Title
Probing the function of the HIV-1 Tat protein

Author
Lynn, Aenoch Jay

Publication Date
1998

Peer reviewed|Thesis/dissertation

PROBING THE FUNCTION OF THE HIV-1 TAT PROTEIN

by

AENOCH JAY LYNN

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

BIOCHEMISTRY

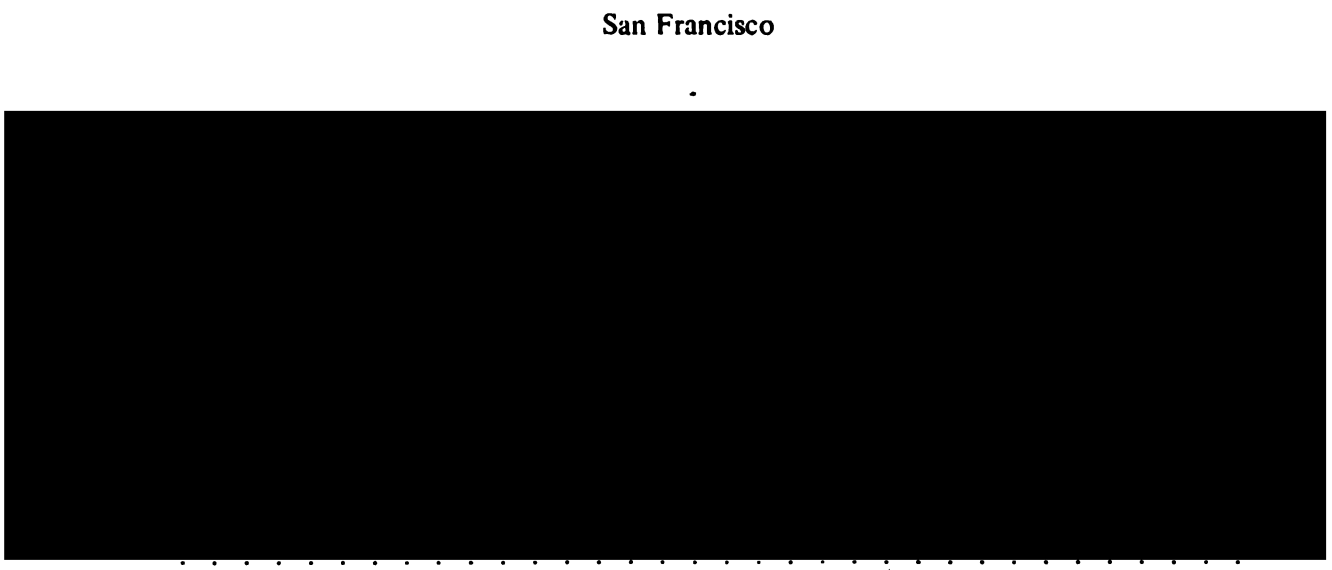
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



. . . . . .

Date University Librarian

Degree Conferred: . . . . . .

ABSTRACT

Probing the Function of the HIV Tat Protein

Aenoch Jay Lynn

The HIV-1 Tat protein plays a critical role in the *trans*-activation of viral transcription. The detailed mechanism of action has been elusive, but recent work has suggested that Tat acts during the elongation phase of transcription. Tat binds several cellular proteins, including TBP and RNA polymerase II, but whether these contacts are important has not been determined. This thesis attempts to describe the role Tat plays in viral transcription using two sets of experiments. The first is an attempt to find cellular proteins that may function in a manner similar to Tat. The second is an attempt to better characterize the spatial relationship between Tat, TAR, and the unknown cellular loop-binding factor.

TABLE OF CONTENTS

Title Page i

Abstract iii

Table of Contents iv

List of Tables vi

List of Figures vii

Chapter 1 — Background and Significance 1

Tat 2

Control of Transcription Elongation 10

Biochemical Experiments on Tat *trans*-Activation 11

BIV Tat-TAR 14

Chapter 2 — Cellular Analogue of Tat? 17

Viruses Take Advantage of Cellular Machinery 17

Targeting the Tat Activation Domain Through the N-Terminus 18

Performing the Screen 28

Chapter 3 — Screening Two Libraries Concurrently 31

The Loop-Binding Factor Aids Tat in Binding to TAR 31

Two-Plasmid Protoplast Fusion 32

Tat-TAR Double Library Experiment 33

TAR Library 34

Tat Library 35

Simultaneous Screen of Two Libraries. 36

Chapter 4 — Discussion 38

cDNA Library Screen 38

Double Library Screen 40

Chapter 5 — Materials and Methods 43
Making Protoplasts 43
Protoplast Fusion 44
Two-Plasmid Frotoplast Fusion 45
FACS 45
Plasmid Recovery 46
Fluorescence Microscopy 47
Electroporation 47
CAT Assays 48
TAR Libraries 49
Tat Library 50
Combinatorial Considerations 54
Bibliography 55

LIST OF TABLES

Table 1. TAR Library Oligonucleotides 51

LIST OF FIGURES

Figure 1. HIV-1 *trans*-Activation 3

Figure 2. HIV-1 Tat Protein Sequence 5

Figure 3. HIV-1 TAR Sequence 6

Figure 4. Organization of HIV-1 and the LTR 8

Figure 5. BIV TAR and BIV Tat RNA-Binding Domain 16

Figure 6. Possible Orientation of cDNA Library Fusion 20

Figure 7. Screen for Tat-Like Factors 21

Figure 8. Linkers of 2, 3, and 4 Glycines 24

Figure 9. Removal of Glycines from Linkers 26

Figure 10. Comparison of Different Linker Lengths 27

Figure 11. FACS Scans of Lipofectin Transfected Plasmids 29

Figure 12. FACS Scans of Two-plasmid Protoplast Fusions 34

CHAPTER 1 — BACKGROUND AND SIGNIFICANCE

Human immunodeficiency virus type-1 (HIV-1 or HIV) is the primary causative agent for acquired immunodeficiency syndrome (AIDS). An early step in viral production is the activation of viral transcription. This step requires the viral protein Tat (*trans*-activator) to produce full-length viral transcripts (Muesing *et al.*, 1987). Without Tat, viral transcription is insufficient for viral production. The mechanism that Tat uses to *trans*-activate the viral promoter is relatively unusual. While there is evidence that Tat increases the rate of transcription initiation (Laspia *et al.*, 1989) and may increase the translation of the viral mRNA (Braddock *et al.*, 1989) Tat seems to function primarily to increase the elongation efficiency of the transcription machinery resulting in transcription of the entire viral genome (Kao *et al.*, 1987; Laspia *et al.*, 1989; Marciniak and Sharp, 1991; Feinberg *et al.*, 1991). Other cellular genes are believed to be controlled at the level of transcription elongation and a detailed knowledge of Tat function may give insights into the control mechanism of these genes along with providing a novel avenue to attack the virus.

Many insights into cellular function have come from investigating viral life cycles. Viruses have limited capacity to carry genetic information, so they exploit cellular proteins and machinery whenever possible. Investigating viral proteins may yield potentially valuable avenues to study cellular mechanisms. The Rev protein from HIV is an example of a viral protein that uses a cellular machinery. Rev binds RRE (Rev response element) RNA in the gag coding region to facilitate export of un-spliced RNAs out of the nucleus. This allows the expression of viral structural genes from the un-spliced message. When bound to the RRE, Rev forms multimers and is actively exported from the nucleus, with its attached RNA cargo (Malim *et al.*, 1989).

Exploration of Rev function has uncovered the interaction of a small viral regulatory protein with a large cellular machinery, highlighting the possibility that investigating Tat may also uncover another such relationship between the virus and its

host cell. It seems unlikely that HIV would *de novo* develop Tat to interact with the transcription machinery to increase dramatically viral transcription. If the ability exists to control transcription during elongation, then the cell likely uses similar mechanism.

Tat

The Tat protein increases the full–length transcription of the viral genome while at the same time reducing short, incomplete transcription. Upon infection and integration into the host genome, the HIV long terminal repeat (LTR) promoter (Figure 1A) generates primarily short, aborted RNAs and few full–length transcripts (Marciniak and Sharp, 1991). These full–length transcripts are spliced and encode regulatory proteins, one of which is Tat (Greene, 1990). Tat binds to TAR (the *trans*–acting response element), an RNA stem–loop at the 5' end of the nascent HIV transcript (Berkhout *et al.*, 1989) and increases the production of full–length transcripts at the expense of short, incomplete RNAs (Figure 1B). Without Tat polymerases have a low processivity and terminate transcription before reaching the end of the viral genome. In the presence of Tat polymerases are highly processive and transcribe through pause and termination sites that would block non–processive polymerases. This effect of Tat has been observed on other promoters besides the HIV LTR promoter. Tat *trans*–activation has been seen on the Rous sarcoma virus (RSV) promoter and the *Drosophilla* alcohol dehydrogenase promoter (unpublished results), both requiring Tat to be delivered to the nascent transcript to function. Tat, then, either aids in the assembly of processive polymerases or it modifies an elongating polymerase to make it processive.

Some early experiments suggested that the Tat protein may also affect the translation of the viral message (Braddock *et al.*, 1989; Braddock *et al.*, 1990). These experiments were performed in Xenopus oocytes where transcriptional effects were minimal. In mammalian cell lines the level of the translational effect was minimal and the

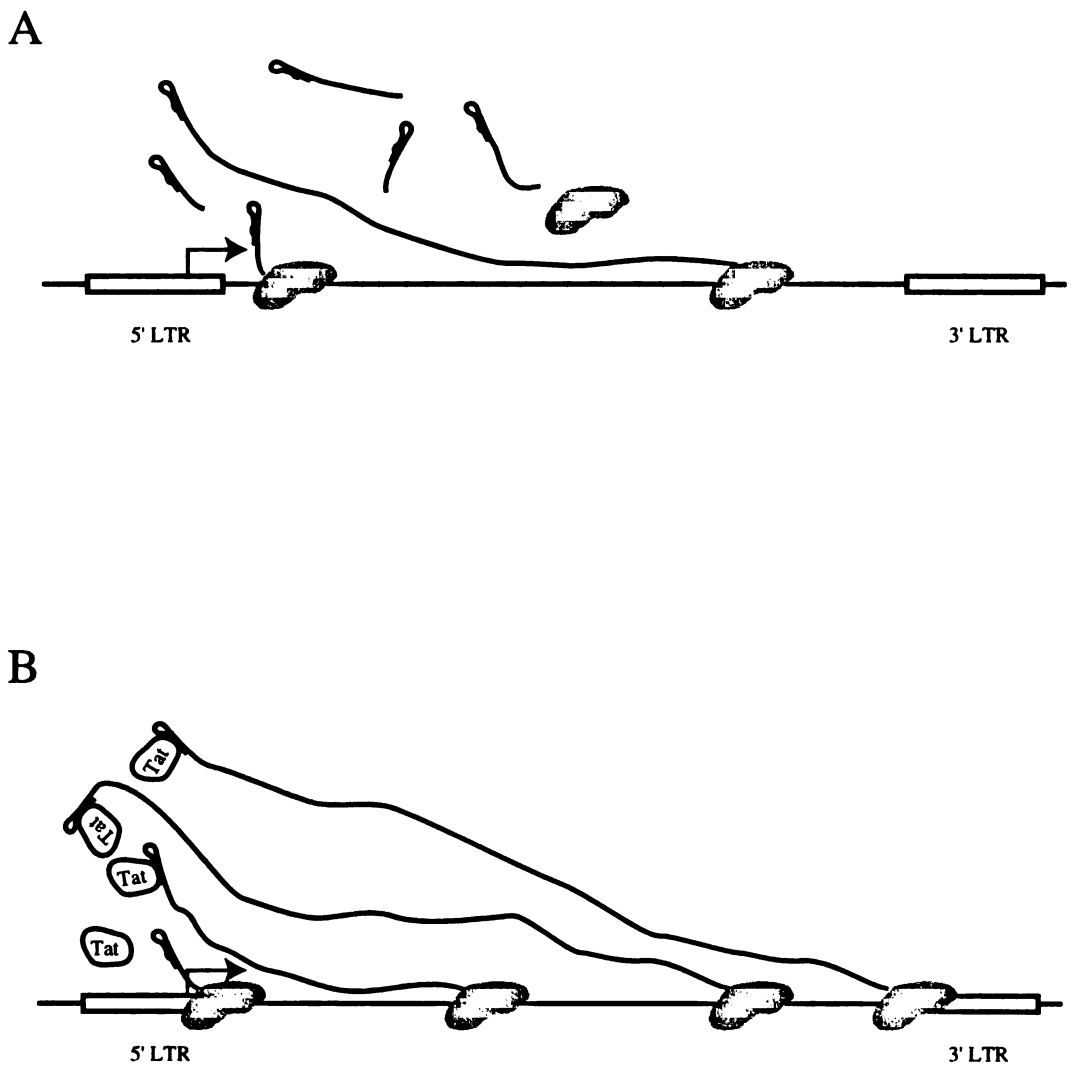


Figure 1. HIV-1 *trans*-Activation.
(A) In the absence of Tat protein the RNA polymerases (in gray) transcribing the viral DNA genome are poorly processive and few make full-length transcripts. (B) When Tat is expressed and correctly targeted to TAR at the 5' end of the nascent transcript it modifies the processivity of the polymerases such that primarily full-length transcripts are made.

transcriptional effect dominated. If Tat does effect translation, it does so minimally in its host organism (Chin *et al.*, 1991).

Tat is encoded by two exons, the first of which is sufficient for Tat *trans*–activation. Tat consists of five domains: an N–terminal activation domain (amino acids 1–48) responsible for transcriptional activation which is further divided into an N-terminal region, a cystine-rich region, and a core domain; a basic domain (amino acids 49–57) that functions as a nuclear localization signal (Hauber *et al.*, 1989) and specifically binds TAR RNA (Weeks *et al.*, 1990); and a dispensable C–terminal domain (amino acids 58–72) (Figure 2). The Tat basic domain binds a three–nucleotide bulge region of TAR (Dingwall *et al.*, 1990; Roy *et al.*, 1990), with a single arginine making specific contacts with a guanine nucleotide and two phosphates (Calnan *et al.*, 1991b) (Figure 3).

While *in vitro* this interaction is sufficient for Tat binding, *in vivo* the TAR loop sequence is required for Tat *trans*–activation (Feng and Holland, 1988). Tat makes no known contacts with this loop, but a host cellular factor most likely binds and presumably assists Tat in binding to TAR (Alonso *et al.*, 1992). Gel shift experiments with the RNA-binding domain of Tat (a synthetic peptide of amino acids 49-58) showed that this peptide binds the wild-type TAR sequence with the same affinity as the loop-mutated TAR (Weeks *et al.*, 1990). However, this loop-mutated TAR does not function in *in vivo* *trans*-activation experiments.

This loop binding protein is not absolutely required for Tat activity, however, since the activation domain of Tat can be targeted to the 5' end of the transcript through a heterologous RNA–protein interaction (Selby and Peterlin, 1990). For example, the Rev–RRE protein-RNA interaction (Southgate *et al.*, 1990; Tan *et al.*, 1993), or R17 coat protein-RNA interaction (Selby and Peterlin, 1990) can be substituted for the HIV Tat-TAR interaction, suggesting that this cellular factor is unnecessary for Tat *trans*–activation *per se* and this factor functions by assisting Tat binding to TAR.

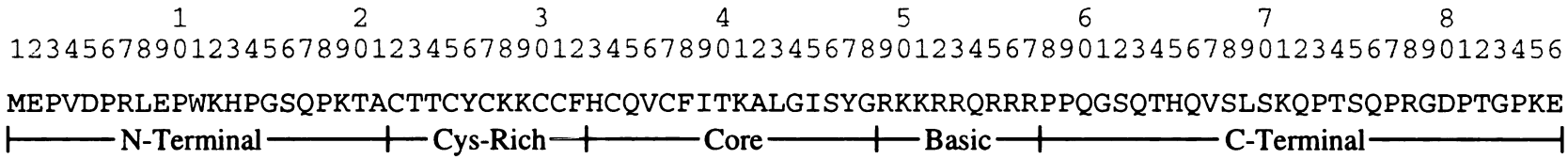


Figure 2. HIV-1 Tat Protein Sequence.
The Tat gene has 5 domains. The N-terminal, Cysteine-rich, and Core domains are responsible for the *trans*-activation of the LTR promoter and are used as Tat_{1-48} in experiments. The basic domain (Tat_{49-57}) binds to the TAR RNA stem-loop in the nascent transcript making a specific contact with arginine 52. The C-terminal domain plays a role in viral production.

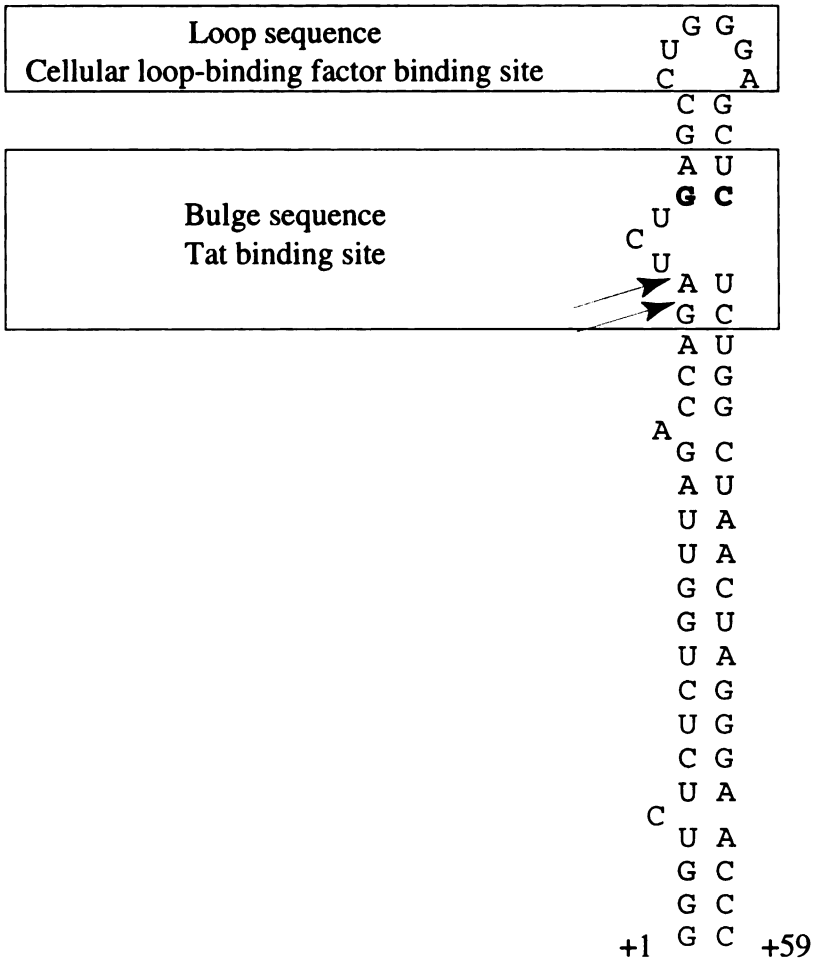
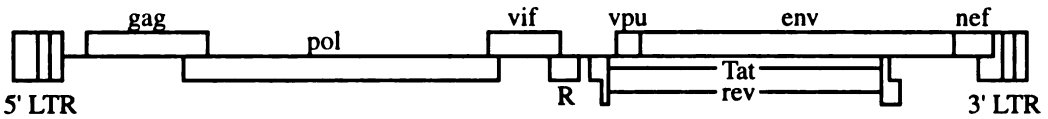


Figure 3. HIV-1 TAR Sequence
Primary and secondary structure of HIV-1 TAR, nucleotides +1 – +59. Tat binds the tri-nucleotide bulge and the cellular loop-binding factor binds the hexa-nucleotide loop. Both sequences are required for Tat function *in vivo*. Arginine 52 from Tat makes a specific contact with the base pair G26•C39 (in bold) and the phosphates of A22 and U23 (indicated by arrows).

The HIV LTR is divided into three regions: U3, R, and U5. The known regulatory sequences for initiation and elongation reside in U3 and R (Frankel, 1992; Jones and Peterlin, 1994) (Figure 4). Three elements in the LTR are sufficient for Tat *trans*–activation. These are three Sp1 binding sites (Harrich *et al.*, 1989), the TATA–box (Garcia *et al.*, 1989; Olsen and Rosen, 1992), and TAR (Muesing *et al.*, 1987; Feng and Holland, 1988; Hauber and Cullen, 1988). The Sp1 sites can be replaced with Gal4 binding sites and the Gal4 protein can substitute for Sp1, suggesting that all that is required is an enhancer of transcriptional initiation. Removal of one or two of the Sp1 binding sites moderately reduces the ability of the promoter to be *trans*-activated and removal of all three severely reduces *trans*-activation (Williams *et al.*, 1996). The HIV TATA-box differs from canonical sequences, and mutations towards the consensus reduce or abolish the ability of the promoter to be *trans*-activated (Garcia *et al.*, 1989), suggesting that the HIV TATA-box sequence is partly responsible for assembly of non-processive RNA polymerase II complexes that can be acted upon by Tat. TAR forms an RNA stem–loop between nucleotides +1 and +59 and Tat binds this structure making contacts around a tri-nucleotide bulge in the upper portion of the stem (Figure 3) (Dingwall *et al.*, 1990; Roy *et al.*, 1990).

The interaction of the Tat RNA-binding domain and TAR has been functionally replaced by the R17 coat protein RNA binding domain and the R17 RNA stem-loop structure (Selby and Peterlin, 1990). This interaction provides good *trans*-activation using HIV LTR CAT reporters when the Tat basic domain is replaced with the coat protein and TAR is replaced with the R17. The R17 interaction does not require a cellular factor to assist in the binding, so the use of high affinity RNA-protein interactions allows *trans*-activation to be loop-factor independent. Besides the R17 protein-RNA interaction, the HIV Rev-RREIIB (Tan *et al.*, 1993) and the BIV Tat-TAR (Chen and Frankel, 1994) interactions can also functionally replace the HIV Tat-TAR interaction. Both of these

A. HIV-1



B. LTR

HIV-1 3' LTR nucleotides 9085–9718
isolate HXB2 accession K03455

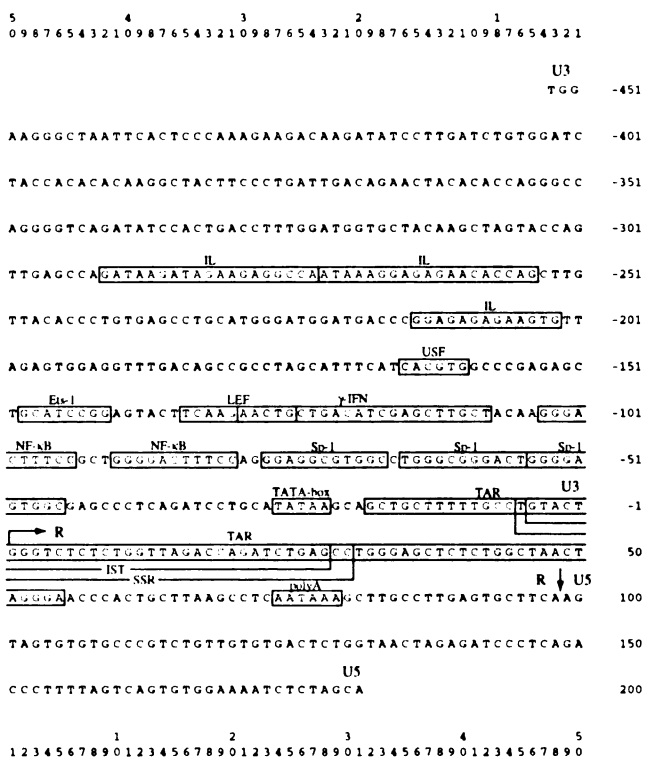


Figure 4. Organization of HIV-1 and the LTR.
(A) The genomic organization of HIV-1. Transcription initiates in the 5' LTR (first vertical line in LTR box) and terminates in the 3' LTR (second vertical line). (B) Nucleotide sequence of the 3' LTR from isolate HXB2 that was used in these studies. U3 (-453 – -1), R (+1 – +98), and U5 (+99 – +181) are the major divisions in the LTR. Binding sites for regulatory proteins are boxed and labeled. Transcription start is at +1, delineates the boundary between U3 and R, and is marked by a horizontal arrow. Polyadenylation cleavage site is marked with a vertical arrow and delineates the boundary between R and U5. TAR was originally defined via mutagenesis, but only the RNA stem-loop is required.

interactions are high affinity and do not require the loop-binding factor to support Tat function.

There are also other sequences in the LTR that are involved in promoter regulation (Figure 4). These include three NFAT sites (Jones *et al.*, 1986), two NF–κ B binding sites (Nabel and Baltimore, 1987), and a LEF binding site (Sheridan *et al.*, 1995). Along with an RNA binding site for Tat, TAR also encompasses an initiator sequence at the start of transcription and the initiator of short transcripts (IST) (Ratnasabapathy *et al.*, 1990; Sheldon *et al.*, 1993). Both the initiator and IST function as DNA sites and presumably bind initiation factors (Zenzie-Gregory *et al.*, 1993; Sheldon *et al.*, 1993). While these additional regulatory sites do not appear critical for *trans*–activation, they may increase the level of activation that can be achieved.

The ability of Tat to *trans*-activate the LTR promoter has been measured by placing the chloramphenicol acetyl transferase (CAT) gene after the LTR in one plasmid and expressing Tat from a second plasmid (Gendelman *et al.*, 1986). When the two plasmids are co-transfected into HeLa or COS cells, CAT activity can be measured through an enzymatic assay (Chapter 5). Such assays show dramatic increases in CAT activity when the Tat expressing plasmid is co-transfected along with the reporter plasmid. Without the Tat expressing plasmid the level of CAT activity is barely detectable.

In addition, nuclear run-on experiments have been performed to investigate the transcriptionally active regions downstream of the LTR promoter (Kao *et al.*, 1987; Selby *et al.*, 1989; Feinberg *et al.*, 1991; Marciniak and Sharp, 1991). Without Tat, transcription of sequences within 100 nt of the start of transcription is high but the level of transcription decreases rapidly further down the template. Around nt 1500 the number of transcription complexes is just 2% of the number at 100 nt (Marciniak and Sharp, 1991). The transcription complexes that make it this far are processive and capable of transcribing all the way around the reporter plasmid. When the level of transcription in

the presence of Tat is compared, 20% of the complexes that transcribe 100 nt are capable of transcribing 1500 nt (Marciniak and Sharp, 1991). There was no observed increase in the amount of 100 nt transcription.

These results indicate that the LTR promoter is an active promoter with many transcripts being initiated, but few being completed. The virus is capable of quickly initiating transcription at the LTR promoter, but the vast majority of these complexes are incapable of productive transcription. Only in the presence of Tat are RNA polymerase II complexes capable of transcribing the entire HIV genome.

Control of Transcription Elongation

Until recently, control of transcription of eukaryotic elongation was relatively unknown. Prokaryotic examples were known (Chamberlin, 1995), but only a few eukaryotic genes were suspected of being controlled at the level of elongation (Spencer and Groudine, 1990; Chamberlin, 1995). The majority of transcriptional activators are known to function by increasing the level of transcriptional initiation. Several genes in eukaryotic cells now have been shown to affect the ability of RNA polymerase II to transcribe through transcriptional pause sites and extend the ability of the polymerase to elongate.

Three general elongation factors have been identified in mammalian cells that have been shown to increase the overall rate at which RNA polymerase II transcribes. SII is a single ~38 kDa protein that promotes the read-through of transcriptional pause and arrest sites (Reines *et al.*, 1989; SivaRaman *et al.*, 1990). SII accomplishes this through a reiterative cleavage and re-extension of the nascent transcript in the ternary complex. The cleavage occurs primarily in 2 nt increments and is followed by a repositioning of the catalytic site of the polymerase (Izban and Luse, 1992). This allows RNA polymerase II to re-initiate transcription after encountering transcriptional impediments. SII is able to greatly reduce the frequency of stalled and aborted transcripts in *in vitro* experiments and increase the amount of full-length transcription products.

TFIIF is a heterodimer of ~70 kDa (RAP74) and ~30 kDa (RAP30) proteins. TFIIF is a required general transcription factor that aids in the accurate initiation of transcription. This factor has been shown to have DNA helicase activity. During elongation, TFIIF suppresses transient pausing by RNA polymerase II (Tan *et al.*, 1995).

Elongin is a heterotrimer composed of ~110 kDa (elongin A), ~18 kDa (elongin B), and ~15 kDa (elongin C) proteins (Aso *et al.*, 1995). Elongin appears to function by suppressing the transient pausing by RNA polymerase II at various sites along the template DNA. Control of elongin function has been tied to familial von Hippel-Lindau (VHL) disease (Duan *et al.*, 1995; Kibel *et al.*, 1995), a rare predisposition to a variety of cancers. Affected individuals have a germline mutation in the VHL gene, a tumor suppressor gene that binds the two smaller subunits of elongin. When VHL binds to elongin BC it may prevent elongin A from binding elongin BC and thus inhibit the formation of an active elongin complex (Duan *et al.*, 1995; Kibel *et al.*, 1995). Not only does the VHL protein prevent elongin A from binding, but the pVHL-elongin BC complex binds Hs-CUL-2, a member of the Cdc53 family (Pause *et al.*, 1997). Of the VHL families, 70% have mutations in the VHL gene that may disrupt Hs-CUL-2 binding to the pVHL-elongin BC trimer (Pause *et al.*, 1997). How the disruption of either elongin A binding to elongin BC or Hs-CUL-2 binding to pVHL-elongin BC leads to transformation is unknown.

Biochemical Experiments on Tat *trans*-Activation

A variety of biochemical techniques have been used to study Tat function. Among these have been nuclear run-on transcriptions, nuclear extract fractionation and *in vitro* transcription, Tat affinity columns, Tat cross-linking, TAR affinity columns, and TAR cross-linking. Several proteins that may cooperate with Tat have been isolated using these techniques (Wu *et al.*, 1991; Marciniak *et al.*, 1990; Gatignol *et al.*, 1991; Gatignol *et al.*, 1989; Gaynor *et al.*, 1989; Kashanchi *et al.*, 1994; Jeang *et al.*, 1993; Nelbock *et al.*,

1990; Desai *et al.*, 1991; Herrmann and Rice, 1993), but none has yet been shown to be a critical factor for Tat function. This is partly due to the lack of a highly defined *in vitro* assay for Tat *trans*-activation. The crude systems available are insufficient to prove whether these factors are absolutely required.

As described earlier, nuclear run-on experiments show a distinct transcription polarity in the absence of Tat (Feinberg *et al.*, 1991; Marciniak and Sharp, 1991). Oligonucleotide probes complementary to the 5' end of the transcript anneal to a large amount of RNA, showing that transcription initiation from the LTR promoter is active. In contrast, probes to the 3' end of the message show that there is little transcription. When Tat is present in the nuclear run-on assays, the levels of 5' and 3' transcripts are nearly the same, demonstrating that Tat modifies the transcriptional apparatus such that the polymerase is able to reach the end of the template.

Addition of excess Tat squelches transcription, suggesting that Tat binds cellular factors required for RNA polymerase II (Song *et al.*, 1994). Some of the factors that bind or are cross-linked to Tat include TBP (Kashanchi *et al.*, 1994), Sp1 (Jeang *et al.*, 1993), TBP-1 (Nelbock *et al.*, 1990), TAK (Yang *et al.*, 1996; Herrmann and Rice, 1995), and a 36-kDa factor (Desai *et al.*, 1991). Of these, interactions with TBP (TATA-binding protein) and Sp1 are unusual in that they are DNA-bound transcription factors whereas Tat is normally bound to RNA and functions poorly when bound to DNA (Berkhout *et al.*, 1990; Southgate and Green, 1991). TBP-1 (Tat binding protein), isolated by its affinity with Tat, is a nuclear protein which when over expressed suppresses Tat *trans*-activation (Nelbock *et al.*, 1990). A family of factors has homology to TBP-1 (Ohana *et al.*, 1993); some of these increase Tat *trans*-activation. The 36-kDa nuclear protein was isolated using Tat peptide chromatography and was shown to allow modest *trans*-activation in rodent cells when microinjected along with a plasmid containing an LTR reporter and another encoding Tat (Desai *et al.*, 1991). Rodent cells are normally non-permissive for Tat *trans*-activation (Hart *et al.*, 1989) because they lack the appropriate

TAR loop-binding factor found in mammalian cells (Alonso *et al.*, 1992). The mechanism of function of the 36-kDa factor is unknown.

TAK (Tat-associated kinase) was isolated by its kinase activity when associated with Tat (Herrmann and Rice, 1995; Yang *et al.*, 1996). TAK has been shown to phosphorylate the C-terminal domain (CTD) of RNA polymerase II (Herrmann and Rice, 1995), the hyperphosphorylation of which is correlated with actively transcribing polymerases and with processive transcription. Cloning and sequencing of TAK has shown it to be PITALRE, a previously known kinase. It also has homology with P-TEFb (positive transcription elongation factor b) (Zhu *et al.*, 1997), a nuclear factor from *Drosophila* that has CTD-kinase activity (Marshall and Price, 1995) and is involved in an early step in transcription. Immunodepletion of TAT in nuclear extracts abolishes Tat *trans*-activation, but add-back experiments with purified TAK have not been shown to restore Tat activity (Zhu *et al.*, 1997; Mancebo *et al.*, 1997).

Trans-activation can be specifically inhibited by the purine nucleoside analog 5,6-dicloro-1-β-D-ribofuranosylbenzimidazole (DRB) (Marciniak and Sharp, 1991) that decreases the elongation properties of RNA polymerase II. DRB also inhibits the CTD-kinase property of P-TEFb (Marshall *et al.*, 1996). A screen for drugs that inhibit Tat *trans*-activation *in vitro* identified inhibitors of protein kinases. When these compounds were tested for specificity it was found that they all strongly inhibited P-TEFb CTD-kinase activity (Mancebo *et al.*, 1997).

Attempts to characterize factors that bind TAR have failed to isolate the cellular loop-binding factor. The more promising candidates, are the previously mentioned 36-kDa nuclear factor (Desai *et al.*, 1991), RNA polymerase II, and TRP-185 (Wu *et al.*, 1991; Wu-Baer *et al.*, 1995; Wu-Baer *et al.*, 1996). TRP-185 was isolated by fractionation of HeLa nuclear extracts and TAR gel mobility assays. While RNA polymerase II requires both the loop and the bulge sequences, TRP-185 requires only the loop sequence from TAR. Both require cellular cofactors to bind TAR, making it unclear

whether RNA polymerase II and TRP-185 make direct contact with TAR RNA, or whether they are associated through these cofactors.

In vitro transcription assays have been used in attempts to isolate Tat cofactors. The general transcription factors required for normal activated transcription are insufficient for Tat mediated *trans*-activation (Zhou and Sharp, 1995). Addition of nuclear extracts can restore *trans*-activation, suggesting that Tat does not interact with the common transcription factors. Fractionation of the nuclear extracts has isolated Tat-SF1 (Tat simulatory factor) which when added back to purified transcription reactions is able to allow Tat *trans*-activation (Zhou and Sharp, 1995). Associated with Tat-SF1 is a kinase that phosphorylates Tat-SF1 in the presence of Tat (Zhou and Sharp, 1996). It has been suggested that the kinase might phosphorylate the CTD, but this has not yet been demonstrated.

TFIIH has been implicated in Tat *trans*-activation by its isolation as a CTD kinase that is stimulated by Tat (Parada and Roeder, 1996; Garcia-Martinez *et al.*, 1997). The hyperphosphorylation of RNA polymerase II C-terminal domain has been correlated with actively transcribing polymerase (O'Brien *et al.*, 1994). This TFIIH kinase activity is different from the TAK described above and is sensitive to DRB (Parada and Roeder, 1996; Garcia-Martinez *et al.*, 1997). A peptide fragment of CDK8 that inhibits TFIIH kinase (CDK7) inhibits Tat *trans*-activation in both *in vitro* and *in vivo* assays, but it also may inhibit a general step in transcription (Cujec *et al.*, 1997).

BIV Tat-TAR

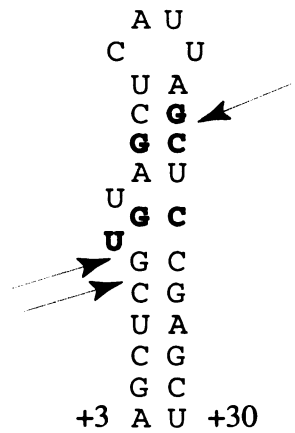
The bovine immunodefficiency virus (BIV) is a lentivirus related to HIV causing persistent lymphocytosis, lymphadenopathy, central nervous system lesions, weakness, and emaciation. BIV contains the structural genes in common with HIV and other retroviruses, as well as regulatory genes that are specific to lentiviruses. Analogues to vif,

tat, rev, vpr, and vpu/vpx have either been isolated or inferred. No nef gene, which is specific to primate lentiviruses, has been found (Gonda, 1992).

The BIV Tat gene is 103 amino acids long and contains a cystine-rich region, a core domain, and a basic region similar to HIV Tat (Garvey *et al.*, 1990; Fong *et al.*, 1997). The cystine-rich and core domains share 54% amino acid homology with the corresponding domains in HIV Tat. BIV Tat is capable of weakly activating the HIV LTR promoter through HIV TAR (Liu *et al.*, 1992; Fong *et al.*, 1997). The BIV Tat arginine-rich domain (SGPRPRGTRGKGRRIRR) spans amino acids 65-81. This region is responsible for BIV Tat binding to BIV TAR (Chen and Frankel, 1994).

The BIV Tat-TAR interaction differs from the HIV Tat-TAR interaction in important ways. Mutagenisis of BIV Tat shows that eight residues are important for BIV Tat-TAR binding (Figure 5) (Chen and Frankel, 1995), whereas in HIV Tat, a single arginine makes a sequence specific contact to HIV TAR. BIV Tat has an isoleucine at position 79 whereas there are no hydrophobic residues in the HIV Tat RNA-binding domain. Biochemical experiments have suggested that the hydrophobic nature of position 79 is important for BIV Tat binding (Chen and Frankel, 1995). NMR also suggests that there are hydrophobic contacts made by I79 to a bulge U at position 10 of BIV TAR (Puglisi *et al.*, 1995). There is also the possibility that I79 may help to stabilize R73 and R77 through contacts to the aliphatic parts of the side chains. There are also critical glycines in BIV Tat suggesting that they may help BIV Tat bind deeply into the major groove of the RNA (Chen and Frankel, 1995). NMR has shown that BIV Tat binds to its TAR in a beta-sheet conformation, making 9 specific contacts with both bases and phosphates in the RNA (Puglisi *et al.*, 1995). HIV Tat, on the other hand, makes a single specific contact through R52 (Calnan *et al.*, 1991b). Even though BIV and HIV Tat proteins perform similar functions their RNA binding properties are quite distinct. The BIV interaction provides advantages for studying the mechanism of *trans*-activation, as described in this thesis.

A



B

RP**RGTRG**K**GR**R**I**RR

65 81

Figure 5. BIV TAR and BIV Tat RNA-Binding Domain
(A) Primary and secondary structure of BIV TAR. Nucleotides +3 – +30 were used to replace HIV TAR upper stem-loop (HIV TAR +15 – +46) in reporter constructs. The nucleotides required for specific binding of BIV Tat are shown in bold, and the phosphates whose ethylation interferes with binding are indicated by arrows (Chen and Frankel, 1994). (B) Sequence of BIV Tat RNA-binding domain (BIV Tat amino acids 65-81). Amino acids whose mutation causes decreased binding are shown in bold (Chen and Frankel, 1995).

CHAPTER 2 — CELLULAR ANALOGUE OF TAT?

Viruses Take Advantage of Cellular Machinery

Many viral proteins appear to be derived from cellular proteins. Viral genomes are highly compact; viruses take advantage of cellular machinery for many of their replicative and control functions (Dimmock and Primrose, 1994). The HIV Rev protein, for instance, uses a cellular nuclear export machinery to export full-length un-spliced viral messages out of the nucleus before splicing can occur (Malim *et al.*, 1989). The SV40 large T antigen binds to the viral DNA (Johnston *et al.*, 1996) perhaps allowing RNA polymerase II to initiate quickly at SV40 promoters. It allows DNA replication independent of cellular factors that are normally required for host cell replication by interacting with DNA polymerase and the SV40 origin of replication (Dimmock and Primrose, 1994).

Viruses have provided many insights into cellular pathways and machinery. The *trans*-activation of the HIV LTR has been difficult to establish, but as described previously, many experiments suggest that the LTR may be controlled by a previously un-characterized elongation mechanism. Only recently has work on other systems suggested that control of transcription elongation may be an important regulatory mechanism for eukaryotic cells (Spencer and Groudine, 1990; Krumm *et al.*, 1993; Duan *et al.*, 1995). Given that transcription of several cellular genes is controlled at the elongation step and that *trans*-activation of the LTR involves switching products from short to full-length, the possibility that HIV has taken advantage of a cellular regulatory mechanism to control its own transcription seems reasonably high.

Several observations support the hypothesis that Tat takes advantage of a cellular elongation control mechanism. HIV Tat functions poorly when an HIV Tat-Gal4 hybrid is co-transfected with an HIV LTR reporter containing five Gal4 binding sites in the LTR (Southgate and Green, 1991). Only the VP16 activation domain functions when bound to

the RNA via HIV Rev-RRE protein-RNA interaction (Madore and Cullen, 1995). These data suggest that activators of initiation do not generally function when bound to RNA and enhancers of elongation do not function when bound to DNA. This may explain why few examples of elongation control are known—to date few have searched for proteins that control transcription when bound to RNA.

It is possible to imagine Tat functioning in several ways. Tat might simply recruit a cellular factor involved in elongation control to the transcriptional machinery, thereby making RNA polymerase II more processive. Tat might also function by mimicking an existing cellular elongation control protein. Such a protein might enhance or activate transcription of specific genes and might be expected to bind RNA polymerase and modify the complex to increase its processivity.

It has been difficult to pursue these questions largely because Tat control of transcriptional elongation has only been observed in mammalian cells; attempts to reconstruct LTR activation in yeast have met with little success (Subramanian *et al.*, 1994). A human cell line is the only setting where the full *trans*-activation of the HIV LTR can be examined. As described here, using the green fluorescence protein (GFP) reporter gene will allow easy detection of activated LTR promoters through fluorescence microscopy and fluorescence activated cell sorting (FACS) (Chalfie *et al.*, 1994). A screen can then be established to search a cDNA library for proteins that can function the way Tat does, by delivering the cDNA library to the HIV LTR promoter via the same route that Tat is delivered.

Targeting the Tat Activation Domain Through the N-Terminus

As a first step in the design of such a library, I first asked whether Tat could be delivered to the promoter using an RNA binding domain located at the opposite end of the protein from its normal location. Heterologous protein-RNA interactions have been used previously to target the HIV transcriptional activation domain to the nascent RNA by

fusing the RNA binding domain onto Tat and replacing HIV TAR with the RNA target (Southgate *et al.*, 1990; Selby and Peterlin, 1990). All these experiments have fused the RNA binding domain to the C-terminus of Tat, either replacing Tat's basic domain or simply appending the new binding domain to the end. In all cases the HIV Tat activation domain was at the N-terminus, however for cDNA libraries it will be advantageous to position the RNA binding domain at the N-terminus.

Placing the library before the RNA binding domain creates two possible frameshift locations. A consensus Kozak sequence (Kozak, 1986) and a start codon are required in the expression vector to allow cDNA clones to be expressed, since randomly primed cDNAs will not necessarily carry their own translation initiation sequences. There would be, therefore, one possible frameshift position between the translation start sequences and the cDNA library, and another between the cDNA library and the RNA binding domain (Figure 6A). In addition, there is the possibility that the cDNA would have a translation stop at the end of the protein sequence, thereby preventing expression of the fused RNA binding domain.

To avoid some of these problems, the expression vector has been designed to fuse the cDNA library to the C-terminus of the RNA binding domain (Figure 6B). This orientation has one possible frameshift location because the RNA binding domain will also have the translation start site and there is only one junction between the open reading frame and cloned sequence. In addition, stop codons at the C-terminus of the cloned sequence will not prevent expression of the RNA binding domain. It is still possible that stop codons might occur between the RNA binding domain and the open reading frame of the cloned sequence, particularly if a long 5' untranslated region is cloned.

I designed a screen to identify proteins that when targeted to the HIV LTR via RNA are able to *trans*-activate transcription (Figure 7). The HIV LTR has only been observed to function in mammalian cells, so experiments must be performed in these cells. Several difficulties arise when using mammalian tissue culture cells for genetic

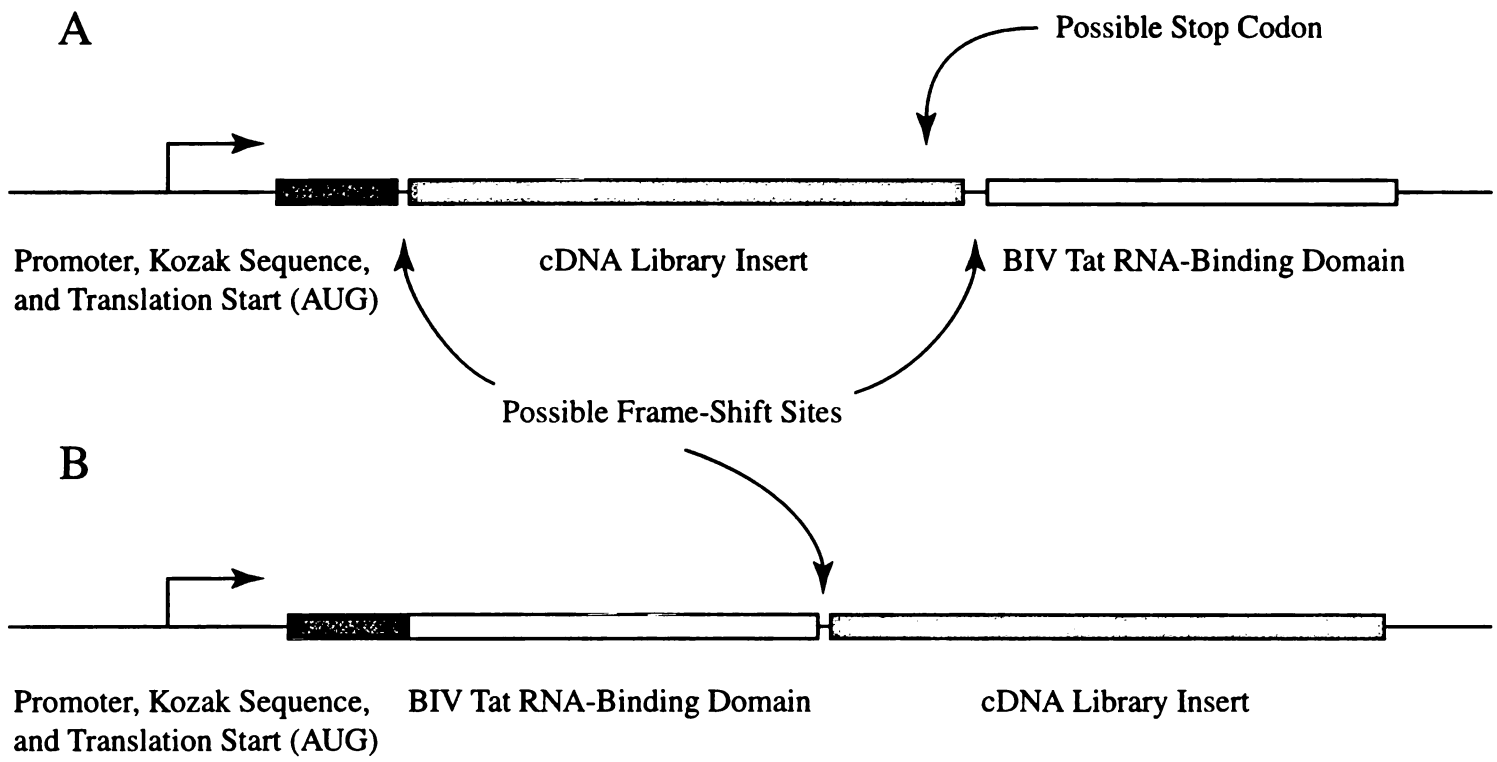


Figure 6. Possible Orientation of cDNA Library Fusion
(A) Placing the cDNA library before the RNA-binding domain will create two possible frame shift sites and have the possibility of naturally occurring stop codons terminating translation before the RNA-binding domain is expressed. (B) With the cDNA library at the C-terminus of the fusion there is only one possible frame shift site and naturally occurring stop codons would not interfere with protein expression.

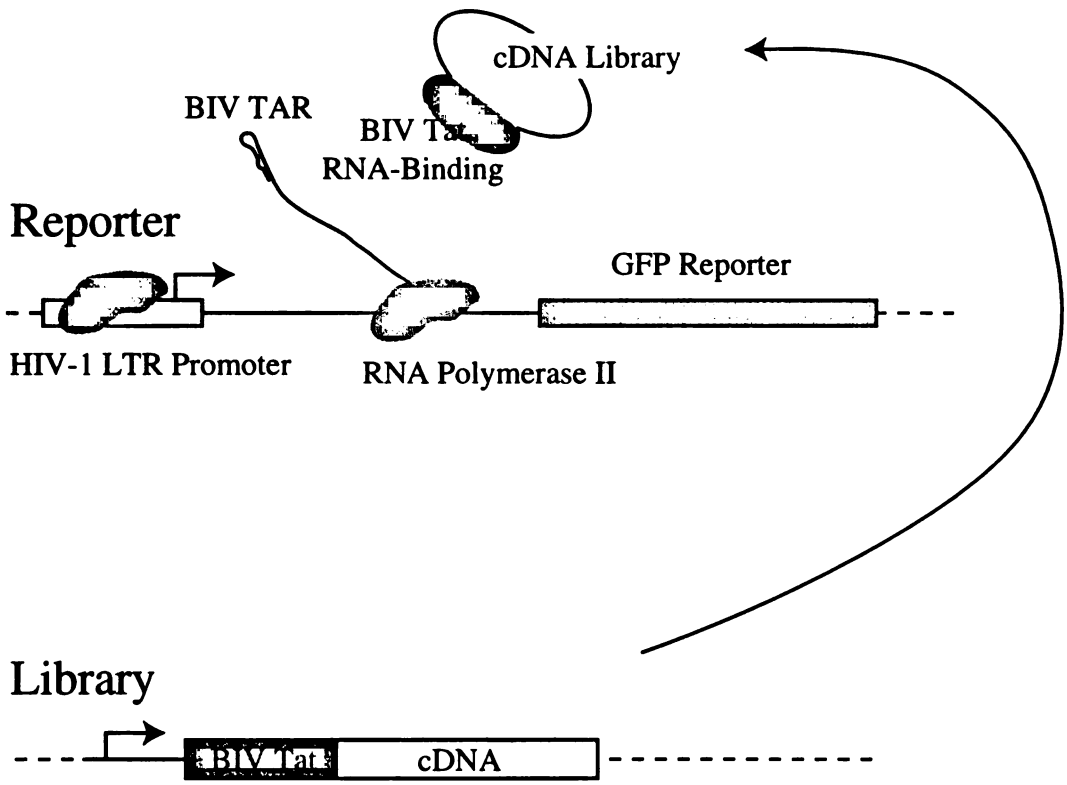


Figure 7. Screen for Tat-Like Factors.
Cartoon of cDNA library screen for factors that are able to trans-activate the HIV LTR promoter when targeted to the nascent transcript. The BIV Tat-TAR protein-RNA interaction is used to bind the cDNA to the RNA by fusing the BIV RNA-binding domain to the N-terminus of the cDNA library. GFP expression is then dependent upon binding a cellular homologue of Tat to TAR. The reporter is stably integrated into the cell line and the library is introduced by protoplast fusion.

screens, but a method developed by R. Tan in the lab overcomes some of these difficulties (Tan and Frankel, 1997). The method uses protoplast fusion to deliver a large number of clonal plasmids to a mammalian tissue culture cell, and is carried out by removing the bacterial cell wall and fusing the resulting protoplasts to tissue culture cells through a brief PEG exposure (Weiss, 1978; Tan and Frankel, 1997). This procedure has been used in some expression cloning strategies to deliver plasmids to tissue culture cells and to make stable cell lines (Seed and Aruffo, 1987; Sambrook *et al.*, 1989).

Plasmid delivery by protoplast fusion is essentially clonal. Approximately 15% of HeLa cells can be transformed with plasmids that constitutively express GFP, and clonal delivery was demonstrated by fusing mixtures of protoplasts containing GFP and CD4 expression plasmids. Fusing equal mixtures to HeLa cells showed that very few cells expressed both GFP and CD4 (Tan and Frankel, 1997). Since the number of protoplasts is in vast excess to HeLa cells, it appears that the ability of protoplasts to fuse is the limiting factor in plasmid delivery.

For this screen, a reporter was designed to use an HIV LTR promoter (HIV nucleotides -453 – +80, relative to the start of transcription) driving expression of GFP. The upper portion of the TAR stem-loop (HIV nucleotides +14 – +44) was replaced with the corresponding BIV Tat-binding stem-loop (BIV nucleotides +3 – +33). A stable HeLa cell line was constructed by R. Tan in which the HIV LTR-BIV TAR-GFP was inserted into pcDNA3, transformed into HeLa cells, and selected using G418 (Tan and Frankel, 1997). This reporter line has been successfully tested using an HIV Tat_{1-48} activation domain fused to the BIV Tat_{65-81} basic domain ($pSV_2Tat_{1-48}BIV_{65-81}$). In addition, the reporter plasmid has been tested in transient transfections in which the reporter and Tat were delivered by lipofection.

The HIV-BIV Tat fusion has the HIV transcriptional activation domain at the N-terminus and the BIV RNA-binding domain at the C-terminus (Chen and Frankel, 1994). As described, the cDNA library fusions will be placed at the C-terminus. Therefore, I

tested whether the HIV Tat activation domain and the BIV Tat basic domain could be switched by expressing BIV Tat_{65-81}-HIV Tat_{1-48}. All previous work has been with the HIV Tat activation domain at the N-terminus of the protein, and while there is no obvious reason why this orientation should not work, I expressed those fusions with several different linker lengths between the basic domain and the activation domain.

At first only three different linkers were tested. Two, three, and four glycine linkers were positioned before a NotI cloning site and HIV Tat_{1-48} and Tat_{1-72}. The cDNA library was designed to be cloned into a NotI site. In addition, HIV Tat_{2-48} was also cloned into the constructs to determine if the additional AUG codon caused problems. Oligonucleotides that contained the consensus Kozak sequence, initiation codon, and BIV Tat_{64-81}, glycine linkers, NotI site, and transcriptional termination codons in all three reading frames were cloned into the NotI site of a pcDNA3 vector (plasmids pcBb2G, pcBb3G, and pcBb4G).

HIV Tat_{1-48}, Tat_{1-72}, and Tat_{2-48} were cloned as PCR fragments (Sambrook *et al.*, 1989) using pSV_2tat- as the template. The PCR products were purified (Qiagen QIAquick gel extraction kit) and cut with EagI, which has a recognition sequence contained within the NotI recognition sequence and leaves the same overhang. The EagI fragments from the Tat_{1-48}, Tat_{1-72}, and Tat_{2-48} PCRs were cloned (Sambrook *et al.*, 1989) into all the pcBb vectors and sequenced (USB Sequenase v2).

The results with these fusions are shown in Figure 8. The CAT assay (see Chapter 5 for details) shows that Tat_{1-48} fusions trans-activated the HIV LTR better than fusions with Tat_{1-72} or Tat_{2-48}. It is possible that the HIV basic domain of Tat_{1-72} interferes with the BIV Tat basic domain binding to BIV TAR. For each version of Tat, the level of CAT activity was greatest with the 4 glycine linker followed by the 2 glycine linker and then the 3 glycine linker. Because CAT activity does not increase or decrease systematically with linker length, there may be some conformational problem not simply related to linker length. The 4 glycine linker is presumably the longest and most flexible, but it still

Figure 8. Linkers of 2, 3, and 4 Glycines.
CAT assays of BIV Tat_{65-81}-Glycine linker-Tat *trans*-activators. Activator constructs are diagramed, followed by the fold *trans*-activation (activity relative to the no activator CAT assay) (lane 12). See Chapter 5 for a description of the CAT assay. Two hundred and fifty ng of each activator plasmid were tested against 25 ng of pHIVLTR-BIVTAR-CAT reporter plasmid. Five µl of extract were assayed for 1 hour. Two, 3, and 4 glycine linkers ("gly"s in sketch) were inserted between the BIV RNA-binding domain (in gray) and the HIV Tat activation domain (in white). Three versions of the HIV Tat protein were tested: full-length protein, Tat_{1-72}; transcription activation domain, Tat_{1-48}; and activation domain without the initial AUG codon, Tat_{2-48}. Full-length HIV Tat_{1-72} does not trans-activate the BIV TAR reporter; Tat_{1-48}-BIV Tat_{65-81} strongly *trans*-activates the reporter.

	Activator Construct	fold *trans*-Activation	CAT Assay
1	BIV Tat_{65-81} -gly-gly- Tat_{1-72}	15	
2	BIV Tat_{65-81} -gly-gly- Tat_{1-48}	97	
3	BIV Tat_{65-81} -gly-gly- Tat_{2-48}	50	
4	BIV Tat_{65-81} -gly-gly-gly- Tat_{1-72}	12	
5	BIV Tat_{65-81} -gly-gly-gly- Tat_{1-48}	54	
6	BIV Tat_{65-81} -gly-gly-gly- Tat_{2-48}	44	
7	BIV Tat_{65-81} -gly-gly-gly-gly- Tat_{1-72}	3	
8	BIV Tat_{65-81} -gly-gly-gly-gly- Tat_{1-48}	97	
9	BIV Tat_{65-81} -gly-gly-gly-gly- Tat_{2-48}	66	
10	Tat_{1-72}	0.7	
11	Tat_{1-48} BIV Tat_{65-81}	260	
12	No Activator	1	

did not approach the level of *trans*-activation when the domains were in the opposite orientation (HIV Tat_{1-48}-BIV Tat_{65-81}).

Additional linkers were next tested (Figure 9). The sequence of the BIV Tat basic domain was extended to the end of the gene (BIV Tat_{65-81}-gly-gly-BIV Tat_{82-103}) followed by the NotI cloning site into which Tat_{1-48} was cloned. Also, a linker from the Oct1 homeodomain protein was used ($Oct1_{72-81}$, or NLSSDSSLSS) which has been shown to be highly flexible. Crystal structures of Oct1 bound to its DNA target showed no electron density for the linker sequence (Klemm *et al.*, 1994) and the second DNA binding domain is capable of binding its DNA site on either upstream or downstream of the first DNA binding site (Cleary and Herr, 1995). Additionally, the Oct1 sequence was cloned at the end of the BIV Tat_{65-81}2G construct, so the expressed protein was BIV Tat_{65-81}-2G-Oct1-NotI site-HIV Tat_{1-48}. The NotI site introduced PGGR into the protein sequence. Of these constructs the BIV Tat_{65-81}-gly-gly-BIV Tat_{82-103} worked best, showing 70% *trans*-activation compared to HIV Tat_{1-48}-BIV Tat_{65-81}. These results suggest that having a more flexible linker does not necessarily allow Tat_{1-48} to better *trans*-activate the LTR promoter. It is possible that the flexible linker sequence inhibits *trans*-activation, but it is not clear how this might be.

For cloning reasons, the BIV Tat_{65-81}-gly-gly-BIV Tat_{82-103} described above contained 2 glycines between residues 81 and 82. BIV Tat_{65-103} was remade without glycines and was compared with the previous constructs (Figure 10). *Trans*-activation was slightly higher with the two additional glycines, but was still lower than $pSV_2Tat_{1-48}BIV_{65-81}$. In addition, the Oct1 sequence was cloned between the BIV Tat82-103 and the HIV Tat1-48 to increase the linker length. In the construct with the additional glycines it appeared to increase trans-activation, but it did not do so in the construct with the Oct1 sequence. The BIV Tat_{65-103} fusions use the CMV promoter for expression while the $pSV_2Tat_{1-48}BIV_{65-81}$ uses the SV40 promoter, possibly accounting for some differences in the levels of activation.

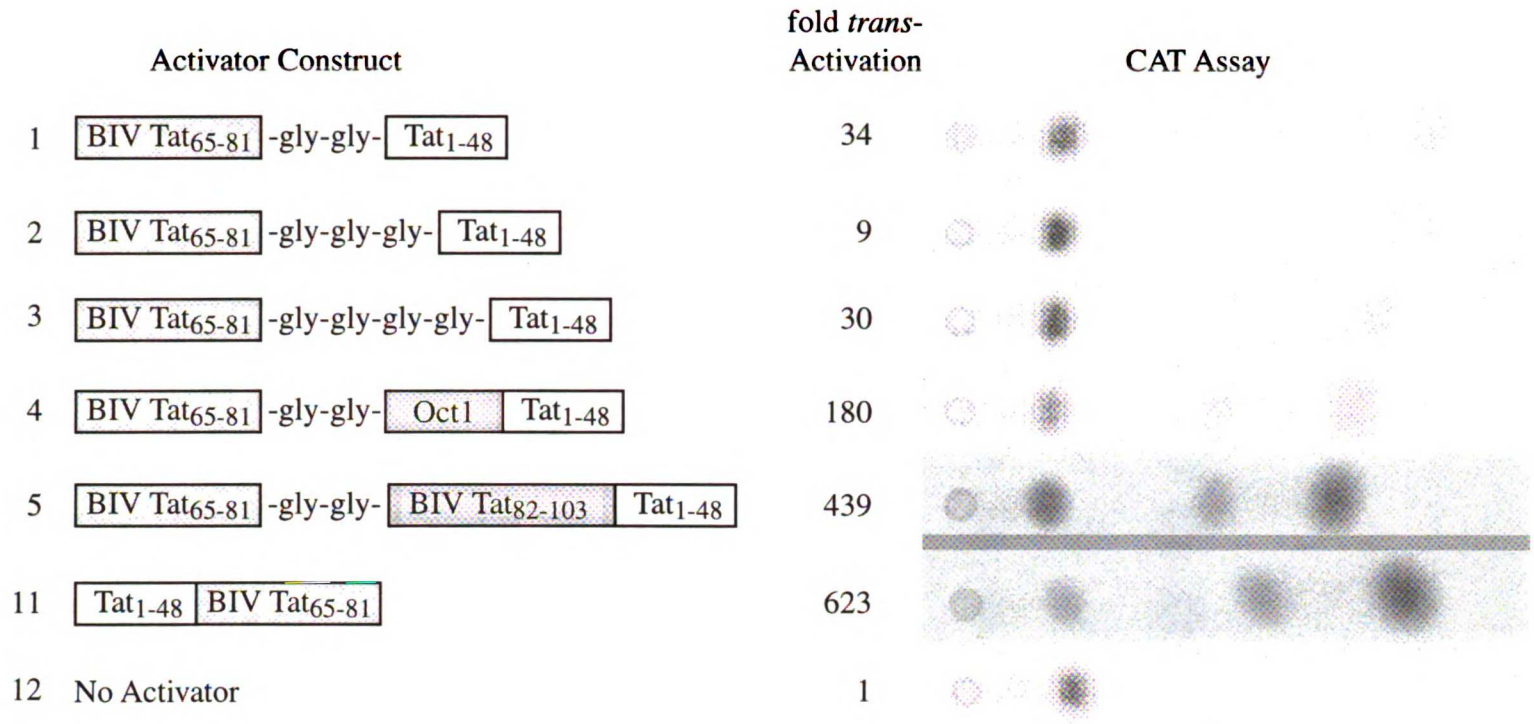


Figure 9. Comparison of Different Linker Lengths.
CAT assays of BIV Tat RNA-binding domain-HIV Tat activation domain with five different linker lengths. Shown are diagrams of the activator constructs, the fold *trans*-activation relative to no activator (lane 12), and the TLC of the CAT assay. Assays were performed using 25 ng of HIV LTR-BIV TAR-CAT reporter plasmid and 50 ng of activator plasmid. Assays were performed on 5 µl of extract for 1 hour. Lanes 1, 2, and 3 are CAT assays of 2, 3, and 4 glycine linkers, respectively. Lane 4 is an assay of a 2 glycine-Oct1 linker. Lane 5 is an assay of a 2 glycine-BIV Tat_{82-103} linker. Lane 11 is an assay of HIV Tat activation domain-BIV Tat RNA-binding domain. Lane 12 had no activator.

	Activator Construct	fold *trans*-Activation	CAT Assay
1	No Activator	1	
2	Tat_{1-48} | BIV Tat_{65-81}	316	
3	BIV Tat_{65-81} | BIV Tat_{82-103} | Tat_{1-48}	207	
4	BIV Tat_{65-81} -gly-gly- BIV Tat_{82-103} | Tat_{1-48}	262	
5	BIV Tat_{65-81} | BIV Tat_{82-103} | Oct1 | Tat_{1-48}	248	
6	BIV Tat_{65-81} -gly-gly- BIV Tat_{82-103} | Oct1 | Tat_{1-48}	186	

Figure 10. Removal of Glycines from Linkers.
CAT assays of BIV Tat RNA-binding domain-HIV Tat activation domain removing the glycine linkers in the BIV Tat_{82-103} and the BIV Tat_{82-103} -Oct1 linker constructs. Shown are diagrams of the activator constructs, the fold *trans*-activation, and the TLC from the CAT assay. Assays performed using 25 ng of HIV LTR-BIV TAR-CAT reporter plasmid and 50 ng of activator plasmid. Assays were performed on 5 µl of extract for 1 hour. Lane 2 shows the CAT activity when the HIV Tat activation domain was in its normal N-terminal location. Four linkers were tested between the BIV Tat RNA-binding domain at the N-terminus and the HIV Tat activation domain at the C-terminus. Lane 3 had the BIV Tat_{82-103} directly fused to BIV Tat_{65-81}. Lane 4 had an additional 2 glycine linker between the BIV Tat domains. Lane 5 had the same BIV Tat sequence as lane 3 but had an additional Oct1 linker after the BIV sequence. Lane 6 was the same as lane 5 but had an additional 2 glycine linker between the BIV Tat domains.

The plasmid encoding the activator BIV Tat_{65-103}-HIV Tat_{1-48} was transiently transfected into HeLa cells along with the HIV LTR-BIV TAR-GFP reporter and the cells analyzed with FACS (Figure 11). The background without an activator was 1.1% GFP positive cells. The activator with the BIV Tat RNA-binding domain at the N-terminus (11.5% positive) functioned just as well as the activator with the RNA-binding domain at the C-terminus (13.3% positive). These results suggest that the construct BIV Tat_{65-103}-HIV Tat_{1-48} trans-activated the HIV LTR promoter almost as well as the HIV Tat_{1-48}-BIV Tat_{65-81}.

These results show that the HIV Tat transcriptional activation domain can be delivered to the nascent RNA through an RNA-protein interaction at the N-terminus of the activation domain. While activity is not quite as high, these results demonstrate that there is no strict spatial arrangement required for the activation of the HIV LTR and that a cellular protein might be able to activate the LTR when targeted to the nascent RNA. While the library screen could have been performed with the RNA binding domain placed at the C-terminus of the cDNA, the placement at the N-terminus avoids potential problems that might require screening more clones to achieve the same complexity.

Performing the Screen

I have not completed the screen, but will briefly describe the plan. The design called for the manufacture of a cDNA library from HeLa cell poly(A)$^+$ RNA primed using random hexamers (Chanda, 1997). NotI adaptors are added to blunt-ended cDNAs and then inserted into the NotI site in pcBbB103 (BIV Tat_{65-103}-NotI site-Stop codons in pcDNA3 backbone) creating fusions at the C-terminus of the BIV RNA-binding domain. This library is transformed into DH5α and the transformants used to make protoplasts (see Chapter 5).

The FACS screen and enrichment procedure are detailed in Chapter 5. Protoplasts are fused to HeLa cells that have the HIV LTR-BIV TAR-GFP reporter stably integrated

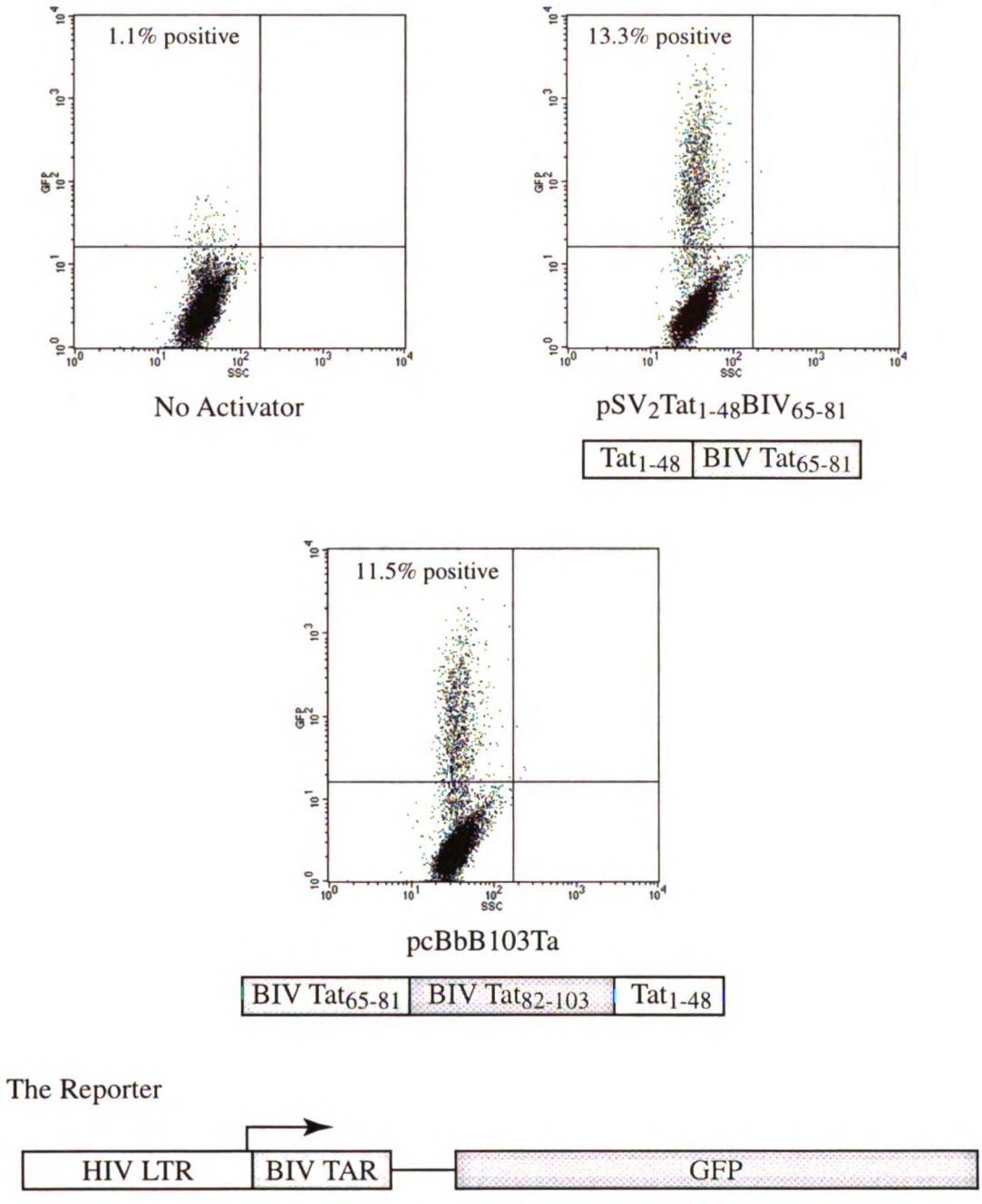


Figure 11. FACS Scans of Lipofectin Transfected Plasmids
Shown are FACS scans of HeLa cells that were lipofectin transfected with the HIV LTR-BIV TAR-GFP reporter (bottom) and activator plasmids simultaneously. The x-ordinate is side scatter (SSC), a measure of the complexity of the cell; the y-ordinate is the GFP fluorescence intensity. Cells in the upper left quadrant were counted as positive GFP-expressing cells. Shown are FACS scans for No Activator (1.1% positive), $pSV_2Tat_{1-48}BIV_{65-81}$ (13.3% positive), and pcBbB103Ta (11.5% positive).

and after 48 hours the HeLa cells are sorted by FACS for GFP expressing cells. The high fluorescing cells are collected and their plasmids recovered. This population of plasmids should be enriched for clones that are able to trans-activate the LTR promoter. These plasmids are then electroporated into freshly made electro-competent DH5α from which protoplasts are made, fused with HeLa reporter cells, and sorted by FACS. This cycle is repeated until the population of plasmids that activate the HIV LTR promoter is above 50%.

Negative screens are performed to remove non-specific clones by testing individual clones against HIV LTR-no TAR-CD4 or HIV LTR (TATA box mutant)-BIV TAR-CD4. These reporters are on pACYC177 backbones, compatible with the cDNA expression vectors. The enriched cDNA library is then transformed into these bacteria and used in a two-plasmid protoplast fusion (explained in Chapter 3 and detailed in Chapter 5). A two-plasmid protoplast fusion will introduce both the negative control reporter plasmid and the cDNA library fusion plasmid to the HeLa cells via the same protoplast. Protoplasts are fused with stable HIV LTR-BIV TAR-GFP reporter HeLa cells and analyzed by FACS. If a cDNA fusion clone activates the expression of both GFP and CD4, it will be discarded as a false positive. After a round to remove false-positives, the remaining clones will be analyzed individually using CAT assays. The clones that are best able to trans-activate the LTR promoter will be sequenced and used for further analysis.

The Loop-Binding Factor Aids Tat in Binding to TAR

Experiments with HIV Tat have shown that the interaction between Tat's basic domain and HIV TAR is not sufficient for *in vivo* function (Feng and Holland, 1988). Loop sequences in TAR are necessary for HIV Tat to function, but are unimportant for *in vitro* binding assays where the interaction of the basic domain and TAR is measured. It appears that the loop-binding factor binds to both the TAR loop sequence and the Tat activation domain.

Mutagenisis experiments with HIV TAR have shown that the distance between the bulge and the loop is important to maintain Tat function (Berkhout *et al.*, 1989). Increasing the distance by just a single base pair greatly reduces the level of *trans*-activation, suggesting that either the spacing between the bulge and the loop is critical, or the orientation between them is important.

J. Tao performed experiments in which the RNA-binding domain of HIV Tat (amino acids 49-57) was replaced with R52, a stretch of lysines with a single arginine at position 52 (Tao and Frankel, 1993). This arginine is responsible for the specific contact between HIV Tat and TAR. The minimal basic domain was determined to be an arginine at position 52 with three charged residues on either side (KKKRKKK, spanning positions 49-55). When the arginine is moved to position 51 or 53, the level of *trans*-activation decreases, and as it is moved further the level of CAT activity continues to decrease (Calnan *et al.*, 1991b; Tao and Frankel, 1993).

L. Chen in the lab then performed a series of experiments in which both the position of the critical arginine was varied and the distance between the HIV TAR bulge and loop was increased (Chen, personal communication). These experiments showed that when the arginine was moved one amino acid away from the activation domain the

optimal *trans*-activation occurred when TAR had one additional base pair between the bulge and the loop. Similarly, when the arginine was moved two amino acids away the optimal *trans*-activation occurred with two additional base pairs in TAR.

These experiments suggest that there is some flexibility in the spacing between Tat's basic domain and the bulge in TAR and the activation domain and its interaction with the loop-binding factor. The type of flexibility may suggest a distinct spatial relationship in the HIV Tat-TAR complex. For example, if a complex can be found that activates transcription when a full helical turn is inserted into TAR, that might suggest a relatively rigid structure and that the loop of TAR must be in a specific orientation to correctly trans-activate the LTR promoter. If many different spacings of TARs and Tat-linkers are possible, that might suggest a less restrictive orientation.

A library experiment has been designed to explore the flexibility of the Tat and TAR interactions. In this screen a protein linker was designed between Tat's activation domain and its RNA binding domain. A linker between the TAR bulge and loop sequences was also designed. While it is technically possible to screen the Tat library against each member of the TAR library individually, it would be more efficient to screen both libraries at the same time, and I have devised a method to do that.

Two-plasmid Protoplast Fusions

A procedure for introducing both the reporter and activator plasmid in the same protoplast (see Chapter 5 for details) was tested using the $pSV_2Tat_{1-48}BIV_{65-81}$ Tat expressor and pAHLBTGa reporter (HIV LTR-BIV TAR-GFP on a pACYC177 backbone). $pSV_2Tat_{1-48}BIV_{65-81}$ uses the pMB1 replicon and pAHLBTGa uses the p15A replicon, compatible origins that allow both plasmids to be stably maintained in the bacterium (Sambrook *et al.*, 1989). The reporter plasmid is on a low copy-number replicon, averaging 15 copies per bacterium. This replicon cannot be amplified by

inhibiting protein production. The pMB1 replicon is derived from pBR322 and is a high copy-number replicon that can be amplified (Clewell, 1972).

DH5α cells were transformed with the pAHLBTGa reporter, selected for transformants, then transformed with $pSV_2Tat_{1-48}BIV_{65-81}$ and selected for doubly transformed clones. A single colony was grown up to prepare protoplasts and fused to HeLa cells (Chapter 5). Figure 12 shows the results of a FACS scan of protoplast fused HeLa cells. The background of GFP-positive cells was 0.3% whereas the protoplast fused sample had 15.2% GFP-positive cells. This demonstrates that the two-plasmid protoplast fusion works well in delivering both the activator and reporter to HeLa cells.

Tat-TAR Double Library Experiment

The concomitant screen of the Tat and TAR libraries was designed using the two-plasmid protoplast delivery system. The initial design was to have the cDNA fusion on the pcDNA3 backbone and the HIV LTR reporter on the pACYC177 backbone. While this is likely to work given the experiments described above, experiments to use a lower copy number pSV_2 backbone have also been planned.

TAR library

The TAR library will consist of 11 different TARs with different numbers of base pairs between the Tat binding site and the loop-factor binding site. The A-form helix will be extended by inserting G•C or C•G base pairs between G28•C37 and C29•G36 in the upper stem of TAR. HIV Tat binds in the major groove at the UCU bulge making specific contacts with G26. Mutations to the upper stem seem to have minor effects, as long as the C•G base pair just before the loop sequence is maintained. This base pair is presumably needed to clamp the upper stem and maintain the loop in a particular conformation for the loop-binding factor.

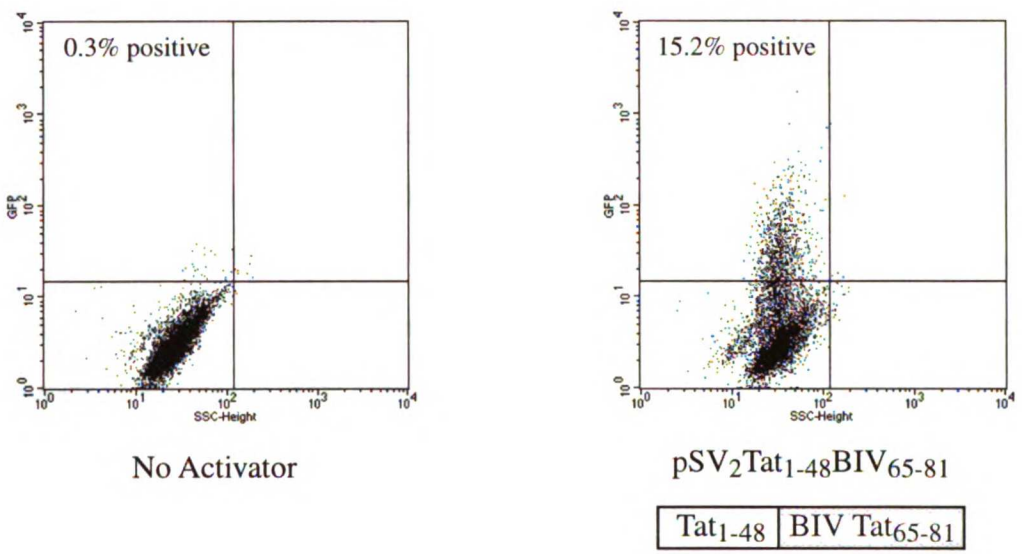


Figure 12. FACS Scans of Two Plasmid Protoplast Fusions.
Protoplasts containing both the activator (below each plot) and the reporter (bottom) plasmids were fused to HeLa cells. Scatter plots of the FACS scans are shown where the x-ordinate is side scatter (SSC), a measure of cellular complexity and the y-ordinate is GFP fluorescence intensity. Cells in the upper left quadrant were counted as positive GFP-expressing cells. Shown are FACS scans of protoplast fusions with no activator (0.3% positive) and $pSV_2Tat_{1-48}BIV_{65-81}$ (15.2% positive).

Oligonucleotides encoding each TAR (with 1 – 11 base pairs inserted) will be synthesized and cloned into the reporter HIV LTR-MCS-GFP (HIV LTR-multiple cloning site-GFP on the pACYC177 backbone), as described in Chapter 5 and listed in Table 1. Each TAR construct will be cloned and transformed into DH5α individually allowing the library to be screened in two ways. Each TAR can be screened against the Tat linker library individually, or all TARs can be mixed together and screened at the same time. While one of the goals of this project is to demonstrate the feasibility of screening two libraries at the same time, it is important to determine whether screening just one library identifies the same clones.

Tat Library

The linker region between the Tat activation domain and the basic domain will consist of a library of 1 – 8 amino acids of glycine, alanine, or proline. Glycines were chosen because they are small and have greater flexibility than other amino acids, alanines for their neutral side-chain and helical propensity, and prolines to add the ability to form rigid structures. This linker will have more sequence complexity than poly-glycine and may avoid problems with poly-glycines observed in the lab in which Tat expression levels may be reduced (D. Campisi). The linker length will be varied to allow the shortest linker to be selected. A random library of 1 – 8 positions consisting of these three amino acids will have a complexity of 9,840 ($\sum_{i=1}^{8} 3^i$).

The Tat library will be constructed by modifying a previously used protocol (Harada *et al.*, 1996) to search for peptides that can bind HIV RRE stem-loop IIB in bacterial cells. Oligonucleotide synthesis will be carried out in special columns that can be opened and the resin removed. This allows a codon-based rather than nucleotide-based library design, which more evenly samples sequence space (described in Chapter 5). Briefly, the resin will be split into three columns after each linker codon, in this case either synthesizing glycine, alanine, or proline. After synthesis of one codon, the resins

are mixed together and the procedure repeated for all linkers positions. After the library codons have been synthesized, the oligonucleotides are processed as normal, the complementary strand is synthesized using an oligonucleotide primer and Sequenase v2 (Harada *et al.*, 1996), the fragments are cut with BsiWI and SgrAI and are purified on an agarose gel. The fragments are then ligated into a prepared vector.

Unlike the previous library, the Tat library will have variable randomized lengths. Each length has its own complexity, and in order for all the different lengths to be equally represented the longer lengths must be in greater proportion than the shorter lengths. The one amino acid linker has a complexity of 3, and therefore should be 3/9,840 of the total population. Likewise, for linkers with seven amino acids the total complexity should be 2,187, and should be 2,187/9,840, or ~22.2%, of the population. Chapter 5 has a discussion of this issue and some solutions that allow even representation.

The TAR library complexity is 11 and the Tat linker library complexity is 9,840, making the total complexity 108,240, or 1.1×10^5. Libraries are commonly screened such that there is a 99% probability of screening a given clone. Chapter 5 has a description of the method to calculate the number of clones that need to be screened for a given probability of screening a given clone. For this double library with its 1.1×10^5 complexity 498,462 clones (5.0×10^5) need to be screened, within the limits of FACS analysis (~1×10^6 cells per hour) assuming a 10% protoplast fusion efficiency.

Simultaneous Screen of Two Libraries.

The design of the concomitant screen of the Tat and the TAR libraries follows the same general plan as the cDNA library screen described in Chapter 2, but because two libraries are to be screened there are some significant differences. The design of the two-library screen is outlined below in brief.

The transformation of bacteria, protoplast preparation, FACS sorting, and plasmid recovery are described in Chapter 5. After plasmids are recovered from GFP expressing

cells, fresh DH5α are transformed and are allowed to grow in a large volume of LB (no antibiotics) for 4 hours, or approximately 2 doublings. The bacteria are then concentrated and split into 2 portions, one to be plated on LB agar plates with the antibiotic to select for the TAR containing plasmids, and the other to be plated onto plates with the antibiotic to select for Tat encoding plasmids. The bacteria are allowed to grow until colonies are just visible and then washed off the plates and inoculated into liquid culture. A large scale plasmid prep is performed on a saturated culture and each type of plasmid is isolated on gels. Plasmids carrying the TAR library are larger than the Tat encoding plasmids. Each plasmid is then retransformed and the screen repeated until the fraction of GFP expressing cells is significant. Combinations of individual plasmids identified in the final round will be tested to determine the corresponding pairs that show activity.

CHAPTER 4 — DISCUSSION

This thesis describes two genetic screens in mammalian tissue culture cells to study trans-activation by HIV Tat. The first, in Chapter 2, is an attempt to identify cellular proteins that are able to replace Tat in trans-activating the LTR promoter. By tethering the cDNA library to the BIV Tat RNA-binding domain it is possible to target the cDNA library to BIV TAR in the nascent transcript without the complication of the cellular loop-binding factor. The second screen, in Chapter 3, was designed to simultaneously screen two libraries. The Tat molecules encoded in the first library, which have spacers between the activation domain and the RNA-binding domain resulting in poor binding to wild-type TAR, are screened against a library of TAR molecules that have altered spacings between the bulge and the loop sequences. This technique has the potential for being used to screen libraries of RNA and proteins against each other in a search for RNA-protein interactions.

cDNA Library Screen

The search for factors able to trans-activate the HIV LTR promoter when targeted to the nascent transcripts is in part an attempt to identify proteins potentially involved in the control of elongation of transcription. It was hypothesized that Tat takes advantage of a cellular control mechanism for elongation control and that by targeting a cDNA library to the location that Tat functions it might be possible to identify components of the control machinery. The experiments described in Chapter 2 show that such a screen is possible. The HIV Tat activation domain can be targeted to the 5' end of the transcript through BIV Tat-TAR RNA-protein interaction located at the N-terminus of Tat_{1-48}. Having the cDNA library at the C-terminus of the fusion prevents the premature termination of the protein before the RNA-binding domain can be translated. It also

eliminates a possible frameshift junction that might be present if cDNAs were placed at the N-terminus.

There are several classes of proteins one might expect from such a screen. Simple recruitment of a general transcription factor may be all that HIV needs; recruitment of one of the transcription factors may nucleate the formation of transcription complexes at the promoter. The phosphorylation of the CTD of RNA polymerase II has been shown to be correlated with actively transcribing polymerases (O'Brien *et al.*, 1994), and with processive polymerases (Lee and Greenleaf, 1997). If the limiting step in HIV transcription is poor phosphorylation of the CTD then binding of a CTD kinase to TAR may allow the formation of processive polymerase complexes. Several studies suggest the involvement of such kinases (Herrmann and Rice, 1995; Marshall *et al.*, 1996; Parada and Roeder, 1996; Garcia-Martinez *et al.*, 1997; Cujec *et al.*, 1997; Zhu *et al.*, 1997). Another possible step is the opening of duplex DNA. While there is no direct evidence for duplex sequence slowing RNA polymerase II it seems plausible that a transcription complex that lacks the ability to separate the strands might transcribe poorly. A DNA helicase, therefore, might help such complexes become more processive.

In addition to screening a fusion library for factors capable of trans-activating the LTR promoter, such a library might also be used to screen for proteins that have other functions. For example, if BIV TAR is placed in the middle of a spliced message, the library might be screened for proteins that inhibit splicing by including GFP in an intron. Such a screen might isolate factors involved in splicing or factors that actively export the un-spliced message from the nucleus. BIV TAR might also be placed between two open reading frames in a bi-cistronic message. Normally, the second open reading frame would not be translated, but if BIV TAR is placed before the second open reading frame (GFP, for example) a screen could be performed to identify factors involved in internal ribosome entry.

The ability to position the BIV Tat_{65-103} RNA-binding domain at the N-terminus of HIV Tat_{1-48} adds more evidence that Tat_{1-48} is an independent activation domain. While wild-type HIV Tat requires a cellular loop-binding factor to bind HIV TAR *in vivo*, that requirement is obviated when a heterologous RNA-protein interaction is substituted (Selby and Peterlin, 1990; Southgate *et al.*, 1990). Tat is able to function, albeit poorly, when bound to a site on the DNA (Berkhout *et al.*, 1990; Southgate and Green, 1991). These experiments suggest that Tat_{1-48} is able to function when it is close to the promoter, and functions optimally when bound to the RNA. The weak interaction between HIV Tat and HIV TAR and the requirement for a cellular factor to aid in binding is probably a regulatory device that the virus has adopted, either to further refine its control of Tat *trans*-activation, or to restrict viral replication to cells that express the cellular loop-binding factor.

Double Library Screen

The concomitant Tat and TAR library screen was designed to find a pair of spacers in both the protein and RNA that supports Tat *trans*-activation. The spacing requirement in the HIV Tat-TAR complex appears relatively strict, because the cellular loop-binding factor is required for tight Tat binding to TAR and the spacing between the bulge and loop appears to be important. The library screen described in Chapter 3 was an attempt to find alternative spacings between the Tat activation domain and the RNA-binding domain. This information may yield insight into the orientation of the activation domain and the loop-binding factor with relation to the RNA-binding domain and TAR.

Experiments in the lab have suggested that poly-glycine linkers may be detrimental to Tat activity, perhaps due to expression difficulties. Alanines and prolines were included in the Tat linker library to help overcome such problems and to add structural variety to the linker. The modest complexity of the Tat linker library should give enough possible linkers to allow the distance between the bulge and the loop to be

spanned. The 8 amino acid linker in Tat would be able to span ~30.4 Å (3.8 Å x 8) (Pauling, 1960) if the peptide bonds were in an extended conformation. This might be sufficient to reach across the 30.8 Å for 11 base pairs of A-form duplex RNA.

The technique of screening two libraries against each other has several additional applications. In particular, the plasmids developed for this screen may be adapted to screen for protein-RNA interactions. BIV TAR can be replaced with either random sequence, or with a specific RNA sequence library, and the BIV RNA-binding domain can be replaced with a random peptide sequence, or with peptide sequences of interest. The two libraries can then be screened against one another for RNA-protein interactions. If a protein-RNA pair binds with a similar affinity as BIV Tat-TAR, cells will express GFP and can be sorted by FACS.

Several types of libraries can be screened in this manner. Of notable interest is the possibility of screening viral genomes for RNA-protein interactions. It should be possible to fragment a retroviral genome into ~200 nt fragments and clone them into the multiple cloning site in the HIV LTR-MCS-GFP reporter construct. An activator library would then be constructed with the fragments cloned after Tat_{1-48}. Thus, putative RNA-binding domains wold be located after the activation domain for the same reasons that the cDNA library was designed to be placed after the BIV RNA-binding domain (preventing one additional frameshift location and eliminating fortuitous stop codons prematurely terminating translation; see Figure 6). This design would have fragments of the entire RNA genome at the 5' end of the nascent transcript and fragments of all viral proteins attached to the Tat activation domain. If any viral RNA-protein interactions exist, this screen should detect them.

One advantage of the two-plasmid protoplast fusion technique is that stable cell lines are not needed for the types of screens described, saving several months of cloning. Another advantage is highlighted by the screen described in Chapter 3; while stable lines could be established for all 11 TAR lengths, there would be no way to determine if each

reporter was working correctly because there are no forms of Tat that are able to trans-activate the altered TAR reporters. Thus, by using the two-plasmid procedure it is possible to avoid making stable lines and to use reporters that do not have positive controls.

This second feature is also important for counter screens against false positives. A counter screen described in Chapter 2 is to screen the enriched cDNA pool against a non-functional BIV TAR. An HIV LTR-mutated BIV TAR-CD4 would have no specific activator and would express CD4 if the cDNA is capable of activating transcription without being targeted to the nascent transcript. Using the two-plasmid protoplast fusion to introduce both the cDNA library and this mutated BIV reporter allows the screen to sort out CD4 expressing cells while still selecting for GFP expression from the integrated reporter. This procedure can also be used to test the specificity of the cDNA activator for the LTR promoter by introducing an HIV LTR mutant-BIV TAR-CD4 along with the cDNA library.

Preparing Protoplasts

Plasmids are transformed into *E. coli* and selected on appropriate antibiotic LB agar plates. A single colony is used to inoculate 5 ml LB cultures with antibiotics. After overnight growth at 37°C with agitation, 0.5 ml of stationary culture is used to inoculate a 50 ml LB culture (Tan and Frankel, 1997); these cultures are grown at 37°C with agitation to OD_{600} = 0.7 – 0.8, and chloramphenicol is added to a final concentration of 250 μg/ml. Chloramphenicol inhibits bacterial protein synthesis and as a result inhibits the replication of the bacterial genome but does not inhibit replication of plasmids with pMB1 or ColE1 replicons (Clewell, 1972). These plasmids continue to replicate for 10-12 hours and reach copy numbers as high as 3000 copies per cell. At this point the bacteria are chilled to 0°C, pelleted for 10 minutes at 2000 x g, re-suspended in 10 ml 50 mM Tris-Cl (pH 8.0) at 0°C, and re-centrifuged.

Using the method of Sandri-Goldin (Sandri-Goldin *et al.*, 1984) with further refinements (Tan and Frankel, 1997), the bacterial outer cell wall is removed to form protoplasts. The pelleted bacteria are suspended in 2.5 ml 20% sucrose, 50 mM Tris-Cl (pH 8.0), and 0.5 ml 5 mg/ml lysozyme (freshly prepared in 250 mM Tris-Cl (pH 8.0)), and the suspension incubated for 5 minutes on ice. Then 1 ml 250 mM EDTA•Na_2 is slowly added and the cells incubated for an additional 5 minutes on ice. Next 1 ml 50 mM Tris-Cl (pH 8.0) is added slowly and the mixture incubated at 37°C until all the bacteria are converted to protoplasts. Protoplast conversion is monitored by phase-contrast microscopy; rod shaped bacteria become converted to round protoplasts, usually in 15 minutes for DH5α bacteria. At this point 20 ml DMEM, 10% sucrose, and 10 mM $MgCl_2$ at room temperature are added slowly and carefully mixed. After a 15 minute

incubation at room temperature the solution is ready for fusion. This final mixture contains ~1.5 x 10^9 protoplasts per ml.

Protoplast Fusion

Protoplast fusion is performed using HeLa cells containing an appropriate HIV LTR reporter grown in 6-well plates to ~70% confluence (Tan and Frankel, 1997). The old medium is aspirated and the cells washed with 4 ml DMEM per well. Two or 4 ml of the prepared protoplasts are added to each well and the protoplasts centrifuged at 1650 x g for 10 minutes at room temperature onto the surface of the HeLa cells. At this point the protoplasts are clearly visible in the HeLa culture as small black points covering the HeLa cells. The supernatant is carefully aspirated and 2 ml of 37°C 50% (w/v) PEG 1500 (polyethylene glycol 1500 average molecular weight) in DMEM is added carefully and allowed to sit on the cells for no more than 120 seconds. The PEG solution causes membrane fusion and long exposure is toxic to the cells. The PEG solution is promptly aspirated and the cells washed carefully three times with 2 ml DMEM. Finally, 4 ml of DMEM-10 (10% (v/v) FBS, penicillin, streptomycin, and spectinomycin in DMEM) are added and the cells and returned to the 37°C, 5% CO_2, and 100% relative humidity incubator. The antibiotic spectinomycin is added to kill the un-fused bacteria, contrary to previous protocols where kanamycin was used, because plasmids with pACYC177 replicons also had the kanamycin resistance gene. The medium is changed after 24 hours and the cells allowed to grow for another 24 hours before examining GFP fluorescence.

The layer of protoplasts should not be disturbed by the PEG exposure and subsequent washes. If the washes are too vigorous then not only would the protoplast layer be removed, but many of the HeLa cells become detached and wash away. Spectinomycin present in the DMEM is sufficient to kill the remaining protoplasts.

Two-Plasmid Protoplast Fusion

Protoplast fusion where the bacteria harbor two compatible plasmids differs only slightly from the above protocol. After the first plasmid is transformed into DH5α and transformants selected on LB agar plates with the appropriate antibiotics, the bacteria are grown in liquid culture to OD_{600} = 0.5 and are made competent for electroporation (see below). The second plasmid is then transformed into the bacteria and double transformants are selected on LB agar plates with the two appropriate antibiotics. Single colonies are then used to inoculate 5 ml LB overnight cultures. For library screens the electroporated bacteria are plated onto 150 mm plates at a targeted density of 100,000 colonies per plate and grown for 6 hours to select for double transformants. Bacteria are then collected by washing the plates with 5 ml of LB and the pooled mixture is used directly to inoculate 50 ml LB cultures.

One concern with the growth of DH5α cells containing two-plasmids was that some cultures did not grow well in the 50 ml cultures and the efficiency of protoplast fusion was generally low. If a culture had difficulty growing the frequency of observed GFP fluorescence was generally 1/10 of the expected value. Occasionally, no fluorescence was observed using positive controls that should have given 20% fluorescencing cells.

FACS

HeLa cells either transfected or fused to protoplasts are grown for 48 hours and then collected by trypsinization. No visible bacteria remain in the culture. The HeLa cells are pelleted at 1650 x g for 10 minutes and re-suspended in 10% cell dissociation buffer (GIBCO/BRL), 1 μg/ml propidium iodide in PBS. The cells were kept on ice until they were scored by FACS.

The FACSTAR$^+$ cell sorter (Becton Dickinson) FACS machine at the Howard Hughes Medical Institute at UCSF was used for all these studies. An argon 488 nm laser is used to excite the GFP and the fluorescence is detected through a 575 ± 12.5 nm band pass filter. FACS scans are performed on 5,000-10,000 cells and FACS sorts are performed at 2,000-3,000 cells per second. Propidium iodide-stained cells are detected by the FACS and these cells are sent into a waste stream. Propidium iodide stains DNA, but if the plasma membrane of the cell is intact, no staining occurs. Threshold level for GFP fluorescence is set by scanning 10,000 negative control cells and setting the threshold level above the level of background fluorescence. Cells that were not stained with propidium iodide and had higher than background levels of GFP fluorescence were directed into a eppendorf tube containing 0.5 ml DMEM. FACS data were analyzed using CellQuest software on Macintosh computers.

Plasmid Recovery

Plasmids from FACS sorted cells are recovered by alkaline lysis (Tan and Frankel, 1997). Briefly, cells are pelleted with 20,000 carrier HeLa cells at 15,000 x g for 5 minutes at 4°C. The cell pellet is re-suspended in 10 µl TE (pH 8.0) containing 0.2 mg/ml tRNA as a carrier and the cells lysed by adding 20 µl 1% SDS, 0.2 N NaOH. After a 5 minute incubation on ice 15 µl 3 M NaOAc (pH 4.8) is added and the mixture incubated on ice for an additional 10 minutes. The cell membranes and denatured proteins are centrifuged at 15,000 x g for 5 minutes at 4°C and the supernatants transferred into fresh tubes. The supernatants are then extracted with an equal volume of phenol:chloroform::1:1 pH 7.5, followed by a chloroform extraction. DNA is precipitated with 1 µl of 20 mg/ml glycogen (Sigma), 0.1 volume of 3 M NaOAc (pH 4.8), and 3 volumes of absolute ethanol. The solution is mixed and incubated at -80°C for 1 hour, and then centrifuged at 15,000 x g for 30 minutes. Pellets are then washed with 1 ml 70% ethanol (-20°C) and air dried. The pellet is re-suspended in 1 µl distilled water.

Fluorescence Microscopy

Analysis of GFP expression was also examined by fluorescence microscopy. HeLa cells are grown on glass cover slips in 6-well plates and either transfected with lipofectin or fused to protoplasts. After 48 hours of growth, the coverslips are removed from the wells and placed cell-side down on glass microscope slides. Cell fluorescence was observed on a phase-contrast Olympus fluorescence microscope using a 515 nm long pass filter.

Electroporation

The appropriate bacteria are colony purified on LB agar plates and a single colony is used to inoculate 25 ml LB overnight cultures. Ten ml of this overnight culture are used to inoculate 500 ml LB cultures in 2.8 L fernback flask and the culture is incubated at 37°C with vigorous agitation until the $OD_{600} = 0.5$. The culture is then chilled quickly in an ice bath and left for 10 minutes to cool completely. To achieve high transformation efficiencies it is important to keep the bacteria chilled from this step forward, so all manipulations are performed in the cold room (4°C) and all equipment is pre-chilled.

Electroporation (Sambrook *et al.*, 1989) requires that the bacterial suspension have very high resistance to keep the current passing though the sample low. To achieve this the majority of free ions in the culture are removed through repeated washings with water. The bacteria are pelleted in 250 ml conical centrifuge tubes (Corning 25350-250) and centrifuged using a Sorvall RC-3B centrifuge and a H6000A rotor at 5 krpm for 20 minutes at -2°C. The supernatants are removed by aspiration and the pellet is re-suspended in 100 ml mqH_2O (de-ionized water purified through a Mili-Qplus PF filter unit to 18.2 MΩcm) at 0°C. The bacteria are pooled into a single centrifuge tube, re-centrifuged and the supernatant aspirated. This bacterial pellet is washed thrice in 250 ml mqH_2O at 0°C with centrifugations done as before. The final pellet is re-suspended in an equal volume of mqH_2O and used directly for electroporation. If the bacteria are to be

frozen, then 10% glycerol at 0°C is used and 100 µl aliquots are frozen slowly in a -80°C freezer. Freezing decreased the transformation efficiency by about 50%.

CAT Assays

HeLa cells are grown in 12-well plates and transiently transfected with lipofectin according to the manufacturer's directions. The cell extract was made and CAT assays performed as described (Calnan *et al.*, 1991a). Forty-eight hours post transfection the medium is removed and the cells washed with 2 ml PBS per well and the media removed. The cells are detached from the wells by adding 1 ml per well of 1 mM EDTA•2Na in TN (40 mM Tris-Cl, pH 7.5; 150 mM NaCl) and allowing the cells to sit for ~5 minutes. The cells are removed by pipetting and transferred to 1.5 ml eppendorf tubes. The wells are washed with 400 ml TN and added to the eppendorf tubes. The cells are centrifuged at 7 krpm in a table-top eppendorf centrifuge for 30 seconds and the media removed by aspiration. The cells are suspended in 1 ml TN, gently re-suspended, and centrifuged again at 7 krpm for 30 seconds and the media removed. The cells are then suspended in 100 µl 250 mM Tris-Cl, pH 7.5 by gentle vortexing and are lysed through three freeze-thaw cycles in ethanol-dry ice and 37°C water baths. Cell debris is removed by centrifugation for 5 minutes in a table-top centrifuge and the supernatant transferred to a new 1.5 ml eppendorf tube.

The level of CAT in the extract is determined using an enzymatic assay (the transfer of acetyl groups from acetyl co-A onto chloramphenicol). The cell extract is diluted with 250 mM Tris-Cl, pH 7.5 (usually 5:45::extract:Tris) and used in the following reaction:

2 µl	200µCi ^{14}C Chloramphenicol / ml (New England Nuclear)
20 µl	10 mg Acetyl Co-A / 3 ml (Sigma)
25 µl	1 M Tris-Cl, pH 7.5
53 µl	dH_2O
50 µl	diluted extract
150 µl	total volume

Reactions are typically incubated at 37°C for 1 hour. Chloramphenicol is extracted using 500 µl ethyl acetate and vortexing and separating the phases through a brief centrifugation. Then 400 µl of the ethyl acetate is transferred to new eppendorf tubes and the ethyl acetate removed through a 2 hour spin in a Speed-Vac (Sorval) with vacuum. The remaining chloramphenicol is re-suspended in 10 µl fresh ethyl acetate and spotted on thin layer chromotography (TLC) plates (silica gel 60, 20 x 20 cm, 0.2 mm, from EM separations) and run in chromatography chambers equilibrated with 190 ml chloroform and 10 ml methanol and a 25 x 25 cm sheet of 3MM Whatman paper. When the solvent is ~1 cm from the top of the TLC plate the plate is removed from the chamber, air dried, and exposed to X-ray film overnight. Alternatively, the TLC plates are exposed to phosphoimager screens for 2 hours and the screens read by a Molecular Dynamics Storm 680 Phosphoimager. Chloramphenicol migrates slowly on the TLC. An acetyl group can be added in two positions on the molecule; each acetylated form has a slightly different mobility and both migrate faster than chloramphenicol. The di-acetylated form migrates the fastest, but it is not always observed.

Quantitation is performed using phosphoimager scans and Molecular Dynamics ImageQuant software on Apple Macintosh computers. The level of CAT activity is determined by summing the acetylated chloramphenicol levels and dividing that value by the total amount of chloramphenicol for that sample. Fold *trans*-activation is calculated by dividing the CAT activity of a given sample with the CAT activity for the "No Activator" sample.

TAR Libraries

TAR libraries were designed to be inserted into HIV LTR-Multiple Cloning Site-GFP (a gift of R. Tan) reporter plasmids. These reporters contain the HIV LTR promoter with TAR replaced with a multiple cloning site of 11 sites, 3 of which are unique in both the

pcDNA3 and the pACYC177 backbones. The TAR libraries and positive controls were designed to be cloned into the AflII – BamHI.

TAR libraries were to be synthesized using complementary oligonucleotides and cloned into the AflII-BamHI site in the HIV LTR MCS GFP reporter constructs. The oligonucleotide pairs for each length upper stem are shown in Table 1.

Tat Library

Tat linker libraries were designed to be inserted into HIV Tat derived from pSV_2tat- (the source of Tat_{1-72} in our lab). The basic region has few restriction sites and the closest AvrII site in the core region cannot be used if the Tat gene is placed into the pcDNA3 vector because there is an AvrII site at nucleotide position 2069 in pcDNA3. The GCG Map program (Genetics Computer Group, 1997) was used to search for restriction sites in HIV Tat. The program identified restriction sites that do not cut in pcDNA3, or in pSV2tat-, the source of the HIV Tat sequence. In the core region a BsiWI site (C/GTACG) is 2 nucleotides from the core-basic region junction and in the basic region an SgrAI site (CR/CCGGYG) is 19 nucleotides from the junction. The sites were introduced by synthesizing complementary oligonucleotides that spanned both restriction sites. These oligonucleotides were then used in PCR (Sambrook *et al.*, 1989) with either T7 or Sp6 primers to generate overlapping PCR products. These were then purified and used in a subsequent PCR with only the T7 and Sp6 primers to amplify the final fragment. This fragment contained the entire HIV Tat gene to the T7 and Sp6 sites. This fragment was cut with SanDI (GG/GTCCC) and BamHI (G/GATCC) which cut within the ends of the Tat coding region, and was ligated into pcDNA3HIVTat_{72} cut with SanDI and BamHI and purified. Clones were checked by restriction digests with EcoRI, SanDI, BamHI, BsiWI, and SgrAI. Correct clones were sequenced and one with the correct Tat sequence was chosen and renamed pcDNA3Tat_{72}c*b.

Number of Nucleotides Between the Loop and the Bulge	Oligonucleotides
0	5' TTAAGGAGCTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTCGGACCCTCGAGAGACCGATTCGAGCCTAG 5'
1	5' TTAAGGAGCTTAGACCAGATCTGAG**G**CCTGGGAG**C**CTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTC**C**GGACCCTC**G**GAGAGACCGATTCGAGCCTAG 5'
2	5' TTAAGGAGCTTAGACCAGATCTGAG**GC**CCTGGGAG**GC**CTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTC**CG**GGACCCTC**CG**GAGAGACCGATTCGAGCCTAG 5'
3	5' TTAAGGAGCTTAGACCAGATCTGAG**GCG**CCTGGGAG**CGC**CTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTC**CGC**GGACCCTC**GCG**GAGAGACCGATTCGAGCCTAG 5'
4	5' TTAAGGAGCTTAGACCAGATCTGAG**GCGC**CCTGGGAG**GCGC**CTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTC**CGCG**GGACCCTC**CGCG**GAGAGACCGATTCGAGCCTAG 5'
5	5' TTAAGGAGCTTAGACCAGATCTGAG**GCGCG**CCTGGGAG**CGCGC**CTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTC**CGCGC**GGACCCTC**GCGCG**GAGAGACCGATTCGAGCCTAG 5'
6	5' TTAAGGAGCTTAGACCAGATCTGAG**GCGCGC**CCTGGGAG**GCGCGC**CTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTC**CGCGCG**GGACCCTC**CGCGCG**GAGAGACCGATTCGAGCCTAG 5'
7	5' TTAAGGAGCTTAGACCAGATCTGAG**GCGCGCG**CCTGGGAG**CGCGCGC**CTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTC**CGCGCGC**GGACCCTC**GCGCGCG**GAGAGACCGATTCGAGCCTAG 5'
8	5' TTAAGGAGCTTAGACCAGATCTGAG**GCGCGCGC**CCTGGGAG**GCGCGCGC**CTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTC**CGCGCGCG**GGACCCTC**CGCGCGCG**GAGAGACCGATTCGAGCCTAG 5'
9	5' TTAAGGAGCTTAGACCAGATCTGAG**GCGCGCGCG**CCTGGGAG**CGCGCGCGC**CTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTC**CGCGCGCGC**GGACCCTC**GCGCGCGCG**GAGAGACCGATTCGAGCCTAG 5'
10	5' TTAAGGAGCTTAGACCAGATCTGAG**GCGCGCGCGC**CCTGGGAG**GCGCGCGCGC**CTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTC**CGCGCGCGCG**GGACCCTC**CGCGCGCGCG**GAGAGACCGATTCGAGCCTAG 5'
11	5' TTAAGGAGCTTAGACCAGATCTGAG**GCGCGCGCGCG**CCTGGGAG**CGCGCGCGCGC**CTCTCTGGCTAAGCTCG 3' 3' CCTCGAATCTGGTCTAGACTC**CGCGCGCGCGC**GGACCCTC**GCGCGCGCGCG**GAGAGACCGATTCGAGCCTAG 5'

Table 1. TAR Library Oligonucleotides

HIV TAR inserts with spacing library between the bulge and loop sequences. Additional base-pairs are in bold. Oligonucleotides are shown as annealed complexes for ligation. TAR with no additional base-pairs between the loop and the bulge was not designed to be part of the library screen but is shown for comparison.

Sense strand oligonucleotides were designed that span the BsiWI site in the core region through the SgrAI site in the basic domain. On the 5' side of BsiWI was a clamp sequence and at the 3' side of SgrAI was the complement to the Sp6 primer binding site. Sp6 primers are used to synthesize the complementary strand (Harada *et al.*, 1996) using Sequenase v. 2.0 (USB); there was no significance to the selection of primer sequence. The linker library insert is located between codons for Gly48 and Arg49. The design of the oligonucleotide is:

5' ctaggtatcTCGTACGGC*CGTAAAAAACGTCGTCAGCGCCGGCGctatagtgtcacctaaat 3'

with the linker library inserted at the asterisk (*). The library can be coded for by GGN for glycine, GCN for alanine, and CCN for proline. To avoid creating SgrAI sites (CR/CCGGYG) the library is synthesized using either A or T for the last nucleotide in each codon. This also reduces some of the G/C bias inherent in the library sequence.

The Tat linker library contains linkers of different lengths. One way to construct the library would be to combine the resin from all three columns after synthesis of each codon, setting aside a fractioin for the final synthesis. After the first codon was synthesized, 3/9,840 of the resin would be set aside and the rest would be used to synthesize the second codon. After the synthesis of the second codon 9/(9,840 - 3) of the resin would be removed and set aside. The fraction that was removed was calculated by dividing the complexity of the number of codons synthesized by the total remaining complexity (the total complexity minus what had already been removed). The general formula for calculating the fraction of resin to be removed is:

$$f = (3^n)/(9{,}840 - (\sum_{i=1}^{n} 3^{(i-1)})),$$

where n = the number of codons of the library that had been synthesized. Following the above example, after seven codons of the library had been synthesized 2,187/(9,840 - 1,092) or ~25% of the resin is removed and set aside. This method has the disadvantage that small amounts of resin are required in the early cycles. After the synthesis of the first

codon 0.03% of the resin would have to be removed (about 0.3 µl if the resin was suspended in 1 ml of acetonitrile) which is beyond the limit of pipetting accuracy.

Another method is to synthesize each length linker separately and process each oligonucleotide independently, then combine the libraries in the correct proportions at the time of the protoplast fusion. The bacteria could be easily diluted or concentrated (pelleted) to mix the libraries in the correct ratio. This procedure would entail synthesizing 8 oligonucleotides, one for each length linker. The first oligonucleotide would have 1 amino acid linker, randomized for alanine, glycine, and proline. The second would have a 2 amino acid linker with a complexity of 9, and the n amino acid linker would have a complexity of 3^n. Bacteria containing each library would be mixed in a $1:3:\ldots:3^n$ ratio to achieve the correct representation of each library member.

While this method makes it easy to achieve the correct proportions for each different length of library linker, it entails 8 separate synthesis for each library length. To simplify the synthesis, a combination of these two procedures was chosen for the final protocol to keep the ability to mix the different length libraries in the correct proportions while reducing the synthesis time. The different library lengths were designed to be synthesized in two sets of 3 and one set of 2 using the first technique. After the first length of the three was synthesized 1/9 of the resin would be reserved, and after the second length is finished 1/4 would be removed. After the third length was finished all the resin would be combined and the synthesis finished. The next synthesis would then start at the next linker length and synthesize the next three length linkers. The last synthesis would have only two linker lengths, so half would be reserved after the seven amino acid linker was synthesized. When all oligonucleotides were synthesized, ligated into the vector, and cloned, then they would be mixed in a 1:27:2,187 ratio to achieve the final correct proportions. Dilutions of this magnitude are simple when one is working with bacteria in large volumes of media or with 1/9 of the resin, and only three synthesies are required.

Combinatorial Considerations

Screening two libraries simultaneously requires that the complexity of each library must be smaller than what is possible when only one library is being screened. The capacity of the FACS procedure is ~1 x 10^7 individual HeLa cells per hour. The protoplast fusion efficiency is ~10%, so the library complexity that can be screened is ~1 x 10^6 /hr. Screening the complexity of the library will not guarantee screening all possible elements in the library. In order to determine how many clones need to be examined, probabilities were calculated as follows:

The probability (q) that a given clone is not screened after examining a single element from the library is,

$$q = 1 - 1/N,$$

where N is the total number of clones in the library. After B selections the probability is,

$$q = (q - 1/N)^B.$$

Since q can also be written as,

$$q = 1 - P,$$

where P is the probability that a given clone is examined, we can write the equation,

$$1 - P = (1 - 1/N)^B.$$

Solving for B gives,

$$B = \ln (1 - P) / \ln (1 - 1/N),$$

the number of clones that have to be screened to have the probability P of screening a given clone from a library with N members.

BIBLIOGRAPHY

Alonso, A., D. Derse, and B. M. Peterlin (1992). "Human chromosome 12 is required for optimal interactions between Tat and TAR of human immunodeficiency virus type 1 in rodent cells." *J Virol* **66**(7): 4617-21.

Aso, T., W. S. Lane, J. W. Conaway, and R. C. Conaway (1995). "Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II." *Science* **269**(5229): 1439-43.

Berkhout, B., A. Gatignol, A. B. Rabson, and K. T. Jeang (1990). "TAR-independent activation of the HIV-1 LTR: evidence that tat requires specific regions of the promoter." *Cell* **62**(4): 757-67.

Berkhout, B., R. H. Silverman, and K. T. Jeang (1989). "Tat trans-activates the human immunodeficiency virus through a nascent RNA target." *Cell* **59**(2): 273-82.

Braddock, M., A. Chambers, W. Wilson, M. P. Esnouf, S. E. Adams, A. J. Kingsman, and S. M. Kingsman (1989). "HIV-1 TAT "activates" presynthesized RNA in the nucleus." *Cell* **58**(2): 269-79.

Braddock, M., A. M. Thorburn, A. Chambers, G. D. Elliott, G. J. Anderson, A. J. Kingsman, and S. M. Kingsman (1990). "A nuclear translational block imposed by the HIV-1 U3 region is relieved by the Tat-TAR interaction." *Cell* **62**(6): 1123-33.

Calnan, B. J., S. Biancalana, D. Hudson, and A. D. Frankel (1991a). "Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition." *Genes Dev.* **5**(2): 201-10.

Calnan, B. J., B. Tidor, S. Biancalana, D. Hudson, and A. D. Frankel (1991b). "Arginine-mediated RNA recognition: the arginine fork." *Science* **252**(5010): 1167-71.

Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher (1994). "Green fluorescent protein as a marker for gene expression." *Science* **263**(5148): 802-5.

Chamberlin, Michael J. (1995). New Models for the Mechanism of Transcription Elongation and Its Regulation. *The Harvey Lectures*. New York, Wiley-Liss, Inc. **88:** 1-21.

Chanda, V. B., Ed. (1997). *Current protocols in molecular biology*. New York, John Wiley & Sons, Inc.

Chen, L., and A. D. Frankel (1994). "An RNA-binding peptide from bovine immunodeficiency virus Tat protein recognizes an unusual RNA structure." *Biochemistry* **33**(9): 2708-15.

Chen, L., and A. D. Frankel (1995). "A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA." *Proc Natl Acad Sci U S A* **92**(11): 5077-81.

Chin, D. J., M. J. Selby, and B. M. Peterlin (1991). "Human immunodeficiency virus type 1 Tat does not transactivate mature trans-acting responsive region RNA species in the nucleus or cytoplasm of primate cells." *J Virol* **65**(4): 1758-64.

Cleary, M. A., and W. Herr (1995). "Mechanisms for flexibility in DNA sequence recognition and VP16-induced complex formation by the Oct-1 POU domain." *Mol Cell Biol* **15**(4): 2090-100.

Clewell, D. B. (1972). "Nature of Col E 1 plasmid replication in Escherichia coli in the presence of the chloramphenicol." *J Bacteriol* **110**(2): 667-76.

Cujec, T. P., H. Okamoto, K. Fujinaga, J. Meyer, H. Chamberlin, D. O. Morgan, and B. M. Peterlin (1997). "The HIV transactivator TAT binds to the CDK-activating kinase and activates the phosphorylation of the carboxy-terminal domain of RNA polymerase II." *Genes Dev* **11**(20): 2645-2657.

Desai, K., P. M. Loewenstein, and M. Green (1991). "Isolation of a cellular protein that binds to the human immunodeficiency virus Tat protein and can potentiate transactivation of the viral promoter." *Proc Natl Acad Sci U S A* **88**(20): 8875-9.

Dimmock, N. J., and S. B. Primrose (1994). *Introduction to Modern Virology.* Cambridge, Blackwell Science, Ltd.

Dingwall, C., I. Ernberg, M. J. Gait, S. M. Green, S. Heaphy, J. Karn, A. D. Lowe, M. Singh, and M. A. Skinner (1990). "HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure." *Embo J* **9**(12): 4145-53.

Duan, D. R., A. Pause, W. H. Burgess, T. Aso, D. Y. Chen, K. P. Garrett, R. C. Conaway, J. W. Conaway, W. M. Linehan, and R. D. Klausner (1995). "Inhibition of transcription elongation by the VHL tumor suppressor protein." *Science* **269**(5229): 1402-6.

Feinberg, M. B., D. Baltimore, and A. D. Frankel (1991). "The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation." *Proc Natl Acad Sci U S A* **88**(9): 4045-9.

Feng, S., and E. C. Holland (1988). "HIV-1 tat trans-activation requires the loop sequence within tar." *Nature* **334**(6178): 165-7.

Fong, S. E., J. D. Greenwood, J. C. Williamson, D. Derse, L. A. Pallansch, T. Copeland, L. Rasmussen, A. Mentzer, K. Nagashima, G. Tobin, and M. A. Gonda (1997). "Bovine immunodeficiency virus tat gene: cloning of two distinct cDNAs and identification, characterization, and immunolocalization of the tat gene products." *Virology* **233**(2): 339-57.

Frankel, A. D. (1992). "Activation of HIV transcription by Tat." *Curr Opin Genet Dev* **2**(2): 293-8.

Garcia, J. A., D. Harrich, E. Soultanakis, F. Wu, R. Mitsuyasu, and R. B. Gaynor (1989). "Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation." *Embo J* **8**(3): 765-78.

Garcia-Martinez, L. F., G. Mavankal, J. M. Neveu, W. S. Lane, D. Ivanov, and R. B. Gaynor (1997). "Purification of a Tat-associated kinase reveals a TFIIH complex that modulates HIV-1 transcription." *EMBO J.* **16**(10): 2836-2850.

Garvey, K. J., M. S. Oberste, J. E. Elser, M. J. Braun, and M. A. Gonda (1990). "Nucleotide sequence and genome organization of biologically active proviruses of the bovine immunodeficiency-like virus." *Virology* **175**(2): 391-409.

Gatignol, A., A. Buckler-White, B. Berkhout, and K. T. Jeang (1991). "Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR." *Science* **251**(5001): 1597-600.

Gatignol, A., A. Kumar, A. Rabson, and K. T. Jeang (1989). "Identification of cellular proteins that bind to the human immunodeficiency virus type 1 trans-activation-responsive TAR element RNA." *Proc Natl Acad Sci U S A* **86**(20): 7828-32.

Gaynor, R., E. Soultanakis, M. Kuwabara, J. Garcia, and D. S. Sigman (1989). "Specific binding of a HeLa cell nuclear protein to RNA sequences in the human immunodeficiency virus transactivating region." *Proc Natl Acad Sci U S A* **86**(13): 4858-62.

Gendelman, H. E., W. Phelps, L. Feigenbaum, J. M. Ostrove, A. Adachi, P. M. Howley, G. Khoury, H. S. Ginsberg, and M. A. Martin (1986). "Trans-activation of the human immunodeficiency virus long terminal repeat sequence by DNA viruses." *Proc. Natl. Acad. Sci. U.S.A.* **83**(24): 9759-9763.

Genetics Computer Group (1997). *Wisconsin Package Version 9.1*. Madison, Wisc., Genetics Computer Group, Inc.

Gonda, M. A. (1992). "Bovine immunodeficiency virus." *Aids* **6**(8): 759-76.

Greene, W. C. (1990). "Regulation of HIV-1 gene expression." *Annu Rev Immunol* **8**(453): 453-75.

Harada, K., S. S. Martin, and A. D. Frankel (1996). "Selection of RNA-binding peptides *in vivo*." *Nature* **380**(6570): 175-179.

Harrich, D., J. Garcia, F. Wu, R. Mitsuyasu, J. Gonazalez, and R. Gaynor (1989). "Role of SP1-binding domains in *in vivo* transcriptional regulation of the human immunodeficiency virus type 1 long terminal repeat." *J Virol* **63**(6): 2585-91.

Hart, C. E., C. Y. Ou, J. C. Galphin, J. Moore, L. T. Bacheler, J. J. Wasmuth, S. R. Petteway, Jr., and G. Schochetman (1989). "Human chromosome 12 is required for elevated HIV-1 expression in human-hamster hybrid cells." *Science* **246**(4929): 488-91.

Hauber, J., and B. R. Cullen (1988). "Mutational analysis of the trans-activation-responsive region of the human immunodeficiency virus type I long terminal repeat." *J Virol* **62**(3): 673-9.

Hauber, J., M. H. Malim, and B. R. Cullen (1989). "Mutational analysis of the conserved basic domain of human immunodeficiency virus tat protein." *J Virol* **63**(3): 1181-7.

Herrmann, C. H., and A. P. Rice (1993). "Specific interaction of the human immunodeficiency virus Tat proteins with a cellular protein kinase." *Virology* **197**(2): 601-8.

Herrmann, C. H., and A. P. Rice (1995). "Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal

domain of the large subunit of RNA polymerase II: candidate for a Tat cofactor." *J Virol* **69**(3): 1612-1620.

Izban, M. G., and D. S. Luse (1992). "The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'→5' direction in the presence of elongation factor SII." *Genes Dev* **6**(7): 1342-56.

Jeang, K. T., R. Chun, N. H. Lin, A. Gatignol, C. G. Glabe, and H. Fan (1993). "*In vitro* and *in vivo* binding of human immunodeficiency virus type 1 Tat protein and Sp1 transcription factor." *J Virol* **67**(10): 6224-33.

Johnston, S. D., X. M. Yu, and J. E. Mertz (1996). "The major transcriptional transactivation domain of simian virus 40 large T antigen associates nonconcurrently with multiple components of the transcriptional preinitiation complex." *J. Virol.* **70**(2): 1191-1202.

Jones, K. A., J. T. Kadonaga, P. A. Luciw, and R. Tjian (1986). "Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1." *Science* **232**(4751): 755-9.

Jones, K. A., and B. M. Peterlin (1994). "Control of RNA initiation and elongation at the HIV-1 promoter." *Annu Rev Biochem* **63**(717): 717-43.

Kao, S. Y., A. F. Calman, P. A. Luciw, and B. M. Peterlin (1987). "Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product." *Nature* **330**(6147): 489-93.

Kashanchi, F., G. Piras, M. F. Radonovich, J. F. Duvall, A. Fattaey, C. M. Chiang, R. G. Roeder, and J. N. Brady (1994). "Direct interaction of human TFIID with the HIV-1 transactivator tat." *Nature* **367**(6460): 295-9.

Kibel, A., O. Iliopoulos, J. A. DeCaprio, and W. G. Kaelin, Jr. (1995). "Binding of the von Hippel-Lindau tumor suppressor protein to Elongin B and C." *Science* **269**(5229): 1444-6.

Klemm, J. D., M. A. Rould, R. Aurora, W. Herr, and C. O. Pabo (1994). "Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules." *Cell* **77**(1): 21-32.

Kozak, M. (1986). "Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes." *Cell* **44**: 283-292.

Krumm, A., T. Meulia, and M. Groudine (1993). "Common mechanisms for the control of eukaryotic transcriptional elongation." *Bioessays* **15**(10): 659-65.

Laspia, M. F., A. P. Rice, and M. B. Mathews (1989). "HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation." *Cell* **59**(2): 283-92.

Lee, J. M., and A. L. Greenleaf (1997). "Modulation of RNA polymerase II elongation efficiency by C-terminal heptapeptide repeat domain kinase I." *J. Biol. Chem.* **272**(17): 10990-10993.

Liu, Z. Q., D. Sheridan, and C. Wood (1992). "Identification and characterization of the bovine immunodeficiency-like virus tat gene." *J Virol* **66**(8): 5137-40.

Madore, S. J., and B. R. Cullen (1995). "Functional similarities between HIV-1 Tat and DNA sequence-specific transcriptional activators." *Virology* **206**(2): 1150-4.

Malim, M. H., J. Hauber, S. Y. Le, J. V. Maizel, and B. R. Cullen (1989). "The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA." *Nature* **338**(6212): 254-7.

Mancebo, H. S., G. Lee, J. Flygare, J. Tomassini, P. Luu, Y. Zhu, J. Peng, C. Blau, D. Hazuda, D. Price, and O. Flores (1997). "P-TEFb kinase is required for HIV Tat transcriptional activation *in vivo* and *in vitro*." *Genes Dev* **11**(20): 2633-2644.

Marciniak, R. A., M. A. Garcia-Blanco, and P. A. Sharp (1990). "Identification and characterization of a HeLa nuclear protein that specifically binds to the trans-activation-response (TAR) element of human immunodeficiency virus." *Proc Natl Acad Sci U S A* **87**(9): 3624-8.

Marciniak, R. A., and P. A. Sharp (1991). "HIV-1 Tat protein promotes formation of more-processive elongation complexes." *Embo J* **10**(13): 4189-96.

Marshall, N. F., J. Peng, Z. Xie, and D. H. Price (1996). "Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase." *J Biol Chem* **271**(43): 27176-83.

Marshall, N. F., and D. H. Price (1995). "Purification of P-TEFb, a transcription factor required for the transition into productive elongation." *J Biol Chem* **270**(21): 12335-12338.

Muesing, M. A., D. H. Smith, and D. J. Capon (1987). "Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein." *Cell* **48**(4): 691-701.

Nabel, G., and D. Baltimore (1987). "An inducible transcription factor activates expression of human immunodeficiency virus in T cells." *Nature* **326**(6114): 711-3.

Nelbock, P., P. J. Dillon, A. Perkins, and C. A. Rosen (1990). "A cDNA for a protein that interacts with the human immunodeficiency virus Tat transactivator." *Science* **248**(4963): 1650-3.

O'Brien, T., S. Hardin, A. Greenleaf, and J. T. Lis (1994). "Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation." *Nature* **370**(6484): 75-7.

Ohana, B., P. A. Moore, S. M. Ruben, C. D. Southgate, M. R. Green, and C. A. Rosen (1993). "The type 1 human immunodeficiency virus Tat binding protein is a transcriptional activator belonging to an additional family of evolutionarily conserved genes." *Proc Natl Acad Sci U S A* **90**(1): 138-42.

Olsen, H. S., and C. A. Rosen (1992). "Contribution of the TATA motif to Tat-mediated transcriptional activation of human immunodeficiency virus gene expression." *J Virol* **66**(9): 5594-7.

Parada, C. A., and R. G. Roeder (1996). "Enhanced processivity of RNA polymerase II triggered by Tat-induced phosphorylation of its carboxy-terminal domain." *Nature* **384**(6607): 375-378.

Pauling, L. (1960). *The nature of the chemical bond.* Cornell, NY, Cornell University Press.

Pause, A., S. Lee, R. A. Worrell, D. Y. Chen, W. H. Burgess, W. M. Linehan, and R. D. Klausner (1997). "The von Hippel-Lindau tumor-suppressor gene product forms a stable complex with human CUL-2, a member of the Cdc53 family of proteins." *Proc Natl Acad Sci U S A* **94**(6): 2156-61.

Puglisi, J. D., L. Chen, S. Blanchard, and A. D. Frankel (1995). "Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex." *Science* **270**: 1200-1203.

Ratnasabapathy, R., M. Sheldon, L. Johal, and N. Hernandez (1990). "The HIV-1 long terminal repeat contains an unusual element that induces the synthesis of short RNAs from various mRNA and snRNA promoters." *Genes Dev* **4**(12A): 2061-74.

Reines, D., M. J. Chamberlin, and C. M. Kane (1989). "Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene *in vitro*." *J Biol Chem* **264**(18): 10799-809.

Roy, S., U. Delling, C. H. Chen, C. A. Rosen, and N. Sonenberg (1990). "A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation." *Genes Dev* **4**(8): 1365-73.

Sambrook, J., E. F. Fritsch, and T. Maniatis (1989). *Molecular Cloning.* Cold Spring Harbor, Cold Spring Harbor Laboratory Press.

Sandri-Goldin, R. M., A. L. Goldin, J. Glorioso, and M. Levine (1984). "Herpes simplex virus thymidine kinase gene is stably maintained and expressed in cells transformed by protoplast fusion." *Somat Cell Mol Genet* **10**(2): 129-38.

Seed, B., and A. Aruffo (1987). "Molecular cloning of the CD2 antigen, the T-cell erythrocyte receptor, by a rapid immunoselection procedure." *Proc Natl Acad Sci U S A* **84**(10): 3365-3369.

Selby, M. J., E. S. Bain, P. A. Luciw, and B. M. Peterlin (1989). "Structure, sequence, and position of the stem-loop in tar determine transcriptional elongation by tat through the HIV-1 long terminal repeat." *Genes Dev* **3**(4): 547-58.

Selby, M. J., and B. M. Peterlin (1990). "Trans-activation by HIV-1 Tat via a heterologous RNA binding protein." *Cell* **62**(4): 769-76.

Sheldon, M., R. Ratnasabapathy, and N. Hernandez (1993). "Characterization of the inducer of short transcripts, a human immunodeficiency virus type 1 transcriptional element that activates the synthesis of short RNAs." *Mol Cell Biol* **13**(2): 1251-63.

Sheridan, P. L., C. T. Sheline, K. Cannon, M. L. Voz, M. J. Pazin, J. T. Kadonaga, and K. A. Jones (1995). "Activation of the HIV-1 enhancer by the LEF-1 HMG protein on nucleosome- assembled DNA *in vitro*." *Genes Dev* **9**(17): 2090-2104.

SivaRaman, L., D. Reines, and C. M. Kane (1990). "Purified elongation factor SII is sufficient to promote read-through by purified RNA polymerase II at specific termination sites in the human histone H3.3 gene." *J Biol Chem* **265**(24): 14554-60.

Song, C. Z., P. M. Loewenstein, and M. Green (1994). "Transcriptional activation *in vitro* by the human immunodeficiency virus type 1 Tat protein: evidence for specific interaction with a coactivator(s)." *Proc. Natl. Acad. Sci. U.S.A.* **91**(20): 9357-9361.

Southgate, C., M. L. Zapp, and M. R. Green (1990). "Activation of transcription by HIV-1 Tat protein tethered to nascent RNA through another protein." *Nature* **345**(6276): 640-2.

Southgate, C. D., and M. R. Green (1991). "The HIV-1 Tat protein activates transcription from an upstream DNA-binding site: implications for Tat function." *Genes Dev* **5**(12B): 2496-507.

Spencer, C. A., and M. Groudine (1990). "Transcription elongation and eukaryotic gene regulation." *Oncogene* **5**(6): 777-85.

Subramanian, T., C. D'Sa-Eipper, B. Elangovan, and G. Chinnadurai (1994). "The activation region of the Tat protein of human immunodeficiency virus type-1 functions in yeast." *Nucleic Acids Res* **22**(8): 1496-9.

Tan, R., L. Chen, J. A. Buettner, D. Hudson, and A. D. Frankel (1993). "RNA recognition by an isolated alpha helix." *Cell* **73**: 1031-1040.

Tan, R., and A. D. Frankel (1997). "A novel glutamine-RNA interaction identified by screening libraries in mammalian cells." *Submitted.*

Tan, S., R. C. Conaway, and J. W. Conaway (1995). "Dissection of transcription factor TFIIF functional domains required for initiation and elongation." *Proc Natl Acad Sci U S A* **92**(13): 6042-6046.

Tao, J., and A. D. Frankel (1993). "Electrostatic interactions modulate the RNA-binding and transactivation specificities of the human immunodeficiency virus and simian immunodeficiency virus Tat proteins." *Proc Natl Acad Sci U S A* **90**(4): 1571-5.

Weeks, K. M., C. Ampe, S. C. Schultz, T. A. Steitz, and D. M. Crothers (1990). "Fragments of the HIV-1 Tat protein specifically bind TAR RNA." *Science* **249**(4974): 1281-5.

Weiss, R. L. (1978). "Methods for protoplast formation in *Escherichia coli*." *Methods Cell Biol* **20**: 141-7.

Williams, R. D., B. A. Lee, S. P. Jackson, and N. J. Proudfoot (1996). "Activation domains of transcription factors mediate replication dependent transcription from a minimal HIV-1 promoter." *Nucleic Acids Res.* **24**(4): 549-557.

Wu, F., J. Garcia, D. Sigman, and R. Gaynor (1991). "tat regulates binding of the human immunodeficiency virus trans-activating region RNA loop-binding protein TRP-185." *Genes Dev* **5**(11): 2128-40.

Wu-Baer, F., W. S. Lane, and R. B. Gaynor (1995). "The cellular factor TRP-185 regulates RNA polymerase II binding to HIV-1 TAR RNA." *EMBO J.* **14**: 5995-6009.

Wu-Baer, F., W. S. Lane, and R. B. Gaynor (1996). "Identification of a group of cellular cofactors that stimulate the binding of RNA polymerase II and TRP-185 to human immunodeficiency virus 1 TAR RNA." *J Biol Chem* **271**(8): 4201-4208.

Yang, X., C. H. Herrmann, and A. P. Rice (1996). "The human immunodeficiency virus Tat proteins specifically associate with TAK *in vivo* and require the carboxyl-terminal domain of RNA polymerase II for function." *J. Virol.* **70**(7): 4576-4584.

Zenzie-Gregory, B., A. Khachi, I. P. Garraway, and S. T. Smale (1993). "Mechanism of initiator-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence." *Mol Cell Biol* **13**(7): 3841-9.

Zhou, Q., and P. A. Sharp (1995). "Novel mechanism and factor for regulation by HIV-1 Tat." *Embo J* **14**(2): 321-328.

Zhou, Q., and P. A. Sharp (1996). "Tat-SF1: Cofactor for stimulation of transcriptional elongation by HIV-1 Tat." *Science* **274**: 605-610.

Zhu, Y., T. Pe'ery, J. Peng, Y. Ramanathan, N. Marshall, T. Marshall, B. Amendt, M. B. Mathews, and D. H. Price (1997). "Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation *in vitro*." *Genes Dev* **11**(20): 2622-2632.