Parren (2001). "Neutralization synergy of human immunodeficiency virus type 1 primary isolates by cocktails of broadly neutralizing antibodies." *J Virol* **75**(24): 12198-208.

Appendix 1: Aptamer binding to the gp120-CB fusion protein



Appendix 2:

Extensive Rsmatch results

Appendix 2: Full RSmatch results using the full-length sequence or variable region of aptamers B40, B4, B62, B84, B116, B141, B156, B163 and B229 as the query sequence to probe the full-length or variable sequence aptamer library.

B40 full vs full

```
#=== Query ===#
>B40:          B40

      .((((((.(((((((.((((.....((((((...)))...))...((...
1  GGGAGACAAGACTAGACGCTCAATGTGGGCCACGCCGATTTTACGCTTT 50

      .....))..)))..)))))))))...(((..((((((...((...))
51 TACCCGCACGCGATTGGTTTGTTCGACATGAGACTCACAAACAGTTCCC 100

      )...)))))..)))..
101 TTTAGTGAGGGTTAAT 116
```

#=== Hits ===#

Rank: 34 Score: 6 Query: 9 (ss:3,ds:6)
Identity: str: 100%; seq:66% (ss:0%, ds:100%)
Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:3, ds:0)

```

B40::                28  GGCCACGCC  36
                        |||      |||
B146:1-113:         47  GGCACAGCC  55

```

Rank: 35 Score: 5 Query: 13 (ss:9,ds:4)
Identity: str: 100%; seq:61% (ss:44%, ds:100%)
Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:5, ds:0)

			.((.....))	
			.((.....))	
B40::	45	CGCTTTTACCCGC		57
B86:1-117:	17	CGCTCAAACGTGC		29

Rank: 44 Score: 4 Query: 9 (ss:3,ds:6)
Identity: str: 90%; seq:88% (ss:66%, ds:100%)
Gap: 1 (ss:1, ds:0) Mismatch: 1 (ss:1, ds:0)

```

      ((... ))
      (((...)))
B40::      28 GGCCAC-GCC  36
             ||| || |||
B165:1-117: 31 GGCTACTGCC 40

```

Appendix 2 continued.

B40 full vs variable

```

===== Query =====#
>B40:      B40

      .(((((((.(((((((.((((.....((((((...)))))).....((...
1  GGGAGACAAGACTAGACGCTCAATGTGGGCCACGCCGATTTTACGCTTT 50

      .....))..)))).)))))))))))).)...(((..((((((...((...))
51  TACCCGCACGCGATTGGTTTGTTCGACATGAGACTCACAAACAGTTCCC 100

      )...)))).))))...
101 TTTAGTGAGGGTTAAT 116

```

#=== Hits ===#

```

Rank: 3   Score: 10   Query: 13 (ss:3,ds:10)
Identity: str: 100%; seq:61% (ss:0%, ds:80%)
Gap: 0 (ss:0, ds:0)  Mismatch: 5 (ss:3, ds:2)
                (((((...))))))
                (((((...))))))
B40::          26 TGGGCCACGCCCG 38
                ||||:   :||||
B28:1-50:     30 TGGGATTTTCCCG 42

```

```

Rank: 4   Score: 9   Query: 11 (ss:3,ds:8)
Identity: str: 100%; seq:63% (ss:33%, ds:75%)
Gap: 0 (ss:0, ds:0) Mismatch: 4 (ss:2, ds:2)
          (((...)))
          (((...)))
B40::      27 GGGCCACGCCC 37
          ||:| | |:||
B11:1-49:  5 GGGCTAAGTCC 15

```

```

Rank: 6   Score: 7   Query: 11 (ss:3,ds:8)
Identity: str: 100%; seq:45% (ss:33%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:2, ds:4)
          (((...)))
          (((...)))
B40::      27 GGGCCACGCCC 37
          |::|  ||::|
B19:1-42:  24 GCTCTTCGAGC 34

```

Appendix 2 continued.

```

Rank: 7   Score: 6   Query: 14 (ss:10,ds:4)
Identity: str: 100%; seq:57% (ss:60%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:4, ds:2)
      ((.....))..
      ((.....))..
B40::      46 GCTTTTACCCGCAC 59
      |:|| || :|||
B31:1-50:   1 GTTTATATATACAC 14

```

```
=====
```

```

Rank: 7   Score: 6   Query: 15 (ss:5,ds:10)
Identity: str: 93%; seq:46% (ss:100%, ds:20%)
Gap: 1 (ss:1, ds:0) Mismatch: 8 (ss:0, ds:8)
      ((((((. ..))))))..
      ((((((...))))))..
B40::      26 TGGGCC-ACGCCCGAT 40
      :::|:| ||:|:::|
B116:1-50:  1 GACGGCAACCCGTTAT 16

```

```
=====
```

```

Rank: 12  Score: 5   Query: 7 (ss:3,ds:4)
Identity: str: 100%; seq:71% (ss:33%, ds:100%)
Gap: 0 (ss:0, ds:0) Mismatch: 2 (ss:2, ds:0)
      ((...))
      ((...))
B40::      29 GCCACGC 35
      ||  |||
B9:1-48:    7 GCACCGC 13

```

```
=====
```

```

Rank: 12  Score: 5   Query: 15 (ss:11,ds:4)
Identity: str: 100%; seq:60% (ss:45%, ds:100%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:6, ds:0)
      ...((.....))
      ...((.....))
B40::      43 TACGCTTTTACCCGC 57
      |||||  || ||
B141:1-50:  30 TACGCACCACCCTGC 44

```

```
=====
```

```

Rank: 12  Score: 5   Query: 7 (ss:3,ds:4)
Identity: str: 100%; seq:57% (ss:66%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:1, ds:2)
      ((...))
      ((...))
B40::      29 GCCACGC 35
      |:|| :|
B197:1-50:  38 GACAATC 44

```

```
=====
```

Appendix 2 continued.

```

Rank: 25  Score: 4  Query: 13 (ss:3,ds:10)
Identity: str: 92%; seq:53% (ss:33%, ds:60%)
Gap: 1 (ss:1, ds:0)  Mismatch: 6 (ss:2, ds:4)
      (((((... )))))
      (((((....)))))
B40::      26 TGGGCCAC-GCCCG 38
      |:|| | ||::|
B4:1-50:   19 TAGGCAGCAGCTTG 32

```

```

=====

Rank: 25  Score: 4  Query: 14 (ss:10,ds:4)
Identity: str: 100%; seq:50% (ss:50%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:5, ds:2)
      ((.....))..
      ((.....))..
B40::      46 GCTTTTACCCGCAC 59
      |:|| | :|||
B137:1-50:  6  GGTCATCATTCAC 19

```

```

=====

Rank: 25  Score: 4  Query: 9 (ss:3,ds:6)
Identity: str: 90%; seq:88% (ss:66%, ds:100%)
Gap: 1 (ss:1, ds:0)  Mismatch: 1 (ss:1, ds:0)
      (((... )))
      (((....)))
B40::      28 GGCCAC-GCC 36
      ||| || |||
B165:1-50:  8  GGCTACTGCC 17

```

```

=====

Rank: 25  Score: 4  Query: 12 (ss:8,ds:4)
Identity: str: 100%; seq:50% (ss:50%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 6 (ss:4, ds:2)
      ((.....))
      ((.....))
B40::      46 GCTTTTACCCGC 57
      :| || |||:
B194:1-50:  33 CCACTTCACCGG 44

```

Appendix 2 continued.**B40 variable vs variable**

```

#=== Query ===#
>B40_variable: aptamer_B40_variable_region

..((((((...))))).(((.....))).....
1 TGTGGGCCACGCCCGATTTACGCTTTTACCCGCACGCGATTGGTTTGT 50

#=== Hits ===#

Rank: 2 Score: 10 Query: 13 (ss:3,ds:10)
Identity: str: 100%; seq:61% (ss:0%, ds:80%)
Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:3, ds:2)
          (((((...))))
          (((((...))))
B40_variable:: 3 TGGGCCACGCCCG 15
          ||||: :||||
B28:1-50:      30 TGGGATTTTCCCG 42

=====
=

Rank: 3 Score: 9 Query: 11 (ss:3,ds:8)
Identity: str: 100%; seq:63% (ss:33%, ds:75%)
Gap: 0 (ss:0, ds:0) Mismatch: 4 (ss:2, ds:2)
          (((((...))))
          (((((...))))
B40_variable:: 4 GGGCCACGCCCG 14
          ||:| | |:|
B11:1-49:      5 GGGCTAAGTCC 15

=====
=

Rank: 4 Score: 7 Query: 11 (ss:3,ds:8)
Identity: str: 100%; seq:45% (ss:33%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:2, ds:4)
          (((((...))))
          (((((...))))
B40_variable:: 4 GGGCCACGCCCG 14
          |:| | |:|
B19:1-42:      24 GCTCTTCGAGC 34

=====
=

Rank: 6 Score: 6 Query: 14 (ss:10,ds:4)
Identity: str: 100%; seq:57% (ss:60%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:4, ds:2)
          ((.....))..
          ((.....))..
B40_variable:: 23 GCTTTTACCCGCAC 36
          |:|| || :||
B31:1-50:      1 GTTTATATATACAC 14

```

Appendix 2 continued.

```

=====
Rank: 6   Score: 6   Query: 15 (ss:5,ds:10)
Identity: str: 93%; seq:46% (ss:100%, ds:20%)
Gap: 1 (ss:1, ds:0) Mismatch: 8 (ss:0, ds:8)
          ((((((. ..))))))..
          ((((((....))))))..
B40_variable:: 3 TGGGCC-ACGCCCGAT 17
          :::|:| ||:|::||
B116:1-50:      1 GACGGCAACCCGTTAT 16
=====

```

```

=====
Rank: 11  Score: 5   Query: 15 (ss:11,ds:4)
Identity: str: 100%; seq:60% (ss:45%, ds:100%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:6, ds:0)
          ...((.....))
          ...((.....))
B40_variable:: 20 TACGCTTTTACCCGC 34
          |||||      || ||
B141:1-50:      30 TACGCACCACCCTGC 44
=====

```

```

=====
Rank: 11  Score: 5   Query: 7 (ss:3,ds:4)
Identity: str: 100%; seq:57% (ss:66%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:1, ds:2)
          ((...))
          ((...))
B40_variable:: 6 GCCACGC 12
          |:|| :|
B197:1-50:      38 GACAATC 44
=====

```

```

=====
Rank: 26  Score: 4   Query: 13 (ss:3,ds:10)
Identity: str: 92%; seq:53% (ss:33%, ds:60%)
Gap: 1 (ss:1, ds:0) Mismatch: 6 (ss:2, ds:4)
          ((((((... )))))
          ((((((....))))))
B40_variable:: 3 TGGGCCAC-GCCCG 15
          |::|| | ||::|
B4:1-50:        19 TAGGCAGCAGCTTG 32
=====

```

```

=====
Rank: 26  Score: 4   Query: 14 (ss:10,ds:4)
Identity: str: 100%; seq:50% (ss:50%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 7 (ss:5, ds:2)
          ((.....))..
          ((.....))..
B40_variable:: 23 GCTTTTACCCGCAC 36
          |:|| | :|||
B137:1-50:      6 GGTTCATCATTAC 19
=====

```

Appendix 2 continued.

```

Rank: 26  Score: 4  Query: 9 (ss:3,ds:6)
Identity: str: 90%; seq:88% (ss:66%, ds:100%)
Gap: 1 (ss:1, ds:0)  Mismatch: 1 (ss:1, ds:0)
                (((... )))
                (((....)))
B40_variable::  5 GGCCAC-GCC 13
                ||| || |||
B165:1-50:      8 GGCTACTGCC 17

```

```

=====

Rank: 26  Score: 4  Query: 12 (ss:8,ds:4)
Identity: str: 100%; seq:50% (ss:50%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 6 (ss:4, ds:2)
                ((.....))
                ((.....))
B40_variable::  23 GCTTTTACCCGC 34
                :|  ||  |||:
B194:1-50:      33 CCACTTCACCGG 44

```

Appendix 2 continued.**B4 full vs full**

```

#=== Query ===#
>B4: aptamer_B4

      .(.....)....((((((((((.....)))))).....((((((((((((.....
1   GGGAGACAAGACTAGACGCTCAAGAGCGGTTAAGGGAGATTTAGGCAGCA 50

      )))))))).....((((((((.....)))))).....))))))
51  GCTTGGACAGTGTATCGGCTGAGTTCGACATGAGACTCACAACAGTTCCC 100

      )))))))).....
101 TTTAGTGAGGGTTAAT 116

#=== Hits ===#

Rank: 49   Score: 9   Query: 21 (ss:9,ds:12)
Identity: str: 91%; seq:71% (ss:100%, ds:50%)
Gap: 0 (ss:0, ds:0)   Mismatch: 6 (ss:0, ds:6)
      .((((((((.....))))))..
      .((((((((.....  ))))))..
B4::          69 CTGAGTTCGACATGAGACTCACA 91
      |:|::| | | | | | | | | | | |::|:| |
B230:1-117:   69 CGGTCTTCGACAT--GAGACTCA 89

=====

```

Appendix 2 continued.**B4 structure 2 full vs full**

```

#=== Query ===#
>B4: aptamer_B4
.....((((((((.....((((((((((((((((.....
1  GGGAGACAAGACTAGACGCTCAAGAGCGGTTAAGGGAGATTTAGGCAGCA 50

    )))))))).....((((((((.....)))))).....)))))
51  GCTTGGACAGTGTATCGGCTGAGTTCGACATGAGACTCACAACAGTTCCC 100

    ))...))))))...
101 TTTAGTGAGGGTTAAT 116

#=== Hits ===#
=====

Rank: 7  Score: 10  Query: 16 (ss:4,ds:12)
Identity: str: 100%; seq:37% (ss:50%, ds:33%)
Gap: 0 (ss:0, ds:0) Mismatch: 10 (ss:2, ds:8)
      (((((((.....))))))
      (((((((.....))))))
B4::          41 TTAGGCAGCAGCTTGG 56
      ::|::||  ||::|::
B145:1-117:   61 GGATTCAATAGGGTTC 76

=====

Rank: 9  Score: 9   Query: 18 (ss:4,ds:14)
Identity: str: 100%; seq:27% (ss:25%, ds:28%)
Gap: 0 (ss:0, ds:0) Mismatch: 13 (ss:3, ds:10)
      (((((((.....))))))
      (((((((.....))))))
B4::          40 TTAGGCAGCAGCTTGA 57
      :::::|  ||:::
B167:1-117:   33 GATCTGCTTGAGCGGATT 50

=====

Rank: 9  Score: 9   Query: 21 (ss:9,ds:12)
Identity: str: 91%; seq:71% (ss:100%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:0, ds:6)
      .((((((((.....))))))..
      .((((((((..... ))))))..
B4::          69 CTGAGTTCGACATGAGACTCACA 91
      |:|::| |||||  ||::|:|
B230:1-117:   69 CGGTCTTCGACAT--GAGACTCA 89

=====

Rank: 59  Score: 6   Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:62% (ss:75%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:1, ds:2)
      ((.....))
      ((.....))
B4::          45 GCAGCAGC 52
      :|| |||:
B5:1-116:     88 ACAACAGT 95

```

Appendix 2 continued.**B4 full vs variable**

#=== Query ===#

>B4: aptamer_B4

```

      .(.....)....((((((((((.....)))))).....((((((.....
1   GGGAGACAAGACTAGACGCTCAAGAGCGGTTAAGGGAGATTTAGGCAGCA 50
      ))))))).....((((((.....)))))).....))))))
51  GCTTGGACAGTGTATCGGCTGAGTTCGACATGAGACTCACAACAGTTCCC 100
      ))))))).....
101 TTTAGTGAGGGTTAAT 116

```

#=== Hits ===#

```

=====
=

```

```

Rank: 3   Score: 9   Query: 18 (ss:4,ds:14)
Identity: str: 100%; seq:27% (ss:25%, ds:28%)
Gap: 0 (ss:0, ds:0) Mismatch: 13 (ss:3, ds:10)
      (((((((.....))))))
      (((((((.....))))))
B4::          40 TTTAGGCAGCAGCTTGGA 57
      :::::||   ||:::
B167:1-50:    10 GATCTGCTTGAGCGGATT 27

```

```

=====

```

```

Rank: 6   Score: 7   Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:50% (ss:75%, ds:33%)
Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:1, ds:4)
      (((.....))
      (((.....))
B4::          44 GGCAGCAGCT 53
      :|:| ||:|:
B156:1-49:    3  CGAACCATCG 12

```

```

=====

```

```

Rank: 6   Score: 7   Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:28% (ss:50%, ds:20%)
Gap: 0 (ss:0, ds:0) Mismatch: 10 (ss:2, ds:8)
      (((((((.....))))))
      (((((((.....))))))
B4::          42 TAGGCAGCAGCTTG 55
      :|:::  ||:::|:
B163:1-49:    10 TAGATCCCAATCTA 23

```

```

=====

```

Appendix 2 continued.

```

Rank: 8   Score: 6   Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:75% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0) Mismatch: 2 (ss:2, ds:0)
      ((....))
      ((....))
B4::      45 GCAGCAGC 52
      ||| |||
B84:1-49: 12 GCACAAGC 19

```

```

=====
Rank: 17  Score: 5   Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:50% (ss:25%, ds:66%)
Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:3, ds:2)
      (((....)))
      (((....)))
B4::      44 GGCAGCAGCT 53
      :|| | ||:
B165:1-50: 8  GGCTACTGCC 17

```

```

=====
Rank: 17  Score: 5   Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:30% (ss:75%, ds:0%)
Gap: 0 (ss:0, ds:0) Mismatch: 7 (ss:1, ds:6)
      (((....)))
      (((....)))
B4::      44 GGCAGCAGCT 53
      ::: |||:::
B223:1-50: 18 TAGCGCACTG 27

```

```

=====
Rank: 29  Score: 4   Query: 24 (ss:12,ds:12)
Identity: str: 96%; seq:33% (ss:66%, ds:0%)
Gap: 0 (ss:0, ds:0) Mismatch: 16 (ss:4, ds:12)
      ((((((.....)))))).....
      ((((((.. ..))))).....
B4::      70 TGAGTTCGACATGAGACTCACAACA 94
      ::::: || ||| ::::: | ||
B201:1-48: 16 TTGTGCCG-CATTTGCACAGCCTCA 39

```

Appendix 2 continued.**B4 structure 2 full vs variable**

```

#=== Query ===#
>B4: aptamer_B4

      .....(((((((.....(((((((.(((((((.....
1   GGGAGACAAGACTAGACGCTCAAGAGCGGTTAAGGGAGATTTAGGCAGCA 50

      ))))))).....(((((((.....))))).....)))))
51  GCTTGGACAGTGTATCGGCTGAGTTGACATGAGACTCACAACAGTTCCC 100

      ))...)))))...
101 TTTAGTGAGGGTTAAT 116

#=== Hits ===#

Rank: 2   Score: 9   Query: 18 (ss:4,ds:14)
Identity: str: 100%; seq:27% (ss:25%, ds:28%)
Gap: 0 (ss:0, ds:0) Mismatch: 13 (ss:3, ds:10)
      (((((((.....))))))
      (((((((.....))))))
B4::          40 TTTAGGCAGCAGCTTGGA 57
      :::::||   ||:::
B167:1-50:    10 GATCTGCTTGAGCGGATT 27

=====

Rank: 4   Score: 7   Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:50% (ss:75%, ds:33%)
Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:1, ds:4)
      (((.....)))
      (((.....)))
B4::          44 GGCAGCAGCT 53
      :|:| ||:|:
B156:1-49:    3  CGAACCATCG 12

=====

Rank: 4   Score: 7   Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:28% (ss:50%, ds:20%)
Gap: 0 (ss:0, ds:0) Mismatch: 10 (ss:2, ds:8)
      (((((((.....))))))
      (((((((.....))))))
B4::          42 TAGGCAGCAGCTTG 55
      :|:::  ||:::|:
B163:1-49:    10 TAGATCCCAATCTA 23

=====

```

Appendix 2 continued.

```

Rank: 6  Score: 6  Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:75% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0)  Mismatch: 2 (ss:2, ds:0)
      ((....))
      ((....))
B4::      45 GCAGCAGC 52
      |||  |||
B84:1-49: 12 GCACAAGC 19

```

```

=====

Rank: 9  Score: 5  Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:28% (ss:0%, ds:40%)
Gap: 0 (ss:0, ds:0)  Mismatch: 10 (ss:4, ds:6)
      ((((((....))))))
      ((((((....))))))
B4::      42 TAGGCAGCAGCTTG 55
      :|:|:      :|:|:
B116:1-50: 1  GACGGCAACCCGTT 14

```

```

=====

Rank: 9  Score: 5  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:50% (ss:25%, ds:66%)
Gap: 0 (ss:0, ds:0)  Mismatch: 5 (ss:3, ds:2)
      (((....)))
      (((....)))
B4::      44 GGCAGCAGCT 53
      :||  |  ||:
B165:1-50: 8  GGCTACTGCC 17

```

```

=====

Rank: 9  Score: 5  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:30% (ss:75%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:1, ds:6)
      (((....)))
      (((....)))
B4::      44 GGCAGCAGCT 53
      ::: |||:::
B223:1-50: 18 TAGCGCACTG 27

```

```

=====

Rank: 28  Score: 4  Query: 13 (ss:3,ds:10)
Identity: str: 92%; seq:53% (ss:33%, ds:60%)
Gap: 0 (ss:0, ds:0)  Mismatch: 6 (ss:2, ds:4)
      ((((((....))))))
      ((((((... ))))))
B4::      42 TAGGCAGCAGCTTG 55
      |:|||  |  ||::|
B40:1-50: 3  TGGGCCAC-GCCCG 15

```

Appendix 2 continued.

```
Rank: 28  Score: 4  Query: 16 (ss:8,ds:8)
Identity: str: 100%; seq:31% (ss:37%, ds:25%)
Gap: 0 (ss:0, ds:0)  Mismatch: 11 (ss:5, ds:6)
      (((.....)))
      (((.....)))
B4::      72 AGTTCGACATGAGACT 87
      :|::|  |  |  ::|:
B146:1-46: 24 GGCACAGCCTTTTGCC 39
```

Excluded as does not form in full-length structure

```
Rank: 28  Score: 4  Query: 24 (ss:12,ds:12)
Identity: str: 96%; seq:33% (ss:66%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 16 (ss:4, ds:12)
      ((((((.....)))))).....
      ((((((.....)))))).....
B4::      70 TGAGTTCGACATGAGACTCACAACA 94
      :::::| | | |  :::::|  | |
B201:1-48: 16 TTGTGCCG-CATTTGCACAGCCTCA 39
```

Only structure from base 18 to 32 forms in full-length structure

Appendix 2 continued.**B4 variable vs variable**

#=== Query ===#

>B4: B4

```

...((((..(.....((((((((.....)))))).....).....))..
1  GAGCGGTTAAGGGAGATTAGGCAGCAGCTTGGACAGTGTATCGGCTGAG 50

```

#=== Hits ===#

```

Rank: 2  Score: 9  Query: 18 (ss:4,ds:14)
Identity: str: 100%; seq:27% (ss:25%, ds:28%)
Gap: 0 (ss:0, ds:0)  Mismatch: 13 (ss:3, ds:10)
      (((((((.....))))))
      (((((((.....))))))
B4::      17 TTAGGCAGCAGCTTGA 34
      ::::||  ||:::
B167:1-50: 10 GATCTGCTTGAGCGGATT 27

```

```

Rank: 3  Score: 7  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:50% (ss:75%, ds:33%)
Gap: 0 (ss:0, ds:0)  Mismatch: 5 (ss:1, ds:4)
      (((.....)))
      (((.....)))
B4::      21 GGCAGCAGCT 30
      :|:|  ||:|:
B156:1-49: 3  CGAACCATCG 12

```

```

Rank: 3  Score: 7  Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:28% (ss:50%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 10 (ss:2, ds:8)
      (((((((.....))))))
      (((((((.....))))))
B4::      19 TAGGCAGCAGCTTG 32
      :|:::  ||:::|:
B163:1-49: 10 TAGATCCCAATCTA 23

```

```

Rank: 5  Score: 6  Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:75% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0)  Mismatch: 2 (ss:2, ds:0)
      ((.....))
      ((.....))
B4::      22 GCAGCAGC 29
      |||  |||
B84:1-49: 12 GCACAAGC 19

```

Appendix 2 continued.

```

Rank: 6  Score: 5  Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:28% (ss:0%, ds:40%)
Gap: 0 (ss:0, ds:0)  Mismatch: 10 (ss:4, ds:6)
      ((((((....))))))
      ((((((....))))))
B4::      19 TAGGCAGCAGCTTG 32
      :|:|:      :|:|:
B116:1-50: 1  GACGGCAACCCGTT 14

```

```

Rank: 6  Score: 5  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:50% (ss:25%, ds:66%)
Gap: 0 (ss:0, ds:0)  Mismatch: 5 (ss:3, ds:2)
      (((....)))
      (((....)))
B4::      21 GGCAGCAGCT 30
      :||  |  ||:
B165:1-50: 8  GGCTACTGCC 17

```

```

Rank: 6  Score: 5  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:30% (ss:75%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:1, ds:6)
      (((....)))
      (((....)))
B4::      21 GGCAGCAGCT 30
      ::: |||:::
B223:1-50: 18 TAGCGCACTG 27

```

```

Rank: 11  Score: 4  Query: 16 (ss:4,ds:12)
Identity: str: 100%; seq:6% (ss:25%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 15 (ss:3, ds:12)
      ((((((....))))))
      ((((((....))))))
B4::      18 TTAGGCAGCAGCTTG 33
      ::::: | :::::
B1:1-47: 26 GGTTTTTTCGGGAAC 41

```

```

Rank: 11  Score: 4  Query: 13 (ss:3,ds:10)
Identity: str: 92%; seq:53% (ss:33%, ds:60%)
Gap: 0 (ss:0, ds:0)  Mismatch: 6 (ss:2, ds:4)
      ((((((....))))))
      ((((((... )))))
B4::      19 TAGGCAGCAGCTTG 32
      |:|:|  |  |:|:|
B40:1-50: 3  TGGGCCAC-GCCCG 15

```

Appendix 2 continued.

```

Rank: 11  Score: 4  Query: 12 (ss:4,ds:8)
Identity: str: 100%; seq:25% (ss:25%, ds:25%)
Gap: 0 (ss:0, ds:0)  Mismatch: 9 (ss:3, ds:6)
      (((.....)))
      (((.....)))
B4::          20 AGGCAGCAGCTT 31
      ::|:  | :|::
B207:1-50:    1  GTGTTACTGCAC 12

```

```

=====
Rank: 26  Score: 3  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:40% (ss:0%, ds:66%)
Gap: 0 (ss:0, ds:0)  Mismatch: 6 (ss:4, ds:2)
      (((.....)))
      (((.....)))
B4::          21 GGCAGCAGCT 30
      :||      ||:
B145:1-50:    7  TGCCTTCGCG 16

```

```

=====
Rank: 26  Score: 3  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:30% (ss:25%, ds:33%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:3, ds:4)
      (((.....)))
      (((.....)))
B4::          21 GGCAGCAGCT 30
      |:: |  ::|
B175:1-50:    17 GCTCGTTGGT 26

```

```

=====
Rank: 26  Score: 3  Query: 18 (ss:4,ds:14)
Identity: str: 94%; seq:16% (ss:75%, ds:0%)
Gap: 1 (ss:1, ds:0)  Mismatch: 15 (ss:1, ds:14)
      (((((((... .)))))))
      (((((((.....)))))))
B4::          17 TTTAGGCAGC-AGCTTGGA 34
      :::::| | |:::
B188:1-50:    18 CTTGTTTACCGAGAACAAG 36

```

```

=====
Rank: 26  Score: 3  Query: 6 (ss:4,ds:2)
Identity: str: 100%; seq:66% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0)  Mismatch: 2 (ss:2, ds:0)
      (....)
      (....)
B4::          23 CAGCAG 28
      || | |
B195:1-49:    25 CATCGG 30

```

Appendix 2 continued.**B62 full vs full**

#=== Query ===#

>B62: B62

```

      ((((((...(((.((((.....((.....)).....(((.....)
1   GGGAGACAAGACTAGACGCTCAACCCGTACCACCACACCCTATGCACATC 50

      ))))))))....)))))).....(((.((((.....((.....)
51  GTTGTTTGTCGTCTTTCCCGCATTTTCGACATGAGACTCACAACAGTTCCC 100

      )...))))))....
101 TTTAGTGAGGGTTAATT 117

```

#=== Hits ===#

```

Rank: 44   Score: 6   Query: 10 (ss:6,ds:4)
Identity: str: 100%; seq:70% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0)   Mismatch: 3 (ss:3, ds:0)

```

```

      ((.....))
      ((.....))
B62::      27 GTACCACCAC 36
      || | || ||
B211:1-116: 34 GTCCTACAAC 43

```

=====

Appendix 2 continued.**B62 full vs variable**

```
#=== Query ===#
>B62:      B62

      ((((((...(((.((((.....((.....)).....(((.....)
1  GGGAGACAAGACTAGACGCTCAACCCGTACCACCACACCCTATGCACATC 50

      ))))))))....)))))).....(((..(((.....((.....)
51 GTTGTTTGTGCTCTTTCCCGCATTTTCGACATGAGACTCACAACAGTTCCC 100

      )....))))..))....
101 TTTAGTGAGGGTTAATT 117
```

```
#=== Hits ===#
```

```
Rank: 5  Score: 6  Query: 11 (ss:5,ds:6)
Identity: str: 100%; seq:36% (ss:80%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:1, ds:6)
      (((.....))
      (((.....))
B62::      42 ATGCACATCGT 52
      :::|||  :::
B3:1-50:   23 TGCCACAAGTA 33
```

```
=====

Rank: 5  Score: 6  Query: 12 (ss:8,ds:4)
Identity: str: 100%; seq:50% (ss:75%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 6 (ss:2, ds:4)
      .((.....)).
      .((.....)).
B62::      26 CGTACCACCACA 37
      |::|| | |::|
B197:1-50:  36 CCGACAATCCGA 47
```

```
=====

Rank: 5  Score: 6  Query: 10 (ss:6,ds:4)
Identity: str: 100%; seq:70% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0)  Mismatch: 3 (ss:3, ds:0)
      (((.....))
      (((.....))
B62::      27 GTACCACCAC 36
      || | || ||
B211:1-49:  11 GTCCTACAAC 20
```

Appendix 2 continued.

```

Rank: 32  Score: 4  Query: 10 (ss:6,ds:4)
Identity: str: 100%; seq:40% (ss:66%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 6 (ss:2, ds:4)
      ((.....))
      ((.....))
B62::      27 GTACCACCAC 36
      :: |||| ::
B207:1-50: 30 AATCCACGTT 39

```

=====

```

Rank: 32  Score: 4  Query: 10 (ss:6,ds:4)
Identity: str: 100%; seq:50% (ss:50%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 5 (ss:3, ds:2)
      ((.....))
      ((.....))
B62::      27 GTACCACCAC 36
      |:||  |:|
B229:1-50: 36 GGACACTCCC 45

```

=====

Appendix 2 continued.**B62 variable vs variable**

```

#=== Query ===#
>B62:      B62

      ...((.....)).....(((.....))).....
1  CCGTACCACCACACCCTATGCACATCGTTGTTTGTCTCTTTCCCGCAT 50

#=== Hits ===#

Rank: 5  Score: 6  Query: 11 (ss:5,ds:6)
Identity: str: 100%; seq:36% (ss:80%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:1, ds:6)
      (((.....)))
      (((.....)))
B62::      19 ATGCACATCGT 29
      :::||||| :::
B3:1-50:   23 TGCCACAAGTA 33

=====

Rank: 5  Score: 6  Query: 11 (ss:5,ds:6)
Identity: str: 100%; seq:54% (ss:40%, ds:66%)
Gap: 0 (ss:0, ds:0)  Mismatch: 5 (ss:3, ds:2)
      (((.....)))
      (((.....)))
B62::      19 ATGCACATCGT 29
      :|| | |||:
B33:1-50:   15 TTGAATCTCGA 25

=====

Rank: 5  Score: 6  Query: 12 (ss:8,ds:4)
Identity: str: 100%; seq:50% (ss:75%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 6 (ss:2, ds:4)
      .((.....)).
      .((.....)).
B62::      3  CGTACCACCACA 14
      |::|| | |::|
B197:1-50:  36 CCGACAATCCGA 47

=====

Rank: 5  Score: 6  Query: 10 (ss:6,ds:4)
Identity: str: 100%; seq:70% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0)  Mismatch: 3 (ss:3, ds:0)
      ((.....))
      ((.....))
B62::      4  GTACCACCAC 13
      || | || |
B211:1-49:  11 GTCCTACAAC 20

```

Appendix 2 continued.

```
=====
Rank: 30  Score: 4  Query: 10 (ss:6,ds:4)
Identity: str: 100%; seq:50% (ss:50%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 5 (ss:3, ds:2)
      ((.....))
      ((.....))
B62::      4  GTACCACCAC 13
      |:||  |:|
B229:1-50: 36 GGACACTCCC 45
=====
```

Appendix 2 continued.

B84 full vs full

No significant hits

B84 full vs variable

#=== Query ===#

>B84: B84

```

1      (((((( (((((( ((.((( (.((.(.....))))))....((.....)).....)
GGGAGACAAGACTAGACGCTCAAATGACGTACCCGCACAAGCCACCACAA 50

51      )))))).)).)).))((((((.((.....((.(((.....)))..))..
GTCTTAATCGCGCCACCCTTGCTTCGACATGAGACTCACAACAGTTCCT 100

101      ..))..)))))).....
TTAGTGAGGGTTAATT 116

```

```
#=== Hits ===#
```

```

Rank: 7   Score: 6   Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:75% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0)   Mismatch: 2 (ss:2, ds:0)
                ((...))
                ((...))
B84::          35 GCACAAGC 42
                |||  |||
B4:1-50:       22 GCAGCAGC 29

```

```

Rank: 7   Score: 6   Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:62% (ss:75%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:1, ds:2)
          ((.....))
          ((.....))
B84::      35 GCACAAGC 42
             |:| |:|
B156:1-49: 4  GAACCATC 11

```

```

Rank: 32   Score: 4   Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:50% (ss:50%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 4 (ss:2, ds:2)
                ((.....))
                ((.....))
B84::          35 GCACAAGC 42
                |:  ||:|
B167:1-50:    39 GGTTAACC 46

```

Appendix 2 continued.**B84 variable vs variable**

```
#=== Query ===#
>B84:      B84
```

```
..(((.....((.....)).....))).....
1  ATGACGTACCCGCACAAGCCACCACAAAGTCTTAATCGCGCCACCCTTGC 49
```

```
#=== Hits ===#
```

```
Rank: 5  Score: 6  Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:75% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0)  Mismatch: 2 (ss:2, ds:0)
```

```
      ((.....))
      ((.....))
B84::      12 GCACAAGC 19
           |||  |||
B4:1-50:   22 GCAGCAGC 29
```

```
=====
```

```
Rank: 5  Score: 6  Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:62% (ss:75%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 3 (ss:1, ds:2)
```

```
      ((.....))
      ((.....))
B84::      12 GCACAAGC 19
           |:|| |:|
B156:1-49:  4  GAACCATC 11
```

```
=====
```

Appendix 2 continued.

B116 full vs full

```
#=== Query ===#
>B116:      B116

      ((((((...((( (((((((...((( (((((((...
1  GGGAGACAAGACTAGACGCTCAAGACGGCAACCCGTTATAACCTCCCACT  50

      )))...)))))))))...((( (((((((...(((
51 GGCTATCCCGTTAAGCTTCCCTATTCGACATGAGACTCACAAACAGTTCCC 100

      )...))))))...))...
101 TTTAGTGAGGGTTAATT 117
```

```
#=== Hits ===#
```

Rank: 33 Score: 8 Query: 17 (ss:7,ds:10)
Identity: str: 100%; seq:35% (ss:57%, ds:20%)
Gap: 0 (ss:0, ds:0) Mismatch: 11 (ss:3, ds:8)

			.((((((...))))))..	
			.((((((...))))))..	
B116::	23	AGACGGCAACCCGTTAT		39
		::: :::		
B217:1-117:	58	ACTTCGCAATCGAGGGT		74

Rank: 39 Score: 7 Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:60% (ss:50%, ds:66%)
Gap: 0 (ss:0, ds:0) Mismatch: 4 (ss:2, ds:2)

(((...)))
 (((...)))
 B116:: 26 CGGCAACCCG 35
 :||| ||:
 B82:1-117: 30 TGGCACGCCG 39

Appendix 2 continued.**B116 full vs variable**

```
#=== Query ===#
>B116:      B116

      ((((((.....(((((((.....(((((((.....)))))).....((((.....
1  GGGAGACAAGACTAGACGCTCAAGACGGCAACCCGTTATAACCTCCCACT 50

      )).....))))))))).....(((((((.....(((((((.....)))))).....
51 GGCTATCCCGTTAAGCTTCCCTATTCGACATGAGACTCACAACAGTTCCC 100

      ).....)))))).....
101 TTAGTGAGGGTTAATT 117
```

```
#=== Hits ===#
```

```
Rank: 3  Score: 8  Query: 15 (ss:5,ds:10)
Identity: str: 100%; seq:33% (ss:60%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 10 (ss:2, ds:8)
      .(((((((.....))))))
      .(((((((.....))))))
B116::      23 AGACGGCAACCCGTT 37
      |:|::|  |:|::|
B217:1-50:  35 ACTTCGCATGCGAGG 49
```

```
=====  
Rank: 7  Score: 6  Query: 17 (ss:7,ds:10)
Identity: str: 100%; seq:29% (ss:42%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 12 (ss:4, ds:8)
      .(((((((.....))))))..
      .(((((((.....))))))..
B116::      23 AGACGGCAACCCGTTAT 39
      |:|::|  |:|::|
B11:1-49:  22 ATCCTTCCTAAAGGACT 38
```

```
=====  
Rank: 7  Score: 6  Query: 15 (ss:5,ds:10)
Identity: str: 93%; seq:46% (ss:100%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 8 (ss:0, ds:8)
      ((((((.....))))))..
      ((((((.....))))))..
B116::      24 GACGGCAACCCGTTAT 39
      ::|::|  ||:|::|
B40:1-50:  3  TGGGCC-ACGCCCCGAT 17
```

Appendix 2 continued.

Rank: 7 Score: 6 Query: 8 (ss:4,ds:4)
 Identity: str: 100%; seq:75% (ss:50%, ds:100%)
 Gap: 0 (ss:0, ds:0) Mismatch: 2 (ss:2, ds:0)

```

      ((....))
      ((....))
B116::      27 GGCAACCC 34
      ||| | ||
B86:1-50:   34 GGCTAGCC 41

```

=====

Rank: 7 Score: 6 Query: 8 (ss:4,ds:4)
 Identity: str: 100%; seq:62% (ss:75%, ds:50%)
 Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:1, ds:2)

```

      ((....))
      ((....))
B116::      27 GGCAACCC 34
      |:|| |:|
B163:1-49:  28 GACATCTC 35

```

Exclude as not found in full-length structure

=====

Rank: 7 Score: 6 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:33% (ss:50%, ds:25%)
 Gap: 0 (ss:0, ds:0) Mismatch: 8 (ss:2, ds:6)

```

      (((((....))))))
      (((((....))))))
B116::      25 ACGGCAACCCGT 36
      :|:: ||::|:
B197:1-50:  24 GCTGTTACTAGT 35

```

=====

Rank: 7 Score: 6 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:41% (ss:25%, ds:50%)
 Gap: 0 (ss:0, ds:0) Mismatch: 7 (ss:3, ds:4)

```

      (((((....))))))
      (((((....))))))
B116::      25 ACGGCAACCCGT 36
      |:::|| |:::|
B199:1-50:  20 ATCGCTTTCGAT 31

```

=====

Rank: 7 Score: 6 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:25% (ss:75%, ds:0%)
 Gap: 0 (ss:0, ds:0) Mismatch: 9 (ss:1, ds:8)

```

      (((((....))))))
      (((((....))))))
B116::      25 ACGGCAACCCGT 36
      ::::| ||:::
B207:1-50:  29 TAATCCACGTTA 40

```

=====

Appendix 2 continued.

Rank: 7 Score: 6 Query: 8 (ss:4,ds:4)
 Identity: str: 100%; seq:62% (ss:75%, ds:50%)
 Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:1, ds:2)

((....))
 ((....))
 B116:: 27 GGCAACCC 34
 :|||| |:
 B212:1-50: 24 CGCAAACG 31

=====

Rank: 7 Score: 6 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:33% (ss:50%, ds:25%)
 Gap: 0 (ss:0, ds:0) Mismatch: 8 (ss:2, ds:6)

(((((....))))
 (((((....))))
 B116:: 25 ACGGCAACCCGT 36
 :|::|| ::|:
 B229:1-50: 3 GCCTCACGAGGC 14

=====

Rank: 18 Score: 5 Query: 14 (ss:4,ds:10)
 Identity: str: 100%; seq:28% (ss:0%, ds:40%)
 Gap: 0 (ss:0, ds:0) Mismatch: 10 (ss:4, ds:6)

(((((....))))
 (((((....))))
 B116:: 24 GACGGCAACCCGTT 37
 :|::| :|::|
 B4:1-50: 19 TAGGCAGCAGCTTG 32

=====

Rank: 18 Score: 5 Query: 11 (ss:7,ds:4)
 Identity: str: 100%; seq:54% (ss:57%, ds:50%)
 Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:3, ds:2)

((.....))
 ((.....))
 B116:: 42 CCTCCCACTGG 52
 |: | | ||:|
 B84:1-49: 38 CGCCACCCTTG 48

=====

Rank: 18 Score: 5 Query: 10 (ss:4,ds:6)
 Identity: str: 100%; seq:40% (ss:50%, ds:33%)
 Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:2, ds:4)

(((((....))))
 (((((....))))
 B116:: 26 CGGCAACCCG 35
 :|:| |:|:
 B145:1-50: 7 TGCCTTCGCG 16

=====

Rank: 18 Score: 5 Query: 13 (ss:9,ds:4)

Appendix 2 continued.

Identity: str: 92%; seq:76% (ss:88%, ds:50%)
 Gap: 1 (ss:1, ds:0) Mismatch: 3 (ss:1, ds:2)
 ((.))..
 ((.....))..
 B116:: 42 CCT-CCCACTGGCT 54
 :|| |||| ||:||
 B157:1-50: 24 TCTCCCAATGACT 37

=====

Rank: 18 Score: 5 Query: 20 (ss:10,ds:10)
 Identity: str: 95%; seq:45% (ss:70%, ds:20%)
 Gap: 1 (ss:1, ds:0) Mismatch: 11 (ss:3, ds:8)
 ...(((((....))))..
 ...((((.....))))..
 B116:: 21 CAAGACGG-CAACCCGTTATA 40
 || :::|: ||| :|:::| |
 B165:1-50: 29 CAGACTGTGCAAAACGGTAAA 49

=====

Rank: 18 Score: 5 Query: 11 (ss:7,ds:4)
 Identity: str: 100%; seq:54% (ss:57%, ds:50%)
 Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:3, ds:2)
 ((.....))
 ((.....))
 B116:: 42 CCTCCCACTGG 52
 :| || |||:
 B184:1-49: 14 ACCCATCTGT 24

=====

Rank: 37 Score: 4 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:16% (ss:50%, ds:0%)
 Gap: 0 (ss:0, ds:0) Mismatch: 10 (ss:2, ds:8)
 (((.....)))
 (((.....)))
 B116:: 25 ACGGCAACCCGT 36
 :::: | |::::
 B19:1-42: 5 CTTATACCTGAG 16

=====

=

Appendix 2 continued.**B116 variable vs variable**

```

#=== Query ===#
>B116:      B116

      ((((((.....))))). .... ((.....)) .....
1  GACGGCAACCCGTTATAACCTCCCACTGGCTATCCCGTTAAGCTTCCCTA 50

#=== Hits ===#

Rank: 3  Score: 7  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:60% (ss:50%, ds:66%)
Gap: 0 (ss:0, ds:0)  Mismatch: 4 (ss:2, ds:2)
      (((.....)))
      (((.....)))
B116::      3  CGGCAACCCG 12
      |:|  |||:|
B141:1-50:  9  CTGACACCGG 18

=====

Rank: 3  Score: 7  Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:28% (ss:50%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 10 (ss:2, ds:8)
      ((((((.....))))))
      ((((((.....))))))
B116::      1  GACGGCAACCCGTT 14
      ::::|||  |::::
B217:1-50:  36  CTTTCGCATGCGAGG 49

=====

Rank: 7  Score: 6  Query: 15 (ss:5,ds:10)
Identity: str: 93%; seq:46% (ss:100%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 8 (ss:0, ds:8)
      ((((((.....)))))..
      ((((((. . . . .)))))..
B116::      1  GACGGCAACCCGTTAT 16
      ::::|:|  ||:|:::|
B40:1-50:   3  TGGGCC-ACGCCCGAT 17

=====

Rank: 7  Score: 6  Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:75% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0)  Mismatch: 2 (ss:2, ds:0)
      ((.....))
      ((.....))
B116::      4  GGCAACCC 11
      |||  |  ||
B86:1-50:   34  GGCTAGCC 41

=====

```

Appendix 2 continued.

Rank: 7 Score: 6 Query: 8 (ss:4,ds:4)
 Identity: str: 100%; seq:62% (ss:75%, ds:50%)
 Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:1, ds:2)

```
((....))
((....))
B116::      4  GGCAACCC 11
             |:|| |:|
B163:1-49:  28 GACATCTC 35
```

Exclude as not found in full-length structure

Rank: 7 Score: 6 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:33% (ss:50%, ds:25%)
 Gap: 0 (ss:0, ds:0) Mismatch: 8 (ss:2, ds:6)

```
(((((....))))
(((....)))
B116::      2  ACGGCAACCCGT 13
             :|:: ||::|:
B197:1-50:  24 GCTGTTACTAGT 35
```

Rank: 7 Score: 6 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:41% (ss:25%, ds:50%)
 Gap: 0 (ss:0, ds:0) Mismatch: 7 (ss:3, ds:4)

```
(((((....))))
(((....)))
B116::      2  ACGGCAACCCGT 13
             |:|| |:|
B199:1-50:  20 ATCGCTTTCGAT 31
```

Rank: 7 Score: 6 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:25% (ss:75%, ds:0%)
 Gap: 0 (ss:0, ds:0) Mismatch: 9 (ss:1, ds:8)

```
(((((....))))
(((....)))
B116::      2  ACGGCAACCCGT 13
             ::::| ||:::
B207:1-50:  29 TAATCCACGTTA 40
```

Rank: 7 Score: 6 Query: 8 (ss:4,ds:4)
 Identity: str: 100%; seq:62% (ss:75%, ds:50%)
 Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:1, ds:2)

```
((....))
((....))
B116::      4  GGCAACCC 11
             :|||| |:
B212:1-50:  24 CGCAAACG 31
```

Appendix 2 continued.

```

Rank: 7   Score: 6   Query: 12 (ss:4,ds:8)
Identity: str: 100%; seq:33% (ss:50%, ds:25%)
Gap: 0 (ss:0, ds:0) Mismatch: 8 (ss:2, ds:6)
          (((.....)))
          (((.....)))
B116::      2 ACGGCAACCCGT 13
          :|::||  ::|:
B229:1-50:  3 GCCTCACGAGGC 14

```

```

=====

```

```

Rank: 18  Score: 5   Query: 11 (ss:7,ds:4)
Identity: str: 100%; seq:54% (ss:57%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:3, ds:2)
          ((.....))
          ((.....))
B116::      19 CCTCCCACTGG 29
          |: | | ||:|
B84:1-49:   38 CGCCACCCTTG 48

```

```

=====

```

```

Rank: 18  Score: 5   Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:40% (ss:50%, ds:33%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:2, ds:4)
          (((.....)))
          (((.....)))
B116::      3 CGGCAACCCG 12
          :|:|  |:|:
B145:1-50:  7 TGCCTTCGCG 16

```

```

=====

```

```

Rank: 18  Score: 5   Query: 13 (ss:9,ds:4)
Identity: str: 92%; seq:76% (ss:88%, ds:50%)
Gap: 1 (ss:1, ds:0) Mismatch: 3 (ss:1, ds:2)
          ((. ....))..
          ((.....))..
B116::      19 CCT-CCCACTGGCT 31
          :|| |||| ||:|
B157:1-50:  24 TCTCCCAATGACT 37

```

```

=====

```

```

Rank: 18  Score: 5   Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:50% (ss:25%, ds:66%)
Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:3, ds:2)
          (((.....)))
          (((.....)))
B116::      3 CGGCAACCCG 12
          :||  | ||:
B167:1-50:  38 AGGTTAACCT 47

```

```

=====

```

Appendix 2 continued.

Rank: 18 Score: 5 Query: 11 (ss:7,ds:4)
 Identity: str: 100%; seq:54% (ss:57%, ds:50%)
 Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:3, ds:2)

```

      ((.....))
      ((.....))
B116::      19 CCTCCCACTGG 29
      :| ||  |||:
B184:1-49:  14 ACCCCATCTGT 24

```

=====

Rank: 36 Score: 4 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:16% (ss:50%, ds:0%)
 Gap: 0 (ss:0, ds:0) Mismatch: 10 (ss:2, ds:8)

```

      (((.....)))
      (((.....)))
B116::      2 ACGGCAACCCGT 13
      :::: | |::::
B19:1-42:   5 CTTATACCTGAG 16

```

=====

Rank: 36 Score: 4 Query: 17 (ss:7,ds:10)
 Identity: str: 94%; seq:41% (ss:71%, ds:20%)
 Gap: 1 (ss:1, ds:0) Mismatch: 10 (ss:2, ds:8)

```

      (((((( .....))))))...
      ((((((.....))))))...
B116::      1 GACGG-CAACCCGTTATA 17
      :::|: ||| :|:::| |
B165:1-50:  32 ACTGTGCAAAACGGTAAA 49

```

=====

Appendix 2 continued.

B141 full vs full

```
#=== Query ===#  
>B141:      aptamer B141  
  
      .((((((.....(((((((.....(((.....)))))))))..)  
1    GGGAGACAAGACTAGACGCTCAATGCTAACCCGTGACACCGGTGTTGAGCT 50  
  
      .....(((.....))).....))))....(((.....(((.....(  
51 TATACGCACCACCCTGCAAGCTCTTCGACATGAGACTCACAAACAGTTCCTC 100  
  
      )...)))).))..)  
101 TTTAGTGAGGGTTAATT 117
```

```
#=== Hits ===#
```

Rank: 41 Score: 9 Query: 12 (ss:6,ds:6)
Identity: str: 100%; seq:75% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:3, ds:0)

		(((.....)))	
		(((.....)))	
B141::	56	GCACCACCCTGC	67
B45:1-112:	31	GCACCACCCTGC	42

Rank: 58 Score: 6 Query: 13 (ss:7,ds:6)
Identity: str: 100%; seq:46% (ss:57%, ds:33%)
Gap: 0 (ss:0, ds:0) Mismatch: 7 (ss:3, ds:4)

		(((.....))) .	
		(((.....))) .	
B141::	56	GCACCACCCCTGCA	68
		: : : :	
B114:1-116:	43	GCCATACACGGTA	55

Appendix 2 continued.

B141 full vs variable

```
#==== Query ===#
>B141:      B1

      .((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((
1      GGGAGACAAGACTAGACGCTCAATGCTAACCTTGACACCGGTGTTGAGCT 50
      .....((((((((((((((((((((((((((((((((((((((((((((((((((((((((
      .....((((((((((((((((((((((((((((((((((((((((((((((((((((((((
51      TATACGCACCACCCTGCAAGCTCTTCGACATGAGACTCACAAACAGTTCCC 100
      .....((((((((((((((((((((((((((((((((((((((((((((((((((((((((
      )...)))))...))...))...
101      TTTAGTGAGGGTTAATT 117
```

```
#==== Hits ====#
Rank: 3   Score: 9   Query: 12 (ss:6,ds:6)
Identity: str: 100%; seq:75% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:3, ds:0)
                (((.....)))
                (((.....)))
B141::          56 GCACCACCTGC 67
                ||||  || |||
B45:1-50:       8  GCACACCCTTGC 19
```

```

Rank: 3   Score: 9   Query: 18 (ss:12,ds:6)
Identity: str: 94%; seq:77% (ss:83%, ds:66%)
Gap: 1 (ss:1, ds:0) Mismatch: 4 (ss:2, ds:2)
          .....(((.....)))
          .....(((.....)))
B141::      50 TTATACGCACCACCC-TGC 67
              |||  |||:||||| :||
B84:1-49:   31 TTAATCGCGCCACCCTTGC 49

```

Excluded as does not form in the full-length structure

```
Rank: 6   Score: 7   Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:60% (ss:50%, ds:66%)
Gap: 0 (ss:0, ds:0) Mismatch: 4 (ss:2, ds:2)
          (((.....)))
          (((.....)))
B141::      32 CTGACACCGG 41
              |:| |||:|
B116:1-50:   3 CGGCAACCCG 12
```

```

Rank: 8   Score: 6   Query: 11 (ss:5,ds:6)
Identity: str: 100%; seq:45% (ss:60%, ds:33%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:2, ds:4)
          .(((.....)))
          .(((.....)))
B141::      31 CCTGACACCGG 41
              ||::||  ::|
B38:1-49:   24 CCTGACTGTAG 34

```

Appendix 2 continued.

```

=====
Rank: 15  Score: 5  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:40% (ss:50%, ds:33%)
Gap: 0 (ss:0, ds:0)  Mismatch: 6 (ss:2, ds:4)
      (((.....)))
      (((.....)))
B141::      32 CTGACACCGG 41
      |:  ||::|
B197:1-50:  25 CTGTTACTAG 34

```

```

=====
Rank: 15  Score: 5  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:30% (ss:75%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:1, ds:6)
      (((.....)))
      (((.....)))
B141::      32 CTGACACCGG 41
      ::: |||:::
B207:1-50:  30 AATCCACGTT 39

```

```

=====
Rank: 15  Score: 5  Query: 12 (ss:6,ds:6)
Identity: str: 100%; seq:33% (ss:66%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 8 (ss:2, ds:6)
      (((.....)))
      (((.....)))
B141::      56 GCACCACCCTGC 67
      ::: ||| |:::
B229:1-50:  35 TGGACACTCCCG 46

```

```

=====
Rank: 27  Score: 4  Query: 13 (ss:7,ds:6)
Identity: str: 100%; seq:38% (ss:42%, ds:33%)
Gap: 0 (ss:0, ds:0)  Mismatch: 8 (ss:4, ds:4)
      (((.....)))
      (((.....)))
B141::      56 GCACCACCCTGCA 68
      :|: | | :|:|
B3:1-50:    21 ACTGCCACAAGTA 33

```

```

=====
Rank: 27  Score: 4  Query: 13 (ss:7,ds:6)
Identity: str: 100%; seq:30% (ss:57%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 9 (ss:3, ds:6)
      (((.....)))
      (((.....)))
B141::      56 GCACCACCCTGCA 68
      :::|| | :|:|
B211:1-49:  10 TGTCTTACAACGA 22

```

Appendix 2 continued.**B141 variable vs variable**

```

#=== Query ===#
>B141:      B141

      ....(((.(((.....)))..)))((((((.....(((.....))))))))))
1  TGCTAACCCCTGACACCGGTGTTGAGCTTATACGCACCACCCTGCAAGCTC 50

#=== Hits ===#

Rank: 2  Score: 9   Query: 12 (ss:6,ds:6)
Identity: str: 100%; seq:75% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0)  Mismatch: 3 (ss:3, ds:0)
      (((.....)))
      (((.....)))
B141::      33 GCACCACCCTGC 44
      ||||  ||  |||
B45:1-50:   8  GCACACCCTTGC 19

=====

Rank: 3  Score: 8   Query: 13 (ss:7,ds:6)
Identity: str: 92%; seq:84% (ss:100%, ds:66%)
Gap: 1 (ss:1, ds:0)  Mismatch: 2 (ss:0, ds:2)
      .(((..... )))
      .(((..... )))
B141::      32 CGCACCACCC-TGC 44
      |||:||||| :||
B84:1-49:   36 CGCGCCACCCTTGC 49

Excluded as does not form in the full-length structure
=====

Rank: 4  Score: 7   Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:60% (ss:50%, ds:66%)
Gap: 0 (ss:0, ds:0)  Mismatch: 4 (ss:2, ds:2)
      (((.....)))
      (((.....)))
B141::      9  CTGACACCGG 18
      |:|  |||:|
B116:1-50:  3  CGGCAACCCG 12

=====

Rank: 4  Score: 7   Query: 16 (ss:6,ds:10)
Identity: str: 100%; seq:31% (ss:50%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 11 (ss:3, ds:8)
      ((.(((.....)))..))
      ((.(((.....)))..))
B141::      6  ACCCTGACACCGGTGT 21
      ::|::|  | |::|::
B167:1-50:  35 TGCAGGTTAACCTTTA 50

=====

```

Appendix 2 continued.

Rank: 6 Score: 6 Query: 11 (ss:5,ds:6)
 Identity: str: 100%; seq:45% (ss:60%, ds:33%)
 Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:2, ds:4)

```

      .(((.....)))
      .(((.....)))
B141::      8 CCTGACACCGG 18
      ||::||  ::|
B38:1-49:   24 CCTGACTGTAG 34

```

Rank: 6 Score: 6 Query: 14 (ss:6,ds:8)
 Identity: str: 93%; seq:71% (ss:66%, ds:75%)
 Gap: 1 (ss:1, ds:0) Mismatch: 4 (ss:2, ds:2)

```

      (.(((.....)))
      (.(((.....)))
B141::      7 CCCTG-ACACCGGTG 20
      | :|| ||| ||:|
B62:1-50:   17 CTATGCACATCGTTG 31

```

Rank: 6 Score: 6 Query: 11 (ss:5,ds:6)
 Identity: str: 100%; seq:45% (ss:60%, ds:33%)
 Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:2, ds:4)

```

      (((.....)))
      (((.....)))
B141::      9 CTGACACCGGT 19
      |:: ||::|
B197:1-50:  25 CTGTTACTAGT 35

```

Rank: 11 Score: 5 Query: 12 (ss:6,ds:6)
 Identity: str: 100%; seq:50% (ss:33%, ds:66%)
 Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:4, ds:2)

```

      (((.....)))
      (((.....)))
B141::      33 GCACCACCCTGC 44
      ||: || :||
B31:1-50:   22 GCGTAACTTCGC 33

```

Rank: 11 Score: 5 Query: 12 (ss:6,ds:6)
 Identity: str: 100%; seq:41% (ss:50%, ds:33%)
 Gap: 0 (ss:0, ds:0) Mismatch: 7 (ss:3, ds:4)

```

      (((.....)))
      (((.....)))
B141::      33 GCACCACCCTGC 44
      :|: || |:|
B114:1-50:  20 GCCATACACGGT 31

```

Appendix 2 continued.

Rank: 11 Score: 5 Query: 10 (ss:4,ds:6)
 Identity: str: 100%; seq:40% (ss:50%, ds:33%)
 Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:2, ds:4)

((.....))

((.....))

B141:: 9 CTGACACCGG 18

|::|| ::|

B156:1-49: 3 CGAACCATCG 12

Rank: 11 Score: 5 Query: 10 (ss:4,ds:6)
 Identity: str: 100%; seq:30% (ss:75%, ds:0%)
 Gap: 0 (ss:0, ds:0) Mismatch: 7 (ss:1, ds:6)

((.....))

((.....))

B141:: 9 CTGACACCGG 18

::: |||:::

B207:1-50: 30 AATCCACGTT 39

Rank: 11 Score: 5 Query: 12 (ss:6,ds:6)
 Identity: str: 100%; seq:33% (ss:66%, ds:0%)
 Gap: 0 (ss:0, ds:0) Mismatch: 8 (ss:2, ds:6)

((.....))

((.....))

B141:: 33 GCACCACCCTGC 44

::: ||| |:::

B229:1-50: 35 TGGACACTCCCG 46

Appendix 2 continued.**156 full vs full**

```
#=== Query ===#
>B156:      B156

      .((((....((.....))....(((....))).....))))).
1  GGGAGACAAGACTAGACGCTCAACACGAACCATCGACCACACCAACTCCA 50

      .((((....(((....((.....).))))))..((((....((.....)
51 AGACCAAATCATTGTGGCAGACTTCGACATGAGACTCACAACAGTTC CCT 100

      ...)))))))).....
101 TTAGTGAGGGTTAATT 116
```

```
#=== Hits ===#
```

```
Rank: 35  Score: 10  Query: 21 (ss:15,ds:6)
Identity: str: 100%; seq:61% (ss:60%, ds:66%)
Gap: 0 (ss:0, ds:0)  Mismatch: 8 (ss:6, ds:2)
      .((((....))).....
      .((((....))).....
B156::      25 ACGAACCATCGACCACACCAA 45
      |||: | |:|||  || || |
B28:1-117:  16 ACGCTCAAGCGAAAACTCCGA 36
```

```
=====  
Rank: 38  Score: 7   Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:50% (ss:75%, ds:33%)
Gap: 0 (ss:0, ds:0)  Mismatch: 5 (ss:1, ds:4)
      (((....)))
      (((....)))
B156::      26 CGAACCATCG 35
      :|:| ||:|:
B4:1-116:   44 GGCAGCAGCT 53
```

```
=====
```

Appendix 2 continued.**B156 full vs variable**

```

#=== Query ===#
>B156:      B156

      .((((....((.....))....(((.....)).....))))).
1  GGGAGACAAGACTAGACGCTCAACACGAACCATCGACCACACCAACTCCA 50

      .((((...(((.(.....).))))))..((((....((.....))
51  AGACCAAATCATTGTGGCAGACTTCGACATGAGACTCACAACAGTTCCCT 100

      ...)))))))).....
101 TTAGTGAGGGTTAATT 116

#=== Hits ===#

Rank: 5  Score: 7  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:50% (ss:75%, ds:33%)
Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:1, ds:4)
      (((.....)))
      (((.....)))
B156::      26 CGAACCATCG 35
      |:| |||:|
B4:1-50:    21 GGCAGCAGCT 30

=====

Rank: 6  Score: 6  Query: 11 (ss:5,ds:6)
Identity: str: 100%; seq:45% (ss:60%, ds:33%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:2, ds:4)
      (((.....))).
      (((.....))).
B156::      26 CGAACCATCGA 36
      |::|| ::||
B38:1-49:    25 CTGACTGTAGA 35

=====

Rank: 13  Score: 5  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:40% (ss:50%, ds:33%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:2, ds:4)
      (((.....)))
      (((.....)))
B156::      26 CGAACCATCG 35
      |:: | |::|
B11:1-49:    25 CTCCTAAAG 34

=====

```

Appendix 2 continued.

```

Rank: 13  Score: 5  Query: 13 (ss:9,ds:4)
Identity: str: 100%; seq:53% (ss:55%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 6 (ss:4, ds:2)
      ((.....)).
      ((.....)).
B156::      10 GACTAGACGCTCA 22
      |: |  || |:||
B40:1-50:   23 GCTTTTACCCGCA 35

```

```

=====

Rank: 13  Score: 5  Query: 12 (ss:6,ds:6)
Identity: str: 100%; seq:41% (ss:50%, ds:33%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:3, ds:4)
      .(((.....))).
      .(((.....))).
B156::      25 ACGAACCATCGA 36
      |: |: |  |: |:|
B145:1-50:  24 AGGGTCGTTCCA 35

```

```

=====

Rank: 13  Score: 5  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:30% (ss:75%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:1, ds:6)
      (((.....)))
      (((.....)))
B156::      26 CGAACCATCG 35
      ::: |||:::
B163:1-49:  12 GATCCCAATC 21

```

Appendix 2 continued.**B156 variable vs variable**

#=== Query ===#

>B156: B156

```

      ..(((.....))).((((.....))))).
1  CACGAACCATCGACCACACCAACTCCAAGACCAAATCATTGTGGCACAC 49

```

#=== Hits ===#

```

Rank: 3  Score: 7  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:50% (ss:75%, ds:33%)
Gap: 0 (ss:0, ds:0) Mismatch: 5 (ss:1, ds:4)
      (((.....)))
      (((.....)))
B156::      3  CGAACCATCG 12
      :|:| ||:|:
B4:1-50:    21 GGCAGCAGCT 30

```

```

Rank: 5  Score: 6  Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:62% (ss:75%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:1, ds:2)
      ((.....))
      ((.....))
B156::      4  GAACCATC 11
      |:|| |:|
B84:1-49:    12 GCACAAGC 19

```

```

Rank: 12  Score: 5  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:40% (ss:50%, ds:33%)
Gap: 0 (ss:0, ds:0) Mismatch: 6 (ss:2, ds:4)
      (((.....)))
      (((.....)))
B156::      3  CGAACCATCG 12
      |::|| |:|
B141:1-50:    9  CTGACACCGG 18

```

```

Rank: 12  Score: 5  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:30% (ss:75%, ds:0%)
Gap: 0 (ss:0, ds:0) Mismatch: 7 (ss:1, ds:6)
      (((.....)))
      (((.....)))
B156::      3  CGAACCATCG 12
      ::: |||:::
B163:1-49:    12 GATCCCAATC 21

```

Appendix 2 continued.**B163 full vs full**

No significant hits

B163 full vs variable

#=== Query ===#

>B163: B163

```

      .(....).....(((((((..(((((.....)))))).(
1   GGGAGACAAGACTAGACGCTCAACTGTTACGATAGATCCCAATCTAGCAT 50

      ((..(((.(.(((.....)))))))))))).))))))(((
51  GACATCTCTATGCTTGTTACGCTTCGACATGAGACTCACAACAGTTCCT 100

      (.....))))).....
101 TTAGTGAGGGTTAATT 116

```

#=== Hits ===#

```

Rank: 4   Score: 12   Query: 15 (ss:5,ds:10)
Identity: str: 100%; seq:53% (ss:80%, ds:40%)
Gap: 0 (ss:0, ds:0)   Mismatch: 7 (ss:1, ds:6)

```

```

      .((((.....))))
      .((((.....))))
B163::      32 ATAGATCCCAATCTA 46
      ||::||| ||::|
B11:1-49:   22 ATCCTTCCTAAAGGA 36

```

```

Rank: 5   Score: 10   Query: 22 (ss:10,ds:12)
Identity: str: 100%; seq:40% (ss:50%, ds:33%)
Gap: 0 (ss:0, ds:0)   Mismatch: 13 (ss:5, ds:8)

```

```

      .(((.(.....))))
      .(((.(.....))))
B163::      59 TATGCTTGTTACGCTTCGACAT 80
      ||:||::| ||::|
B62:1-50:   18 TATGCACATCGTTGTTTGTTCGT 39

```

Excluded as does not form in full-length structure

```

Rank: 6   Score: 9   Query: 14 (ss:8,ds:6)
Identity: str: 100%; seq:57% (ss:75%, ds:33%)
Gap: 0 (ss:0, ds:0)   Mismatch: 6 (ss:2, ds:4)

```

```

      (((.....)))
      (((.....)))
B163::      64 TTGTTACGCTTCGA 77
      :|:| ||||:|:
B30:1-50:   29 GTATCCCGCTTTGC 42

```

Appendix 2 continued.

```

=====
Rank: 9  Score: 7  Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:28% (ss:50%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 10 (ss:2, ds:8)
      ((((((.....))))))
      ((((((.....))))))
B163::      33 TAGATCCCAATCTA 46
      :|::: ||:::|:
B4:1-50:    19 TAGGCAGCAGCTTG 32

```

```

=====
Rank: 9  Score: 7  Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:28% (ss:50%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 10 (ss:2, ds:8)
      ((((((.....))))))
      ((((((.....))))))
B163::      33 TAGATCCCAATCTA 46
      :::|:| | :|:::
B86:1-50:    35 GCTAGCCTCTTAGC 48

```

```

=====
Rank: 14  Score: 6  Query: 12 (ss:4,ds:8)
Identity: str: 100%; seq:33% (ss:50%, ds:25%)
Gap: 0 (ss:0, ds:0)  Mismatch: 8 (ss:2, ds:6)
      ((((((.....))))))
      ((((((.....))))))
B163::      34 AGATCCCAATCT 45
      :::|| | | |:::
B229:1-50:   3  GCCTCACGAGGC 14

```

```

=====
Rank: 20  Score: 5  Query: 16 (ss:6,ds:10)
Identity: str: 100%; seq:18% (ss:50%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 13 (ss:3, ds:10)
      .(((((((.....)))))).
      .(((((((.....)))))).
B163::      32 ATAGATCCCAATCTAG 47
      |:::| | | |:::|
B217:1-50:   35 ACTTCGCATGCGAGGG 50

```

```

=====
Rank: 34  Score: 4  Query: 12 (ss:8,ds:4)
Identity: str: 100%; seq:41% (ss:62%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:3, ds:4)
      ((.....))
      ((.....))
B163::      65 TGTTACGCTTCG 76
      :: || |||::
B45:1-50:    8  GCACACCCTTGC 19

```

Appendix 2 continued.

```

Rank: 34  Score: 4  Query: 12 (ss:4,ds:8)
Identity: str: 100%; seq:16% (ss:50%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 10 (ss:2, ds:8)

```

```

      (((.....)))

```

```

      (((.....)))

```

```

B163::          34 AGATCCCAATCT 45

```

```

      ::::||  ::::

```

```

B141:1-50:    11 GACACCGGTGTT 22

```

```

=====

```

```

Rank: 34  Score: 4  Query: 12 (ss:4,ds:8)
Identity: str: 100%; seq:25% (ss:25%, ds:25%)
Gap: 0 (ss:0, ds:0)  Mismatch: 9 (ss:3, ds:6)

```

```

      (((.....)))

```

```

      (((.....)))

```

```

B163::          34 AGATCCCAATCT 45

```

```

      |:::|  :::|

```

```

B199:1-50:    20 ATCGCTTTCGAT 31

```

```

=====

```

Appendix 2 continued.**B163 variable vs variable**

```

#=== Query ===#
>B163:      B163

      ..((.((((((((.....))))))((((.....)))))))).)
1  CTGTTACGATAGATCCCAATCTAGCATGACATCTCTATGCTTGTTACGC 49

#=== Hits ===#

Rank: 2  Score: 11  Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:50% (ss:75%, ds:40%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:1, ds:6)
      (((((.....))))))
      (((((.....))))))
B163::      10 TAGATCCCAATCTA 23
      |:::||| ||:::|
B11:1-49:   23 TCCTTCCTAAAGGA 36

=====

Rank: 3  Score: 7   Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:28% (ss:50%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 10 (ss:2, ds:8)
      (((((.....))))))
      (((((.....))))))
B163::      10 TAGATCCCAATCTA 23
      :|:::  ||:::|:
B4:1-50:    19 TAGGCAGCAGCTTG 32

=====

Rank: 3  Score: 7   Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:28% (ss:50%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 10 (ss:2, ds:8)
      (((((.....))))))
      (((((.....))))))
B163::      10 TAGATCCCAATCTA 23
      :::|:|  |::::
B86:1-50:   35 GCTAGCCTCTTAGC 48

=====

Rank: 3  Score: 7   Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:50% (ss:75%, ds:33%)
Gap: 0 (ss:0, ds:0)  Mismatch: 5 (ss:1, ds:4)
      (((.....)))
      (((.....)))
B163::      12 GATCCCAATC 21
      :|:|  ||:|:
B223:1-50:  18 TAGCGCACTG 27

=====

```

Appendix 2 continued.

```

Rank: 6  Score: 6  Query: 12 (ss:4,ds:8)
Identity: str: 100%; seq:33% (ss:50%, ds:25%)
Gap: 0 (ss:0, ds:0)  Mismatch: 8 (ss:2, ds:6)
      (((.....)))
      (((.....)))
B163::      11 AGATCCCAATCT 22
      ::|||  :|::
B207:1-50:  29 TAATCCACGTTA 40

```

```

Rank: 6  Score: 6  Query: 12 (ss:4,ds:8)
Identity: str: 100%; seq:33% (ss:50%, ds:25%)
Gap: 0 (ss:0, ds:0)  Mismatch: 8 (ss:2, ds:6)
      (((.....)))
      (((.....)))
B163::      11 AGATCCCAATCT 22
      ::|||  | |::
B229:1-50:   3  GCCTCACGAGGC 14

```

```

Rank: 8  Score: 5  Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:21% (ss:25%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 11 (ss:3, ds:8)
      ((((((.....))))))
      ((((((.....))))))
B163::      10 TAGATCCCAATCTA 23
      :|:::|  ::|:
B116:1-50:   1  GACGGCAACCCGTT 14

```

```

Rank: 8  Score: 5  Query: 10 (ss:4,ds:6)
Identity: str: 100%; seq:30% (ss:75%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:1, ds:6)
      (((.....)))
      (((.....)))
B163::      12 GATCCCAATC 21
      ::: |||:::
B156:1-49:   3  CGAACCATCG 12

```

```

Rank: 8  Score: 5  Query: 14 (ss:4,ds:10)
Identity: str: 100%; seq:21% (ss:25%, ds:20%)
Gap: 0 (ss:0, ds:0)  Mismatch: 11 (ss:3, ds:8)
      ((((((.....))))))
      ((((((.....))))))
B163::      10 TAGATCCCAATCTA 23
      |:::  |:::|
B167:1-50:  12 TCTGCTTGAGCGGA 25

```

Appendix 2 continued.

Rank: 15 Score: 4 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:16% (ss:50%, ds:0%)
 Gap: 0 (ss:0, ds:0) Mismatch: 10 (ss:2, ds:8)

(((((....))))))

(((((....))))))

B163:: 11 AGATCCCAATCT 22

::::|| ::::

B141:1-50: 11 GACACCGGTGTT 22

=====

Rank: 15 Score: 4 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:25% (ss:25%, ds:25%)
 Gap: 0 (ss:0, ds:0) Mismatch: 9 (ss:3, ds:6)

(((((....))))))

(((((....))))))

B163:: 11 AGATCCCAATCT 22

|:::| :::|

B199:1-50: 20 ATCGCTTTCGAT 31

=====

Rank: 38 Score: 3 Query: 14 (ss:4,ds:10)
 Identity: str: 100%; seq:7% (ss:25%, ds:0%)
 Gap: 0 (ss:0, ds:0) Mismatch: 13 (ss:3, ds:10)

(((((....))))))

(((((....))))))

B163:: 10 TAGATCCCAATCTA 23

:::::| :::::

B217:1-50: 36 CTTGCATGCGAGG 49

=====

Appendix 2 continued.**B229 full vs full**

No significant hits

B229 full vs variable

#=== Query ===#

>B229: B229

```

      ((((((.....((((((((.....)))).....)))..((
1   GGGAGACAAGACTAGACGCTCAACAGCCTCACGAGGCCACTTATGCAAAT 50

      (((((..((((.....)))).....))))))....))....((
51  GTCGCAATGGACACTCCCGTACCTTCGACATGAGACTCACACCAGTTCCC 100

      ((.....)))).....
101 TTTAGTGAGGGTTAATT 117

```

#=== Hits ===#

```

Rank: 4   Score: 9   Query: 15 (ss:7,ds:8)
Identity: str: 100%; seq:53% (ss:57%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 7 (ss:3, ds:4)
      .((((.....))))
      .((((.....))))
B229::      56 AATGGACACTCCCGT 70
      ||::|  |||  |::|
B31:1-50:   20 AAGCGTAACTTCGCT 34

```

```

=====

Rank: 4   Score: 9   Query: 19 (ss:11,ds:8)
Identity: str: 95%; seq:63% (ss:90%, ds:25%)
Gap: 1 (ss:1, ds:0) Mismatch: 7 (ss:1, ds:6)
      ..((((.....)))).....
      ..((((.....)))).....
B229::      24 CAGCCT-CACGAGGCCACTT 42
      ||::|  |||||:::|  |||
B185:1-50:   7  CATGGTGCACGACCACTCTT 26

```

```

=====

Rank: 7   Score: 8   Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:75% (ss:100%, ds:50%)
Gap: 0 (ss:0, ds:0) Mismatch: 2 (ss:0, ds:2)
      ((.....))
      ((.....))
B229::      28 CTCACGAG 35
      |:||||:|
B40:1-50:   32 CGCACGCG 39

```

Appendix 2 continued.

Rank: 7 Score: 8 Query: 10 (ss:6,ds:4)
 Identity: str: 100%; seq:70% (ss:83%, ds:50%)
 Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:1, ds:2)

```

      ((.....))
      ((.....))
B229::      59 GGACACTCCC 68
      :|||| |||:
B197:1-50:  37 CGACAATCCG 46

```

=====

Rank: 15 Score: 6 Query: 10 (ss:6,ds:4)
 Identity: str: 100%; seq:70% (ss:50%, ds:100%)
 Gap: 0 (ss:0, ds:0) Mismatch: 3 (ss:3, ds:0)

```

      ((.....))
      ((.....))
B229::      59 GGACACTCCC 68
      || || | ||
B33:1-50:   5  GGCCAATTCC 14

```

=====

Rank: 15 Score: 6 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:33% (ss:50%, ds:25%)
 Gap: 0 (ss:0, ds:0) Mismatch: 8 (ss:2, ds:6)

```

      (((.....)))
      (((.....)))
B229::      26 GCCTCACGAGGC 37
      ::|:| |:|::
B141:1-50:  11 GACACCGGTGTT 22

```

=====

Rank: 15 Score: 6 Query: 12 (ss:4,ds:8)
 Identity: str: 100%; seq:33% (ss:50%, ds:25%)
 Gap: 0 (ss:0, ds:0) Mismatch: 8 (ss:2, ds:6)

```

      (((.....)))
      (((.....)))
B229::      26 GCCTCACGAGGC 37
      ::|| | |:|:
B163:1-49:  11 AGATCCCAATCT 22

```

=====

Rank: 15 Score: 6 Query: 18 (ss:10,ds:8)
 Identity: str: 100%; seq:38% (ss:50%, ds:25%)
 Gap: 0 (ss:0, ds:0) Mismatch: 11 (ss:5, ds:6)

```

      (((.....))).....
      (((.....))).....
B229::      26 GCCTCACGAGGCCACTTA 43
      ::|:| |:|: ||||
B199:1-50:  20 ATCGCTTTCGATTTCCTTA 37

```

Appendix 2 continued.

```

=====
Rank: 39  Score: 4  Query: 14 (ss:6,ds:8)
Identity: str: 100%; seq:21% (ss:50%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 11 (ss:3, ds:8)
      (((.....)))
      (((.....)))
B229::      57 ATGGACACTCCCGT 70
      ::::|  ||:::
B36:1-50:   17 GGACAACTTCGTTC 30

=====
Rank: 39  Score: 4  Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:50% (ss:50%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 4 (ss:2, ds:2)
      ((.....))
      ((.....))
B229::      28 CTCACGAG 35
      |:||  :|
B212:1-50:  24 CGCAAACG 31

=====

```

Appendix 2 continued.**B229 variable vs variable**

```

#=== Query ===#
>B229:      B229

      ..((((.....)))).....((((.....))))....
1  CAGCCTCACGAGGCCACTTATGCAAATGTCGCAATGGACACTCCCGTACC 50

#=== Hits ===#

Rank: 2  Score: 9  Query: 15 (ss:7,ds:8)
Identity: str: 100%; seq:53% (ss:57%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:3, ds:4)
      .((((.....))))
      .((((.....))))
B229::      33 AATGGACACTCCCGT 47
      ||::|  |||  |::|
B31:1-50:   20 AAGCGTAACTTCGCT 34

=====

Rank: 2  Score: 9  Query: 19 (ss:11,ds:8)
Identity: str: 95%; seq:63% (ss:90%, ds:25%)
Gap: 1 (ss:1, ds:0)  Mismatch: 7 (ss:1, ds:6)
      ..((((.....)))).....
      ..((((.....)))).....
B229::      1 CAGCCT-CACGAGGCCACTT 19
      ||::|  |||||::|  |||
B185:1-50:   7 CATGGTGCACGACCACTCTT 26

=====

Rank: 4  Score: 8  Query: 8 (ss:4,ds:4)
Identity: str: 100%; seq:75% (ss:100%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 2 (ss:0, ds:2)
      ((.....))
      ((.....))
B229::      5 CTCACGAG 12
      |:||||:|
B40:1-50:   32 CGCACGCG 39

=====

Rank: 4  Score: 8  Query: 10 (ss:6,ds:4)
Identity: str: 100%; seq:70% (ss:83%, ds:50%)
Gap: 0 (ss:0, ds:0)  Mismatch: 3 (ss:1, ds:2)
      ((.....))
      ((.....))
B229::      36 GGACACTCCC 45
      :||||  |||:
B197:1-50:   37 CGACAATCCG 46

=====

```

Appendix 2 continued.

```

Rank: 4  Score: 8  Query: 12 (ss:4,ds:8)
Identity: str: 100%; seq:41% (ss:75%, ds:25%)
Gap: 0 (ss:0, ds:0)  Mismatch: 7 (ss:1, ds:6)
      (((.....)))
      (((.....)))
B229::      3  GCCTCACGAGGC 14
      ::||| |::
B217:1-50:  37 TTCGCATGCGAG 48

```

```
=====
```

```

Rank: 7  Score: 7  Query: 13 (ss:5,ds:8)
Identity: str: 100%; seq:38% (ss:60%, ds:25%)
Gap: 0 (ss:0, ds:0)  Mismatch: 8 (ss:2, ds:6)
      (((.....))).
      (((.....))).
B229::      3  GCCTCACGAGGCC 15
      |:: || ::||
B207:1-50:  1  GTGTTACTGCACC 13

```

```
=====
```

```

Rank: 9  Score: 6  Query: 10 (ss:6,ds:4)
Identity: str: 100%; seq:70% (ss:50%, ds:100%)
Gap: 0 (ss:0, ds:0)  Mismatch: 3 (ss:3, ds:0)
      ((.....))
      ((.....))
B229::      36 GGACACTCCC 45
      || || | ||
B33:1-50:   5  GGCCAATTCC 14

```

```
=====
```

```

Rank: 9  Score: 6  Query: 12 (ss:4,ds:8)
Identity: str: 100%; seq:33% (ss:50%, ds:25%)
Gap: 0 (ss:0, ds:0)  Mismatch: 8 (ss:2, ds:6)
      (((.....)))
      (((.....)))
B229::      3  GCCTCACGAGGC 14
      ::|| |::
B141:1-50:  11 GACACCGGTGTT 22

```

```
=====
```

```

Rank: 9  Score: 6  Query: 12 (ss:4,ds:8)
Identity: str: 100%; seq:33% (ss:50%, ds:25%)
Gap: 0 (ss:0, ds:0)  Mismatch: 8 (ss:2, ds:6)
      (((.....)))
      (((.....)))
B229::      3  GCCTCACGAGGC 14
      ::|| |::
B163:1-49:  11 AGATCCCAATCT 22

```

```
=====
```

Appendix 2 continued.

```

=====
Rank: 9  Score: 6  Query: 18 (ss:10,ds:8)
Identity: str: 100%; seq:38% (ss:50%, ds:25%)
Gap: 0 (ss:0, ds:0)  Mismatch: 11 (ss:5, ds:6)
      (((((.....)))).....
      (((((.....)))).....
B229::      3  GCCTCACGAGGCCACTTA 20
      ::|||  :|::  ||||
B199:1-50: 20 ATCGCTTTCGATTTCCTTA 37

```

```

=====
Rank: 32  Score: 4  Query: 14 (ss:6,ds:8)
Identity: str: 100%; seq:21% (ss:50%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 11 (ss:3, ds:8)
      (((((.....))))
      (((((.....))))
B229::      34 ATGGACACTCCCGT 47
      ::::|  ||:::
B36:1-50: 17 GGACAACTTCGTTC 30

```

```

=====
Rank: 32  Score: 4  Query: 14 (ss:6,ds:8)
Identity: str: 93%; seq:57% (ss:66%, ds:50%)
Gap: 1 (ss:1, ds:0)  Mismatch: 6 (ss:2, ds:4)
      (((((.... ..))))
      (((((.....))))
B229::      34 ATGGACAC-TCCCGT 47
      |::|  |||  ||::|
B45:1-50: 30 ATCGCCACAACCGAT 44

```

```

=====
Rank: 32  Score: 4  Query: 16 (ss:8,ds:8)
Identity: str: 94%; seq:43% (ss:87%, ds:0%)
Gap: 0 (ss:0, ds:0)  Mismatch: 9 (ss:1, ds:8)
      .(((.....)))).
      .(((.... ..)))).
B229::      33 AATGGACACTCCCGTAC 49
      |::::  ||  ||::::|
B137:1-50: 4  AGGGTTCA-TCATTAC 19

```