Expression and Purification of Recombinant HIV-1 BH10 Tat Protein

BY

KELVIN B. LUTHER

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Chemistry
University of Manitoba
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

Expression and Purification of Recombinant HIV-1 BH10 Tat Protein

BY

Kelvin B. Luther

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree of Master of Science

Table of Contents

Abstract

The Human Immunodeficiency Virus (HIV) is the causative agent of acquired immune deficiency syndrome (AIDS). The HIV-1 BH10 genome encodes a two-exon gene for an 86 amino acid viral transactivator protein (Tat) that markedly enhances viral transcription. The objective of the present research was to develop an expression and purification system to produce HIV-1 BH10 Tat_{1-86} for structural analysis. Large amounts (mg quantities) of pure, conformationally homogenous protein are required for high-resolution structural analysis by X-ray diffraction and NMR spectroscopy. High-resolution structures of Tat, alone and in complex with viral and cellular targets, are necessary for a detailed atomic level explanation of the role of Tat in the transactivation of transcription. Tat contains a cysteine-rich region, a core hydrophobic region, and a basic region, which comprise the minimal transactivation domain of the protein. These sequences also present several problems for purification. The numerous cysteine residues may give rise to micro-heterogeneity caused by oligomerization and the basic nature of Tat increases the likelihood of nucleic acid contamination. We first attempted to express Tat protein with a PelB leader sequence for secretion into the periplasmic space of *E. coli*. Purification using the PelB leader would be potentially very efficient due to low levels of contaminating proteins and the absence of nucleic acids in the periplasmic space. Unfortunately, this system failed to produce Tat. We next expressed Tat with a cleavable 6-histidine purification tag at the amino terminus. This tag allowed purification of Tat to greater than 90 % using immobilized metal-affinity chromatography. This system produced approximately1/2 mg of purified Tat per litre of culture. The presence of several codons in the HIV-1 BH10 *tat* cDNA that are rarely used in *E. coli* was investigated as a cause of low levels of protein expression. To this end, a *tat* cDNA was produced by splicing nucleotides 1-186 of a Tat gene composed of frequently used *E. coli* codons with nucleotides 187-258 of the HIV-1 BH10 *tat* cDNA. *E. coli* harbouring the

new plasmid yielded amounts of His_6-Tat_{1-86} several fold higher than the original construct. In summary, His_6-Tat_{1-86} has been purified in one step using immobilised metal-affinity chromatography, and expression levels of Tat have been elevated using a codon sequence substantially optimised for expression in *E. coli*.

Acknowledgements

First I would like to thank Dr. Joe O'Neil for giving me an opportunity to do research and for his infinite patience in guiding my education. It is only now that I truly realise how little I knew when I began my project, and how very much I have learned under his tutelage. I thank Dr. Secco and Dr. Court for agreeing to be on my committee, and for all the effort they put into critiquing my thesis. I also owe an enormous debt of gratitude to Dr. Gillian Henry who helped guide my research and was an unending source of encouragement and advice. I apologize for having made your eyes role so many times. Darren and Xing deserve special thanks for their friendship and for making the lab a great place to be each day. The Duckworth, Jamieson, and Loewen labs have earned my gratitude for the many times they loaned equipment, resources, and advice, and in particular, Dr. Linda Donald and Mrs. Kosala Sivananthan for helpful discussions. I would also like to thank Dr. Wasim Haque for his advice and many interesting discussions. Thanks go to Dr. Annie Scoot for her encouragement, advice, and support; you have made me a better teacher. I thank the staff in the Chemistry Main Office who were always willing to help when it was needed, and did it with a laugh and a smile. I would like to thank all of the graduate students who have offered me their friendship and support while I was here, in particular the climbing club (you know who you are), and most of all Pam. Finally, I would like to thank my parents. Without their unending and unquestioning support I could never have attended university, and would never have made it to graduate school.

I would like to acknowledge and thank MRC and NSERC for funding.

List of Figures

List of Tables

List of Abbreviations

Å	Angstrom
µg	micro-gram
µL	micro-Litre
µM	micro-Molar
A_{260}	ultra-violet/visible absorbance at 260 nm
A_{600}	ultra-violet/visible absorbance at 600 nm
AIDS	Acquired Immune Deficiency Syndrome
ARM	arginine rich motif
ATP	adenosine triphosphate
bp	base pair(s)
BSA	Bovine serum albumin
CA	capsid protein
CE	cellulose ester
CD4	cluster of differentiation 4
CD26	cluster of differentiation 26
CDK	cyclin dependent kinase
CIP	calf intestinal alkaline phosphatase
cm	centi-metre
CTD	carboxy-terminal domain
DAB	3,3' diaminobenzidine tetrahydrochloride
ddH_2O	doubly deionised water
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	ethylenediamine tetra-acetic acid
EF-1 δ	elongation factor-1 delta
EIAV	equine infectious anaemia virus
g	gravity
HF	high fidelity
HIV-1 BH10	Human Immunodeficiency Virus 1 strain BH10
IEC	International Equipment Company
IgG	immuno-globulin G
IN	integrase
IPTG	isopropyl β-D thiogalactopyranoside
LB	Luria-Bertani
LTR	long terminal repeat
M	molar
MA	matrix protein
mg	milli-gram
mL	milli-Litre
mM	milli-Molar
MHC	major histo-compatibility complex
ng	nano-gram
nL	nano-Litre

nM	nano-Molar
NC	nucleocapsid protein
NF-κB	nuclear factor kappa B
NIAID	National Institute of Allergy and Infectious Diseases
NIH	National Institutes of Health
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NTA	nitrilotriacetate
NTP	nucleoside triphosphate
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PelB	pectate lyase B
pmol	pico-mole
PMSF	phenylmethylsulfonyl fluoride
PR	protease
psi	pounds per square inch
pTEFb	positive transcriptional elongation factor b
RCF	relative centrifugal force
RGD	arginine-glycine-aspartate
RNA	ribonucleic acid
RNAP II	RNA polymerase II
RNase A	ribonuclease A
RNase H	ribonuclease H
rpm	revolutions per minute
RT	reverse transcriptase
SDS	sodium dodecylsulphate
SU	surface protein
TAE buffer	Tris-acetate EDTA buffer
TAK	Tat-associated kinase
TAR	transactivation response region
Tat	transactivator of transcription
TB	Terrific Broth
TBE buffer	Tris-borate EDTA buffer
TCEP	Tris (2-carboxyethyl) phosphine-HCl
TE buffer	Tris-EDTA buffer
TFIIH	transcription factor II H
TM	trans-membrane protein
Tris-HCl	Tris (hydroxymethyl) aminomethane hydrochloride
$tRNA^{lys}$	transfer-RNA lysine
U	units
UCDNA	University Core DNA
UV	ultra-violet

Chapter One: Introduction

1.0 Composition of the HIV-1

The Human Immunodeficiency Virus-1 (HIV-1) is a member of the family *Retroviridae* and the genus *Lentivirus*. HIV-1 is the aetiological agent of Human Acquired Immunodeficiency Syndrome (Jeang *et al.*, 1999). The HIV-1 virion contains a diploid, single-stranded plus-sense RNA genome. Each chromosome contains a structurally identical hairpin. The loop sequences are complementary and form a six base pair helix joining the two chromosomes together (Frankel and Young, 1998). The RNA is contained within an icosahedral nucleocapsid cone composed of the CA (capsid) protein (see Figure 1.1) (White and Fenner, 1994). The RNA is coated by a viral protein named nucleocapsid (NC) (see Figure 1.1) (Frankel and Young, 1998). The association of NC and the viral RNA is known as nucleoprotein (Frankel and Young, 1998). The core also contains a reverse transcriptase with RNase H activity (RT), an integrase (IN), a protease (PR), as well as viral accessory proteins Vpr, Vif, Nef and p6 (White and Fenner 1994, Frankel and Young 1998). The core is surrounded by a lipid membrane from which 72 oligomers of the SU (surface) and TM (transmembrane) glycoproteins protrude (White and Fenner 1994) (see Figure 1.1). In addition, the viral envelope and capsid may contain cellular host proteins, particularly class I and II MHC (Major Histo-compatibility Complex) antigens (White and Fenner 1994, Folks and Hart 1997). A myristoylated protein named MA (matrix) lines the inner surface of the viral membrane (see Figure 1.1) (Frankel and Young, 1998).

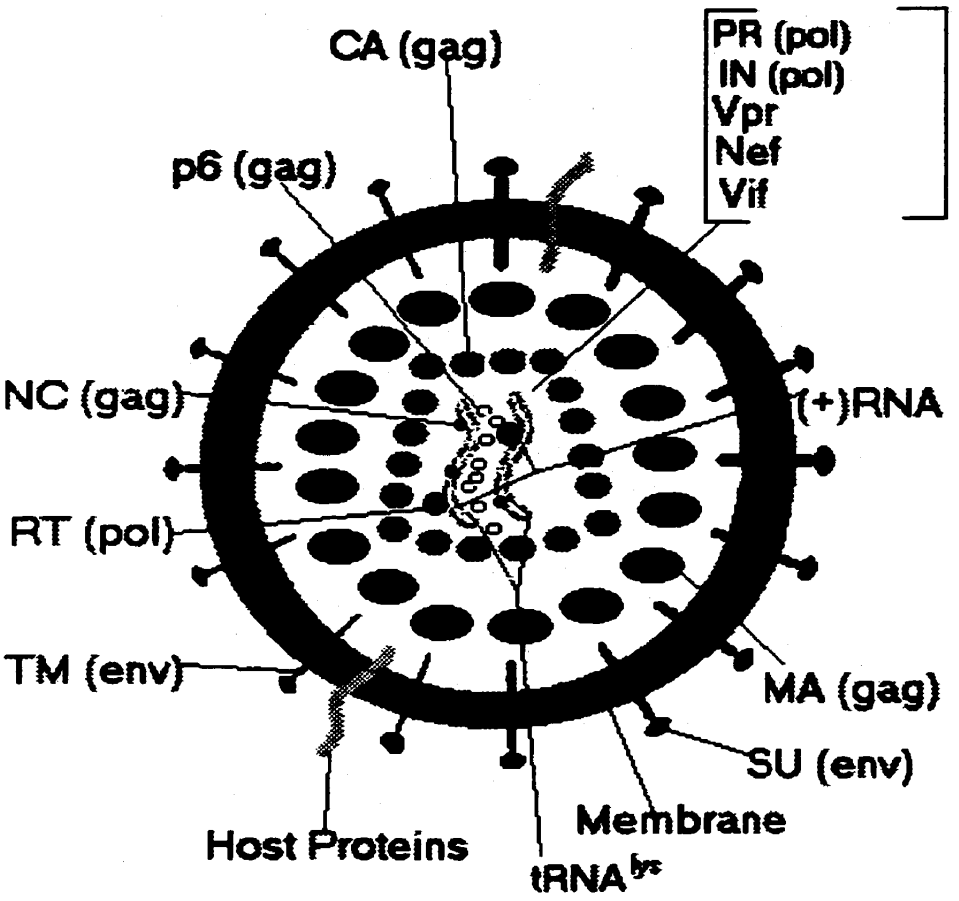

Figure 1.1 The HIV-1 Virion (White and Fenner, 1994; Frankel and Young, 1998, redrawn). A cartoon of the HIV-1 virion, showing the diploid single stranded plus sense RNA genome, associated reverse transcriptase, viral structural proteins, and viral envelope and associated host cell-derived proteins. The protein abbreviations are capitalised and their origin in the HIV-1 genome is indicated in brackets. MA is also known as p17, CA is known as p24, and NC is known as p7.

The genomes of all retroviruses contain the gag, pol, and env genes flanked by Long Terminal Repeats (LTR) required for integration into the host DNA as well as regulation of transcription (see Figure 1.2) (White and Fenner, 1994). In addition to these three genes, the HIV-1 genome encodes regulatory/accessory proteins called Tat (transactivator of transcription), Rev (regulator of expression of the virion), Nef (negative factor), Vpu (viral protein U), Vpr (viral protein R), and Vif (viral infectivity factor) (see Figure 1.2) (White and Fenner, 1994; Frankel and Young, 1998). Many of the HIV mRNAs are multi-cistronic with reading frames that overlap (see Figure 1.2). Thus, many HIV-1 proteins are produced by ribosomal frame-shifting and initiation site scan through. Three types of mRNA result from transcription of the HIV-1 genome. Full length (9.2 kb), unspliced mRNAs serve as templates for the Gag and Pol proteins and are produced late in gene expression (White and Fenner, 1994). The gag reading frame is translated as a Gag polyprotein, which is cleaved into its constituent proteins by the viral PR enzyme. Approximately 10% of the time, the ribosome slips at the 3' end of the gag mRNA during translation, causing a -1 base pair frameshift and the translation of a Gag-Pol polyprotein, which is later, cleaved into its constituent proteins by the PR enzyme (White and Fenner, 1994; Folks and Hart, 1997). The gag gene encodes core structural proteins (MA, CA, NC, and p6) (White and Fenner 1994, Frankel and Young 1998, Folks and Hart 1997). The pol gene encodes the enzymes RT, IN, and PR (White and Fenner 1994, Folks and Hart 1997, Frankel and Young 1998). A shorter (4.5 kb), singly spliced mRNA is also produced late in replication and encodes the Env, Vif, Vpu, Vpr and Tat proteins (White and Fenner, 1994). The env gene encodes gp160, the viral envelope protein precursor, which is cleaved by a host cell protease into the TM and SU

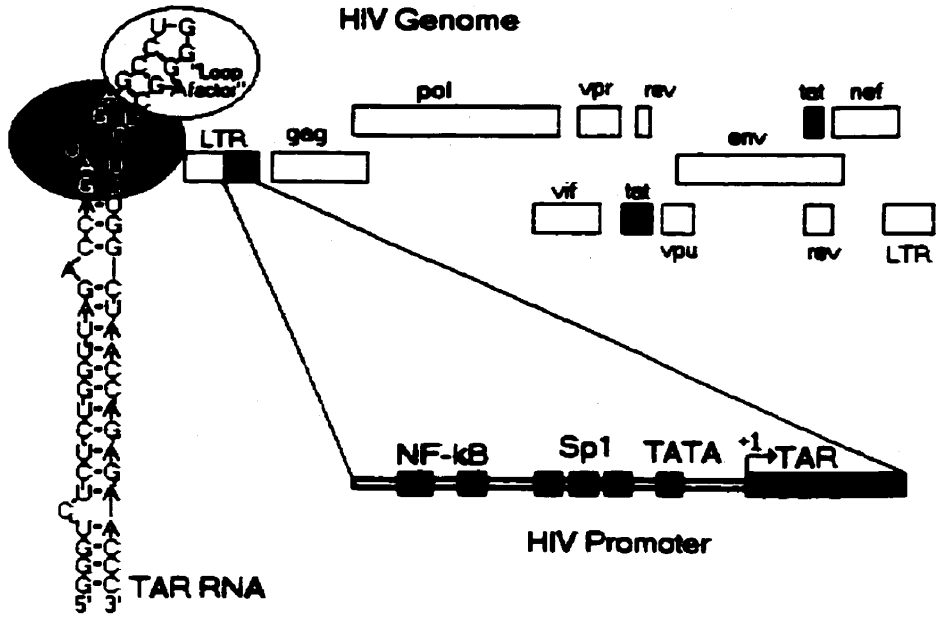

Figure 1.2 The HIV-1 Genome (Figure 2 from Karn, 1999, redrawn). The HIV-1 genome and promoter showing the position of the LTR, the three Sp1 binding sites in tandem, the TATA box, and the position of the TAR stem-loop structure. Also pictured is the TAR RNA stem-loop. The dark oval represents Tat, and the white oval the binding site for cyclin T_1 (see Section 1.1) (Wei *et al.*, 1998). The bold nucleotides in the Tat binding site are those that are critical for Tat binding. NF-κB is an inducible host-cell transactivation protein that can bind the HIV LTR. The effect of NF-κB on viral transcription has not been completely elucidated, but it is likely that it transactivates viral transcription only under certain circumstances, and is not responsible for the initial levels of viral transcription after the host cell has been infected.

glycoproteins (White and Fenner 1994, Frankel and Young 1998, Folks and Hart 1997). The open reading frame for the Vpu mRNA has a sub-optimal start site that the ribosome generally passes over for the stronger Env start site from which an Env polyprotein is translated (Folks and Hart, 1997). Thus, the amount of Vpu protein produced will depend on the frequency with which the ribosome initiates translation from the sub-optimal initiation site of the *vpu* mRNA. Tat, Rev, and Nef proteins are encoded by shorter (2 kb) doubly-spliced mRNAs that are produced early in the replication cycle (White and Fenner, 1994).

1.1 The HIV-1 Life Cycle

The transmission of HIV from one person to another is dependent on the exchange of infectious bodily fluids (White and Fenner 1994). However the virus must attach to specific cell receptors to achieve a productive infection of a host cell. There are two main host cell types in humans, the CD4+ helper T-cell, and the macrophage (White and Fenner 1994). The HIV SU protein mediates attachment of the virus to the CD4 protein expressed on these cells (Dalgleish *et al.*, 1984; Klatzmann *et al.*, 1984a and 1984b; McDougal *et al.*, 1986). In order for the virus to penetrate the cell there must also be available a specific co-receptor protein on the cell surface (Folks and Hart 1997). The two most important co-receptors are CCR5 in macrophages and CXCR4/fusin in CD4+ T-cells (Feng *et al.*, 1996; Moore, 1997). In general, an individual virion will be capable of infecting only one of the two cell types (Frankel and Young 1998, Folks and Hart 1997).

Once inside the host cell cytoplasm the nucleoprotein complex consists of the Gag and Pol proteins and genomic viral RNA. Reverse transcription of viral RNA is primed by a cellular-derived $tRNA^{Lys}$ packaged with the genome in the virion (White and Fenner 1994). The viral RT enzyme produces a DNA copy of the single-stranded RNA template (White and Fenner 1994). The RNase H activity of the enzyme (in a second domain) degrades the RNA portion of the RNA/DNA duplex during reverse transcription (White and Fenner 1994). In the process, a small oligomer of RNA is left behind, forming a short double-stranded RNA/DNA region (White and Fenner 1994, Folks and Hart 1997). This short stretch of RNA then serves as a primer for synthesis of the second strand of the DNA duplex and the viral RT enzyme polymerises the complementary DNA strand (White and Fenner, 1994; Folks and Hart, 1997; Frankel and Young, 1998). After production of a double-stranded DNA copy of the genome is completed the pre-integration complex contains the viral Gag matrix, Pol integrase, Pol RT, and Vpr proteins. The matrix and Vpr proteins mediate transport of the pre-integration complex to the nucleus. Next, the viral IN enzyme inserts the DNA into the host cell genome in several locations (White and Fenner 1994, Folks and Hart 1997, Frankel and Young 1998). It was long accepted that sites of pro-virus integration are random. However, Shiramizu *et al.* (1994), investigating B-cell lymphomas, have found that there may indeed be non-random integration in at least some instances, possibly contributing to cases of AIDS-related diseases such as Kaposi's sarcoma.

Once the DNA copy of the viral genome is integrated into the host cell, low levels of viral proteins are produced through transcription initiated by constitutive host cell

factors, such as Sp1 and the TATA-binding factors (White and Fenner, 1994; Folks and Hart, 1997). These transcription-inducing proteins bind specific regions of the 5' LTR (long terminal repeat) (refer to Figure 1.2) (White and Fenner, 1994; Folks and Hart, 1997). The LTR is a region of repeated sequences at each end of the genome, and contains the viral transcription promoter site at the 5' end (Folks and Hart, 1997). The LTRs are also important for reverse transcription and insertion of the viral DNA into the host cell genome (Folks and Hart, 1997.) Initially only 2 kb *tat*, *rev*, and *nef* transcripts, produced from multiply-spliced mRNAs are exported to the cytoplasm (White and Fenner, 1994; Folks and Hart, 1997). Tat protein is a potent viral transactivator that, once produced, greatly enhances the production of viral mRNA (White and Fenner, 1994; Folks and Hart, 1997). The MA portion of the Gag polyprotein targets Gag to the plasma membrane for eventual assembly of new virus particles (Schultz and Oroszian, 1983; Chazal *et al.*, 1995). TM forms a trimer that spans the plasma membrane, and each TM monomer is non-covalently linked to an SU molecule (which extends out from the membrane) (Folks and Hart, 1997; Frankel and Young, 1998). MA assists TM and SU as they insert into the membrane (Freed *et al.*, 1995; Mammano *et al.*, 1995) The PR enzyme cleaves the Gag-Pol polyprotein (Kaplan *et al.* 1994), and the structural proteins assemble into a mature virion and a new virus buds from the host cell membrane (Folks and Hart, 1997; Frankel and Young, 1998). In this final process, the p6 protein appears to play a role in incorporating Vpr into the mature virion (Lu *et al.*, 1995; Kondo and Gottlinger, 1996).

1.2 Tat Protein Function

The integrated HIV-1 genome is subject to transcriptional regulation by host factors as well as by viral factors including Tat. Tat is now known to regulate transcription by controlling the level of elongation but evidently has no effect on initiation of transcription. Recently, the mechanism of Tat transactivation of viral mRNA transcription has begun to be elucidated (see Figure 1.3 for the mechanism). All HIV-1 mRNAs contain a stem loop secondary structure at the 5' end known as TAR (transactivation response), which has a uridine-rich bulge of three residues (Dingwall *et al.*, 1989) (see Figure 1.2). Tat binds the uridine-rich bulge contacting uridine 23 at the base of the bulge (see Figure 1.2) (Dingwall *et al.*, 1989; Delling *et al.*, 1992; Churcher *et al.*, 1993). In the absence of Tat, the RNA polymerase II complex (RNAP II) is stalled by the secondary structure of TAR RNA, preventing transcription from continuing (see Figure 1.3) (Karn, 1999). The carboxy-terminal domain (CTD) of RNAP II must be phosphorylated by CDK7 (cyclin-dependent kinase 7), which is part of the transcription factor IIH (TFIIH) complex in order for transcription to be initiated and for the transcription complex to clear the promoter (Karn, 1999). Tat remains associated with the transcription complex following transactivation. Identification of a Tat-associated kinase (TAK) that is analogous to a component of the *Drosophila* positive transcription elongation factor b (pTEFb) began to clarify the role of Tat (Karn, 1999). Important to note is that pTEFb can phosphorylate the CTD of RNAP II only after TFIIH has phosphorylated the CTD (Marshal *et al.*, 1996). A molecule termed PITALRE was shown to be identical to the TAK, and was renamed CDK9 when it was later recognized as a member of the family of cyclin-dependent kinases. The significance of Tat-mediated

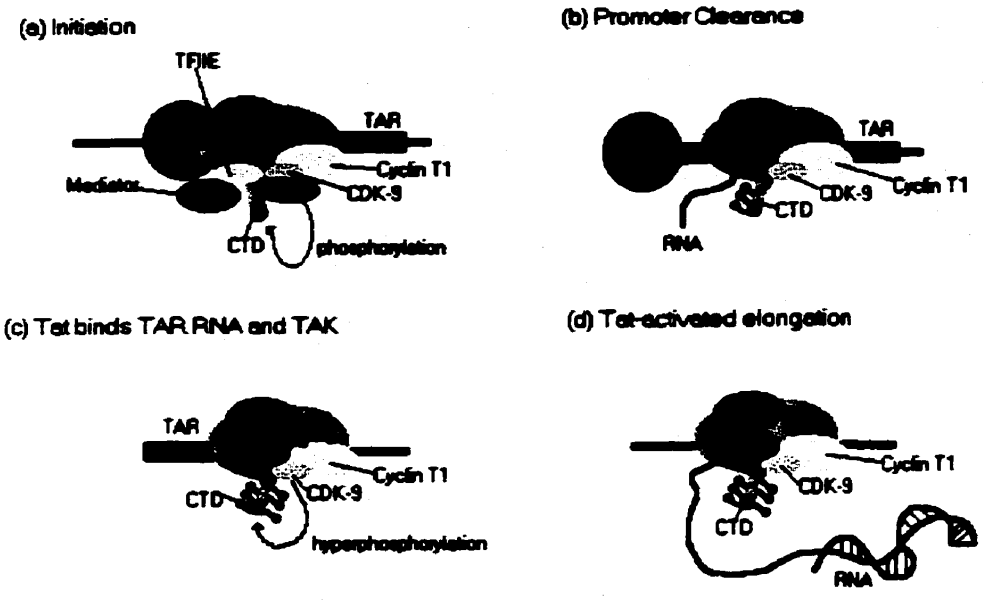

Figure 1.3 Transactivation of the HIV-1 LTR by Tat Protein. (Figure 6 from Karn, 1999, redrawn). a) TFIID recruits the RNAP II complex to the HIV LTR. After CDK7 of TFIIH phosphorylates the CTD of the RNAP II complex, the promoter is cleared and transcription begins. b) When TAR is transcribed the stem-loop structure forms blocking further transcription. c) Tat binds TAR allowing a complex to form between TAR, Tat, Cyclin T_1, and CDK9. The result is hyper-phosphorylation of the CTD of RNAP II. d) The processive RNAP II complex efficiently completes transcription of the viral mRNA.

hyper-phosphorylation of the CTD of RNAP II by CDK9 is that the polymerase complex, unstable upon encountering the TAR stem-loop, becomes a fully processive enzyme in the hyper-phosphorylated state (Karn, 1999). A processive enzyme is capable of clearing the TAR stem loop and completing a full transcript (Karn, 1999) (see Figure 1.3). It is unknown as yet whether the phosphorylation of the CTD by CDK9 is residue specific (Karn, 1999). Another component of the TAK complex, cyclin T_1, modulates the interaction of Tat and TAR (Wei *et al.*, 1998). Cyclin T_1 is not believed to bind the loop directly; rather it appears likely that it alters Tat structure to allow interaction with the TAR loop by Tat (Bieniasz *et al.*, 1998; Garber *et al.*, 1998). With the identification of cyclin T_1 as the loop factor (Wei *et al.*, 1998), the functional significance of some of the cysteine residues of Tat may now have been determined. In one model, the interaction between human cyclin T_1 and Tat involves chelation of Zn^{2+} ions by cysteine residues on both proteins (Garber *et al.*, 1998). Several groups have shown that the murine cyclin T_1, with a cysteine to tyrosine substitution in the sequence compared to the human protein, is unable to bind Zn^{2+}, and as such, does not interact in the same manner with Tat (Kwak *et al.*, 1999; Ivanov *et al.*, 1999; Garber *et al.*, 1998; Chen *et al.*, 1999; Bieniasz *et al.*, 1998; Fujinaga *et al.*, 1999). It is believed that cyclin T_1 and CDK9 are normally associated with the RNAP II complex (Isel and Karn, 1999; Ping and Rana, 1999; Bieniasz *et al.*, 1998, 1999).

1.3 Localisation of Function to Segments of Tat

The *tat* gene contains two exons. The first exon corresponds to the first 72 amino acids in the sequence, while the second exon encodes a variable number of amino acids

dependent upon the strain of virus (Jeang *et al.* 1999). The *tat* genes isolated from vast majority of infected individuals code for a 101 amino acid protein (Jeang *et al.* 1999). The HIV-1 BH10 strain contains a *tat* gene encoding an 86 amino acid protein, and has been used extensively by researchers as a standard source of Tat protein. However, it would now seem that this is a truncated form of the protein that is due to *in vitro* passage of the virus, and thus BH10 is considered to be a lab strain (Jeang *et al.* 1999). Most research into the function of Tat protein has focussed on the transactivation of viral transcription and the first 72 amino acids of the protein derived from the first exon, which are fully functional in transactivation assays (Jeang *et al.* 1999).

The Tat protein sequence can be divided into six segments (see Figure 1.4) (Gregoire and Loret, 1996; Jeang *et al.* 1999). The amino-terminal twenty amino acids compose the first segment, which is rich in proline and acidic residues. Residues 21 to 31 compose the cysteine-rich segment that contains five of the seven highly-conserved cysteine residues in the sequence. Residues 32 to 48 compose a core hydrophobic segment (Jeang *et al.* 1999). Perhaps the most important (with regards to transactivation), and certainly the most studied segment is number four, which includes residues 49 to 58, is arginine rich, and is responsible for binding TAR RNA (Gregoire and Loret, 1996, Jeang *et al.* 1999). Region five consists of residues 59 to 72 of the protein and is glutamine rich (Jeang *et al.* 1999). Segment six consists of the carboxyl terminus of the protein and is composed of residues 73 to 86 or 101 dependent upon the strain of the virus (Gregoire and Loret, 1996, Jeang *et al.* 1999).

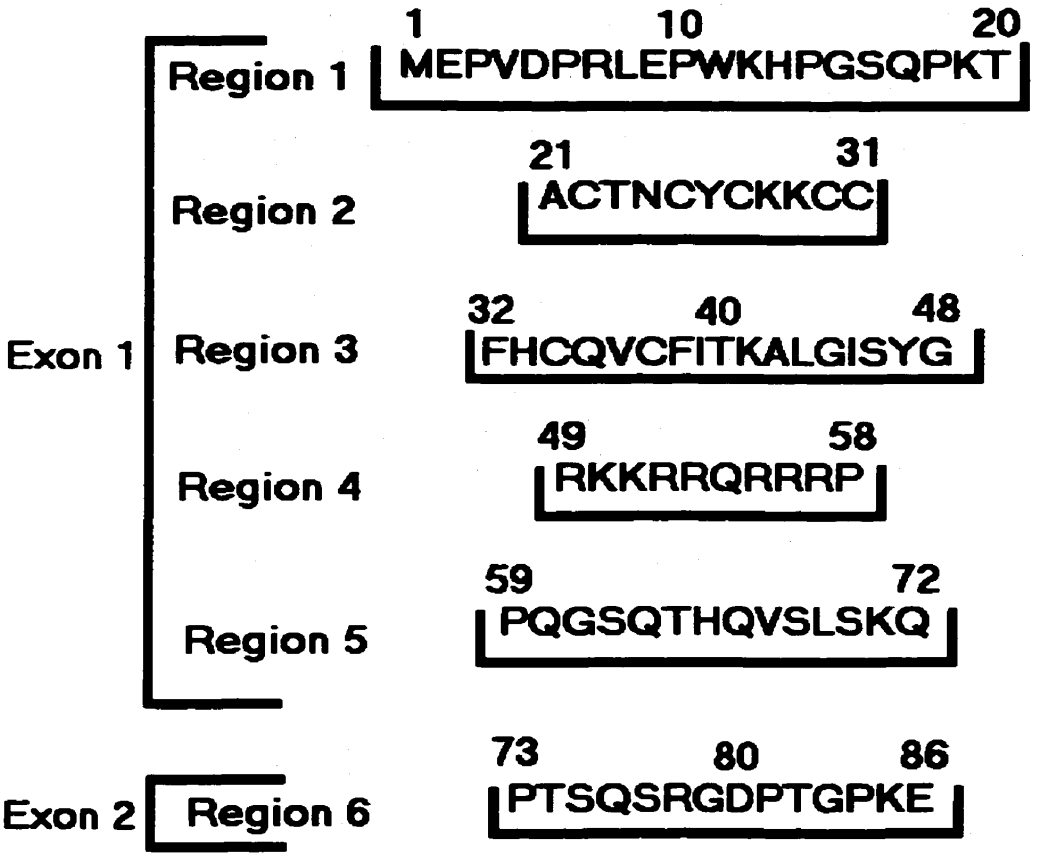

Figure 1.4 The Six Regions of HIV-1 Tat Protein.
1: proline rich amino terminus, 2: cysteine rich region, 3: highly conserved core domain, 4: basic region, 5: glutamine rich region, 6: carboxyl terminus. The diagram shows which exon the six regions are translated from. Exon one is composed of 216 nucleotides encoding 72 amino acids while exon 2 is composed of 42 nucleotides encoding a further 14 amino acids.

CD26 (dipeptidyl peptidase IV) is a transmembrane proline-specific serine protease responsible for the degradation and post-translational processing of bioactive peptides in human/mammalian cells (Mrestani-Klaus *et al.*, 1998). Inhibition of CD26 function by Tat protein has been shown to suppress mitogen-induced proliferation of T-cells, presumably contributing to immune suppression (Wrenger *et al.*, 1996; Mrestani-Klaus *et al.*, 1998). Substitution of aspartic acid 5 with isoleucine, or proline 6 with leucine, was shown to reduce this inhibition and in addition affect the solution conformation of the peptide significantly, suggesting that the amino terminus of Tat is responsible for CD26 inhibition (Mrestani-Klaus *et al.*, 1998). Further investigation has shown that a hypersialylated isoform of CD26 appears during HIV infection (Smith *et al.*, 1998). Investigation of this phenomenon with Tat, and cationic peptides, appears to show that the binding of positively-charged molecules to the hyper-sialylated (highly-negatively charged) form of the enzyme produced greater inhibition than observed for the native CD26 (Smith *et al.*, 1998).

The amino terminus of the Tat protein has also been implicated as necessary for maximal efficiency in reverse transcription (Ulich *et al.*, 1999). A variety of point mutations in the 20 amino terminal residues of the protein severely decreased the efficiency of reverse transcription. Conversely it was observed that mutations of amino acid residues absolutely required for transactivation such as cysteine 27 and lysine 41, as well as replacement of the basic region (lysine/arginine residues 50 to 57) with glycine had no significant affect on reverse transcription. This suggests that binding of TAR RNA is not a factor in reverse transcription and that the effect of Tat on reverse

transcription is not due to increases in gene expression caused by the transactivation function of Tat (Ulich *et al.*, 1999).

The cysteine-rich region is absolutely necessary for transactivation (Garcia *et al.*, 1988; Kuppuswamy *et al.*, 1989; Sadaie *et al.*, 1990; 1996; Jeang *et al.*, 1999). Mutation of 6 of the 7 conserved cysteine residues alone will seriously affect the level of or abolish transactivation by the protein (Garcia *et al.*, 1988; Sadaie *et al.*, 1990; Jeang *et al.*, 1999). Potential zinc binding motifs have long been suggested for this region, and there have been suggestions of a role for the cysteine residues in metal-linked dimerization (Frankel *et al.*, 1988; Garcia *et al.*, 1988; Sadaie *et al.*, 1990; Kuppuswamy *et al.*, 1989) but there is little structural evidence for this and few metal binding studies have been conducted.

The hydrophobic domain consisting of residues 32 to 48 (Kuppuswamy *et al.*, 1989; Bayer *et al.*, 1995; Gregoire and Loret, 1996) is necessary for the transactivation function in the protein, however, it would seem that only the lysine 41 residue is of critical importance in the highly conserved KALGI (lysine, alanine, leucine, glycine, isoleucine) segment (Aboul-ela *et al.*, 1995; Gregoire and Loret, 1996; Bayer *et al.*, 1995). It has been suggested that this residue may be involved in contacts with TAR RNA (Aboul-ela *et al.*, 1995; Gregoire and Loret, 1996; Bayer *et al.*, 1995). However, significant decreases in transactivation have been documented for mutants of glycine 44 (Kuppuswamy *et al.*, 1989). NMR structural studies coupled with molecular modelling have concluded that the conformational flexibility of this residue is pivotal in the turn-like nature of the core domain (Bayer *et al.*, 1995).

The basic region is responsible for binding the 3-nucleotide bulge in the stem of TAR RNA (see Figure 1.4) (Aboul-ela *et al.*, 1995; Gregoire and Loret, 1996; Puglisi *et al.*, 1992; Puglisi *et al.*, 1993; Calnan *et al.*, 1991a and 1991b). A single arginine residue has been implicated in binding the bulge of the TAR RNA stem, making contacts with a guanine residue and a uridine residue (Aboul-ela *et al.*, 1995; Puglisi *et al.*, 1992 and 1993; Gregoire and Loret, 1996; Calnan *et al.*, 1991b). The optimal location of the arginine residue appears to be at position 52 or 53 (Calnan *et al.*, 1991a and 1991b). Structural studies using circular dichroism spectropolarimetry (Gregoire and Loret, 1996), as well as NMR spectroscopy (Puglisi *et al.*, 1992 and 1993; Bayer *et al.*, 1995) and molecular modelling calculations (Bayer *et al.*, 1995) indicate the likelihood that a structured core domain (region III) as well as a structured glutamine-rich domain (region V) are required for the arginine-rich domain to exist in an extended structure that allows TAR binding. Binding of Tat to TAR RNA appears to induce a change in local conformation of the bulge region, allowing Tat to make further contacts with the two base pairs on either side of the bulge (Aboul-ela *et al.*, 1995; Puglisi *et al.*, 1993; Gregoire and Loret, 1996; Calnan *et al.*, 1991b).

The functional significance of region six, the carboxyl terminus of the protein has until now received much less attention than the first four regions. This segment is composed of residues 73 to 86 or 101, depending on the strain of virus. One group has reported that the carboxyl terminus of Tat protein, while often considered unnecessary for transactivation, does contribute to replication in a significant manner (Verhoef *et al.*, 1998). The manner in which the carboxyl terminus contributes to viral replicative fitness

is as yet unknown however. The carboxyl terminus of Tat has been implicated as necessary for depression of manganese superoxide dismutase in cells (Westendorp *et al.*, 1995) as well as for T-cell activation (Ott *et al.*, 1997). Through the use of a yeast two-hybrid system, one group was able to show that the carboxyl terminus of Tat is responsible for specific binding to human elongation factor-1 delta (EF-1 δ) which results in suppression of translation from host cell mRNA, but not from viral transcripts (Xiao *et al.*, 1998). Perhaps the most obvious question about the function of the carboxyl terminus surrounds the arginine-glycine-aspartate (RGD) sequence. The RGD sequence is found in proteins that bind the integrin family of proteins responsible for cell surface adhesion. Several studies have shown that, although there is significant cellular uptake and internalization of Tat protein in a truncated form lacking the RGD sequence (Tat_{1-72}), the presence of the carboxyl terminus containing the RGD sequence increases internalization as much as tenfold (Ensoli *et al.*, 1993; Ma and Nath, 1997).

1.4 The Structures of Tat and TAR

A high-resolution structure of the HIV-1 Tat protein has not yet been determined. Nevertheless, there have been several attempts to obtain structural information for various truncated segments of the protein, as well as chimeric fusions of functional regions from HIV-1 Tat protein and other related Lentiviral transactivators (Echetebu and Rice, 1993; Mujeeb *et al.*, 1994; Klostermeier *et al.*, 1997; Metzger *et al.*, 1997). There is a high resolution NMR solution structure of TAR RNA alone, as well as TAR RNA complexed with the minimal recognition region of the Tat protein (residues 37-72)

(Aboul-ela *et al.*, 1995 and 1996) as well as a 1.3 Å X-ray diffraction structure of TAR with three bound Ca^{2+} ions (Ippolito and Steitz, 1998).

A limited investigation of the structure of an 86 amino acid HIV-1 Tat protein using NMR restraints and molecular dynamics calculations has been published (see Figure 1.5) (Bayer *et al.*, 1995). The authors report that long-range nuclear Overhauser effects (NOEs) were observed to originate only in the glutamine rich and core regions of the protein (Bayer *et al.*, 1995). Equine infectious anaemia virus (EIAV) Tat protein for which some NMR structural information has been published (Sticht *et al.*, 1993; Willbold *et al.*, 1993), and HIV-1 Tat protein share a highly conserved core region with one difference, the HIV-1 Tat protein core region has a flexible hinge between two rigid parts of the core (Bayer *et al.*, 1995). Comparison of the structural assignments of the rigid parts of the core regions for both proteins resulted in a nearly ideal superposition of the backbone structure. From the limited data available the authors concluded that the glutamine rich and core regions form rigid structures, between which the amino terminus of the protein is sandwiched, with the carboxyl terminus of the protein contacting the glutamine rich region (see Figure1.5). The RGD loop, present in the carboxyl terminus, is solvent-exposed and relatively rigid in comparison to most RGD loop-containing proteins (Bayer *et al.*, 1995). The cysteine-rich and basic regions were reported to be flexible and unstructured (Bayer *et al.*, 1995).

Several bases in the bulge/stem loop structure of TAR have been shown to be absolutely necessary for Tat binding. The guanine 26/cytidine 39 and adenine 27/uridine

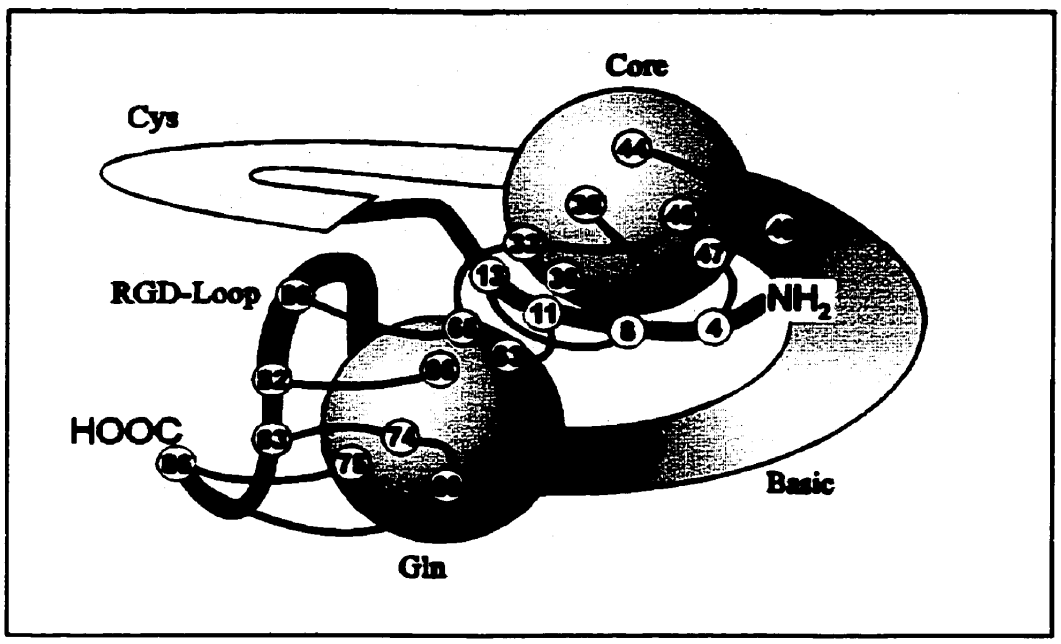

Figure 1.5 Structural Aspects of the Six Regions of HIV-1 Tat Protein. (Figure 3 from Bayer *et al.* 1995, with permission) This is a cartoon of Tat protein structure using data from NMR solution studies and molecular dynamics calculations. Note that the cartoon indicates rigid structures in the core and glutamine rich domains, extended structures for the cysteine rich and basic domains and numerous contacts between the amino terminus and the structured domains.

38 base pairs above the bulge are essential for Tat binding (see Figure 1.6) (Delling *et al.*, 1992; Churcher *et al.*, 1993). In addition, experiments have shown that two base pairs below the bulge, adenine 22/uridine 40 and guanine 21/cytidine 41 contribute to efficient binding (Aboul-ela *et al.*, 1995) while several phosphates situated on both strands below the bulge, P22, P23, and P40 are involved in pivotal contacts with Tat (Calnan *et al.*, 1991b; Tao and Frankel, 1992; Churcher *et al.*, 1993; Hamy *et al.*, 1993; Pritchard *et al.*, 1994). Finally, there are important hydrogen bonding contacts between Tat and N^7 of guanine 26 and adenine 27 as well as N^3 and O^4 of uridine 23 (Churcher *et al.*, 1993).

NMR spectroscopy has shown that ligand-free TAR RNA has an unusually wide major groove induced by the uridine-rich (U-rich) bulge region in comparison to A-form RNA (Aboul-ela *et al.*, 1996). The NMR data show that the stems above and below the U-rich bulge are composed of A-form helical RNA (Aboul-ela *et al.*, 1996). However, the adenine 22/uridine 40 base pair immediately below the U-rich bulge appears to be relatively unstable. Across from the U-rich bulge, sequential NOE interactions indicate that uridine 40 and cytidine 39 resemble A-form RNA, although an unusually strong NOE between uridine 40 H6 and cytidine 39 H1′ indicates a distortion in the structure. The NMR data also show that uridine 40 stacks normally on cytidine 41 while the interaction with cytidine 39 is weakened (Aboul-ela *et al.*, 1996). These distortions in the twist of the strand opposite the U-rich bulge are posited as the reason for a weakening of the adenine22/uridine 40 base pairing interaction, which is unobservable on the NMR timescale (Aboul-ela *et al.*, 1995 and 1996). The NOE connectivities show that typical A-form RNA base stacking is continuous through the first two residues (uridine 23 and

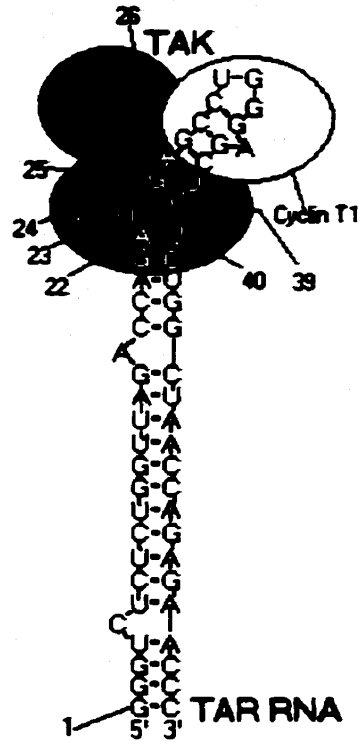

Figure 1.6 Cartoon of the binding of Tat, Cyclin T_1, and CDK9 to TAR RNA. (Figure 4 from Karn, 1999, redrawn) Cyclin T_1 is thought to bind Tat in a metal dependent manner through cysteine residues on each protein. Upon binding Tat, Tat structure is altered to allow Tat to interact with the loop. Any direct contact between Cyclin T_1 and the loop has yet to be elucidated. The residues important for Tat binding are shown in white. Phosphate groups important for affinity and specificity of Tat binding are shown with black dots.

cytidine 24) of the U-rich bulge (Aboul-ela *et al.*, 1996). NOE connectivities for the uridine 25 residue are limited in number and weak (Aboul-ela *et al.*, 1996). NOE connectivities between guanine 26 and cytidine 24 flanking uridine 25 show that uridine 25 is not base stacked between these two residues, but is looped out of the helix (Aboul-ela *et al.*, 1996). The absence of NOE cross-peaks between the aromatic resonances of guanine 26 and the sugar or aromatic resonances of cytidine 24 also indicate that uridine 25 is looped out of the helix transiently on the NMR timescale (Aboul-ela *et al.*, 1996). The distortion in TAR RNA caused by the bulge residues produces a significant bend in the RNA helical structure between the stem regions above and below the bulge (see Figure 1.7) (Aboul-ela *et al.*, 1995; Zacharias and Hagerman, 1995).

The NMR data show that in the presence of ligand (a Tat peptide encompassing the core and basic domains) TAR undergoes a conformational change resulting in an energetically stabilised Tat-TAR structure (see Figure 1.7) (Aboul-ela *et al.*, 1995). The overall effect of Tat binding to TAR is to make the major groove of TAR more accessible, while the phosphate groups P22, P23 and P40 are positioned on one face of the molecule (Aboul-ela *et al.*, 1995). The NMR structure shows that it is the absence of a phosphate backbone in the bulge region that allows Tat to bind TAR, while the widened major groove created by the bulge structure in the unbound form serves to increase accessibility further (Aboul-ela *et al.*, 1995 and 1996). The precise positioning of the phosphates upon binding allows interaction between the phosphates and amino acid residues of the Tat protein (Aboul-ela *et al.*, 1995). It is the interaction with these phosphates that provides Tat with the ability to bind TAR with both high affinity and

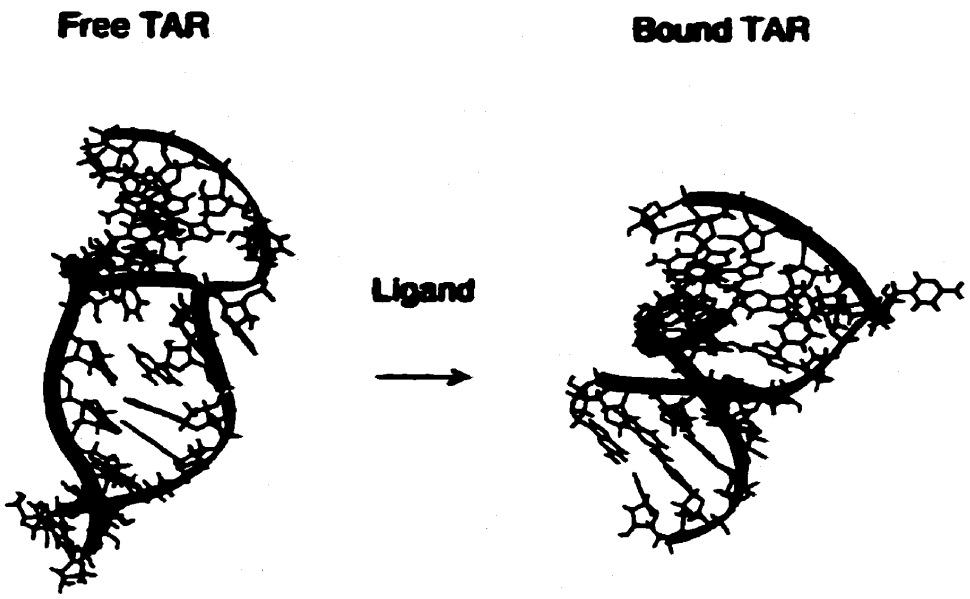

Figure 1.7 Structural Changes in HIV-1 TAR RNA Induced by Tat Binding. (Figure 11 from Aboul-ela *et al.* 1995, with permission). Diagram showing the change in structure of a TAR RNA segment upon binding of a peptide encompassing the basic region of Tat as the bulge region folds into a locally different conformation.

specificity (Aboul-ela *et al.*, 1995). Upon binding of Tat, base stacking between adenine 22 and uridine 23 as well as between uridine 23 and cytidine 24 is disrupted (Aboul-ela *et al.*, 1995). This allows all three bulge nucleotides to loop out of the structure in the bound form, with the result that the bend in the helical structure of unbound TAR is virtually eliminated (Aboul-ela *et al.*, 1995; Zacharias and Hagerman, 1995). When bound TAR is superimposed over an idealised co-ordinate set for A-form RNA produced from the lower stem structure, the bulge residues are observed to superimpose on residues displaced by three base pairs. The result is maintenance of helical continuity across the bulge region (Aboul-ela *et al.*, 1995). The disruption of the base stacking in the bulge region allows the ring of uridine 23 to be drawn near guanine 26, where a single arginine side chain interacts with both residues. Both the guanidinium group and ϵ-amino of the arginine side chain are within hydrogen bonding distance of N7 of guanine 26 (Aboul-ela *et al.*, 1995). The aliphatic portion of the arginine side chain is stacked underneath the ring of uridine 23 (Aboul-ela *et al.*, 1995). A further result of the disruption in base stacking in the bulge region is the formation of a stable Watson-Crick base pair between adenine 22 and uridine 40, witnessed by the observation of a NOE between the imino ^{1}H of uridine 40 and H2 of adenine 22 that is absent in the free form of TAR (Aboul-ela *et al.*, 1995). In the bound form, adenine 22 points toward N7 of guanine 26 with the result that base stacking between these residues observed in the free form is disrupted. The authors report that a dyad axis running through the uridine 40/adenine 22 base pair is rotated by 45° to 60° with regards to the guanine 26/cytidine 39 base pair. They posit that this may allow stacking interactions with amino acid residues of the highly conserved core region of Tat (Aboul-ela *et al.*, 1995).

From the structural information available, a model of Tat binding to TAR has been suggested. An arginine residue (probably in position 52 or 53) is able to enter the widened major groove of the bulge region of TAR RNA where there is no phosphate backbone to prevent access (Aboul-ela *et al.*, 1995; Huq and Rana, 1997). The interaction of this arginine with uridine 23 and guanine 26 causes a conformational change in TAR RNA whereby the RNA folds around the Tat protein, positioning phosphates P22, P23, and P40 which are important for binding affinity and specificity, on one surface of the molecule where they can presumably interact favourably with charged amino acid residues such as lysine and arginine (Aboul-ela *et al.*, 1995; Huq and Rana, 1997). The aliphatic portion of the arginine side chain stacks underneath the uridine 23 base (Aboul-ela *et al.*, 1995). The conformational change also allows adenine 22 and uridine 40 to form a stable Watson-Crick base pair that is not present in the unbound TAR structure (Aboul-ela *et al.*, 1995 and 1996). This new base pair is oriented so that adenine 22 now points towards N7 of guanine 26, with the result that the adenine 22-uridine 40 base pair is placed at an angle with respect to the guanine 26:cytidine 39 base pair (Aboul-ela *et al.*, 1995 and 1996). This disrupts base stacking interactions between adenine 22 and guanine 26 present in the unbound TAR RNA and may allow stacking interactions with amino acid residues of the core region of Tat (Aboul-ela *et al.*, 1995 and 1996). The conformational change disrupts base stacking interactions between the first two bases of the U-rich bulge, as well as between adenine 22 and uridine 23 (Aboul-ela *et al.*, 1995). This allows all three bulge residues to loop out of the structure unlike the situation in unbound TAR where only uridine 25 loops out of the bulge region (Aboul-ela *et al.*, 1995 and 1996). This looping out of all three bulge residues almost completely

eliminates a significant bend between the helical stem regions above and below the bulge that is present in the unbound form of Tat (Aboul-ela *et al.*, 1995; Zacharias and Hagerman, 1995). The study by Ma and Nath (1997) positing that the RGD loop in the carboxyl terminus increases cellular uptake of Tat protein due to a tertiary interaction in the protein fits well with the proposed Tat structure (Bayer *et al.*, 1995) where the carboxyl terminus contacts the glutamine rich region. The absence of the carboxyl terminus could alter or disrupt the structural features necessary for internalisation. The loss of transactivation function for mutant Tat proteins missing amino acids aspartate 2 to proline 6 is also supported well by the suggested model. The sandwiching of the amino terminus of the protein between the rigid core and glutamine rich regions suggests the absence of the amino terminus could disrupt the structure of Tat and affect transcription (Kuppuswamy *et al.*, 1989, Bayer *et al.*, 1995).

The specific binding of RNA by a single arginine residue places Tat (along with Rev) in the class of Arginine-Rich-Motif (ARM) proteins (Burd and Dreyfuss, 1994). The ARM class of proteins consists of ribosomal, viral and bacteriophage proteins that contain arginine rich sequences of 10 to 20 amino acids (Burd and Dreyfuss, 1994). Examples include the λ, P22, and ϕ21 bacteriophage antiterminator N proteins which bind stem loops, HIV-1 Rev protein which binds internal loops, and HIV-1 Tat which binds bulges (Burd and Dreyfuss, 1994). Whereas peptides composed of the arginine rich domain bind with high affinity, the neighbouring amino acid residues and domains of this class of protein tend to increase affinity and specificity through their own contacts with the RNA (Burd and Dreyfuss, 1994). It has been proposed that the positive charge

of an ARM domain allows the protein to probe the negatively charged backbone for a high affinity site (Burd and Dreyfuss, 1994). Presumably, the arginine is then best suited to hydrogen bond with the RNA structure due to its greater number of atoms with potential for hydrogen bonding in comparison with lysine (Burd and Dreyfuss, 1994).

1.5 Thesis Objectives

The objective of the present research was to develop an expression and purification system for HIV-1 BH10 Tat_{1-86} protein for structural analysis. Large amounts (mg quantities) of pure protein are required for high-resolution structural analysis of proteins by X-ray diffraction and NMR spectroscopy. High-resolution structures of Tat alone and in complex with TAR and the TAK complex are necessary for a detailed atomic level explanation of the role of Tat in the transactivation of transcription. Because Tat is required for viral gene expression and the mechanism of transactivation appears to be unique to the Lentiviruses it is an attractive therapeutic target (Rhim *et al.*, 1993). With a high-resolution structure assigned, structure-based drug design would become possible. Finally, as indicated above, Tat has been implicated in a variety of cellular and viral processes apart from transactivation. Large amounts of purified Tat are needed for analysis to begin to explain the structural and functional aspects of Tat's interactions with those viral and cellular targets. Reports in the literature indicate that production of large amounts of Tat protein could be hampered by instability of the tat gene when expressed in *E. coli* as well as toxic effects of Tat in the cells (Ciccarelli *et al.*, 1990; McKenna *et al.*, 1994).

Chapter Two: Materials and Methods

2.0 Equipment and Basic Methods

2.0.0 Equipment

Absorbance measurements were made using a Cary 1 ultra-violet/visible spectrophotometer, with a 1 cm path length cuvette. Centrifugation of 10 mL sterile tubes used for preparation of competent cells was performed using a Dynac centrifuge. Centrifugation of 1.5 mL Eppendorf tubes was performed using an International Equipment Company Micro-MB centrifuge. Dr. H. Duckworth, Department of Chemistry, The University of Manitoba generously permitted the use of a Sorvall RC-5B refrigerated superspeed centrifuge for centrifugation of 250 mL bottles and 50 mL centrifuge tubes used for protein purification.

2.0.1 Calculation of Relative Centrifugal Force

Calculation of relative centrifugal force (RCF) was made with the following formula (Tedeschi, 1993): $RCF = 11.17\ (r) \times (n \div 1000)^2$

where r = the radius in centimetres from the centreline of the rotor to the point in the tube where the RCF value is determined, and n = the rotor speed in rpm.

2.0.2 Preparation of DNase-free RNase A

A 10 mg/mL deoxyribonuclease-free (DNase-free) ribonuclease A (RNase A) (Calbiochem) stock solution was prepared according to Sambrook *et al.* (1989). RNase A was dissolved in 10 mM Tris-HCl (pH 7.5), 15 mM NaCl to a concentration of 10 mg/mL. The solution was heated to 100 °C for 15 minutes in a water bath. After cooling

to room temperature the DNase-free RNase A was divided into aliquots and stored at - 20 °C.

2.0.3 Ethanol Precipitation of DNA

Concentration of plasmid DNA was accomplished with an ethanol precipitation following Sambrook *et al.* (1989). A one tenth volume of 3 M sodium acetate pH 5.2 was added to the plasmid DNA in TE buffer in a 1.5 mL Eppendorf tube. After mixing, between 2 and 3 volumes of ice cold 95 % ethanol was added. The tube was centrifuged for 15 minutes at 4 °C. The supernatant was carefully removed with a micropipette. The precipitated DNA was washed with 500 µL of 70 % ethanol, and centrifuged for 5 minutes. The ethanol was removed and the DNA pellet allowed to air dry at room temperature. After drying, the pellet was resuspended in an appropriate volume of TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0).

2.0.4 Agarose Gel Electrophoresis of DNA

Samples for electrophoresis were prepared by adding one-fifth volume of bromophenol blue gel loading solution (0.25% bromophenol blue, 40% sucrose in water). A 0.7 % electrophoresis grade agarose (Fisher) solution was prepared in Tris Acetate EDTA (TAE) buffer (40 mM Tris free base, pH to 7.7 with acetic acid, 1mM EDTA). The cooled gel was immersed in an electrophoresis tank in TAE buffer containing 20 µg/mL of ethidium bromide. Ethidium bromide intercalates with DNA and glows orange when illuminated with UV light. The samples were loaded and electrophoresed at 5 to 100 Volts for varying periods (generally 50 V for several hours or 5 V overnight). The

DNA bands were then visualized under UV light. Estimation of DNA fragment length was accomplished by electrophoresis of 1 Kb Plus DNA Ladder purchased from Life Technologies. The ladder consisted of fragments of double stranded linear DNA of 100, 200, 300, 400, 500, 650, 850, 1000, 1650, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10 000, 11 000, and 12 000 base pairs (bp). Measurements of the mobility of each band in cm were plotted against the logarithm of the DNA base pair length using the program Kaleidagraph running on a MacPlus computer. Estimations of the lengths of unknown DNA molecules were obtained by interpolation assuming a linear relationship between mobility and the logarithm of the DNA length (Helling *et al.*, 1974).

2.0.5 Polyacrylamide Gel Electrophoresis of DNA

Polyacrylamide gel electrophoresis (PAGE) was used to analyse *tat* DNA due to greater resolving power compared to agarose gels. Eight per-cent polyacrylamide gels were prepared according to Sambrook *et al.* (1989). DNA samples contained 20% by volume of xylene cyanol gel loading solution (0.25 % w/v xylene cyanol FF, 40 % w/v sucrose, in ddH_2O). The gels were electrophoresed in TBE buffer at less than 8 Volts per cm of gel. Estimation of DNA fragment length was accomplished by electrophoresis of 1 Kb Plus DNA Ladder (see above).

2.0.6 Production of Recombinant Plasmids-Overview

Figures 2.1 and 2.2 detail the production of pTatKL1 and pUCTat respectively. pTatKL1 and pUCTat served as a source of *tat* DNA for the subsequent production of recombinant expression plasmids. Figure 2.3 outlines the production of pTatKL19 from

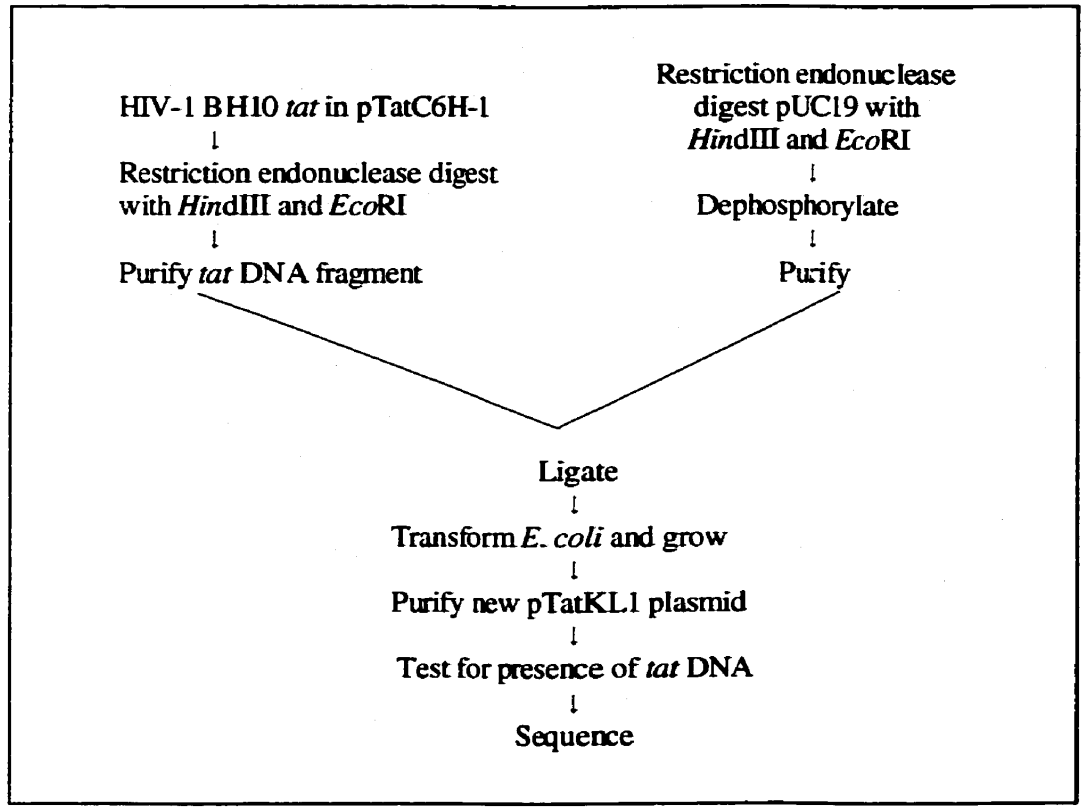

Figure 2.1 Flow Chart for the Production of the pTatKL1 Cloning Plasmid. The pTatC6H-1 plasmid was produced by Purvis *et al.*, (1995). The *tat* DNA contained in the plasmid was produced by Adams *et al.*, (1988).

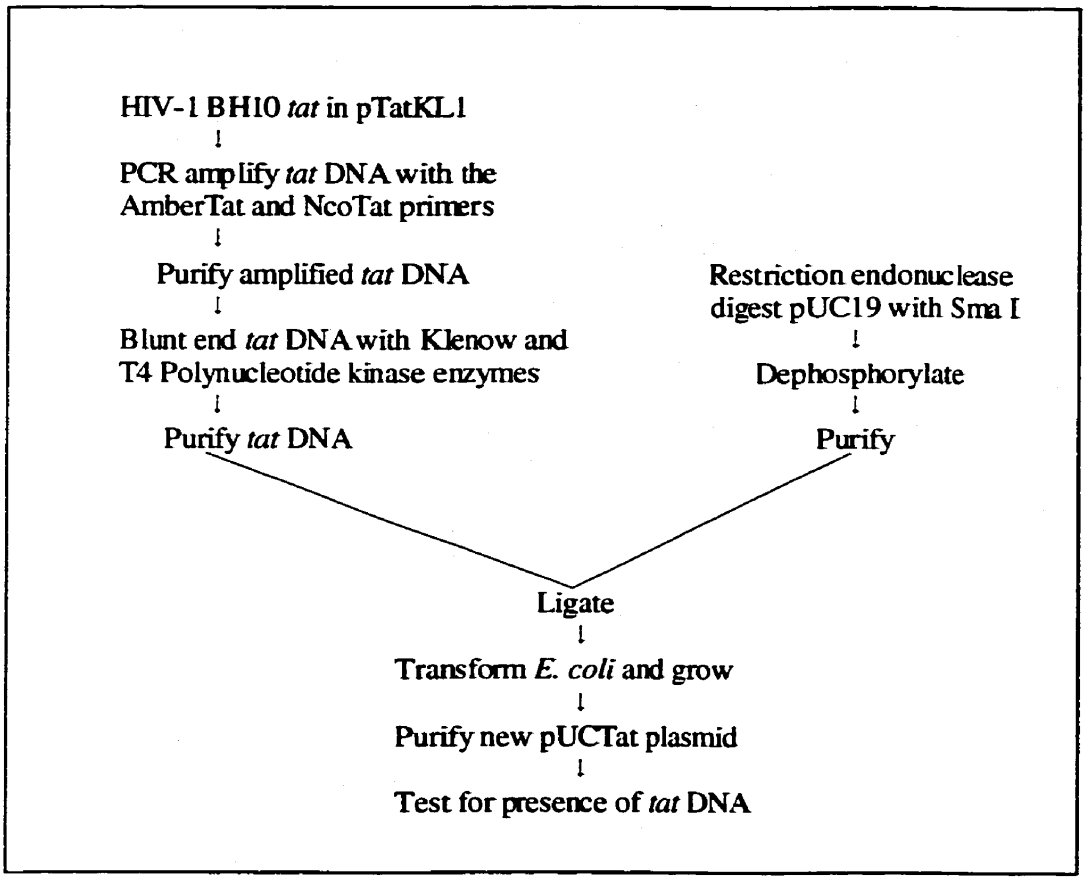

Figure 2.2 Flow Chart for the Production of the pUCTat Cloning Plasmid.

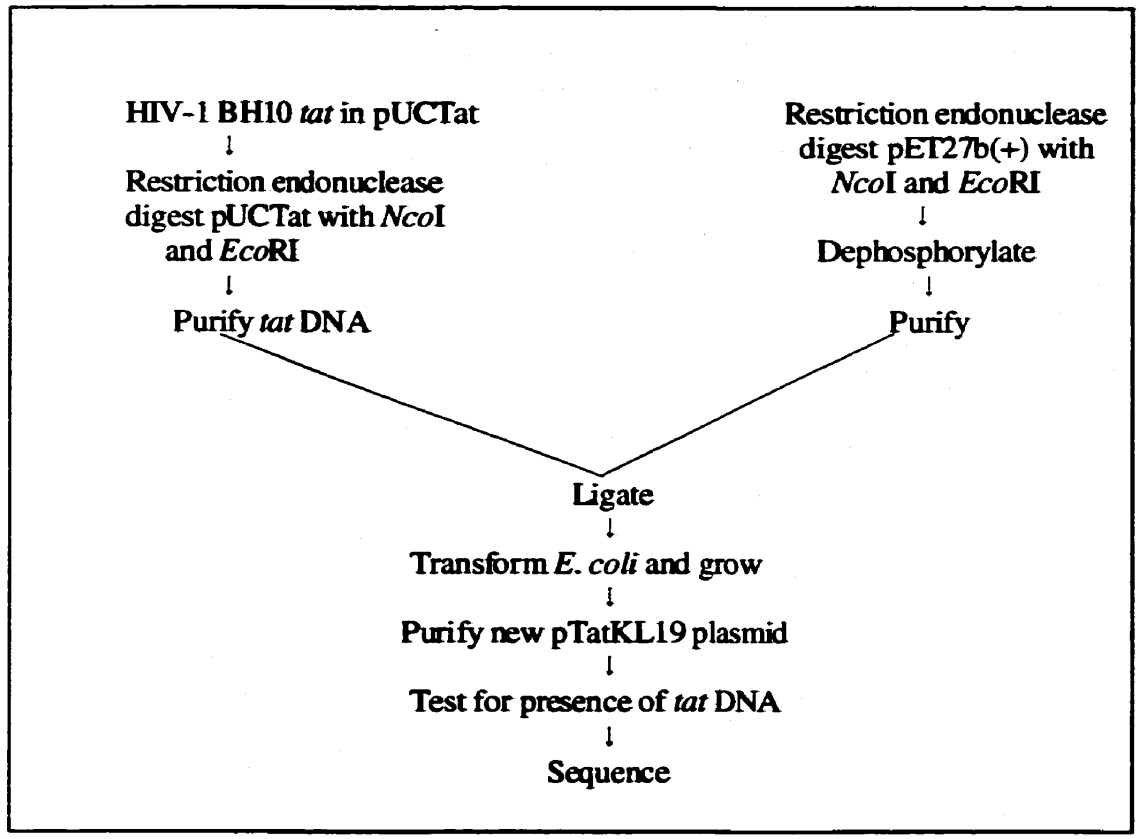

Figure 2.3 Flow Chart for the Production of the pTatKL19 Expression Plasmid.

the pET27b(+) plasmid. This plasmid was designed to express a protein with a pectate lyase B (PelB) leader sequence allowing secretion of recombinant Tat protein into the periplasmic space. Figures 2.4 and 2.5 detail the production of the pTatKL16 and pTatKL26 plasmids using the pET28b(+) vector. These two plasmids express recombinant Tat protein with an amino terminal six-histidine tag, allowing purification of the Tat protein using Ni^{2+} metal-chelate affinity chromatography. The pTatKL16 plasmid contains *tat* DNA with a sequence partially optimised for expression in *E. coli* (Adams *et al.*, 1988), whereas pTatKL26 contains *tat* DNA with a sequence more substantially optimised for expression in *E. coli*. The specifics of the production of these plasmids will be discussed in section 2.1. In general, *tat* DNA was PCR amplified from a suitable template using primers incorporating specific restriction endonuclease sites. The DNA was purified, restriction endonuclease digested and purified again. The chosen plasmid vector was restriction endonuclease digested with the same enzymes used to digest the *tat* DNA, and then dephosphorylated to prevent the digested plasmid from ligating together without a *tat* DNA insert. The plasmid was purified after dephosphorylation and then ligated with the purified *tat* DNA. The recombinant plasmid was then transformed into *E. coli*, the *E. coli* were cultured, and the plasmid purified. The plasmid was then tested to confirm that it contained *tat* DNA prior to sequencing.

2.0.7 Restriction Endonuclease Digests

Digestion of plasmid DNA with restriction endonucleases followed protocols found in Sambrook *et al.* (1989) or those supplied with the enzyme. The recommendations of the suppliers were followed to ensure the enzymes were incubated in

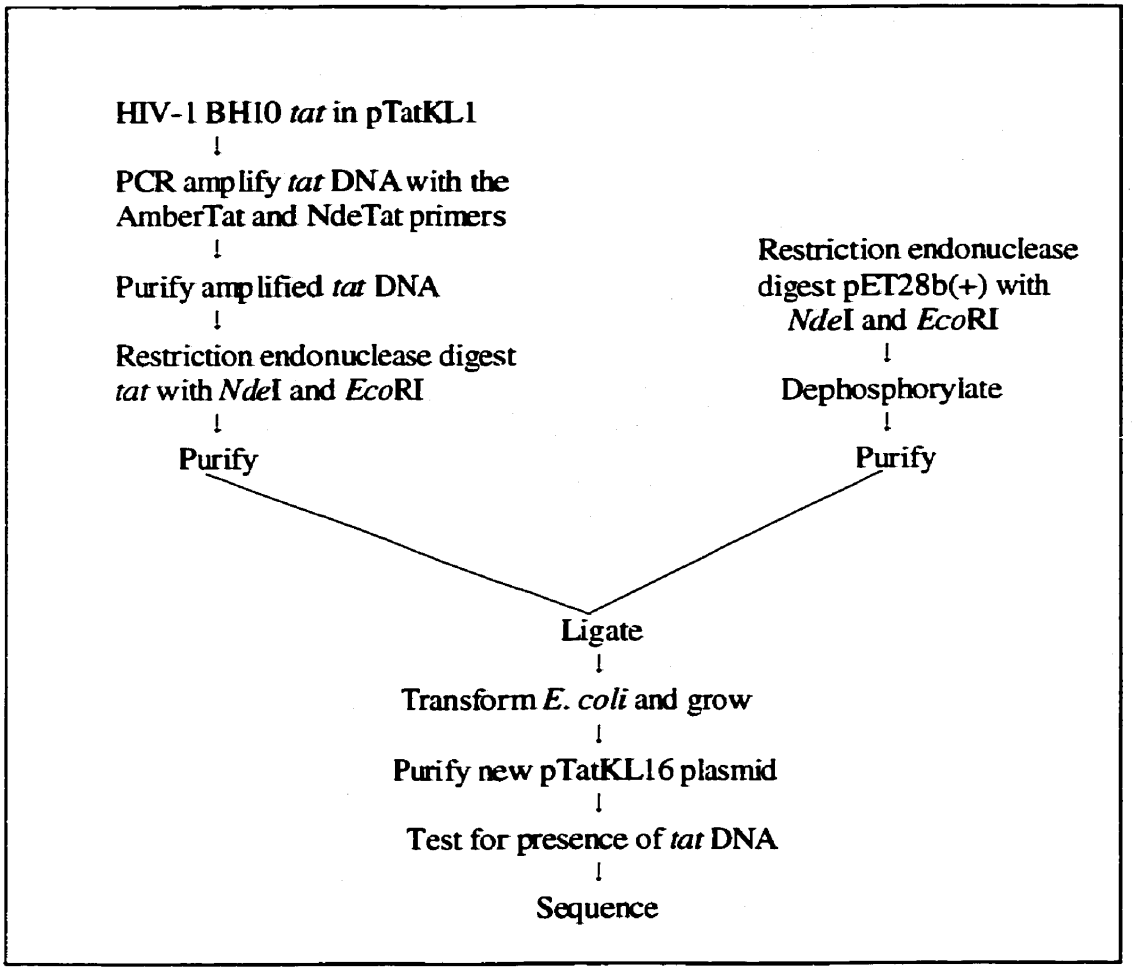

Figure 2.4 Flow Chart for the Production of the pTatKL16 Expression Plasmid.

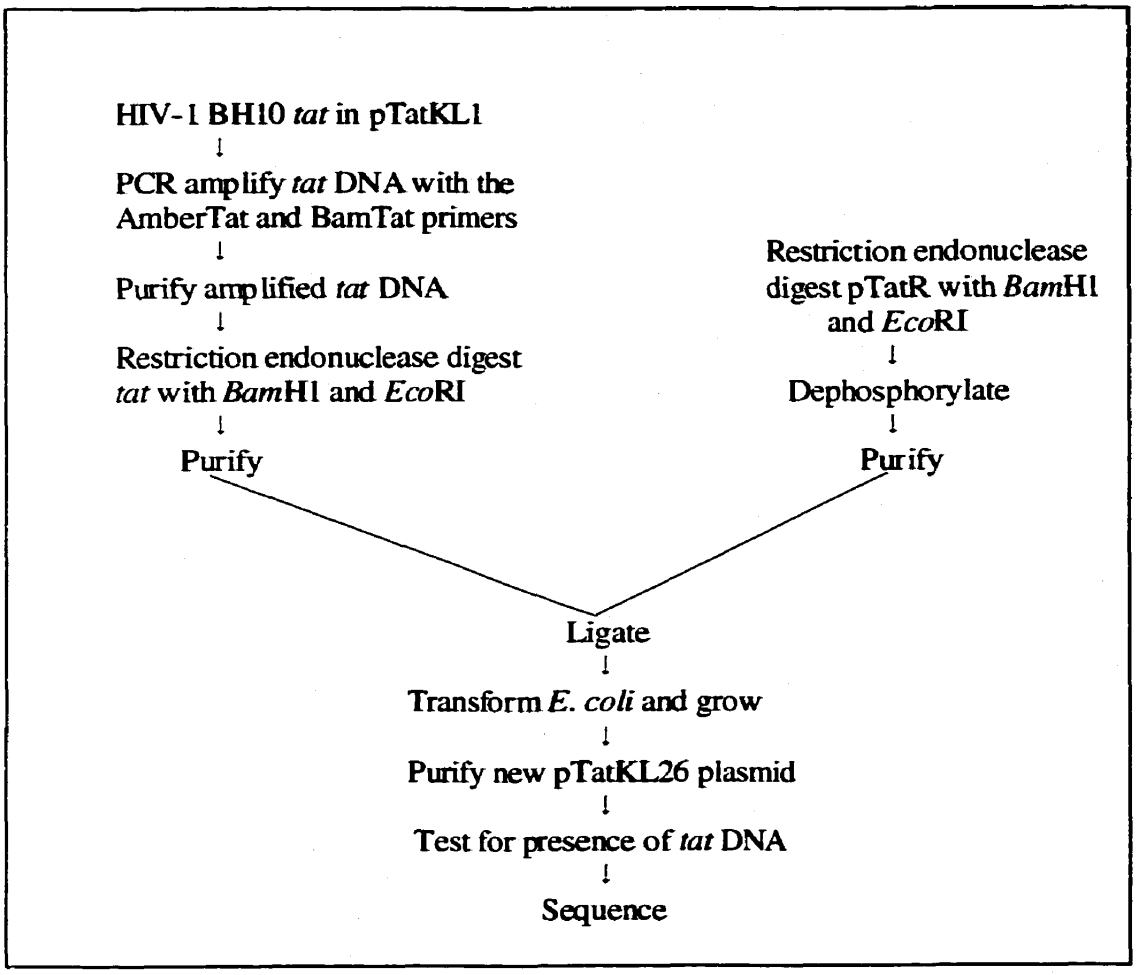

Figure 2.5 Flow Chart for the Production of the pTatKL26 Expression Plasmid.

a buffer in which their activity was acceptable (see Table 2.1 for a list of enzymes used). In addition, some enzymes are adversely affected when glycerol comprises more than a total of 5% by volume of the reaction mixture. This was controlled by dilution of reaction mixtures. Solutions contained 0.1 mg/mL acetylated BSA (Sigma) (to prevent potential denaturation of the enzyme in dilute solutions), 2 to 10 units of enzyme (1 μL), the appropriate reaction buffer, and ddH_2O to a final volume of 20 μL. The mixture was incubated in a recirculating water bath at the suppliers' recommended temperature for periods based on suppliers' recommendations.

2.0.8 Dephosphorylation of DNA

Plasmid vector was prepared for ligation by restriction endonuclease digestion. When the restriction endonuclease digestion was complete, calf intestinal alkaline phosphatase (CIP) was added, following protocols from Sambrook *et al.* (1989) and the Novagen pET System Manual. CIP removes phosphate groups from the 5' termini of DNA, so the amount of CIP used was based on the number of pmol ends after restriction endonuclease digestion. CIP incubation requires 0.05 units (U) of enzyme per pmol DNA ends. DNA was incubated with CIP for 30 minutes at 37°C. This prevents restriction endonuclease digested plasmid from ligating together without first incorporating the *tat* DNA insert, which was not dephosphorylated. The circular plasmid that results from the ligation of insert and dephosphorylated vector has two single-stranded nicks, which are repaired by the T4 DNA ligase enzyme or by bacterial enzymes after transformation.

Table 2.1 Restriction Endonucleases Used

Enzyme	Buffer	Temp.	Supplier
*Aat*II	0.5 mM DTT, 10 mM Mg-acetate, 33 mM Tris-acetate, 66 mM K-acetate, pH 7.9	37° C	Boehringer
*Bam*H1	10 mM $MgCl_2$, 50 mM Tris-HCl, 100 mM NaCl, pH 8.0	37° C	GibcoBRL
*Eco*RI	10 mM $MgCl_2$, 50 mM Tris-HCl, 100 mM NaCl pH 8.0	37° C	GibcoBRL
*Hin*dIII	10 mM $MgCl_2$, 50 mM Tris-HCl, 50 mM NaCl pH 8.0	37° C	GibcoBRL
*Nco*I	10 mM $MgCl_2$, 50 mM Tris-HCl, 100 mM NaCl, pH 8.0	37° C	GibcoBRL
*Nde*I	10 mM $MgCl_2$, 50 mM Tris-HCl, 50 mM NaCl pH 8.0	37° C	GibcoBRL
*Pvu*II	6 mM $MgCl_2$, 50 mM Tris-HCl, 50 mM NaCl, 50 mM KCl, pH 7.4	37° C	GibcoBRL
*Sma*I	5 mM $MgCl_2$, 20 mM Tris-HCl, 50 mM KCl, pH 7.4	30° C	GibcoBRL
*Xho*I	10 mM $MgCl_2$, 50 mM Tris-HCl, 50 mM NaCl pH 8.0	37° C	GibcoBRL

2.0.9 DNA Purification

Plasmid Purification with Diatomaceous Earth

1.5 mL of an overnight culture of *E. coli* grown from a single colony was transferred to a 1.5 mL Eppendorf tube. The tube was centrifuged for 1 minute. The supernatant was discarded, and a fresh 1.5 mL aliquot of *E. coli* was added to the tube. The tube was then centrifuged a second time and the supernatant discarded. The cell pellet was resuspended in 200 µL of sterile Cell Resuspension Buffer (50 mM glucose, 25 mM Tris-HCl, 10mM EDTA, pH 8.0). 200 µL of freshly prepared Alkaline Lysis Solution (0.2 M NaOH, 1% SDS) was added to the tube (Birnboim and Doly 1979). The solution was mixed thoroughly without vortexing to minimize shearing of DNA and incubated at room temperature for 5 minutes. To precipitate the chromosomal DNA and SDS, 200 µL of 2.55 M potassium acetate pH 4.8 was added to the tube. The tube was centrifuged for 5 minutes and the supernatant was transferred to a clean Eppendorf tube. The centrifugation process was repeated until all the cellular debris and chromosomal DNA was removed. To eliminate cellular RNA, DNase-free RNase A was then added to the tube at a minimum concentration of 10 µg/mL and the solution was incubated for 1 hour at 37°C in a recirculating water bath. After incubation, an equal volume (600 µL) of 6 M Guanidine-HCl was added to the tube to denature all protein.

Diatomaceous earth (Sigma) was suspended in ddH_2O to a concentration of 50 mg/mL (Kim and Pallaghy, 1996). The solution was left to stand overnight. The next day, the milky suspension above the sediment was decanted. The washing procedure was repeated 3 times. The diatomaceous earth was shaken until completely resuspended and

a 600 µL aliquot was added to a 1.5 mL Eppendorf tube that was then centrifuged for 30 seconds. The supernatant was removed until the total remaining volume was 200 µL. This step is necessary to concentrate the diatomaceous earth so that 1.5 mL Eppendorf tubes could be used throughout the purification process. The diatomaceous earth was resuspended in the remaining liquid and added to the plasmid prepared as described above. To allow binding of plasmid DNA to the diatomaceous earth the solution was mixed gently at room temperature for 5 minutes (Kim and Pallaghy 1996). Half of the suspension (approximately 700 µL) was added to a spin filter in a clean Eppendorf tube. The tube was centrifuged for 3 minutes. The filtrate was discarded and the remainder of the plasmid was added to the spin filter, and the centrifugation repeated. Two centrifugations are necessary because the spin filter only holds 700 µL. The filtrate was again discarded, and the spin filter containing the diatomaceous earth was placed in a clean Eppendorf tube. The diatomaceous earth was washed three times with 500 µL of 80% ethanol, each time centrifuging 1 minute and removing the spin filter to a clean Eppendorf tube. After the final wash, the spin filter was again placed in a clean Eppendorf tube and centrifuged for 1 minute to remove any vestiges of liquid. The plasmid DNA was eluted by placing the spin filter in a clean Eppendorf tube, adding 50 µL of TE buffer, and incubating at room temperature for 10 minutes. The filter was then centrifuged for 3 minutes and the eluate containing the plasmid DNA was stored at -20 °C. A sample of the plasmid DNA was digested with a restriction enzyme that cuts at a single site to produce linear DNA, which was electrophoresed on an agarose gel to assess purity. This purification protocol was often scaled up to purify plasmid from as much as 50 mL of *E. coli* culture.

Freeze-Squeeze DNA Purification

The freeze-squeeze method of DNA purification was used to purify *tat* DNA and some restriction endonuclease-digested plasmid DNAs according to the protocols of Thuring *et al.* (1975) and Tautz and Renz, (1983). The DNA was first electrophoresed on a purification-grade agarose (Fisher) gel. A solution was prepared by adding EDTA pH 8.0 to a final concentration of 1mM to a solution of 0.3 M sodium acetate pH 5.2. The band of interest was excised from the gel and immersed in the solution using a volume in mL equal to 10 × the weight of the agarose plug in grams. The agarose plug was soaked for 15 minutes with shaking, frozen in liquid N_2, and crushed between layers of wax paper. The crushed plug was placed in a spin filter in a 1.5 mL Eppendorf tube before it melted. The tube was then centrifuged for 15 minutes. The eluted liquid containing DNA was precipitated with 2 to 3 volumes of ice-cold 95 % ethanol.

Prep-a-Gene DNA Purification

The Prep-A-Gene DNA purification kit (BioRad) was used to purify some restriction endonuclease-digested plasmid DNA (Willis *et al.*, 1990). The protocol was provided with the kit. The DNA to be purified was electrophoresed on a 0.7% purification grade agarose gel. An estimate of the amount of DNA present in the gel bands of the ethidium bromide stained gel was made by visual comparison under UV-light with DNA standards of known concentration. The band of interest was excised from the gel with a clean scalpel blade and centrifuged for several seconds to bring the gel slice to the bottom of a 1.5 mL Eppendorf tube. The volume of the gel slice was

estimated by pipetting water into a second tube until the water level equalled that of the gel pellet. Approximately three volumes of Prep-A-Gene DNA purification kit binding buffer (6 M sodium perchlorate, 50 mM Tris-HCl pH 8.0, 10 mM EDTA) were then added to the gel pellet. Agitation and heating to 55°C for several minutes assisted the dissolution of the agarose plug in the sodium perchlorate binding buffer. Prep-A-Gene matrix (high purity diatomaceous earth in deionised water) was then added at 5 μL of matrix per μg of DNA, and the suspension was mixed gently by inverting the tube continually for 10 minutes at room temperature. The matrix was pelleted by centrifugation at 12,700 × *g* for 30 seconds. The matrix-bound plasmid DNA was then resuspended in a volume of binding buffer equal to 25 times the volume of matrix added. The matrix was again pelleted by centrifugation for 30 seconds, and the supernatant removed and discarded. The pellet was then washed twice in 25 times the matrix volume of Prep-A-Gene wash buffer (2 mM EDTA, 20 mM Tris-HCl pH 7.5, 400 mM NaCl, 50% ethanol). After the second wash, centrifugation, and removal of supernatant, a final centrifugation was performed to pellet the matrix, and allow elimination of all vestiges of liquid. The removal of all wash buffer supernatant prior to elution of the DNA is important since high salt concentrations and ethanol can inhibit enzymes. The DNA was then eluted using a minimum of one matrix volume of Prep-A-Gene elution buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and incubation at 50°C for 5 minutes. The matrix was then pelleted by centrifugation and the DNA-containing supernatant removed to a clean Eppendorf tube. The plasmid was stored at -20°C.

DNA Purification for Sequencing

Purification of plasmid DNA for sequencing was performed according to a protocol supplied by University Core DNA Services (UCDNA Services) at the University of Calgary. Alkaline lysis of a 25 mL *E. coli* culture in LB medium was performed as previously described (Birnboim and Doly, 1979). RNase A was then added to the supernatant to a total concentration of 20 µg/mL and the mixture was incubated at 37°C for 1 hour. Then an equal volume of a 25:24:1 solution of phenol:chloroform:isoamyl alcohol mixture was added to each Eppendorf tube and mixed by inversion to form an emulsion. The mixture was centrifuged for 1 minute at room temperature, and the aqueous phase was transferred to a clean Eppendorf tube. This extraction was repeated once. The aqueous phase was then extracted twice with an equal volume of chloroform, and the aqueous phase was removed to a clean Eppendorf tube. The DNA was then precipitated by adding an equal volume of 100% isopropanol and centrifuged at room temperature for 20 minutes. The DNA pellet was then washed with 500 µL of 70% ethanol, the ethanol removed, and the pellet dried under vacuum for 3 minutes. After dissolving the pellet in 32 µL of ddH_2O, the DNA was precipitated by adding NaCl to a final concentration of 0.4 M and then polyethylene glycol 8000 (PEG 8000) to 6.5% in a total volume of 80 µL. The solution was mixed thoroughly and then incubated on ice for 20 minutes. The precipitated DNA was pelleted by centrifugation at 4°C for 15 minutes. The pellet was rinsed with 500 µL of 70% ethanol, and dried under vacuum for 3 minutes. The pellet was then resuspended in 12 µL of ddH_2O and stored at -20°C. A sample of the plasmid DNA was then electrophoresed on an agarose gel to assess purity.

2.0.10 Polymerase Chain Reaction Primer Design

Four primers were designed to amplify *tat* DNA from a pUC19 vector containing HIV-1 BH10 *tat*. The reverse primer (AmberTat) for the 3' terminus of the *tat* DNA had a non-complementary tail incorporating an *Eco*RI restriction endonuclease site and a TAG (amber) termination codon (see Table 2.2). Two forward primers were designed to add an *Nco*I restriction endonuclease site or an *Nde*I restriction endonuclease site at the 5' terminus of the *tat* DNA (see Table 2.2). The *tat* first exon sequence in pTatR contains a *Bam*H1 site 181 nucleotides from the 5' end. To prepare for ligation of the first exon of the *tat* DNA in pTatR and the second exon of *tat* from pTatKL1, the BamTat primer (see Table 2.2) was designed and two mutations were introduced into the pTatKL1 *tat* DNA sequence. This created a *Bam*H1 restriction endonuclease site 181 nucleotides from the 5' end of the *tat* DNA. Restriction endonuclease digestion of pTatR with *Bam*H1 and *Eco*RI would remove the 3' end of the tat gene and permit ligation with similarly digested amplified *tat* DNA from pTatKL1. The resultant pET28b(+) plasmid contains a 258-nucleotide *tat* DNA coding for an 86 amino acid protein. All four primers were obtained from Life Technologies (Gibco). Primers arrived lyophilized and were dissolved in 300 µL of TE buffer pH 8.0. Concentrations of the reconstituted primers were determined by measuring the absorbance at 260 nm and calculating the amount of DNA present based on the estimate of 20 µg/mL per O.D. unit for single stranded DNA oligomers (Sambrook *et al.*, 1989).

Table 2.2 PCR Primers

Non-complementary portions of the sequences are in lower case letters, the portion complementary to the *tat* gene is in upper case letters, and the restriction endonuclease recognition sites are underlined. The melting points were calculated from the following equation: $T_M = 67.5°C + 0.34$ (the G + C content as a whole number) - (395/the length of the oligomer). Only the complementary portion of the primers was used for the calculation.

Primer Name	**Sequence**	**T_M (°C)**
AmberTat	5' gcg aat tct aTT CCT TAG GAC CTG TCG GGT CAC C 3'	70.8
NcoTat	5' gcc ggc cAT GGA ACC AGT CGA CCC TAG ACT GG 3'	70.7
NdeTat	5' tag atc gtc atA TGG AAC CAG TCG ACC CTA GAC TGG 3'	70.7
BamTat	5' aat agc GGa TCc CAG ACT CAT CAA GTT TCC TTG TCC AAG CAA C 3'	70.7

2.0.11 The Polymerase Chain Reaction

The Polymerase Chain Reaction (Mullis *et al.*, 1986) was performed according to protocols provided with the Expand High Fidelity PCR System (Boehringer Mannheim). This system uses a mixture of Taq DNA polymerase and Pwo DNA polymerase. The Pwo enzyme contains a 3' to 5' proofreading capability not present in Taq polymerase, which reduces the incorporation of errors. Each PCR reaction contained each dNTP in a concentration of 200 μM, each primer in a concentration of 300 nM, an estimated 0.1 to 0.75 μg of template DNA, 2.6 U of Expand High Fidelity PCR System Enzyme mix and Expand HF buffer. Boehringer does not supply any information on the components of the Expand HF buffer other than the presence of $MgCl_2$ at a final concentration of 1.5 mM. The PCR mixture was placed in an MJ Research Inc. PTC-100 programmable thermal controller programmed as outlined in Table 2.3. The PCR product was electrophoresed on an 8% polyacrylamide gel to determine if amplification was successful.

2.0.12 Preparation of *tat* DNA for Ligation

Preparation of *tat* DNA for cohesive end ligation involved purifying the PCR product using the freeze-squeeze method, digestion with the appropriate restriction endonucleases, and a further freeze-squeeze purification. When *tat* DNA was prepared for a blunt end ligation, the cohesive ends produced by the PCR had to first be blunted. Blunting the termini of the PCR product involved using the Klenow enzyme (a fragment of the *E. coli* polymerase lacking the 5' to 3' exonuclease activity) that polymerizes the addition of mononucleotides to the 5' hydroxyl terminus of primer DNA, thus producing blunt ends. T4 polynucleotide kinase then adds phosphate to the 5' hydroxyl group of the

Table 2.3 Polymerase Chain Reaction Program for the thermal cycler.

Step	Length	Temp. (°C)	Explanation
1	2 min.	94	Initial melting period
2	1 min.	94	Melting period per cycle
3	1 min.	55	Annealing period
4	1 min. + 20 sec. per cycle	72	Amplification period per cycle
5	return 19 times to step 2		Cycling of program
6	7 min.	72	Final period of amplification
7	indefinite	4	Refrigeration of sample prior to recovery
8	end		End of program

terminal nucleotides, which is necessary for ligation to occur. In order to inactivate the polymerase enzymes remaining in the reaction mixture, it was placed in the thermal cycler, and the program TaqDeath was run (outlined in Table 2.4). The reaction mixture was then incubated at 37° C for 30 minutes after addition of 4 U of Klenow enzyme (Boehringer), ATP disodium salt to a concentration of 2 mM, and dNTPs to 0.5 mM each (the PCR buffer contained the Mg^{2+} required by the Klenow enzyme). After 30 minutes, the temperature was increased to 70° C for 15 minutes to inactivate the Klenow enzyme. The PCR product was then purified using the freeze-squeeze method. A sample of the purified PCR product was electrophoresed on a 0.7% agarose gel and visualized under UV light after ethidium bromide staining. Concentration was estimated by comparison of band intensity with a known standard. When enough product was available, spectrophotometric estimation of DNA concentration was used with the estimate of 50 µg/mL of DNA per unit of absorbance at 260 nm (Sambrook *et al.*, 1989). The PCR product DNA was then incubated at 37° C for 30 minutes with 10 U of T4 Polynucleotide Kinase (GibcoBRL) per µg of DNA in the supplied reaction buffer. The solution was then incubated for a further 10 minutes at 70° C to inactivate the enzyme.

2.0.13 DNA Ligation

Ligation of restriction endonuclease-digested plasmid DNA and PCR amplified *tat* DNA was accomplished using protocols from Sambrook *et al.* (1989), the Novagen pET System Manual and information supplied with the T4 DNA ligase (Boehringer Mannheim). The plasmid vector was agarose gel-purified to eliminate any residual nicked or supercoiled plasmid. The production of recombinant plasmids involved both

Table 2.4 TaqDeath Program for the thermal cycler.

Step	Length	Temp. °C	Explanation
1	10 mins.	99°C	Heat inactivation of Polymerase.
2	50 mins., 40 secs.	0.1°C/4 secs.	Slow cooling to room temperature.
3	Indefinite	4°C	Refrigeration of sample prior to recovery.
4	End		End of program.

cohesive end and blunt end ligation. The ligation reaction contained an estimated 0.015 pmol of restriction endonuclease digested, dephosphorylated plasmid for a cohesive end ligation, and an estimated 0.15 pmol of similarly prepared vector for a blunt end ligation. An estimated 0.2 pmol of gel-purified *tat* DNA was added in a cohesive end ligation, while an estimated 2 pmol of insert was added for a blunt end ligation reaction. In addition, T4 DNA ligase buffer (66 mM Tris-HCl, 5 mM $MgCl_2$, 1 mM dithioerythritol, 1 mM ATP, pH 7.5), 1 unit of T4 DNA ligase (1 µL), and ddH_2O to a final volume of 10 µL or 20 µL was added. The enzyme was added last to the solution and it was incubated at 16°C overnight. Each ligation was performed in multiples, with varying volumes of *tat* DNA present. This was necessary because of uncertainty in the amount of DNA present in the *tat* DNA preparation. In addition, each set of ligation reactions contained a control, with no *tat* DNA present, to test for any background of ligated plasmid that does not contain *tat* DNA.

2.0.14 Preparation of Competent *E. coli*

All manipulations of *E. coli* were performed using aseptic technique (Tartoff and Hobbs, 1987). Luria-Bertani (LB) or Terrific Broth (TB) medium and LB agar plates were prepared using standard methods (Sambrook *et al.*, 1989). Ampicillin (Sigma), kanamycin (Sigma), chloramphenicol (Sigma), and tetracycline (Sigma) were added to media to concentrations of 50 µg/mL, 30 µg/mL, 34 µg/mL, and 12.5 µg/mL respectively, unless stated otherwise. Antibiotic stock solutions (typically 1000 × final concentration) were prepared in water (ampicillin and kanamycin) or 95 % ethanol (chloramphenicol and tetracycline). Aqueous solutions were filter sterilised using a 0.22

micron, surfactant free cellulose acetate membrane (Nalgene). Preparation of competent *E. coli* was accomplished using the protocol described in Sambrook *et al.* (1989). Briefly, *E. coli* were grown in LB medium to mid-log phase (A_{600} 0.4-0.8), incubated on ice for 5 minutes, centrifuged, resuspended in one half volume of ice cold $CaCl_2$, and incubated on ice for 30 minutes. The cells were then pelleted by centrifugation, the supernatant removed, and the cells resuspended in one tenth volume of ice cold $CaCl_2$. The cells were then stored at 4 °C for up to one week, or prepared as frozen cell stocks in the presence of 10 % glycerol and kept at -60 °C.

2.0.15 Transformation of *E. coli*

The methods of Pope and Kent (1996) were used to accomplish rapid transformation of competent cells with plasmid DNA. When plasmids were not carrying kanamycin resistance the following protocol was used. Two aliquots of competent *E. coli* in 1.5 mL screw-top tubes were recovered from storage, thawed, and placed on ice. An aliquot of purified plasmid DNA solution containing between 1 and 10 ng (approximately 2 µL) was added to one aliquot of cells, gently mixed with a pipette tip, and left to incubate on ice for 5 minutes. The transformed cells were then spread in 50 µL, 10 µL, and 2 µL aliquots, and the non-transformed cells in a single 100 µL aliquot, on LB agar plates containing the appropriate selective antibiotics. A 10 µL aliquot of transformed cells was spread on an LB plate without antibiotics to assess the viability of the cells. The plates were left for 15 minutes at room temperature to allow the liquid to absorb. The plates were then inverted and incubated overnight at 37 °C.

When the plasmid carries kanamycin resistance, a heat shock and recovery step is required for efficient transformation and optimum growth (Pope and Kent 1996). Two aliquots of competent cells in 1.5 mL screw-top tubes were recovered from storage and kept on ice. One aliquot of competent cells was mixed with plasmid DNA and left on ice for 5 minutes. The competent cells were then incubated in a recirculating water bath at 42 °C for 2 minutes after which they were placed on ice for a further 5 minutes. The cells were then added to 500 µL of antibiotic-free LB medium in a sterile 12 mL plastic culture tube and incubated for 1 hour at 37 °C with vigorous gyrotatory shaking (approximately 150 rpm). The cells were then returned to the 1.5 mL screw-top tubes, centrifuged, and the supernatant removed until the total volume was approximately 100 µL. The cell pellets were then resuspended and spread on LB agar plates containing kanamycin and any other appropriate antibiotics, and incubated inverted at 37 °C overnight. The next day LB medium was inoculated with a single colony from the plates and plasmid was purified from this culture using diatomaceous earth. The recombinant plasmids and the original plasmid vectors lacking a *tat* DNA insert were then restriction-digested to produce linear DNA. The digested plasmids were electrophoresed on an agarose gel and their relative mobilities compared to determine if *tat* DNA was present in the recombinant plasmid. Alternatively, digestion with a restriction endonuclease that cleaves a single site in the *tat* DNA and does not cleave the plasmid vector DNA was used to determine whether the *tat* DNA was present.

2.1 Sub-cloning the HIV-1 BH10 *tat* DNA

2.1.0 General Information

The source of the HIV-1 *tat* DNA was pTatC6H-1. The pTatC6H-1 plasmid contains a two-exon HIV-1 *tat* DNA coding for an 86 amino acid Tat protein, in a pDS56, RB, 6× His expression plasmid carrying an ampicillin resistance gene (see Table 2.5 and Figure 2.6) (Purvis *et al.*, 1995). The pTatC6H-1 plasmid was supplied as a frozen cell stock of pTatC6H-1 transformed *E. coli* M15::pDMI,1 (Certa *et al.*, 1986). The pDMI,1 plasmid carries kanamycin resistance and is integrated into the *E. coli* genome in this strain (Certa *et al.*, 1986).

2.1.1 Production of pTatKL1

It was not possible to PCR amplify directly the *tat* DNA from pTatC6H-1 because the sequence and size of the *tat* DNA insert was not supplied with the plasmid, therefore, the *tat* DNA was subcloned into pUC19. After determining that the *tat* DNA in pTatC6H-1 was between *Hin*dIII and *Eco*RI sites (see Figure 2.6), the plasmid was digested with these enzymes, electrophoresed on an agarose gel, and the putative *tat* DNA was then extracted from the gel using the freeze-squeeze method. This fragment was then ligated with freeze-squeeze purified, dephosphorylated pUC19 vector that had been digested with *Eco*RI and *Hin*dIII restriction endonucleases. The plasmid was then sequenced. The new plasmid pTatKL1 served as template for PCR amplifications of the *tat* DNA for the production of recombinant pET expression plasmids (see Figure 2.7).

Table 2.5 Cloning and Expression Plasmids

Plasmid	Type of Plasmid
‡pUC19	High copy number plasmid
†pET28b(+)	Expression vector
†pET27b(+)	Expression vector
*pTatC6H-1	Expression vector

‡The pUC19 high copy number plasmid was generously supplied by the laboratory of Dr. Peter Loewen (Department of Microbiology, University of Manitoba). †The pET27b(+) and pET28b(+) plasmids were purchased from Novagen. *pTatC6H-1 was prepared by Purvis *et al.* (1995) from a pDS56, RB, 6 × His expression plasmid and provided to the NIH AIDS Reference and Reagent Program.

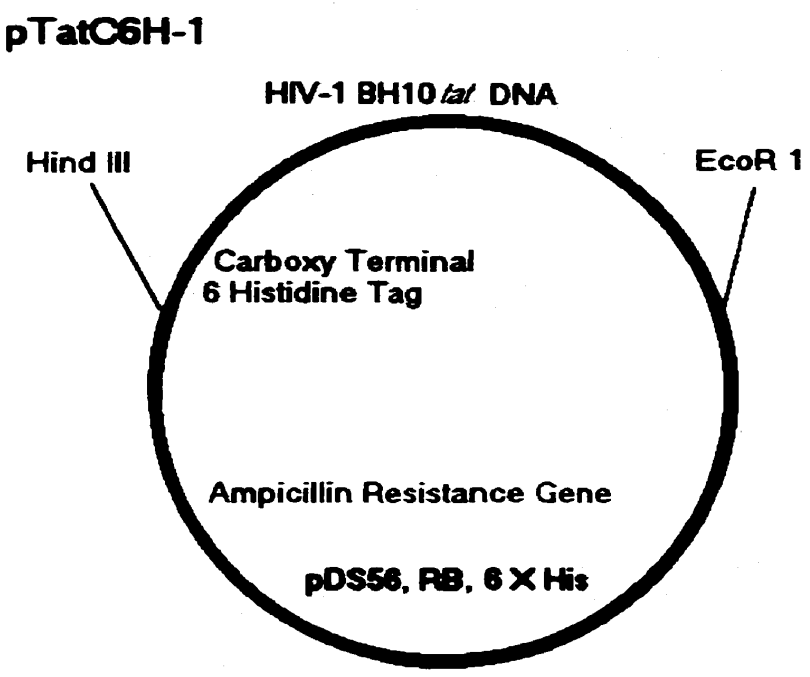

Figure 2.6 The pTatC6H-1 Plasmid. pTatC6H-1 is a recombinant pDS56, RB, 6 × His plasmid expressing Tat protein from HIV-1 BH10 *tat* DNA (red) with a carboxy-terminal six-histidine tag (green). It contains a gene conferring ampicillin resistance (brown).

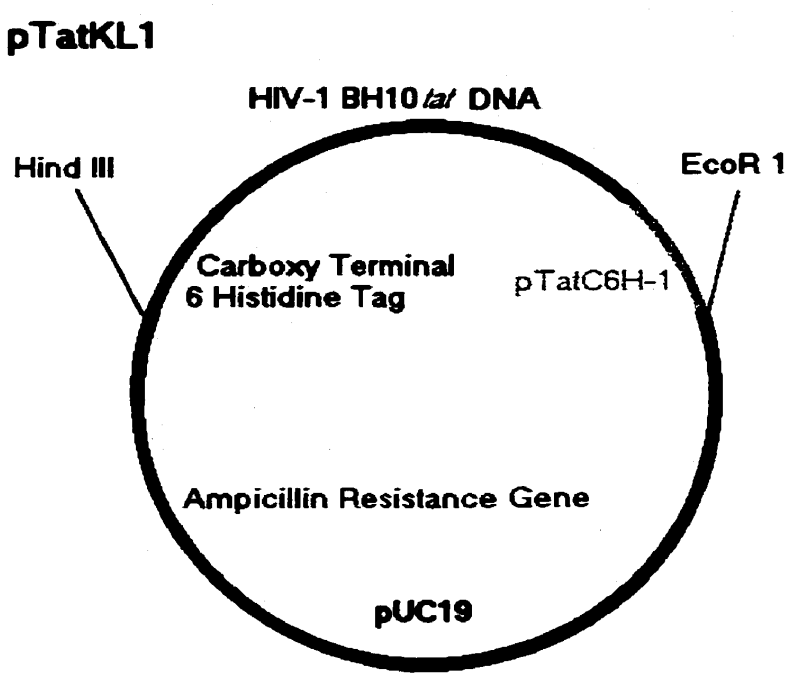

Figure 2.7 The pTatKL1 Plasmid. pTatKL1 is a recombinant pUC19 plasmid containing the *Eco*R1 and *Hin*dIII restriction endonuclease fragment from pTatC6H-1 (orange, red and green). This fragment contains HIV-1 BH10 *tat* DNA (red) with a carboxy terminal six-histidine tag (green). The plasmid contains a gene conferring ampicillin resistance (brown).

2.1.2 Production of pUCTat

The NcoTat and AmberTat PCR primers were used to amplify *tat* DNA from pTatKL1 with an *Nco*I restriction endonuclease site at the 5' end of the DNA, and an amber termination codon followed by an *Eco*RI restriction endonuclease site at the 3' end of the *tat* sequence. However, incubation of this PCR product with the *Nco*I enzyme failed to result in restriction endonuclease digestion at the 5' terminus. Therefore, the termini of the PCR product were blunted, and the DNA fragment was freeze-squeeze purified. The blunt ended *tat* DNA was then ligated into *Sma*I restriction endonuclease digested (produces blunt ends), dephosphorylated, Prep-A-Gene purified pUC19. pUCTat served as a source of *tat* DNA that could be removed by digestion with *Nco*I and *Eco*RI (see Figure 2.8).

2.1.3 Preparation of pET Plasmid Vectors

Production of pTatKL19 was accomplished by ligation of *Nco*I and *Eco*RI restriction endonuclease digested, dephosphorylated, Prep-A-Gene purified pET27b(+) vector with similarly digested *tat* DNA obtained from the pUCTat plasmid. The pTatKL19 plasmid has *tat* DNA inserted in register with the *pelB* leader sequence intended to cause secretion of the recombinant protein into the periplasmic space of the *E. coli* cell (see Figure 2.9) (Power *et al.*, 1992; Brown *et al.*, 1998).

Plasmid pTatKL16 was produced by ligating *Nde*I and *Eco*RI restriction endonuclease digested *tat* DNA obtained by PCR amplification of pTatKL1, with similarly digested pET28b(+) plasmid. This construct has the *tat* DNA inserted in codon

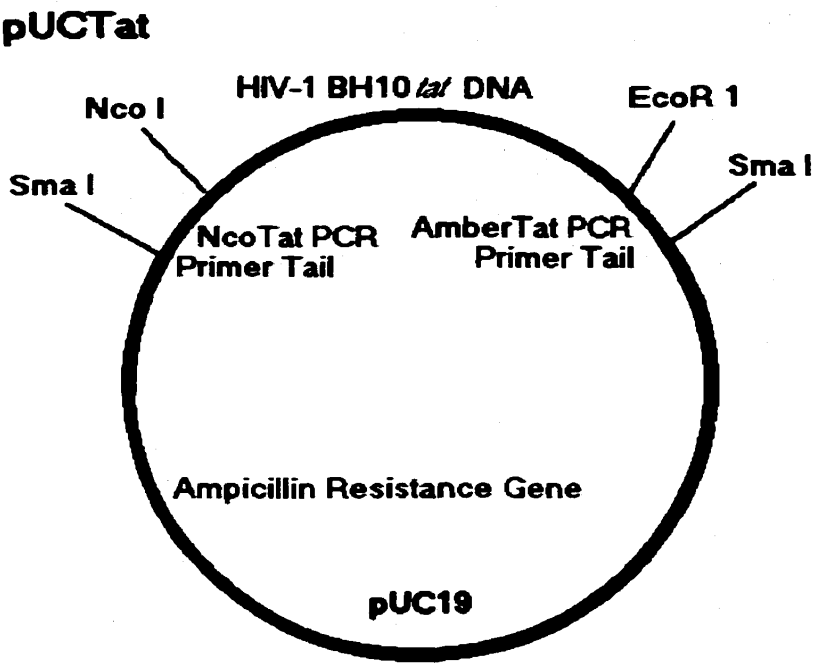

Figure 2.8 The pUCTat Plasmid. pUCTat is a recombinant pUC19 plasmid containing *tat* DNA (red) PCR amplified from pTatKL1 using the NcoTat and AmberTat primers (blue). The plasmid contains a gene conferring ampicillin resistance (brown).

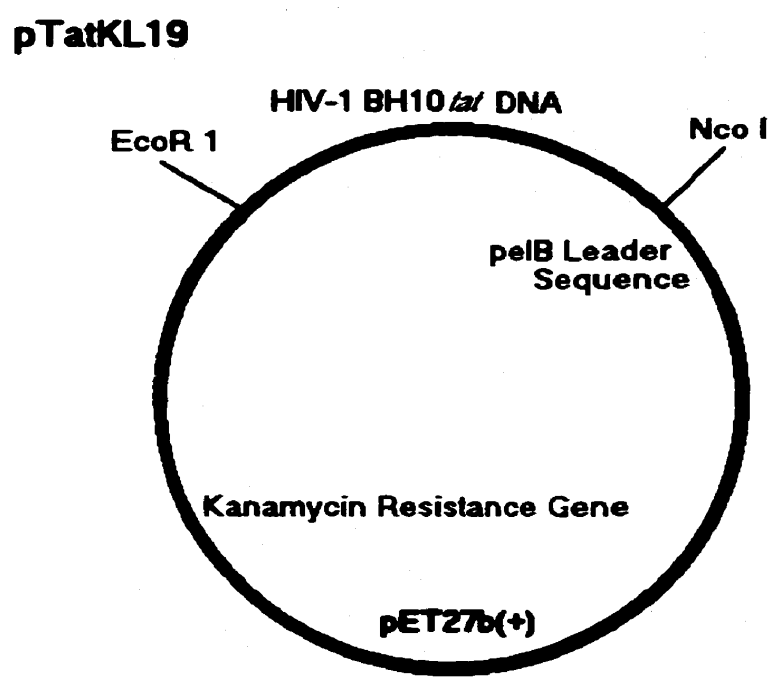

Figure 2.9 The pTatKL19 Plasmid. pTatKL19 is a recombinant pET27b(+) plasmid expressing Tat protein from HIV-1 BH10 *tat* DNA (red) with an amino terminal *pelB* leader sequence. The plasmid contains a gene conferring kanamycin resistance (brown).

register with a thrombin cleavable amino terminal six-histidine tag, which will bind a Ni^{2+} metal-chelate affinity column for purification (Hochuli *et al.*, 1987; Smith *et al.*, 1988) (see Figure 2.10).

Amplification of pTatKL1 with the BamTat and AmberTat primers yielded DNA encompassing part of the first and all of the second exon of *tat*. This was digested with *Bam*H1 and *Eco*RI and ligated with similarly digested pTatR plasmid (see Figure 2.11). The pTatR plasmid was produced by Dr. Gillian Henry, and consists of a pET28b(+) plasmid which contains the first exon of HIV-1 BH10 *tat* DNA producing Tat protein with an amino terminal 6 histidine tag. The sequence of *tat* DNA in the pTatR plasmid contains mostly codons that belong to the most abundant tRNAs in *E. coli* (Ikemura, 1981). Therefore it appears the sequence has been optimized for expression in *E. coli*. Thus, the pTatKL26 plasmid contains the first 186 nucleotides of the *tat* DNA from the pTatR plasmid and the final 72 nucleotides of the HIV-1 BH10 *tat* DNA from pTatKL1 (see Figure 2.12) for the purpose of producing an 86 amino acid residue Tat protein.

The pTatKL16 and pTatKL19 plasmids were digested with the restriction endonuclease AatII and electrophoresed on an agarose gel. The *tat* DNA in these constructs contains a single AatII site, and the pET plasmids do not. Observation of a single band of linear plasmid at the appropriate M_r on the gel is evidence the *tat* DNA had been inserted into the pET vectors.

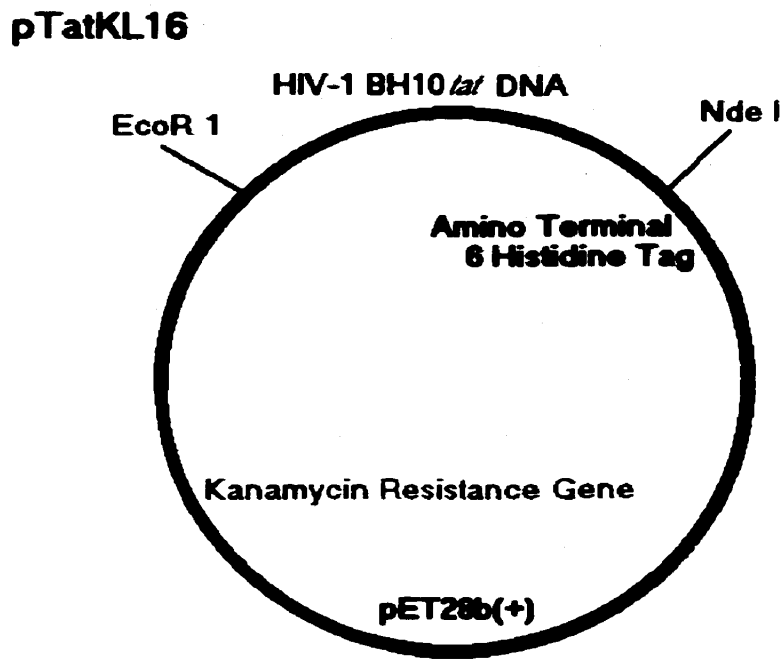

Figure 2.10 The pTatKL16 Plasmid. pTatKL16 is a recombinant pET28b(+) plasmid expressing Tat protein from HIV-1 BH10 *tat* DNA (red) with an amino terminal six-histidine tag (blue). The plasmid contains a gene conferring kanamycin resistance (brown).

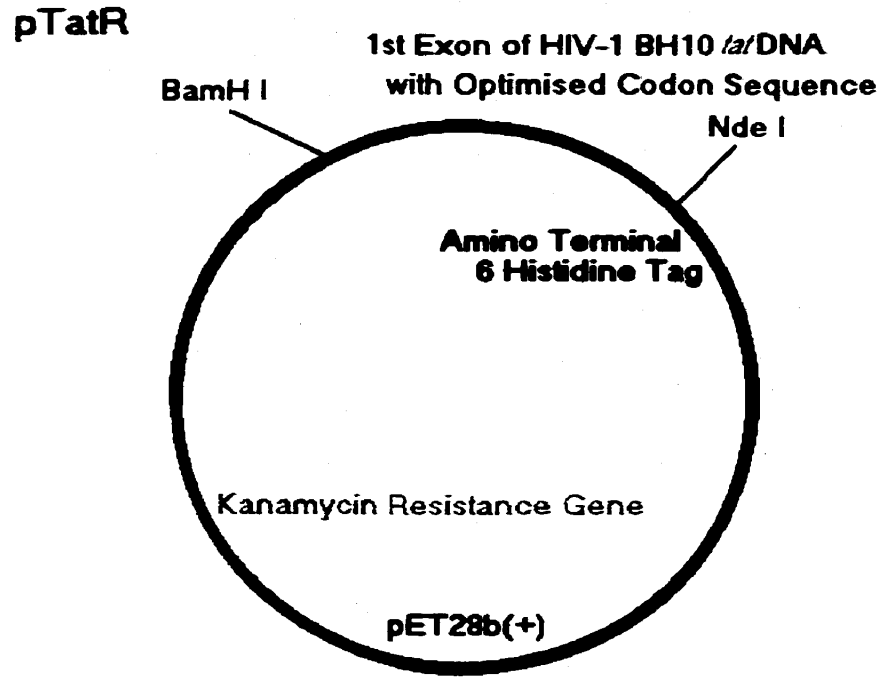

Figure 2.11 The pTatR Plasmid. pTatR is a recombinant pET28b(+) plasmid expressing a 72 amino acid Tat protein from the first exon of *tat* DNA (red). The *tat* DNA has a codon sequence optimised for expression in *E. coli*. The protein is expressed with an amino terminal six-histidine tag (blue) and the plasmid contains a gene conferring kanamycin resistance (brown).

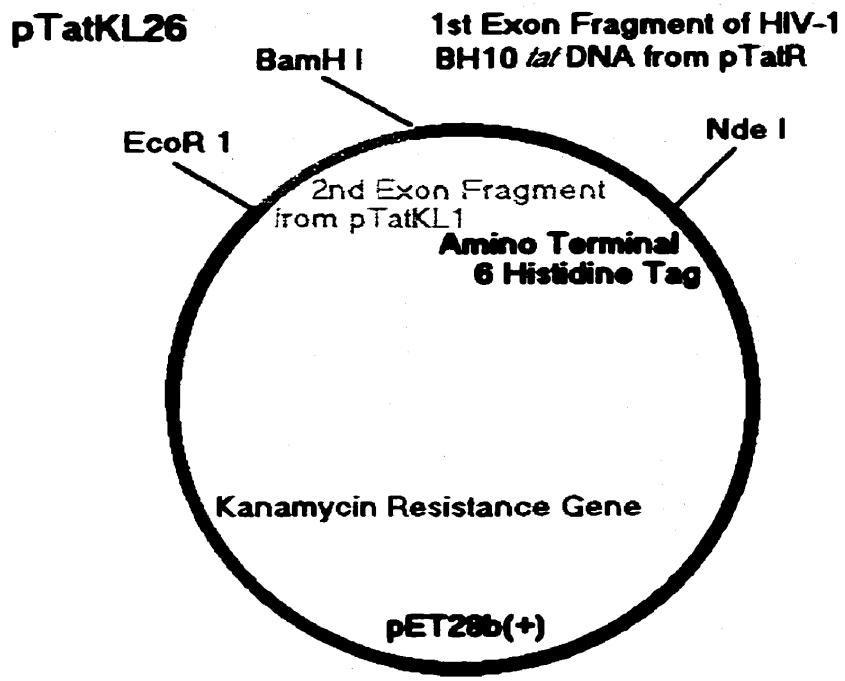

Figure 2.12 The pTatKL26 Plasmid. pTatKL26 is a recombinant pET28b(+) plasmid expressing an 86 amino acid Tat protein with an amino terminal six-histidine tag (blue). It contains a fragment of *tat* DNA with a codon sequence optimised for expression in *E. coli* encompassing the majority of the first exon (red) ligated to a fragment of HIV-1 BH10 *tat* DNA encompassing the remainder of the gene (grey). The plasmid contains a gene conferring kanamycin resistance (brown).

2.2 Protein Expression Methods and Reagents

2.2.0 Induction of Recombinant Protein Expression and Growth Curves

Cells were grown at 37° C with shaking until they reached mid-log phase. Then expression of recombinant Tat protein in *E. coli* BL21(DE3)pLysS was induced using 1 mM isopropylthio-β-D galactopyranoside. A 250 mL sidearm flask containing 50 mL of culture medium was inoculated with 1 mL of an overnight culture of *E. coli* and incubated with shaking at approximately 100 rpm. The turbidity of the growing cultures was measured with a Klett-Summerson colourimeter, and the data plotted using the program KyPlot on an IBM compatible computer.

2.2.1 Analysis of Tat Production by SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed following the method of Laemmli (1970). Coomassie staining of gels followed standard procedures (Sambrook *et al.*, 1989). SDS-PAGE was used to measure protein levels. Samples were prepared by centrifuging 500 µL aliquots of cell cultures for 3 minutes, the cell pellets were resuspended in 50 µL of ddH_2O, mixed with 50 µL of 2× SDS gel loading buffer (125 mM Tris-HCl pH 6.8, 4% SDS, 3.1% dithiothreitol (DTT), 20% glycerol) and stored at -20°C. Prior to being loaded on the gel, samples were heated at 70°C for 10 minutes. Preparation of whole cell samples with this method often led to difficulty in loading gels due to high viscosity. A solution was to resuspend the cells in 500 µL of 50 mM Tris-HCl pH 8.3, and mix an 80 µL aliquot of the resuspended cells with 20 µL of 5× SDS gel loading buffer.

When attempts were made to produce Tat with a PelB leader sequence, it was necessary to test for secretion of Tat protein into the culture medium. Samples of LB medium were prepared by lyophilization of 5 mL of medium and resuspension in 400 µL of dialysis buffer (50 mM Tris-HCl pH 7.6, 20 mM β-mercaptoethanol, 10 % glycerol). The samples were then dialysed overnight to remove salts concentrated by the lyophilization. Small dialysis clips were produced by cutting the bottoms off clean 1.5 mL Eppendorf tubes. Protein was pipetted into the cap (between 100 and 200 µL), dialysis tubing with a molecular weight cut-off of 5000 was placed across the opening (Spectra/Por CE cellulose ester, Spectrum Medical Industries), and the bottom of the Eppendorf tube was clipped over the tubing and cap to seal the apparatus. The dialysis clip was then placed in 500 mL of dialysis buffer that was stirred overnight at 4°C. The next day, the sample volumes were measured, and the appropriate sample volume to load on the gel was calculated based on the increase in volume after dialysis. The samples were then mixed with 5 × SDS gel loading buffer and stored at -20°C. Since guanidine will precipitate dodecylsulphate, any solutions containing guanidine were also dialysed as previously described, prior to loading on the gel.

Recombinant protein expressed in bacteria can often be sequestered in insoluble inclusion bodies. To test for Tat in inclusion bodies, soluble and insoluble fractions from whole cells were prepared by following a protocol from the pET System Manual. At four hours post-induction, 5 mL of cell culture was centrifuged in the Dynac centrifuge at the highest speed setting for 3 minutes. Cells were resuspended in 500 µL of 50 mM Tris-HCl, 2 mM EDTA pH 8.0, placed in a 1.5 mL Eppendorf tube, and 50 µL of a 1%

solution of Triton X-100 were added. The suspension was incubated on ice for 10 minutes and then sonicated with a Fisher Sonic Dismembrator (Model 300), using a microtip at a 35% power setting for two 30-second pulses. The solution was then centrifuged at 12,700 × *g* for 15 minutes at 4°C. The supernatant was removed and retained as the soluble cell fraction and the pellet retained as the insoluble fraction. The soluble material was mixed with an equal volume of 2 times concentrated SDS-gel loading buffer, while the insoluble material was resuspended in 100 µL of 1 × SDS gel loading buffer (62.5 mM Tris-HCl pH 6.8, 2% SDS, 1.55% dithiothreitol (DTT), 10% glycerol). The samples were heated at 70°C for 10 minutes prior to loading on the gel.

Dalton MK VII-L M_r markers (Sigma) were used for estimation of the M_r of Tat in SDS-PAGE. The markers are α-lactalbumin from bovine milk (M_r 14,200), soybean trypsin inhibitor (M_r 20,100), PMSF-treated trypsinogen from bovine pancreas (M_r 24,000), carbonic anhydrase from bovine erythrocytes (M_r 29,000), glyceraldehyde-3-phosphate dehydrogenase from rabbit muscle (M_r 36,000), egg albumin (M_r 45,000) and bovine albumin (M_r 66,000). After the gel was Coomassie-stained and destained, the mobilities of the protein bands were measured and used to estimate the M_r of Tat using standard procedures (Shapiro *et al.*, 1967; Weber and Osborn, 1969).

2.2.2 Western Immuno-Blotting of SDS-PAGE Gels

Western immuno-blotting of polyacrylamide gels followed standard procedures (Sambrook *et al.*, 1989). The primary antibody (mouse Monoclonal Anti-poly Histidine clone HIS-1 from Sigma) was diluted 1:3000 in phosphate buffered saline (PBS) (1.9

mM $NaH_2PO_4 \cdot H_2O$, 8.1 mM Na_2HPO_4, 154 mM NaCl, pH 7.4 in ddH_2O). The secondary antibody horseradish peroxidase conjugate (Fc specific anti-mouse IgG peroxidase conjugate developed in goat, adsorbed with bovine, horse and human serum proteins, from Sigma) was diluted 1:2000 in PBS. Incubation in the primary antibody was at room temperature with gentle agitation for 30 minutes. Incubation with the secondary antibody was at room temperature for 30 to 60 minutes with gentle agitation. After incubation with the secondary antibody, the membrane was developed with 3,3' diaminobenzidine tetrahydrochloride solution (Sigma) (0.6 mg/mL 3,3' DAB, 9 mM Tris-HCl pH 7.6, 0.03% $CoCl_2$, 0.03% $NiCl_2$, with 0.03% fresh H_2O_2 added last) until bands appeared. The membrane was air dried and stored protected from light.

2.3 Purification of Recombinant Tat Protein

2.3.0 Ni^{2+} Metal Chelate Affinity Chromatography

Preparation of Cells

A 10 mL overnight culture of *E. coli* BL21(DE3)pLysS transformed with pTatKL16 was used to inoculate 500 mL of LB medium. The culture was incubated at 37°C with gyrotatory shaking (approximately 75 rpm). Tat protein expression was induced by addition of IPTG to a final concentration of 1 mM when the A_{600} was between 0.4 and 0.6. Four hours after induction, samples for SDS-PAGE were prepared and the remaining culture was divided into two 250 mL centrifuge bottles. The bottles were centrifuged at 2800 × *g* for 10 minutes at 4°C, the pelleted cells flash frozen in liquid N_2, and stored at -20°C.

Purification of Recombinant Tat Under Denaturing Conditions

To purify recombinant Tat protein under denaturing conditions, the previously frozen cell pellets were thawed and resuspended in 50 mL of denaturing lysis buffer (1mM phenyl methyl sulphonyl fluoride (PMSF), 10 mM Tris-HCl, 10 mM β-mercaptoethanol, 50 mM sodium phosphate, 6 M guanidine-HCl, pH 8.0). The PMSF was dissolved in 95% ethanol prior to being added to the lysis buffer. To shear DNA and decrease viscosity, the solution was expelled through a 20 gauge needle twice, then twice through a 26 gauge needle, and finally once through a 27 gauge needle. The suspension was flash frozen in liquid N_2, thawed, and then sonicated on ice using a micro-tip with the output set to 35% for two 2-minute bursts, with a 30 second cooling interval. The suspension was then flash frozen in liquid N_2, thawed, and centrifuged at 7800 × *g* for 1 hour at 4°C. Samples of the supernatant and pellet were retained and stored at -20°C for later analysis. Meanwhile, 4 mL of Ni^{2+}-nitrilotriacetate (Ni-NTA) matrix in 30% ethanol (Qiagen) was poured into an 18 mL glass column and equilibrated with 10 mL of lysis buffer. Ni-NTA matrix was then added to the supernatant and gently stirred for two hours at room temperature. After two hours, the Ni-NTA suspension was centrifuged on the Dynac centrifuge at the lowest speed setting for 30 seconds. The majority of the supernatant was removed from the pelleted resin, and the remaining Ni-NTA resin was poured into a glass column 10 cm long with an internal diameter of 1.5 cm. A sample of the unbound supernatant was stored at -20°C for later analysis. The resin was washed with 50 mL of lysis buffer, and then with 50 mL to 100 mL of lysis buffer without PMSF. Samples of the column eluate were monitored for protein content by measuring the A_{280}. When the absorbance was 0.01 or less, the resin was washed with 50 mL of wash buffer

(10 mM β-mercaptoethanol, 50 mM sodium phosphate, 300 mM NaCl, pH 6.0). A final wash with 20 mL of 10 mM sodium phosphate buffer pH 6.0 containing 10 mM β-mercaptoethanol preceded the elution. The recombinant Tat protein was eluted with 50 mM sodium-acetate/acetic acid pH 4.5 into 30 tubes containing approximately 2 mL each and the A_{280} of each tube was measured. Samples of eluate were analysed by SDS-PAGE and Western immuno-blotting and the protein-containing fractions were combined and stored at -20°C.

Purification of Recombinant Tat under Non-Denaturing Conditions

The purification of recombinant Tat protein using non-denaturing conditions was similar to purification using denaturing conditions except for the following changes. All steps were performed at 4 to 8°C in the cold room. The lysis buffer contained 1mM PMSF, 10 mM β-mercaptoethanol, 10 mM imidazole, 50 mM sodium phosphate and 300 mM NaCl at pH 8.0. Wash buffer contained 10 mM β-mercaptoethanol, 20 mM imidazole, 50 mM sodium phosphate and 300 mM NaCl at pH 8.0. Elution buffer contained 10 mM β-mercaptoethanol, 50 mM sodium phosphate, 250 mM imidazole and 300 mM NaCl at pH 8.0.

Chapter Three: Results

3.0 Production of Recombinant Tat Protein

3.0.1 Strategy

Development of a strategy to express and purify a recombinant Tat protein had to take into account potential problems due to the nature of the Tat protein sequence. The HIV-1 BH10 Tat protein contains seven cysteine residues that are highly conserved in all HIV-1 strains. In addition, there is a highly basic region from arginine 49 to proline 58. The large number of cysteine residues could give rise to microheterogeneity as a result of inter-molecular and/or intra-molecular disulphide formation. A reducing environment during purification, to prevent disulphide formation, was chosen as the most suitable strategy for the production of homogenous Tat. The basic region of the protein and the known RNA binding properties of the Tat protein suggested that the removal of nucleic acids from the protein would also be necessary during purification. Finally, the level of purification needed is high as the ultimate goal, once sufficient amounts of protein are produced, is biochemical and structural characterization using NMR spectroscopy.

A widely used vector for expression of recombinant protein in *E. coli* is the multi-copy pBR322-based pET plasmid originally designed by Studier and Moffat (1986). As the name suggests (plasmid for elongation by T7 RNA polymerase) (Rosenberg *et al.*, 1987), pET plasmids take advantage of the fact that the RNA polymerase of bacteriophage T7 elongates about 5 times faster than *E. coli* polymerase and is highly selective for promoters that are rarely found in non-T7 DNA (Studier and Moffat, 1986; Studier *et al.*, 1990). In addition to the T7 RNA polymerase promoter and transcription

initiation site, pET plasmids also contain the highly efficient ribosome binding site from the T7 phage major capsid protein (Studier and Moffat, 1986). Commercially available pET vectors (Novagen) also incorporate various amino and carboxy terminal purification tags into the recombinant product. These include a six-histidine tag at either terminus of the protein for Ni^{2+} metal-chelate affinity chromatography (Porath *et al.*, 1975), and a PelB leader sequence at the amino terminus for secretion into the periplasmic space (Better *et al.*, 1988). Purification of recombinant protein from the periplasmic space should be very efficient due to relatively low levels of contaminating protein. In addition, the amino terminal six-histidine tag contains a thrombin cleavage site for removal of the purification tag. The plasmids are used with *E. coli* bacteriophage λDE3 lysogens that carry the T7 RNA polymerase gene under the control of the *lacUV5* promoter that is inducible by the addition of IPTG. A few hours after induction, up to 50 % of total cell protein is recombinant with the use of this system. Some pET plasmids also carry DNA coding for the *lac* repressor that represses transcription of both T7 RNA polymerase and recombinant DNA because of the presence of the T7*lac* promoter. This provides strict repression of basal levels of T7 RNA polymerase activity (Dubendorff and Studier, 1991).

There are potential problems that must be considered concerning the use of thrombin to cleave the six-histidine tag. First, the carboxy terminus of the Tat protein contains an RGD thrombin cleavage site (amino acids 78-80). This site is however not expected to be a good substrate for thrombin since a thrombin inhibitor contains a similar sequence (Theunissen *et al.*, 1993), thus limited proteolysis might work. The use of

thrombin to cleave an amino terminal 6 histidine tag could be avoided however by using cyanogen bromide, which will cleave at the carboxy terminal side of methionine resulting in the loss of the methionine residue at the amino terminus. The only methionine residues in the sequence of the expressed Tat protein result from the initiation codons at the beginning of both the six-histidine tag sequence and the Tat sequence. A second potential problem of using a metal affinity purification step is that even minute levels of paramagnetic metal ions such as Ni^{2+} are unacceptable in a sample prepared for NMR spectroscopy. Despite the potential for contamination by metal, this method has become relatively common in the preparation of samples for NMR spectroscopy as metal contamination can generally be eliminated in subsequent purification steps (Kirsch *et al.*, 1996).

Because of the potential problems with removal of the amino terminal six-histidine tag, and contamination by nickel, the PelB leader sequence was an attractive avenue of investigation. The *E. coli* chaperone protein SecB binds proteins containing a PelB leader sequence (Kumamoto, 1989). The protein is then shuttled to the membrane where it binds the protein SecA (Hartl *et al.*, 1990), which is part of a protein complex involving the SecY and SecE integral membrane proteins (Brundage *et al.*, 1990). The SecA protein functions as an ATPase when bound by the protein containing the PelB leader (Lill *et al.*, 1989). The hydrolysis of ATP, and production of a proton motive force then drives the translocation of the protein across the membrane (Economou *et al.*, 1995). The PelB leader peptide is cleaved from the protein in the process (Wickner *et al.*, 1991). The elimination of thrombin and cyanogen bromide cleavage by using a PelB leader

would mean one less opportunity for loss of Tat protein during purification. The purification of recombinant Tat protein using the PelB leader would be potentially very efficient due to the lower levels of contaminating proteins in the periplasmic space, as well as the absence of nucleic acids there. The greatest concern in using this method would be the effect the periplasmic environment might have on the oxidative state of the seven cysteine residues.

During the course of the present work it was observed that Tat protein expression from pTatR was much higher than from pTatKL16. Both are pET28b(+) expression plasmids producing proteins with an amino terminal six-histidine tag. pTatR produces a one-exon Tat_{1-72} whereas pTatKL16 produces a two-exon Tat_{1-86}. The *tat* DNA in pTatKL16 was optimised for expression in both yeast and *E. coli*, with the necessity that some compromises in codon usage were necessary (Adams *et al.*, 1988). Examination of the nucleotide sequence of the *tat* DNA in pTatR shows that unlike the *tat* DNA in pTatKL16, most of the codons correspond to the most frequently used codons in *E. coli* (see Table 3.1) (Wada *et al.*, 1991). Furthermore, the frequency of codon usage has been correlated to the abundance of tRNAs specific for each codon (Ikemura, 1981). This suggests that the nucleotide sequence of the *tat* DNA in pTatR has been more substantially optimised for expression in *E. coli* compared to the *tat* DNA in pTatKL16. For example, in pTatR only 1 rare codon is used for amino acids 50 to 60 whereas in pTatKL16 *tat* 5 amino acids are encoded by low frequency codons (see Table 3.2). It was therefore decided that the second exon from pTatKL16 should be added to pTatR in an attempt to increase expression levels.

Table 3.1 Codon Frequencies in HIV and *E. coli*

Amino Acid	Codon	# Codons /1000		pTatR	pTat KL16 1st exon	Amino Acid	Codon	# Codons /1000		pTatR	pTat KL16 1st exon
		HIV	*E. coli*					*HIV*	*E. coli*		
ARG	CGA	6.3	3.1	1		GLY	GGG	18.2	9.6	1	
	CGC	2.6	22.0				GGU	7.1	28.0	1	4
	CGG	3.1	4.6		1	VAL	GUA	24.7	11.8		
	CGU	0.9	24.7	6	3		GUC	7.7	14.3	1	1
	AGA	39.9	2.0		3		GUG	15.0	25.3		
	AGG	17.7	1.3				GUU	7.3	20.4	2	2
LEU	CUA	15.8	3.0	1		LYS	AAA	31.1	36.9	8	4
	CUC	11.3	9.8				AAG	25.0	11.9		4
	CUG	15.3	54.8	2	1	ASN	AAC	18.2	24.2	1	1
	CUU	10.7	9.9				AAU	31.0	15.9		
	UUA	20.5	10.3			GLN	CAA	27.0	13.0		4
	UUG	13.6	11.2		2		CAG	27.2	30.1	7	3
SER	UCA	12.8	6.3			HIS	CAC	8.7	11.0	3	2
	UCC	8.0	9.6	2	4		CAU	15.5	11.5		1
	UCG	2.7	7.9			GLU	GAA	41.1	43.7	2	2
	UCU	7.5	10.4	3	1		GAG	25.6	19.3		
	AGC	14.6	15.0			ASP	GAC	17.7	22.3	1	1
	AGU	14.0	7.1				GAU	19.3	32.0		
THR	ACA	28.4	6.4			TYR	UAC	8.9	13.4	2	2
	ACC	14.4	24.6	4	2		UAU	18.4	14.9		
	ACG	4.4	12.5			CYS	UGC	10.0	6.2	7	2
	ACU	15.2	10.5		2		UGU	15.0	4.7		5

Table continued on next page.

Table 3.1 (continued) Codon Frequencies in HIV and E. coli											
Amino Acid	**Codon**	**# Codons /1000**		**pTatR**	**pTat KL16 1st Exon**	**Amino Acid**	**Codon**	**# Codons /1000**		**pTatR**	**pTat KL16 1st Exon**
		HIV	*E. coli*					HIV	*E. coli*		
PRO	CCA	25.4	8.1	1	2	PHE	UUC	10.6	18.2	2	2
	CCC	11.5	4.2	1			UUU	15.6	18.5		
	CCG	4.6	24.2	5	4	ILE	AUA	29.3	3.8		
	CCU	15.3	6.5		1		AUC	12.8	27.1	2	2
ALA	GCA	28.7	20.6		1		AUU	16.6	27.0		
	GCC	11.8	23.7	1		MET	AUG	21.9	26.5	1	1
	GCG	4.8	33.3	1		TRP	UGG	28.5	12.8	1	1
	GCU	15.3	17.8		1	END	UAA	1.4	2.0		
GLY	GGA	31.6	6.7	1			UAG	1.6	0.2		
	GGC	12.6	30.7	1			UGA	0.6	0.8		

The values given in the table indicate the number of each codon found per 1000 codons in each genome. The authors used 1187 genes for *E. coli*, and 205 genes for HIV, all in Genbank (Wada *et al.*, 1991). The codon composition of *tat* DNA in pTatR and the corresponding 1st exon of *tat* DNA in pTatKL16 are also given for comparison.

Table 3.2 Rare Codons in the pTatKL16 and pTatR *tat* DNA Sequences.

Codon Position in the Gene	pTatKL16 Codons	pTatR Codons	#Codons/1000 in *E. coli**
Pro 3	CCA	CCG	CCC 4.2, CCT 6.5, CCA 8.1, CCG 24.2
Pro 6	CCT	CCG	CCC 4.2, CCT 6.5, CCA 8.1, CCG 24.2
Lys 50	AAG	AAA	AAG 11.9, AAA 36.9
Arg 53	AGA	CGT	AGG 1.3, AGA 2.0, CGA 3.1, CGG 4.6, CGC 22.0, CGT 24.7
Arg 56	AGA	CGT	AGG 1.3, AGA 2.0, CGA 3.1, CGG 4.6, CGC 22.0, CGT 24.7
Pro 58	CCA	CCG	CCC 4.2, CCT 6.5, CCA 8.1, CCG 24.2
Gln 60	CAA	CAG	CAA 13.0, CAG 30.1

*Codon frequencies obtained from a table compiled by Wada *et al.* (1991) from 1187 *E. coli* genes found in Genbank. The values indicate the distribution of degenerate codons for the listed amino acid in the 1187 *E. coli* genes surveyed. The codons in bold typeface are rare codons present in the pTatKL16 *tat* DNA sequence.

3.0.2 Production of Recombinant Plasmids

The response to our communication with Purvis *et al.*, (1995) regarding the exact location of the *tat* DNA in pTatC6H-1 indicated that the corresponding author did not possess this information. Examination of the sequence of the source plasmid pDS56, RB, 6×His (Certa *et al.*, 1986), suggested that *Hin*dIII and *Eco*RI are good choices for restriction enzymes likely to excise the *tat* DNA inserted into the plasmid by Purvis *et al.* (1995). Digestion of pDS56, RB, 6×His with these two enzymes should yield a 60 base pair and a 3356 base pair fragment, whereas digestion of pTatC6H-1 would be expected to yield fragments of a length greater than 258 bp (the size of the *tat* DNA) and 3356 bp (see Table 3.3). To characterize pTatC6H-1 the plasmid was digested with restriction endonucleases *Pvu*II, *Eco*RI, *Hin*dIII, *Hin*dIII and *Eco*RI and, *Pvu*II and *Xho*I and then electrophoresed on a 0.7% agarose gel (not shown). The observed and predicted fragment sizes for the restriction digests are listed in Table 3.3. The results suggest that *tat* DNA is inserted into pDS56, RB, 6×His between *Hin*dIII and *Eco*RI restriction sites.

The putative *tat* DNA was inserted into the pUC19 high copy number plasmid to produce pTatKL1. The region encompassing the DNA insert was then sequenced by UCDNA services at the University of Calgary. Figure 3.1 shows the DNA nucleotide sequence results and the corresponding amino acid sequence for Tat. The protein sequence was compared with the SwissProt protein sequence database using the BLAST program (Altschul *et al.*, 1990). It was determined that the Tat protein sequence is identical to that of the HIV-1 BH10 strain (see Figure 3.1) (Ratner *et al.*, 1985).

Table 3.3 Predicted and Observed Sizes of Restriction Fragments of pTatC6H-1 and pDS56, RB, 6×His.

	***Pvu*II**	***Eco*RI**	***Hin*dIII**	***Hin*dIII & *Eco*RI**	***Pvu*II & *Xho*I**
Predicted pDS56, RB, 6×His size	3416 bp	3416 bp	3416 bp	60 & 3356 bp	412 & 3004 bp
Predicted* pTatC6H-1 size	~3600 bp	~3600 bp	~3600 bp	~260 & 3350 bp	~670 & 2950 bp
Observed pTatC6H-1 size	~3500 bp	~3500 bp	~3500 bp	~400 & ~3100 bp	~600 & ~2800 bp

* The predicted lengths of the fragments are estimates because the size of the DNA inserted into pDS56, RB, 6×His to produce pTatC6H-1 was unknown to us, and could not be provided by the authors. bp = base pairs.

Nucleotide Sequence

.....ACGACGTTGTAAAACGACGGCCAGTGAATTCATTAAAGAGGAGAAATTA

ACT**ATGGAACCAGTCGACCCTAGACTGGAACCGTGGAAACACCCGGGTT**

CCCAGCCGAAAACTGCATGCACCAACTGTTACTGTAAAAAGTGTTGCTTC

CACTGTCAAGTTTGTTTCATCACCAAGGCTTTGGGTATCTCCTACGGTCG

TAAGAAACGTAGACAGCGCAGACGTCCACCGCAAGGTTCTCAGACTCAT

CAAGTTTCCTTGTCCAAGCAACCGACCTCCCAATCTCGCGGTGACCCGA

CAGGTCCTAAGGAAAGATCTCATCACCATCACCATCACTAAGCTTGGCGTA

ATCATGGTCATAGCTNT.....

Amino Acid Sequence

MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKALGISYGRKKRRQ

RRRPPQGSQTHQVSLSKQPTSQSRGDPTGPKE

Figure 3.1 The Nucleotide and Amino Acid Sequences of HIV-1 Tat in pTatKL1. The *tat* DNA sequence is in bold typeface. The amino acid sequence of HIV-1 BH10 Tat protein from the SwissProt database is identical to the Tat amino acid sequence from pTatKL1.

PCR primers were used to amplify the DNA from pTatKL1 as described in Methods. This amplified DNA was inserted into a suitably prepared pET27b(+) expression plasmid to produce pTatKL19, and the insert region was sequenced. Figure 3.2 shows that the nucleotide sequence of the *tat* DNA in pTatKL19 is identical with the *tat* DNA in pTatKL1, indicating that no mutations were introduced into the sequence by the PCR (see Figure 3.2). The *pelB* leader sequence is present in its entirety at the 5' end of the *tat* DNA and the codons are in frame with the *tat* DNA.

pTatKL16 was produced in a similar fashion to pTatKL19. The amplified DNA was inserted into a pET28b(+) expression plasmid and the insert region was sequenced. Figure 3.3 shows that the *tat* DNA was inserted in frame with the amino terminal six histidine purification tag as intended, and that the nucleotide sequence is identical to the *tat* DNA contained in pTatKL1.

PCR primers were designed to amplify the second exon from pTatKL1, which was inserted into pTatR to produce pTatKL26 (see Figure 3.4). The sequence shows that the *tat* DNA in pTatKL26 contains the first 186 nucleotides of the *tat* DNA from pTatR and the final 72 nucleotides of the *tat* DNA from pTatKL1.

3.1 Induction of Tat Protein Expression From pET/*tat* Plasmids

3.1.0 Induction of Tat Protein Expression from pTatKL19

Figure 3.5 shows the growth curves for pTatKL19-transformed *E. coli* BL21(DE3)pLysS. The curves show that induction of recombinant Tat has no significant

Nucleotide Sequence

.....CTATTNATTNCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATAT

ACATATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTG

CCCAGCCGGCGATGGCCATGGAACCAGTCGAC.....GGTCCTAAGGAATA

GAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGATCAAACGGGCT

AGCCAGCCAGAACTC.....

Amino Acid Sequence

MKYLLPTAAAGLLLLAAQPAMAMEPVDPRLEPWKHPGSQPKTACTNCYCKKC

CFHCQVCFITKALGISYGRKKRRQRRRPPQGSQTHQVSLSKQPTSQSRGDPTGPK

E

Figure 3.2 The Nucleotide and Amino Acid Sequences of HIV-1 Tat in pTatKL19. The *pelB* leader sequence is underlined. The *tat* DNA sequence is in bold typeface, and identical to that shown in Figure 3.1.

Nucleotide Sequence

.....NCGGANACATCCCTCTAGAAATTAATTTTGTTTAACTTTAAGAAGGAGATA

TACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCG

CGCGGCAGCCAT**ATGGAACCAGTCGAC.....GGTCCTAAGGAA**TAGAATT

CGAGCTCCGTCNACNAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCA

CTN....

Amino Acid Sequence

MGSSHHHHHHSSGLVPRGSHMEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFH

CQVCFITKALGISYGRKKRRQRRRPPQGSQTHQVSLSKQPTSQSRGDPTGPKE

Figure 3.3 The Nucleotide and Amino Acid Sequences of HIV-1 Tat in pTatKL16. The six-histidine tag sequence is underlined. The *tat* DNA is in bold type face and identical to that shown in Figure 3.1.

Nucleotide Sequence

.....GATTTGATTGANACACTTCNCCTCTAGAAATAATTTTGTTTAACTTTAAG

AAGGAGATATACCatgggcagcagccatcatcatcatcacagcagcggcctggtgccgcgcggcagccatA

TGGAACCGGTCGACCCGCGTCTGGAACCATGGAAACACCCCGGGTCCCA

GCCGAAAACCGCGTGCACCAACTGCTACTGCAAAAAATGCTGCTTCCAC

TGCCAGGTTTGCTTCATCACCAAAGCCCTAGGTATCTCTTACGGCCGTAA

AAAACGTCGTCAGCGACGTCGTCCGCCGCAGGGATCCCAGACTCATCAA

GTTTCCTTGTCCAAGCAACCGACCTCCCAATCTCGCGGTGACCCGACAGGTCC

TAAGGAATAGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCAC

CACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTG

AG.....

Amino Acid Sequence

MGSSHHHHHHSSGLVPRGSH**MEPVDPRLEPWKHPGSQPKTACTNCYCKKCC**

FHCQVCFITKALGISYGRKKRRQRRRPPQGSQTHQVSLSKQPTSQSRGDPTGP

KE

Figure 3.4 The Nucleotide and Amino Acid Sequences of HIV-1 Tat in pTatKL26. The six-histidine tag sequence is shown in lower case. The first exon fragment from pTatR is in bold typeface, and the fragment containing the second exon and part of the first exon from pTatKL1 is underlined.

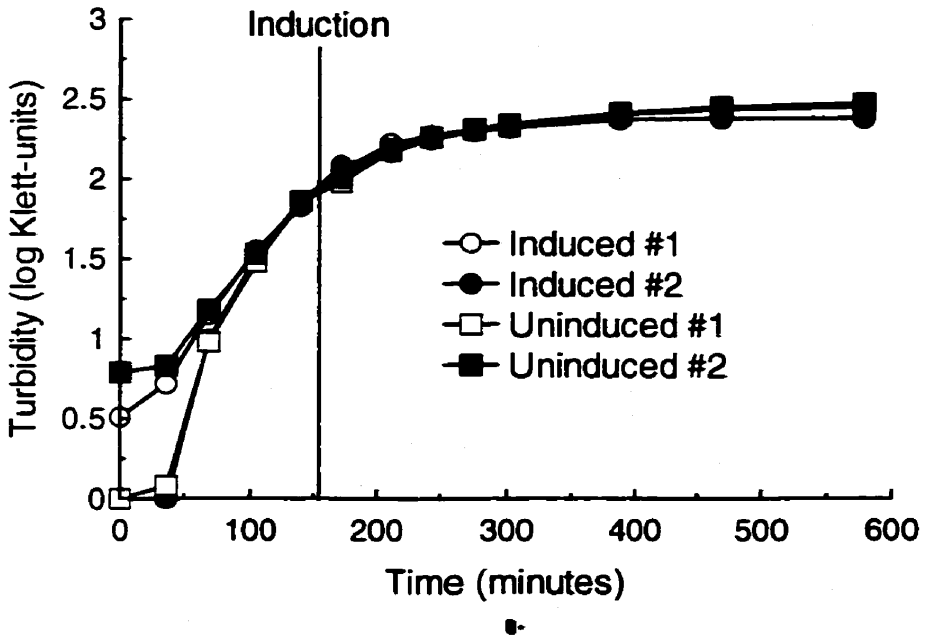

Figure 3.5 Graph of the logarithm of growth measured as turbidity of *E. coli* BL21(DE3)pLysS/pTatKL19 cultured in LB medium. The cells were cultured in LB medium containing chloramphenicol and kanamycin at 37°C with shaking. All cultures were inoculated from a single overnight culture.

detrimental effect on the growth or viability of the cells. Figure 3.6 is a Coomassie-stained electrophoregram of solubilised cells. Lanes 2 and 4 contain uninduced controls sampled 2.25 and 4.5 hours after induction respectively, while lanes 3 and 5 contain the corresponding induced samples. Comparison of these samples shows no obvious protein bands present in the induced lane that are not present in the control sample indicating that any Tat protein expression achieved is very small. Lane 6 contains a sample of lyophilized culture medium from the uninduced cells equivalent to 470 μL of the original medium while lane 7 contains the corresponding induced sample. Comparison of these samples does not show any putative Tat protein bands. This indicates that Tat protein is not being secreted into the medium. This shows that the PelB leader system is not expressing Tat protein in amounts suitable for purification on the scale necessary to produce an NMR sample. Lanes 8 and 9 contain purified His_6-Tat_{1-72} provided by Dr. G. Henry. It should be noted that purified His_6-Tat_{1-72} electrophoreses as two bands with apparent M_r of 14,250 and 26,500 as indicated in lane 8. Both protein bands are recombinant Tat as indicated by reactivity in a Western immuno-blot (not shown). Tat is known to migrate anomalously slowly using SDS-PAGE (Wright *et al.*, 1986) and the apparent M_r determined for His_6-Tat_{1-72} suggests that the protein electrophoreses as a monomer protein and as a dimer.

3.1.1 Induction of Tat Protein Expression from pTatKL16

Figures 3.7 and 3.8 show growth curves in LB and TB medium respectively, for BL21(DE3)pLysS cells harbouring the pTatKL16 plasmid. Induction of Tat protein expression had no observable effect on the growth or viability of the cells in either case.

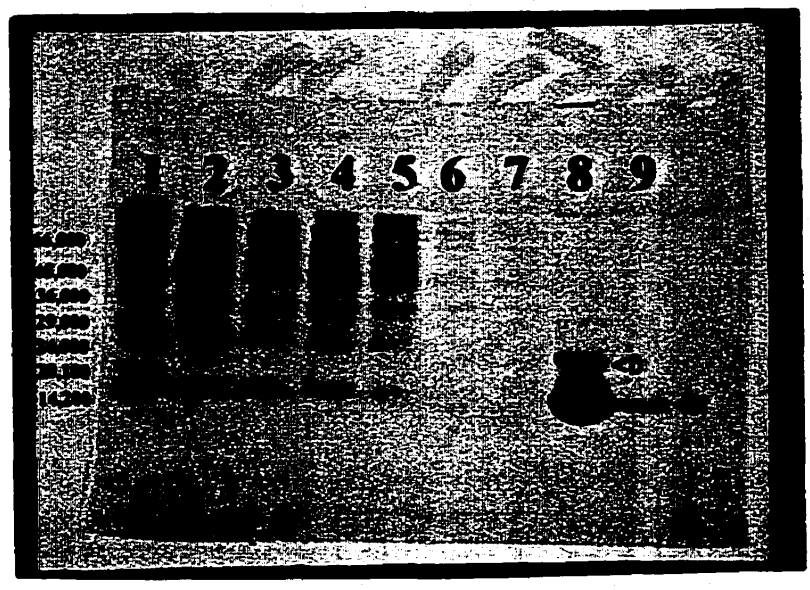

Figure 3.6 SDS-PAGE analysis of *E. coli* BL21(DE3)pLysS/pTatKL19 whole cells. Lane 1: Dalton Mk VII protein M_r markers. Lanes 2 through 5 contain 48 µL samples of the cell cultures. Lanes 2 and 4 contain uninduced cells, while lanes 3 and 5 show induced cells. Lanes 2 and 3 were sampled 2.25 hours after induction, while lanes 4 and 5 were sampled 4.5 hours after induction. Lane 6: 470 µL lyophilized medium from the uninduced culture, 4.5 hours post induction, Lane 7: 470 µL lyophilized medium from the induced culture, 4.5 hours post induction, Lane 8: approximately 27.4 µg of purified Tat_{1-72} protein standard, Lane 9: approximately 1.7 µg of purified Tat_{1-72} protein standard. The arrows indicate the Tat monomer and dimer bands in lanes 8 and 9.

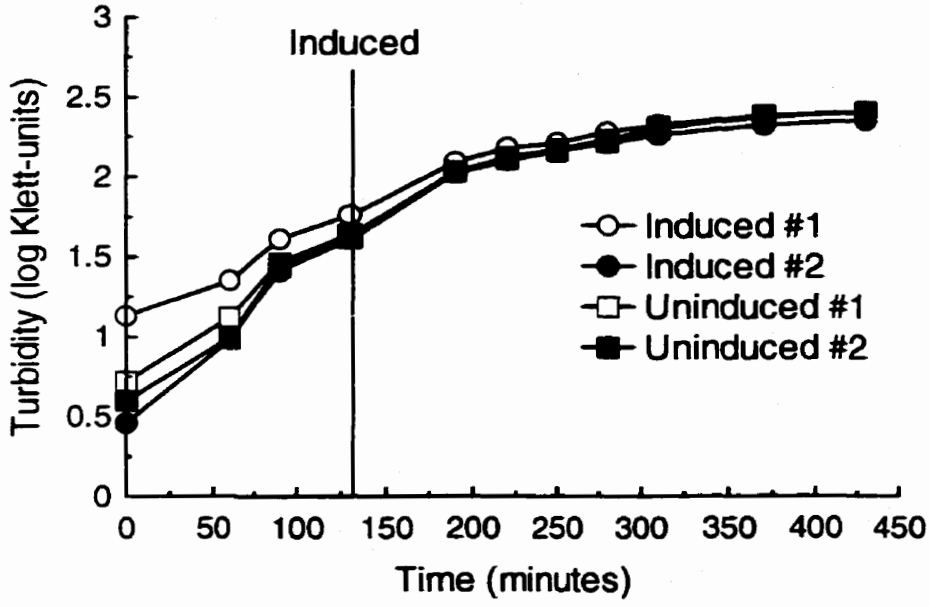

Figure 3.7 Graph of the logarithm of growth measured as turbidity of *E. coli* BL21(DE3)pLysS/pTatKL16 cultured in LB medium. The medium contained chloramphenicol and kanamycin and the cells were grown at 37°C with shaking. All cultures were inoculated from a single overnight culture.

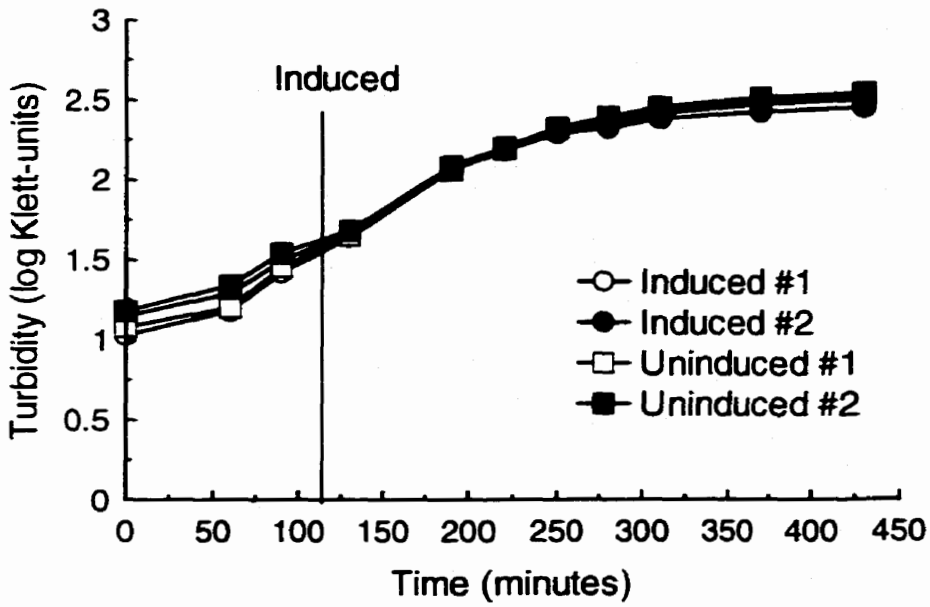

Figure 3.8 Graph of the logarithm of growth measured as turbidity of *E. coli* BL21(DE3)pLysS/pTatKL16 in TB medium. The medium contained chloramphenicol and kanamycin and the cells were incubated at 37°C with shaking. All cultures were inoculated from a single overnight culture.

The cultures grown in TB medium reached a slightly greater cell density at the end of their growth curves. This small increase in cell growth was not judged to be sufficient to justify the use of TB medium in place of LB medium.

To determine the optimal period for harvesting recombinant Tat protein, a time course of expression was performed. Whole cell samples were collected at various times after induction and electrophoresed. Figure 3.9 shows an electrophoregram of cells at 0, 1, 2, 3, 4, and 5 hours after induction of Tat expression. The figure shows a steady increase in cell protein over the course of the experiment. A putative Tat band migrating at an apparent M_r of 17,900 appears to increase at a slightly greater rate than most other protein bands. This band was determined to be recombinant Tat protein by Western immuno-blotting.

To determine if the observation of variable levels of Tat expression from pTatKL16 was due to different *E. coli* clones yielding different levels of Tat, eight colonies were selected, grown in LB medium, and analysed by SDS-PAGE. Figure 3.10 shows two Coomassie-stained electrophoregrams of uninduced and induced whole cell samples. Some cells (e.g. lane 13) appear to produce very little Tat protein, whereas other cells (lane 17) appear to have more intense Tat bands.

Figure 3.11 compares the levels of Tat production by BL21(DE3)pLysS cells containing pTatKL16 (lanes 2-7) and pTatR (lanes 8 and 9). It must be noted that twice the volume of pTatKL16-containing cells were electrophoresed compared to pTatR-cells.

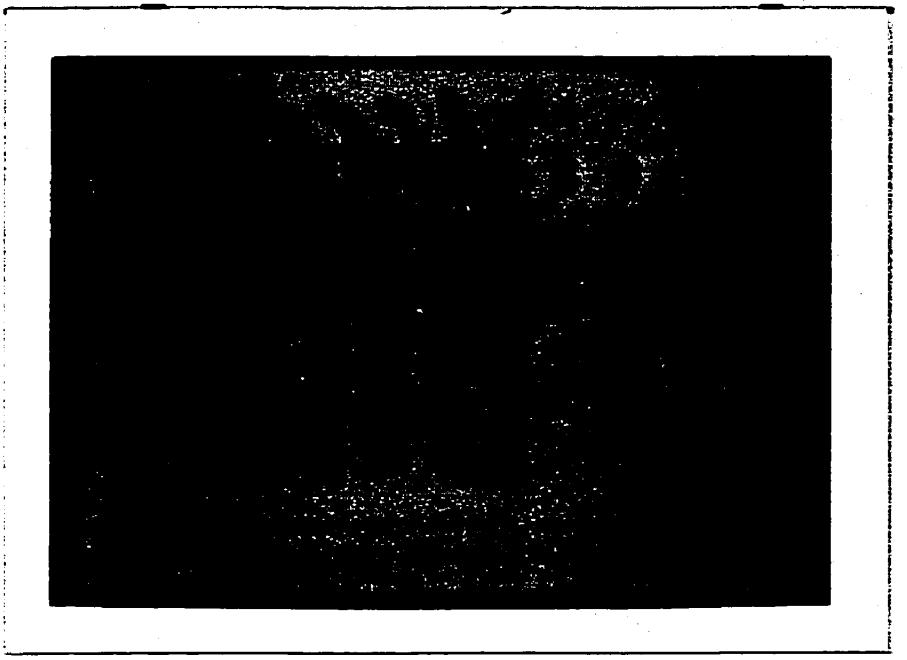

Figure 3.9 SDS-PAGE analysis of 25μL of BL21(DE3)pLysS/pTatKL16 whole cell samples at various times after induction of Tat protein expression. Lane 1: Dalton Mk VII protein M_r markers, Lane 2: pre-induction, Lane 3: 1 hour post-induction, Lane 4: 2 hours, Lane 5: 3 hours, Lane 6: 4 hours, Lane 7: 5 hours. Lane 9: Dalton Mk VII protein M_r markers. The arrow in lane 7 indicates the 86 amino acid Tat protein monomer.

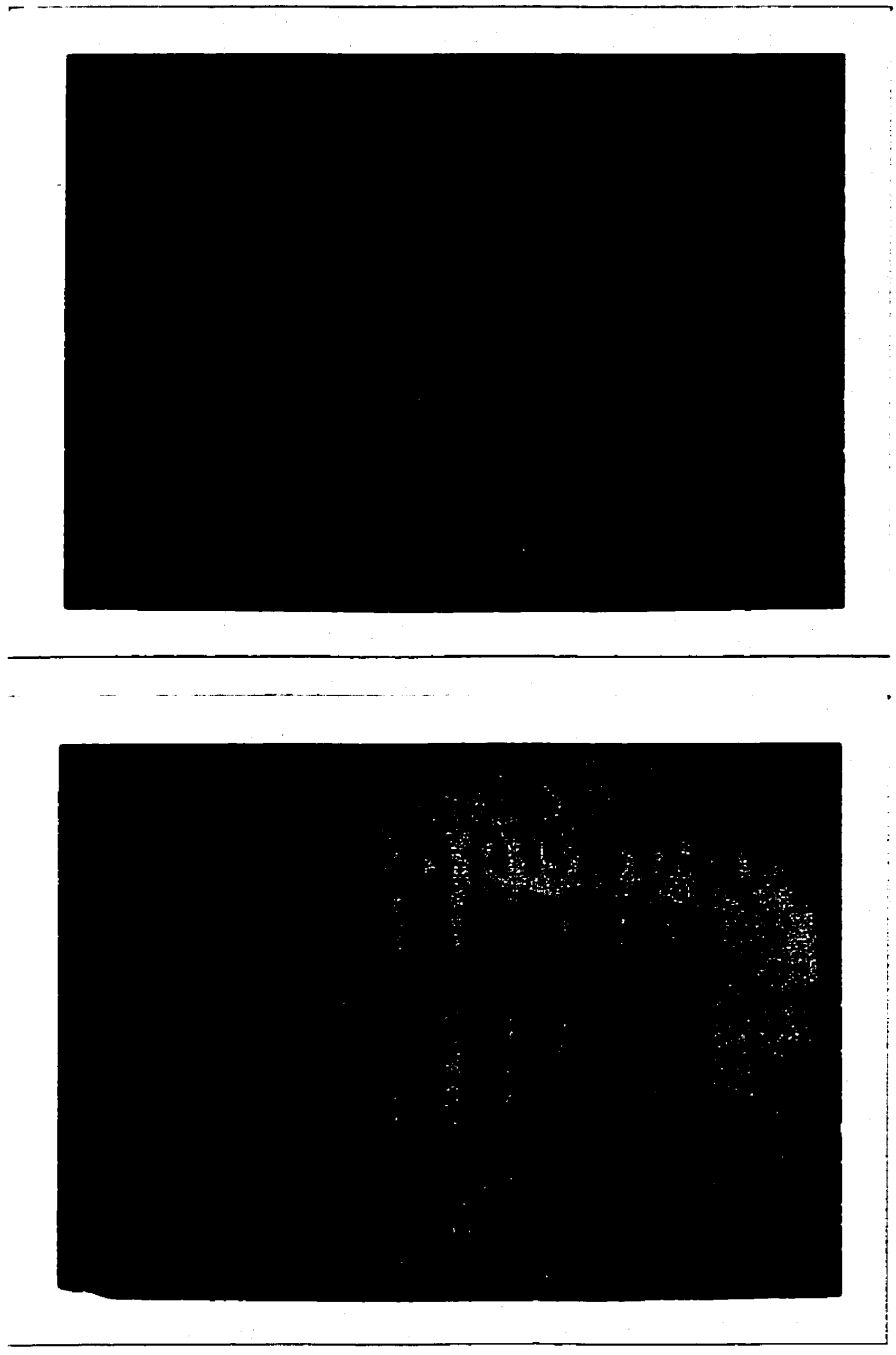

Figure 3.10 Samples of induced and uninduced *E. coli* BL21(DE3)pLysS/pTatKL16 cells. Lanes 1 & 10: Dalton Mk VII protein M_r markers, Lanes 2, 4, 6, 8, 12, 14, 16, and 18 contain 25 µL of uninduced cells, whereas lanes 3, 5, 7, 9, 11,13, 15, and 17 contain 25 µL of induced cells. The arrows in lanes 3 and 11 indicate the Tat protein monomer band.

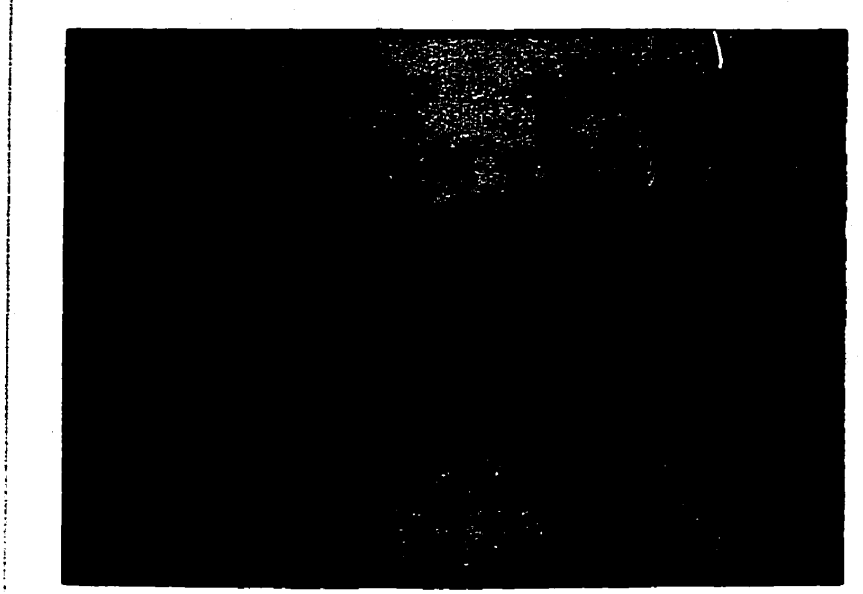

Figure 3.11 Comparison of Tat production by pTatKL16-and pTatR-transformed *E. coli* by SDS-PAGE. Lanes 1 & 10: Dalton Mk VII protein M_r markers, Lanes 2, 4 and 6 contain 25 μL of uninduced pTatKL16 cells from three cultures. Lanes 3, 5, and 7 contain 25 μL of the corresponding induced pTatKL16 cells. Lanes 8 and 9 contain 12.5 μL of uninduced and induced pTatR cells respectively. The arrow in lane 3 indicates the 86 amino acid Tat monomer and the arrow in lane 9 indicates the 72 amino acid Tat monomer identified on the basis of Western analysis (see Figure 3.12). Note that, as expected, His_6-Tat_{1-72} (lane9) electrophoreses more rapidly than Tat_{1-86} (lane3).

Despite this difference, the intensity of the Tat band in pTatR (lane 9) indicates that a larger proportion of cell protein is Tat than in the pTatKL16 cells (lanes 3, 5, and 7). The high level of Tat_{1-72} production by pTatR is confirmed by a Western analysis of a polyacrylamide gel containing the same samples as Figure 3.11 as the signal intensity for Tat_{1-72} (lane 8; Figure 3.12) is far greater than for any of the Tat_{1-86} samples (lanes 3, 5 and 7; Figure 3.12) Lanes 3, and 5 show the presence of a dimeric Tat_{1-86} migrating with an apparent M_r of 30,565. Note that there is a faint band migrating faster than the monomer band indicating a possible degradation product of the monomer in lane 3. The fact that this putative degradation product reacts with an antibody to the six-histidine tag at the amino terminus of the protein implies that degradation is most likely occurring at the carboxyl terminus of the protein. Note also that the Tat_{1-72} protein exhibits multiple bands corresponding to oligomers of the protein (dimer, trimer, tetramer, pentamer, hexamer).

In order to investigate whether the variable level of Tat protein production from pTatKL16 in *E. coli* BL21(DE3)/pLysS (Figure 3.10) was due to some clones harbouring greater numbers of plasmid than others, cells were grown in increasing concentrations of kanamycin to explore the possibility that higher antibiotic concentrations would favour cells harbouring multiple plasmids and that these clones would produce more Tat per cell. Figure 3.13 is a Coomassie-stained electrophoregram of whole cell samples from cultures containing increasing amounts of kanamycin. From left to right on the gel the concentrations are 1 times (30 μg/mL), 1.5 times, 2 times, 2.5 times, and 3 times the standard concentration. No noticeable difference in the intensity of the Tat monomer

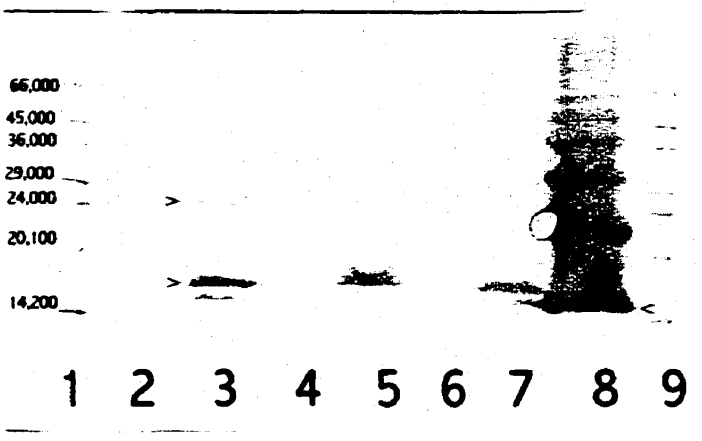

Figure 3.12 Comparison of Tat production by pTatKL16-and pTatR-transformed *E. coli* by Western immuno-blot analysis. Lanes 1 & 9: positions of Dalton Mk VII protein M_r markers indicated in pencil, Lanes 2, 4 and 6 contain 75 µL of uninduced pTatKL16 cells from three cultures. Lanes 3, 5, and 7 contain 75 µL of the corresponding induced pTatKL16 cells. Lane 8: 75 µL of induced pTatR cells. The primary antibody was mouse Monoclonal Anti-poly Histidine clone HIS-1 from Sigma. The secondary antibody was Fc specific anti-mouse IgG horseradish peroxidase conjugate.

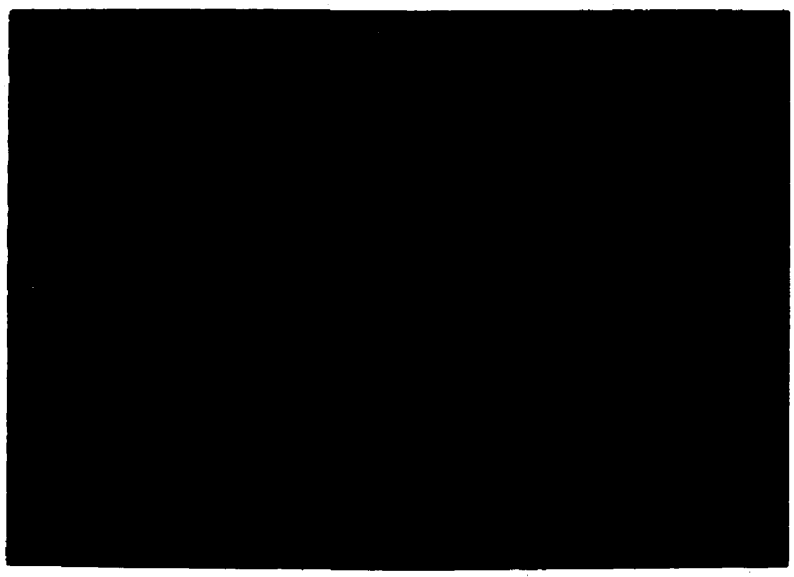

Figure 3.13 Effect of kanamycin on Tat production by BL21(DE3)pLysS/pTatKL16 by SDS-PAGE. The kanamycin concentration increases from left to right across the gel. The uninduced samples were loaded in the odd numbered lanes, with the corresponding sample from the induced culture in the subsequent even numbered lane. All samples were 25 μL of cell culture. Lanes 1 and 2 contained the standard concentration of kanamycin (30 μg/mL). Lanes 3 and 4: 1.5 times concentration, Lanes 5 and 6: 2 times concentration, Lanes 7 and 8: 2.5 times concentration, Lanes 9 and 10: 3 times concentration. The arrow in lane 10 indicates the Tat protein monomer.

band is observable between any of the induced samples. Figure 3.14 is a Western immuno-blot of a polyacrylamide gel containing identical samples to Figure 3.13. This figure suggests a slight decrease in Tat production at higher concentrations of kanamycin. Thus, increased concentrations of kanamycin do not produce an increase in Tat protein production.

3.1.2 Induction of pTatKL26

The low level of Tat protein production from pTatKL16, and the lack of success in determining why expression levels are low justified production of pTatKL26. pTatKL26 was constructed by inserting the second exon and part of the first exon from the *tat* DNA contained in pTatKL16 into the pTatR plasmid containing most of the first exon of *tat* DNA. Figure 3.15 shows the growth curves for *E. coli* BL21(DE3)pLysS/pTatKL26. Induction of protein expression in these cells did not result in a decrease in the optical density of the induced cultures in comparison with the uninduced control cultures indicating that expression of Tat protein did not affect the growth of the cells significantly. Comparison of the growth curves of induced cultures of pTatKL16 (Figure 3.7) and pTatKL26 (Figure 3.15) transformed *E. coli* shows no difference in cell growth, indicating that the pTatKL26-transformed cells do not reach a greater cell density than the pTatKL16-transformed cells.

Figure 3.16 is a Coomassie-stained electrophoregram containing induced (odd numbered lanes) and uninduced (even numbered lanes) whole cell samples from cultures of *E. coli* BL21(DE3)pLysS transformed with pTatKL16 (lanes 2 & 3), pTatKL26 (lanes

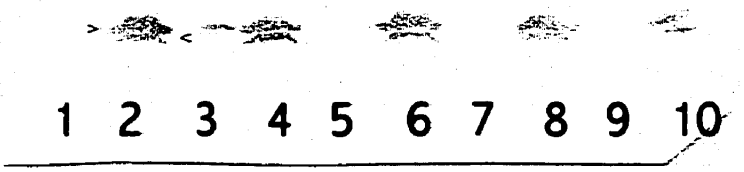

Figure 3.14 Effect of kanamycin on Tat production by BL21(DE3)pLysS/pTatKL16 by Western immuno-blot analysis. The kanamycin concentration increases from left to right across the gel. The uninduced samples were loaded in the odd numbered lanes, with the corresponding sample from the induced culture in the subsequent even numbered lane. All samples were 75 µL of cell culture. Lanes 1 and 2 were in the standard concentration of kanamycin (30 µg/mL). Lanes 3 and 4: 1.5 times concentration, Lanes 5 and 6: 2 times concentration, Lanes 7 and 8: 2.5 times concentration, Lanes 9 and 10: 3 times concentration. The arrows in lane 10 indicate the Tat protein monomer and a faster migrating putative degradation product. The primary antibody was mouse Monoclonal Anti-poly Histidine clone HIS-1 from Sigma. The secondary antibody was Fc specific anti-mouse IgG horseradish peroxidase conjugate.

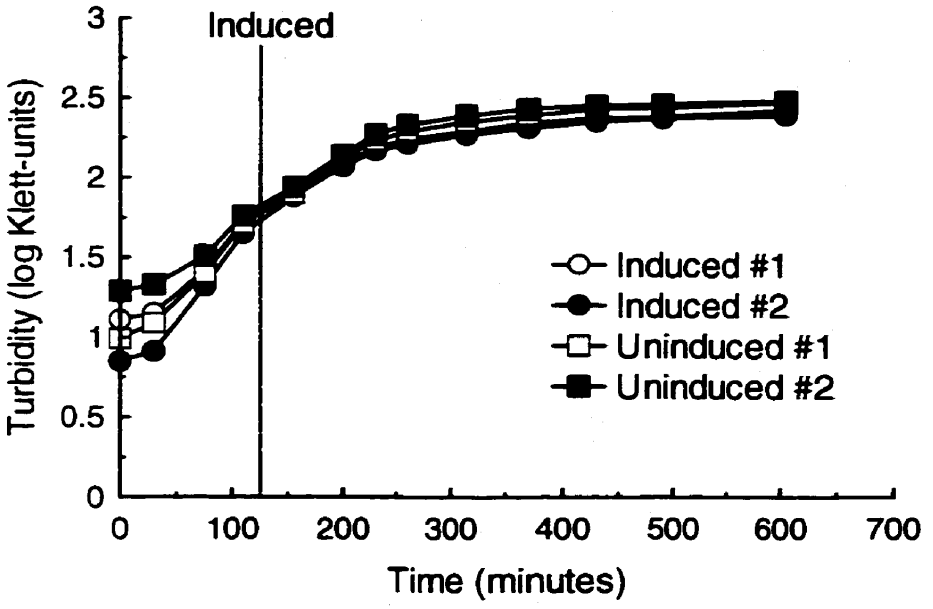

Figure 3.15 Graph of the logarithm of growth measured as turbidity of *E. coli* BL21(DE3)pLysS/pTatKL26. The cultures were induced with 1 mM IPTG in LB medium containing chloramphenicol and kanamycin incubated at 37°C. All cultures were inoculated from a single overnight culture.

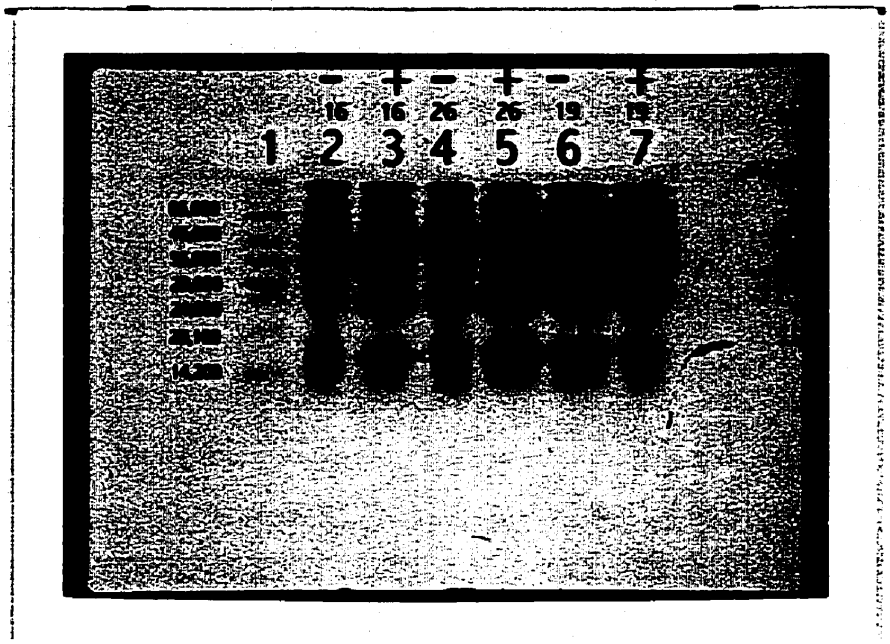

Figure 3.16 SDS-PAGE analysis of Tat protein production in *E. coli* harbouring pTatKL16, pTatKL26, or pTatKL19. Lane 1: Dalton Mk VII protein M_r markers. All cell samples were 24 μL of cell culture. Cells were sampled from pTatKL16 and pTatKL26 cultures at 4 hours after induction, while the pTatKL19 cells were sampled at 5 hours after induction. Lane 2: uninduced pTatKL16 cells, Lane 3: induced pTatKL16 cells, Lane 4: uninduced pTatKL26 cells, Lane 5: induced pTatKL26 cells, Lane 6: uninduced pTatKL19 cells, Lane 7: induced pTatKL19 cells. Arrows mark 6 × His tagged Tat protein monomer visible in lanes 3 and 5.

4 & 5) and pTatKL19 (lanes 6 & 7). Comparison of lanes 6 and 7 show no bands in the induced lane (lane 7) that are not present in the uninduced sample (lane 6), indicating that Tat protein is not detectable in the pTatKL19 cells as shown earlier in Figure 3.6. Comparison of the Tat protein monomer bands in lanes 3 and 5 suggests that pTatKL26-transformed cells (lane 5) produce significantly more Tat monomer than pTatKL16-transformed cells (lane 3). To confirm the difference in Tat protein production between the pTatKL16- and pTatKL26-transformed cells, serial dilutions of cell samples from induced cultures were electrophoresed using SDS-PAGE, and then analysed by Western immuno-blotting. Figure 3.17 shows the Coomassie-stained electrophoregram of the induced whole cell samples. The electrophoregram clearly shows higher Tat production from cells carrying pTatKL26 (compare lanes 2 and 3 to lanes 6 and 7). Figure 3.18 shows the Western immuno-blot of a duplicate gel identical to the one in Figure 3.17. Comparison of lane 2 containing the largest amount of pTatKL16 sample with the corresponding pTatKL26 sample in lane 6 shows that Tat protein production from the pTatKL26 plasmid is much greater than the production of protein from pTatKL16. Comparison of lane 5 and lane 9 which contain the smallest amount of sample for each of the transformed cell lines shows that strong signal remains for the pTatKL26-transformed cells, whereas Tat protein has been diluted to an unmeasurable level in the case of the pTatKL16-transformed whole cell sample. Oligomerization of Tat_{1-72} was noted earlier (Figure 3.12). Figure 3.18 indicates that Tat_{1-86} can also form a number of high apparent M_r oligomeric species.

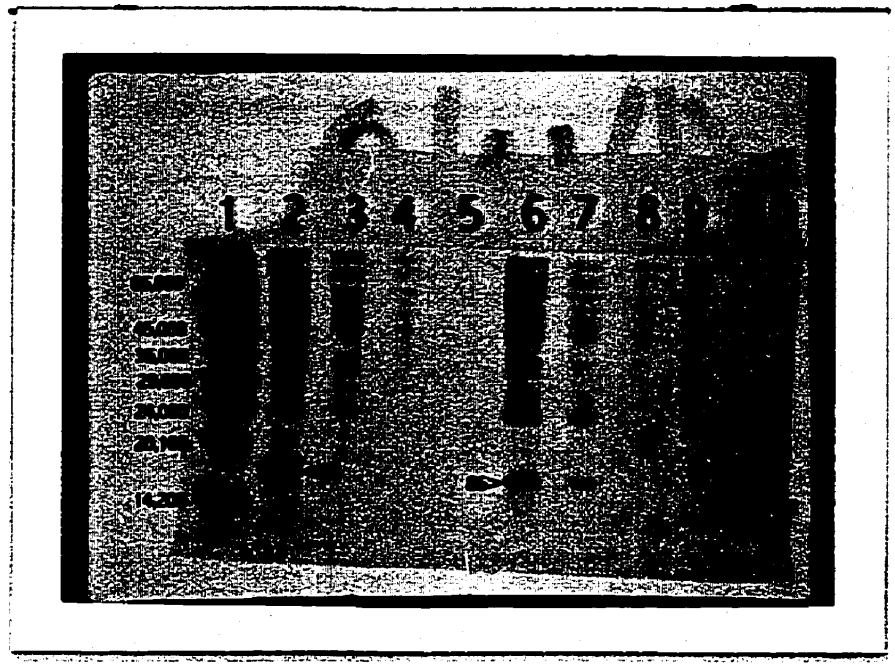

Figure 3.17 SDS-PAGE analysis of serial dilutions of induced cells transformed with pTatKL16 or pTatKL26. Lane 1: Dalton Mk VII protein M_r markers. Lanes 2 and 6 were loaded with 4 μL of cells, lanes 3 and 7 with 1.6 μL, lanes 4 and 8 with 0.8 μL, and lanes 5 and 9 with 0.4 μL. Lanes 2 through 5 contain pTatKL16 cells, and lanes 6 through 9 contain pTatKL26 cells. Lane 10: approximately 1.7 μg of purified Tat_{1-72}. The number 72 in lane 10 marks the 72 amino acid Tat protein monomer. In lanes 2 and 6, the 86 amino acid Tat protein monomer is indicated accordingly.

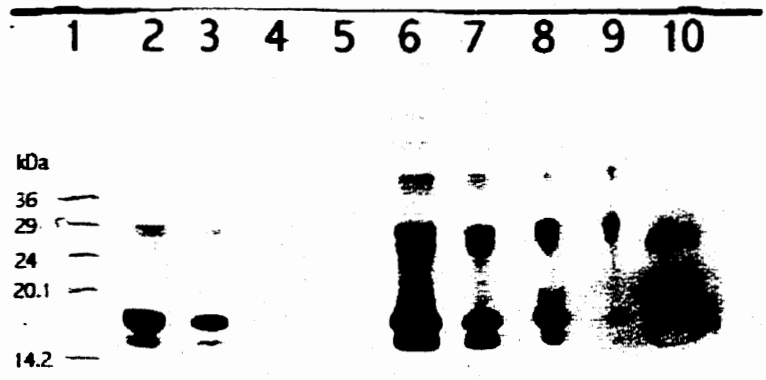

Figure 3.18 Western immuno-blot analysis of serial dilutions of induced cells transformed with pTatKL16 or pTatKL26. Lane 1: positions of Dalton Mk VII protein M_r markers indicated in pencil. Lanes 2 and 6 were loaded with 4 μL of cells, lanes 3 and 7 with 1.6 μL, lanes 4 and 8 with 0.8 μL, and lanes 5 and 9 with 0.4 μL. Lanes 2 through 5 contain pTatKL16 cells, and lanes 6 through 9 contain pTatKL26 cells. Lane 10: approximately 1.7 μg of purified Tat_{1-72}. The primary antibody was mouse Monoclonal Anti-poly Histidine clone HIS-1 from Sigma. The secondary antibody was Fc specific anti-mouse IgG horseradish peroxidase conjugate.

3.2 **Purification of Recombinant Tat Protein**

3.2.0 Identification of Tat Protein Localisation in *E. coli* Lysates

To determine if Tat protein is sequestered in inclusion bodies in the *E. coli* cells, soluble and insoluble fractions of culture samples were prepared and analysed using SDS-PAGE. Figure 3.19 shows the Coomassie-stained gel. The insoluble fraction in lane 5 appears to contain much more protein than the soluble fraction in lane 4. However, a Tat monomer band is distinguishable only in the soluble fraction. Lane 6 contains a sample of the insoluble fraction after 1 wash with 1 % Triton X-100, and shows no significant Tat protein. Lanes 7 and 8 contain a second and third wash respectively and show no protein bands. This gel shows that the major proportion of Tat protein is not sequestered in inclusion bodies. This evidence, coupled with reports in the literature that Tat protein is found in the soluble fraction (Orsini *et al.*, 1996; Kirsch *et al.*, 1996), eliminated inclusion bodies as a source of Tat protein.

3.2.1 Purification of Tat Protein Under Non-denaturing Conditions

Purifying protein under non-denaturing conditions has the potential benefit that in the absence of denaturants, refolding steps after purification should not be necessary. Figure 3.20 is an electrophoregram of samples collected at different stages during the purification of pTatKL16-derived Tat_{1-86} protein under non-denaturing conditions in a reducing environment. Lane 3 shows a prominent Tat monomer band in the induced whole cell sample. Lane 4 shows that the soluble cell extract also contains a prominent Tat protein band, whereas the insoluble fraction shows a significantly fainter Tat monomer band in lane5. Lane 6 shows that there is a prominent Tat monomer band in

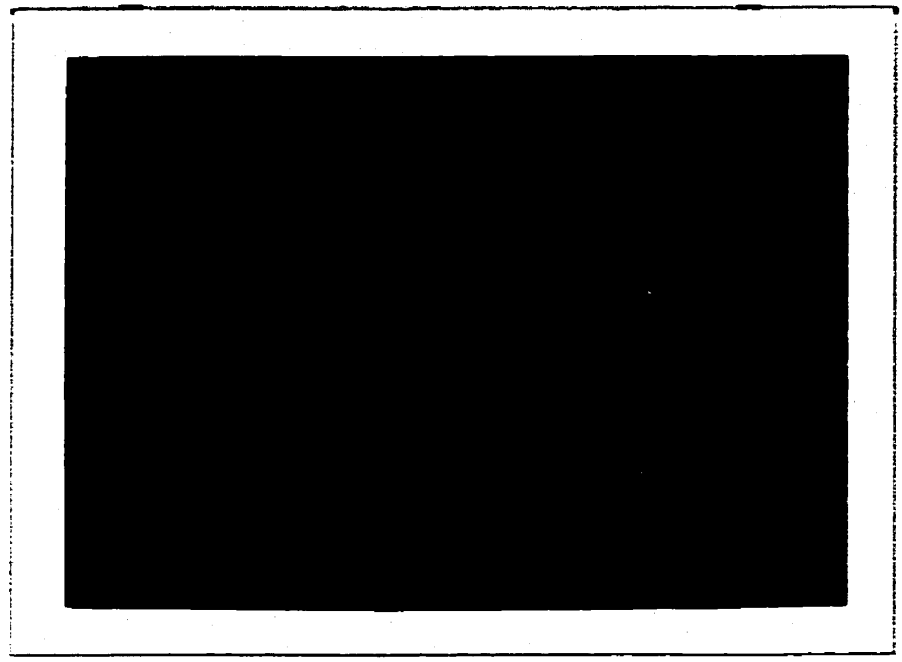

Figure 3.19 SDS-PAGE analysis of soluble and insoluble whole cell fractions of *E. coli* BL21(DE3)pLysS/pTatKL16 induced with IPTG. Lane 1: Dalton Mk VII protein M_r markers, Lane2: 12.5 μL of uninduced cells, Lane 3: 12.5 μL of induced cells, Lane 4: 62.5 μL of the soluble protein fraction, Lane 5: 62.5 μL of the insoluble protein fraction, Lane 6: 187.5 μL of the insoluble fraction after 1 wash in Triton X-100, Lane 7: 187.5 μL of the insoluble fraction after 2 washes, Lane 8: 187.5 μL of the insoluble fraction after 3 washes, Lane 9: 62.5 μL of the insoluble fraction after 1 wash, Lane 10: 125 μL of the insoluble fraction after 1 wash. The arrow in lane 4 shows Tat protein monomer.

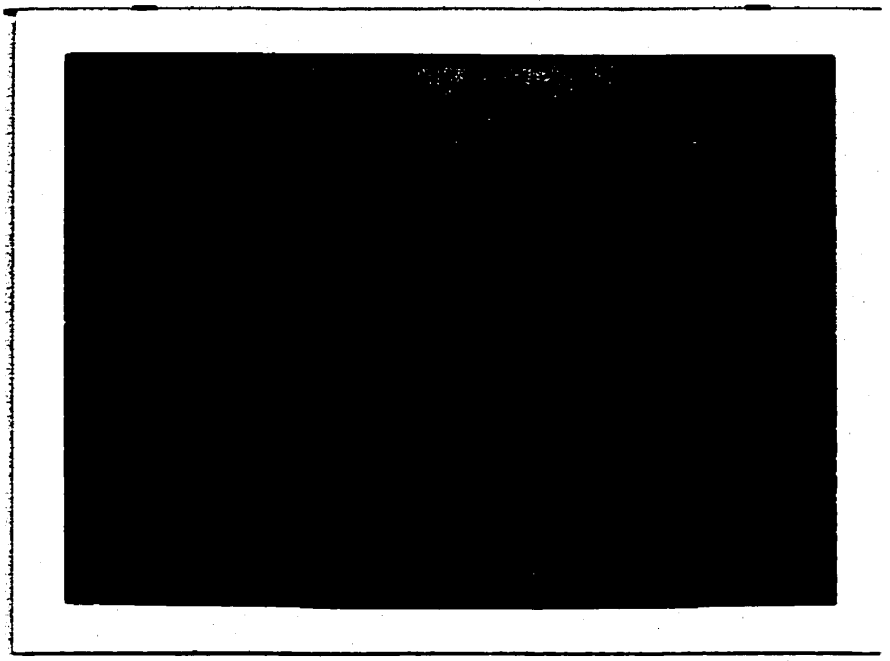

Figure 3.20 SDS-PAGE analysis the purification of Tat protein using non-denaturing conditions. Lane 1: Dalton Mk VII protein M_r markers, Lane 2: uninduced cells, Lane 3: induced cells, Lane 4: the soluble lysate, Lane 5: the insoluble fraction, Lane 6: the unbound material from the column, Lane 7: tube # 2 of 250 mM imidazole elution, Lane 8: tube #3 of the imidazole elution, Lane 9: 100 mM EDTA eluate, Lane10: 6 M urea eluate.

the material that does not bind to the Ni^{2+} column. This indicates that the column has not bound all of the available Tat protein. Lanes 7 and 8 contain samples of the protein eluted from the column using 250 mM imidazole. They show prominent Tat monomer bands as well as higher M_r bands, later shown by Western immuno-blot analysis to be mostly Tat multimers. After elution with imidazole, the column was washed with 100 mM EDTA to determine if all bound Tat protein had eluted. Lane 9 contains a sample of this eluate, which shows no protein bands. Finally, the column was eluted with 6 M urea to remove any protein that had precipitated on the column, and again, the sample in lane 10 shows that no protein was recovered in this elution. Figure 3.21 is a Western immuno-blot of a gel identical to the one in Figure 3.20. Lanes 3 through 8 indicate the presence of Tat monomer, a possible degradation product, as well as many higher M_r multimer species. Lane 7 shows that a very significantly purified Tat protein exists in the elution, as almost all bands on the Coomassie-stained SDS-PAGE gel (Figure 3.20) can be identified as Tat protein from the Western immuno-blot (Figure 3.21). There are potentially some high M_r contaminant bands at the top of the gel, but these cannot be definitively identified as contaminants because the highest M_r proteins on the SDS-PAGE gels did not routinely transfer to the nitrocellulose membrane with high efficiency. Figure 3.22 shows a Coomassie-stained gel with the high M_r proteins that did not transfer to the nitrocellulose.

Tat is a highly basic protein known to bind nucleic acid. Would a purification procedure using non-denaturing conditions remove all contaminating nucleic acid from the protein? Figure 3.23 is an ethidium bromide-stained polyacrylamide gel of samples

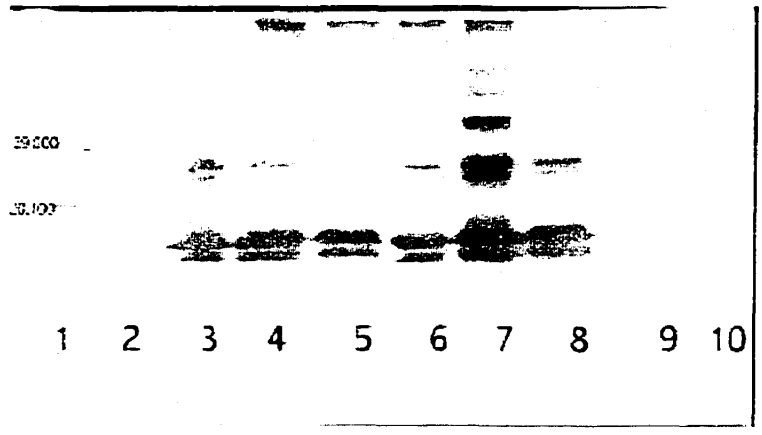

Figure 3.21 Western immuno-blot analysis of a purification of pTatKL16-derived Tat protein under non-denaturing conditions. This is a Western blot of a gel identical to the one shown in Figure 3.20. Lane 1: positions of Dalton Mk VII protein M_r markers indicated in pencil, Lane 2: 8 μL of uninduced whole cells, Lane 3: 8 μL of induced whole cells, Lane 4: the soluble fraction, Lane 5: the insoluble fraction, Lane 6: the material that did not bind the column, Lane 7: tube # 2 of the 250 mM imidazole elution, Lane 8: tube #3 of the imidazole elution, Lane 9: 100 mM EDTA elution, Lane10: 6 M urea elution. The primary antibody was mouse Monoclonal Anti-poly Histidine clone HIS-1 from Sigma. The secondary antibody was Fc specific anti-mouse IgG horseradish peroxidase conjugate.

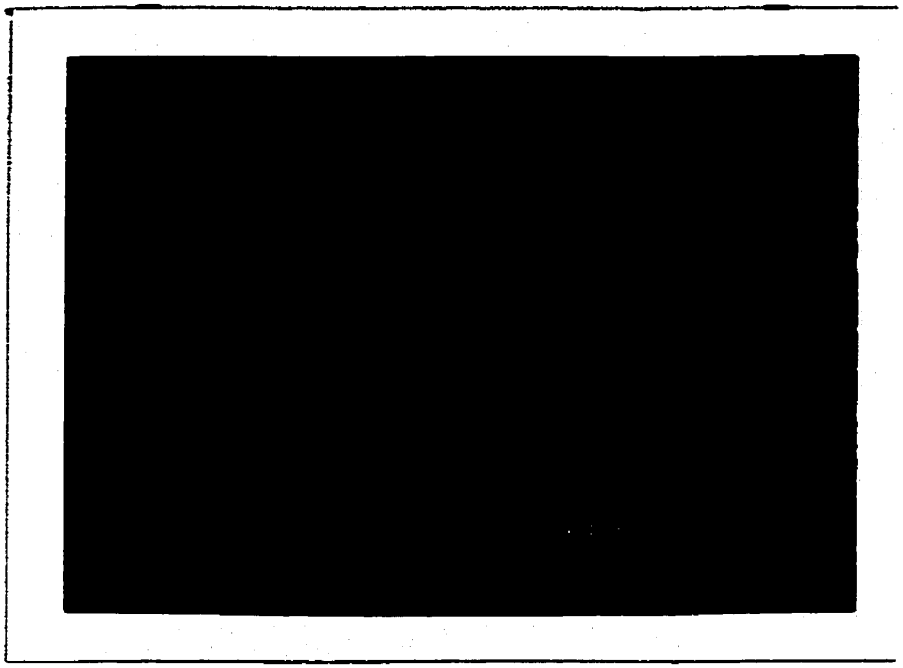

Figure 3.22 Coomassie-stained polyacrylamide gel after electrophoresis onto a nitrocellulose membrane. Lane 1: Dalton Mk VII protein M_r markers, Lanes 2 through7: cell samples of *E. coli* harbouring pTatKL16.

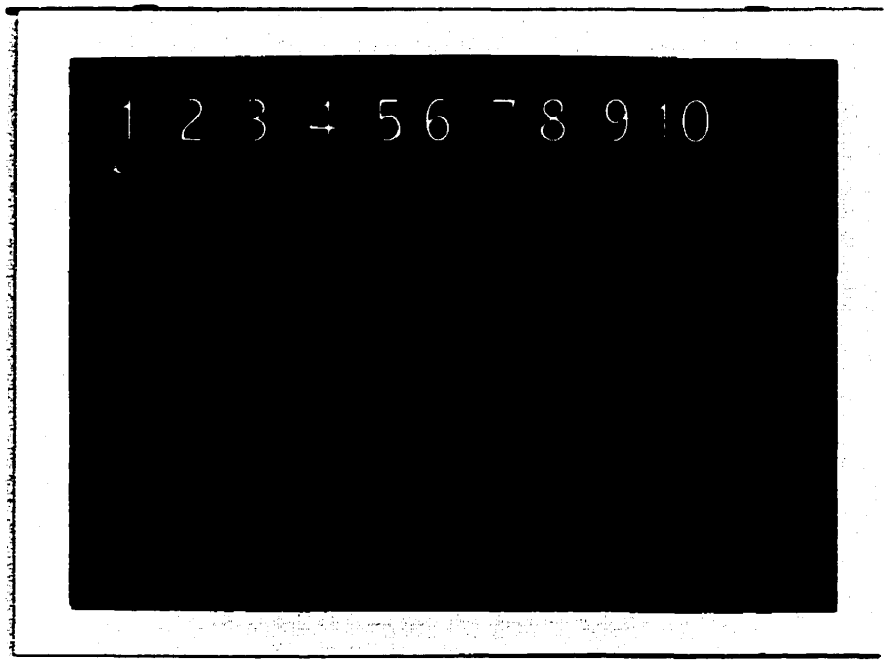

Figure 3.23 Ethidium bromide test for the presence of nucleic acid. Lane 1: soluble cell lysate corresponding to lane 4 of Figure 3.20, Lane 2: Tat protein eluted from column (lane 7 Figure 3.20), Lane 3: 60 minute RNase digestion of soluble cell lysate, Lane 4: 30 minute RNase digestion of soluble cell lysate, Lane 5: 10 minute RNase digestion of soluble cell lysate, Lane 6: 0 minute RNase digestion of soluble cell lysate, Lane 7: 60 minute RNase digestion of eluate, Lane 8: 30 minute RNase digestion of eluate, Lane 9: 10 minute RNase digestion of eluate, Lane 10: 0 minute RNase digestion of eluate.

identical to those in lane 4 (supernatant) and lane 7 (Tat containing eluate) of Figures 3.20 and 3.21. Lanes 2 and 10 show that none of the bands corresponding to Tat bind RNA. At the top of the gel there is a small amount of ethidium bromide-reactive material that may have been bound to the Tat protein prior to electrophoresis. Lane 1 indicates a large amount ethidium bromide-reactive material present in the soluble cell extract. This material was shown to be RNA by incubation with RNase except for the band that remains after incubation (lanes 3-6).

3.2.2 Purification of Tat Protein Under Denaturing Conditions

The decision to use denaturing conditions for purification of Tat was made based on the observation of Tat protein in the insoluble fraction of cells as indicated in lane 5 of Figure 3.21 in an attempt to increase the yields of purification. Figure 3.24 is an electrophoregram of a purification of Tat protein under denaturing conditions. Lane 6 shows the insoluble cell lysate. The amount of protein in this lane is quite small indicating that the denaturant is solubilising most of the protein in the lysate (compare with lane 5 in Figure 3.20). There are no visible protein bands of M_r less than 36,000 indicating that the majority of Tat protein in the lysate was solubilised. This was confirmed in a Western immuno-blot that shows only a trace of Tat in the insoluble fraction (lane 6 Figure 3.25). Lane 4 shows a large number of proteins in the soluble cell fraction and the corresponding lane in Figure 3.25 indicates the presence of Tat protein. Lane 5 in Figure 3.24 shows that most proteins do not bind the Ni^{2+} resin whereas lane 5 in Figure 3.25 shows that very little Tat protein does not bind the resin. That only trace amounts of Tat protein remain in the insoluble cell lysate and the unbound material from

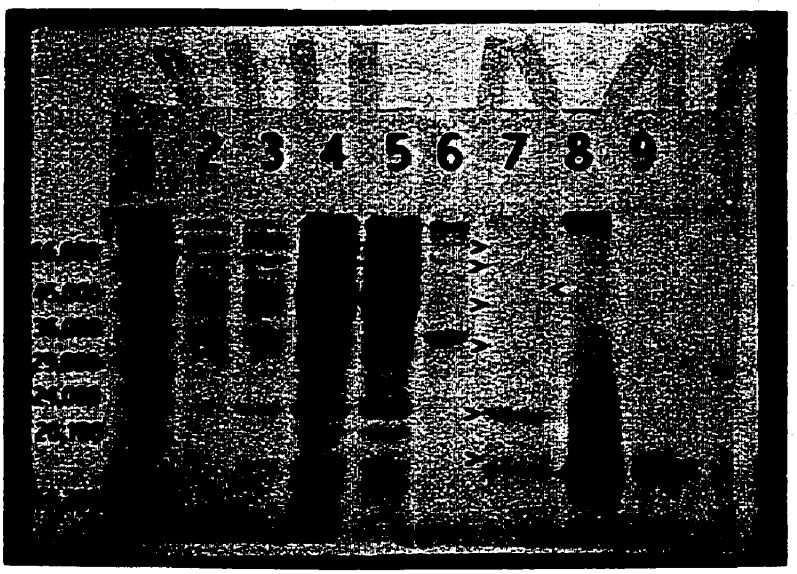

Figure 3.24 SDS-PAGE Analysis of a purification of pTatKL16-derived Tat protein under denaturing conditions. Lane 1: Dalton Mk VII protein M_r markers, Lane 2: uninduced cells, Lane 3: induced cells, Lane 4: soluble lysate, Lane 5: unbound material from the column, Lane 6: insoluble cell fraction, Lane 7: lyophilised Tat_{1-86} protein from 20 tubes eluted using pH 4.5 buffer. Lane 8: approximately 27 µg of purified Tat_{1-72}. Lane 9: approximately 1.7 µg of purified Tat_{1-72}. The arrows in lane 7 mark the protein bands producing signal on the immuno-blot starting at the bottom with the degradation product of the Tat monomer. The arrow in lane 9 marks the Tat_{1-72} monomer band.

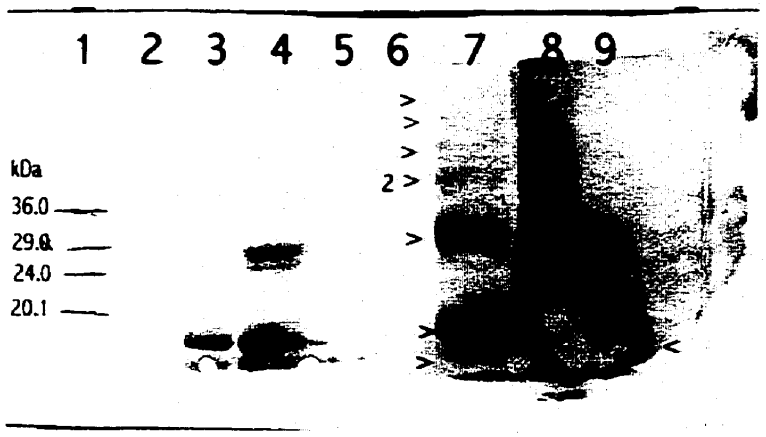

Figure 3.25 Western immuno-blot analysis of a purification under denaturing conditions of pTatKL16-derived Tat protein using Ni^{2+} metal-chelate affinity chromatography. Lane 1: positions of Dalton Mk VII protein M_r markers indicated in pencil, Lane 2: uninduced cells, Lane 3: induced cells, Lane 4: soluble cell lysate, Lane 5: unbound material from the column, Lane 6: insoluble cell fraction, Lane 7: lyophilised Tat_{1-86} protein from 20 tubes eluted using pH 4.5 buffer. Lane 8: approximately 27 µg of purified Tat_{1-72}. Lane 9: approximately 1.7 µg of purified Tat_{1-72}. The arrows in lane 7 mark the bands visible in the SDS-PAGE gel starting at the bottom with the degradation product of the Tat monomer. The arrow marked with a 2 indicates a thick band on the blot that probably encompasses two bands on the SDS-PAGE gel. The arrow in lane 9 marks the Tat_{1-72} monomer band. The primary antibody was mouse Monoclonal Anti-poly Histidine clone HIS-1 from Sigma. The secondary antibody was Fc specific anti-mouse IgG horseradish peroxidase conjugate.

the column shows that the majority of Tat protein is bound to the column. Lane 7 of Figure 3.24 shows the protein eluted from the Ni^{2+} column with Na-acetate/acetic acid pH 4.5. That the major bands correspond to Tat dimer, Tat monomer, and a degradation product, is confirmed by the Western immuno-blot (lane 7 Figure 3.25). Some very faint high M_r aggregates also appear to be recombinant Tat judging from the immuno-blot (compare lane 7 from Figures 3.24 and 3.25). That nearly all the Coomassie-stained protein is immuno-reactive suggests that the Tat protein has been purified to greater than 95 %. Lane 9 in Figure 3.24 contains approximately 1.7 µg of pure 72 amino acid Tat protein. This was used to estimate that approximately 280 µg of total Tat protein was obtained from this purification. The original volume of the culture was 500 mL, so the pTatKL16 plasmid can produce after purification, approximately 560 µg of 95% pure Tat protein per litre of culture. To obtain 0.5 mL of a 1 mM solution of Tat protein needed for NMR analysis would require approximately 6 mg of Tat from 10 to 15 litres of culture.

3.2.3 Purification of Tat_{1-72} Protein Under Denaturing Conditions

To test my purification procedures I purified Tat_{1-72} using the method I developed for Tat_{1-86}. Figures 3.26-3.28 show electrophoregrams from a purification of Tat_{1-72} from pTatR-transformed cells. Lane 3 in Figure 3.26 shows a prominent Tat monomer in cell lysate. Lane 4 shows the soluble cell lysate containing Tat protein and lane 5 shows that some Tat is present in the insoluble fraction. Figure 3.27 shows the protein eluted from the Ni^{2+} column. Lanes 3 through 6 contain samples from tubes 2 through 5 of the Na-acetate/acetic acid pH4.5 elution, and show mostly monomeric Tat_{1-72} with a trace of

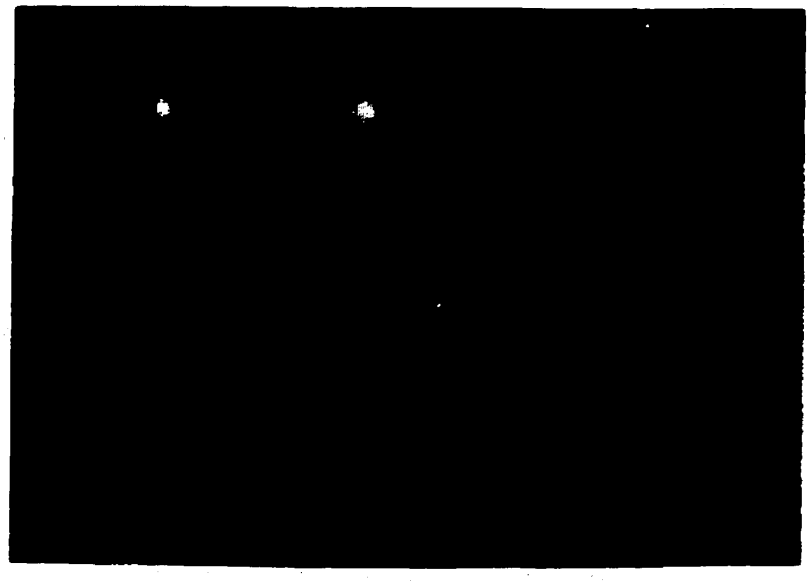

Figure 3.26 SDS-PAGE analysis of a purification of Tat_{1-72} under denaturing conditions. Lane 1: Dalton Mk VII protein M_r markers, Lane 2: uninduced cells, Lane 3: induced cells, Lane 4: the soluble cell lysate, Lane 5: insoluble fraction, Lane 6: the unbound material from the column. The arrow in lane 3 indicates the 72 amino acid Tat monomer.

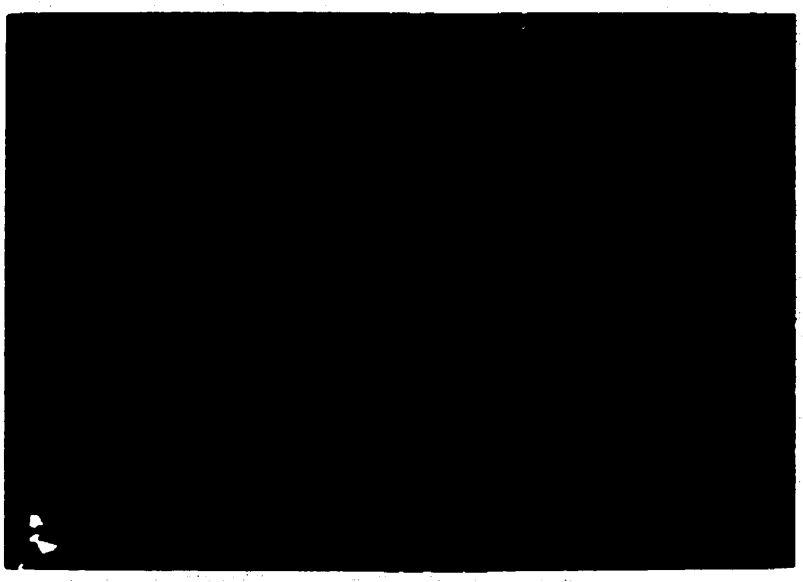

Figure 3.27 SDS-PAGE analysis of Tat_{1-72} purified under denaturing conditions. Lane 1: Dalton Mk VII protein M_r markers, Lane 2: induced cells, Lane 3: tube # 2 of the 50 mM Na-acetate/acetic acid pH 4.5 elution, Lane 4: tube #3, Lane 5: tube #4, Lane 6: tube #5, Lane 7: tube #2 of the 6 M urea, 100 mM EDTA elution, Lane 8: tube #3, Lane 9: tube #4, Lane 10: #5. The arrow in lane 2 indicates the 72 amino acid Tat monomer.

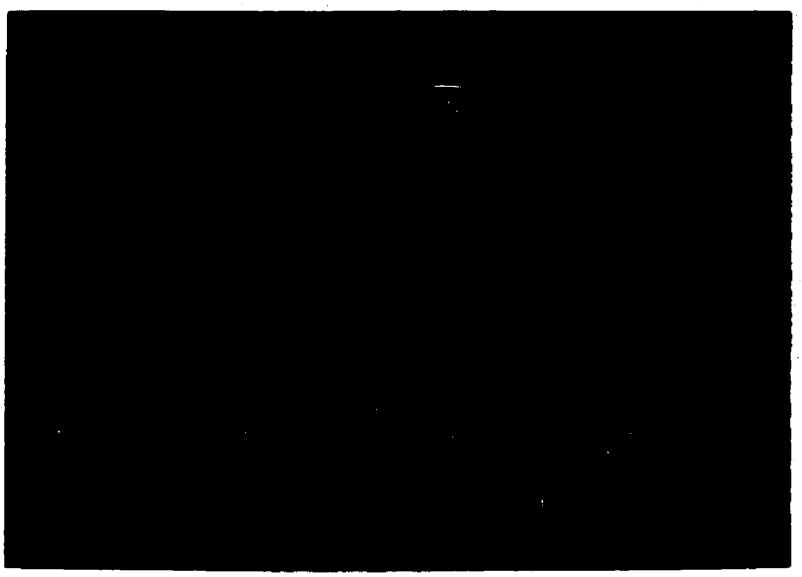

Figure 3.28 SDS-PAGE analysis of Tat_{1-72} purified under denaturing conditions. Lane 1: Dalton Mk VII protein M_r markers, Lane 2: induced cells, Lanes 3-10 contain samples of tubes 6-13, respectively, of the 50 mM Na-acetate/acetic acid pH 4.5 elution. The arrow in lane 2 indicates the Tat_{1-72} monomer.

dimeric protein. Figure 3.28 shows samples of tubes 6 through 13 of the pH 4.5 elution. This gel shows that the amount of Tat protein eluted in each tube is roughly equivalent, and is present over a large elution volume. The samples of the protein eluate also show that the Tat protein has been significantly purified as it appears there are only Tat monomer and dimer bands, and perhaps a small amount of contaminating protein at molecular weights intermediate between these two species. Lanes 7 to 10 of Figure 3.27 show the protein eluted from the column with a 6 M urea, 100 mM EDTA solution. These samples all have prominent Tat monomer bands, some dimer, and traces of higher M_r protein. The successful purification of Tat_{1-72} protein shows that my purification methods are sound. The yield of Tat_{1-72} from this purification is estimated to be, at a minimum, several fold higher than from pTatKL16 expressed Tat_{1-86}.

Comparison of Tat Purification Under Non-denaturing and Denaturing Conditions

Comparison of samples of insoluble cell lysate from denaturing (lane 6 Figure 3.25) and non-denaturing purifications (lane 5 Figure 3.21) shows that while most of the Tat protein binds the column under denaturing conditions, significant amounts of Tat fail to bind under non-denaturing conditions. The insoluble cell lysates from the denaturing (lane 6 Figure 3.25) and non-denaturing purifications (lane 5 Figure 3.21) show that all Tat protein is solubilised under denaturing conditions, while non-denaturing conditions fail to solubilise significant amounts of Tat. Lanes 7 and 8 from Figures 3.20 and 3.21 show the eluate from the non-denaturing purification. Comparison with eluate from a denaturing purification (lane 7 from Figures 3.24 and 3.25) shows that the eluate from both purifications contain prominent bands for Tat monomer, dimer, and a degradation

product, as well as higher M_r Tat multimer species. However, the Western immuno-blot analysis of these two purifications (Figures 3.21 and 3.25) shows that all of the bands corresponding to Coomassie-stained material in the electrophoregram of the denaturing purification contain recombinant protein, whereas the electrophoregram of the non-denaturing purification contains bands for protein contaminants judging from the weakness of the Western immuno-blot.

Chapter Four: Discussion

4.0 Purpose of the Work

Determination of high-resolution structures of HIV-1 Tat, and the Tat-TAR complex as well as Tat-TAR bound to the TAK complex is necessary for a description of the function of Tat in viral transactivation. In addition, a Tat structure at atomic resolution would make structure based design of anti-Tat pharmaceuticals possible. There is perhaps no more attractive target in the HIV genome for pharmaceuticals than the Tat protein. Without Tat, levels of viral transcription would be much lower. One potential problem with the drugs currently in use against HIV infection and AIDS is the ability of the highly mutable HIV virus to evolve resistance to them. Despite the use of these drugs in concert with each other to prevent the HIV virus from escaping their effects, reservoirs of virus in quiescent cells remain even after prolonged treatment (Finzi *et al.*, 1997; Wong *et al.*, 1997).

4.1 Sub-cloning HIV-1 *tat*

Four expression vectors for the production of recombinant Tat were designed and synthesised. Two plasmids were constructed as cloning vehicles. DNA sequencing showed that all but one of the recombinant plasmids was made successfully. Thus, incorporation of restriction endonuclease sites using non-complementary tails on the PCR primers was a fast and useful method of sub-cloning the *tat* DNA. Unfortunately the incorporation of an *Nco*I restriction endonuclease site into one of the PCR primers did not result in DNA that could be cleaved by the enzyme. Despite the lack of *Nco*I digestion at the 5' terminus of the *tat* DNA, the DNA ligated into the pET27b(+) plasmid, presumably

by a blunt end ligation at the 5' terminus. However, this resulted in a 5-nucleotide frameshift mutation and a premature stop codon close to the 5' terminus of the *tat* DNA. An attempt was made to repeat this ligation with a suitably restriction endonuclease-digested pUC19 plasmid. Unfortunately, the blunt end ligation was not reproducible. The solution was to blunt end ligate the amplified *tat* DNA into pUC19 (pUCTat) prior to digestion with *Nco*I. This was the only major problem encountered in attempting to sub-clone *tat* DNA.

4.2 Expression of Recombinant Tat Protein

There are reports in the literature that the tat gene is unstable when expressed in *E. coli* and that the protein is toxic to cells (Ciccarelli *et al.*, 1990; McKenna *et al.*, 1994). This is a major impediment to expression of Tat by *E. coli* and is usually detected as cell growth arrest immediately following induction of recombinant protein synthesis. Growth curves of pTatKL16, pTatKL19, and pTatKL26 transformed *E. coli* showed no such arrest upon induction of Tat expression (see Figures 3.5, 3.7 and 3.15). One possible explanation for the lack of toxicity is the low levels of protein expression of pTatKL16 and pTatKL19. However, we observe no evidence of Tat toxicity even for the higher expressing pTatKL26. Another possibility to be considered is that small differences in the Tat sequences reported in the literature account for the differences in toxicity. For example, McKenna *et al.* (1994) report toxicity of Tat_{1-72}, whereas we expressed Tat_{1-86}.

Interestingly, the levels of Tat expression from the three plasmids varied widely. Protein expression was lowest in the vector (pTatKL19) designed to produce Tat with a

PelB leader sequence for protein secretion. No Tat protein was observed in cell lysates or in lyophilised and dialysed samples of medium from the pTatKL19-transformed cells (see Figure 3.6). The reasons for this are unclear. Higher levels of expression were observed from the two vectors designed to express His_6-Tat_{1-86} (pTatKL16 and pTatKL26). The highest level of expression was obtained for the vector comprised of nucleotides 1-186 of pTatR and nucleotides 187-258 from HIV-1 BH10 *tat* (pTatKL26). The *tat* DNA in pTatR differs from HIV-1 BH10 *tat* at several codons including several at the 5' terminus and some encoding the basic region. The sequence differences might account for higher protein expression by pTatKL26 because that vector contains codons that are more frequently used by *E. coli* as indicated in the following Table:

Table 4.1 Frequency in *E. coli* of Codons in the *tat* DNA Sequence of the pTatKL16 and pTatKL26 Plasmids

Amino Acid		Pro-3	Pro-6	Arg-7	Lys-50	Arg-53	Arg-56	Pro-58	Gln-60
pTatKL26	Codon	CCG	CCG	CGT	AAA	CGT	CGT	CCG	CAG
	E. coli frequency	24.2%	24.2%	24.7%	36.9%	24.7%	24.7%	24.2%	30.1%
pTatKL16	Codon	CCA	CCT	AGA	AAG	AGA	AGA	CCA	CAA
	E. coli frequency	8.1%	6.5%	2.0%	11.9%	2.0%	2.0%	8.1%	13.0%

In the future, when confronted with chronically low levels of expression that cannot be easily explained by other factors, an attempt should be made to remove rare codon sequences, especially those at the beginning of the coding sequence of interest.

4.3 Purification of HIV-1 Tat Protein

A review of the literature indicates that Tat is highly susceptible to oxidation and aggregation (Waszak *et al.*, 1992; Rhim *et al.*, 1993; Orsini *et al.*, 1996). There is also a

concern that denaturants may irreversibly unfold the protein (Waszak *et al.*, 1992; Orsini *et al.*, 1996). Finally, Kirsch *et al.* (1996) report the use of Tris (2-carboxyethyl) phosphine-HCl (TCEP) to reduce the cysteines in Tat because they identified covalent adducts of DTT and Tat by mass spectrometry. Covalently modified protein should be avoided as the adducts might interfere both with folding and biological activity. In our group, DTT has been observed to affect the aggregation state of Tat protein (G. Henry, personal communication). In the present work, two strategies were attempted to overcome these difficulties. In both, β-mercaptoethanol was used in place of DTT as a reducing agent. It has the added advantage of being compatible with the Ni^{2+}-chelate matrix. We purified Tat protein under strongly denaturing conditions (6 M guanidine-HCl) and under non-denaturing conditions (no guanidine-HCl). Non-denaturing conditions resulted in low extraction yields without any clear reduction in the level of Tat oligomerization (see Figures 3.20 and 3.21). Purification of Tat protein under denaturing conditions had several benefits over non-denaturing conditions. Perhaps most important, was the ability to solubilise all of the Tat protein as evidenced by SDS-PAGE and/or Western immuno-blot analysis after lysis of the cells. Comparison of lane 5 in Figure 3.21 and lane 6 in Figure 3.25 shows the complete absence of Tat protein bands in the insoluble fraction of the cell lysate during denaturing purification, while significant amounts of Tat protein remain in the insoluble fraction during non-denaturing purification. Since the expression levels of Tat protein were low to begin with, the recovery of all available Tat protein from the lysates was deemed essential. A further concern is that Tat protein is capable of binding nucleic acid due to its high positive charge, a potential problem with the purification under non-denaturing conditions that

would require further purification steps. Incubation of the eluate and supernatant from a purification under non-denaturing conditions with RNase showed the presence of RNA in both samples (Figure 3.23). Elution of His_6-Tat_{1-86} from the Ni^{2+} affinity column, whether under non-denaturing or denaturing conditions, resulted in gradual removal of the Tat protein from the column in approximately 25 tubes containing approximately 1.5 to 2 mL. This pattern was observed with both imidazole and Na-acetate/acetic acid elution buffers (for an example see Figures 3.27 and 3.28) and is an indication that the protein has formed multiple oligomeric species. Elution of the columns with EDTA after elution with imidazole or Na-acetate/acetic acid routinely showed that little or no nickel bound Tat protein was being retained on the column. Subsequent elution with 6 M urea often washed significant protein, including SDS-PAGE visible amounts of Tat protein, off of the column. Thus, although pure Tat protein could be eluted from the column, a smaller, but significant amount of Tat protein was precipitating on the column.

4.4 Comparison of Reported HIV-1 Tat Protein Purification Procedures

Reports in the literature of production of relatively large amounts of pure Tat protein varies from 200 µg/L of culture (GST-Tat_{1-86}) (Rhim *et al.*, 1993) to approximately 5 mg/L (Tat_{1-76}) (Waszak *et al.*, 1992). Waszak *et al.* (1992) report the production of 5 mg/L of pure monomeric Tat protein expressed without any purification tags, but containing two non-native amino acids at the carboxy-terminus of the protein. Dr. Gillian Henry has purified His_6-Tat_{1-72} in pET28b(+) (personal communication) and our attempt to purify His_6-Tat_{1-72} resulted in an estimated 2-5 mg/L of purified Tat (see figures 3.27 and 3.28). Only one group has shown the production of Tat protein from *E.*

coli in an amount and purity suitable for NMR (Tat_{1-86}) (Kirsch *et al.*, 1996). Kirsch *et al.*, incorporated an amino-terminal six-histidine purification tag in recombinant Tat_{1-86} protein for purification by Ni^{2+} metal-chelate affinity chromatography, and produced approximately 2.5 mg/L of pure Tat protein. Subsequent to purification, Kirsch *et al.* (1996) reduced the protein using solid TCEP in a ratio of TCEP:Tat of 25:1. Thus, the ability to reduce Tat protein without the use of DTT is a potential improvement in our current protocol.

My attempt to produce HIV-1 Tat protein in an amount and purity suitable for NMR solution studies was only partly successful. With the use of denaturing conditions, Tat protein expressed from the pTatKL16 plasmid was purified in one step to greater than 90 % purity as judged by Coomassie-stained SDS-PAGE. The yield is estimated to be approximately 1/2 mg/L, or roughly 20 % of the necessary yield to produce NMR suitable quantities from small-scale cultures. At this point, it does not appear that pTatKL16 can produce enough Tat protein to make NMR studies feasible. However, pTatKL26 has been shown to rival the level of Tat protein expression achieved from pTatR, and as judged by Western immuno-blot and SDS-PAGE analysis, is producing several fold more Tat protein than pTatKL16. Many of the problems encountered during purification of Tat produced from pTatKL16 were attributed to low levels of protein production. Future attempts to purify Tat_{1-86} from pTatKL26 should no longer be hampered by low levels of protein expression, and will hopefully allow further refinements in the purification process.

4.5 Future Avenues of Inquiry

Once suitable amounts of pure protein are available, one of the first avenues of research that should be pursued is determination of the metal binding properties of the Tat protein. The protein contains seven cysteine residues and three histidines. A review of the literature shows that no one has conducted a controlled, thorough study of the metal binding characteristics of the protein. Measurement of structural information from CD spectropolarimetry studies of Tat protein in the presence and absence of Zn^{2+} would be a natural first step toward an NMR solution structure determination. The second avenue of inquiry that should be pursued is the production and purification of HIV-1 TAR RNA. It may well be necessary to bind Tat protein to TAR in order to obtain a solution structure of functional Tat protein. In addition, the production of TAR RNA would allow CD spectropolarimetry studies of Tat structure in the presence and absence of TAR RNA and the influence, if any, of metals on their interaction. Finally, site-directed mutagenesis of the Tat protein would be a logical third step once structural information has been obtained for Tat. Metal binding studies in concert with a complete set of cysteine and histidine mutant variants of the Tat protein should be an informative line of research to pursue. A review of the literature shows that no one has systematically mutated the histidines, and no one has thoroughly investigated their potential role in metal binding. Mutation of each of the cysteines and histidines that are potential metal ion ligands would allow preliminary identification of the residues involved in metal ion chelation by the Tat protein, providing insights into their structural role.

References

Aboul-ela-F, Karn-J, Varani-G (1995) Journal of Molecular Biology, **253**: 313-32.

Aboul-ela-F, Karn-J, Varani-G (1996) Nucleic Acids Research, **24** (20): 3974-81.

Adams-SE, Johnson-ID, Braddock-M, Kingsman-AJ, Kingsman-SM, Edwards-RM (1988) Nucleic Acids Research, **16**: 4287-99.

Altschul-SF, Gish-W, Miller-W, Myers-EW, Lipman-DJ (1990) Journal of Molecular Biology, **215** (3): 403-10.

Bayer-P, Kraft-M, Ejchart-A, Westendorp-M, Frank-R, Rösch-P (1995) Journal of Molecular Biology, **247**: 529-35.

Better-M, Chang-CP, Robinson-RR, Horwitz-AH (1988) Science, **240**:1041-43.

Bieniasz-PD, Grdina-TA, Bogerd-HP, Cullen-BR (1998) EMBO Journal, **17** (23): 7056-7065.

Bieniasz-PD, Grdina-TA, Bogerd-HP, Cullen-BR (1999) Proceedings of the National Academy of Sciences USA, **96**: 7791-96.

Birnboim-HC, Doly-J (1979) Nucleic Acids Research, **7** (6): 1513-23.

Brown-JC, Brown-BA II, Li-Y, Hardin-CC (1998) Biochemistry, **37**: 16338-348.

Brundage-L, Hendrick-JP, Schiebel-E, Driessen-AJM, Wickner-W (1990) Cell, **62**: 649-657.

Burd-CG, Dreyfuss-G (1994) Science, **265**: 615-621.

Calnan-BJ, Biancalana-S, Hudson-D, Frankel-AD (1991a) Genes & Development, **5**: 201-10.

Calnan-BJ, Tidor-B, Biancalana-S, Hudson-D, Frankel-AD (1991b) Science, **252**: 1167-71.

Certa-U, Bannwarth-W, Stüber-D, Gentz-R, Lanzer-M, Le Grice-S, Guillot-F, Wendler-I, Hunsmann-G, Bujard-H, Mous-J (1986) EMBO Journal, **5**: 3051-56.

Chazal-N, Gay-B, Carrière-C, Tournier-J, Boulanger-P (1995) Journal of Virology, **69** (1): 365-75.

Chen-D, Fong-Y, Zhou-Q (1999) Proceedings of the National Academy of Sciences USA, **96**: 2728-733.

Churcher-MJ, Lamont-C, Hamy-F, Dingwall-C, Green-SM, Lowe-AD, Butler-PJG, Gait-MJ, Karn-J (1993) Journal of Molecular Biology, **230**: 90-110.

Ciccarelli-RB, Loomis-LA, McCoon-PE, Holzschu-DL (1990) Nucleic Acids Research, **18** (5): 1243-48.

Dalgleish-AG, Beverley-PCL, Clapham-PR, Crawford-DH, Greaves-MF, Weiss-RA (1984) Nature, **312**: 763-67.

Delling-U, Reid-LS, Barnett-RW, Ma-MYX, Climie-S, Sumner-Smith-M, Sonenberg-N (1992) Journal of Virology, **66** (5): 3018-25.

Dingwall-C, Ernberg-I, Gait-MJ, Green-SM, Heaphy-S, Karn-J, Lowe-AD, Singh-M, Skinner-MA, Valerio-R (1989) Proceedings of the National Academy of Sciences USA, **86**: 6925-929.

Dubendorff-JW, Studier-FW (1991) Journal of Molecular Biology, **219**: 45-59.

Echetebu-CO, Rice-AP (1993) Journal of Acquired Immune Deficiency Syndromes, **6**: 550-57.

Economou-A, Pogliano-JA, Beckwith-J, Oliver-DB, Wickner-W (1995) Cell, **83**: 1171-81.

Ensoli-B, Buonaguro-L, Barillari-G, Fiorelli-V, Gendelman-R, Morgan-RA, Wingfield-P, Gallo-RC (1993) Journal of Virology, **67** (1): 277-287.

Feng-Y, Broder-CC, Kennedy-PE, Berger-EA (1996) Science, **272**: 872-77.

Finzi-D, Hermankova-M, Pierson-T, Carruth-LM, Buck-C, Chaisson-RE, Quinn-TC, Chadwick-K, Margolick-J, Brookmeyer-R, Gallant-J, Markowitz-M, Ho-DD, Richman-DD, Siliciano-RF (1997) Science, **278**: 1295-1300.

Frankel-AD, Bredt-DS, Pabo-CO (1988) Science, **240**: 70-73.

Frankel-AD, Young-JAT (1998) Annual Reviews in Biochemistry, **67**: 1-25.

Freed-EO, Englund-G, Martin-MA (1995) Journal of Virology, **69** (6): 3949-54.

Freund-J, Vértesy-L, Koller-K, Wolber-V, Heintz-D, Kalbitzer-HR (1995) Journal of Molecular Biology, **250**: 672-688.

Folks-TM, Hart-CE (1997) Chapter 2 in AIDS: Biology, Diagnosis, Treatment and Prevention, fourth edition, Lippincott-Raven (New York) pp 29-43.
Fujinaga-K, Taube-R, Wimmer-J, Cujec-TP, Peterlin-BM (1999) Proceedings of the National Academy of Sciences USA, **96**: 1285-90.

Garber-ME, Wei-P, KewalRamani-VN, Mayall-TP, Herrmann-CH, Rice-AP, Littman-DR, Jones-KA (1998) Genes & Development, **12**: 3512-527.

Garcia-JA, Harrich-D, Pearson-L, Mitsuyasu-R, Gaynor-RB (1988) EMBO Journal, **7** (10): 3143-3147.

Gregoire-CJ, Loret-EP (1996) Journal of Biological Chemistry, **271** (37): 22641-646.

Hamy-F, Asseline-U, Grasby-J, Iwai-S, Pritchard-C, Slim-G, Butler-PJG, Karn-J, Gait-MJ (1993) Journal of Molecular Biology, **230**: 111-123.

Hartl-FU, Lecker-S, Schiebel-E, Hendrick-JP, Wickner-W (1990) Cell, **63**: 269-279.

Helling-RB, Goodman-HM, Boyer-HW (1974) Journal of Virology, **14** (5): 1235-44.

Hochuli-E, Dobeli-H, Schacher-A (1987) Journal of Chromatography, **411**: 177-84.

Huq-I, Rana-TM (1997) Biochemistry, **36**: 12592-599.

Ikemura-T (1981) Journal of Molecular Biology, **146**: 1-21.

Ippolito-JA, Steitz-TA (1998) Proceedings of the National Academy of Sciences USA, **95**: 9819-824.

Isel-C, Karn-J (1999) Journal of Molecular Biology, **290**: 929-41.

Ivanov-D, Kwak-YT, Nee-E, Guo-J, García-Martínez-LF, Gaynor-RB (1999) Journal of Molecular Biology, **288**: 41-56.

Jeang-K, Xiao-H, Rich-EA (1999) Journal of Biological Chemistry, **274** (41): 28837-840.

Karn-J (1999) Journal of Molecular Biology, **293**: 235-54.

Kim-K, Pallaghy-CK (1996) U.S. Department of Commerce/NOAA/NMFS/NWFSC/ Molecular Biology Protocols

Kirsch-T, Boehm-M, Schuckert-O, Metzger-AU, Willbold-D, Frank-RW, Rösch-P (1996) Protein Expression and Purification, **8**: 75-84.

Klatzmann-D, Barré-Sinoussi-F, Nugeyre-MT, Dauguet-C, Vilmer-E, Griscelli-C, Brun-Vezinet-F, Rouzioux-C, Gluckman-JC, Chermann-J, Montagnier-L (1984) Science, **225**: 59-63.

Klostermeier-D, Bayer-P, Kraft-M, Frank-RW, Rösch-P (1997) Biophysical Chemistry, **63**: 87-96.

Kondo-E, Göttlinger-HG (1996) Journal of Virology, **70** (1): 159-64.

Kumamoto-CA (1989) Proceedings of the National Academy of Sciences USA, **86**: 5320-24.

Kuppuswamy-M, Subramanian-T, Srinivasan-A, Chinnadurai-G (1989) Nucleic Acids Research, **17** (9): 3551-561.

Kwak-YT, Ivanov-D, Guo-J, Nee-E, Gaynor-RB (1999) Journal of Molecular Biology, **288**: 57-69.

Laemmli-UK (1970) Nature **227**: 680-85.

Lill-R, Cunningham-K, Brundage-LA, Ito-K, Oliver-D, Wickner-W (1989) EMBO Journal, **8** (3): 961-66.

Lu-Y, Bennett-RP, Wills-JW, Gorelick-R, Ratner-L (1995) Journal of Virology, **69** (11): 6873-79.

Ma-M, Nath-A (1997) Journal of Virology, **71** (3): 2495-2499.

Mammano-F, Kondo-E, Sodroski-J, Bukovsky-A, Göttlinger-HG (1995) Journal of Virology, **69** (6): 3824-30.

Marshal-NF, Peng-J, Xie-Z, Price-DH (1996) Journal of Biological Chemistry, **271** (43): 27176-183.

McDougal-JS, Kennedy-MS, Sligh-JM, Cort-SP, Mawle-A, Nicholson-JKA (1986) Science, **231**: 382-85.

McKenna-MC, Muchardt-C, Gaynor-R, Eisenberg-D (1994) Protein Expression and Purification, **5**: 105-11.

Metzger-AU, Bayer-P, Willbold-D, Hoffmann-S, Frank-RW, Goody-RS, Rösch-P (1997) Biochemical and Biophysical Research Communications, **241**: 31-36.

Moore-JP (1997) Science, **276**: 51-52.

Mrestani-Klaus-C, Fengler-A, Brandt-W, Faust-J, Wrenger-S, Reinhold-D, Ansorge-S, Neubert-K (1998) Journal of Peptide Science, **4**: 400-410.

Mujeeb-A, Bishop-K, Peterlin-BM, Turck-C, Parslow-TG, James-TL (1994) Proceedings of the National Academy of Sciences USA, **91**: 8248-52.

Mullis-K, Faloona-F, Scharf-S, Saiki-R, Horn-G, Erlich-H (1986) Cold Spring Harbor Symposia on Quantitative Biology, **LI**: 263-73.

Orsini-MJ, García-Martínez-LF, Mavankal-G, Gaynor-RB, Debouck-CM (1996) Protein Expression and Purification, **8**: 238-46.

Ott-M, Emiliani-S, Van Lint-C, Herbein-G, Lovett-J, Chirmule-N, McCloskey-T, Pahwa-S, Verdin-E (1997) Science, **275**: 1481-85.

Ping-Y, Rana-TM (1999) Journal of Biological Chemistry, **274** (11): 7399-404.

Pope-B, Kent-HM (1996) Nucleic Acids Research, **24** (3): 536-537.

Porath-J, Carlsson-J, Olsson-I, Belfrage-G (1975) Nature, **258**: 598-99.

Power-BE, Ivancic-N, Harley-VR, Webster-RG, Kortt-AA, Irving-RA, Hudson-PJ (1992) Gene, **113**: 95-99.

Pritchard-CE, Grasby-JA, Hamy-F, Zacharek-AM, Singh-M, Karn-J, Gait-MJ (1994) Nucleic Acids Research, **22** (13): 2592-2600.

Puglisi-JD, Tan-R, Calnan-BJ, Frankel-AD, Williamson-JR (1992) Science, **257**: 76-80.

Puglisi-JD, Chen-L, Calnan-BJ, Frankel-AD, Williamson-JR (1993) Proceedings of the National Academy of Sciences USA, **90**: 3680-84.

Purvis-SF, Jacobberger-JW, Sramkoski-M, Patki-AH, Lederman-MM (1995) AIDS Research and Human Retroviruses, **11** (4): 443-450.

Ratner-L, Haseltine-W, Patarca-R, Livak-KJ, Starcich-B, Josephs-SF, Doran-ER, Rafalski-JA, Whitehorn-EA, Baumeister-K, Ivanoff-L, Petteway-SR-Jr., Pearson-ML, Lautenberger-JA, Papas-TS, Ghrayeb-J, Chang-NT, Gallo-RC, Wong-Staal-F (1985) Nature, **313**: 277-84

Rhim-H, Chan-F, Echetebu-CO, Rice-AP (1993) Protein Expression and Purification, **4**: 24-31.

Rosenberg-AH, Lade-BN, Chui-D, Lin-S, Dunn-JJ, Studier-FW (1987) Gene, **56**: 125-35.

Sadaie-MR, Mukhopadhyaya-R, Benaissa-ZN, Pavlakis-GN, Wong-Staal-F (1990) AIDS Research and Human Retroviruses, **6** (11): 1257-63.

Sambrook-J, Fritsch-EF, Maniatis-T (1989) Molecular Cloning: A Laboratory Manual, second edition, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Schultz-AM, Oroszlan-S (1983) Journal of Virology, **46** (2): 355-61.

Shapiro-AL, Viñuela-E, Maizel-JV Jr. (1967) Biochemical and Biophysical Research Communications, **28** (5): 815-20.

Smith-MC, Furman-TC, Ingolia-TD, Pidgeon-C (1988) Journal of Biological Chemistry, **263** (15): 7211-215.

Smith-RE, Talhouk-JW, Brown-EE, Edgar-SE (1998) AIDS Research and Human Retroviruses, **14** (10): 851-868.

Sticht-H, Willbold-D, Bayer-P, Ejchart-A, Herrmann-F, Rosin-Arbesfeld-R, Gazit-A, Yaniv-A, Frank-R, Rösch-P (1993) European Journal of Biochemistry, **218**: 973-976.

Studier-FW, Moffatt-BA (1986) Journal of Molecular Biology, **189**: 113-30.

Studier-FW, Rosenberg-AH, Dunn-JJ, Dubendorff-JW (1990) Methods in Enzymology, **185**: 60-89.

Studier-FW (1991) Journal of Molecular Biology, **219**: 37-44.

Tao-J, Frankel-AD (1992) Proceedings of the National Academy of Sciences USA, **89**: 2723-26.

Tartof-KD, Hobbs-CA (1987) Bethesda Research Laboratory Focus, **9**: 12.

Tautz-D, Renz-M (1983) Analytical Biochemistry, **132**: 14-19.

Tedeschi-H (1993) page 11 in Cell Physiology: Molecular Dynamics, second edition (Dubuque).

Theunissen-HJM, Dijkema-R, Grootenhuis-PDJ, Swinkels-JC, de Poorter-TL, Carati-P, Visser-A (1993) Journal of Biological Chemistry, **268** (12): 9035-40.

Thuring-RWJ, Sanders-JPM, Borst-P (1975) Analytical Biochemistry, **66**: 213-20.

Ulich-C, Dunne-A, Parry-E, Hooker-CW, Gaynor-RB, Harrich-D (1999) Journal of Virology, **73** (3): 2499-2508.

Verhoef-K, Bauer-M, Meyerhans-A, Berkhout-B (1998) AIDS Research and Human Retroviruses, **14** (17): 1553-59.

Wada-K, Wada-Y, Doi-H, Ishibashi-F, Gojobori-T, Ikemura-T (1991) Nucleic Acids Research, **19** (supplement): 1981-86.

Waszak-GA, Hasler-JM, McQuade-TJ, Tarpley-WG, Deibel-MR Jr. (1992) Protein Expression and Purification, **3**: 126-33.

Weber-K, Osborn-M (1969) Journal of Biological Chemistry, **244** (16): 4406-12.

Wei-P, Garber-ME, Fang-S, Fischer-WH, Jones-KA (1998) Cell, **92**: 451-62.

Westendorp-MO, Shatrov-VA, Schulze-Osthoff-K, Frank-R, Kraft-M, Los-M, Krammer-PH, Dröge-W, Lehmann-V (1995) EMBO Journal, **14** (3): 546-54.

White-DO, Fenner-FJ (1994) Chapter 35 in Medical Virology, fourth edition, Academic Press (San Diego).

Wickner-W, Driessen-AJM, Hartl-F (1991) Annual Reviews of Biochemistry, **60**: 101-24.

Willbold-D, Krüger-U, Frank-R, Rosin-Arbesfeld-R, Gazit-A, Yaniv-A, Rösch-P (1993) Biochemistry, **32**: 8439-445.

Willis-EH, Mardis-ER, Jones-WL, Little-MC (1990) Biotechniques, **9** (1): 92-99.

Wong-JK, Hezareh-M, Günthard-HF, Havlir-DV, Ignacio-CC, Spina-CA, Richman-DD (1997) Science, **278**: 1291-95.

Wrenger-S, Reinhold-D, Hoffmann-T, Kraft-M, Frank-R, Faust-J, Neubert-K, Ansorge-S (1996) FEBS Letters, **383**: 145-49.

Wright-CM, Felber-BK, Paskalis-H, Pavlakis-GN (1986) Science **234**: 988-92.

Xiao-H, Neuveut-C, Benkirane-M, Jeang-K (1998) Biochemical and Biophysical Research Communications **244**: 384-89.

Zacharias-M, Hagerman-PJ (1995) Proceedings of the National Academy of Sciences USA **92**: 6052-56.