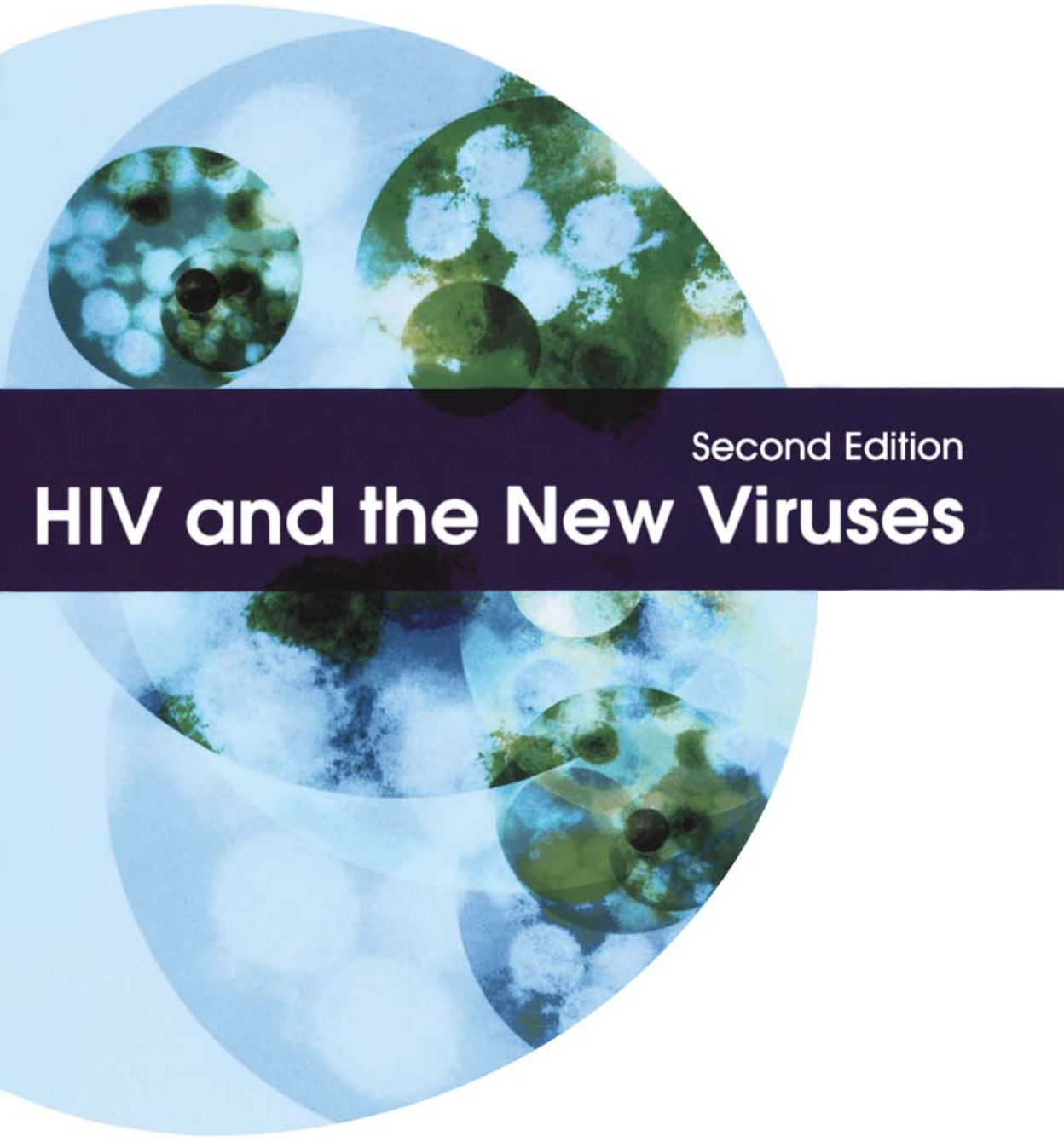

Second Edition

HIV and the New Viruses

Edited by Angus G. Dalgleish & Robin A. Weiss

HIV AND THE NEW VIRUSES

This Page Intentionally Left Blank

HIV AND THE NEW VIRUSES

Second Edition

Angus Dalgleish
Division of Oncology,
Department of Cellular and Molecular Sciences,
St George's Hospital Medical School,
Cranmer Terrace,
London SW17 0RE, UK

Robin Weiss
Windeyer Institute of Medical Sciences,
University College London,
46 Cleveland Street,
London W1P 6DB, UK

ACADEMIC PRESS
Harcourt Brace & Company, Publishers
San Diego London Boston New York Sydney Tokyo Toronto

ACADEMIC PRESS
525 B Street, Suite 1900, San Diego,
California 92101–4495, USA
http://www.apnet.com

ACADEMIC PRESS
24–28 Oval Road
LONDON NW1 7DX
http://www.hbuk.co.uk/ap/

This book is printed on acid-free paper

A catalogue record for this book is available from the British Library

Library of Congress Card Number: 98-83211

ISBN 0–12–200741–7

Typeset by J&L Composition Ltd, Filey, North Yorkshire
Printed in Great Britain by MPG Books Ltd, Bodmin, Cornwall
99 00 01 02 03 04 MP 9 8 7 6 5 4 3 2 1

CONTENTS

LIST OF CONTRIBUTORS

Alicia Alonso Centre for HIV Research, ICAPB, University of Edinburgh, Edinburgh EH9 3JN, UK

Harold Baum Division of Life Sciences, King's College Hospital, Campden Hill Road, London W8 7AH, UK

Neil Berry Division of Retrovirology, NIBSC, Blanche Lane, Potters Bar EN6 3QG, UK

Andrew J Leigh Brown Centre for HIV Research, Institute of Cell Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN, UK

William F Carman Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, UK

Mario Clerici Cattedra di Immunologia, Centro Ricerche LITA, H.L. Sacco, Universita di Milano, Via Venezian, 1, 21033 Milano, Italy

Alberto Clivio Dipartimento di Biologia e Genetica per le Scienze Mediche, Centro Ricerche LITA, H.L. Sacco, Universita di Milano, Via Venezian, 1, 21033 Milano, Italy

Angus G Dalgleish Division of Oncology, Department of Cellular and Molecular Sciences, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK

Klaus-Michael Debatin University Children's Hospital, Prittwitzstrasse 43, D-89075 ULM, Germany

Barbara Ensoli Laboratory of Virology, Istituto Superiore di Sanità, Viale Regina Elena, 299, 00161 Rome, Italy

Daniela Fenoglio Unit of Retroviral Immunology, Department of Immunology, San Martino Hospital, University of Genoa and Advanced Biotechnology Center, Largo Benzi, 10, 16132 Genoa, Italy

Jeremy A Garson Department of Virology, University College London Medical School, Windeyer Building, 46 Cleveland Street, London W1P 6DB, UK

Antoine Gessain Unité d'Oncologie Virale, Départment du SIDA et des Rétrovirus, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris Cedex 15, France

Frances Gotch Department of Immunology, Chelsea and Westminster Hospital, 369 Fulham Palace Road, Chelsea, London SW10 9NH, UK

Marie-Lise Gougeon Département SIDA et Rétrovirus, Unité d'Oncologie Virale, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris, Cedex 15, France

Elizabeth Hounsell Department of Biochemsitry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK

Paul Kellam Department of Virology, Institute of Cancer Research, Royal Cancer Hospital, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

Jonathan R Kerr Department of Virology, University College London Medical School, Windeyer Building, 46 Cleveland Street, London W1P 6DB, UK

Giuseppina Li Pira Unit of Retroviral Immunology, Department of Immunology, San Martino Hospital, University of Genoa and Advanced Biotechnology Center, Largo Benzi, 10, 16132 Genoa, Italy

Paolo Lusso Unit of Human Virology, DIBIT, San Raffaele Scientific Institute, Milano 20132, Italy

Renaud Mahieux Laboratory of Gene Expression and Receptor Biology, NCI/NIH, Building 41, Bethesda, MD 20892, USA

Fabrizio Manca Unit of Retroviral Immunology, Department of Immunology, San Martino Hospital, University of Genoa and Advanced Biotechnology Center, Largo Benzi, 10, 16132 Genoa, Italy

Dean L Mann Division of Immunogenetics, University of Maryland Medical System, 22 South Greene Street, P2F01E, University Center, Baltimore, MD 21201-595, USA

Stephen McAdam Department of Immunology, Chelsea and Westminster Hospital, 369 Fulham Palace Road, Chelsea, London SW10 9NH, UK

Myra McClure Division of Medicine, Jefferies Research Trust, Imperial College of Medicine at St Mary's, Praed Street, London W2 1NY, UK

Paola Monini Laboratory of Virology, Istituto Superiore di Sanità, Viale Regina Elena, 299, 00161 Rome, Italy

Michael Norcross Division of Haematological Products, Center for Biologics Evaluation and Research, FDA, Bethesda, MD20892, USA

Clive Patience Biotransplant Incorporated, Building 75, 3rd Avenue, Charlestown Navy Yard, MA 02129, USA

B. Matija Peterlin Howard Hughes Medical Institute, Mount Zion Cancer Center, University of California, San Francisco, CA 94143-0703, USA

Andrew Phillips Department of Primary Care and Population Sciences, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, UK

Lisa Rosenblum Division of Medicine, Jefferies Research Trust, Imperial College of Medicine at St Mary's, Praed Street, London W2 1NY, UK

Hanneke Schuitemaker Department of Clinical Viro-Immunology, Central Laboratory of the Netherlands Red Cross Blood Transfusion Service, Plesmanlaan 125, 1006 AD, Amsterdam, The Netherlands

Cecilia Sgadari Laboratory of Virology, Istituto Superiore di Sanità, Viale Regina Elena, 299, 00161 Rome, Italy

Gene M Shearer Experimental Immunology Branch, NCI, NIH, Bethesda, MD 20892, USA

Peter Simmonds Department of Medical Microbiology, University of Edinburgh, Medical School, Teviot Place, Edinburgh EH8 9AG, UK

Donald B Smith Department of Medical Microbiology, University of Edinburgh, Medical School, Teviot Place, Edinburgh EH8 9AG, UK

Simon J Talbot Department of Medical Microbiology, Edinburgh University, Medical School, Teviot Place, Edinburgh EH8 9AG, UK

Richard Tedder UCL Medical School, Department of Virology, The Windeyer Building, 46 Cleveland Street, London W1P 6DB, UK

Christian Trautwein Abteilung Gastroenterologie und Hepatologie, Medischinische Hochsschule Hannover, Hannover, Germany

Mark A Wainberg McGill University AIDS Centre, Lady Davis Institute–Jewish General Hospital, 3755 Chemin Cote Ste-Catherine, Montreal, Quebec, Canada H3T 1E2

Ian V D Weller Department of Sexually Transmitted Diseases, University College London Medical School, Mortimer Market Centre off Capper Street, London W1CE 6AU, UK

Denise Whitby Institute of Cancer Research, Chester Beatty Research Institute, 237 Fulham Road, London SW3 6JB, UK

Mark K Williams Unit of Human Virology, DIBIT, San Raffaele Scientific Institute, Milano 20132, Italy

This Page Intentionally Left Blank

PREFACE TO THE FIRST EDITION

A decade ago mankind seemed to be on the brink of conquering viral disease. Smallpox had finally been eradicated, safe and efficacious vaccines were becoming available for most of the childhood infections and there was a general air of confidence that virally induced cancers would similarly yield to preventive measures. The recognition of acquired immune deficiency syndrome (AIDS) in 1981 shattered our complacency. Nevertheless, there has been remarkable progress in our understanding of AIDS since the causative agent, human immunodeficiency virus (HIV) was first isolated in 1983. The first six chapters in the volume chart the course of the AIDS epidemic, the molecular and cellular biology and pathology of HIV infection, and the prospects and problems in harnessing this knowledge for therapeutic and preventive means.

During the past decade other human viruses came to light for the first time, even though they may have an ancient provenance in human populations. The human T-cell leukaemia viruses (HTLV-I, HTLV-II) were discovered in 1980 and 1982 representing the first human retroviral pathogens, new human papilloma virus genomes (HPV-16, HPV-18) were identified in 1983 associated with cervical cancer, human herpes virus type 6 was recognized in 1987 and most recently in 1989, hepatitis C virus (HCV) has been identified among the non-A, non-B hepatitis infections. Some of these viruses (HTLV, HHV6) have been isolated by the classic method of propagation in culture, while others (HPV, HCV) were identified by molecular cloning.

The chapters contained in this volume were written to enable scientists and clinicians interested in human viral infection to obtain topical reviews that critically evaluate relevant data and concepts from the rapidly burgeoning literature. The chapters are contributed by leading investigators who have been enjoined to review a wide field without stifling their personal views. The subject matter is necessarily selective. We have, for example, omitted discussion of human papovaviruses or of the interesting animal lentiviruses related to HIV that have come to light since the discovery of HIV.

We are most grateful to the authors for their contributions and for ensuring that they are up to date including 1990 references. We thank Sue King at Academic Press for her understanding and persistence in producing this volume.

Angus G. Dalgleish
Robin A. Weiss

This Page Intentionally Left Blank

PREFACE TO THE SECOND EDITION

A decade ago we agreed to edit a book of authoritative and easy-to-read reviews on the major aspects of the human immunodeficiency virus (HIV), the causative agent of acquired immune deficiency syndrome, which itself had been recognized for less than a decade at the time of editing. It became apparent while planning the book that there would be much reference to the first human retrovirus, human T cell lymphotrophic virus type 1 (HTLV-1) and its relative HTLV-2, and we therefore commissioned a comprehensive review of these viruses. It became apparent that HIV and HTLV-1 were not the only new viruses to be discovered in the 1980s, and we broadened the remit to include reviews of herpesvirus 6 and the new hepatitis viruses.

In the last decade there has been a tremendous amount of information on HIV, its molecular structure, and the way it interacts with the host. The reviews presented in this volume are timely and summarize the major issues that have emerged since the first book. In order to accommodate this amount of information, the scope of the book is much broader than its predecessor.

It has been brought to our attention that the first book was referred to as much for the chapters on the new viruses as for those on HIV and AIDS. Indeed, it appeared pertinent to address the fact that viruses other than HIV were being discovered for the first time. In the past ten years a number of viruses have been discovered and we have taken the opportunity to include comprehensive reviews not only of the new viruses such as human herpesvirus 8 (HHV-8), the causative agent of Kaposi's sarcoma, but also of new aspects about other viruses only recently discovered at the time of publication of the first edition. This book also contains new information about older viruses such as hepatitis B virus, which can appear enigmatic to those not intricately acquainted with its molecular structure and pathogenesis, and also serves as an introduction to the new hepatitis viruses.

The last decade has seen the discovery and characterization of new neurological diseases and the identification of new associated viruses. We have therefore included a review on this topic as well as on the increasing documentation and association of some clinical conditions associated with endogenous retroviruses.

Although this is not a comprehensive textbook on virology, the issues covered here are of major interest to clinicians and scientists alike who are interested in the emergence of these new infectious agents. We are not aware of any other source of such comprehensive reviews of the major issues in HIV and the new viruses.

We thank Tessa Picknett, Duncan Fatz, Emma White and Lilian Leung at Academic Press for working patiently with us on this volume.

Angus G. Dalgleish
Robin A. Weiss

This Page Intentionally Left Blank

Chapter 1

HIV-1: Control of gene expression by the viral regulatory proteins Tat and Rev

Alicia Alonso and B. Matija Peterlin

INTRODUCTION

Expression of the human immunodeficiency virus type 1 (HIV-1) in infected cells results from a complex interplay between cellular proteins and viral components (DNA, RNA and proteins). Prominent among virally encoded proteins are Tat and Rev, which are unique to HIV-1 and related lentiviruses. After extensive mutagenesis of the viral genome, both proteins were found to be essential for viral gene expression, replication and cytopathology (Dayton et al, 1986; Feinberg et al, 1986; Fisher et al, 1986; Sodroski et al, 1986). Although both proteins mediate their functions through viral RNA elements, Tat via the transactivation response element (TAR) and Rev via the Rev response element (RRE), the control points at which Tat and Rev act and their mechanisms of action are fundamentally different. Tat modifies the nascent transcription complex allowing for efficient elongation of transcription, which increases levels of full-length viral transcripts. Rev acts post-transcriptionally to export incompletely spliced viral mRNAs, which code for structural proteins, from the nucleus to the cytoplasm. Thus, both proteins are required to activate fully the quiescent provirus. Since they are indispensable for viral replication, they are also attractive candidates for therapeutic intervention. Therefore, unraveling their mechanism of action is of great biological and clinical interest. This chapter reviews the roles of these regulatory proteins in the viral life cycle, as well as structures and functions of Tat and Rev and their target sequences TAR and RRE. Their interactions with cellular proteins and forms of therapeutic intervention are also addressed.

Tat, Rev AND THE PROVIRAL LIFE CYCLE

Integration of the viral DNA into the host genome generates a provirus, which is a necessary step in the life cycle of all retroviruses. Once this event has occurred, the HIV-1 provirus behaves as a cellular gene, with the promoter and transcriptional start site located in the 5′ long terminal repeat (LTR), and a termination/polyadenylation site located in the 3′ LTR (Figure 1.1A; see also Figure 1.4 and The promoter region). Viral replication depends on the activation of the LTR and on the expression, from the 9 kb

HIV and the New Viruses Second Edition
ISBN 0-12-200741-7

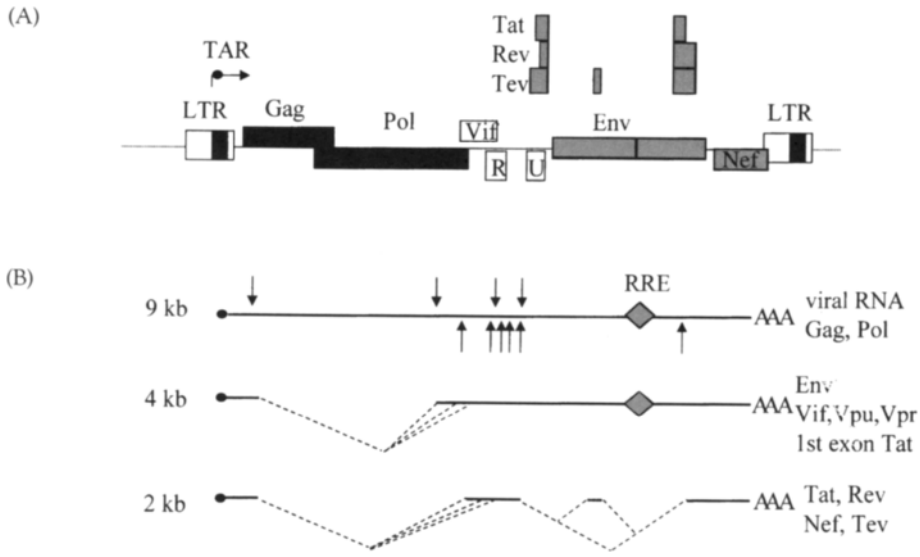

Figure 1.1 ***HIV-1 genomic organization and mRNA expression***
(A) The 9 kb proviral genome is shown schematically flanked by cellular sequences (dotted lines). The 5′ LTR sequences contain the promoter. Transcription, indicated by the arrow, starts within R (solid black). Positioned within nucleotides 1 to 60 is TAR, indicated by a solid dot. Coding sequences for the viral proteins are depicted by rectangles. The solid line within *env* separates gp120 from gp41, R= *vpr*, U= *vpu*. Spliced exons of *tat*, *rev* and *tev* are indicated above. The 3′ LTR contains termination signals.
(B) The three mRNA species present during active infection are depicted. TAR is present in all of them (solid dot); a diamond indicates the presence of Rev. Splice donor sites are indicated above the line representing the full-length transcript; splice acceptor sites are indicated below. Their positions are approximate. Sizes of the unspliced, singly spliced and multiply spliced mRNA are indicated on the left, and on the right the proteins they encode. Dotted lines represent different splicing events.

provirus, of nine different open reading frames (ORFs). Of these, three encode structural proteins, and six encode regulatory and accessory proteins.

Viral mRNAs

In T cells or macrophages which support replication of HIV-1, three sizes of viral mRNAs can be detected by northern blot analysis performed with the 3′ LTR polyA probe. They are 9 kb, 4 kb and 2 kb in length respectively (Figure 1.1B). The 9 kb transcripts, which represent the genomic mRNA, are homogeneous in size. They are unspliced and serve two purposes: they are the viral RNA which is packaged into new virions; and they translate the structural Gag proteins and the enzymatic Pol polyprotein, encoding reverse transcriptase, protease and integrase. In contrast, the 4 kb and 2 kb transcripts are heterogeneous in size and their existence depends on the choice of alternative splicing signals located along the 9 kb transcript (Figure 1.1B). The 4 kb transcripts are singly spliced and code for Env, Vpr, Vpu, Vif and the first exon of Tat. The 2 kb transcripts are multiply spliced, and code for Tat, Rev, Nef and Tev. To date, 15 alternatively spliced transcripts have been described (Kim et al, 1989; Robert-Guroff et al, 1990; Schwartz et al, 1990a,b; Neumann et al, 1994).

Temporal analyses of these viral mRNAs, in T cells that had undergone a single round of HIV-1 replication, determined that their expression is tightly regulated. The appearance of the 2 kb species precedes that of the 4 kb and 9 kb species: they are thus referred to as early and late transcripts respectively (Feinberg et al, 1986; Kim et al, 1989). Since Tat and Rev were translated from early transcripts, they were considered likely candidates for controlling late events in viral replication. Moreover, mutations in either of these two proteins abrogated viral replication (Dayton et al, 1986; Feinberg et al, 1986; Fisher et al, 1986; Sodroski et al, 1986). Both proteins control coordinately the expression and localization of viral transcripts.

Tat phenotype

Analyses of the total viral RNA content in cells infected with viruses containing mutations in Tat demonstrated low abundance of all three mRNA species. This defect was corrected when cells were complemented with the wild-type Tat (Feinberg et al, 1991; Adams et al, 1994). By performing steady-state RNA and nuclear run-on analyses, where the changes in the density of promoter proximal and promoter distal transcripts were monitored, it was determined that Tat increased rates of transcription from the HIV-1 LTR (hence its name *transactivator of transcription*). Tat allows RNA polymerase II (RNAPII) to traverse the length of the viral genome, thus dramatically increasing levels of elongated transcripts (Kao et al, 1987; Laspia et al, 1989, 1990; Feinberg et al, 1991). This effect of Tat was dependent on a viral *cis*-acting RNA sequence, called the transactivation response element (TAR) which is located precisely at the start of all viral transcripts. TAR extends from nucleotides +1 to +60 and forms a stable stem-loop structure (Rosen et al, 1985; Kao et al, 1987; Feng and Holland, 1988; Hauber and Cullen, 1988; Jakobovits et al, 1988; Berkhout and Jeang, 1989; Garcia et al, 1989; Selby et al, 1989; Roy et al, 1990a,b).

In the absence of Tat, the predominant viral transcripts present in infected or transfected cells are non-polyadenylated RNAs of about 60 bp, which contain TAR. They are referred to as short transcripts. Since TAR is resistant to nuclease degradation, short transcripts are probably the result of nuclease degradation of incompletely elongated transcripts which are constantly being initiated at the HIV-1 LTR by the inducer of short transcripts (IST) (see The promoter region; Kao et al, 1987; Selby et al, 1989; Toohey and Jones, 1989). When Tat is present, it binds to TAR as nascent RNA (Berkhout et al, 1989), represses the synthesis of short transcripts (Pendergrast and Hernandez, 1997), and modifies the transcription complex allowing complete elongation of viral transcripts. These events are discussed in detail in Tat and the transcription complex.

Rev phenotype

When viral transcripts from cells infected with viruses bearing mutations in *rev* were analyzed, a significant difference was apparent when nuclear and cytoplasmic RNAs were fractionated and compared. All three mRNA species were present in the nucleus, but only 2 kb transcripts were observed in the cytoplasm (Emerman et al, 1989; Felber et al, 1989; Hammarskjold et al, 1989; Malim et al, 1989). The appearance of 9 kb and 4 kb transcripts in the cytoplasm was restored only after the coexpression of Rev in the same cells (Knight et al, 1987). Thus, Rev allows for the movement of the unspliced and

singly spliced mRNA species from the nucleus to the cytoplasm. Rev derives its name, the *regulator of expression of virion proteins*, because these late transcripts encode the structural and enzymatic proteins needed for virion assembly.

Splicing versus nuclear export

Some groups have hypothesized that Rev also inhibits aspects of pre-mRNA splicing, which leads indirectly to enhanced pre-mRNA export (Hammarskjold et al, 1989; Kjems and Sharp, 1993). However, it has now been demonstrated that Rev interacts directly with the nuclear export machinery (see Rev and the nuclear transport of mRNAs). In contrast to cellular transcripts, which do not exit the nucleus until all introns have been removed, the two late transcripts of HIV-1 contain at least one complete unexcised intron, i.e. the second major HIV-1 intron which codes for Env (Figure 1.1B). Like Tat, the activity of Rev is dependent on the presence of a *cis*-acting RNA element. This sequence, the Rev response element, is between 240 and 351 nucleotides long. It is located precisely in the second major intron of HIV-1, just 3′ to the junction between the gp120 and gp41 subunits of *env* (Rosen et al, 1988; Felber et al, 1989; Hadzpoulou-Cladaras et al, 1989; Malim et al, 1989b; Mann et al, 1994). The RRE sequence is therefore present only in unspliced and singly spliced mRNA species. How interactions between Rev and the RRE lead to nuclear export of the late viral transcripts is discussed on pp. 13–18.

The proviral life cycle

Tat and Rev control the switch from latency to active replication of the provirus in the following manner (Figure 1.2). After integration, the virus remains in a transcriptionally latent state, characterized by the presence of short TAR transcripts in the cytoplasm (Feinberg et al, 1991; Adams et al, 1994). Cellular activation signals overcome this latency by increasing rates of initiation of HIV-1 transcription from the LTR (see Tat and the transcription complex; Jones and Peterlin, 1994). All introns from these primary transcripts are removed in the nucleus by the splicing machinery. Thus, multiply spliced 2 kb transcripts encoding Tat and Rev accumulate (Kim et al, 1989). These early transcripts are exported rapidly to the cytoplasm, where Tat and Rev are translated and imported into the nucleus (Figure 1.2A) (Malim et al, 1989a; Perkins et al, 1989; Ruben et al, 1989; Siomi et al, 1990). Tat then interacts with TAR and the transcription complex and increases tremendously the amount of full-length polyadenylated transcripts in the nucleus (Feinberg et al, 1991; Adams et al, 1994). Because of Rev and its interaction with the nuclear export pathway, these unspliced and singly spliced transcripts are now translocated to the cytoplasm through the RRE (Knight et al, 1987; Fornerod et al, 1997a; Stade et al, 1997). In the cytoplasm, translation of structural proteins and assembly of new virions ensue (Figure 1.2B).

THE Tat PROTEIN: A MODULAR STRUCTURE

Tat is encoded in two exons. The first exon is located just 5′ of *env* and encodes 72 amino acids. The second exon is located 3′ of *env*, and is of variable size, from 14 to 29 amino acids, depending on the viral isolate (Arya et al, 1984; Sodroski et al, 1985). Within early transcripts, three species of multiply spliced mRNAs encode Tat. Within

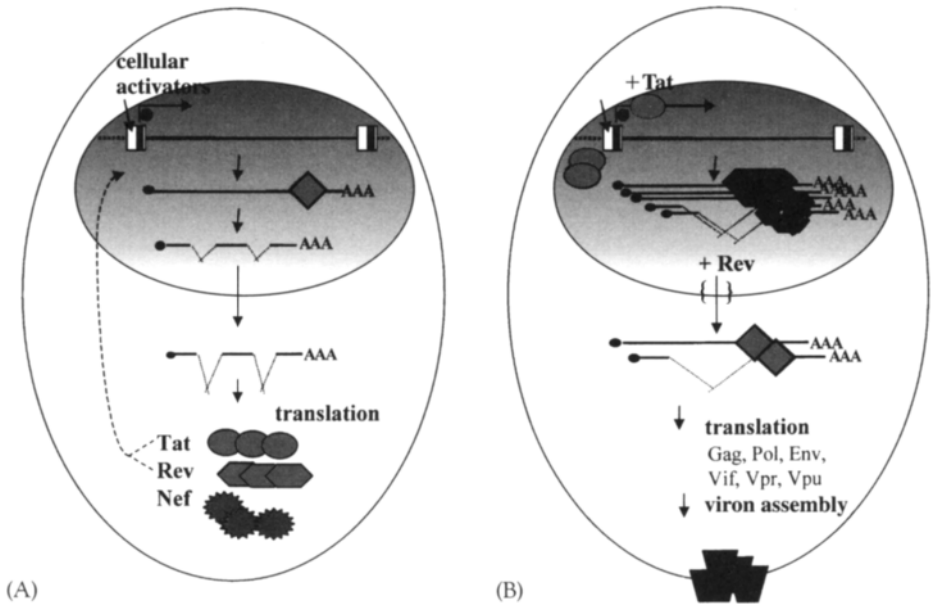

Figure 1.2 ***The proviral life cycle***
(A) Early events. Proviral transcription is activated by cellular transcription factors acting on the 5′ LTR. Low levels of mRNA accumulate in the nucleus (shaded area), from which only the multiply spliced mRNAs (2 kb) are detected in the cytoplasm. These code for Tat (circles), Rev (hexagons) and Nef (porcupines). Tat and Rev relocalize into the nucleus.
(B) Late events. In the nucleus, Tat binds to TAR and modifies the transcription complex, indicated by the drawn-out arrow. This results in increased levels of mRNAs. In addition, Rev binds to the RRE and allows for the unspliced and singly spliced mRNAs to be exported to the cytoplasm. After translation, virion assembly and extrusion occur.

late transcripts, Tat is encoded by a singly spliced first exon (Schwartz et al, 1990b). In addition, Tat is expressed as a hybrid molecule, Tev, the product of the first exon of *tat* spliced to 114 nucleotides from *env* and the second exon of *rev* (see Figure 1.1) (Benko et al, 1990).

Transient transfection assays into tissue culture cells, using plasmids containing the HIV-1 LTR fused to a reporter gene (chloramphenicol acetyltransferase or luciferase) and expression plasmids containing either wild-type or mutated *tat*, have demonstrated that the first exon of Tat suffices for high levels of transactivation (Garcia et al, 1988; Kuppuswamy et al, 1989). The second exon mediates cellular functions, such as lymphocyte activation (Howcroft et al, 1993; Chang et al, 1995; Ott et al, 1997).

At the amino acid level, the first exon of *tat* is highly conserved among different isolates (Myers et al, 1991), as well as among different lentiviruses (Peterlin et al, 1993). This similarity led to the preliminary delineation of five structural domains, namely the N-terminal, cysteine-rich, core, basic and C-terminal domains of Tat (Figure 1.3A). Their roles have been defined by mutagenesis followed by functional analysis. Like well-known DNA transactivator proteins, Tat has a modular structure, with one domain that interacts with cellular factors and one domain that binds to its cognate RNA sequence. However, the RNA-binding domain of Tat cannot be separated from its activation domain for efficient interaction with TAR.

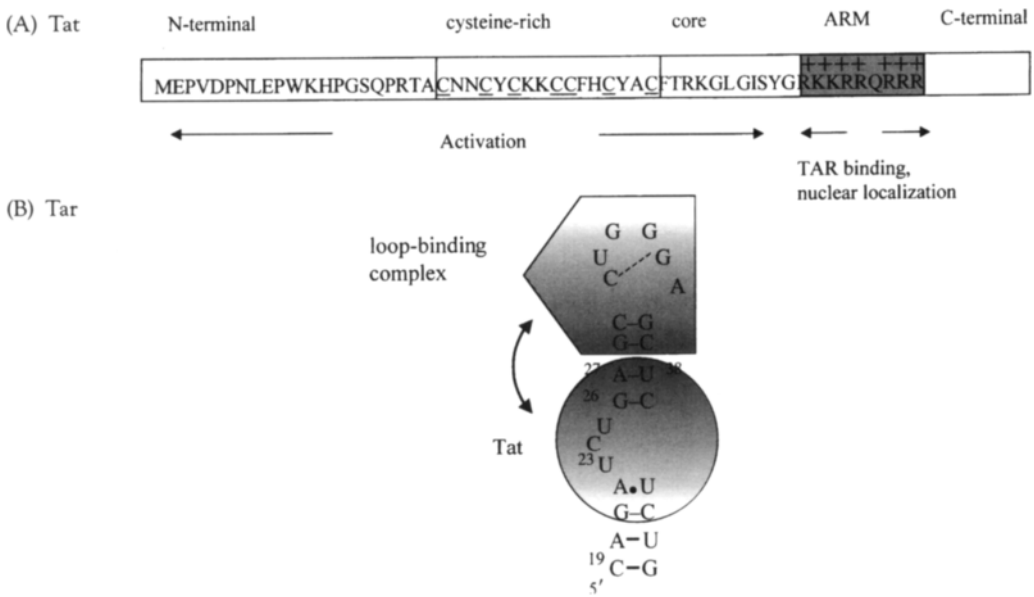

Figure 1.3 ***Domain structure of Tat and RNA structure of TAR***
(A) The first exon of Tat. The five structural domains with their relevant amino acid sequences are indicated. The activation domain extends from amino acids 1 to 47. Positive charges in the ARM domain are indicated above the corresponding amino acids. A solid unfilled block represents the C-terminal domain.
(B) The minimal functional TAR sequence, from positions +19 to +43, and its proposed secondary structure are depicted. Within TAR, the lower stem, the bulge, the upper stem and the loop are indicated. The loop is shown with base pairing of C30·G34 and a bulged-out A35 (Jones and Peterlin, 1994). A22·U40 are thought to pair after binding, which is indicated by a dot. Tat binds to the U23 in the major groove, occupying the bulge and 2 base pairs above and below the bulge. A putative loop-binding complex is shown which may interact and stabilize Tat binding to TAR, indicated by the arrow.

The activation domain of Tat

The first three structural domains of Tat are necessary for transcriptional activation. The N-terminal domain corresponds to amino acids 1 to 21. It contains PXXXP repeats, and acidic amino acids at positions 2, 5, 9 which appear to be necessary for transactivation. These amino acids can be replaced, provided the overall structure remains amphipathic (Kuppuswamy et al, 1989; Rappaport et al, 1989; Tiley et al, 1992). The cysteine-rich domain corresponds to amino acids 22 to 37, and contains seven cysteines, of which three cysteine pairs are indispensable (Garcia et al, 1988; Sadaie et al, 1988; Kuppuswamy et al, 1989; Rice and Carlotti, 1990). The core domain corresponds to amino acids 38 to 48. It is essential and has no known similarities to other activation domains (Kuppuswamy et al, 1989; Rice and Carlotti, 1990). The use of chimeric Tat proteins on chimeric HIV-1–LTRs has corroborated that these 47 amino acids correspond to the activation domain. These amino acids were fused to each one of two known RNA-binding proteins: the coat protein of the bacteriophage MS2 and Rev; and TAR was replaced by the MS2 operator RNA target or the stem-loop IIB (SLIIB) of the RRE (Selby and Peterlin, 1990; Southgate et al, 1990). These fusion proteins were able

to transactivate the HIV-1 LTR from these hybrid promoters, thus confirming that the activation domain is contained within these 47 N-terminal residues, and that Tat acts by binding to RNA. Further domain swapping with Tat from the equine infectious anemia virus (EIAV) determined that a minimal lentiviral Tat can be constructed with the cysteine-rich and core domains of Tat of EIAV (15 amino acids) and the basic domain of Tat from HIV-1 (Derse et al, 1991).

To prove that Tat can also function at the level of initiation of transcription, Tat was tethered to the promoter as a DNA-binding protein (Berkhout et al, 1990; Kamine et al, 1991; Southgate and Green, 1991). Although these attempts have also been successful, the mechanism of Tat action via DNA and RNA is different. RNA-targeted Tat inhibits the synthesis of short transcripts and enhances that of elongated transcripts, whereas DNA-targeted Tat increases both processes (Pendergrast and Hernandez, 1997). It should be noted that Tat acts via RNA binding in preference to DNA binding.

The basic domain of Tat (arginine-rich motif)

The basic domain of Tat extends from amino acids 49 to 57, and consists of an arginine-rich motif (ARM), RKKRRQRRR. This motif has a dual role: it acts as a nuclear localization signal (Hauber et al, 1989; Ruben et al, 1989; Siomi et al, 1990), and binds to TAR (Dingwall et al, 1989; Roy et al, 1990a; Weeks et al, 1990). In cells, both the ARM and the activation domain of Tat are required to bind TAR efficiently, suggesting that a cellular protein interacts simultaneously with the activation domain of Tat and the loop in TAR (Jones and Peterlin, 1994) (see pp. 8–13).

The C-terminal domain of Tat

The C-terminal domain of Tat, which extends from amino acids 58 to 72, has no known motifs. Since deletions in the region past amino acid 60 cause only a small decrease in the activity of Tat, this domain is thought to act as an auxiliary domain (Muesing et al, 1987; Garcia et al, 1988; Frankel et al, 1989; Kuppuswamy et al, 1989).

Tat structure

Nuclear magnetic resonance (NMR) spectroscopy (Bayer et al, 1995) has solved the structure of Tat from HIV-1Z2, an isolate with 86 amino acids. This analysis has revealed that the N-terminus of Tat is sandwiched between core and C-terminal domains, leaving a flexible cysteine and basic domain accessible for protein–protein and protein–RNA interactions. Bayer et al found no α-helices in Tat, in contrast with the α-helical structure observed with the minimal lentiviral Tat (Mujeeb et al, 1994).

Cellular factors that bind to Tat

Since the activation domain of Tat must interact with some component of the transcription complex, biochemical approaches have been used to identify the factor(s) that interacts with Tat. About 15 proteins, which range from transcription factors to kinases, have been isolated. Among the former are TBP (Kashanchi et al, 1994, 1996); the p62 subunit of TFIIH (Blau et al, 1996); TatSF1 (Zhou and Sharp, 1996); the largest subunit

of RNAPII (Mavankal et al, 1996) and Sp1 (Jeang et al, 1993). The latter include CAK, the CDK-activating kinase associated with the TFIIH complex (Cujec et al, 1997; Garcia-Martinez et al, 1997) and CDK9/PITALRE, a kinase associated with the Tat-associated kinase/positive transcription elongation factor b complex (TAK/P-TEFb) (Herrmann and Rice, 1995; Yang et al, 1996; Mancebo et al, 1997; Zhu et al, 1997). The relevance of these kinases in mediating transactivation by Tat is discussed on pp. 11–13.

TRANSACTIVATION RESPONSE ELEMENT, TAR: A DUAL ROLE

The TAR element is the target sequence required for Tat transactivation. This sequence is unique, among all other *cis*-acting elements utilized in transcription, in that it forms an RNA stem loop. Early suggestions that TAR was an RNA element included the following observations: TAR was active only in the sense orientation (Peterlin et al, 1986), computer modeling demonstrated that TAR could form a stable RNA hairpin (Muesing et al, 1987; Feng and Holland, 1988), mutations that affected folding of this putative structure inhibited transactivation, which could be rescued by compensatory mutations (Feng and Holland, 1988; Hauber and Cullen, 1988; Jakobovits et al, 1988; Garcia et al, 1989; Selby et al, 1989; Roy et al, 1990b), Tat could be tethered to the transcription complex via heterologous RNA-binding domains (Selby and Peterlin, 1990; Southgate et al, 1990), and RNA but not DNA TAR decoys could inhibit transactivation (Graham and Maio, 1990; Sullenger et al, 1990, 1991; Lisziewicz et al, 1991).

Computer modeling of TAR exposed two distinctive elements in the TAR sequence (Figure 1.3B): a three-nucleotide bulge between positions +23 and +25 and a six-nucleotide loop between positions +30 and +35. The rest of the sequence provides a stable stem that makes TAR a highly nuclease-resistant RNA structure. Deletions and single point mutations in TAR narrowed the minimal sequences required for transactivation to those located between positions +19 and +43. Within this minimal structure, both the U23 in the bulge and the loop are required for Tat transactivation (Berkhout and Jeang, 1989; Dingwall et al, 1989; Cordingly et al, 1990; Roy et al, 1990a,b; Calnan et al, 1991a,b).

The bulge in TAR

Final proof that Tat binds TAR RNA came with the use of the electrophoretic mobility shift assay (EMSA), which revealed that recombinant Tat protein or Tat peptides bound to ^{32}P-labeled TAR RNA. Binding occurred in a ratio of one TAR to one Tat, was mediated by the ARM, and was specific for U23. In addition, two base pairs in the stem above the bulge, G26·C39 and A27·U38, two base pairs below the bulge, A22·U40 and G21·C41, and two phosphates located between positions 22, 23 and 24 were also involved in binding (Calnan et al, 1991a,b; Weeks and Crothers, 1991; Delling et al, 1992; Tao and Frankel, 1992; Churcher et al, 1993; Hamy et al, 1993; Pritchard et al, 1994).

When in solution, the A22 and U40 in TAR are unpaired making a four-nucleotide bulge, and U23 is stacked between C24 and A22 (Figure 1.3B) (Colvin and Garcia-Blanco, 1992; Critchley et al, 1993). The bulge bends the TAR RNA stem and introduces local distortions that widen the major groove, making it available for Tat binding

(Weeks and Crothers, 1991; Delling et al, 1992). Structural studies using circular dichroism (CD) spectrum analysis suggested that the ARM is unstructured and adopts a fixed conformation only after its binding to TAR (Calnan et al, 1991a). The fact that only one arginine from the ARM is sufficient for binding, albeit at low specificity, led to several studies using NMR spectroscopy, TAR and argininamide, an arginine analog (Calnan et al, 1991b; Puglisi et al, 1992). A simple binding model, the 'arginine fork' was proposed. Upon the binding of arginine, the bases in the bulge become separated and U23 and A27·U38 form a base-triple interaction that stabilizes hydrogen bonding of arginine to G26 and P22·P23 (Puglisi et al, 1993). Recently this model has been challenged by NMR data obtained with TAR and a Tat peptide containing both basic and core domains, which provides maximal specificity of binding (Churcher et al, 1993; Aboul-ela et al, 1995). Contrary to the findings of Frankel and co-workers (Puglisi et al, 1992, 1993), Varani and co-workers (Aboul-ela et al, 1995) found that U23 does not participate in the formation of a base triplet with A27·U38. Instead the conformational change is brought about by interactions of an arginine residue with both U23 and G26, thus repositioning the phosphate groups P22, P23 and P40 on the surface of the molecule.

The loop in TAR

The nucleotides in the loop interact through non-Watson–Crick base pairing, where C30 and G34 base pair and A35 bulges out (Figure 1.3B) (Colvin et al, 1993; Critchley et al, 1993). Notably, although these nucleotides are essential for Tat transactivation, they do not mediate the binding of Tat to TAR (Dingwall et al, 1989, 1990; Cordingly et al, 1990; Roy et al, 1990a; Weeks et al, 1990; Calnan et al, 1991a,b; Weeks and Crothers, 1991; Delling et al, 1992; Churcher et al, 1993). This finding further suggests that cellular factors bind the loop and in conjunction with Tat form an essential part of the modified transcription complex.

Cellular proteins that bind to the loop in TAR have been sought, both biochemically and genetically. Partially purified nuclear extracts from HeLa cells contain two fractions that bind TAR: TRP-185/TRP-1, which binds to the loop, and TRP-2, which binds to the bulge (Sheline et al, 1991; Wu et al, 1991). Their exact roles in transcription mediated by Tat are still under investigation. The fraction TRP-185 binds to TAR through cellular cofactors, of which three – EF-12 – PTB and SRB, have been identified. RNAPII also binds to TAR using these same cofactors (Wu-Baer et al, 1995, 1996).

Genetic evidence indicating that loop-binding proteins are required for Tat transactivation comes from studies in rodent cells which support very low levels of Tat transactivation (Hart et al, 1989; Newstein et al, 1990). Bypassing the requirement of TAR loop-binding proteins, with heterologous tethering of Tat via the coat protein, the EIAV TAR and Gal4 DNA-binding domain, restored Tat transactivation in these cells (Alonso et al, 1992; Madore and Cullen, 1993). Binding studies in vivo suggested that a complex involving Tat, the loop in TAR and a protein encoded in human chromosome 12 are all required for efficient transactivation (Alonso et al, 1994). An 83 kDa protein that binds to the loop and is specific for chromosome 12 has been described (Hart et al, 1995).

Tat AND THE TRANSCRIPTION COMPLEX

Tat acts on non-processive transcription complexes which are set up by the HIV-1 TATA box (Olsen and Rosen, 1992; Lu et al, 1993). In the absence of Tat, these complexes stall and gradually abort transcription 100–200 nucleotides downstream of the initiation site (Kao et al, 1987; Laspia et al, 1989, 1990; Feinberg et al, 1991). Thus, Tat might have effects both on promoter clearance and transcription elongation.

The promoter region

The viral promoter is located in the U3 region of the 5′ LTR (Figure 1.4). It contains upstream regulatory sequences that include binding sites for USF, and T cell-specific activators such as ETS1 and LEF. In addition it contains sites for extracellular signal transducers, such as NF-AT (position −300) and NFkB. All these elements set up the minimal transcription levels that can break cellular latency of the provirus (see The proviral life cycle; Jones and Peterlin, 1994).

The core promoter, which extends from position −80 to the site of initiation of transcription, contains a TATA box, an initiator site (SSR) and three tandem Sp1 sites. The

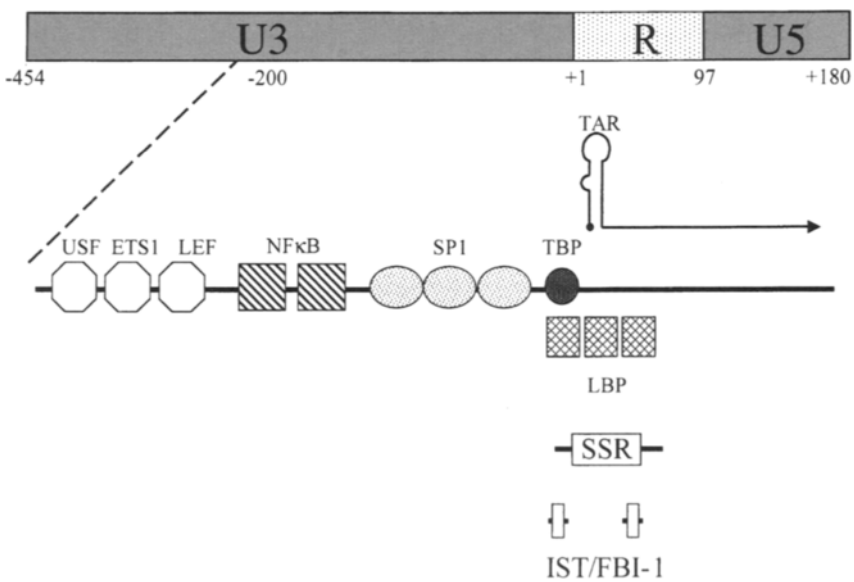

Figure 1.4 ***The HIV-1 promoter region***
The 5′ LTR is schematically represented at the top of the Figure. The coordinates given under U3, R and U5 define the limits of these regions with respect to +1, the first transcribed nucleotide. Below, the DNA sequence and the transcription factors that interact with the sequences, from positions −200 to +100, are shown. USF, ETS1 and LEF are upstream transcription factors; sites for NFkB, an inducible activator, are located within positions −100 to −80. The core promoter is composed of a TATA box (positions −30 to −20), and three Sp1 sites (−80 to −45). Additionally it contains three LBP/UBP sites (−30 to +20) which overlap with the initiator site (SSR). The IST (inducer of short transcripts) is located between −5 and +26 and +40 to +59; FBI-1 binds to IST. Transcription is indicated by the arrow, beginning with TAR, which forms at the 5′ end of all RNAs.

Sp1 sites are required to facilitate the assembly of non-processive complexes, but can be functionally replaced by other activator sequences (Southgate and Green, 1991; Berkhout and Jeang, 1992). Regulatory sequences downstream of the promoter include LBP-1/UBP-1 binding sites at position −38 to +27, which colocalize with the SSR, and the inducer of short transcripts element (IST). The IST is located between position −5 and +26 and position +40 to +59 and overlaps TAR. The role of the LBP-1/UBP-1 complex is still unclear; however, the IST mediates the synthesis of the short TAR transcripts, which recruit Tat to the promoter (Sheldon et al, 1993; see Tat phenotype) Recently, a factor that binds to the IST, FBI-1, has been described (Pessler et al, 1997).

Transcription

Initiation of transcription occurs through binding of a very stable RNAPII holoenzyme containing RNAPII, the general transcription factors (GTFs) TFIIF, TFIIH and mediator (a multisubunit complex containing suppressors of RNAP B proteins and other activators and repressors) onto a preassembled TFIID and TFIIB at the TATA box (Koleske and Young, 1995; Ossipow et al, 1995; Maldonado et al, 1996). After incorporation of a few nucleotides, the complex will pause until a critical regulatory step, the phosphorylation of the C-terminal domain (CTD) of the largest subunit of RNAPII, is achieved. This step represents the transition from initiation to elongation complexes. The CTD contains 52 repeats of a YSPTSPS domain in which serines, threonines and tyrosines are potential targets of phosphorylation. RNAPII thus exists in a non-phosphorylated form in initiation complexes (RNAPIIA) and in a hyperphosphorylated form in elongation complexes (RNAPIIO). Two steps are controlled by phosphorylation of the CTD, promoter clearance (when GTFs and mediator disengage from RNAPII) and elongation. Kinases that control these events are presently under study, and are targeted by Tat.

Cellular kinases

The original evidence that suggested that Tat affected rates of elongation of RNAPII by recruiting a cellular kinase was based on the observation that the purine nucleoside analog 5,6–dichloro-1B-D-ribofuranosylbenzimidazole (DRB) blocked Tat transactivation. DRB inhibits elongation by inhibiting cellular kinases (Marciniak et al, 1990; Marciniak and Sharp, 1991). The search for the kinase, however, split into two fields. While one set of data implicated TFHII, which plays a pivotal role in promoter clearance and has CAK-associated kinase activity, other studies implicated a new kinase complex, TAK. Tat can interact directly with the p62 subunit of TFIIH, and increases the ability of TFIIH to phosphorylate the CTD (Blau et al, 1996; Parada and Roeder, 1996). On the other hand, Tat interacts, through its activation domain, with TAK. TAK is more DRB sensitive than TFHII and can hyperphosphorylate the CTD (Herrmann and Rice, 1995; Yang et al, 1996).

Two subsequent findings have established that Tat targets both kinases. First, Tat has been shown to bind directly to CAK (Cujec et al, 1997; Garcia-Martinez et al, 1997). Second, the 42 kDa kinase subunit in TAK has been identified as the cyclin-dependent kinase CDK9, and is the human homolog of the kinase subunit of the *Drosophila* elongation complex P-TEFb (Zhu et al, 1997); P-TEFb acts at an early elongation step by

preventing RNAPII arrest, is DRB sensitive and can phosphorylate the CTD (Marshall et al, 1996). TAK is therefore the human homolog of P-TEFb.

The role of CDK9 in mediating Tat transactivation has been reinforced by studies in which a significant correlation was found between inhibitors of CDK9 and decreased levels of Tat transactivation (Mancebo et al, 1997).

It has recently been demonstrated that cyclin T, the associated CDK9 cyclin, is a subunit of the TAK/P-TEFb complex. Cyclin T interacts specifically with the activation domain of Tat. This new complex binds to TAR with bulge and loop specificity. Cyclin T has been localized to chromosome 12, and rescues activity in rodent cells (Wei et al, 1998).

Mechanism of Tat transactivation

An attractive two-step regulatory model has been suggested which takes into account all these data (Figure 1.5) (Jones, 1997). First, Tat is recruited to the transcription complex as part of the RNAPII holoenzyme, through its interaction with CAK/TFHII. In this case, Tat increases promoter clearance. In addition, Tat is recruited to the transcription complex via TAR as part of the TAK complex. In this case, either by increasing the local

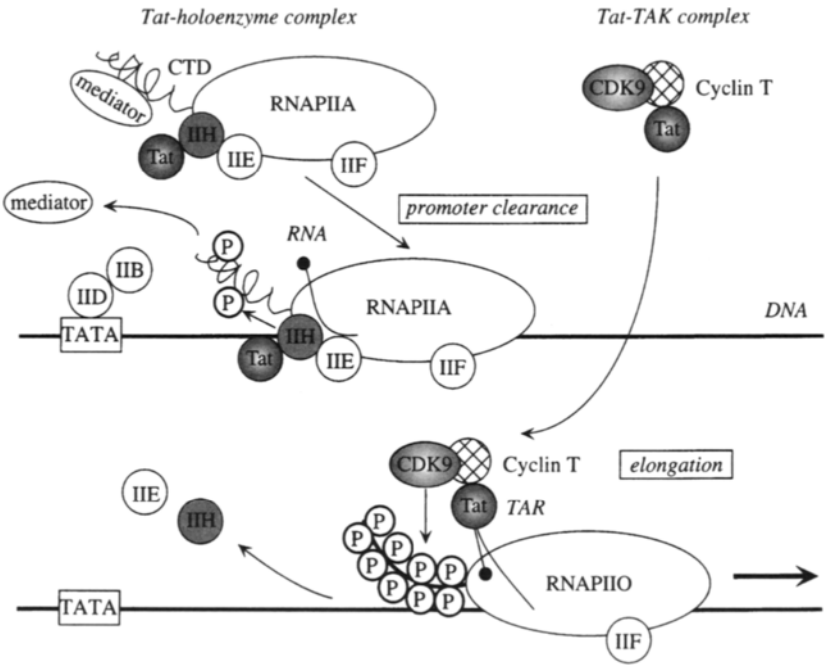

Figure 1.5 ***Proposed mechanism of action of Tat***
Tat exists in two complexes in cells, as Tat-RNAPIIA holoenzyme complex (Tat-holoenzyme complex) and Tat-TAK/P-TEFb complex (Tat-TAK complex). The former complex is part of the preinitiation complex on the HIV-1 LTR, where TFIIH and its kinase CDK7 are brought into proximity of the CTD of RNAPII. At this point, RNAPII is hypophosphorylated (RNAPIIA and RNAPIIO) and clears the promoter. Several general transcription factors (GTFs) and mediator leave RNAPII. TAR is transcribed and recruits the Tat-TAK complex, which binds to RNA. CDK9 hyperphosphorylates the CTD of RNAPII and possibly other proteins, thus converting a non-processive to processive RNAPII (RNAPIIO). TAK, which is a multisubunit complex, is shown as consisting minimally of CDK9 and its associated cyclin T.

concentration of TAK or hyperphosphorylating the CTD, Tat enhances the processivity of the early elongation complex. Initially, Tat travels with the early transcription complex bound to TAR; at later stages it disengages from TAR and stays with the modified transcription complexes (Keen et al, 1996, 1997). It remains to be determined whether the complex formed by Tat and TAK binds TAR with specificity for the loop, and, if so, whether a component of TAK is the factor encoded by human chromosome 12.

THE Rev PROTEIN: A MODULAR STRUCTURE

Rev is encoded by two exons, which overlap those encoding Tat, but are read in a different translation frame. Whereas the first exon codes for 25 amino acids, the second exon codes for 91 amino acids. The final protein has a molecular mass of 13 kDa and is modified by phosphorylation (Feinberg et al, 1986; Sodroski et al, 1986; Hauber et al, 1988). Six doubly or triply spliced mRNAs can direct the translation of Rev at early stages of replication (see Figure 1.1) (Schwartz et al, 1990a).

Transient transfection assays, using target plasmids containing an RRE/reporter gene flanked by an HIV-1 splice donor (SD) and splice acceptor (SA) sites, were utilized to map domains for Rev activity. As with Tat, extensive mutageneses studies identified two functional domains within Rev, a multifunctional domain and an effector domain. These domains, however, are located within dispensable regions (Figure 1.6A). The multifunctional domain is found between residues 14 and 56, and the effector domain between amino acids 75 and 84. Amino acids 1 to 8 and 85 to 91 are dispensable for Rev function (Malim et al, 1991).

The multifunctional domain of Rev

The multifunctional domain is responsible for three Rev functions: localization of Rev to the nucleus, binding of Rev to the RRE and the oligomerization of Rev. The nuclear localization domain is contained within amino acids 40 to 45 and can target proteins efficiently to the nucleus (Malim et al, 1989a,b; Perkins et al, 1989). An ARM from positions 35 to 50, which includes the nuclear localization domain, forms the RNA-binding domain. This ARM can functionally replace that of Tat; however, the tryptophan at position 45 is a strict requirement for binding to the RRE (Hope et al, 1990; Malim et al, 1990; Olsen et al, 1990; Subramanian et al, 1990; Zapp et al, 1991).

Unlike Tat, Rev binds to its target sequence, the RRE, as a multimer, in a ratio of up to 8 Revs to 1 RRE. In fact, Rev can even multimerize in the absence of the RRE both in vitro and in vivo. Sequences involved in oligomerization lie on either side of the arginine-rich motif (amino acids 14 to 34 and 51 to 56) (Heaphy et al, 1990, 1991; Olsen et al, 1990; Malim and Cullen, 1991; Zapp et al, 1991). Deletions of the oligomerization domain result in mutant Rev proteins that bind the RRE in a 1 to 1 ratio, and are inactive in vivo. This indicates that oligomerization is essential for Rev activity (see Oligomerization of Rev on the RRE and Mechanism of Rev action) (Malim and Cullen, 1991).

The effector domain of Rev (nuclear export signal)

The effector domain of Rev, i.e. the domain that interacts with the cellular factor(s) that mediate the nuclear export of RRE/Rev-containing mRNAs, comprises amino acids 73

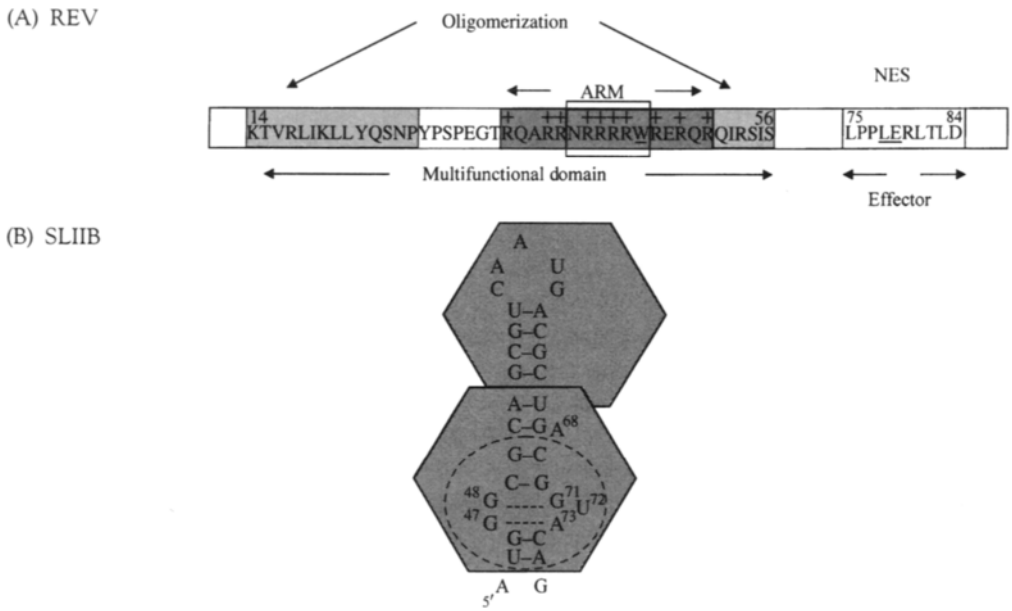

Figure 1.6 ***Domain structure of Rev and RNA structure of the SLIIB***
(A) The two functional domains in Rev are indicated below the rectangle that represents Rev. Inside are the relevant amino acid sequences. The bipartite oligomerization domain spans amino acids 14 to 27 and 51 to 56; the ARM spans amino acids 35 to 50, the positive charges are indicated above the respective amino acids, and tryptophan 45 is underlined. Boxed amino acids within the ARM (40 to 45) comprise the nuclear localization sequence. The NES spans amino acids 75 to 84; underlined are the amino acids that when replaced by DL eliminate the function of Rev (RevM10).
(B) The sequence and proposed secondary structure of the SLIIB, from nucleotides 44 to 76 in the RRE. The purine-rich loop (bubble) is indicated by the dotted circle, non-Watson–Crick base pairing is indicated by a dotted line. Rev binds to the major groove of the SLIIB and interacts with bases on either side of the groove. Two Rev molecules are positioned on the SLIIB to indicate oligomerization occurring after initial binding.

to 84. This is a leucine-rich region (LQLPPLERLTLD). Mutations in these residues abolished the activity of Rev without compromising its binding to RRE, nuclear localization or oligomerization (Perkins et al, 1989; Hope et al, 1990; Venkatesh and Chinnadurai, 1990; Malim et al, 1991). This effector domain has recently been shown to act as a nuclear export signal (NES), and is referred to here as Rev NES (see Nuclear export signals). Data obtained from domain swapping experiments have strengthened the notion that Rev NES interacts with cellular factor(s), and suffices for Rev activity. Rev NES was enough to confer Rev activity when it was assayed as a fusion protein with the MS2 coat protein and its target RNA operator (McDonald et al, 1992). Likewise, Rev NES was interchangeable with the equivalent domain from the Rev-like protein Rex (Hope et al, 1991; Weichselbraun et al, 1991).

Missense mutations in Rev NES, such as those introduced in RevM10 (a change in positions 78 and 79 from LE to DL), resulted in a *trans*-dominant negative protein (Malim et al, 1989a; Venkatesh and Chinnadurai, 1990). RevM10 can inhibit HIV-1 replication; however, instead of interacting with a cellular factor and squelching it, the