CELLULAR FACTORS THAT INTERACT WITH THE NEGATIVE REGULATORY ELEMENT IN THE 5'-LONG TERMINAL REPEAT OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1.

Dr. Kim Orchard

Thesis submitted to the University of London

for the Degree of Doctor of Philosophy.

April 1999

Royal Free and University College Medical School

and the Institute of Cancer Research,

Chester Beatty Laboratories.

ProQuest Number: U641902

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.
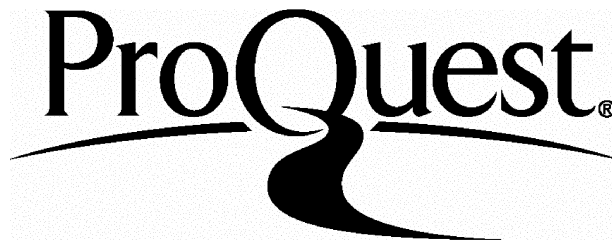
ProQuest.

ProQuest U641902

Published by ProQuest LLC(2015). Copyright of the Dissertation is held by the Author.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

To my Mother and Father

Acknowledgements

Ah! well a-day what evil looks
Had I from old and young!
Instead of the cross, the Albatross
About my neck was hung.

Samuel Taylor Coleridge,
from The Rime of the Ancient Mariner

Abstract.

Transcriptional regulation of HIV-1 gene expression has been shown to be regulated by a combination of viral and cellular proteins which bind to regulatory elements in the viral 5' long terminal repeat (5'LTR) Functionally the LTR can be divided into three regulatory regions: the TAR region, extending from nucleotides 1 to 60 relative to the start site of transcription, contains sequences with which the viral *trans*-activator Tat interacts. The adjacent region from nucleotides -1 to -78 contains the core promoter with elements crucial for both basal and Tat-induced expression. The third region, extending from -79 to -454, contains numerous elements with which a variety of cellular factors may interact, resulting in either positive or negative modulation of LTR-driven transcription.

The work contained within this thesis describes the discovery and delineation of two new transcription factor binding sites, designated as site A and site B, within the 5'-LTR of HIV-1. The majority of the work focused on site B itself, involving the characterisation of cellular proteins that specifically interacted with the nucleotide sequences in this site. Site B was found to contain a palindromic sequence TGACC involved in protein-DNA contact separated by a 9 base-pair spacer sequence that was not important for protein binding. This palindrome resembles the consensus binding site for members of the nuclear hormone receptor super-family of transcription factors. Although several members of this super-family of transcription factors were shown to interact *in vitro* with site B, the predominant protein present in T-lymphocyte nuclear extracts did not correspond to any of those previously characterised. The T-cell protein was shown to have a relative molecular size of 100-110 kD for the monomeric polypeptide and bound to site B as a dimer. Maximal binding to site B required both halves of the palindrome.

Functionally site B was shown to act as a repressor element of both basal transcription and of transcription activated by phorbol ester in T-lymphocytes. Site B was also shown to function as a retinoic acid response element (RARE) in a heterologous promoter. The ability to function as either a positive or negative regulatory element is a recognised characteristic of nuclear hormone response elements and in part is a function of the relative abundance of factors able to interact with the site or to form complexes with one another. The overall effect of site B upon LTR-directed transcription may similarly depend upon the complex interaction of multiple factors which themselves depend on the cell type, cell activation state and degree of differentiation.

List of Contents.

Chapter 6

General discussion and future directions.

List of figures and diagrams

Chapter 1

Introduction.

Chapter 1

Introduction.

1.1 Historical Perspective.

1.1.1 The emergence of AIDS

In the spring of 1981 the Centre for Disease Control (CDC) in the USA reported five cases of an unusual atypical pneumonia that had occurred in previously healthy male homosexuals in the USA (Centers for Disease Control, 1981). The causative agent in each case was a fungal organism, *Pneumocystis carinii*, which had been known to cause infections in patients immunosuppressed either as a result of disease (for example leukaemia and lymphoma) or following organ transplantation with the use of immunosuppressive drugs. These initial communications from the CDC triggered a spate of reports from physicians in the USA describing further cases of *P. carinii* pneumonia and other unusual conditions, such as Kaposi's sarcoma, chronic perianal ulceration due to herpes simplex and mucosal candidiasis, all affecting previously healthy individuals. All the patients reported were either sexually active homosexuals or intravenous drug users (IVDU) and displayed decreased T-cell response to mitogens or antigenic stimulation (Gottlieb et al. 1981; Masur et al. 1981; Siegal et al. 1981).

The term Acquired Immune Deficiency Syndrome, AIDS, was coined for what appeared to be a new disease entity. A number of potential aetiological agents were proposed for this new condition including some of the drugs favoured by homosexuals, such as amyl nitrate, or infection with agents such as cytomegalovirus (Vanley et al 1982). In 1982 AIDS related conditions were reported in haemophiliacs with no other risk factors, supporting the epidemiological evidence for an infective agent transmitted by exposure to blood or cell-free blood products and by sexual contact.

The first significant evidence for an infective agent causing the underlying pathology of AIDS cases was reported simultaneously in 1983 from Montagnier's group in Paris (Barre-Sinoussi et al 1983) and Gallo and co-workers in the USA (Essex et al. 1983; Gallo et al. 1983; Gelmann et al. 1983). Montagnier et al detected reverse transcriptase activity, cytopathic effect and viral particles in activated lymphocytes isolated from a lymph node biopsies taken from a patient with lymphadenopathy. The virus was termed LAV for Lymphadenopathy Associated Virus and similar viral isolates were rapidly obtained by the same group from patients with AIDS. Gallo reported the presence of c-type viral particles in T-lymphocytes, from a patient with AIDS,

14

which were cultured in 'T-cell growth factor' (IL-2). They were also able, using a DNA probe based on the *env* gene of HTLV-1, to detect proviral DNA sequences in T-cells of two patients with AIDS. In retrospect, as there is insufficient homology between the *env* proteins of HIV 1 and HTLV 1 to allow hybridization by blotting, the significance of this report is unclear. Further progress in establishing a causal link between a potential retroviral pathogen and AIDS was hampered by the lack of sufficient purified virus required for full characterisation and the generation of reagents required for the analysis of patients' sera. Improvements in T-lymphocyte culture resulted in the generation of continuous, high-titre virus producer cell lines and sufficient purified virus to allow characterisation of viral isolates. The virus was shown to be a retrovirus with tropism for CD4 positive T-lymphocytes (Gallo et al. 1984). Because of the apparent similarity with the other human T-lymphocyte tropic retroviruses HTLV-I and II already identified by his group, Gallo proposed the term HTLV-III for the new retrovirus, while Levy reported the isolation of a similar retrovirus which was given the designation ARV for AIDS-related virus (Levy et al. 1984). It seemed unlikely that there were several different retroviruses associated with AIDS and within a few months of their initial descriptions LAV, HTLV-III and ARV were shown to be variants of the same retrovirus (Ratner et al 1985). In 1986 an International Committee proposed the name Human Immunodeficiency Virus (HIV) for all isolates of the AIDS virus (Coffin et al. 1986).

From the original report of five cases in 1981, HIV infection and AIDS has exploded into a global pandemic. At the end of 1995 WHO published figures showed that 1.3 million cumulative cases of AIDS had been reported from 193 countries. This figure includes confirmed cases of AIDS, however, because of under-diagnosis and reporting failures, this almost certainly represents only a small proportion of all cases. It has been estimated that the true number of cases of AIDS in adults and children was in the order of 6 million, of these probably 5 million have died. Indeed, AIDS is now the main cause of morbidity and mortality in young adults in the major cities of the USA, Europe and in sub-Saharan Africa. Furthermore the long incubation period of HIV before symptomatic AIDS-related conditions appear means that far greater numbers of individuals have already been infected, the WHO estimate that 24 million adults and 1.5 million children are HIV positive with 10,000 new infections occurring daily (data from WHO 1995).

1.1.2 Origins and spread of HIV.

Four phases of the global HIV pandemic have been described: (Quinn 1996)

1. emergence
2. dissemination
3. escalation
4. stabilisation.

These phases are not unique to HIV-1 but are common for any infectious disease either arising and infecting a new, non-immune, population or reappearing in a population when the natural immunity has been lost. They should not be seen as distinct sequential episodes but overlapping phases, continuing in parallel and concurrently at different points of progression in the countries where HIV has appeared. For example, in the USA, the pandemic has stabilised at 40-60,000 new infections per year, with a similar number of deaths from AIDS, while in Asia (India, Thailand, Myanmar), the rate of new infections is increasing rapidly with an estimated 0.8% of the population of Thailand already seropositive (Weniger et al 1991). Even within the USA the rate of new infections now varies between different risk groups; amongst male homosexuals the rate is falling while in heterosexuals, particularly young women, the rate is rising and as a consequence so is the rate of perinatal infection.

1.1.3 Origins of HIV-1 and HIV-2.

Although HIV-1 and HIV-2 have certain structural, genetic and biological characteristics in common, there are sufficient differences to indicate different origins. Both are capable of causing AIDS but the pathogenicity of HIV-1 is greater than that of HIV-2 (Kanki 1991) with HIV-2 infected individuals displaying a longer 'asymptomatic' period between initial exposure to the virus and the appearance of AIDS defining illness (reviewed by Markovitz 1993) and lower viral titres during this period than seen in individuals with HIV-1 infection (Simon et al. 1993). There also seems to be a slower period of progression after the onset of symptoms (Marlink et al. 1988; Romieu et al. 1990). Finally, vertical and horizontal transmission of HIV-2 are lower than those determined for HIV-1 (Andreasson et al. 1993; Donnelly et al. 1993; Kanki et al. 1994). These differences in pathogenicity between HIV-1 and HIV-2 may in part explain the slower rate of spread of HIV-2 from West Africa to the rest of the world.

Current theories for the origins of HIV-1 and -2 are that the viruses have appeared relatively recently as human pathogens having evolved from simian lentiviruses or SIVs infecting certain species of Old World primates, (Myers et al

1992; Sharp 1994: Dale 1989; Seale 1988) and transferred to the indigenous human populations in rural areas of West Africa (HIV-2) and Sub Saharan Africa (HIV-1). HIV-2 is most closely related to SIV_{SM}, a simian lentivirus isolated from sooty mangabeys (*Cercocebus torquatus atys*) in West Africa where in the wild populations of sooty mangabeys the seroprevalence for SIV_{SM} is as high as 40% (Chen et al. 1996). Interestingly the HIV-2 seroprevalence in the human population in the same areas of West Africa (Sierra Leone and Liberia) was found to be only 0.02% (Chen et al. 1996). HIV-2 is more limited than HIV-1 in its geographical range, being confined to West Africa, while its appearance in other areas can be traced to epidemiological links with individuals of West African origin (De Cock et al 1993; Quinn 1994).

The origin of HIV-1 seems to be less certain, the closest homology of HIV-1 is with SIV strains isolated from captured wild chimpanzees (*Pan troglodytes troglodytes*), three of which have been sequenced, $SIV_{CPZ-GAB}$ $SIV_{CPZ-GAB2}$ and $SIV_{CPZ-ANT}$ (Huet et al. 1990; Janssens et al. 1994; Vanden Haesevelde et al. 1996). Full sequence comparison was possible with $SIV_{CPZ-GAB}$ and $SIV_{CPZ-ANT}$ but only partial sequence comparison of *pol* from $SIV_{CPZ-GAB2}$ has been made. However it is not clear whether chimpanzees form the natural reservoir for these HIV-1 related SIV strains. Seroprevalence in the wild population of chimpanzees is low, in one study only 3/94 wild-captured animals tested were positive for SIV (Peeters et al. 1989). This population sample may be biased by the testing juvenile chimps, certainly in Sooty mangabeys seroprevelence is significantly higher in adults than juveniles, reflecting sexual transmission and possibly by fighting.

Two alternative hypotheses have been suggested to explain the relationship of HIV-1 and SIV_{CPZ} and the apparent lack of a simian reservoir of virus. The first suggests that SIV_{CPZ} arose from a recent human-to-chimpanzee cross infection; the second hypothesis proposes a third, as yet undiscovered, ancestral simian lentivirus that has recently infected both humans and chimpanzees.
The non-human lentiviruses form five distinct clades, SIV_{CPZ} from chimpanzees, SIV_{SM} from the sooty mangabey, SIV_{SYK} from Sykes' monkey, SIV_{AGM} from African Green monkeys and SIV_{MND} from mandrills (Sharp 1994). The human lentiviruses which have been isolated and sequenced, rather than forming a distinct branch of their own, show clustering with two distinct SIV lineages, that of SIV_{CPZ} for HIV-1 and SIV_{SM} for HIV-2 isolates. The same phylogenic analysis has been applied to different HIV-1 isolates from several countries and can be used to trace the possible route of spread of the virus around the world. Based on DNA sequence comparison HIV-1 isolates have been

classified into three major groups; group M ('main'), which is responsible for the global pandemic of AIDS, has been further divided into ten subtypes (A-J) based on sequence comparison of the *env* and *gag* regions; group O ('outlier') which is more divergent and represented by a few viral isolates from the Cameroon, Gabon and Equitorial Guinea (Gurtler et al. 1994; Vanden Haesevelde et al. 1994) and group N ('non-M, non-O'), represented by isolates from 2 individuals in the Cameroon and appears to be a mosaic of SIV_{CPZ-US} and HIV-1. The presence of three distinct HIV-1 lineages has been suggested to demonstrate that cross-species transmission of HIV-1 has occurred at least three times, represented by the three 'founder' lineages. The group M subtype B predominates in the Americas and Europe possibly reflecting an initial small "inoculum" (Gao et al 1999).

The earliest evidence of human infection by HIV-1 has been obtained retrospectively from tissue and serum samples taken from patients who had died from unusual infectious diseases. Sera collected in Zaire in 1959 have been found to be contain antibodies against HIV-1 (Huminer et al 1987). The earliest definite report of HIV-1 positive sera outside of Africa were obtained from a Norwegian sailor, his wife and daughter. The sailor became symptomatic with symptoms suggestive of the early stage of HIV-1 infection, sera stored in 1971 were shown later to contain HIV-1 antibodies (Froland et al 1988). Some early reports of unusual persisting infections in patients with 'late onset immunodeficiency syndrome' may have been in retrospect due to HIV infection (for example Freeman et al 1977).

Further clues as to the origins of the human lentiviruses can be inferred from the interaction of the simian lentiviruses with their hosts. The simian lentiviruses are non-pathogenic for their natural hosts but become pathogenic when cross-species infection occurs. This feature may reflect the length of time a virus and host species have been associated with a gradual evolution from high pathogenicity to asymptomatic but persisting infection. For example, the normal host of Herpes B virus is the macaque monkey, in which it causes a relatively benign but persisting infection rather like Herpes simplex in humans. However, Herpes B causes a lethal illness in humans infected through contact with monkeys carrying the virus. There is no advantage for the virus in killing the host animal, a persistent infection in a healthy host would permit a wider dissemination of progeny virus. The apparent lack of pathogenicity of the chimpanzee SIVs and the failure of HIV-1 to cause symptomatic disease in chimpanzees despite causing sero-conversion favours the initial hypothesis that HIV-1 has recently evolved from SIV_{CPZ}.

In conclusion, HIV-1 probably evolved in the relatively recent past, originating from lentiviruses present in the wild population of chimpanzees in Central Africa. Sequence analysis of HIV-1 isolates shows a clustering of HIV-1 into three major groups, M, N and O, possibly as a consequence of three independent introductions of simian lentiviruses into the human population.

1.1.4 Transmission of HIV.

There are three main modes of transmission of HIV-1 and-2, sexual (both homo- and hetero-), parenteral (including percutaneous) and perinatal. The relative importance of each route of transmission varies between different countries and even within countries. Sexual transmission is the most significant responsible for approximately 75% of infections world wide. Heterosexual contact was probably the important route of transmission during the early phase when HIV was limited to the areas where initial transmission to humans occurred. Homosexual transmission of the virus was important in the rapid escalation of the infection in developed countries, presumably as a direct consequence of the relative promiscuity of male homosexuals at that time. There is epidemiological evidence suggesting that the risk of transmission of the virus is also increased by penetrative anal sex compared with vaginal intercourse (de Vincenzi 1994; European Study Group on Heterosexual Transmission of HIV. 1992; Guimaraes et al 1997). In other parts of the world however, homosexual transmission has not been as important as heterosexual transmission, for example in Thailand, Africa and India. In these countries female prostitutes are a reservoir of the virus and epidemiological evidence indicates a strong association between HIV and sexually transmitted diseases such as Gonorrhoea, Chlamydia, syphilis and Trichomoniasis (Ghys et al. 1997; Laga et al. 1991 and 1993).

Parenteral transmission includes the transfusion of blood products, exposure to blood through needle or syringe-sharing by intravenous drug users (IVDU'S) or in healthcare environments where sterilisation procedures are inadequate. Transmission in IVDU's varies but HIV-1 seropositivity in this risk group is as high as 55-60% in some communities (Des Jarlais et al 1989, 1992). The transmission of HIV via blood products was a significant problem in the early 1980's prior to the onset of routine screening measures. The problem was compounded in the USA and other countries where blood donors were paid for each donation, this encouraged donors from IVDU's. The development of reliable serological tests for HIV-1 and HIV-2 and the routine screening of all blood donations in the USA and the UK in 1985 virtually eliminated this as a source of transmission of the virus. It has been estimated that currently in the

UK the risk of transfusion related HIV-1 infection is between 1 in 750,000 - 1 million donations (McClelland et al. 1996) while in the USA the risk varies between 1 in 40,000 to 1 in 225,000 (Dodd 1992). A particular problem existed for individuals who required blood products that had been pooled from numerous donors e.g. haemophiliacs receiving coagulation factor concentrates, batches of which are produced by pooling the plasma from several hundred blood donors. Unfortunately a large number of haemophiliacs in the UK that received factor concentrates during the early 1980's became HIV-positive (Lee 1996).

Perinatal transmission of HIV-1 is increasing in parallel with the increasing number of women infected with HIV-1 world-wide. Transmission of the virus can occur in utero, at birth or in the post-natal period by via breast milk (Gray et al 1996) with transmission rates ranging from 13% to 52% in studies from different countries (Peckham and Gibb 1995; Dabis 1993; Siena Workshop 1995).

1.1.5 Dissemination of HIV-1 and HIV-2.

As outlined in section 1.1.3, the geographical localisation of HIV-2 infection to the West Coast of Africa suggests that this virus made the transition from monkeys to humans in this area. The distribution of HIV-1 infection in Africa is less clear-cut but the weight of evidence indicates probable origins in rural areas of Uganda, Tanzania, Rwanda, Burundi. The virus probably had been endemic in such rural areas for some time spreading slowly by heterosexual contact. Massive social and political changes began to occur in the 1950's and 60's with a change from the relative stability of colonial rule to a period of rapid flux. Changes in population dynamics, industrialisation and the movement of individuals from rural areas to the more densely populated cities allowed an increased rate of dissemination of the virus through sexually active members of the population. The long relatively asymptomatic period of infection allowed dissemination 'silently' as infected individuals could carry on with the activities associated with spreading the virus. Some HIV-related deaths must have occurred during this period but the high prevalence of other infectious diseases in the regions, for example malaria, parasite infestation and tuberculosis, may have obscured the then infrequent AIDS cases.

Dissemination of the virus world-wide was the result of the increased level of international travel over the last 20 years, particularly air travel, and the increased popularity and cheapness of 'exotic' holidays. In parallel with the dissemination of the virus were the dramatic changes in social order in sub-Saharan Africa with increased availability of commercial sex, decline in

welfare and health and increased poverty. Medical services declined and there was a clear increase in the incidence of other sexually transmitted diseases including gonorrhoea, syphilis, Chlamydia and Trichomoniasis (Laga et al 1991; Schechter and Laga 1995; Vuylsteke et al. 1993) (reviewed by Quinn 1994).

The major escalation in the spread of HIV-1 began in the early to mid 1980's. The virus was transmitted by high risk activities such as sexual contact, both homosexual and heterosexual, intravenous drug abuse, blood product transfusion. The initial appearance of HIV-1 and AIDS related diseases amongst homosexuals in the USA was a result of rapid dissemination of the virus by the promiscuous life-style of individuals particularly in certain cities in the USA. The clustering of unusual infections occurring in an otherwise apparently healthy individuals in a Western nation alerted the health authorities to the presence of a new disease, even though cases of HIV-1 related illness must have occurred in Africa many years earlier these were almost certainly sporadic, affecting individuals in rural communities and masked by the endemic infectious diseases.

The discovery of HIV-1 as the causative agent of AIDS, a growing knowledge of the mode of transmission and the accompanying publicity together with a massive health education programme has resulted in a stabilisation in the spread of HIV in the USA, Western Europe and Australia. In the last two years the annual rate of new cases of HIV infection has fallen in the USA and the pattern of newly infected individuals has also changed. In the USA the initial cases of AIDS were in male homosexuals and AIDS soon became the major cause of death amongst men aged 25-44 years (Centers for Disease Control 1993). However the pattern of infection began to change and although male homosexuals still account for the majority of new cases of HIV-1 infection, if the *proportional* rate of change amongst different risk groups is considered, the highest rate of change is for heterosexuals and indeed that seen in homosexuals is falling (fig 1.1.1).
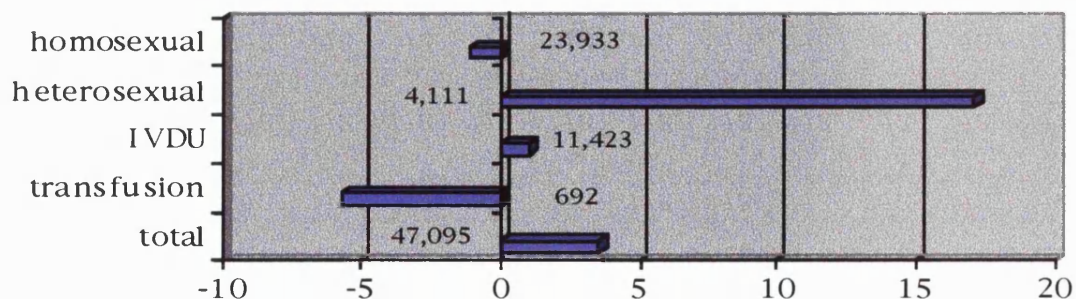
Fig 1.1.1   Changes in AIDS cases 1992 vs 1991
         Percentage proportional change in AIDS cases reported to the
         CDC in 1992 vs 1991, shown by risk category. (from Quinn 1994)

21

In the Americas, about three million cumulative cases of HIV-1 have occurred with one million in North America and two million in Latin America and the Caribbean (WHO data). In the USA there have been 500,000 cases of AIDS reported with 300,000 deaths (Centers for Disease Control 1995).

In the UK the annual rate of newly diagnosed HIV-1 cases continues to rise. In 1996 there were 2896 new HIV-1 infections reported compared with 2683 in 1995, a rise of 7.9%. The cumulative total of reported HIV-1 infection was 28,447 cases up to December 1996. Analysis of the results according to risk category show that in contrast to the results from the USA male homosexual transmission continues to display a year-on-year increase from 1474 cases in 1995 to 1634 in 1996, a rise of 11%. Heterosexual transmission rose from 781 cases in 1995 to 858 in 1996, these were mainly in the 'exposure abroad' category. The number of AIDS cases had also risen to 1862, an increase of 18% from the 1995 figure (data from Communicable Diseases Surveillance Centre).

Within the rest of Western Europe the pattern of HIV-1 infection varies, in Scandinavia the majority of cases of AIDS have been amongst homosexual and bisexual men while in Italy and Spain the majority of HIV-1 cases are in IVDUs. Data relating to Eastern Europe is limited but that which is available indicates a major epidemic is emerging. Sexual transmission of the virus, both homosexual and heterosexual, appears to be most significant but there have been some important iatrogenic infections, particularly in children. In the Czech and Slovak republics the majority of infections are related to homosexual transmission. In Poland, the majority are due to parenteral transmission amongst IVDUs. Most tragically of all are the cases of HIV-1 infection caused by unscreened blood transfusions given indiscriminately to children in government-run orphanages during the 1980's in Romania. An estimated 1,000-2,000 children were infected with HIV-1, because of the lack of screening for virus combined with the multiple use of needles and syringes. The majority of these children have now developed AIDS. A similar but smaller focus of HIV-1 infection occurred in Kalmykia in Russia, where children were exposed to infection by reused unsterilised needles and syringes.

1.1.6 Future predictions of the HIV pandemic.

The current WHO predictions are that, by the year 2000, there will be 26 million people infected with HIV. The cumulative total for HIV infection by 2000 will be 40 million with 90% of these in the developing nations in Africa, the Far East and the Indian sub-continent. The cumulative number of deaths from AIDS by the millennium will be in the order of eight million. The pattern of dissemination and escalation of infection initially seen in Western countries in the 1980's is being repeated in the developing countries. While in the West the peak of new cases has, in the case of the USA, been reached with the number of new cases of HIV infection balanced by deaths due to AIDS, in the developing nations of Africa and Asia the epidemic is still in an expansive phase. The spread of HIV in Sub Saharan Africa is predicted to continue with the peak of AIDS cases at around 750,000 cases per annum in 2005. In Asia the peak of AIDS cases is predicted to be in the order of 850,000 per annum by 2010 (WHO data taken from Quinn 1996).

The major impact of the epidemic will continue to fall on the developing countries, here the effect of antiviral drug therapy will be negligible. Effective, affordable measures to control the spread of infection must continue to come from public health initiatives. Experience in the West has shown that education programs, particularly if targeted at the 'at risk' populations, are effective in increasing the public awareness of HIV. The decline of new cases of HIV-1 in male homosexuals in the USA appears to be the result of such educational policies, together with a highly motivated risk group. Specific health policies such as the distribution of free condoms, health screening of sex-workers and the aggressive treatment of sexually transmitted diseases have already been effective in reducing the spread of infection in high risk groups in Africa and Asia (Hanenberg et al. 1994; Laga et al. 1994). Needle exchange programs have been started in a number of different countries with the aim of reducing the parenteral route of infection amongst IVDUs. Unfortunately the social, emotional and political issues that arose with the initial appearance of HIV in the West has almost certainly handicapped the implementation of conventional public health measures used in controlling the spread of severe infections.

Measures such as population screening with the purpose of identifying seropositive individuals, contact tracing, legal measures to restrict infected individuals disseminating virus, all these measures are used in the effective control of, for example, tuberculosis, syphilis and gonorrhoea. Hopefully, as the public understanding of HIV grows, this virus and the disease associated with it will be viewed in the same way as the 'conventional' infectious diseases and allow the implementation of public health measures to contain the spread of infection. Certainly if, as has been suggested, the early treatment of HIV positive individuals with combination drug therapy has a significant impact on the progression and outcome of the infection then there will be increasing pressure to identify HIV-infected individuals before the onset of symptoms.

The hope of an effective vaccine and the availability of cheap drug therapy remain distant for the majority of the worlds population, for the present time and immediate future efforts must be directed towards education and carefully targeted public health measures to alter the social, behavioural and environmental factors that are involved in the spread of HIV.

1.2 General description of retroviruses and HIV-1

1.2.1 The retroviruses.

Retroviruses are enveloped, RNA-containing viruses that replicate via a DNA intermediate, a process that requires the viral-encoded enzyme RNA-dependent DNA polymerase, also termed reverse transcriptase (RT). They are primarily found in vertebrate hosts, although the presence of retrovirus-like particles in other animals has been reported.

In the classification of viruses, retroviruses form a large family called Retroviridae which is further divided into three subfamilies, the Oncoviruses, Lentiviruses and the Spumaviruses. This classification is based primarily on pathogenicity rather than genome relationships (Teich 1985). Retroviruses are further described according to the following features:

a)   Virion structure under the electron microscope.

This was an early classification based on electron microscopy (EM) studies of virion formation, release and the form of mature viral particles and is still used to describe the early retroviruses discovered. Viruses are placed into one of four groups, A to D, although this classification tends not to be applied to the more recently described viruses. Small A-type particles represent the immature particles of the B- and D-type viruses as seen in the cytoplasm of infected cells. B-type particles bud as mature virions from the cell membrane, the mature virion contains an eccentrically placed electron-dense core. The prototypic B-type virus is Mouse Mammary Tumour Virus (MMTV). No intracellular particles are seen in cells infected with C-type retrovirus, instead mature viral particles bud directly from the cell membrane via a crescentric "c" shaped condensation. Budding of progeny virions can occur from the cell surface or into intracellular vesicles. The majority of retroviruses fall into this morphological group including such diverse viruses as the avian and murine leukaemia viruses, bovine leukaemia virus and the human T-Cell Leukaemia viruses HTLV-1 and HTLV-2. D-type particles resemble B-type in forming ring-like structures within the cytoplasm before budding from the cell membrane. The Mason-Pfizer Monkey Virus is the prototypic D-type virus.

b)   Utilisation of cell receptors.

e.g. amphotrophic, ecotropic, xenotropic and polytropic. This classification has practical uses in defining the range of host cells susceptible to infection by a particular virus.

c)   The presence or absence of an oncogene.

d)   Whether exogenous or endogenous.

e)   Genome structure and nucleotide sequence.

This classification is the most important allowing the grouping of viruses into families based on genome organisation and complexity.

1.2.2 Taxonomic classification of Retroviruses.

All retroviruses belong in the large family Retroviridae which consists of three subfamilies, Oncoviridae, Lentiviridae and Spumaviridae.

The subfamily Oncoviridae is further divided into five sub-groups and contains all the oncogenic retroviruses such as Murine Leukaemia Virus (MuLV), Avian Sarcoma Virus (ASV) and Murine Mammary Tumour Virus (MMTV), together with numerous related retroviruses. This subfamily includes the first two known human retroviruses identified, the Human T-cell Lymphotropic Viruses, HTLV-1 and HTLV-2.

Members of Lentiviridae are the aetiological agents of so called 'slow-viral' infections and include the visna-maedi virus, Caprine Arthritis Encephalitis Virus, Equine Infectious Anaemia Virus and the Immunodeficiency viruses including Human Immunodeficiency Viruses types I and II. Other, related Lentiviruses are associated with immunodeficiency syndromes in other animal hosts such as cats (Feline Immunodeficiency Virus, FIV), non-Human primates (Simian Immunodeficiency Viruses, SIV). The genome of these viruses are characterised by a complex combination of genes with extensive splicing events involved in their expression.

The Spumaviruses are so named because of the effect these viruses have upon infected cells. Extensive vacuolation gives the cells a 'foamy' appearance (spume: derived from the Latin spuma for foam or froth). Spumaviruses are able to cause chronic, persisting infections in mammalian hosts including monkeys, cattle, cats and possibly humans but without any evidence of harm to the host or, in the case of humans, any clinical disease.

1.2.3 Retroviruses and disease.

The first association of a retrovirus with disease came in the early 1900's when the transmission of chicken leukaemia and sarcoma was demonstrated to be mediated by a filterable agent (Ellerman and Bang 1908; Rous 1911), later identified as a retrovirus which was subsequently named after one of the scientists associated with the original observation, the Rous Sarcoma Virus. The first mammalian retrovirus, murine leukaemia virus, was isolated from in-bred mice in 1951 (Gross 1951), over the following three decades numerous other retroviruses were shown to cause a wide variety of malignancies in numerous species but none could be found in human tissue, either normal or

malignant. Experience with animal retroviruses did however provide valuable information crucial in the search for the human counterparts with improvements in methods of cell culture, viral assay techniques, especially RT assays, electron microscopy and genome analysis.

1.2.4 Retroviruses in humans.

The first Human retrovirus was isolated from T-lymphocytes taken from a patient with mycosis fungoides, a chronic T-lymphocyte malignancy, this virus was called Human T-cell Leukaemia Virus, HTLV-1. The *in vitro* culture of HTLV-1 was possible only because of the earlier discovery, cloning and large scale production of T-cell growth factor, now called interleukin 2 (IL-2) (Poiesz et al. 1980). Despite the initial isolation of HTLV-1 from a patient with mycosis fungoides no aetiological association with this disease has been made. However, HTLV-1 is associated with two human disorders, Adult T-cell Leukaemia/Lymphoma, ATLL (Catovsky et al. 1982) and Tropical Spastic Paraparesis, a progressive degenerative neurological disorder (Bangham et al. 1988; Gessain et al. 1985). As described in section 1.1, the discovery that the human acquired immune deficiency syndrome was caused by a retrovirus was made in the early 1980's soon after the initial description of patients with AIDS. Two related lentiviruses, HIV-1 and HIV-2 are associated with AIDS and although there have been some discussions as to whether these human retroviruses are the aetiological agents of AIDS (Duesberg 1989a and 1989b) the overwhelming weight of evidence indicates that these viruses alone are responsible for the profound immunosuppression resulting from the depletion of CD4 T-lymphocytes in infected individuals. The following sections describe the characteristics of HIV-1, those of HIV-2 lie out-side the scope of this thesis.

1.2.5 Physical description of HIV-1.

1.2.5.1 Structural proteins.

Electron microscopy has revealed the HIV-1 virion as an icosahedral core surrounded by a lipid bilayer (Gelderblom et al. 1987; Marx et al 1988).

Fig. 1.2.1.

<u>Schematic Diagram of HIV-1 Virion</u>.

$p7^{gag}$ $p24^{gag}$
$p17^{gag}$
$p9^{gag}$
host-determined membrane proteins
ssRNA genome,one of two separate strands.
reverse transcriptase p55/61
gp 41^{env}
gp120^{env}
bilipid membrane
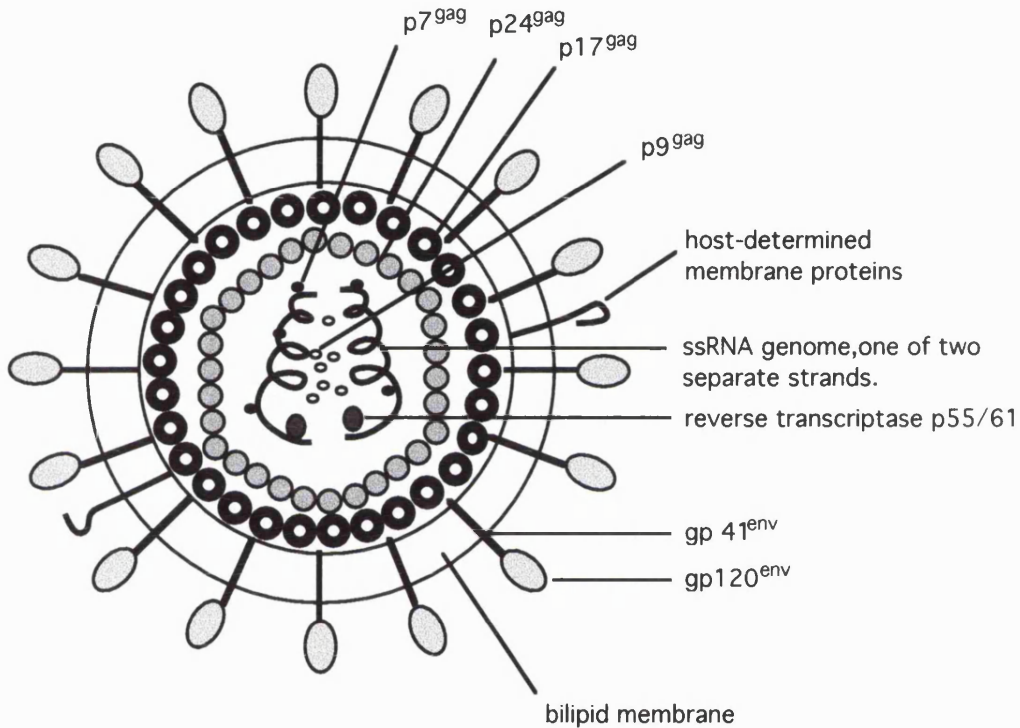
<u>Major Structural Proteins of HIV-1.</u>

(Nomenclature from Leis et al. 1988)

Trivial name.	Standard name.	
gp120 ^{env}	SU	Surface antigen
gp 41 ^{env}	TM	Transmembrane protein
p 24 ^{gag}	CA	Core antigen
p 17 ^{gag}	MA	Matrix antigen (myristylated)
p7/9 ^{gag}	NC	Nucleocapsid

The core structure consists of several proteins derived from the p55gag-precursor which is cleaved by the viral encoded protease (Debouck et al. 1988; Sanchez-Pescador et al. 1985). The p55gag-precursor is vital for the assembly and release of viral particles, described in more detail below. The phosphorylated p24 protein forms the major structural component of the core, while p17 coats the inner layer of the membrane and is myristylated at the N-terminal glycine (Gottlinger, Sodroski, and Haseltine 1989). This modification acts as a membrane-localisation signal and appears to be vital for virion assembly prior to budding, inhibition of myristylation resulted in a complete failure of viral assembly and loss of infectivity (Bryant et al. 1989). The p9 protein is bound directly to the viral genomic single-stranded RNA through a zinc-finger structure and together with p9 forms the nucleoid core.

Also contained within the capsid are two copies of the RNA genome with several viral-encoded enzymes, p66/p51 reverse transcriptase (RT), p32 integrase (IN), and p10 protease (PR). These are derived from the precursor polypeptide p160gag-pol after cleavage by the viral protease.

1.2.5.2   The envelope.

For simplicity the label gp160^{env} will be used for the precursor and gp120^{env} and gp41^{env} for the mature glycoproteins, although the exact molecular weights vary slightly between strains of virus.

The lipid bilayer is derived from the cytoplasmic membrane of the host cell and contains a number of external spikes formed from the two major envelope proteins gp120^{env} and gp41^{env}. These two envelope protein are derived from a common precursor, initially as an 88 kDa polypeptide. The addition to the polypeptide backbone of high-mannose N-linked carbohydrate chains occurs in the rough endoplasmic reticulum, resulting in the formation of gp160^{env} glycoprotein which undergoes endoproteolytic cleavage by a cellular trypsin-like protease to generate the mature envelope glycoproteins gp120^{env} and gp41^{env}. Only 5-15% of gp160^{env} is cleaved to produce the derivative glycoproteins, the majority is delivered to lysosomes and degraded (Willey et al. 1988).

The major glycoprotein gp120 is located on the outside of the virion and is held there by non-covalent interaction with gp41 which spans the membrane, carboxy-terminal on the inside of the virion, amino-terminal externally. For HIV-2 and SIV-mac, env has been found as homodimers (Rey et al. 1990). For HIV-1, evidence for env as dimers, trimers and tetramers has been found by

several groups, the complex held together by non-covalent bonds between gp41 molecules (Pinter et al. 1989; Schawaller et al. 1989; Thomas et al. 1991). Both sub-units are glycoproteins, gp120 having over 30 potential sites for N-linked glycosylation while gp41 has far fewer, limited to the external domain. The sequences of gp120 and gp41 show considerable variation between isolates, but there are regions or structural features that are highly conserved (Modrow et al. 1987; Starcich et al. 1986).

Within gp120 there are 18 cysteines that are conserved together with four highly conserved regions (C1-C4) and regions of greater variability (V1-V5). The significance of these features in generating the 3-dimensional structure of the envelope protein has only been inferred from the results of mutational analysis and epitope mapping (Kowalski et al. 1987) as crystallographic data are not yet available. The two obvious functions of gp120 are CD4 binding and gp41 association. The carboxy-terminus has been shown to contain sequences involved in CD4 binding together with several amino-acids scattered along the entire gp120 protein, suggesting that the binding site is formed by complex folding. Binding of gp120 to gp41 has been mapped to amino-acid residues in both the N- and C-termini (Helseth et al. 1991; Ivey-Hoyle et al 1991) suggesting that gp120 wraps around the external portion of gp41, covering the hydrophobic fusogenic domain.

1.2.5.3   The V3 loop and cell tropism.

A third important domain of gp120 is the V3 loop. This structure is important for the virus-cell fusion event, at least for laboratory strains of HIV-1. Evidence for a further function of the V3 loop has been presented by several groups (Hwang et al. 1991; O'Brien et al. 1990; Shioda et al 1991), sequences within V3 have a significant role in determining the tropism of HIV-1 strains for either T-lymphocytes (T-tropic) or monocytes (M-tropic) (Choe et al. 1996). The role of the V3 loop, the CD4 receptor and the cellular co-receptors necessary for the fusion and entry of the virus will be discussed in later sections.

The transmembrane component of *env*, gp41, has limited sequence variation, presumably because of greater constraints on structure and function. The protein has three distinct regions; an internal sequence which may interact with the structural protein p17. A role in syncytium formation by the virion has been suggested (Ashorn et al 1990). The transmembrane region contains hydrophobic residues and in addition a single lysine indicating that the interaction with the membrane is not simply by hydrophobic interaction.

The external region has several functions which are determined by different domains; those that interact with gp120; a domain involved in oligomerization; the amino-terminal fusion domain which has homology to the fusogenic · domain of paramyxovirus fusion proteins (Gonzalez-Scarano et al. 1987). The N-terminus contains two highly conserved cysteines both of which are required for correct processing of the precursor gp160 and for infectivity (Syu et al. 1991), these probably form a disulphide bridge to maintain an immunodominant loop.

1.2.6   Genome structure.

1.2.6.1   Physical   structure.

The HIV-1 genome has physical features common to all retroviral species, essentially a positive sense mRNA molecule containing all of the viral coding and non-coding sequences. Modifications to the nascent viral full-length RNA are performed by cellular machinery. The 5' end is capped during synthesis with m7G5'ppp5'Gmp and the 3' end polyadenylated, there is a conserved polyadenylation signal sequence. The importance of these modifications to the genomic RNA, more appropriate for mRNA which will be translated, is unknown but are probably required for stability and to avoid degradation in the cytoplasm.

The viral genomic single-stranded RNA is present as a dimeric structure with two identical molecules joined at their 5' region by the Dimer Linkage Structure (DLS). The DLS has been shown to be formed by the non-covalent association of sequences in the 5'LTR with a core sequence PuGGAPuA common to many retroviral species (Ehresmann et al. 1991). The two genomes are also associated at other points along their length but by weaker interactions than in the DLS.

Associated with the viral genomic RNA are small RNA molecules, the most important of these is a single ^{lys}tRNA molecule per genome. This is bound by Watson-Crick pairing of its 3'-terminal to a specific sequence near to the 5'-end of the genome, termed the primer-binding site (PBS). This tRNA molecule serves as a primer from which the viral reverse transcriptase initiates synthesis of the positive strand of DNA (section 1.2.8.4 and fig.1.2.5).

1.2.6.2   Genome organisation.

Figure 1.2.2

Comparison of prototypic retroviral and HIV-1 viral genomes:

A. Prototypic oncoviral genome MMTV
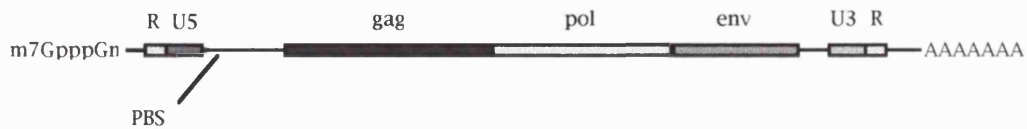
B. HIV-1 viral genome
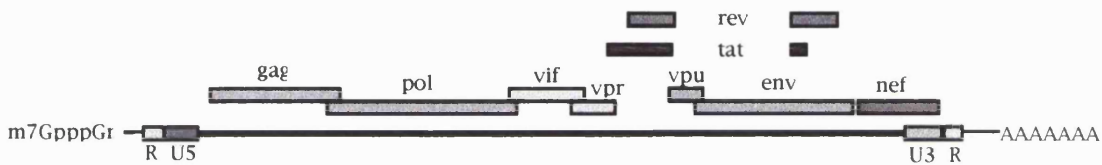
Fig. 1.2.2:

Although the overall organisation of the genomes of HIV-1 and MMTV are similar, that of HIV-1 shows far greater complexity.

In addition to the three genes encoding *gag*, *pol* and *env*, HIV-1 also contains ORFs for six additional genes between *pol* and the 3' poly A. These are *vif*, *vpr*, *vpu*, *rev*, *tat* and *nef*. These are discussed in detail below.

1.2.6.3   The structural genes.

The *gag* gene codes for the various structural proteins present in the viral nucleocapsid. It is translated from full length unspliced 9kb mRNA that appears late in the infection producing the p55gag precursor protein, this is cleaved by the viral protease producing, from the amino terminus, p17 (MA), p24 (CA) with p9 (NC) and p7 from the carboxy terminus. The components of these gag proteins have important functions in the assembly and maturation of progeny virions.

The *pol* gene codes for the viral protease, reverse transcriptase and integrase. It is also translated from unspliced 9kb mRNA producing a gag-pol fusion protein p160gag-pol. The gag and pol sequences overlap by 241 bases, pol is produced by a -1 base pair ribosomal frame-shift (Jacks et al. 1988). The pol gene product is cleaved by the viral protease to release the p10 protease (PR), p66/p55 reverse transcriptase (RT) and the p32 integrase (IN).

1.2.6.4   Viral genes encoding the 'accessory' proteins.

Retroviruses contain three long contiguous open reading frames (ORFs) encoding the *gag, pol* and *env* sequences from which the corresponding proteins are translated forming the entire requirement for structural and enzymatic function needed to produce progeny viral particles. Members of the Lentivirus sub-family characteristically carry additional ORFs which code for auxiliary proteins detectable in the infected cell. HIV-1, HIV-2 and the SIVs all have six additional ORFs with mRNA produced by complex splicing events. Two of these genes, *Tat* and *Rev* are absolutely required for productive viral infection, mutations affecting these genes severely affect viral replication (reviewed by Cullen 1992), Tat functions as a powerful *cis*-acting transactivator of the viral LTR while Rev stabilises full-length mRNA, an essential requirement for the transition from the early phase of infection to the late phase with production of progeny viral particles. The remaining genes consist of, in HIV-1, *nef, vif, vpr, vpu,* and in HIV-2 and SIV *vpx* which appears to be equivalent to *vpr* with which it has significant homology. Early mutational analysis of these genes indicated that they were not required, at least *in vitro*, for viral replication, and hence the proteins encoded by these genes were called 'accessory' proteins. However it has become increasingly apparent that these genes have important roles in the full natural history of the virus, possibly playing subtle but important functions in the interaction of virus with the host cell, and in particular with host defence mechanisms (reviewed by Trono 1995 and Subbramanian 1998). All are highly conserved in primary isolates again indicating the importance of these genes to the virus.

The Tat, Rev and Nef proteins are synthesised early from Rev-independent mRNAs that have been multiply spliced, all approximately 2 kb in size, while Vif, Vpr and Vpu are expressed late from singly spliced ~4 kb Rev-dependent mRNA (Cullen 1992).

1.2.7 Accessory proteins.

1.2.7.1   Nef.

Initially termed 3'-orf, E', F, orfB, *nef* is located at the 3' end of the proviral genome partially overlapping the U3 region (Ratner, Gallo, and Wong Staal 1985; Wain-Hobson et al. 1985). The ORF codes for a 25-27kDa protein which is myristylated at the N-terminus and associates with the cell membrane and cytoskeleton (Harris and Coates 1993; Niederman et al. 1993). Antibodies against nef are present in 50-80% patients with HIV-1 infection, appearing early in the course of infection (Cheingsong-Popov et al. 1990; Lefrere et al. 1991). *Nef* derives it's now inappropriate name from early studies using mutational analysis and transient transfections, i.e. **negative factor**. Mutation of *nef* in infectious recombinant virus resulted in the generation of viruses that replicated more rapidly and to higher titre than the 'wild type' strains (Luciw et al 1987; Terwilliger et al. 1986). Transient transfections using nef-expression vectors cotransfected with an LTR-driven CAT reporter gene appeared to show in inhibitory effect of Nef on LTR expression; (Ahmad and Venkatesan 1988; Niederman et al 1989). Subsequently, in more carefully designed experiments, no inhibitory effect could be demonstrated either on viral replication or LTR-driven transcription (Hammes et al. 1989; Kim et al. 1989a). The key report indicating the vital function of nef *in vivo* came from Desrosiers group in 1991 (Kestler et al. 1991). They infected rhesus monkeys with SIVmac, a virus that normally causes an AIDS like illness in these monkeys associated with CD4 depletion. Three infectious clones were used, the originating infectious viral clone was SIVmac239, which had a premature stop signal at codon 93 caused by a single base-pair change, a corresponding *nef* +ve virus was engineered, 'SIVmac239/nef-open', by correcting this point mutation. A third infectious clone was created with a 182 bp deletion in the coding sequence for *nef*. No differences in the replication rate of these three infectious clones was shown *in vitro* when transfected into T-cell line CEMX174, in primary monkey T-lymphocytes or monkey alveolar macrophages. However the behaviour of these three viral clones differed markedly *in vivo* when used to infect rhesus monkeys. The monkeys infected with the SIVmac239/nef-open strain demonstrated features of active and progressive infection with high viral loads, lymphadenopathy and death from AIDS. The animals infected with SIVmac239/nef-stop also developed active disease, virus isolated and sequenced

through *nef* showed loss of the stop mutation and the recovery of the full *nef* ORF, this was demonstrated as soon as two weeks following the initial infection. The animals that received the *nef*-deletion mutant strain had consistently low viral load at least 100-fold less than the other groups, and remained asymptomatic for the duration of the study. Virus could be recovered from the animals over one year after the initial exposure and strong antibody responses demonstrated. Thus, virus lacking a functional *nef* gene and unable to revert to nef-positivity, possessed a much reduced pathogenicity for the host even though the animals were persistently infected. These results explained the apparent heterogeneity of *nef* in laboratory strains of HIV and SIV many of which have mutations causing premature termination. There is no selective pressure to conserve *nef* when passaged *in vitro*, whereas *in vivo* a strong selective pressure operating to favour the viral progeny with functional nef protein.

Since this important observation a number of functions for nef have been proposed. More importatly, some indications of how these different functions relate to the pathogenicity associated with this protein, have been obtained. Nef specifically down-regulates CD4 expression by rapid endocytosis and lysosomal degradation and appears to depend on 20 residues from the cytoplasmic domain of CD4 that are membrane proximal (Aiken et al. 1994; Garcia et al 1993) although it is not clear whether Nef interacts directly or indirectly with CD4.

A second property of Nef is an apparent disturbance of T-cell activation, with both inhibitory and enhancing effects have been reported. Baur et al (Baur et al. 1994) constructed a CD8-Nef chimera, which in Jurkat cells produced an accumulation of the chimeric molecule in the cytoplasm which resulted in an inhibition of activation via the T-cell receptor. In a sub-population of transfected cells selected for high membrane localisation of the CD8-nef chimera, spontaneous activation and apoptosis was seen.

The third function of Nef is as an enhancer of HIV-1 infectivity which is independent of CD4 down-regulation and appears to operate during viral particle formation. However the effect is functionally effective after the progeny virus enters a new target cell host (Miller et al. 1994; Spina et al. 1994) and one study has reported increased efficiency of reverse transcription although this has not been confirmed (Schwartz et al. 1995). Finally, it has been reported that, *in vitro*, Nef protects infected primary T-lymphocytes against killing by allogeneic cytotoxic T-lymphocytes (CTLs) from HIV-1 positive individuals (Collins et al. 1998). The same group demonstrated a

reduction of surface MHC class I expression in Nef-expressing cells, an effect of Nef that had previously been reported (Kerkau et al. 1989; Scheppler et al. 1989), and attributed to increased endocytosis of MHC molecules by Nef-expressing lymphocytes (Schwartz et al. 1996) and has also been demonstrated to occur with the viral proteins Tat and Vpu (Howcroft et al. 1993; Kerkau 1997) but by different mechanisms. It is possible that HIV has evolved several separate mechanisms to avoid CTL killing and hence prolonging the survival of infected T cells.

Nef must operate by interaction with cellular proteins, either in the cytosol or, more likely, with membrane associated proteins. Nef has been immunoprecipitated with a serine kinase (Sawai et al. 1994) although the normal substrates and function of this kinase is not known. In addition, a conserved proline-rich (PXXP) repeat in nef binds specifically with the src homology region 3 (SH3) domains of Hck and Lyn, cell src kinases, a property that is required for the enhanced infectivity effect of Nef but not related to CD4 down-regulation (Saksela, Cheng, and Baltimore 1995). Thus the pathogenicity of Nef demonstrated *in vivo* may be the net result of several different functions of this intriguing viral protein. As Nef is present in all lentiviruses found in primates it is interesting to speculate why the pathogenic properties are only manifest if cross-species infection occurs, as in the case of SIVmac, the natural host for which is sooty mangabey (Chen et al. 1996), and HIV-1 and -2 which have recently infected humans (section 1.1) from primate origins.

1.2.7.2 vpu.

(Viral protein U) is unique to HIV-1 and the closely related monkey lentivirus SIV$_{CPZ}$. It is a 16 kDa phosphoprotein appearing late in the infection of a cell and is primarily associated with the internal membranes. Vpu has two functions, to reduce CD4 expression by increased degradation in the endoplasmic reticulum (Willey et al. 1992a and 1992b) and to enhance the release of progeny viral particles from the cytoplasmic membrane (Klimkait et al. 1990). The first effect is of functional importance to the virus as CD4 binds with nascent envelope precursor molecules in the ER inhibiting their transposition to the cell surface and potentially reducing the number of progeny viral particles. Vpu in the phosphorylated form interacts specifically with sequences in the cytoplasmic domain of CD4 triggering accelerated proteolytic degradation (Bour et al 1995) thus allowing processing and transport of env protein. The precise mechanism of this effect is unknown but as Vpu can form into oligomers, by binding to CD4 in large complexes the ER degradation pathways are invoked. The second function of Vpu, to increase the

release of progeny virions, can be separated functionally and genetically from the effect on CD4. The absence of Vpu resulted in the accumulation of virions with multiple cores and viral budding into vacuolar compartments rather than from the cell surface (Klimkait et al. 1990). This effect may not be specific for HIV-1 as Vpu also stimulated the budding of HIV-2, visna and murine leukaemia virus, retroviruses that normally lack vpu (Gottlinger et al. 1993) possibly indicating a function related to gag processing and transport.

1.2.7.3   Vif.

Vif (virion infectivity factor) is a 23 kDa protein produced as a late gene product from a single ORF, *vif* (previously called *sor*). Initial functional studies on Vif indicated a role in producing fully infective progeny viral particles, Vif-negative mutants were able to efficiently produce and release progeny virus but these were as much as 1,000 fold less efficient in establishing a new infection (Fisher et al. 1987; Strebel et al. 1987). Cell-to-cell transmission of virus was only slightly impaired. Vif functions during the assembly of new viral particles, but only trace quantities of the protein are present in mature virions (Sova and Volsky 1993; von Schwedler et al. 1993). Vif-mutant HIV produce progeny virions with abnormally packaged nucleoprotein (Höglund et al. 1994) and it appears likely that Vif is important in the processing and packaging of the RNA-protein complex in a form that allows the events that occur following the entry of the virus into the target cell, i.e. uncoating, reverse transcription, transfer to the nucleus, to proceed efficiently. The phenotype determined by Vif is also dependent upon the producer cell, thus *vif*-mutated HIV-1 is only functionally impaired when released from primary peripheral blood lymphocytes, macrophages and some T-cell lines such as H9, when other transformed cell lines are used as producer cells the effect of vif-mutation is usually not apparent.

Vif appears to show considerable diversity between isolates, sequence comparison of *vif* from HIV-1 isolates in Uganda showed that only 52% of the derived amino-acids were conserved, however there was good conservation of sequences when different isolates from the same individual were compared (Wieland et al. 1997). Another study showed a high frequency of *vif*-defective isolates, up to 31%, which did not correlate with disease stage (Tominaga et al. 1997). These results suggest that vif not only appears to display considerable variation but Vif-expressing and Vif-negative viruses may exist together *in vivo* possibly reflecting the *in vitro* observation of permissive and non-permissive cells.

1.2.7.4 Vpr.

Vpr (viral protein R) is a 14 kDa basic protein that is present in HIV-1, HIV-2 and most SIV isolates. The protein is coded by a single ORF lying between *pol* and *env* and shares some sequences with *vif*. The gene product is a late protein, accumulating in the nucleus of infected cells and is also found in mature viral particles where it is associated with the C-terminus of gag (Lu et al 1993).

Vpr displays two distinct functions; in non-dividing target cells, such as macrophages, Vpr facilitates the translocation of the newly reverse-transcribed proviral pre-integration complex into the nucleus. Secondly, in proliferating T-cells vpr induces cell cycle arrest at the G2/M phase of the cell cycle, an activity associated with an activation of transcription.

Vpr-expressing HIV-1 were shown to grow faster in culture and produce higher titres of virus than Vpr-negative infectious clones (Cohen et al. 1990; Ogawa et al. 1989). Vpr works in concert with the viral matrix protein (MA) to ensure the delivery of the preintegration complex to the nuclear import pathways. The mechanism by which Vpr localises the pre-integration complex to the nucleus is unclear but it is distinct from the karyopherin pathway used by MA. This function of the vpr-MA combination is of particular importance in non-dividing cells such as monocytes and dendritic cells (Heinzinger et al. 1994) where the nuclear envelope remains intact in the absence of mitosis. HIV-1 strains with mutations in vpr and MA are unable to infect macrophages or growth arrested cells due to a block in nuclear transport.

Vpr also appears to have an effect upon the cell cycle and LTR-driven transcription. Expression of Vpr in rhabdomyosarcoma and osteosarcoma cell lines induces growth arrest and differentiation (Levy et al. 1993), an effect that was repeated in a wide range of tumour cell lines of different origins (Mahalingam et al. 1997). The expression of Vpr in several different human tumour cell lines resulted in an accumulation of cells at G2/M phase associated with altered cell morphology. Vpr also increased the permissiveness of cells to HIV replication, even if the protein was added exogenously to cells, and to reactivate latent virus (Levy et al 1995). These properties may have significance *in vivo* by extending the range of cells able to sustain a productive infection.

The mechanisms by which Vpr influences the cell cycle and activates transcription have not been fully determined but appear to be linked functions. Vpr has been shown to directly interact directly with several important cellular proteins including those involved in the control of the cell

cycle, such as cyclin B1.Cdc2 and HHR23A (human homologue of the yeast DNA-repair protein RAD23), the interaction with cyclin B1.Cdc2 has been shown to also potentiate the activity of p300, a key transcriptional co-activator (Felzien et al. 1998; Gragerov et al. 1998). Vpr has also interacts directly with Sp1 and TFIIB, enhancing transcription from minimal promoter constructs (Agostini et al. 1998; Wang et al. 1998), but as Vpr lacks a transactivating domain it is not clear how these direct interactions can influence transcription. The G2/M phase arrest and transcription activation properties of vpr are functionally determined by the same regions of the protein while the nuclear localisation function is distinct (Subbramanian et al. 1998). Thus Vpr demonstrates again the functional economy of HIV 1, using not only the same DNA sequences to encode several proteins (Vif, Vpr and Tat) but also the derived protein has multiple functions.

1.2.7.5   Tat.

The *tat* gene consists of two exons, the first precedes *env* the second within *env* (fig 1.2.2), the protein is translated from multiply spliced mRNAs to produce two forms, predominantly an 86 amino-acid 16kDa protein from a 2kb mRNA and a lesser quantity of a 72 amino acid 14kDa protein translated from a singly spliced 4 kb mRNA and consisting of the first exon alone. Tat appears early following infection. Functional Tat protein is absolutely required for viral replication, Tat-deficient mutants are unable to replicate unless Tat is supplied in *trans* either via an expression vector or as exogenous protein (Dayton et al. 1986; Fisher et al. 1986). Tat functions as a powerful transactivator of LTR-directed gene expression, able to act in the context of the integrated provirus or when the LTR is driving a reporter gene, expression increasing by 40-3,000 fold depending upon the experimental conditions (Arya et al. 1985; Cullen 1986; Sodroski et al. 1985). Tat exerts its effect on LTR-driven expression by interacting with the TAR region in the 5'-end of nascent mRNA. Initial activation of viral mRNA expression occurs by the interaction of cellular factors with elements in the enhancer and the core promoter in the LTR, short multiply spliced mRNAs are exported and translated producing tat protein. This translocates to the nucleus and specifically enhances further viral expression; thus although the viral LTR is dependent upon cellular transcription apparatus and must initially compete with cellular genes, the combination of tat and TAR favours expression of the viral genes. The mechanisms by which tat and TAR influence transcription are detailed in section 1.5.

Three functional domains have been identified in Tat (fig 1.2.3),

Fig. 1.2.3   Functional domains of Tat

2	11	22	37	48	57	▼	86
	acidic		cysteine rich		basic	GLN rich	

CxxCxCxxCCxxCxxC   GRKKRRQRRR
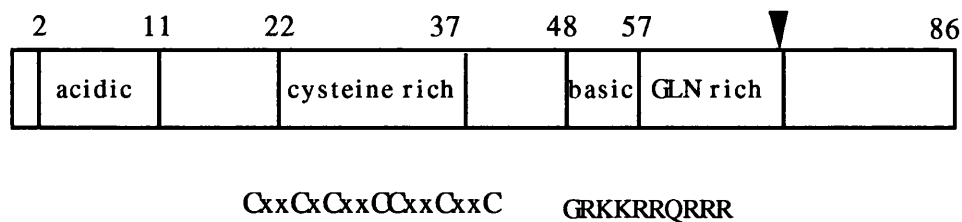
Fig. 1.2.3. Diagrammatic representation of Tat protein showing the main functional domains. The amino acid sequences of the cysteine-rich and basic domains are detailed. The boundary between exons 1 and 2 is indicated by the arrow.

The acidic amino-terminal domain (residues 2-11) contains an antipathic α-helix which can function as an autonomous transcription-activation domain when fused to a heterologous DNA-binding protein analogous to the activation domains of many well characterised transcription factors (Ptashne 1988; Ptashne and Gann 1997). Mutation of this domain dramatically reduces Tat function. Residues 22 to 37 contains 7 conserved cysteine residues, mutation of these (except C 31) destroys Tat function (Ruben et al. 1989; Sadaie et al. 1988), this domain may be involved in dimerisation of tat together with zinc ions, but dimer formation may not be important functionally.

Residues 48 to 57 is an (Arg+Lys)-rich basic domain essential for binding to TAR with the arginine at position 52 making direct contact. This region also contains the nuclear localisation signal GRKKR (residues 48-52). The residues 72-86 are encoded by the second exon and are not required for transactivation, it does however contain a conserved RGD motif that mediates binding to integrins allowing exogenous tat to be taken up by cells, a characteristic of Tat that can be demonstrated *in vitro* but the importance *in vivo* is unclear.

1.2.7.6   Rev.

The second important non-structural protein of HIV-1 is Rev, a 19 kDa protein of 116 amino acids which is predominantly localised in the nucleus of infected cells (Cullen et al. 1988). Rev is expressed from two exons which overlap with those coding for Tat, suggesting that these two RNA-binding proteins have co-evolved. The function of Rev is to regulate the export of viral mRNA from the nucleus acting as the effector of post-transcriptional gene expression. *Rev* deficient mutants of HIV-1 (*rev* $^-$) are unable to produce progeny virus; high levels of spliced RNA species are found in the cytoplasm of cells infected with

rev ¯ provirus, but only low levels of singly spliced mRNAs coding for *env,*
vpu, vif and *vpr* are present and virtually no full-length unspliced mRNA
coding for the structural proteins *gag* and *pol* and progeny genomic RNA
(Feinberg et al. 1986; Malim et al. 1989 and 1990; Cullen 1994). Full length (9kb)
and partially spliced (4kb) viral mRNAs contain a 234-nucleotide sequence
within *env* coding region that is the specific target for Rev, this is the Rev-
response element (RRE). The sequence of the RRE allows Watson-Crick pairing
forming a complex helical-loop structure to part of which Rev specifically
binds. A minimal 90 nucleotide element from the 5'-end of the RRE is sufficient
to allow Rev binding but alone does not confer Rev functionality *in vivo.* Rev
binding to the RRE is dependent upon both the primary sequence and
secondary structure (Dayton et al. 1992; Holland et al. 1992).

Fig 1.2.4   Functional domains of rev.
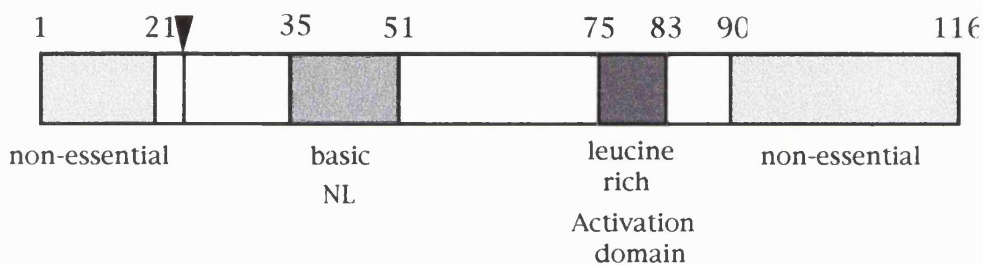
Diagrammatic representation of Rev showing domains.
NL = nuclear localisation signal;

The region of rev from amino-acids 35-50 is highly basic, containing 10
arginine residues and is required for RNA binding and nuclear localisation,
mutations within this region prevents rev accumulating in the nucleus and
abolishes RNA-binding (Malim and Cullen 1991). Flanking the basic domain
are several amino-acid residues required for multimerisation of rev mutation
of which result in the loss of function (Malim and Cullen 1991). A second
important domain lies towards the C-terminus between amino acids 75-83 and
contains three closely spaced leucine residues, mutation of any one of these
results in loss of rev function but nuclear localisation and RRE binding are not
affected. The binding of Rev to the RRE is complex; the initial interaction is
with a high affinity site on stem-loop IIB, this site is characterised by non-
Watson-Crick base pairs in a stem-bulge-stem structure (Bartel et al. 1991).
Following the initial interaction co-operative binding of a further 6-8 rev
monomers occurs onto the RRE structure. The leucine-rich activation domain
of Rev specifically interacts with nucleoporin-like proteins which mediate the
export of RRE containing RNA species from the nucleus, exploiting transport
pathways used for 5SRNA and snRNAs (Fischer et al. 1995).

1.2.8   Life cycle of HIV-1

1.2.8.1   Cell tropism: CD4 as receptor for HIV-1 and 2.
Early in the study of HIV-1 and the effect of the virus upon infected
individuals a consistent finding was the selective depletion of CD4 positive
peripheral blood T-lymphocytes indicating tropism for these cells. Cultured
CD4 +ve cells specifically supported the replication of HIV-1 (Klatzmann et al.
1984a). Further evidence for the CD4 molecule as receptor for HIV-1 came from
experiments using monoclonal antibodies against specific epitopes of CD4
which blocked viral cytopathic effect and syncytia formation (Dalgleish et al.
1984; Klatzmann et al. 1984b). Binding of gp120 to CD4 was demonstrated by
coprecipitation using antibodies directed against either of the proteins.
Conclusive evidence was the conversion of CD4 -ve HeLa cells, which are not
normally infected by HIV-1, into a permissive cell by the transfection of a CD4
expression vector. The CD4 receptor is expressed on T-helper lymphocytes and
monocyte-macrophages making these cells the most obviously affected cell
targets, it is also present on antigen presenting cells (follicular dendritic cells,
Langerhans cells), microglial cells, haemopoietic cells such as mega-
karyocytes and bone marrow stromal cells, these cells are also infected in vivo
and accounts for the wide range of pathogenic effects of HIV upon infected
individuals.
However, expression of human CD4 alone did not seem to be sufficient to
confer infectivity, when transfected into murine cells they remained resistant
to infection and although virus could bind to the cells fusion and entry did not
follow. This provided tantalising evidence for the requirement of a second
component, present constitutively on human cells, which was required for
full infectivity (Maddon et al. 1986). Recently, two other cellular receptors
have been shown to be important in the entry of virus into target cells. The
discovery of these co-receptors also provided insight to the observed cell
tropism of different HIV isolates.

1.2.8.2   T-lymphocyte vs. macrophage tropism of HIV.
The majority of laboratory strains of HIV-1, which have been passaged in T-
cell lines, easily infect such cells and can also infect primary T-cells but not
primary monocytes or macrophages. Some primary isolates of virus i.e. not
passaged in culture, also show a preference for T-cells rather than monocytes,
this seems to be a feature particularly when virus has been taken from
patients late in their disease. These strains of virus are termed T-cell tropic (T-
tropic). Conversely, the majority of primary viral isolates appear to infect T-
cell lines with some difficulty but infect and replicate well in primary T-cells

and monocytes/macrophages. These strains are called macrophage tropic (M-tropic). M-tropic strains are present early in the progression of infection and persist throughout the asymptomatic period of the disease. The viral determinant of cell tropism maps to the V3 loop of gp120 (Choe et al. 1996) and recently the cellular determinant has been shown to involve the type of co-receptor expressed. The first co-receptor identified was fusin (Feng et al. 1996), also termed LESTR (leucocyte-expressed seven-transmembrane-domain receptor), is the receptor for the stromal cell-derived chemokine SDF-1 (Bleul et al. 1996; Oberlin et al. 1996). Fusin is widely expressed on peripheral blood monocytes, neutrophils and PHA-stimulated lymphocytes as well as on Jurkat cells, HeLa cells and H9 cells. The co-expression of fusin with CD4 rendered cells susceptible to infection by T-tropic strains of HIV-1 but not by macrophage-tropic strains. The interaction of fusin with the natural ligand SDF-1 blocks infection by HIV-1. The co-receptor for macrophage-tropic strains of HIV-1 has been identified as CCR-5 which is also a member of the seven-transmembrane, G-protein-coupled chemokine receptor family. The natural ligands for CCR-5 are the chemokines macrophage inflammatory protein-1α and 1β (MIP-1α, MIP-1β) and RANTES (regulated on activation normal T-cell expressed and secreted) and as was found for fusin and SDF-1 the presence of the natural ligands for CCR-5 prevents HIV-1 infection (Alkhatib et al. 1996; Dragic et al. 1996). For some time it has been known that the V3 loop of gp120 determined cell tropism, it appears that the same region of gp120 determines chemokine receptor usage (Choe et al. 1996). In addition, the CC-chemokine receptor family has several members and although CCR-5 is used by all M-tropic HIV strains some are able to utilise other CCRs. Presumably there is some advantage to HIV in having this elaborate method of gaining access to target cells, perhaps the requirement for a chemokine receptor ensures that a viral particle bound to CD4 only undergoes fusion with the cell membrane if the cell is a suitable host able to sustain a productive infection, the presence of a chemokine receptor may act as a marker of activation or impending proliferation in the case of T-lymphocytes. The observation that the CMV gene product US28 mimics the β-chemokine receptor and allows uptake of HIV in cells infected by CMV, potentially increasing the range of cell types infected by HIV *in vivo* (Pleskoff et al. 1997).

1.2.8.3   Binding, fusion and entry of virus.

The primary cellular receptor for HIV-1 is the CD4 membrane protein to which the viral gp120 (env) binds with an affinity constant of 10^{-9}M. A conserved region at the carboxy-terminus of gp120 is involved in the binding with CD4. Binding of env to CD4 induces conformational changes in env, however these

changes alone are not sufficient to permit virus-cell fusion, for this to occur a co-receptor must be present. The exact function of the co-receptor has not been determined but it does not appear to be important in the initial binding process, nor in the conformational changes induced in gp120 following binding but without a co-receptor, fusion of viral envelope and cell membrane does not occur and the viral capsid is unable to enter into the cytoplasm. It has been suggested that the chemokine co-receptor molecule, either by direct interaction (V3 loop) or indirect influence, induces further conformational changes in env thus exposing the fusogenic domain of gp41. Fusion occurs in a pH-independent manner and does not involve receptor-mediated endocytosis, electron microscopy revealed direct fusion of the viral envelope with the cytoplasmic membrane (Stein et al. 1987).

1.2.8.4   Reverse transcription.

The process of proviral DNA synthesis from two single strands of viral RNA proceeds in a similar manner in HIV as in other retroviruses (reviewed by Coffin 1991). Following entry of the HIV core into the cytoplasm of a susceptible cell reverse transcription of the viral RNA becomes activated and using the structure, local environment and enzymatic activities contained in the viral core to produce linear, proviral dsDNA. Evidence from other retroviruses suggests that the process requires components of the capsid rather than occurring in free solution within the cytoplasm. The enzymes present are the RNA-dependent DNA polymerase (RT), which also functions as ribonuclease (RNase H) (Arti et al 1996) and an integrase, the endonuclease function of which allows the insertion of the proviral DNA into the host cell DNA. All are derived from the viral *pol* gene. Fig 1.2.5 illustrates the important stages of reverse transcription. The viral genomic RNA is shown diagrammatically in fig 1.2.5a with the short repeat sequences (R) at both 5' and 3' extremes and the tRNA molecule attached to the primer binding site (PB). RT uses the 5' end of tRNA primer to initiate DNA synthesis and creates a DNA strand complementary to U5 and R while RNase H activity of RT removes the RNA sequences for 5' R and U5. Now the first of two transfers occurs moving the point of DNA synthesis from the 5' R site to the 3' R site (fig 1.2.5b), this allows further elongation of DNA synthesis (fig 1.2.5c). At this point a heteroduplex of RNA and DNA exists, in order to create the positive strand of DNA the RNA strand is degraded by RNaseH activity of the pol enzyme approximately 20 nucleotides behind the newly synthesised DNA. A specific polypurine tract present at the 5'-end of U3 is resistant to RNaseH attack and this residual RNA serves as the primer for second strand synthesis from 5' to 3' running up to and over the tRNA primer. The second transfer occurs, possibly

involving a circularised intermediate form held by complimentary sequences in the PB sites (fig 1.2.5e). Synthesis of the positive strand continues and the tRNA molecule is removed by RNase H. The resulting linear dsDNA molecule has identical 5' and 3' sequences, U3-R-U5 which form the long terminal repeats or LTRs flanking the viral genes which are in the same order as in the viral genome.

---◆---

Legend for figure 1.2.5:

Reverse transcription from viral genomic RNA to proviral dsDNA (from Coffin, 1991).

a.
Initiation of 1st strand synthesis from 5' end of tRNA bound to primer binding site (PB). RNase H activity of reverse transcriptase removes the 5' R and U5 sequences.

b.
Transfer of 1st strand replication from 5' R site to the 3' R. Further elongation of the 1st strand follows with formation of an intermediary RNA/DNA heteroduplex.

c.
RNA/DNA heteroduplex is shown but a full heteroduplex may never be formed as the RNA strand is degraded approximately 20 base pairs behind the newly synthesised DNA.

d. and e.
The polypurine tract in the 3' U3 resists RNase H degradation and serves as the primer for second strand synthesis running up to and over the tRNA primer. Second transfer occurs with the formation of a circular intermediate held by base-pairing between complimentary PB sites. Synthesis of the positive DNA strand continues and the tRNA molecule is removed by RNase H activity.

f.
Final linear dsDNA provirus. The transition from ssRNA to dsDNA results in the formation of identical 5' and 3' sequences in U3-R-U5 forming the 5' and 3' long terminal repeats.

Figure 1.2.5

Reverse transcription from viral genomic RNA to proviral dsDNA.

a.

R   U5   PB   gag   pol   env   U3   R

R'   U5'

b.

PB   gag   pol   env   U3   R

R'

c.

PB   gag   pol   env   U3   R

PB'   gag'   pol'   env'   U3'   R'

d.

U3   R   U5

PB'   gag'   pol'   env'   U3'   R'   U5'

e.

U3   R   U5   PB

U3'   R'   U5'

f.

U3   R   U5   PB   gag   pol   env   U3   R   U5

5' LTR   3' LTR
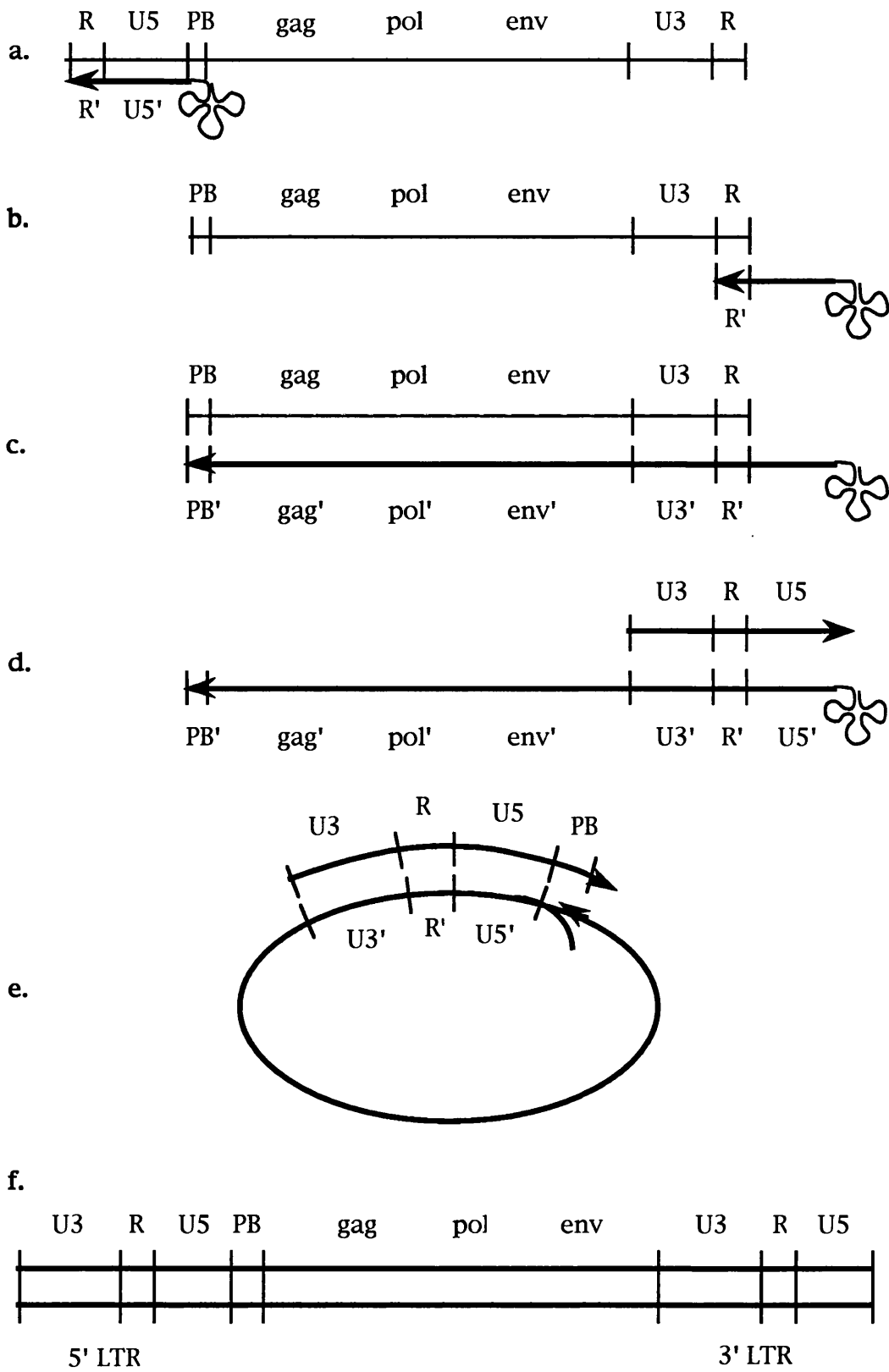
Retroviruses in general are unable to replicate in non-dividing cells although the precise site of the block is unclear and may differ for different viruses and target cells. For HIV the potential target cells may be relatively quiescent, e.g. nonactivated T-cell, non-dividing monocytes, or activated cells. Infection of quiescent T-cells is nonproductive and only incomplete proviral sequences can be identified (Zack et al. 1990) which have a half life of about one day. Other workers have identified transcriptionally active proviral DNA that exists as a stable episomal element (Stevenson et al. 1990) which is able to integrate into the cell genome upon activation of the T-cell. The situation appears to differ in cultured nondividing monocytes that are readily infected by M-tropic strains of HIV with integration of proviral DNA and release of progeny virus. Either the state of 'quiescence' is relative or HIV has evolved different strategies to cope with the different cell types that it may infect and replicate successfully.

1.2.8.5   Integration.

The viral nucleoprotein complex is activly transported to the nucleus. This complex consists of MA, p9, vpr and incompletely reverse transcribed viral RNA. Full formation of the dsDNA provirus is completed in the nucleus, it then becomes integrated into the host genomic DNA by a unique process catalysed by the viral integrase. The integration site is random but tends to occur at regions of DNA that are transcriptionally active. Possibly the open chromatin structure in such regions is more readily approached by the proviral structure. The process of integration by HIV appears to be similar to other retroviruses, indeed highly conserved dinucleotide sequence 5'TG----CA3' are present at each end of the genome, a feature of all retroviruses. Integration may require a degree of cellular 'activation' as differences in the efficiency of integration has been demonstrated between quiescent versus activated T-cells.

1.2.8.6   Expression of viral genes.

The integrated retroviral genome is essentially a genetic 'cassette' carrying all the necessary genetic information to direct the production of progeny virus. The majority of the viral genes encode the structural proteins and enzymes required in the completed viral particles. The function of the integrated viral genome is to produce mRNA to be translated into these proteins and also produce full length mRNA for packaging in the progeny viral particles as genomic RNA. Expression of the proviral mRNA is dependent upon the cellular transcriptional machinery contained in the RNA pol II complex, cellular trans-acting transcription factors and cis-acting virally encoded factors (section 1.4).

1.2.8.7   Patterns of mRNA expression: role of Tat and Rev.

The HIV-1 genome is transcribed by cellular RNA pol II transcriptional machinery but depends upon the two virally determined factors Rev and Tat for efficient and regulated expression. These proteins function through interactions with *cis* elements present in the transcribed mRNA, the Rev responsive element (RRE) and the transactivation response element (TAR). Several splice donor and acceptor sites have been identified providing the potential for over 20 differentially spliced mRNAs (Schwartz et al. 1990), however, three major sizes of mRNA are produced, a multiply spliced 2kb transcript, a partially spliced 4kb transcript and a full genome length 9kb transcript (fig 1.2.6). These appear in a time sequential manner (Kim et al. 1989b) analogous to the regulated expression of many eukaryotic viruses that produce immediate early, early and late mRNA species co-ordinating the requirements of the virus for proteins to control viral replication in the early stages of replication and structural proteins and viral genomes in the later stages ready to assemble into progeny virion.

Figure 1.2.6 HIV-1 mRNA splicing and gene products.

HIV-1 proviral genome
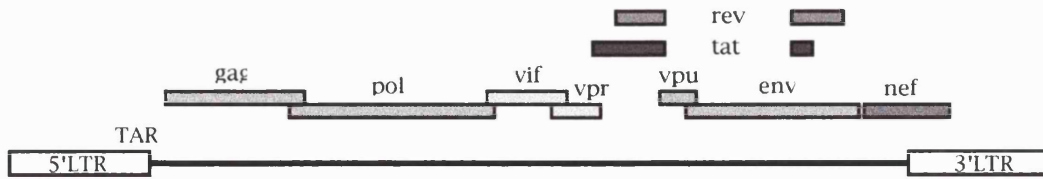
Major mRNA transcripts
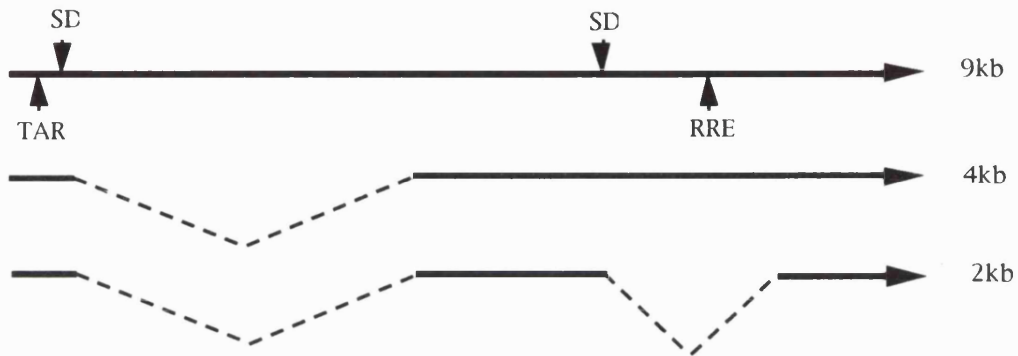
Proteins translated from major transcripts (in sequence of appearance

2kb mRNA ⟶ Tat (p16), Rev, Nef

4kb mRNA ⟶ Env, Vif, Vpu, Vpr, Tat (p14)

9kb mRNA ⟶ Gag and Pol
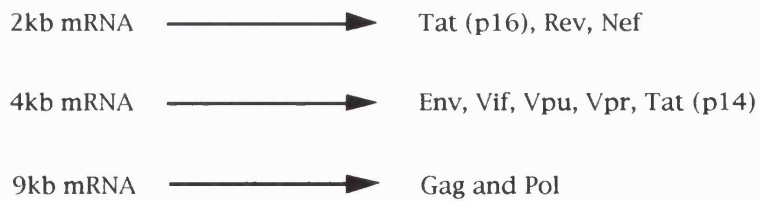
Fig. 1.2.6: The HIV-1 proviral genome and major mRNA transcripts. Three major classes of mRNA are produced, the relative proportions depend upon splicing events and the presence of Rev. Fixed splice donor sites (SD) are indicated, several splice acceptor sites exist (not shown).

The earliest mRNAs to appear in the cytoplasm following integration of the provirus are fully spliced 2kb mRNA producing high levels of the viral proteins Nef, Tat and Rev. All three proteins have important functions in the early phase of infection, that of Nef is unclear but may involve modulation of the immune response against the virus and is detailed in section 1.2.7.1. Tat acts as a powerful transactivator of viral LTR-driven expression increasing the production of viral mRNA by 10-100-fold (section 1.5). Rev is required for the switch from multiply spliced transcripts present in the early phase of viral expression to the 4kb and 9kb transcripts that mark the later phases leading to the production of structural proteins and full length genomic RNA. A critical threshold of Rev concentration in the nucleus is required for the transition to partially spliced and unspliced mRNA species (Pomerantz et al 1992) a process that involves the interaction of Rev with the RRE which allows the transport of these unprocessed mRNAs from the nucleus to the translational apparatus in the cytoplasm. The incompletely spliced 4kb transcripts are translated into the structural protein Env, a smaller but functional Tat protein of 14kDa (p14) and the 'accessory' proteins Vif, Vpu and Vpr (section 1.2.7). Rev also allows the survival and transfer of full-length 9kb mRNA to the cytoplasm, this mRNA is multifunctional, translated into the p55gag precursor protein and p160gag-pol which provides all of the structural proteins required for assembly of viral capsid and nucleoprotein core. The Pol protein is translated from the same 9kb transcript by a ribosomal frame shift (Jacks et al. 1988), the frame shift is inefficient with 20-fold more Gag-precursor produced than Pol but this suits the requirement for the production of progeny viral particles.

1.2.8.7   Assembly of viral particle.

The assembly of progeny virion occurs in a co-ordinated fashion and requires the appearance of structural proteins, virally determined enzyme activity and full genome length unspliced mRNA ready to be packaged. The assembly and budding of HIV-1 and other lentiviruses is similar to that of the C-type retroviruses, no complete intracellular forms are seen (unless budding into intracellular vesicles or cisternae), the first forms visible by electron microscopy are crescent-shaped patches at the sites of budding. The process of viral assembly can be divided into 5 stages:

1)   interaction of genomic RNA with gag

2)   translocation of nucleoprotein complex to cell membrane and association with env

3)   full assembly of nucleocapsid core components

4)   maturation of viral particle by processing of proteins contained in the

nucleoprotein core

5)   budding of particles

These events do not occur in a strictly sequential fashion but over-lap, assembly and budding occurring almost simultaneously (Coffin 1991). The polyprotein precursors, products of the *gag* and *pol* genes, accumulate in the cytoplasm as p55gag; NH2-p17(MA)-p24(CA)-p9(NC)-p7-COOH and p160 gag-pol; NH2-p17(MA)-p24(CA)-p9(NC)-p7-p10(PR)-p66(RT)-p32(IN)-COOH.

The interaction of gag and the genomic RNA is specific depending upon a packaging signal residing in the leader region between the major splice donor in U5 and the gag gene initiation codon, deletion of a 19-base sequence in this region results in a 98% reduction of viral RNA packaging (Lever et al. 1989). As these sequences are downstream of the splice donor site it will be absent in all sub-genomic spliced mRNAs explaining the preferential incorporation of unspliced RNA in the assembling viral particle. The p9 NC domain of gag binds to the packaging sequence probably before cleavage by the protease, using a conserved zinc-finger like motif deletion of which reduces virion associated RNA (Gorelick et al. 1990). How the virus ensures that only two RNA molecules are incorporated per viral particle is not known. The assembly process is mediated by specific functional properties of the components. Myristilation of the NH2-terminal (MA) of p55gag and p160gag-pol targets these proteins to the plasma membrane, the genomic RNA is brought in by the interaction of p9 NC with the packaging sequence. The nucleocapsid assembly continues at the membrane with the auto-assembly of p24 MA sub-units to form an enclosing tube-shaped core. No function has been determined for p7. The gag and gag-pol precursors are cleaved during the process of assembly and further maturation of the core occurs after budding.

The envelope glycoproteins, synthesised as a precursor gp160 molecule, are incorporated into viral particles by a separate pathway. The nascent protein undergoes extensive glycosylation in the ER, the precursor is cleaved into gp120 and gp41 either before or after insertion into the cell membrane (McCune et al 1988) . The targeting of env product to the membrane is probably mediated by gag proteins, the expression of gag alone using baculo-virus or vaccinia based expression systems resulted in the formation and release of virus-like particles (Gheysen et al. 1989). These results also indicate that properties of the gag proteins direct and mediate the process of budding from the plasma membrane. The viral accessory proteins Vpu and Vif also appear to be important in the production of fully infectious viral particles, Vpu appears to enhance budding and Vif may be involved in the maturation of the nucleoprotein core (sections 1.2.7.4 and 1.2.7.5).

1.3 Pathogenesis of HIV-1.

1.3.1 Introduction.

HIV-1 is similar to the majority of lentiviruses in causing a persistent and, at this moment in time, incurable infection. However, a particular characteristic of HIV-1, shared with some SIVs and to a lesser extent HIV-2, is a marked pathogenicity for the host. HIV-1 infection of humans was made apparent by the appearance of a new disease entity characterised by unusual infections associated with a marked impairment of cell-mediated immune response (Gottlieb et al. 1981). The discovery of HIV-1 in these patients led to the not unexpected hypothesis that the virus was responsible for the destruction of CD4-positive cells. The problem with this hypothesis was that in patients infected with HIV-1 few of the circulating CD4-positive cells were expressing viral RNA, only 1/100,000 to 1/10,000 of total mononuclear cells (Biberfield et al. 1986; Harper et al. 1986), and low frequencies of HIV-1 infected CD4+ cells could be detected (in the order of 1/50,000) until the late stages of the disease when AIDS had developed and >1/100 CD4+ carried integrated proviral genome (Ho et al 1989; Schnittman et al. 1989). Although *in vitro* a direct cytopathic effect of the virus could be demonstrated there was no evidence for such an effect operating *in vivo*.

The dichotomy between the observed effect of the virus clinically and *in vitro* generated a number of theories to explain the loss of CD4-cells, for example apoptotic cell death initiated by gp120/CD4 interaction, autoimmune destruction of uninfected CD4+ cells, syncytia formation killing uninfected cells, immune-complexes inducing complement activation, the cytotoxic effect of viral proteins such as Tat, the presence of a viral super-antigen causing Vβ specific clonal deletion (Imberti et al. 1991).

A clearer understanding of the relationship between virus and host became apparent over the period 1991-1995 with the discovery of three vital pieces of evidence. Firstly, the realisation that although there were few infected CD4-positive cells in the blood, the lymphoid organs (the lymph nodes, spleen, gut-associated lymphoid tissue) all had high levels of virally infected cells from early in the infection. This was first reported by Racz in 1986 but at the time the relevance was ignored until a report from Pantaleo in 1991 (Pantaleo et al. 1991) and the virtually simultaneous reports by three groups in 1993 (Embretson et al. 1993; Pantaleo et al. 1993; Piatak et al. 1993). Not only are the lymphoid organs an important site of infection, the distribution of infection at a cellular level has been reassessed. The focus on CD4+ lymphocytes as the target for infection has shifted to other CD4+ cell types, in particular the

monocyte/macrophage lineage, dendritic cells and microglial cells in the CNS, cells in which HIV-1 can establish chronic productive infection.

The second piece of evidence was obtained by the application of mathematical modelling to interpret the effects of potent new anti-viral drugs on the levels of virus and CD4 cells in HIV-1 infected patients (Ho et al. 1995; Wei et al. 1995), revealing a much higher rate of viral production and destruction in the chronic phase of the disease. These analyses were made possible by the development of more sensitive, accurate and reproducible methods to assay HIV-1 virion RNA in plasma, the 'titre' of virion RNA also providing an indirect measure of active viral replication in lymphoid tissue, the largest reservoir and most important site of viral replication.

The third important piece in the puzzle was the long awaited characterisation of the coreceptors required for viral fusion/entry to target cells, which also provided an explanation for the cell tropism displayed by different isolates of HIV-1 and the inhibitory effect of certain soluble factors released by CD8+ T-cells that inhibited viral replication in culture, the β-chemokines RANTES, MIP-1α, MIP-1β and SDF-1.
These three major discoveries profoundly altered the understanding of the interaction of HIV-1 with the host and had important implications for the treatment of patients.

The present view of HIV-1 infection is that of a highly active infective process that continues throughout the clinically asymptomatic period. An active humoral and cell mediated response rapidly removes both free viral particles and infected cells, with a balanced replacement of CD4 cells and progeny virus during the 'steady-state' period. A gradual exhaustion of the pool of uninfected CD4+ cells and involution of lymphoid tissue ultimately occurs with the progression to AIDS an inevitable consequence for the majority of patients. The following describes in detail some of the factors, viral and host, that are thought to play a part in the pathogenesis associated with HIV-1 infection (reviewed by Coffin 1996; Feinberg 1996; Haynes, Pantaleo, and Fauci 1996; Pantaleo and Fauci 1996).

1.3.2 HIV-1 infection at the cellular level.
As described in section 1.2 the initial phase of the life cycle of HIV-1 is broadly similar to that of any retrovirus; following binding of the viral particle to the CD4 receptor and the appropriate co-receptor of a susceptible target cell the virus enters the cytoplasm, proviral DNA is synthesised by reverse transcription and the formation of the preintegration complex completed.

There are four possible resulting types of infection,

1) abortive; 2) latent/reactivation; 3) acute; 4) persistent

The fate of the infection depends upon the type of cell infected and the activation state of that cell. HIV-1 can infect quiescent, proliferating and terminally differentiated CD4-positive cells present in a wide number of tissues, including the peripheral blood (CD4-positive T-lymphocytes, monocytes), lymph nodes and spleen (T-lymphocytes, monocytes/macrophages, follicular dendritic cells) and central nervous system (microglial cells and astrocytes). Amongst these cell types are naive, unprimed T-cells which are blocked in Go phase of the cell cycle and are essentially quiescent. *In vitro*, infection of these cells is abortive, halting soon after the entry of virus, either without completing reverse transcription (Zack et al. 1992) or after the generation of a proviral complex that is unable to enter the nucleus but can remain as a stable episomal element or which is lost (Chou et al 1997; Spina et al 1995). If the cell is activated before the arrested provirus is lost the entry of the cell into S phase and mitosis allows integration. If the necessary activation signals are removed or the accessory cytokines are absent the cell may become quiescent as a memory T-cell, harbouring an integrated HIV-1 proviral genome in a 'latent' form. Further exposure of the memory cell to specific antigen and/or cytokines will activate the cell to proliferate, in the process activating the HIV genome which utilises the same cellular signalling pathways (see section 1.5). If actively dividing T-cells are infected, the initial events occur within a few hours resulting in proviral integration and productive infection with expression of viral genes and release of progeny virions within 36 hours continuing until cell death. HIV-1 and other lentiviruses are unique in their ability to establish productive infection in terminally differentiated non-dividing cells such as monocytes and dendritic cells that are held at the G1/S or G2 phases of cell cycle a property possibly attributable to the presence of viral accessory proteins peculiar to lentiviruses such as Vpu and Vif (section 1.2.6).

1.3.3 The concept of latency: clinical and cellular.

It is important to differentiate clinical latency, the relatively asymptomatic period that follows initial infection, from latency at a cellular level, where the proviral genome may be silent or show a low level of basal expression. As discussed in previous sections, evidence shows that viral replication in the host is very active from the onset. Large numbers of viral particles are produced throughout the infection with, in the steady state, a balanced destruction by the host's immune response. Destruction of viral particles and virally infected cells is balanced by replacement with fresh, uninfected CD4-

positive cells. The following section describes the events *in vivo* that are thought to occur following infection of an individual with HIV-1. The description by necessity is generalised, on an individual scale the course of an infection will be influenced by a number of interacting factors, viral, cellular and host determined. For example the predominant subtype of virus present in the initial inoculum, the route of infection, the viral dose, the presence of other infective agents and intrinsic properties of the host, such as coreceptor phenotype, age (reviewed by Feinberg 1996; Pantaleo and Fauci 1996).

1.3.4 Clinical course of infection.

The clinical course of HIV infection is divided into four stages:

1)   primary infection.
2)   asymptomatic phase a period of clinical latency, with a median
        duration 10 years before progression to
3)   symptomatic phase
4)   late phase with onset of AIDS-defining illness.

These four stages are characteristic for the majority of individuals infected with HIV-1, termed 'typical-progressors' which constitute 70-80% of HIV-1 infected patients. In addition a further three patterns of clinical progression have been described based on retrospective studies of large groups of patients. In some individuals (10-15%) the so-called 'rapid progressors', the disease course is shortened, the peripheral CD4+ cell count falling to levels associated with a high-risk of AIDS within a few years of the initial infection. A third group of patients remain asymptomatic for a prolonged period, have slowly falling CD4 levels but appear to establish a prolonged equilibrium with the virus and although peripheral blood CD4+ counts are in the range of 200-500 cells per ml remain relatively well, these have been termed the 'long-term survivors'. The final group of patients identified, the 'long-term non-progressors', show no signs of disease and have relatively stable CD4 counts and low levels of virus. This group forms less than 5% of all HIV-1 infected individuals.

1.3.4.1   Primary infection: viral dynamics and host response.

The primary infection may be associated with a mononucleosis-like acute illness with malaise, fevers and lymphadenopathy which occurs within four weeks of infection. These may be ignored at the time of primary infection but it is estimated that 50-70% of HIV-1 infected individuals experience symptoms during the primary infection (Tindall and Cooper 1991). The primary infection is associated with a rapid rise in circulating virus with, at the peak of the acute viraemia, up to 10^7 HIV-1 RNA copies per ml of plasma and high levels of HIV-

1 infected CD4+ lymphocytes corresponding to the period of systemic dissemination of virus. Evidence for the dissemination and tissue targets for HIV-1 has been obtained from experimentally infecting monkeys with SIV and from studies in humans by cross-sectional analysis of virus distribution in lymph node biopsies taken soon after seroconversion. In the SIV model, large numbers of virally infected cells were detected in lymph nodes by day seven following inoculation, the peak of virus expression in the lymph nodes preceding, or was simultaneous with, the peak seen in the peripheral blood, suggesting that the lymph nodes are the major site of viral replication during this stage. A similar pattern of virus distribution was found in HIV-1 infected patients, with large numbers of infected cells in lymph nodes during the primary infection (Fauci 1993).

1.3.4.2   Events following acute infection.

Following the early dissemination of virus there is a brisk immunological response by the host, both humoral and cell mediated. This results in a fall in plasma viral titres and clearance of viral particles (fig 1.3.1). This is associated with marked changes in lymph nodes, follicular hyperplasia occurs and viral particles are trapped by follicular dendritic cells, possibly by the binding of virus-antibody complexes via Fc receptor interactions. Such bound virus may not be removed but could infect CD4+ lymphocytes that associate with the dendritic cells (Heath et al. 1995).
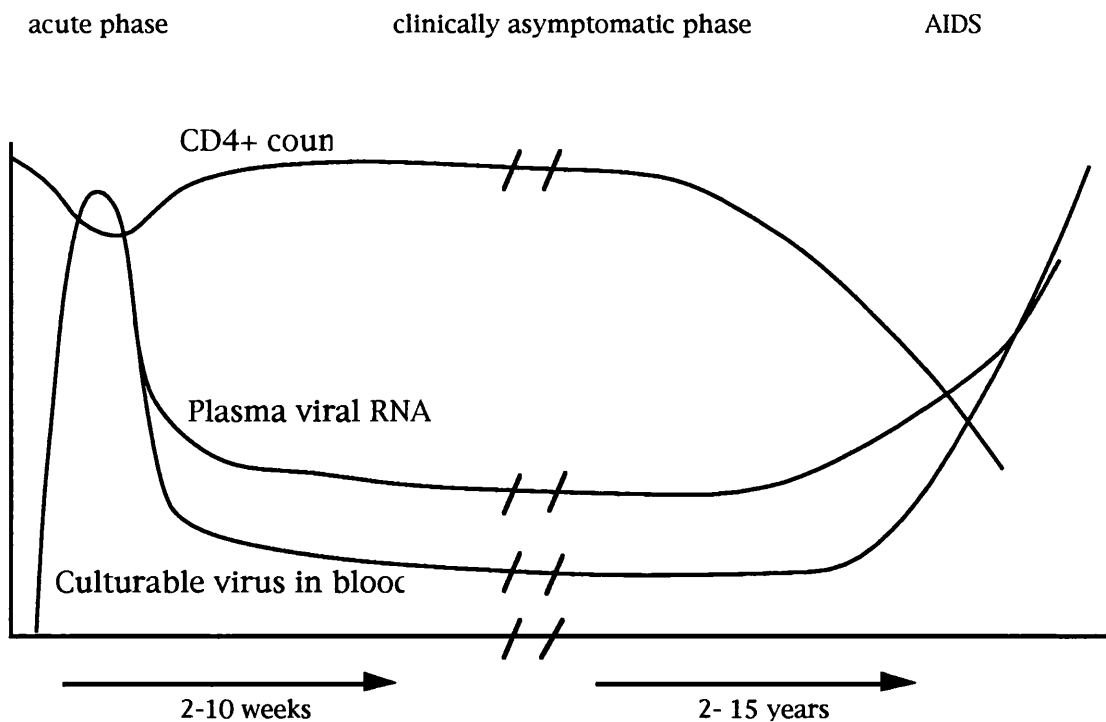
Figure 1.3.1   **Pattern of 'typical' HIV infection.**

The course of infection for the 'typical progressor' pattern of HIV-1 infection. For details of other patterns of infection see Pantaleo 1996.

1.3.5 Factors involved in viral persistence.

Despite the demonstrable and specific immunological response to viral antigens the virus avoids total elimination but instead establishes a stable, chronic infection as the disease enters the asymptomatic, steady state phase. The mechanisms by which HIV-1 is able to avoid complete elimination and establish chronic infection are poorly understood, a number of factors have been implicated: (from Pantaleo and Fauci review 1996).

viral factors:

- formation of a large pool of latently infected cells
- trapping of virus in the follicular dendritic cell network in lymphoid germinal centres
- changes in viral phenotype by rapid mutation and selection
- direct cell to cell transfer of virus avoiding neutralising Abs

host factors:

- inadequate immune response, both quantitative and qualitative
- deletion of HIV-specific cytotoxic CD8+ Vβ restricted clones
- lag between generation of neutralising antibody response and mutation of viral epitopes
- lag in generation of HIV-specific CTLs as viral antigens change.

1.3.5.1   Viral   factors.
Mutation rate of HIV-1.

An important feature of HIV-1 is the high degree of polymorphism
demonstrated between viral isolates. Such polymorphism has been shown for
viral isolates from separate individuals and also for isolates from the same
individual, from early in the infection. It has been suggested that at any point
during the course of infection there exists a population of viruses or pseudo-
species with sequence fidelity to the original infecting strains of virus but
varying by up to 15%. The polymorphism is not uniform throughout the
genome but certain sequences show a degree of conservation reflecting
functional importance, with *env* showing the greatest variation particularly
in certain hypervariable domains. Base changes in *gag* and *pol* are
predominantly at the third base position resulting in silent mutations while
more than 50% of those in env are at first or second bases resulting in amino-
acid changes in gp120. These observations are the result of pressure from the
requirement to conserve functionally important amino-acids and the selective
pressure exerted by the immune response favouring 'escape' mutants.

The high degree of polymorphism is a consequence of 1) base incorporation
infidelity by the virally encoded RNA dependent DNA polymerase, reverse
transcriptase (RTase) and 2) the generation of new strains by the
'recombination' that occurs during the process of reverse transcription,
creating a single dsDNA proviral genome from two ssRNA virion genomes.
Mutation rates have been estimated from sequence analysis of viral isolates in
patients, for *env* as 1.58×10^{-2} to 3.17×10^{-3} per base per year and for *gag*
between 1.85×10^{-3} and 3.70×10^{-4} per base per year (Hahn et al. 1986; Saag et al.
1988). The rate of mutation occurring during a single cycle of replication has
been demonstrated to be similar in T-cell lines and HeLa cells with a rate of 3-4
$\times 10^{-5}$ mutations per base pair per cycle (Mansky 1996) indicating that the rate
was independent of cellular context and presumably a property of the virus. It
has been calculated that, assuming 10^9 cells are infected daily each and every
possible point mutation occurs 10^4-10^5 times per day in an HIV-1 infected
individual (Coffin 1995).

The rapid rate of mutation has important consequences for the dynamic
interaction of viral population present in an infected individual with immune
defence mechanisms. For the viral population the benefits would be the
creation of escape mutants able to elude both humoral and cell mediated
immune destruction, and variation of coreceptor usage may extend the
susceptible cell range (Clapham and Weiss 1997). There may also be advantages
to the virus in the generation of large numbers of defective viral particles
(Coffin 1995).

An important consequence of the mutation rate of HIV-1 is the rapid appearance of drug-resistance following the introduction of antiviral agents. Strains resistant to nucleoside analogues have been shown to be present at very low frequency even before the introduction of the agent (Najera et al. 1995), these variants will have a growth disadvantage before the introduction of the agent compared to the 'wild-type' virus but when exposed to the drug rapidly dominate the population. The use of multiple drug combinations (RTase inhibitors, protease inhibitors) have been shown to improve the suppression of viral replication and improve CD4+ lymphocyte numbers. The presence of chronically infected cells such as monocytes, dendritic cells and microglial cells may prevent the complete elimination of virus from an individual and act as reservoirs allowing the inevitable appearance of strains resistant to the multiple agents.

Trapping of viral particles within the follicular dendritic cell network of lymphoid tissue appears to be important in the initial fall of viral titres following acute infection and is important in maintaining infection during the 'latent' phase (Pantaleo and Fauci 1996) possibly by facilitating the infection of CD4+ cells by 'concentrating' viral particles in an environment where cell-cell contact occurs.

1.3.5.2   Host factors.

The host response to the acute HIV-1 infection is complex consisting of a specific cell mediated (mainly CD8+, Vβ restricted T-cells) response, HIV-specific antibody response and cytokine response. The role of each of these arms of the immune response in the early suppression of viral replication has been difficult to dissect out and it is likely that all contribute to reduce viral growth.

Humoral response.

A specific humoral response to HIV-1 antigens appears early in the infection, with antibodies to Env, Nef and Gag proteins. These form the basis for the diagnosis of HIV-1 infection by ELISA and western blot assays. However, despite showing neutralisation of laboratory strains of HIV-1 in culture, the sera from infected individuals during the acute phase do not effectively neutralise autologous virus particularly as the infection progresses (Koup et al. 1994; Moore et al. 1994). This may reflect the rapid mutation of HIV-1 with the emergence of new variants advantaged by the selective pressure of the immune response. The situation is rather like that of Alice with the Red Queen, having to run faster and faster to stay in the same place. A randomised trial of passive immunotherapy using heat-inactivated plasma from HIV-1 sero-

positive individuals was reported to slow the rate of progression to AIDS, possibly by undermining the ability of the virus to keep ahead of the immune response (Vittecoq et al. 1992).

The HIV-1 specific antibodies may still contribute to the clearance of viral particles from the peripheral blood by the formation of virus-antibody-complement complexes which are removed by reticuloendothelial cells or the FDC network in the hyperplastic lymph nodes (Pantaleo and Fauci 1996). As the infection enters the chronic steady state phase, often months after the initial seroconversion, neutralising antibodies that are reactive against autologous viral isolates appear (Koup et al. 1994). Why these are only detectable at this stage but not during the acute infection is unclear but may be a result of changes in the rate of viral replication and the generation of 'escape' mutations allowing the lagging humoral response to 'catch-up'.

Cell mediated response.

During the acute infection there is an oligoclonal expansion of CD8+ T-lymphocytes with restricted Vb usage, within this population are HIV-specific cytotoxic T-lymphocytes (CTLs) (Pantaleo et al. 1994). The increase of CD8+ CTLs may be 20-fold above the normal range and occurs at the time of the fall in viral titres (Pantaleo and Fauci 1996). A delayed CTL response has been associated with a prolonged acute phase and persistently elevated viral titres (Koup et al. 1994). A more detailed analysis of CTL $V\beta$ restriction during the acute response showed a correlation between the $V\beta$ repertoire and clinical progression of disease; in subjects displaying a major expansion of a single Vb family generally had a rapid progression of disease with AIDS occurring within 2-4 years; individuals with a moderate expansion of two $V\beta$ families showed an intermediate rate of disease progression; finally a third group of individuals with multiple, minor or no expansion of $V\beta$ families had a slow rate of progression to AIDS. HIV-specific CTL clones can be eliminated in the absence of viral epitope changes, a limited number of clones showing high levels of proliferation followed by death despite persistence of the antigenic target, HIV virally infected cells and this may be an important mechanism for the observed clinical differences between rapid progressors and long-term survivors.

In the acute phase CTL are directed against epitopes present on the structural proteins Gag and Env and to a lesser extent the non-structural proteins Tat and Rev. The HIV-specific CTL precursor (CTLp) frequency varies between individuals from 1 CTLp in 100 cells to 1 in 10,000 (Koup et al. 1994; Pantaleo and Fauci 1996). HIV-1 variants appear during the course of infection that are

not recognised by the prevailing CTL clones in vivo, these escape mutants appear early in the infection and are almost certainly a result of the rapid viral replication and selection exerted by the immune response (Phillips et al. 1991).

In addition to a direct cytotoxic effect, CD8+ T lymphocytes also secrete a variety of soluble factors, including cytokines and β-chemokines, that inhibit viral replication.

Cytokine networks and HIV-1.

The immune system is exquisitely regulated by a complex system of signals, mediated by cell-cell contact via surface molecular interactions and by the release and recognition of soluble factors, mainly cytokines. These factors are released by a wide variety of cell types, including T-lymphocytes (CD4+ and CD8+), monocytes/macrophages and stromal cells. These factors have a complex effect upon HIV-1 influencing the replication of the virus and the destruction of viral particles and virally infected cells. The main approaches to investigating the role of cytokines in the pathology and progression of the infection have been determined *in vitro*, using cell lines, primary cell isolates and *in vivo* by assaying levels of specific cytokines in HIV infected individuals. Many of the *in vitro* experiments have used laboratory derived strains of HIV-1 and cell lines chronically infected with HIV and how closely these experimental systems reflect the *in vivo* situation is debatable. Cytokines and other factors can influence the binding of virus and the cell tropism (e.g. RANTES, MIP-1α, SDF-1), alter cell signalling pathways and indirectly modulate proviral expression via elements within the 5'-LTR (section 1.5). In addition, cytokines, by influencing the immune response to the virus and systemic effects (malaise, weight loss), will play a part in determining the clinical progression of the disease (reviewed by Fauci 1996).

Numerous individual cytokines induce HIV-1 expression when added to acutely or chronically infected cells in culture (Fauci 1996). These include IL-1β, IL-2, IL-3, IL-12, TNF-α and TNF-β, colony stimulating factors M-CSF and GM-CSF. The cytokines TGF-β, IL-4, IL-10 , IL-13 and IFN-γ can either induce or repress HIV replication depending upon the particular culture conditions employed whereas IFN-α and IFN-β only appear to repress. The inflammatory cytokines TNF-α, IL-1β and IL-6 are the most potent inducers of HIV-1 expression, IL-10 has been shown to exert a marked inhibitory effect on HIV-1 replication *in vivo*, in part because of down-regulating the expression of TNF-α and IL-1β. The effect of specific cytokines and combinations of cytokines will also depend upon the tissue context, the concentration of cytokine producing cells within lymphoid tissue is higher with the potential for paracrine loops to operate.

The cytokine microenvironment in tissues may not only modulate viral replication but may have an effect on the expression of the co-receptors CCR5 and CXCR4 augmenting viral entry, conversely the release of the natural chemokine ligands for these co-receptors will potentially block HIV entry in target cells. These effects are particularly relevant if there is active inflammation with release of the inflammatory cytokines TNF-α, TNF-β, IL-1β and IL-6 which stimulate the production of β-chemokines and may up-regulate the receptors. This could be a contributing mechanism for the increased risk of HIV transmission to individuals with active STDs such as gonorrhoea (Quinn et al. 1987). Conversely, individuals with consistently elevated levels of the β-chemokines appear to be relatively resistant to infection by HIV-1 (Paxton et al. 1996) possibly because the co-receptor molecules in these individuals are occupied by the natural ligands thus blocking HIV-1 binding, an effect demonstrated in vitro.

1.3.6 Events during the asymptomatic phase.

Using sensitive HIV RNA assay methods, quantitative RT-PCR, nucleic acid sequence-based amplification and branched-DNA assay, viral RNA levels in plasma can be accurately measured and monitored throughout the course of infection. Viral RNA assay has produced important information regarding the behaviour of HIV-1 during the clinical phases of the infection and provides results useful in predicting the future progress of infection.

Within a few weeks of the primary infection, concomitant with the mounting immune response, viral RNA falls by 2-3 logs and, after a period of fluctuation that may last up to 6 months, stabilises at what has been called the 'set-point'. The set point of HIV-1 RNA level varies between individuals and reflects the steady-state rate of viral replication (Henrard et al. 1995). There is an inverse relationship of HIV-1 RNA and CD4+ counts and a correlation with disease progression, individuals with higher steady-state levels are at risk of a more rapid progression to AIDS (reviewed by Saag et al. 1996). The number of HIV-1 RNA copies range from 200 to 10^6 per ml but most individuals show levels during the asymptomatic phase of 10^3 to 10^5 copies per ml. The rapid decline of plasma viral RNA levels seen after starting effective anti-viral therapy allows the half-life of plasma viral particles to be calculated, this was found to be in the order of only 6 hours and to maintain the steady-state levels of virus the replication rate must be 10^8 to 10^9 viral particles per day, an equal number being removed (Ho et al. 1995; Perelson et al. 1995; Wei et al. 1995). Virus production from infected cells has been calculated to proceed for only 2 days with the half-life of an infected cell being only 1.6 days with cell death limiting viral production, by several potential mechanisms as discussed above.

The estimated time for a completed viral replication cycle, from infection of a cell to the release of progeny viral particles, was 2.5 days translating into a rate of 140 viral life cycles per year of infection (Perelson et al. 1995). These results have led to a complete re-evaluation of the importance of chronically infected cells *in vivo*, with 99% of viral progeny resulting from recently infected and short-lived CD4+ cells there appears to be a small (~1%) contribution from monocytes/macrophages, dendritic cells and CNS microglial cells. The rate of viral production by these chronically infected cells may under certain conditions increase. Exogenous factors such as cytokines, infective organisms such as tuberculosis, mycoplasma and other viruses (HSV; CMV) can dramatically increase the rate of viral replication by these cells (Finnegan et al 1996).

In parallel with the rapid destruction and replication of HIV virions during the steady-state asymptomatic phase the numbers of CD4+ cells remain relatively stable after an initial fall during the acute infection with a brief surge as plasma viral titres fall. There appears to be a poorly understood homeostatic mechanism to maintain T-lymphocyte numbers which results in replacement of those cells destroyed by either direct cytopathic effects of HIV-1 infection or because of immune-mediated destruction. The rate of replacement of CD4+ cells has been calculated from the rate and magnitude of CD4+ recovery following anti-viral therapy, as many as 2×10^9 cells per day being destroyed and replaced which represents 5% of the total number of CD4+ cells estimated to be about 2×10^{11} in an individual (Ho et al. 1995). As HIV-1 more efficiently infects and establishes a productive infection in CD4+ that are activated in response to antigen or cytokine signals, viral replication is also greatly increased in activated CD4+ cells, presumably hastening the rate of destruction. These CD4+ cells are potentially the most important to the host, replicating to replace those already destroyed by HIV-1 or because they are responding to infection by other agents such as CMV, mycobacteria. There have been several epidemiological studies demonstrating a more rapid progression to AIDS in individuals with concurrent infections (Orenstein, Fox, and Wahl 1997), conversely, the early treatment of infections and the use of prophylactic therapy delays disease progression (Stein et al. 1994). Immunisations also transiently increase the level of plasma HIV-1 RNA, presumably by CD4+ cell activation and enhanced HIV-1 replication (Stanley et al. 1996; Staprans et al. 1995).

The source of these uninfected CD4+ cells is unclear, the majority of T-lymphocytes in a normal individual are present in lymphoid tissues (98%), whether CD4+ cells destroyed by HIV-1 infection are replaced from this pool or

63

from a smaller, replication competent, pool of cells is not known. In either case, at the rapid rate of cell destruction and replacement, the source of available uninfected CD4+ cells becomes depleted resulting in the steady fall in CD4+ count seen after the period of stability, and an involution of lymphoid tissue. As the CD4+ count falls below 500×10^9 per ml the risk of opportunistic viral, bacterial, fungal and protozoan infections increases and the patient enters the symptomatic phase of the infection (fig 1.3.1). The absolute number of CD4+ cells may not be the only determinant for opportunistic infection. The repertoire of immune response would be restricted by selective clonal attrition predisposing the individual to the same agents, such as CMV, HSV, mycobacteria and mycoplasma (Blanchard and Montagnier 1994) which caused the persisting and relapsing infections responsible for specific CD4+ T-cell activation.

As the humoral response to HIV-1 antigens continues, neutralising antibody effective against autologous viral isolates appear and persist throughout the asymptomatic phase together with antibodies specific for p24, Nef, Rev, Tat, Vpr and Vpu.

1.3.7 Final phases of infection.
Symptomatic illness and progression to AIDS.
In the late phase of infection and preceding the fall of CD4+ cell number there is a shift from M-tropic viral strains present in the early and chronic phases of infection, to T-tropic strains, the viral and host factors that underlay this change of tropism are unknown but presumably reflect a survival/replication advantage of these strains during the extreme immunological dysfunction present in the late phase of infection. In part the shift in tropism is due to a change in co-receptor usage, Fauci has suggested that a crucial factor underlying the shift is a change in the concentration of blocking β-chemokines, in particular SDF-1, in lymphoid tissue that results from the involution of this tissue in the late stages of infection (Fauci 1996).

Before the development of AIDS complex diseases CD4+ cell numbers fall more rapidly with a steady rise in HIV-1 RNA levels (fig 1.3.1). The risk of developing AIDS related disease or dying within 24 months is <5% for individuals with CD4+ counts >500 cells per ml and rises to >70% for those with CD4+ counts <50 per ml (Stein et al 1992). As plasma levels of virus rise, virus associated with lymphoid tissue decreases, presumably as a result of the disruption of normal structure and involution that occurs in these tissues with cell loss and fibrosis.

Profound immunosuppression is a feature of the late stages of infection, in part due to the loss of functional CD4+ cells but other factors such as decrease

of immunoregulatory and pro-inflammatory cytokines may be important (Fauci 1996). Humoral and cell mediated responses are impaired; HIV-1 specific CTLs disappear and neutralising antibodies become undetectable (Haynes et al 1996). There are high levels of certain cytokines, for example IL-10 and IFN-γ possibly due to the excess of CD8+ T-lymphocytes as CD4+ cells are lost. Recurring infections from opportunistic organisms occur with increasing frequency and severity, becoming more difficult to control with drugs. Kaposi's sarcoma (KS), an AIDS defining condition, occurs more commonly in male homosexuals infected with HIV-1 and has been shown to be associated with a newly isolated herpes virus, human herpes virus 8. Other features of the late stage of HIV-1 infection are cachexia, due to a combination of factors such as chronic opportunistic infections, malabsorption due to gut parasites (Giardiasis, cryptosporidium), CMV enteritis and high levels of TNF-α. Death usually results from overwhelming infection or from the AIDS related malignancies Kaposi's sarcoma and high grade B cell non-Hodgkin's lymphoma (Aboulafia 1994).

1.3.8 Factors influencing disease progression.

There has been considerable interest in determining the factors that influence the progression of infection, clearly such information has important implications for the management and treatment of infected individuals (reviewed by Haynes, Pantaleo, and Fauci 1996).

As indicated in section 1.3.4, not all individuals infected with HIV-1 show the same rate of progress through the clinical stages of infection. Early epidemiological studies revealed some host determined factors that appeared to influence the rate of progression to AIDS; male homosexuals progressed to AIDS more rapidly than haemophiliacs, although the difference was significantly reduced when Kaposi's sarcoma was removed as an AIDS-defining illness (Biggar 1990). Age at the time of initial exposure was shown to be the most important determinant of time for progression to AIDS in haemophiliacs, those infected in preadolescence and early adulthood progressing more slowly to AIDS than individuals infected after 45 years of age (Biggar 1990). Children infected in the neonatal period show three patterns of disease progression, a proportion progress rapidly developing AIDS-defining illnesses within weeks or months, a further group develop AIDS by 24 months and a small number of infants remain asymptomatic for several years (Blattner 1991).

Four patterns of disease progression have more recently been described: 'Typical progressors'; 'rapid progressors'; 'long-term nonprogressors' and 'long-term survivors'. Detailed descriptions of these disease patterns and the

possible factors involved in determining the disease course in infected individuals can be found in the reviews by Pantaleo and Haynes (Haynes et al 1996; Pantaleo and Fauci 1996). Some of the host and viral factors involved in the 'rapid progressor', long-term non-progressor' and 'long-term survivor' patterns will be discussed.

1.3.8.1 Rapid progressors.

The most important factor associated with this clinical sub-group are persistently elevated HIV-1 RNA levels following the acute infection. It is not yet clear whether the clinical severity of the acute infection or the level of viraemia are in any way predictive of clinical outcome but the duration of the acute illness is prolonged. Evidence indicates a relatively inefficient immune response to HIV-1, in particular cell mediated responses show a limited CTL Vβ repertoire which quickly disappears. Rapid progressors have lower levels of anti-HIV antibodies and high levels of enhancing antibodies. Certain genetic factors may be important, such as human leucocyte antigens expression (HLA) some of which have been associated with rapid progression. These associations may not be due to the HLA alleles *per se* but result from linkage dysequilibrium with closely linked genetic loci, for example complement factors and TNF-α. The importance of the natural ligands for the co-receptors utilised by HIV-1 in influencing the progression of disease in these individuals is not clear but theoretically a poor response by the effector cells secreting these factors (primarily CD8+ T-cells) would favour HIV-1 binding and entry into uninfected CD4+ cells.

There is less evidence for the role of viral factors in rapid progressors. Some in vitro evidence indicates the presence of more rapidly replicating 'virulent' strains in viral isolates from these individuals.

In addition to the factors described the influence of other infective agents has profound effects upon the progression of infection. Active infection with Mycobacterium tuberculosis accelerates the progression to AIDS (Pape et al. 1993), and the course of the infection shows a more rapid progression in Sub-Saharan Africa than in the West (Colebunders and Latif 1991; Nagelkerke et al. 1990) possibly reflecting the effect of untreated or sub-optimally treated concurrent infections.

1.3.8.2 Long-term non-progressors and long-term survivors.

There appear to be both viral and host determined factors influencing progression of infection in these groups of individuals. Of particular interest are the small percentage (5%) of HIV-1 infected individuals that are long term non-progressors (LTNPs).

Evidence indicates an efficient and sustained immune response to the initial infection with rapid clearance of virus. High levels of MHC class 1 restricted anti-HIV specific CTLs appear soon after exposure and are sustained, strong CD8+ non-MHC restricted HIV suppressor activity (possibly mediated by soluble factors, β-chemokines) and high levels of anti-HIV antibodies (reviewed by Haynes 1996). The architecture of lymph node germinal centres and follicular dendritic cells is preserved (Pantaleo 1996). An extended dominance of M-tropic strains of virus has been demonstrated in LTNPs.

These observations, although important, do not provide an explanation of the underlying mechanisms favouring the hosts immune response over the replication of the virus. Individuals with persistently elevated levels of β-chemokines could disadvantage the infection of CD4+ cells by blocking viral binding, slowing down the rate of viral replication and CD4+ cell loss. Evidence exists for the presence of less pathogenic strains of HIV-1 in LTNPs; sequencing viral isolates from LTNPs have revealed a variety of defects potentially disabling viral replication in the host; deletion mutations of *nef* have been identified in viral isolates from LTNPs with normal CD4+ counts and no evidence of immunosuppression (Deacon et al. 1995; Kirchoff and al 1995), analogous to macaques infected with the *nef* deletion mutant SIVmac239Δnef reported by Desrosiers group (Kestler et al. 1991). Mutations altering important regulatory sequences in the LTR, at the NF-κB and Sp1 sites have been described (Cao et al. 1995; Huang and Jeang 1993) and although not yet reported, mutations of other important sequences in the LTR or TAR region could also impair viral replication *in vivo*. Investigation of viral strains associated with an attenuated infection has been hampered by the difficulty in isolating virus for analysis, in one study virus was isolated from only 3 of 10 LTNPs (Cao et al. 1995), in another virus was obtained from the lymph nodes from 7 of 15 patients but not from the plasma (Pantaleo et al. 1995).

1.3.9 Resistance to infection.

Epidemiological surveys revealed individuals who were repeatedly exposed to infection with HIV-1, because of persistent high-risk activities such as sex-workers and those with HIV-1 infected partners, but remained uninfected. Considerable interest was taken in these individuals as they appeared to have a constitutional resistance to HIV-1. Peripheral blood MNC from these individuals displayed variable resistance to infection with laboratory strains of virus (Paxton et al. 1996), however the discovery of the inhibitory effects of the β-chemokines suggested that their receptors would be of particular interest. The peripheral blood MNC from two subjects proved to be highly resistant to infection by M-tropic strains of HIV-1, but easily infected with T-

tropic strains. The gene for the M-tropism receptor, CCR5, in these individuals was found to be abnormal with an almost identical 32-base pair deletion (Liu et al. 1996) producing a truncated protein that was not expressed at the cell surface. Both individuals were homozygous for the deletion and had no detectable CCR5 although they were both well with no symptoms indicative of impaired defence mechanisms. Presumably there is a degree of redundancy amongst this group of chemokines which could not be exploited by M-tropic HIV-1 strains. Population studies show that nearly 10-16.8% of Europeans are heterozygous for this deletion, it has not yet been detected in African or Japanese subjects tested (Samson et al. 1996). The frequency of the defective CCR5 gene was less in HIV-1 infected Caucasians than in a similar, uninfected population suggesting that the homozygous state may confer a degree of protection from infection. Most primary infections by HIV-1 appear to involve M-tropic strains and these are the dominant form present in the early phases of the disease (Roos et al. 1992; Zhu et al. 1993) even when both M- and T-tropic strains are present in the infecting partner (Cornelissen et al. 1995). An interesting implication from the finding that the lack of CCR5 co-receptor confers resistance to infection is that the main route of entry of HIV-1 is selectively by M-tropic strains via susceptible cells expressing CD4 and CCR5, such cells are found in mucosal surfaces (tissue macrophages, dendritic cells). Individuals with high levels of b-chemokines also appear to have a relative resistance to infection with HIV-1 (discussed above) possibly by blocking viral binding to the co-receptor.

1.3.10   HIV-1 and CNS infection.

A significant proportion of patient infected with HIV develop neurological complications, in a small number of patients these appear early in the course of infection while in the majority the neurological problems appear late and parallel the fall of CD4 numbers and the development of immunodeficiency. Early manifestations include aseptic meningitis, encephalitis, acute and chronic inflammatory demyelinating polyneuropathies, mononeuritis multiplex and polymyositis. These conditions are often transient and are thought to have an auto-immune basis, responding to treatment with immuno-suppressive or modulating therapies. The main neurological complications of HIV infection occur later in the course of infection and include a specific clinical syndrome characterised by progressive dementia which has been termed the AIDS dementia complex (also called HIV-associated dementia). AIDS dementia complex occurs in approximately 20% of HIV-infected patients and initially appears an impairment of cognitive function with poor concentration, forgetfulness, lethargy, apathy and emotional lability. Motor

function is affected with loss of fine motor control, unsteady gait and tremor. The symptoms progressively worsen at a rate that varies between individuals but in most cases results in severe dementia with mutism, spasticity, incontinence and immobility. The neuropsychological features are similar to those displayed by patients with subcortical dementia caused by Huntingdon's disease, Parkinson's disease or progressive supranuclear palsy. A related condition occurs in children infected with HIV (Epstein et al 1984 and 1985). An HIV-associated myelopathy occurs in ~20% of patients, with clinical features similar to sub-acute degeneration of the cord as caused by severe vitamin B12 deficiency. A painful peripheral polyneuropathy can develop in ~40% of patients with AIDS (reviewed by Glass and Johnson 1996; Price et al. 1988; Sharer 1992).

1.3.10.1 Pathogenesis of HIV-associated neurological syndromes.

The aetiology of HIV-associated neuropathies is complex and under active investigation. Opportunistic infections with agents such as CMV, toxoplasma, polyoma virus cause significant neurological problems, indeed initially the AIDS related dementia was assumed to be caused by such agents, in particular CMV (Snider et al. 1983). However with the discovery of HIV as the causative agent of AIDS lead to the hypothesis that the virus could directly cause CNS disease. Animal lentiviruses cause neurological disease in animals, for example visna in sheep (Nathanson et al. 1985) and caprine arthritis-encephalitis virus (CAEV) in goats (Narayan et al. 1980), supporting the hypothesis that HIV had the potential for infecting the CNS. Improvements in methods of virus culture and detection produced evidence for the presence of infectious virus within brain tissue and CSF (Ho et al. 1985; Levy et al. 1985), in situ studies have been able to localise the specific types of cells infected (Bagasra et al. 1996). Gross changes in the brains of patients with HIV related dementia have been determined from nuclear magnetic resonance imaging and from pathological specimens. Changes include cerebral atrophy, ventricular enlargement and loss of white and deep grey matter.

Details of the neuropathological features of HIV infection are beyond the scope of this thesis and have been reviewed extensively (for example Sharer 1992). The important features include the presence of multinucleated giant cells, usually in perivascular sites but also in isolation, microglial nodules, with an absence of inflammatory lymphocytes and minimal tissue destruction although tissue cavitation and necrosis can rarely occur. White matter changes include diffuse myelin pallor caused by demyelination and a leukoencephalopathy with myelin destruction, infiltration by macrophages

and multinucleated giant cells. A final feature is a reactive astrocytosis in the cortex and sub-cortical grey matter.

1.3.10.2   Cellular targets of HIV in the CNS.

Evidence for the type of cells in the CNS infected by HIV has be indirectly determined using *in vitro* cultures of cell lines and directly from pathological specimens. The *in vitro* results are not fully supported by the direct examination of brain tissue from patients with HIV, cell lines of astrocyte and neuronal origin can be infected *in vitro* even though they do not express CD4 (Clapham et al. 1986), possibly by binding to galactosyl ceramide (Harouse et al. 1991). Evidence from tissue immuno-histochemistry and *in situ* hybridisation for HIV-1 specific RNA suggests that the most productively infected cells in the CNS are microglial cells and tissue macrophages, indeed the multinucleated giant cells are derived from a macrophage related cell (Koenig et al. 1986). There have been a few reports in which infection of astrocytes, neuronal cells and microvascular endothelial cells have been described (Nuovo et al. 1994; Saito et al. 1994; Tornatore et al. 1994). In a report which described the results of *in situ* PCR on brain tissue from 22 HIV infected individuals, viral RNA was detectable in several different cell types, most frequently in microglial cells and macrophages but variable levels of HIV-1 infection could be demonstrated in microvascular endothelial cells, neurones and astrocytes (Bagasra et al. 1996). Several groups have demonstrated the presence of macrophage-tropic but not lymphocyte-tropic strains of HIV or SIV in brain tissues (Cheng-Meyer et al. 1989; Epstein et al. 1991; Power et al. 1994; Sharma et al. 1992).

1.3.10.3   Entry of HIV into the CNS.

Infection of the CNS occurs soon after the initial exposure to virus, a single case report of a patient who died 15 days after iatrogenic exposure to a large inoculum (600-700 TCID) of virus intravenously showed infected cells within the CNS (Davis et al. 1992). The intravenous route is not the 'natural' route of HIV infection and the CNS probably becomes infected during the viraemic phase following infection through a mucosal surface. Virus may enter the CNS either via microvascular endothelial cells (Wiley et al. 1986) or carried into the CNS by infected macrophages.

1.3.10.4   Mechanisms of neurotoxicity.

In the absence of strong evidence for direct viral infection of neuronal cells other mediators of the observed neurotoxicity of HIV have been suggested. These include soluble factors released by infected macrophages and microglial

cells such as cytokines (TNF-α, IL-1β, IL-6), arachidonic acid metabolites and undefined soluble toxins (reviewed by Glass and Johnson 1996). Viral proteins have also been demonstrated to be toxic to neuronal cells and glial cells *in vitro* and *in vivo*. The viral envelope protein gp120 can cause neuronal injury either directly by non-CD4 binding mechanism or indirectly via the activation of macrophages and microglial cells which then secrete cytokines (Harouse et al. 1991). Transgenic mice expressing a truncated gp120 inserted in a gene normally expressed in astrocytes (glial fibrillary acidic protein) developed neuropathological changes in their CNS but gp120 could not be detected (Toggas et al. 1994).

1.3.11   Summary.

Infection of the CNS by HIV appears to occur early in the course of infection possibly due to the entry of infected macrophages into the CNS or from direct infection of microvascular endothelial cells. The disruption of normal immune-regulation in the early stages of infection may lead to the inflammatory and auto-immune neurological syndromes that occur at this stage. Although HIV- infected cells are present virtually from the onset of HIV infection, predominantly in microglial cells and macrophages, few symptoms are manifest . As the course of infection progresses, with falling CD4 counts and worsening immunosuppression, features of a sub-cortical dementia appear in approximately one fifth of patients. Why AIDS related dementia occurs in only a proportion of patients when virus has been shown to infect the CNS in most patients is not known. The pattern of CNS infection demonstrated by HIV is unusual and unlike other viral neurological infections such as Herpes simplex, polyoma virus or polio virus where infection at a cellular level correlate with pathological and clinical findings. Clearly HIV infection of the CNS is more complex than the direct effect of virus on specific cell types and results from the interaction of host determined factors, the involvement of other infectious agents (perhaps as yet undefined) and an indirect effect of due to viral proteins and soluble factors such as cytokines.

1.4 Transcriptional Regulation in Eukaryotes.

1.4.1 Introduction:
General mechanisms of mRNA transcription.

The integrated proviral genome depends upon the cellular RNA pol II transcriptional apparatus, cellular transcriptional enhancers and the virally determined proteins *tat* and *rev* for the fully co-ordinated production of spliced and unspliced mRNAs encoding viral proteins and full length genomic RNA. The following section briefly describes the process and regulation of transcription in eukaryotic cells, the basal transcriptional machinery, transcriptional co-factors and enhancers.

Section 1.5 describes the regulation of transcription of the HIV-1 provirus and the roles of cellular transcription factors.

1.4.1.1   The RNA polymerases.

The nuclear genes of eukaryotes are transcribed into RNA by one of three RNA polymerases; RNA polymerase I (pol I) transcribes from genes (class I genes) encoding 5.8S, 18S and 28S ribosomal RNAs (rRNAs), RNA pol II transcribes class II genes which includes all genes that encode mRNAs and from some genes that determine certain small nuclear RNAs (snRNAs). The third enzyme, RNA pol III transcribes class III genes which include all transfer RNAs (tRNAs), 5S rRNA and some snRNAs. Although many of the basal transcriptional components of these three polymerases are common to all three the HIV-1 provirus is transcribed only by pol II. The following section briefly describes transcription driven by RNA pol II. (reviewed by Latchman 1995; Roeder 1996; Zawel and Reinberg 1995).

1.4.1.2.   Transcription of mRNA by RNA pol II.

The transcription of mRNA in eukaryotes requires the co-ordinated activity of a large number of proteins interacting via protein-DNA and protein-protein contacts. The DNA provides the appropriate milieu for the initiation and regulation of transcription directing the RNA pol II complex to the precise start point for mRNA synthesis. Further specific DNA elements interact with protein factors that influence the activity of the RNA pol II complex in a cell-specific manner and serve to link the production of mRNA (and hence proteins) to events outside the cell via signalling pathways. The regions of DNA that mediate control over class II genes can be identified both upstream and downstream of the start point for mRNA transcription, often regulatory elements can be located many kilobases away from the gene these elements control. These specific DNA elements are recognised in a highly specific

manner by protein factors, transcription factors. Although there are common themes, each gene has its own unique arrangement of these DNA elements, characterised by the type of elements, their positioning and relationship to one another and the gene. Further layers of complexity are achieved by the protein factors that bind to these DNA elements, some are ubiquitous and appear to be necessary for any transcription to occur while others are cell-specific or appear transiently in a tightly regulated fashion determined by the state of differentiation and the environment of the cell.

1.4.2 Promoters, enhancers and silencers.

Promoters are defined as regions of DNA located close to the initiation site of transcription and can extend for several hundred bases upstream (fig 1.4.1) Gene promoters usually consist of two components, a transcription promoter element and the regulatory element. Within the promoter a minimal core-promoter element can normally be identified. The core-promoter elements are defined as 'minimal DNA elements that are necessary and sufficient for accurate transcription initiation by RNA pol II in reconstituted cell-free systems'. Promoters and enhancers can be viewed as containing funtionally distinct 'modules' (Dynan 1989). The commonest of these elements is the TATA box, with the consensus sequence TATAa/tAa/t, located between -30 and -25 bases from the transcriptional start site (by convention, bases upstream of the transcription start site are prefixed by a minus ('-') while those downstream by a plus ('+')) and a pyrimidine-rich initiator (Inr, consensus YYANt/aYY) located at the start site (fig. 1.4.1). The HIV-1 minimal promoter is unusual in that it contains both a TATA box and an initiator region.

Fig. 1.4.1.   Representation of the promoter for a eukaryote gene.
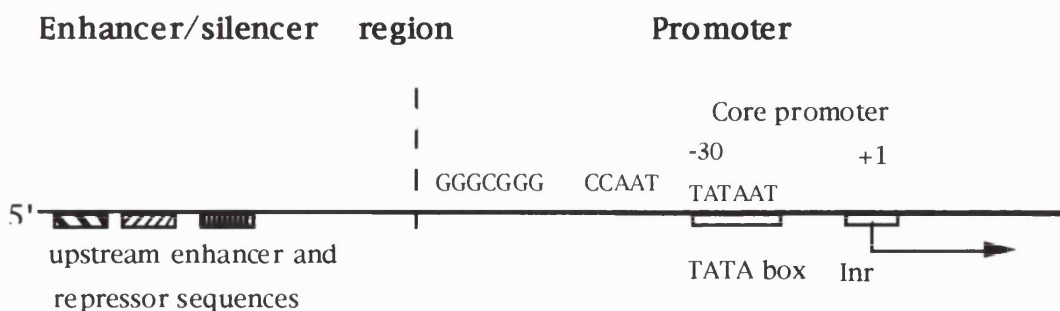
Some promoters lack a TATA-element (so called 'TATA-less' promoters), examples are the mammalian terminal deoxynucleotidyl transferase gene and

the SV40 late genes. Transcription from these genes is dependent upon the INR and flanking sequences (reviewed by Kaufmann and Smale 1994; Roeder 1991).

1.4.2.1 CCAAT box and Sp1 sites.

In many class II genes the promoter also contains other important DNA sequences adjacent to the TATA box, the presence of which are required for full transcription. Examples are the CCAAT box and GC-rich sequences. These sites bind proteins that enhance the formation and stability of the pre-initiation complex (PIC) and can thus increase the basal level of transcription. The factors that bind to these sites are present in most cell types and are constitutively expressed. The CCAAT box binds CBP (CCAAT binding protein), the ubiquitous transcription factor Sp1 binds to the GC box (reviewed by McKnight and Tjian 1986). The HIV-1 promoter contains three Sp1 sites corresponding to a GC-rich region from -46 to -78 bp (Jones et al. 1986).

1.4.2.2 Enhancer and silencer elements.

(reviewed by Latchman 1995; Thompson 1992).

Transcriptional enhancers are defined by four characteristics;

1. action over distance - an enhancer element can activate RNA pol II through a promoter even when positioned many kilobases away from the promoter;
1. the action of an enhancer is independent of its orientation;
1. an enhancer element can operate when positioned upstream or downstream of the promoter, or within the transcribed region of the gene;
1. a transcriptional enhancer alone cannot activate transcription from the minimal core promoter (i.e. TATA box and INR) but requires the presence of functional promoters.

Although transcriptional enhancer elements are functionally distinct from promoters this distinction is largely operational, many protein binding sequences are present in both promoters and enhancers and bind appropriate transcription factors in a similar fashion. Enhancer elements may also be located relatively close to the initiation site of transcription, for example the histone H2A enhancer is just over 100 bases upstream of the start site of transcription (Grosschedl and Birnsteil 1980). The presence of an enhancer element cannot be predicted from the native sequence and structure of regulatory elements, the manipulation of these elements *in vitro* is required to demonstrate the defining characteristics given above.

Enhancer elements for genes appear to be composed of distinct sub-elements or modules (Muller et al 1988) each module containing the binding sites for several different transcription factors which work together as a functional

unit to activate gene transcription. The individual short DNA sequences within a module (the binding sites for transcription factors) are given the general term of enhanson. Thus a transcriptional enhancer has a hierarchical structure, a further level of complexity that has been shown to have important functional implications

Silencers.

Analogous to enhancer elements, silencer elements are regions of DNA that modulate the activity of RNA pol II, but in this case to reduce the rate of transcription. The distinction between enhancer and silencer elements is convenient but may be inappropriate. Although some DNA elements have been shown to act only as transcriptional silencers, other elements can act as either enhancers or silencers, the functional effect of an element determined by the context, flanking sequences, relative abundance of competing DNA-binding factors and co-factors.

1.4.3 The 'general transcription factors'.

The current theories of mRNA synthesis, from the early stages of specific protein-DNA interaction locating the RNA pol II complex to the start site for transcription through to the termination of transcription and release of pol II, have been developed from results obtained from yeast and Drosophila genetics, nuclease protection assay, electrophoretic mobility shift and X-ray crystallography (reviewed by McKnight 1996; Roeder 1996; Chiang et al 1993). Several protein factors, in addition to RNA pol II, are required for what has been termed the minimal transcriptional machinery, (also called the basal transcriptional apparatus). This minimal transcriptional machinery is sufficient for the initiation and transcription of RNA from a consensus TATAA element *in vitro*. The nomenclature of these factors evolved from their sequence of elution from ion-exchange columns and has been retained.

The core promoter directs the assembly of the pre-initiation complex (PIC) consisting of RNA pol II and the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE and TFIIH, details of which are given below. The components of the basal transcriptional apparatus are highly conserved from yeast to humans.

Table 1.4.1 **The 'general eukaryotic transcription factors'.**

Factor		Function
TFIID	⌠ TBP	Core consensus sequence (TATAA-box) binding. TFIIB binding.
	⌡ TAFs	Core promoter recognition in non-TATAA promoters. Possible requirement in some TATA- promoters, via direct DNA interaction. Interaction with factors binding to upstream enhancer/repressor elements
TFIIA		Stabilisation of TBP binding, stabilisation of TAF$_{II}$s-DNA interactions, anti-repressor function.
TFIIB		RNA pol II-TFIIF recruitment; directs RNA pol II to start site.
TFIIF		Complexed with RNA pol II. Promoter targeting of pol II; reduces non-specific RNA pol II-DNA interactions.
RNA pol II		Catalytic activity in RNA synthesis; recruits TFIIE.
TFIIE		TFIIH recruitment; modulation of helicase and kinase activities of TFIIH.
TFIIH		Multiple functions: DNA melting at promoter site (helicase function); phosphorylation of carboxy- terminal domain (CTD) of RNA pol II.

Nomenclature:

TBP → TATA-binding polypeptide

TAF$_{II}$s → TBP-associated factors

The transcription of mRNA can be seen as a repeating cycle of events, each round resulting in the synthesis of a single mRNA transcript. The cycle has been divided into discrete but contiguous stages:

1) pre-initiator complex assembly;

2) initiation and promoter clearance;

3) transcript elongation;

4) termination.

1.4.3.1   Pre-initiation complex (PIC) assembly.

TFIID, consisting of TBP and TAFs, binds to the TATA-box. The factor TBP binds specifically to the TATA-box sequences and can function in the absence of TAFs *in vitro* although it is found complexed with TBP in cell extracts. Binding of TBP induces a sharp bend in the DNA which may physically facilitate subsequent interaction between TFIID-TFIIB and RNA pol II. The precise role of TAFs remains unclear (reviewed by Burley and Roeder 1996; Hampsey and Reinberg 1997), in yeast they appear to be required for transcription *in vivo* from TATA-less promoters but in higher eukaryotes they have acquired more complex functions involving the interaction of TFIID complex with RNA pol II and also act as coactivators modulating transcription by interacting with factors bound to upstream enhancer/repressor elements.

There are two models for the assembly of the PIC following TFIID binding to the TATA-box. In a multi-step model the general transcription factors and RNA pol II assemble at the promoter through protein-DNA and protein-protein interactions in an ordered, hierarchical manner (Zawel and Reinberg 1995). Following the binding of TFIID (TBP) to the TATA-box TFIIB binds, this recruits the RNA pol II/TFIIF complex and serves to locate the complex at the transcriptional start site. Further GTFs then bind until the full complex has assembled. In the second model, the GTFs and RNA pol II are present as a preformed complex consisting of RNA pol II, TFIIB, TFIIF, TFIIH and other polypeptides, SRBs, (suppressers of RNA polymerase B), such complexes have been isolated from cells of several species from yeast to humans. This complex is termed the RNA pol II holoenzyme and it is this complex that binds to the initiation site, directed by interactions with TFIID-TFIIB. The two models are not mutually exclusive, indeed either may operate *in vivo* depending upon the promoter context. Both however result in the assembly of the PIC ready for the initiation of transcription (fig 1.4.2b).

Figure 1.4.2

Formation of pre-initiation complex and promoter clearance.

(details in text)
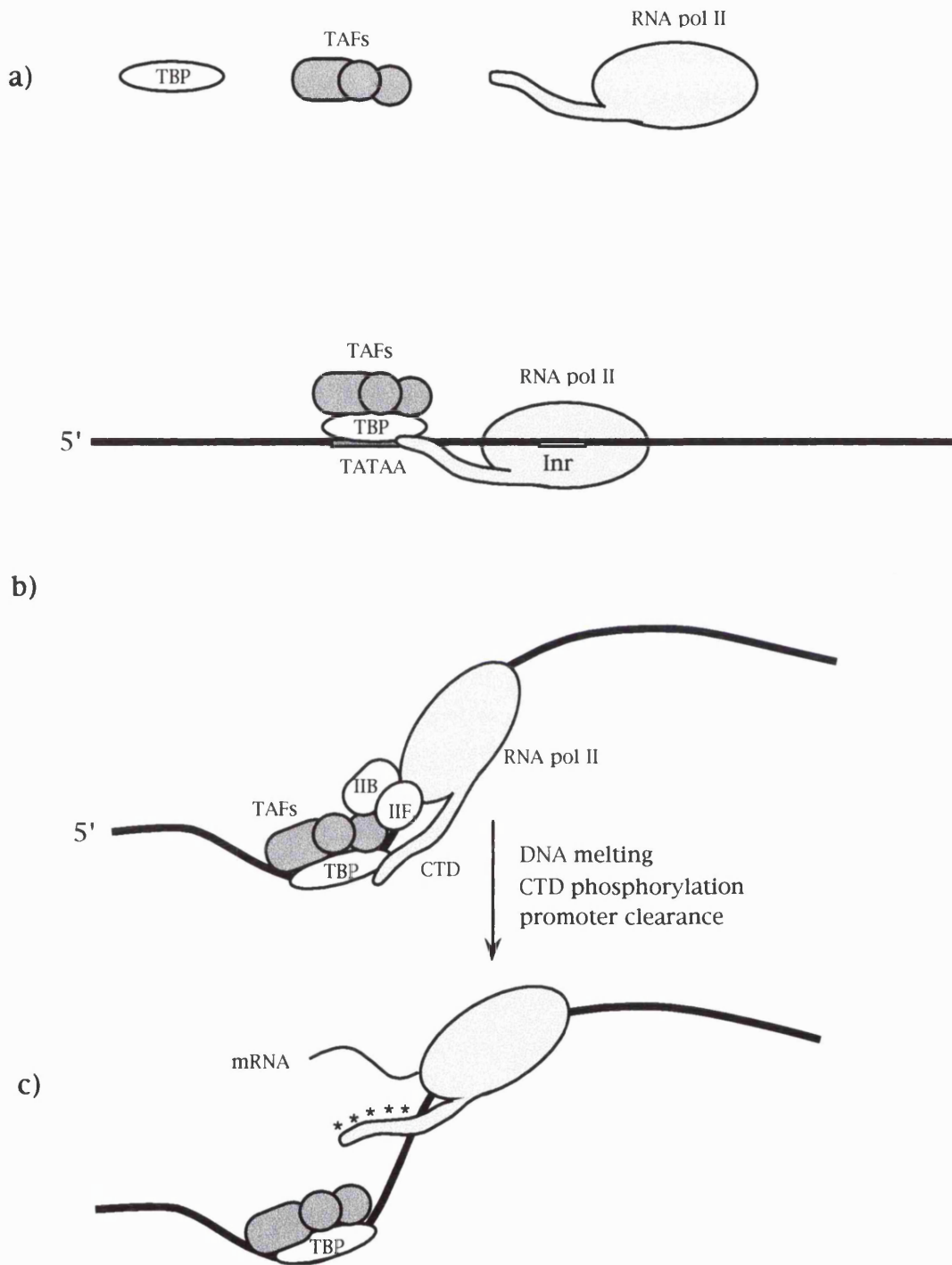
1.4.3.2   Promoter clearance.

Once the PIC has assembled, a series of events occur leading to the release of the RNA pol II complex from TFID/TFIIB. For these steps to happen, ATP and other ribonucleoside triphosphates are required. TFIIH induces DNA melting (PIC activation) ~10bp upstream from the initiation site, the melted region extends downstream and at this time initiation of transcription, the formation of the first phosphodiester bond, occurs. The next step involves hyperphosphorylation of the carboxy terminal domain (CTD) of RNA pol II, mediated either directly by TFIIH or indirectly through TFIIH and cyclin-dependent kinases. Hyperphosphorylation of the CTD is associated with release of RNA pol II-TFIIF complex from TBP/TAFs-TFIIB, the release of factors required for the formation of the PIC and allows binding to the RNA pol II complex of factors required for elongation. TBP/TFIID has been shown *in vitro* to remain associated with the TATA-box facilitating further rounds of transcription initiation. Other factors in the PIC dissociate sequentially, TFIIB and TFIIE followed by TFIIH, although some evidence suggests a role for TFIIH in elongation (reviewed by Svejstrup et al 1996).

1.4.3.3   Elongation.

The RNA pol II complex must physically progress along the DNA transcribing a complementary strand of RNA, in doing so there are physical blocks to procession such as proteins bound avidly to the DNA, nucleosomes, and specific sequences which stall the complex, and functional limitations such as limiting concentrations of ribonucleoside triphosphates. The rate of procession can be influenced by the interaction of a variety of factors some acting as co-factors binding to the complex such as TFIIF, SII (TFIIS) and SIII, these prevent the RNA pol II complex stalling or pausing and facilitate efficient transcription Transcriptional activators, bound to upstream elements in the enhancer region of the promoter, may also influence elongation (reviewed by Reines et al 1996).

1.4.3.4   Termination.

A specific sequence of DNA signals the termination of transcription. The nascent mRNA is released from the complex and the CTD region of RNA pol II is dephosphorylated by a specific phosphatase. The RNA pol II-TFIIF complex can now repeat PIC assembly on the initiation site.

1.4.4   Modulation of mRNA expression.

1.4.4.1   Transcriptional activators, repressors, co-factors and co-repressors.

The level of transcription determined by the core promoter element and the basal transcriptional machinery *in vivo* is modulated by proteins acting upon the RNA pol II complex at any of the stages of transcription described above. Eukaryotic transcription factors are modular with a specific DNA-binding domain and a separate activation domain. The activation domains of eukaryotic transcription factors fall into several groups according to the predominant characteristic; proline-rich (e.g. NF1/CTF, AP2 and Jun), glutamine-rich (e.g. Sp1 and the homeotic proteins) and the acidic activators. The majority of eukaryotic transcription factors contain acidic activating domains, lacking a consistent amino acid sequence homology but containing a high proportion of acidic amino acids. Initially, the model for RNA pol II modulation was based on that of prokaryotes, the general transcription factors were the molecular targets for promoter and enhancer-bound transcriptional activators. Specificity was a function of the DNA-binding domain while the activation domain directly influenced the basal transcriptional apparatus, either by increasing the binding of components of the PIC or altering the conformation of bound factor(s) and in so doing stimulating the activity or stability of the entire complex (reviewed by Bjorklund 1996; Latchman 1995). Such direct interactions have been demonstrated *in vitro* and *in vivo* (Colgan et al 1995; Roberts et al. 1993). However, transcription systems constructed *in vitro* using purified general initiation factors only supported low levels of transcription and could not be activated by the binding of transcriptional activators to elements distal to the PIC (Pugh and Tjian 1990). These results indicated that additional factors were required for activator-enhanced transcription, the function of which was to increase the efficiency of interaction between promoter/enhancer bound transcription factors and components of the basal transcriptional apparatus. These are termed co-activators, of which there are two classes; the TBP-associated factors (TAF$_{II}$s) which are associated with the TATA-binding protein TBP forming TFIID (reviewed by Hampsey and Reinberg 1997; Verrijzer and Tjian 1996) and independent 'free' positive co-factors, PC1, PC2 Dr2/PC3, PC4, PC5, PC6 and HMG2 (reviewed by Kaiser and Meisterernst 1996). The function of co-activators may be to serve as bridging molecules between promoter/enhancer bound activators and the basal transcriptional apparatus and in some manner enhance its function. For example, the co-factor PC4 has been shown to interact with both the general factor TFIIA and the acidic activation domain of VP16 (from Herpes simplex), TFIIA accelerates

and stabilises binding of TFIID to the TATA-box and in this way VP16 enhances transcription (Kaiser et al 1995).

Many other co-activator factors have been identified but one in particular, CREB-binding protein (CBP and a related factor p300) is relevant because it has a central role in mediating transcriptional activation by a number of transcription factors including several involved in the activation of the HIV-1 LTR.

1.4.4.2   Role of CBP/P300 in transcription.

The CREB-binding protein (CBP) and the related protein p300 are general mediators of signal-dependent transcription. They are large nuclear proteins that can interact with a wide variety of DNA-binding factors and the basal transcription apparatus, via TFIIB. The DNA-binding factors are signal-dependent and include CREB, jun/fos, myoD, STATs 1 and 2 and the nuclear hormone receptors. CBP/p300 appears to act as the convergence point for several distinct cellular signalling pathways, mediating and co-ordinating signals that result in the induction of transcription. CBP/p300 also has intrinsic activity as a histone acetyltransferase, acetylation of histone proteins is associated with 'open' chromatin and gene transcription. Nuclear hormone receptors interact directly with CBP/p300 through their ligand binding domains, this is mediated by the steroid-receptor co-activators (SRCs) which are constitutively associated with CBP/p300 (Smith et al. 1996; Yao et al. 1996).

1.4.5 Nuclear hormone receptors.

The nuclear hormone receptor superfamily consists of nearly 200 different but structurally related proteins that are involved with the transduction of extracellular signals to the nucleus and influence gene expression. Generally these proteins serve as receptors for small lipophilic molecules that are able to diffuse across the plasma membrane, they are thus intracellular receptors. The ligands for these receptors include many hormones, such as glucocorticoids, oestrogens, progesterone, androgens, thyroid hormone, vitamin D, morphogenic retinoids such as all-*trans* and 9-*cis* retinoic acid and a variety of other molecules including prostaglandins. In addition to the nuclear hormone receptors with known ligands there are many members of this superfamily that appear not to have an associated ligand, or at least the ligand has not yet been identified. These are termed 'orphan' receptors (reviewed by Enmark and Gustafsson 1996; Mangelsdorf et al. 1995a; Mangelsdorf and Evans 1995b).

The nuclear receptor superfamily has been divided into sub-families based on the ligand binding properties and for the orphan receptors, based on their DNA sequence homologies. The members of this super family are characterised by a similar basic modular structure (figure 1.4.4).

Figure 1.4.4   **Schematic diagram of a typical NHR**
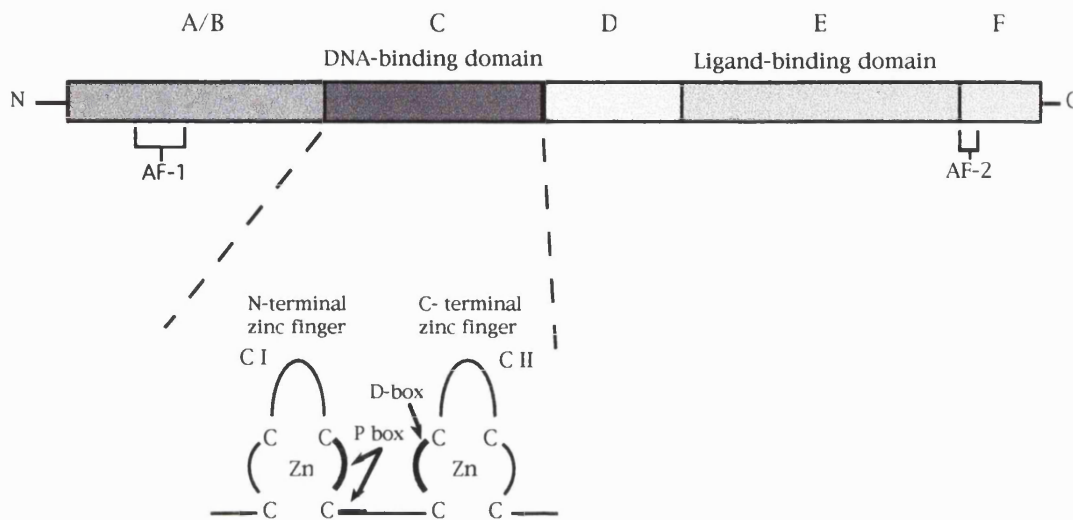
The prototypic NHR can be divided into five regions based on structural and functional similarities, A-F. The regions 'C' and 'E' contain the conserved DNA-binding domain and ligand-binding domain respectively. Transactivation domains (AF-1 and AF-2) are shown. The receptors are found as homo- and hetero-dimers, dimerisation domains are present in regions 'C' and 'E'. The DNA-binding domain contains two highly conserved zinc-finger motifs one of which determines DNA-binding sequence specificity (P-box in C I), the other is involved in dimerisation (sequences in the D-box of C II). The combination of ligand-binding and DNA-binding properties in a single protein appears to be highly successful and has been exploited in a wide range of signal pathways. A simple description of the function of a NHR is as a molecular switch, in the absence of ligand the protein is in a non-permissive state, binding of the specific ligand shifts the receptor to a transcriptionally active state. A full description of this large and interesting superfamily of transcription factors is beyond the scope of this thesis. However, there are two sub-families that are of particular relevance to the experimental work described in the following chapters, these are COUP-TF and the retinoic acid receptor sub-families.

1.4.5.1   Retinoic acid receptors RAR and RXR.

RARs and RXRs are members of the thyroid hormone receptor/RAR/vitamin D3 receptor subfamily and share a highly conserved DNA-binding domain segment called the proximal (P)-box that contains an α-helix and determines specific binding to the half-site sequence RGG/TTCA (reviewed by Linney 1992). Three closely homologous proteins, RARα, β, γ all have all-*trans* retinoic acid as ligand and are expressed in a tissue-specific manner. RXR also exists as three related proteins RXRα, β and γ with 9-*cis* retinoic acid as ligand (reviewed by Mangelsdorf and Evans 1995b). RARs and RXRs regulate gene expression by a complex set of mechanisms involving the basic components of receptor (as homo- or hetero-dimer), the response element in the promoter (HRE) and ligand. These receptors have been shown to function as both transcription activators or repressors, a feature that has made these transcription factors and their response elements particularly difficult to characterise. The net effect on transcription is influenced by the receptor type, the dimer composition, the nucleotide sequence of the response element, the spacing of the response element, the presence or absence of ligand, the presence or absence of specific co-activators and co-repressors. RARs can form homodimers that act as transcription activators of specific retinoic acid response elements (RAREs) in the presence of retinoic acid, however RAR homodimers may have no significance *in vivo*, for high affinity binding to RAREs requires the presence of RXR forming RAR/RXR heterodimers. RXR acts as a master regulator of several related NHRs including RAR, ThR, VD3R and PPAR (reviewed by Green 1993; Mangelsdorf and Evans 1995b). A RXR homologue has been identified in Drosophila, Ultraspiracle, suggesting that the regulation of action of these receptors by heterodimer formation evolved before the divergence of vertebrates and invertebrates and has been retained.

Although RAR/RAR homodimers can bind to RAREs *in vitro*, transcriptional activation is absent. RAR/RXR heterodimers transactivate in the presence of all-*trans* retinoic acid but not 9-*cis* RA. The combination of all-*trans* RA and 9-*cis* RA is synergistic. The transcriptional responses induced by the interaction of ligand are thought to be due to conformational changes which promote interactions between receptor dimers, co-activators, co-repressors and the basal transcriptional machinery. In addition, for direct repeat response elements, the polarity of the RAR/RXR heterodimer influences which HREs is preferred. Thus, 5'-RXR/RAR binds to DR2 and DR5, 5'-RAR/RXR binds to DR1 (discussed below).

1.4.5.2 The hormone response elements.

A further level of complexity is determined by the nucleotide sequence and arrangement of the HRE. For members of the RAR/RXR/ThR/VD3R subfamily variations on the half site element RGG/TTCA determines binding specificity. The HRE may consist of a direct repeat (DR), an inverted repeat (IR) or everted repeat (ER).

Figure 1.4.5 shows the arrangement of these elements.

Figure 1.4.5 Nuclear hormone response elements.

Evidence from *in vitro* experiments have led to the creation of a set of rules for the relationship of NHRs and specific HREs. For example, the HRE for VD3R is a DR with a 3 nucleotide spacer (DR3), that of THR a DR4 and RARE a DR2 or DR5. This has been called the 1-to-5 rule (Mangelsdorf et al 1994). However several HREs have been identified in natural gene promoters that deviate from the 1-to-5 rule, functional RAREs with DR1, IR0 and complex arrangements of DRs and ERs. For example an ER8 has been described in the γF-crystallin gene promoter (Tini et al. 1993) and a complex arrangement of an ER8, a DR0 and an IR4 has been characterised in the medium chain acyl-coenzyme A dehydrogenase gene promoter (Carter et al. 1994). In DR5 elements, RXR preferentially occupies the 5' half-site with RAR on the 3'. On a DR1 element this polarity is reversed and results in a potent repression of transcription (Mangelsdorf and Evans 1995b, reviewed by Gronemeyer and Moras 1995).

1.4.4.2 COUP-TF

This member of the nuclear hormone receptor superfamily has no known ligand and is thus classified as an orphan receptor. COUP-TF was first identified as a transcription factor that binds to a direct repeat element in the chicken ovalbumin promoter (Pastorcic et al. 1986; Wang et al. 1989). There are two closely related proteins, COUP-TF I and II and these show variable homologies with other members of the steroid receptor superfamily (reviewed by Tsai and Tsai 1997). COUP-TF I was shown to be identical to a previously cloned protein termed ear3 which had been isolated from a human placental library using a DNA probe derived from the human estrogen receptor, under low stringency conditions

(Miyajima et al 1988). COUP-TF II (also found in the earlier literature as COUP-TF2) was originally cloned and described by Ladias as a factor binding to the apolipoprotein A1 promoter (Ladias and Karathanasis 1991). COUP-TF I and II display overall 87% amino acid sequence identity, with the greatest differences residing in the A/B domain N-terminal to the DNA-binding domain, with only 45% homology. This contains the transactivation domain TF-1, the sequence diversity may imply differential activation/repression activities for COUP-TF I and II. The remainder of the two proteins show 98% homology. COUP-TF I and II have a Mr of 45 kDa.

COUP-TFs bind to DR1 elements with the highest affinity but can also bind to a wide range of DR, IR and ER elements with variable affinities. COUP-TF binds as a homodimer and although heterodimers can be created *in vitro* with RXR and several other NHRs these are not thought to exist naturally. Most naturally occurring COUP-TF elements are DR1, for example in the promoters of apolipoprotein CIII and A1 genes, mouse lactoferrin gene and the chicken apolipoprotein VLDLII gene (Tsai and Tsai 1997). A DR0 COUP-TF site has been described in the promoters of the oxytocin and hemopexin genes, a DR6 in the rat insulin 2 promoter and ER8 in the acyl-CoA dehydrogenase gene promoter (Tsai and Tsai 1997). Although originally described as an enhancer of transcription · from the chicken ovalbumin gene promoter, in the majority of other cases COUP-TF acts as a potent repressor of transcription. As COUP-TF response elements can also function as RAREs, ThREs or VD3REs the potential exists, and has been demonstrated for many elements, for COUP-TF to act as an inhibitor of transcription mediated by these transcription factors (Cooney et al 1992 and 1993; Qui et al 1996; Kliewer et al 1992; Tran et al 1992).

1.4.6 Summary.

The nuclear hormone receptor superfamily of transcription factors contains a large number of related but functionally diverse members. These factors have important roles in the response of cells to external signals mediating growth, differentiation and morphogenesis. The complexity of the interaction between different members, the influence of specific ligands, the nature of the response element and the relative balance of positive and negative co-factors has only recently been realised and further complexities may yet be discovered. In practical terms, the analysis of promoters that contain these response elements is difficult, the net effect on the transcriptional activity of a reporter gene for example will be influenced by parameters that are difficult to define and may be unknown.

Section 1.5   Transcriptional Regulation of HIV-1

Introduction.

The function of the integrated proviral genome is to produce viral mRNA species for translation into viral proteins and full length RNA for packaging into progeny virions. To achieve these aims the provirus behaves as a genetic cassette plugged into the host's DNA and usurping cellular transcriptional machinery. Transcription of HIV RNA is initiated in the 5' LTR and terminates in the 3' LTR. The expression of the viral RNA species is highly regulated and complex, the pattern of replication employed by the virus is dependent upon the cell type and the activation and proliferative states of the cell. The ability of HIV-1 to adopt different replication patterns in a wide range of cell types results from the complexity of interactions between the three major components involved in regulating expression of the proviral genome. These are:

1.   regulatory elements within the LTR
      (and possibly minor elements outside the LTR)

2.   the virally encoded factors, tat and rev

3.   cellular proteins
      e.g. transcription factors and accessory proteins.

The 5'-LTR of HIV-1 has been the focus of intensive investigation since the virus was first isolated and the genome sequenced. Regulatory sequences within the 5'-LTR have been identified and characterised using techniques such as DNAse I footprint analysis, electrophoretic mobility shift assay (band shift), *in situ* DNAse I-footprinting and sequence analysis for consensus transcription factor binding sites (reviewed by Antoni et al 1993; Cullen 1992; Gaynor 1992; Greene 1990; Kingsman and Kingsman 1996).

Numerous potential transcriptional control elements have been identified using these methods, the functional significance of which have been demonstrated with molecular constructs of the LTR and a reporter gene such as chloramphenicol acetyl transferase (CAT). Such constructs have then been used in transient transfection assays in model cell systems such as Cos or HeLa cells, or potential HIV-target cells such as T-cell, monocytic or neuronal cell lines. Extensive mutagenesis of LTR sequences, by point mutation, deletions, linker-scanning, have further delineated important regulatory elements within the LTR. Using model systems to analyse a transcription regulatory element with the complexity of the HIV-1 LTR, although providing a great deal of insight into the roles of some components of the LTR, they are unlikely to display the subtleties of HIV-1 replication in the infected host. Each analytical model system may accentuate one or two dominant features of LTR function

and lose others that, *in vivo*, may be essential for the survival and replication of the virus. This complexity may underlie the apparently conflicting published data on the regulation of HIV-1 expression. A feature of HIV-1 and of lentiviruses in general is the complexity of the genome and the startling economy of DNA usage, some sequences are used to encode two or even three separate proteins (section 1.2.7). There appears to be no 'unnecessary' sequences and it would seem likely that such economy has also been used in the LTR.

1.5.1 Regulatory elements in the HIV-1 5'-LTR.

The 5'-LTR and 3'-LTR are generated by the transformation of the single stranded virion RNA into dsDNA (section 1.2.8.4). Although the two LTRs are essentially identical transcription of RNA is only initiated from the 5'-LTR. The 5'-LTR consists of the U3 region (-454 to -1) the R region (+1 to +99) and the U5 region (+100 to +184). The majority of functionally important regulatory elements are present in the U3 and R regions, although some isolated elements have been identified in the U5 and in downstream coding sequences (van Lint et al. 1991 and 1997; Verdin 1991; el Kharroubi et al 1994).

The LTR has been divided into functional regions, shown in figure 1.5.1, determined mainly by deletion analysis (reviewed by Antoni et al 1993; Gaynor 1992; Greene 1990; Vaishnav and Wong-Staal 1991). The division of the LTR into these regulatory sub-regions is artificial, some elements overlap physically with shared or competing sequence usage. Sequences may function as positive or negative regulatory elements depending upon the type of cell, the activation state of the cell and the presence or absence of regulatory factors and co-factors. The term 'negative regulatory element' has been used since the initial deletion mapping of the LTR, the deletion limits for these experiments often determined by the presence of convenient restriction enzyme sites. The effect of such deletions would reflect the influence of dominant sequences upon the overall level of transcription under the conditions used for the experiment and could miss the more subtle effects of other sequences. Subsequently, the 'NRE' has been demonstrated to contain binding sites for transcription factors that can function as activators of transcription and an alternative term has been suggested, the more apt 'modulatory element' (Gaynor 1992).

1.5.2.1:   HIV-1 LTR: General description.
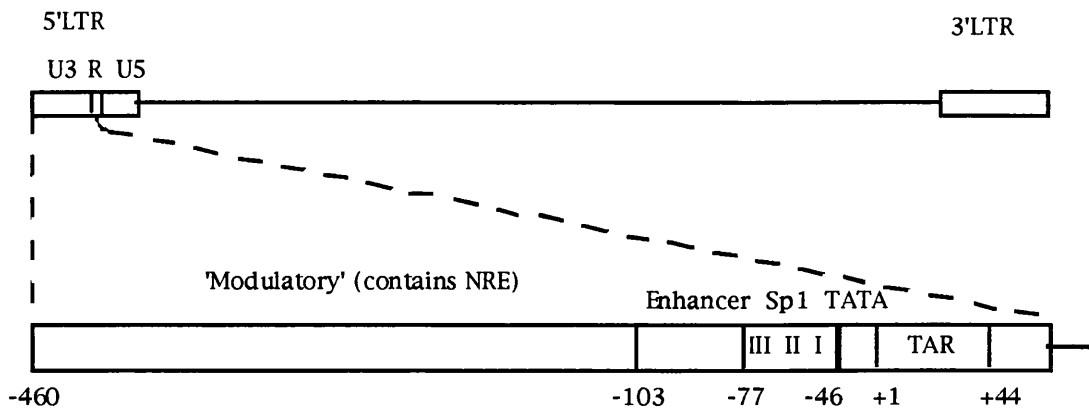
Figure 1.5.1

Diagram of the 5'LTR showing the major functional domains.

Core promoter: TATA box + Initiator region (InR) -27 to +1

(some authors also include TAR)

Basal promoter: core plus Sp1 sites -77 to +1

TAR: binding sequences for Tat +1 to +44

Enhancer: two NF-kB sites -77 to -103

Modulatory region: contains sequences shown to function as enhancer elements or transcriptional repressors. Contains the NRE.

The LTR consists of the core promoter containing the TATA-box (-27 to -23) and a pyrimidine-rich initiation region (-2 to +7) which are sufficient to mediate a low level of transcription in the absence of upstream elements (Figure 1.5.2). The TATA-box is superficially similar to those present in many eukaryotic promoters, however detailed analysis has revealed this region to be complex. The basal transcription factor TFIID, consisting of TBP and TAFs, binds to the TATA-box and is important in the formation of the PIC recruiting additional components of the RNA pol II complex (section 1.4.4). Mutation of the TATA-box reduces basal transcription of otherwise intact LTR-driven transcription in transient transfection assays (Berkhout and Jeang 1992; Olsen and Rosen 1992) but similar mutations in a recombinant proviral DNA result in the loss of infectivity (Lu et al. 1989) these differences in the influence of TATA-box mutations are possibly due to subtle effects on Tat interaction with the PIC. The TATA-box of HIV-1 is flanked by two conserved palindromic sequences termed 'E-boxes' which are the binding sites of two basic-helix-loop-helix (bHLH) transcription factors HTF4 and E47 as has been demonstrated *in vitro* (Zhang et al 1992). The binding of the bHLH factors interferes with TBP binding to the TATA-box but stabilises the interaction with other components of the PIC leading to the suggestion that binding of factors to the E-boxes may be important in the switch from basal (TBP initiated) transcription and activated transcription (involving interaction of Tat and the bHLH factors) (Ou et al. 1994). The complex role of the small region has been indicated by the exchange of the TATA-region of the LTR with the corresponding region from other promoters, this resulted in a marked fall in the level of Tat-induced transcription (Berkhout and Jeang 1992; Olsen and Rosen 1992). Possibly the core promoter can recruit two different transcription initiation complexes, with TBP in the absence of Tat and with bHLH factors binding to the TATA region in the presence of Tat (Lu et al 1993).

1.5.1.2   Initiator Region (InR).

The HIV-1 InR maps from -2 to +7 and is similar to InRs of cellular genes in being pyrimidine rich. Several cellular transcription factors have been shown to bind to sequences overlapping or near to the InR. The factors YY1, LBP-1 (leader binding protein 1, also termed LSF, UBP-1 AND CP-2), USF, and TDP-43 have all been shown to bind to sequences within the -17 to +27 region of the promoter. YY1 is a 68 kDa nuclear matrix protein with a zinc-finger structure and is present in a wide range of cell types. YY1 has been shown to directly interact with a number of viral and cellular factors including Sp1 (Lee et al 1993), TFIID (Chiang and Roeder 1995), adenovirus E1A protein (Shi et al. 1991), CREB (Zhou et al 1995). YY1 has also been shown to interact with

LBP-1, which binds to overlapping sequences in the InR (Romerio et al 1997). YY1 interaction with the InR results in a reduction of LTR-driven transcription and virus production (Margolis et al 1994). LBP-1 (LSF/UBP-1/CP-2) has been identified by several groups as a cellular factor interacting with multiple sites within the core promoter and TAR region of the LTR, a high affinity site map to sequences between -16 to +27, over the InR/TAR, and a low affinity site between -38 to -16 (fig. 1.5.2) (Garcia et al. 1987; Jones et al 1988; Kato et al 1991; Wu et al. 1988a; Yoon et al 1994). LBP-1 binding to the LTR has been shown to have both activating and inhibitory effects on transcription. The addition of LBP-1 has been shown to directly inhibit transcription using *in vitro* transcription assays, binding to the low affinity site blocks TFIID binding to the TATA-box while the addition of LBP-1 after PIC formation inhibits mRNA elongation (Prada et al 1995). However these effects on transcription may be an artefact of the *in vitro* assay as transient expression of LBP-1 has no effect on the *in vivo* activity of the LTR (Yoon et al 1994).

The factor TDP-43 binds to the InR and can block the assembly of Tat-activated transcription complex formation on the site (Ou et al. 1995). The cellular transcription factor USF binds to the InR stimulating transcription (Du et al 1993; di Fagagna et al 1995), an additional USF-binding site has been delineated upstream between -159 to -173 (Garcia et al. 1987) which will be discussed below. A cellular factor CTF/NF1 has been shown to bind to sequences downstream from the transcription start site, at nucleotides 40-45, although no clear functional significance of this interaction has been demonstrated.

The core promoter of HIV-1 LTR, containing the TATA-box and InR, thus appears to display considerable complexity with specific binding sites for cellular proteins with either positive or negative effects on transcription. Transcription appears to be influenced at the stage of TBP/TFIID binding, the assembly of the PIC at the TATA-box or InR, by interaction with factors binding to upstream elements, processivity of the RNA polII complex and by interaction with Tat. In addition to these 'local' events other, less specific determinants may be operating through the core promoter, for example chromatin structure, histone acetylation, the tumour suppressor protein p53 (by interacting with TFIID and other components of the transcription complex), adding yet further layers of complexity (reviewed by Kingsman and Kingsman 1996).

Figure 1.5.2   Binding sites for cellular factors in the InR.
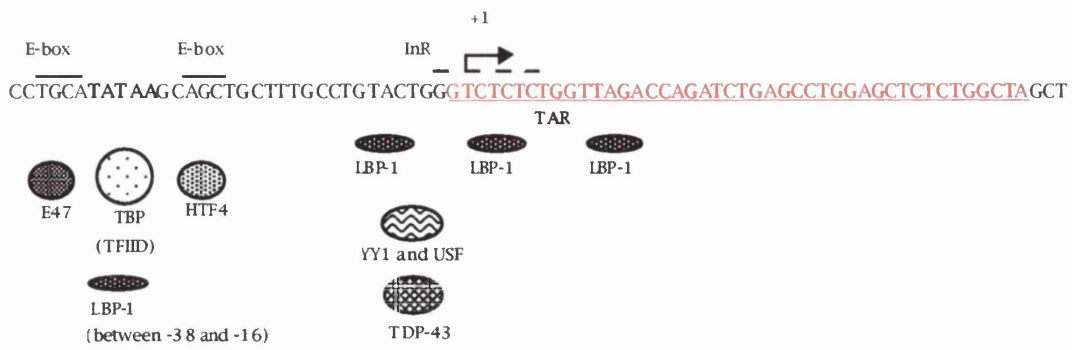
Figure 1.5.2   For details of abbreviations see text.

1.5.1.3   Role of Tat/TAR interaction.

The *tat* gene of HIV-1 encodes Tat, an 86-amino acid protein that functions as a powerful transactivator of the viral genome (Arya et al. 1985; Sodroski et al. 1985; reviewed by Rosen 1991 and in section 1.2.7.7). Tat is required for HIV replication, infectious molecular clones of HIV with mutations in *tat* are unable to replicate (Fisher et al. 1986). Tat is an early viral gene product, translated from small, multiply-spliced RNA species, the primary function of Tat is as a cis-acting transcriptional enhancer increasing the level of LTR-driven expression. A simple model for the role of Tat is that initial viral gene expression, in the absence of rev, is determined by cellular transcription factors and results in the production of increasing levels of Tat protein producing a specific positive feed-back loop further boosting viral gene expression. This a feature of several viral early gene products that have evolved to enhance viral replication through the manipulation of transcription, however Tat has a mechanism of action that is unique to lentiviruses, it interacts with the 5'-end of nascent mRNA binding to specific sequences between nucleotides +1 to +44, the *trans*-activation responsive element or TAR. The TAR sequence forms a stable secondary structure with a stem-loop configuration (figure 1.5.3), a critical bulge is present in the stem (nucleotides +23 to +25). For full tat activation the stem, bulge and terminal loop are required. Nucleotides in the bulge and flanking base-paired nucleotides are required for Tat binding (reviewed by Antoni et al 1993; Cullen 1991; Kingsman and Kingsman 1996). Although Tat binds to the bulge sequences *in vitro* with no requirement for the terminal hexanucleotide loop (nucleotides +30 to +35) (Dingwall et al. 1989; Hamy et al. 1993) or stem, any mutations in the loop results in loss of transactivation by Tat (Feng and Holland 1988; Roy et al. 1990). The function of the terminal loop in tat-dependent transactivation is not clear. Cellular factors have been shown to specifically bind to the loop RNA and may be important by interacting with Tat and stabilising the tat/TAR interaction or by acting as transcriptional co-factors for Tat, in particular a 68 kDa protein (Marciniak et al 1990) and a 185 kDa protein (Sheline et al 1991a; Wu, Garcia, and Sigman 1991). Tat transactivation is inefficient in rodent cells, the introduction of human chromosome 12 in rat-human hybrid cell lines results in high levels of Tat transactivation, and Tat-activation of transcription cannot be reproduced in yeast cells. These results provide indirect evidence for a critical role of human cellular factors in Tat/TAR transactivation (Alonso et al 1992; Hart et al. 1989; Newstein et al. 1990). Human cyclin T1 has been shown to be the factor that is expressed by chromosome 12 (Wimmer et al 1999). The NF-kB elements alo require the presence of a functionally active Tat-TAR complex for full

activation of HIV transcription in peripheral blood CD4 cells (Alcamia et al 1995).

Several other cellular factors have been shown to interact with sequences in TAR, proteins between 70-110 kDa in size bind to the bulge and transactivate HIV-1 LTR *in vitro* (Sheline et al 1991b) other factors bind to sequences in the stem regions (reviewed by Antoni et al 1993). Cellular factors have also been shown to interact directly with Tat, and are likely to be important as co-factors involved in the full *in vivo* function of Tat. Thus Tat has been shown to bind to Sp1 *in vitro* and Tat/Sp1 complex can be immunoprecipitated from HIV-infected cells (Jeang et al. 1993), Tat-associated protein (TAP) binds to the conserved core region of Tat and also binds to TFIIB, suggesting a possible function for TAP as a bridging factor linking Tat to the early events in the formation of the PIC (Yu et al. 1995). Screening of an expression library with Tat identified a factor TBP-1 (Tat binding protein 1) which is a member of a family of related factors including MSS1 and SUG1, factors involved in transcription and cell cycle control. TBP-1 and MSS1 bind to Tat *in vitro* enhancing Tat transactivation (Ohana et al. 1993).

1.5.1.4   Cellular kinases interacting with Tat.

A number of cellular kinases have been shown to bind to Tat, termed Tat-associated kinases (TAKs). Tat has been shown to bind to cyclin-dependent kinase 7 (CDK7) leading to an interaction with CDK-activating kinase (CAK) complex increasing the phosphorylation of the carboxy-terminal domain (CTD) of RNA pol II and thus increasing the rate of RNA elongation (Cujec et al. 1997b). Another cellular kinase that has been shown to interact with Tat is P-TEFb, a Drosophila Cdc2-related kinase that is known to regulate RNA polII activity via the CTD, the human homologue of P-TEFb, PITLRE, associates with the activation domain of Tat (Mancebo et al. 1997; Zhu et al. 1997).

1.5.1.5   Function of Tat and the Tat/TAR complex.

The importance of Tat/TAR interaction in the regulation of HIV-1 gene expression has resulted in considerable investigation of both tat and TAR, and has been extensively reviewed (Antoni et al 1993; Kingsman and Kingsman 1996). Early evidence suggested a role for Tat in both transcription and post-transcription events. Nuclear run-off experiments showed increased initiation of transcription in the presence of Tat (Kao et al. 1987; Laspia et al 1989), however a more significant effect of Tat is the effect on the RNA pol II complex which becomes more processive and generates full-length mRNA (Kao et al. 1987; Garcia-Martinez et al 1995). The current models for Tat/TAR transactivation of the viral LTR is shown in figure 1.5.3.

Fig. 1.5.3   Model of Tat-mediated transactivation.

1. Early transcription: no Tat, short RNA transcripts

short transcripts = TAR elements
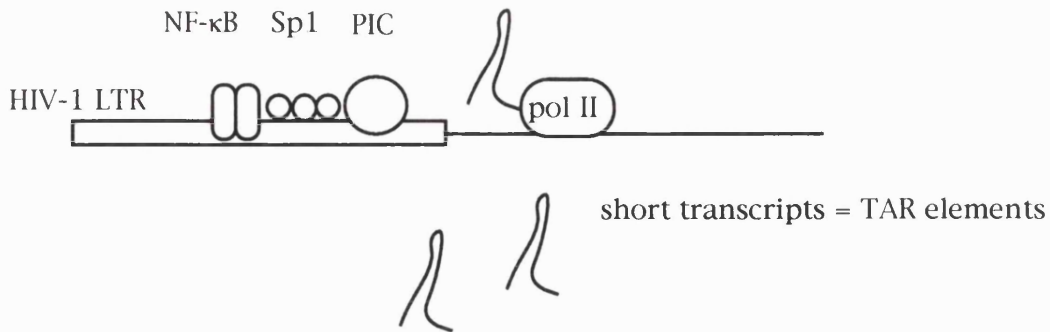
2. [Tat] increasing
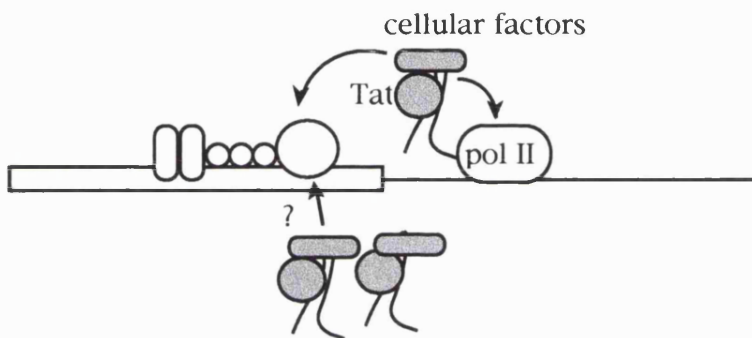
cellular factors

3. Efficient PIC formation and increased processivity of pol II
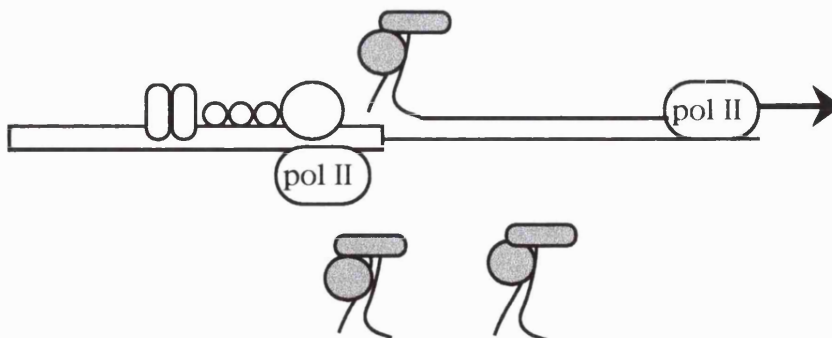
Legend:

Tat   For details see text

cellular factors

In the absence of Tat in a quiescent cell, LTR-driven transcription is low, the initiation of transcription is inefficient with the formation of an RNA pol II complex with poor processivity producing short RNA transcripts (part '1' in fig. 1.5.3). The formation of these short transcripts is in part due to specific sequences within the initiation site from -5 to +26 and a secondary element from +40 to +59. This has been termed the 'inducer of short transcripts' or IST and has characteristics of a positive element that may induce the formation of a PIC with poor processivity. Transcription initiated from the IST terminates at the base of the TAR stem-loop, thus the short transcripts are essentially TAR elements. Activation of the cell increases the level of transcription initiation and is dependent upon cellular factors binding to upstream elements including the Sp1 sites. This produces increasing amounts of short RNA transcripts with a small amount of multiply spliced mRNA reaching the cytoplasm to be translated into Tat and rev proteins. The short RNA transcripts contain the TAR hairpin-loop which is the target for Tat, the Tat/TAR complex now functions as a powerful transactivator initiating a change in the functional character of the PIC and in the rate of PIC formation (part '2' in fig. 1.5.3). The RNA pol II complex thus formed now overcomes the strong transcriptional stop signals downstream of the start site, possibly because of more efficient phosphorylation of the CTD, to generate increasing amounts of full length mRNA (part '3' in fig. 1.5.3). Interestingly, the level of short RNA transcripts does not change even though more full-length RNA species appear (Kao et al. 1987) implying that the IST continues to function independently.

In this model, TAR behaves as a localisation signal to bring Tat, and cellular factors, into proximity of the initiation site. In this role Tat is behaving as a typical positive transactivating factor. In addition, Tat has properties, perhaps by binding kinases, similar to elongation factors such as TFIIF, SII and SIII (section 1.4.4). Tat itself directly interacts with components of the PIC and indirectly alters the functional efficiency of the RNA pol II complex by recruiting important kinases to the complex. Once bound to the transcription complex Tat does not require the presence of TAR RNA (Keen et al 1997) indeed Tat can bind to the transcription complex in the absence of TAR and produce an efficient transcription complex able to generate full length transcripts (Cujec et al. 1997a). However *in vivo* Tat requires the TAR region to be in close proximity of the transcription start site, moving the TAR downstream results in loss of Tat-activation (Selby et al. 1989). It is not clear whether the Tat/TAR complex 'reaches forward' to activate the RNA pol II complex which generated that particular TAR or 'reaches back' to influence the formation and function of subsequent pol II complexes (reviewed by Kingsman and Kingsman 1996).

Why lentiviruses have evolved a transactivation system based on the interaction of the viral transactivator Tat and an RNA target is unknown.

Many viruses, including other retroviruses, have virally-encoded transactivation factors that bind to specific DNA elements within the promoter/enhancer elements of the viral genes or interact with cellular transcription factors. Possibly nascent short RNA transcripts remain in close proximity to the viral transcription initiation site and effectively increase the number of target sites for Tat binding, providing abundant Tat and the cellular factors recruited by Tat and TAR. If transcription initiation is limited by the concentration or availability of transactivation factors, the virus effectively pushes the reaction in its favour, otherwise a single integrated provirus must compete with numerous cellular gene promoter elements for the available components of transcription. The requirement for Tat seems to be transient, after the introduction of Tat into cells containing an LTR-driven reporter gene the increase of new RNA synthesis lasts for approximately four hours and is then switched-off, this is an active process as it requires new protein synthesis (Drysdale and Pavlakis 1991).

1.5.2 The basal promoter; the role of Sp1

Immediately adjacent to the core promoter are three binding sites for the ubiquitous mammalian transcription factor Sp1, originally delineated by DNA footprinting using purified Sp1 from HeLa cells (Jones et al. 1986).

Sp1 is a member of a family of closely related transcription factors consisting of Sp1, Sp3 and Sp4 (reviewed by Hagen et al. 1994). Sp1 and Sp4 function as transcriptional activators while Sp3 has both activating and repressor activity depending upon the promoter and cell type. Sp1 is constitutively expressed in a wide range of cell types. Sp1 is a strong positive transactivating factor, DNA-binding being mediated by three zinc-finger motifs and transcription activation via two distinct glutamine-rich domains. The consensus binding site for Sp1 is an asymmetric decanucleotide sequence (G/TGGGCGGPuPuPy), the affinity of binding determined in part by the match of a sequence with the consensus. Sp1-responsive promoters often contain multiple Sp1-binding sites, although a single site is sufficient to activate transcription. The Sp1 sites are usually located 40 to 150 bp upstream of the transcription start site. The three Sp1 binding sites in the HIV-1 LTR conform to this pattern, they are GC-rich and extend from -78 to -46 bp.

In transient transfection assays mutation or deletion of the Sp1 sites caused a marked reduction of basal LTR-driven expression of a reporter gene and

reduced Tat-activation (Harrich et al. 1989). Replication competent molecular clones of HIV-1 in which the Sp1 sites were deleted were able to replicate in T cells in the presence of high levels of NF-κB, in the absence of NF-κB the same Sp1 deleted virus was unable to replicate indicating a degree of functional redundancy (Parrott et al. 1991). Cooperative interaction between Sp1 and NF-κB through their adjacent binding sites is required for optimal HIV-1 enhancer activation and inducible LTR-driven gene expression (Majello et al. 1994; Perkins et al. 1993). The role of Sp1 appears to be complex involving potential interactions with other cellular factors including COUP-TF (Rohr et al 1997), p53 (Gualberto and Baldwin 1995) and with Tat (Jeang et al. 1993; Kamine and Chinnadurai 1992). In addition, while all three members of the Sp1 family of transcription factors can bind to the GC-rich elements in the HIV-1 LTR, Sp1 and Sp4 function as transcriptional activators while Sp3 represses LTR activity in HeLa and Drosophila SL2 cells. Sp3 has been shown to act as a repressor of the HIV-1 LTR in transient transfection assays in a microglial cell line (Rohr et al 1997) a function that required the DNA-binding domain of Sp3, implying that Sp3 acts by 'squelching', blocking the interaction of Sp1 with TFIIB and other factors.

1.5.3   The enhancer region -82 to -103

A feature of HIV-1 infection is a rapid and substantial induction of viral replication on exposure of host cells to a wide range of activating factors. Initial experimental work used T-lymphocytes as models for both viral infection and in transient transfection assays using LTR-reporter gene constructs. T lymphocytes can be readily activated *in vitro* using a combination of phorbol myristate acetate (PMA) and a plant lectin such as phytohaemagglutinin (PHA), resulting in cell proliferation and the expression of activation markers including the IL-2 receptor and secretion of cytokines such as IL-2 and TNFα. In cell culture, T-cell lines and primary isolates infected with HIV-1 produce low levels of progeny virus, exposure to activating agents markedly increased viral proliferation (Harada et al. 1986; Zagury et al. 1986).

Using an LTR-CAT reporter gene construct in transient transfection into H9 (T cell), Raji (B-cell line), HeLa and feline epithelial cells, Rosen *et al* delineated an enhancer region from -137 to -17 which behaved as a typical enhancer by functioning in an orientation and position independent manner in a heterologous promoter construct (Rosen et al 1985). Nabel and Baltimore identified an element in the enhancer that interacted with a Jurkat nuclear protein present in activated cells but not unstimulated cells (Nabel and

Baltimore 1987). This element contained sequences identical to those of previously identified as binding sites for NF-κB, a transcription factor that activated the transcription of the immunoglobulin kappa (κ) gene in maturing B-lymphocytes. NF-κB activity in Jurkat cells correlated with the induction of LTR-driven expression of a CAT reporter gene while mutation of bases within the NF-κB sites abolished the induction of LTR-CAT on exposure of the cells to PMA and PHA. Nabel and Baltimore identified two NF-κB sites lying adjacent to the Sp1 III site (figure 1.5.4), confirmed by DNA-footprinting (Wu et al. 1988b). Since the original description of the NF-κB sites in the HIV-1 LTR considerable research has been focussed on the role of the sites and NF-κB in regulating the expression of HIV-1 in the full range of potential host cells.

Fig 1.5.4   NF-κB and Sp1 sites in the HIV-1 LTR
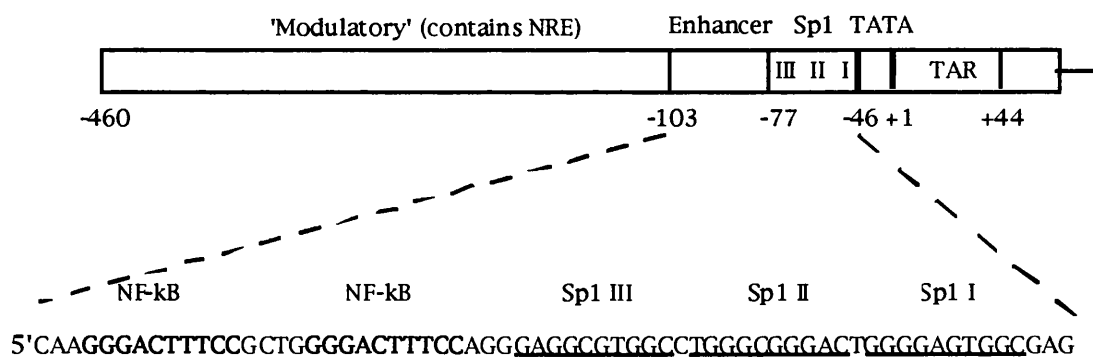
The two NF-κB sites are indicated by boxes, the three Sp1 sites lie downstream.

Before describing the role of NF-κB in the regulation of HIV-1 expression some details on the function of NF-κB in eukaryotic gene expression is required (section 1.5.3.1).

1.5.3.1   The role of NF-κB

NF-κB is present in a wide range of cell types and many genes have regulatory regions containing NF-κB binding sites, (κB sites). In vertebrates these genes are primarily involved in the rapid response of cells to external stimuli, particularly those involved in defence responses to pathogens and stress. Thus κB sites have been identified in the regulatory regions of immune response genes (Igκ, IL-2, IL-2 receptor, MHC class I and II), inflammatory and acute phase proteins (IL-1, IL-6, TNFα, TNFβ, β-interferon), growth control genes (P53, c-myc, Ras) and cell adhesion molecules (I-CAM, V-CAM, E-selectin) (reviewed by Verma et al. 1995). Pathogenic viruses have exploited the central role of NF-κB in cellular activation and carry NF-κB binding elements in their own promoter enhancer elements, for example CMV, SV40, adenovirus and HIV-1.

1.5.3.2   Regulation of NF-κB.

Activation of NF-κB does not require protein synthesis; in unstimulated cells NF-κB is present in the cytoplasm as a complex with an inhibitory protein, IκB. Diverse stimulatory signals all operate through the phosphorylation of IκB by specific cellular kinases (reviewed by Israel 1997; May and Ghosh 1998; Miyamoto and Verma 1995) including PKC-zeta and PKC-beta (Folgueira et al. 1996; Kim et al. 1996). Phosphorylation of IκB results in the phosphorylation-dependent ubiquination of IκB and its degradation in proteosomes while still complexed to the NF-κB dimer. Removal of IκB reveals a nuclear localisation signal on NF-κB which can then translocate to the nucleus and bind to the enhancer elements of NF-κB-responsive genes (figure 1.5.5).

Fig. 1.5.5 Pathways for NF-κB activation.
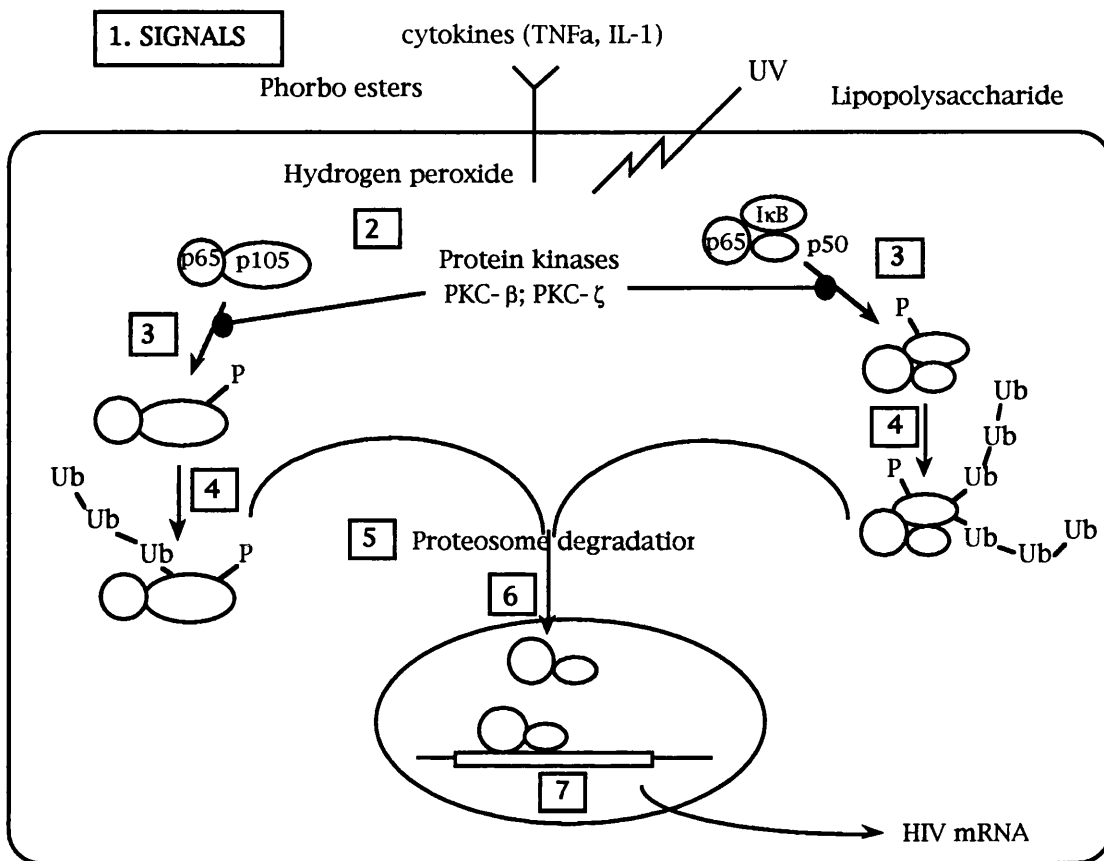
Key.

1. Signals shown to activate HIV-1 expression through NK-κB pathways. (reviewed by Verma et al. 1995). Hydrogen peroxide has a central role in the signal events, the effect of many activating agents can be inhibited by antioxidants (Baeuerle and Henkel 1994; Schreck et al 1991).

2. Activation of cellular kinases; several have been shown to phosphorylate IκB *in vitro* but isoforms of protein kinase C (β and ζ) are important *in vivo*. In addition, kinases specific to IκB isoforms may exist.

3. Phosphorylation of IκB and p105.

4. Modification of the phospho-proteins IkB and p105 by the addition of multiple ubiquitin residues.

5. Rapid proteolytic degradation of IkB and p105 in 26S proteosomes.

6. Release of NF-kB exposes the nuclear localisation signal and allows translocation into the nucleus.

7. Binding of transcriptionally active NF-κB to the two κB sites in the HIV-1 LTR and activation of transcription.

1.5.3.3   The NF-κB/rel family of transcription factors.

The NF-κB transcription factor consists of a family of related proteins that form homo- and hetero-dimers in the completed NF-κB protein (reviewed by Baeuerle and Henkel 1994; Siebenlist et al 1994; Verma et al. 1995). Highly purified NF-kB consists of two subunits, 50 kDa (p50 or NF-κB1) and 65 kDa (p65 or RelA) both of which show sequence homology to the *v-rel* oncogene of the reticuloendotheliosis virus and the *c-rel* proto-oncogene (Ghosh et al. 1990; Kieran et al. 1990; Ruben et al. 1991). These proteins show homology over a 300-amino acid region in the N termini, termed the rel homology region or domain (RHR or RHD) which includes the DNA-binding domains and the nuclear localisation signal, the probable site of interaction with IκB. In total five independent genes have been identified as encoding related proteins in the NF-κB family of transcription factors, these are NF-κB1 (p50), NF-κB2 (p52), RelA (p65), *c-rel* and *relB*. Another related protein, p49, has been described (Schmid et al. 1991) which is derived from alternative splicing of the p100 mRNA and is an isoform of NF-κB2. The polypeptides p50 and p52 contain DNA-binding domains but lack transactivation domains, thus although able to form homo-dimers these are not transcriptionally active and could potentially function as repressors by binding to and blocking κB sites. Only the related proteins RelA (p65) and *c-rel* have transactivation domains, thus the p50/p65 and p52/p65 heterodimers are transcriptionally active. The proteins p50 and p52 are synthesised as larger precursor polypeptides, p105 and p100 respectively which contain repeated units of 33 amino acids showing homology to the erythroid membrane protein ankyrin (Ghosh et al. 1990; Kieran et al. 1990; Nolan and Baltimore 1992). The p100 and p105 precursors cannot bind to DNA unless the ankyrin repeats are removed by proteolytic cleavage, a process that is signal dependent, requires ATP and is mediated by a ubiquitin-dependent proteosome degredation pathway (Mercurio et al. 1993; Palombella et al. 1994; Rice, Mackichan, and Israel 1992).

The regulation of transcription by NF-κB/rel transcription factors operates not only through the interaction of various homo- and heterodimers, precursor polypeptides and IκBs. There is evidence for the differential function of NF-κB factors mediated by the precise nucleotide sequence of the κB binding site possibly by inducing conformational change in the bound NF-κB dimer. The consensus binding sequence for NF-κB/Rel factors is 5'-GGGRNNYYCC-3', the two sites in the HIV-1 enhancer are identical to the κB site in the Igκ enhancer with the sequence 5'-GGGACTTTCC-3'. The different combinations of NF-κB proteins have different effects on HIV-1 activation. The p50/p65 heterodimer has been shown to be the predominant form involved HIV-1 proviral expression in T cells

(Fujita et al. 1992), in addition p52(p49)/p65 also transactivates the LTR (Schmid et al. 1991). The p49 isoform of NF-κB2 has been shown to act in concert with Tat to synergistically activate HIV-1 expression, an effect not shown by the p50 subunit (Liu et al. 1992).

Although the role of NF-κB activation in the stimulation of HIV-1 expression in 'resting' cells carrying an integrated provirus is well established, the importance of NF-κB in chronically infected cells, such as monocytes, dendritic cells and neuronal cells, is less clear. Because of the difficulty in working with primary isolates of these cell types most of the published results are from experiments with cell lines. Cell lines may have constitutively active NF-κB or respond aberrantly to stimuli. The isolation of monocytes (macrophages) and the processes involved in transfecting these cells may induce NF-κB. However there is evidence that NF-κB is important in the expression of HIV-1 by these cells. HIV infection of monocyte/macrophage cell lines can activate NF-κB (Bachelerie et al. 1991) possibly by an autocrine loop involving TNFα. The viral protease can directly cleave the p105 precursor of p50 and generate active NF-κB (Poli et al. 1990), a mechanism that could help to maintain NF-κB activation *in vivo*. Cell lines representing 'immature' monocytic cells (U937, HL-60) do not contain NF-κB in the active form; treatment of these cells with phorbol esters results in differentiation into a more mature phenotype and was associated with the appearance of active NF-κB and a parallel activation of LTR expression. Cell lines with a more 'mature' or differentiated phenotype (THP-1, PU5-1.8 and P388D1) contained constitutively active NF-κB and LTR-driven expression of a CAT reporter gene could not be induced by exposure to phorbol esters (Griffin et al. 1989). However, GM-CSF, which does not activate NF-κB, activates HIV expression in mature macrophages (with constitutively active NF-κB) indicating that additional, alternative cellular signalling pathways have been usurped by the virus (Koyanagi et al. 1988). In contrast, Suzan et al were unable to detect constitutively active NF-κB in primary uninfected isolates of human monocytes or in retinoic acid-treated U937 cells infected by HIV-1 although progeny virions were produced (Suzan et al. 1991). NF-κB was however induced in primary monocytes following infection with HIV-1. These apparently conflicting results may be explained by the utilisation by the virus of multiple activation pathways in host cells - in the 'immature' monocyte line U937, NF-κB is not active and not inducible by viral infection, LTR-driven expression appears to be regulated via alternative transcriptional pathways. Exposure of U937 cells to differentiating agents such as retinoic acid (RA) and in mature monoctes the phenotype of the cell allows NF-κB activation following viral

infection, possibly via cytokine signals. Infectious viral clones with either point mutations or deletions of the NF-κB sites replicate efficiently in T cells; however deletion of the NF-κB and Sp1 sites results in the complete loss of viral replication (Harrich et al. 1989; Leonard et al. 1989). Multiple HIV-1 isolates from an infected individual were found to have deletions of the NF-κB sites, an LTR-reporter gene construct produced inducible gene expression but not as high as with an intact LTR (Zhang et al. 1997). The isolate contained a duplication of another enhancer element, a TCF-1α binding site, which appeared to efficiently substitute for the loss of NF-κB function, supporting the concept of functional redundancy in the HIV-1 LTR.

1.5.3.4 Other cellular factors interacting with the NF-κB sites.

The nucleotide sequences which contain the two NF-κB binding sites have also been shown to interact with a number of other cellular transcription factors. These include PRDII-BF-1, also called MBP-1, a factor that has also been shown to bind to the NF-κB sites in the MHC class I, the interferon-β, and IL-2R enhancer elements (Baldwin et al. 1990; Fan and Maniatis 1990) although the function of PRDII-BF-1 in vivo is unclear (Muchardt et al. 1992). The ubiquitous transcription factor AP-2 binds to an element between and over-lapping the NF-κB sites (Perkins et al. 1994), a similar conjunction of κB and AP-2 sites has found in the promoter of the mouse histocompatibility gene H-2K^b. The significance of these potentially competitive κB binding factors has not been established. AP-2 binding may contribute to the basal expression of HIV but are displaced by NF-κB upon activation, possibly in cells expressing high levels of AP-2 competition with NF-κB could be more significant (Perkins et al. 1994).

Other cellular transcription factors shown to bind to sites overlapping or within the NF-κB sites include ets-1 (Bassuk et al 1997; Holzmeister et al. 1993) and EBP-1 (Wu et al. 1988a). Several inducible genes in T cells contain adjacent or overlapping ets and NF-κB sites. This suggests evolutionary conservation of the mechanisms regulating the induction of these genes which has been retained by HIV-1 and 2. The transcription factors also directly interact, through the 'ets' domain of the Ets protein and the C terminal region of the Rel homology domain of NF-κB and NFAT proteins (Bassuk et al 1997).

Two putative thyroid hormone response elements have been mapped to sequences -104 to -75 and -80 to -45, these correspond respectively to the NF-κB and Sp1 sites. Purified chicken thyroid hormone receptor (T3R-α1) homodimers and T3R-α1/RXR-β heterodimers bound to these sites and excluded NF-kB binding. These sites also functioned as thyroid hormone responsive elements in a heterologous promoter (Desai-Yajnik 1993a and 1993b).

1.5.5 Modulatory region: -107 to -440 and the negative regulatory element.

A large number of sequences distal to the enhancer/basal promoter of the HIV-1 LTR have been described and are the subject of several review articles (Antoni et al 1993; Cullen 1991; Gaynor 1992; Greene 1990; Kingsman and Kingsman 1996; Vaishnav and Wong-Staal 1991). The term 'negative regulatory element' (NRE) has been used for sequences upstream of -155, although the precise limits for the NRE vary. Deletion of these sequences or elements within this region have been shown to result in an increase in the basal level of transcription and increased activation by mitogens. Infectious molecular clones of HIV with NRE deletions also replicated to higher levels. These results led to the conclusion that regulatory elements were present within the LTR upstream of -155 that down-regulated LTR expression and the term 'negative regulatory element' was adopted. However this is a misleading simplification, the LTR in this region has been shown to contain sequences that can interact with a large number of cellular factors exerting both positive and negative effect on LTR expression depending upon the activation state of the cell, the type of cell use for analysis and the particular HIV-1 LTR used. A more appropriate term would be 'modulatory region'.

1.5.5.1
Transcription factor binding sites in the modulatory region

Prior to the publication of the results described in this thesis, no binding sites for cellular factors had been delineated in the modulatory region upstream of the enhancer, LTR-deletion analysis had identified the 'NRE' in broad terms only (Rosen, Sodroski, and Haseltine 1985). At the time of preparing this thesis at least ten elements have been identified in the modulatory region and 15-20 cellular proteins shown to interact with these sites (reviewed by Antoni et al 1993; Kingsman and Kingsman 1996). The apparent conservation of these sites in a virus with a high rate of mutation and complex replication cycle indicates the important functional significance of the region to the virus. The number of sites, the evidence for multiple interactions with individual sites and the potential for interaction of factors binding to adjacent, overlapping or distant sites makes analysis of the region difficult. In addition, factors binding to elements in the modulatory region may recruit transcriptional co-factors which may be present in certain cell types infected by HIV-1. Figure 1.5.6 summarises the sites that have been delineated in this region of the LTR and the cellular factors reported as binding to them.

Figure 1.5.6

Binding sites for cellular transcription factors identified in the 5' LTR of HIV-1

site A   site B/HRRE

```
                                    ⬭  ⬭
  ┌──────────┬─┬─┬──⬭──┬─⬭⬭──────────┐
  └──────────┴─┴─┴─────┴──────────────┘
-450        -400      -350        -300
            GATA 3    GATA 3

                      NF-I    COUP-TF
                      p43     RARs
                              RAR/RXR
                              several nuclear hormone
                              receptors
```

```
      NFAT-1            USF-1  TCF-1    NF-κB x2   Sp1 x3  TFIID

  ┌─┬──⬭⬭──────┬──────⬭──⬭⬭──┬─⬭⬭⬭──⬭⬭⬭──⬭──→
  └─┴──────────┴───────────────┴───────────────
    -250      -200    -150      -100      -50       +1
GATA 3    AP-1        GATA 3    GATA 3

                      myc       PRDII-BF1

                      70 and 110 kDa   ets
                      proteins

                      sp50      EBP-1

                      HIV-TF1   T3R      T3R

                      NF-IL6
```
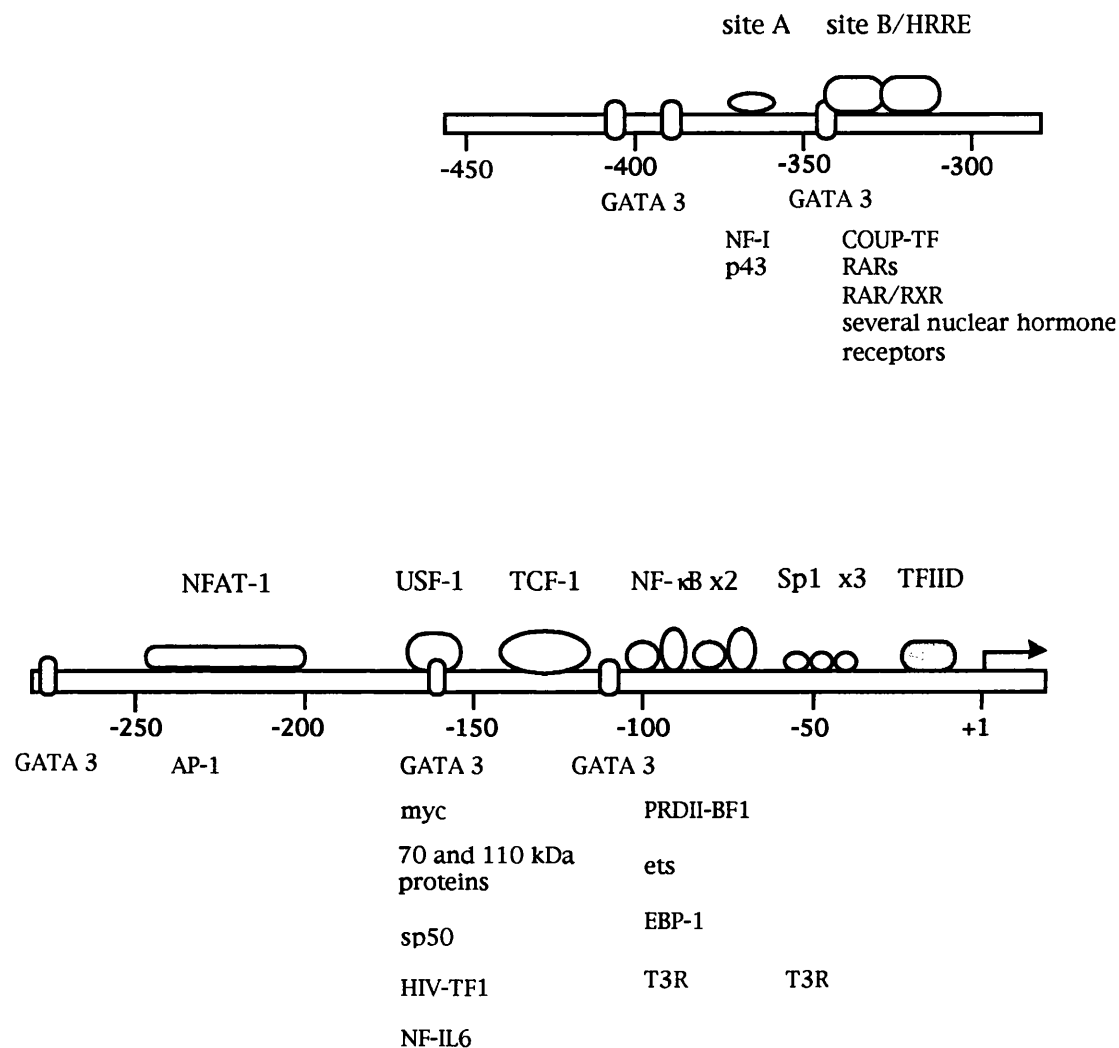
References:

TCF-1 (Golub et al 1990; Waterman et al 1991; Waterman and Jones 1990)

USF-1 (Garcia et al. 1987; Smith and Greene 1989); myc (Murre et al 1989)

70 and 110 kDa proteins (Giacca et al. 1992)

sp50 (Smith and Greene 1989)

HIV-TF1 (Maekawa et al. 1991)

NFAT-1 (Crabtree 1989; Shaw et al. 1988; Li et al 1991)

GATA 3 (Yang and Engel 1993); NF-IL6; (Tesmer et al. 1993)

Ets 1 and 2 (Seth et al. 1993)

site B protein (Orchard et al. 1990)

RAR and RXA (Orchard et al. 1993)

COUP-TF (Cooney et al. 1991)

RAR, RXR, multiple nuclear hormone receptors (Ladias 1994)

NF-1 and p43 (Schwartz et al. 1997)

Details of binding sites in the modulatory region.

1. The upstream regulatory element (URE) -121 to -150.

This region was identified by deletion analysis as a negative control element; in Jurkat cells deletion of this element increased LTR-driven expression while in U937 a small reduction was seen (Nakanishi et al 1991).

A 55 kDa cellular protein called TCF-1α binds to sequences -123 to -138, a T-lymphocyte protein involved in the regulation of the T-cell receptor Cα gene expression where it acts as a positive activator. There is no direct evidence of a functional role for TCF-1α in HIV expression but duplication of this site has been found in primary isolates of HIV-1 from patients and these isolates showed increased replication in culture compared with wild-type virus (Golub et al 1990).

2. USF-1 binding site; -159 to -173.

The sequence CACTGT in this element binds helix-loop-helix transcription factors including c-myc and a 43 kDa protein termed USF-1. The site has been shown to have a negative regulatory effect on LTR expression and was originally termed 'the negative regulatory factor binding site' (Garcia et al. 1987; Lu et al. 1989). Several factors have been shown to bind to sequences in this site, including proteins of 70 and 110 kDa (Giacca et al. 1992) and a 50 kDa T cell factor termed sp50 which also interacts with a putative negative regulatory element in the IL-2R promoter (Smith and Greene 1989) and a 39 kDa phosphoprotein called HIV-TF1 (Maekawa et al. 1991).

This element appears to have both negative and positive effects on LTR expression, linker-scanning mutation of the site reduced LTR activity (Zeichner et al 1991), deletion of the site in an infectious viral molecular clone resulted in higher levels of HIV progeny virus production and multimerisation of the site aused decreased expression from a heterologous promoter construct (Giacca et al. 1992; Lu et al. 1990).

A transcription factor involved in the activation of IL-6 responsive genes, NF-IL6, has been shown to bind to the USF-1 element. NF-IL6 increases expression from promoters containing the binding site and may mediate the IL-6-induced expression of HIV in T cells (Tesmer et al. 1993).

3. NFAT-1 binding site; -216 to -254.

NFAT-1 (Nuclear Factor of Activated T cells) is a transcription factor initially identified as an activator of IL-2 gene expression and interacts with specific sequences in the LTR (Shaw et al. 1988). NFAT-1, like NF-κB, is an inducible transcription factor, stimulation of T cells by phorbol esters and lectins stimulate NFAT-1 binding to its response elements, but unlike NF-κB, NFAT-1

activation is blocked by cyclosporin A indicating alternative signal pathways are utilised, possibly calcineurin-mediated. The role of NFAT-1 in the activation of HIV-1 expression is not clear but it seems probable that it acts in concert with NF-κB to activate LTR-activity. Another cellular factor, ILF-1, has been shown to bind to this site and may function as a repressor of basal LTR expression (Li et al 1991). Linker-scanning mutation across the NFAT-1 site caused a modest increase of basal LTR expression and a small decrease in expression after activation by phorbol ester and PHA (Zeichner et al 1991).

4. The region -260 to -450.

Binding sites for several cellular factors have been delineated in this region of the LTR. An early report described the presence of two potential binding sites for the transcription factor AP-1 (Franza et al. 1988) but the sequences identified were not shown to interact with AP-1 in EMSA or DNA-footprint assays. The results could indicate a possible indirect binding of *fos/jun* to this site, possibly by protein-protein interactions with factors occupying the site. Further evidence for the lack of direct AP-1 interaction with sequences in this region will be presented in chapter 3. Two strong DNAse I footprints were obtained with Jurkat nuclear extracts extending from -323 to -356 (site B) and -365 to -380 (site A); the characterisation of these sites and the cellular factors interacting with site B forms the subject of this thesis.

5. GATA-3 binding sites.

Six binding sites for the T cell specific zinc-finger factor GATA-3 have been described, located by DNAse-footprint analysis of the LTR using partially purified human GATA-3 (Yang and Engel 1993). The position of these sites are shown in figure 1.5.6. GATA 3 is a member of a multigene family of transcription factors which bind to similar WGATAR sequences and includes the erythroid lineage specific factor GATA-1. GATA-3 is abundantly expressed in T cells and is also present in a sub-set of neuronal cells. The sites could be shown to have functional activity, co-transfection of a GATA-3 expression vector and an LTR-reporter construct produced a 7-fold increase in LTR-expression while mutation of all six sites reduced GATA 3 induced expression. In T cells basal and stimulated expression from the mutated LTR were reduced.

1.5.6 Role of chromatin structure and HIV-1 expression.

1.5.6.1: Chromatin structure and nucleosome binding.

Chromatin organisation plays a crucial role in the regulation of expression of eukaryotic genes (reviewed by Felsenfeld 1992; Kadonaga 1998; Latchman 1995). Gene expression in eukaryotes is associated with regions of 'open' chromatin which are relatively sensitive to digestion by pancreatic DNase I. Packaging of nuclear DNA into increasingly complex structures associated with chromatin has been shown to prevent the interaction of transcription factors to their specific binding sites (Hayes and Wolffe 1992) and the binding of a single nucleosome can affect the function of such sites (Simpson 1990). Despite the importance of chromatin structure in the control of eukaryote gene expression, relatively little attention has been given to the role of chromatin in the expression of the integrated HIV-1 provirus. Once integrated into the host genome the proviral DNA will be incorporated into the structural organisation of nuclear DNA starting with nucleosome binding and leading to higher levels of packaging. The degree of organisation of the proviral genome will potentially depend upon the region of the host's DNA that integration occurs. Integration of the HIV-1 provirus following reverse transcription is a random event but is more likely to occur at sites of open, i.e. actively transcribed, DNA. If the region of DNA at which integration occurs remains in the 'open' state then the proviral 5'LTR will be accessible to the transcriptional apparatus and cellular trans-acting factors required for the expression of the early viral mRNA species. If however the host cell becomes quiescent and the provirus has integrated in a region of the host's genome which is no longer transcribed, the proviral genome may become transcriptionally silent. HIV-1 may have evolved a mechanism to reduce this possibility, there is evidence for the acetylation of histone proteins by tat, either directly or by an indirect mechanism involving the binding of a histone acetytransferase (Tip60) to tat (Yamamoto and Horikoshi 1997). Acetylation of histone proteins (in particular H3 and H4) is associated with areas of active ('open') chromatin, the exact mechanisms are not clear but acetylation may affect the association of nucleosomes into larger units or with DNA allowing access of transcription factors. Acetylation of histone proteins increases the binding of basal transcriptional proteins in a Xenopus model (Latchman 1995). By increasing the acetylation of histone proteins associated with the 5'LTR the region may remain in an 'open' state and allow either continued viral mRNA expression or perhaps reduce the 'lag' phase following reactivation of the host cell favouring the early expression of the viral genome over cellular genes. The importance of histone acetylation on HIV-1

expression has been studied using cell lines carrying integrated HIV-1 provirus with little or no viral gene expression in the unstimulated state ('latently infected'). Treatment of these cells with specific inhibitors of deacetylation (trapoxin and trichostatin) caused the global hyperacetylation of cellular histones and resulted in the activation of HIV-1 expression (van Lint et al. 1996). However activation of HIV-1 expression was independent of NF-KB and the same agents had a minimal effect upon a transfected HIV-1 LTR-reporter gene construct. DNase I mapping of the integrated HIV-1 genome in the 'latently infected' cell lines U1 (derived from HIV-1 infected monocytic cell line U937) and ACH2 (derived from CEM T-cell line) revealed three DNase I hypersensitive sites in the 5'LTR, two major sites corresponding to the enhancer, basal promoter and the TAR region. A minor DNase I hypersensitive site was present in the NRE, corresponding to the region containing site B, suggesting that this region of the integrated proviral LTR was open and nucleosome-free (Verdin 1991). Further sites were present downstream of the transcriptional start site and may correspond to downstream regulatory elements (van Lint et al. 1994 and 1997). Mapping with micrococcal nuclease in the same cell lines revealed a consistent nucleosome pattern over the 5'LTR (Verdin et al 1993). A single nucleosome (nuc-0) was present in the U3 region (nucleotides -420 to -260) with the rest of the U3 region open and corresponding to the major Dnase I hypersensitive sites containing the enhancer and basal promoter. A further nucleosome (nuc-1) mapped to nucleotides +11 to +166, a large linker region of 124 nucleotides separated nuc-1 from the next nucleosome (nuc-2) followed by regularly spaced nucleosomes nuc-3 and nuc-4. On activation of the cells with phorbol ester or TNF-α, nuc-1 association with the LTR was rapidly disrupted while nuc-0, nuc-2, -3 and-4 remained associated with the LTR. They proposed a transcription activation model based on the nucleosome positions and interactions between upstream and downstream regulatory elements (Verdin et al 1993).

1.5.6.2: CpG methylation.

Methylation of cytosine residues to form 5'-methylcytosine has been shown to modulate gene expression in both prokaryotes and eukaryotes. Unmethylated CpG sites are associated with actively transcribed genes while methylation of CpG sites in promoter sequences abolishes transcription. Two CpG sites have been described in the 5'LTR of HIV-1, at nucleotides -218 and -146. Enzymatic methylation of these sites abolished LTR-driven expression of tat in a transient transfection assay using a cell line carrying a 'latent' proviral genome (Bednarik et al 1990). This methylation-sensitive block could be overcome by stimulation of the cells with phorbol ester. Methylation at these sites also

abolished the binding of an uncharacterised nuclear factor from a human adenocarcinoma cell line (SW480) while nuclear extracts prepared from SW480 cells constitutively expressing tat appeared to restore binding despite methylation.

The evidence provided for the importance of both chromatin structure and CpG methylation is important and potentially adds further levels of complexity to the regulation of HIV-1 expression. However the experiments were performed with cell lines with specific characteristics, in particular the presence of an integrated viral genome with little or no basal LTR-driven transcription. The cell lines were derived from parental HIV infected cell lines and selected for the low level of viral expression and replication (Clouse et al. 1989; Folks et al. 1988). These derivative lines have been studied extensively as convenient models of viral latency but are unlikely to provide a true representation of the complex, dynamic interaction of virus and host *in vivo*. The low level of viral expression in the derivative cell line U1 has been shown to be due to different mutations in the tat genes of the two integrated proviral genomes harboured by this line (Emiliani et al. 1998).

1.5.6.3 The nuclear matrix and HIV-1 expression.

The nuclear matrix is an important part of the architecture of the nucleus, forming the internal scaffold which not only creates the physical structure but also has important functions including modulation of gene expression (reviewed by Davie 1997). The nuclear matrix is important in the higher-order organisation of chromatin and for the tethering of regions containing actively transcribed genes via matrix-associated regions (MARs) or scaffold-attached regions (SARs). The nuclear matrix may also serve as a repository for cellular transcription factors including steroid/thyroid hormone receptors. There have been no published reports describing the role of the nuclear matrix and the higher orders of chromatin structure on HIV-1 expression. Although the role of the nuclear matrix in the modulation of HIV-1 expression has not been directly investigated a tenuous connection however does exist. The cellular nuclear protein YY1 has been shown to bind to the HIV-1 LTR in the initiator region (-17 to +27) and the expression of YY1 inhibited HIV-1 LTR expression and viral production (Margolis et al 1994). The binding of YY1 to this region of the LTR involves the formation of a complex with LSF (LBP-1/UBP/CP-2) (Romerio et al 1997). The protein YY1 has been shown to be identical to a nuclear matrix protein NMP-1 (Guo et al. 1995) suggesting that at least this region of the LTR can interact with the nuclear matrix.

Chapter 2

Materials and Methods.

Chapter 2

Materials and Methods.

2.1 Preparation and handling of Nucleic acids.
2.1.1 General procedures and buffers.

All solutions were prepared with HPLC-grade water, autoclaved (120°C, 15 minutes) or, if thermolabile, filtered through a 0.22 micron filter. Solutions for RNA were treated with diethylpyrocarbonate (DEPC) as a 0.1% v/v solution overnight at room temperature, then autoclaved to destroy residual DEPC. All plastic tubes and pipette tips were autoclaved.

Buffers:

TE;	10mM Tris-HCl pH 8.0, 1mM EDTA.	
TNE;	TE with 100mM NaCl	
TBE;	1 litre of x5 concentration prepared with 54g Tris base, 27.5g boric acid and 4.6g EDTA. Final pH 8.2	
TAE;	1 litre of x50 concentration with 242g Tris base, 57.1ml glacial acetic acid, 100ml 0.5M EDTA.	
SSC;	1 litre of x20 solution prepared from 175.3g NaCl, 88.2g sodium citrate, pH adjusted to 7.0.	

Organic solvents: phenol was obtained as a water-saturated liquid, equilibrated with TE, stored containing 100mM ß-mercaptoetanol and 0.1% w/v 8-hydroxyquinolone under a layer of TE at 4°C in the dark. Chloroform was prepared as a 25:1 v/v mixture with isoamyl alcohol.

2.1.2 Restriction endonuclease digestion of DNA.

Plasmid DNA was digested with 5-10 units of enzyme per 1 μg DNA in a final volume of 20-30μl for 2-6 hours in buffer supplied by the manufacturer.

2.1.3 Ligation of DNA fragments into vectors.

Fragments and vectors were combined at a molar ratio of approximately 3:1 in 66mM Tris-HCl pH 7.5, 5mM $MgCl_2$, 5mM DTT, 1mM ATP, 0.1 micrograms of BSA and 1 unit of T4 DNA ligase (Boehringer Mannheim) in a final volume of 10 μl, then incubated overnight at 4°C. For optimal ligation the final concentration of 5' ends was between 0.1-1.0μM.

2.1.4 Removal of 5' terminal phosphate from linearized vectors.

Self-ligation of vector was reduced by dephosphorylation of 5' ends using calf intestinal phosphatase (CIP). Following restriction enzyme digestion, a 1/10 volume of CIP buffer (0.5M Tris-HCl pH 9.0, 10mM $MgCl_2$, 1mM $ZnCl_2$, 10mM

spermidine) was added together with 10 units of CIP (Boehringer Mannheim) and incubated at 37°C for 20 minutes. The enzyme was inactivated by heating at 68°C for 15 minutes in 0.5% SDS, followed by phenol/chloroform extraction and ethanol precipitation of DNA.

2.1.5 Preparation of transformation-competent bacteria.

E. coli strain TG1 (Amersham) was maintained on minimal salts medium agar plates. A single colony was picked into L-broth (10g bacto-tryptone, 5g yeast extract, 10g NaCl per litre) and cultured overnight at 37°C with aeration. 0.5ml of the fresh culture was innoculated into 50ml of prewarmed L-broth, supplemented with 10 mM MgSO4 and 0.2% glucose (LBMG), and incubated until the optical density reached 0.3-0.8 units. The cells were chilled on ice for 15 minutes, pelleted (1500g 4°C 10 minutes), gently resuspended in 15ml ice-cold TFB -10mM MES (2-[N-morpholino]-ethanesulphonic acid) pH 6.3, 45mM MnCl2, 10mM CaCl2, 100mM KCl, 3mM hexamminecobalt chloride- and incubated on ice for 15 minutes. The cells were pelleted (1500g 4°C 10 minutes), resuspended in 2ml ice-cold TFB, 70ml of dimethyl formamide added and incubated on ice for 5 minutes. 70µl of 2.25M DTT was added and the cells incubated on ice for 10 minutes. Cells remained competent for 4 hours.

For long-term storage, cells were grown in LBMG to mid-logarithmic phase (OD 0.3-0.8 units), chilled on ice for 10 minutes then pelleted as above. Cells were resuspended in 0.5 ml of ice-cold LBMG and 2.5 ml of storage solution slowly added (36% glycerol, 12% PEG 75, 12mM MgSO4, in LB). Cell suspensions were snap frozen in 0.1ml aliquots in liquid nitrogen and stored in liquid nitrogen vapour phase.

2.1.6 Transformation of competent cells with plasmid DNA.

5ml of ligation reaction was added to 100ml of competent bacteria on ice and incubated for 20 minutes. Cells were heat-shocked for 90 seconds, cooled on ice for 1 minute then 1ml of warmed L-broth added and the cells grown for 1 hour to allow expression of antibiotic resistance. 50µl aliquots of the culture were plated onto nutrient agar plates containing the appropriate antibiotic and incubated over-night. Single colonies were picked off into 5ml of nutrieent broth using a sterile tooth pick and grown over-night. Plasmid preps were made from the over-night culture ('mini-preps') and screened for the cloned insert using either PCR or restriction endonuclease mapping.

2.1.7 Preparation of plasmid DNA.

Solutions:

A: 50mM glucose

10mM EDTA

25mM Tris pH 8.0

lysozyme 5mg/ml

(added immediately before use)

B: 1% SDS

0.2M NaOH

C: 5M KAcetate pH 4.8, 60 ml

acetic acid 11.5 ml

distilled H_2O 28.5 ml

2.1.8 Small scale preparation ('mini-prep').

A single bacterial colony was transferred into 5ml of 'terrific broth' (20g bacto-tryptone, 10g yeast extract, 10g NaCl) containing the appropriate antibiotic and cultured overnight at 37°C with vigorous shaking. Bacteria were harvested from 1ml in a 1.5ml eppendorf tube (1200g, 3 minutes at RT), resuspended in 100µl solution A and incubated for 5 minutes at room temperature. Cells were lysed by adding 200µl freshly prepared solution B with gentle mixing, incubated on ice for 5 minutes. Proteins and chromosomal DNA were precipitated with 375 µl ice-cold solution C and pelleted by centrifugation (1200g, 5 minutes at 4°C). The supernatant was heat treated at 65°C for 10 minutes to inactivate DNases then extracted once with phenol/chloroform. Plasmid DNA was precipitated by adding an equal volume of isopropanol, pelleted by centrfugation (1200g, 10 minutes at RT), washed twice with 70% ethanol, air dried and redissolved in 50µl of TE. The plasmid preparation was treated with DNase-free RNase A (Boehringer), 5µl of 10mg/ml, a 37°C incubation for 15 minutes then extracted once with phenol, then phenol/chloroform and finally chloroform. Plasmid DNA was precipitated with x2 volume of cold ethanol and pelleted by centrfugation. The DNA pellet was washed twice with ice-cold 70% ethanol and finally dissolved in 25µl of TE.

2.1.9 Large scale preparation ('maxi-prep').

Single colony of E. coli was innoculated into 250ml of 'terrific broth' containing an appropriate antibiotic and cultured overnight with vigorous aeration. Cells were pelleted (5000g for 30 minutes at 4°C), resuspended in 8ml of solution A containing 5µg/ml of lysozyme and incubated for 10 minutes at room temperature. To this suspension 16ml of ice-cold solution B was slowly added with gentle mixing, then left on ice for 5 minutes after which 12ml of ice-cold solution C was added and the mixture vortexed briefly. Cell debris was pelleted (9000g 10 minutes at 4°C) and the supernatant carefully removed avoiding the overlying white material. The supernatant was heated at 65°C for 10 minutes to inactivate DNases, the volume of the supernatant recorded and a

0.6x volume of ice-cold isopropanol added to precipitate nucleic acids. Nucleic acids were pelleted (12000 rpm on High Speed MSE for 10 minutes at RT), the pellet resuspended in 1ml of TE and proteins extracted with chloroform, chloroform-phenol and finally phenol. RNA was precipitated with 2 ml of 5M LiCl (kept at -20°C) per 1 ml of solution and kept on ice for 5-10 minutes. Precipitated RNA was pelleted (10000g HS MSE 10 minutes 4°C) the supernatant removed and the plasmid DNA precipitated with a 2x volume of -20°C ethanol and pelleted (12000 rpm HS MSE 10 minutes 4°C), the pellet washed twice with cold 70% ethanol and dissolved in 500ml TE. The DNA was further purified by banding on a caesium chloride gradient under standard conditions.

2.1.10   Preparation of LTR fragments.

The LTR of HIV-1 strain HXB2 (a generous gift from G. Nabel) was digested with Kpn I and Rsa I generating the following fragments.

Fragment	Size (bp)	Comments
I	220 (-306 to -460)	comprising 150 bp of U3 and 70bp of *nef* coding sequences. Contained the NRE.
II	165 (-140 to -305)	
III	135 (-5 to -139)	containing TATA box, SPI and NFkB sites.

These fragments were subcloned into the pUC1813 vector. For the experimental work described in this thesis only the '220' fragment was used as this contained virtually the entire NRE as defined in both resting and activated T-lymphocytes (Siekevitz et al. 1987).

2.2 DNA Mobility Shift Assay.

(electrophoretic mobility shift assay, EMSA).

2.2.1 Preparation of nuclear extracts.

(modified from Dignam, Lebovitz, and Roeder 1983).

Buffer A: 10mM Hepes pH 7.9   Buffer C: 20mM Hepes pH 7.9

          1.0mM $MgCl_2$ 1.5mM $MgCl_2$

          60mM KCl   0.5mM EDTA

          10mM NaCl   0.5mM DTT

          0.5mM DTT   20% glycerol

          0.5mM EDTA

Buffer D: 20mM Hepes pH 7.9

          100mM KCl

          0.25mM EDTA

          0.5mM DTT

          20% glycerol

Protease inhibitors:

stock solns. per ml.		x100 soln. per ml	final concs. per ml
antipain	1mg in water	0.1mg	1.0µg
elastinal	1mg "	0.1mg	1.0µg
leupeptin	2mg "	0.2mg	2.0µg
aprotinin	5mg "	0.01mg	0.1µg
pepstatin A	10mg in methanol	0.5mg	5.0µg

The x100 soln. was prepared with all protease inhibitors combined in water

and stored frozen in small aliquots at -20°C.

In addition:

stock solns.

PMSF	100mM in ethanol	-	1mM
TCLK	10mg in water	-	0.1µg

All protease inhibitors were obtained from Sigma.

Cells were cultured as described in section 2.4.1. Suspension cells were harvested by centrifugation at 600g for 10 minutes, washed once with PBS at room temperature and the volume of the cell pellet determined (Vp). All buffers from this stage contained protease inhibitors and care was taken to maintain the temperature between 0-4°C. The washed cells were resuspended in a volume of ice-cold buffer A equal to 5x Vp and allowed to stand on ice for 10 minutes. The cells were again pelleted at 600g and then resuspended in 3x

Vp of fresh buffer A. Lysis of the cytoplasmic membrane was achieved by homogenisation in a 2ml Dounce glass homogeniser, usually 10-20 strokes with rod B was sufficient to obtain 90% lysis. Cell lysis was determined by comparison of homogenised material with a sample taken before treatment, free nuclei appearing as smaller, darker spheres lacking a phase-bright halo when viewed in the counting chamber of a haemocytometer. The lysate was centrifuged at 600g for 10 minutes to pellet the nuclei, the supernatant removed and kept on ice (cytoplasmic fraction, Cf) and the nuclear pellet resuspended in an equal volume of buffer C. The total volume containing the nuclei was accurately determined and 3M NaCl added dropwise to give a final concentration of 0.32M. The nuclear suspension was kept on ice for 30 minutes with occasional gentle mixing, then centrifuged at 50,000g (17,000 rpm) on a Beckman bench-top ultracentrifuge (TL-100) with the TLS-55 rotor to pellet the chromatin. The nuclear extract (Ne) and cytoplasmic fraction were then dialysed against buffer D at 4°C for 8hr then centrifuged at 50,000g to remove the protein precipitate. Aliquots of the resulting cleared extracts were snap frozen in liquid nitrogen and stored at -70°C. Extracts remained active in EMSA for at least 2 years.

Adherent cells were washed twice with phosphate buffered saline on the culture dish, then scraped into ice-cold PBS with a rubber policeman. The cells were centrifuged at 600g then treated as for suspension cells.

The protocol above was suitable for the preparation of Ne from most cell types used, however the following modifications were found necessary for specific cell lines: Jurkat required 0.6M sucrose in the buffer A to lyse the cytoplasmic membrane and to prevent premature lysis of nuclei. Hela cells were very resistant to mechanical lysis and required the addition of 0.05% Triton X-100 to buffer A. Lysis of HL60 cells required only 2-3 strokes of the homogeniser rod.

2.2.2 Binding reaction.

Nuclear extracts (1μl) were incubated in a binding buffer containing 4% Ficol (Sigma mol. wt. 400,000), 10mM Tris-HCl pH 7.8, 2mM $MgCl_2$, 50 mM KCl, 1 mM DTT, 3μg poly(dI:dC),poly(dI:dC) and 2μg poly(dA:dT),poly(dA:dT) (both from Pharmacia), in a final volume of 20μl, for 15' on ice. Subsequently, 10fmol (-10,000 dpm) of labelled oligonucleotide were added, (together with an unlabelled competitor oligonucleotide if appropriate) and the binding reaction continued for a further 20 minutes on ice. The reaction was loaded on a pre-run, pre-cooled 4.5% polyacrylamide gel with 0.25 x TBE as running buffer.

Gels were run at 4°C, 10V per cm, for 2.5 hrs., then dried and exposed to pre-flashed X-ray film (Kodak X-OMAT).

2.2.3 Acrylamide gel mix.

100ml for a 18 x 15 x 0.3 cm gel

29% acrylamide: 1% bis-acrylamide	14 ml
x5 TBE	5ml
HPLC-grade water or equivalent	81ml
TEMED	45μl
ammonium persulphate (APS)10%	450μl
running buffer	0.25% TBE

The components were mixed together, except the APS, de-gassed for 2 minutes, then the APS added. The gel mix was poured and allowed to polymerise for 1 hour.

It was found to be important to pre-run the gel for at least 90 minutes and change the buffer for fresh 0.25 % TBE before loading and running the assay.

2.2.4 Preparation of oligonucleotides.

Oligonucleotides used in all GMSAs were deprotected and HPLC purified by the manufacturer (Oswel DNA Services, Department of Chemistry, University of Edinburgh). Oligonucleotides were designed to generate $SalI/XhoI$ compatible overhangs when annealed.

The concentration of each single stranded oligonucleotide was determined from the OD_{260} and applying the formula,

$$\text{conc. mmol/ml} = \frac{OD_{260}}{S\!Œ}$$

where SŒ represents the sum of the absorption coefficients (Œ) of each base in the oligonucleotide (adenine=15.4; thymine=8.8; cytosine=7.3; guanine=11.7)

2.2.5 Annealing oligonucleotides.

Equimolar concentrations of complementary oligonucleotides were heated to 95°C in 20mM Tris-HCl pH 7.8; 50mM NaCl, cooled for 2 hr. to 60°C, then to room temperature overnight. The concentration of the double stranded oligonucleotide was determined by measuring the OD_{260} and applying the formula above.

2.2.6 Creation of oligonucleotides multimers.

Oligonucleotides were annealed as described. Multimerisation was performed in seven stages:

1) Addition of 5' terminal phosphate (kinase reaction).

2) Ligation.

3) Partial digestion with *Xho* 1 and *Sal* 1 restriction endonucleases.

4) Agar gel electrophoresis and staining with ethidium bromide to visualise bands. Band(s) of choice representing the appropriate multimer of the original oligonucleotide were identified by size and excised from the agar. DNA extracted and purified.

5) Multimer cloned via *Sal* 1 site into reporter plasmid pSP65.tk.CAT.

6) Transfected bacteria screened for insert of correct size.

7) Clones containing insert sequenced to confirm number and orientation of oligonucleotides in multimer.

2.2.7 Labelling oligonucleotide probes.

Probes were prepared by filling-in the 5'-overhang with AMV reverse transcriptase (Pharmacia) in RT buffer: 50mM Tris-HCl pH 7.6; 60mM KCl; 10mM $MgCl_2$; 1mM of each of dATP, dTTP, dGTP; 20pmol (100 μCi) of ^{32}P-dCTP; 1mM DTT;

1-2 pmol of oligonucleotide. The reaction was incubated for 45 minutes at 37°C, then stopped by adding 70μl TE. Labelled probe was separated from unincorporated ^{32}P-dCTP by chromatography on a 1ml G-50 Sephadex (Pharmacia) column using centrifugation. The efficiency of incorporation of radiolabel was determined on an aliquot of the reaction taken before separation.

2.2.8 UV cross-linking.

Oligonucleotides were constructed with 5-bromo-deoxyuridine (5BrdU) substitutions for specific thymidine positions as indicated:

```
          *    * *          *
5'-TCGACCAGGGGUCAGAUAUCCACTGACCUTTC
   GGUCCCCAGTCUATAGGUGACUGGAAAGAGCT- 5'
     *              *     *    *
```
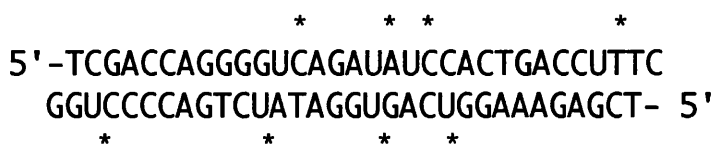
The oligonucleotide was labelled using AMV-RTase to fill-in the 5'-overhang and a preliminary GMRA performed, with appropriate competition controls, to confirm that the substituted oligonucleotide was shifted and competed in a manner identical to the non-substituted oligonucleotide. For competition a 100-fold excess of unlabelled site B or an oligonucleotide of unrelated sequence was used.

A binding reaction was prepared as described previously (section 2.2.2), but the volume of reaction increased to 50µl with 4µl of Ne (approximately 30µg of total protein) and 50 fmol of probe. After 10 minutes incubation on ice with the labelled probe the binding reaction was transfered to a round-bottomed well in a plastic microtitre dish, placed on an ice-tray, the tray covered with Saran-wrap and then exposed to ultra-violet irradiation (310nm) using an inverted transilluminator as shown in the diagram below.
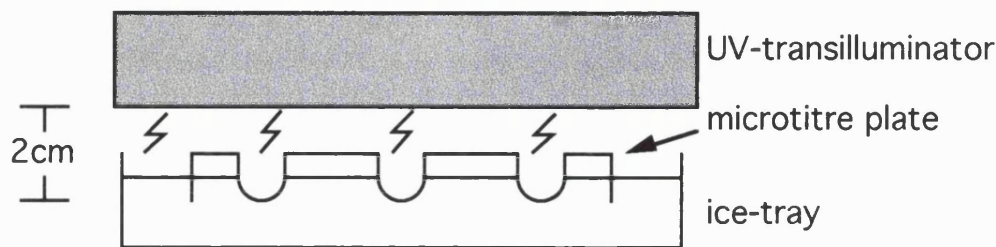
The procedure was performed at 4°C and care was taken to ensure that the microtitre plate wells containing the binding reactions remained in close contact with the ice as considerable heating occurred during the period of exposure.

Following irradiation for variable periods of time, the binding reaction was then treated in one of two ways:

1. Direct method: An equal volume of x2 SDS loading buffer was added, heated to 100°C for 5 minutes then loaded onto an 8% polyacylamide gel containing 0.1% SDS. ^{14}C-labelled protein molecular weight markers (Amersham, Herts.) were also loaded. Details of gel preparation and electrophoresis were as described previously. Following electrophoresis, gels were dried and exposed to X-ray film as before.

2. Indirect method: The binding reaction was loaded onto a 4.5% non-denaturing polyacrylamide gel and an EMSA performed as described in section 2.2.2. Following electrophoresis, one of the gel plates was removed, the gel and remaining plate wrapped in Saran-wrap, then exposed to X-ray film for 4-6 hours. Bands corresponding to specific protein/probe binding were localised on the wet gel and excised in a small gel slice. Gel slices were heated to 100°C in x2 SDS loading buffer for 10 minutes, then loaded into a slot in an 8% polyacrilamide gel containing 2% SDS gel, taking care to maintain the correct orientation of the slice. Gels were run with radiolabelled protein molecular weight markers and analysed in parallel on the gel. The gel was then dried and exposed to X-ray film.

120

Optimal exposure time for cross-lining was determined in a series of timed exposures under controlled conditions. The optimal time was found to be from 60-90 minutes, crossed-linked DNA-protein complexes increasing with increased exposure time to a maximum yield then decreasing, presumably due to local heating and protein denaturation within the microplate wells.

2.3 Ferguson plot.

The technique was modified from that described in the Sigma technical bulletin number MKR-137, which in turn has been modified from the methods originally described by Bryan (1977).

The combination of the Ferguson plot analysis and EMSA provides a powerful method for determining the molecular weight of specific protein/DNA complexes identified on the gel. Thus, complexes that are formed by the non-covalent interaction of two or more proteins may be analysed when associated with the specific DNA sequence. Such information when compared with a molecular weight determined on a denaturing gel (for example after purification or cross-linking) is of obvious importance in determining whether the protein binds to DNA as a monomer or dimer, as a heterodimer, or if other proteins are involved that do not themselves directly bind to DNA.

2.3.1 Principle of Ferguson plot.

The relative electrophoretic mobility of a given protein during non-denaturing-PAGE is determined by several factors, such as the net charge of the protein under the conditions of pH and ionic strength of the buffer used, the conformation, the molecular size, the concentration of polyacryamide and degree of cross-linking in the gel.

The mobility of a protein relative to the tracking dye (Rf) may be determined on a set of gels of increasing polyacryamide concentrations. A graph of 100xlog.(Rfx100) versus gel concentration produces a straight line, the slope of which is the retardation coefficient (K_R). A series of K_R values are determined for proteins of known molecular weights. The logarithm of the negative slope is then plotted against the logarithm of the molecular weight for each protein standard. This generates a linear plot. If the protein of unknown molecular weight is electrophoresed under identical conditions, the K_R determined, then the molecular weight can be derived from the standard plot.

2.3.2 Application to the EMSA.

The original method required the electrophoresis of standard and unknown proteins on discontinuous gels. When applied to the analysis of complexes in a EMSA, 0.25 x TBE as buffer was used with polyacryamide concentrations of 4.0%, 4.5%, 5.0%, 5.5%. The same stock solutions of acrylamide and TBE were used for each set of experiments.

Molecular weight markers for non-denaturing polyacryamide gel electrophoresis (nd-PAGE) were obtained from Sigma and prepared as indicated in the technical bulletin supplied.

α-Lactalbumin, bovine milk	1mg/ml
carbonic anhydrase, (bovine erythrocytes)	1mg/ml
albumin, chicken egg	1mg/ml
albumin, bovine serum	1mg/ml

all the above prepared in 1mM sodium phosphate pH 7.0, 50mM NaCl

urease (Jack bean)   0.2mg/ml in water

all proteins obtained from Sigma.

Polyacrylamide gels were prepared as described previously, pre-run, then dilutions of standard proteins loaded into separate slots. Binding reactions were performed as described and also loaded. The gels were run until the marker dye reached 1cm from the bottom. To locate the position of the dye after staining, the gel was cut through along the line of the dye. The gels were fixed in 15% methanol/7% acetic acid/78% H_2O, stained with a 0.1% solution of Coomassie Brilliant Blue in fixative solution then destained in 15% methanol/7% acetic acid/78% H_2O. The distances from the gel slot to the bottom of the gel and to the stained protein band were measured and Rf calculated thus:

$$Rf = \frac{\text{distance of protein band}}{\text{distance of marker dye}}$$

Gels were dried, exposed to X-ray film and the position of the specific retarded band marked onto the gel. The Rf for the retarded band could then be determined, using the same graphical methods as used for the protein standards, the K_R calculated and hence the molecular weight of the protein/DNA complex.

Table 2.1 showing 100x(logRfx100) for each standard protein at the gel concentration shown:

GEL CONCENTRATION

Protein standard	4.0	4.5	5.0	5.5
carbonic anhy	133.82	133.72	131.94	130.96
alb. chicken egg	185.81	185.13	183.07	181.87
BSA monomer	178.52	177.82	174.13	172.89
BSA dimer	170.94	168.64	163.66	159.86
urease trimer	156.07	149.05	139.11	125.53

Bandshift

	4.0	4.5	5.0	5.5
Jurkat	149.42	141.33	131.12	124.55
Hela 1	158.09	154.40	149.83	145.59
Hela 2	162.39	157.82	153 97	149.07

The Ferguson plot for these data is shown in section 4.2.6 (results)

2.4 Cell Culture and Transfections.

2.4.1 Cell Culture.

Jurkat cells were cultured in RPMI 1640 (supplied by Institute of Cancer Research) supplemented with 10% foetal calf serum (Gibco Tissue Culture Services) L-glutamine (2mM), penicillin (100U/ml) and streptomycin (100U/ml).

HL-60 (a human promyelocyte line), U937 (a human promonocyte cell line), THP-1 (a human monocyte line), were cultured in RPMI 1640/10% FCS supplemented with $5x10^{-5}$ M β-mercaptoethanol.

All the above cell lines were propagated in a 37°C incubator with 5% CO_2 atmosphere. Cell density was maintained between $5-10x10^5$/ml by sub-culture every 4 days. To avoid prolonged periods in culture, stocks of low-passage number cells were kept frozen in liquid nitrogen. Cells in culture were discarded regularly and replaced by fresh cells taken from these stocks. This was important to maintain transfection efficiency.

The adherent cell lines HeLa (human cervical carcinoma) and COS-1 (derived from the simian kidney cell line CV-1 transformed by SV40) were maintained in Dulbecco's minimum essential medium (DMEM) with 10% FCS and antibiotics, incubated in a 10% CO_2 atmosphere. Cells were propagated by trypsinizing sub-confluent cultures and replating at $5x10^4$ cells/ml. The murine embryonal carcinoma cell line was maintained in DMEM/10% FCS supplemented with

antibiotics and L-glutamine, this line required passaging every 2 days, replating at 4×10^5 cells for a 90mm culture dish.

All cell lines were routinely screened at 3 month intervals for contamination by mycoplasma. Any cells found to be positive for mycoplasma were discarded and fresh cells grown from frozen stock. If no stocks were available, infected cells were cultured in the presence of ciprofloxacin for a minimum of three passages then again screened for mycoplasma.

2.4.2 Transfection of DNA into eukaryotic cells.

Plasmid DNA for transfection was prepared as described in section (2.1.9). Following CsCl/ethidium bromide banding the plasmid DNA was treated with proteinase K 50µg/ml in 100mM Tris-HCl pH 7.8, 2mM EDTA, 0.25% SDS at 55°C for 2-4 hours; the DNA was then extracted with phenol/chloroform, ethanol precipitated, redissolved in TE and dialysed against a x1000 volume of TE. The concentration of plasmid DNA was determined from the OD_{260} assuming that an OD of 1 corresponded to 50µg/ml dsDNA. For further accuracy, dilutions of the sample were analysed on an agarose gel together with a standard DNA of known concentration and then stained with ethidium bromide.

A.   DEAE-dextran:

Cells growning as suspension cultures were transfected by a modification of the DEAE-dextran method.

Reagents; DEAE-dextran, mol.wt. 500,000 (Pharmacia), 50mg/ml. Sterilised by autoclaving (15 lbs/sq.in. for 20 minutes) and stored in small aliquots at -20°C. Culture medium: RPMI 1640 without bicarbonate and buffered with HEPES (10mM pH 7.15).

1. 24 hours before transfection, cells were taken from an exponentially growing culture (cell density $6\text{-}8 \times 10^5$/ml) and diluted in fresh medium to produce 2×10^5 cells/ml.

2. After 24 hrs. further culture cells were harvested by centrifugation (1000g for 10 min.), washed twice in serum-free RPMI with HEPES and finally resuspended at 1×10^7/ml in serum-free RPMI/HEPES.

3. For each transfection the following were added in the order given: to 1ml of cells; plasmid DNA in 10ml TE; 1ml DEAE-dextran solution diluted to 600µg/ml added dropwise (final concentration 300µg/ml).

4. The cells were incubated in the DNA/DEAE-dextran for 37°C for 1hour, washed x2 with RPMI and once with RPMI/1% FCS. The cells were finally resuspended in 10ml RPMI/10% FCS, any potential activating agents added, and then incubated for a further 24-36hrs. Cells were harvested by centrifugation,

washed twice with PBS and extracts prepared for the chloramphenicol acetyltransferase assay.

B. Calcium phosphate-mediated tranfection

(Chen and Okayama 1987).

This method was found to be very efficient for adherent cell lines eg. HeLa, COS-1, F9.

1. Preparation of cells: 8hrs. prior to transfection, subconfluent cells were harvested by trypsinization and replated at $5x10^5$ cells in 10cm diameter petri dishes.

2. Preparation of reagents:

1M BES (N,N-bis[2-hydroxyethyl]-2-aminoethanesulphonic acid); this was prepared in 90ml HPLC-grade water, and the pH adjusted to exactly 6.95 with 2M NaOH, the volume then made up to 100ml.

1M NaCl; 500mM Na_2HPO_4; 2.5M $CaCl_2$; this was sterilized by filtration through a

0.22 micron filter, stored at -20°C in 1ml aliquots. A stock solution of 2x BES-buffered saline (x2 BBS) was prepared to give final concentrations of - 50mM BES, 280mM NaCl, 1.5mM Na_2HPO_4 and the pH again adjusted to 6.95. This stock

solution was sterilized by filtration and stored in 10ml aliquots at -20°C.

3. Transfection:

All reagents were warmed to room temperature. The $CaCl_2$ solution was diluted to 0.25M and 0.5ml dispensed into a 5ml perspex Bijou for each transfection point. Plasmid DNA was added as indicated in the footnote for each set of results. A total of 20mg plasmid DNA was required for each point, the balance made up with a plasmid such as pUC 1318. 0.5ml of 2x BSS was added dropwise and the mix allowed to stand at RT for 10-20 minutes. The entire 1ml was then added dropwise to 10ml of DMEM/10% FCS in a 9cm tissue-culture dish containing the prepared cells. The transfection mix remained in contact with the cells overnight in an atmosphere of 3% CO_2, ideally a fine precipiate would have formed overlying the cells. The medium and precipitate were removed and the cell monolayer washed x2 with DMEM. Fresh DMEM/10% FCS was added and the cells incubated for a further 24-48 hrs. in 10% CO_2.

Cells were harvested by scraping after washing twice with PBS.

2.4.3 Chloramphenicol acetyltransferase (CAT) assay.

A.	*Preparation of cells:* cells grown in suspension culture were harvested by centrifugation (1000g for 10 minutes at 4°C), washed three times in PBS and finally pelleted in a 1.5ml eppendorf tube. The cell pellet was then resuspended in 100μl of solution 1 (Tris-HCl pH 7.5; DTT 5mM; EDTA 5mM;

glycerol 15%; PMSF 0.5mM). Cells were disrupted by three freeze/thaw cycles (liquid nitrogen and a 37°C waterbath), cellular debris then pelleted by centrifugation (2 minutes in a microfuge at 18000g) and the supernatant removed. If a β-galactosidase expression plasmid had been used in the transfection an aliquot of the extract was removed at this point for assay. The remaining extract was heated to 65°C for 15 minutes to inactivate endogenous deacetylases, any precipitate of denatured proteins removed by centrifugation (microfuge for 2 minutes) and the remaining extract either immediately assayed for CAT activity or stored frozen at -20°C.

B. *β-Galactosidase assay:* As described in Maniatis p16.66 (Sambrook *et al* 1989).

C. *CAT assay:* The volume of extract to be assayed was normalised from the β-galactosidase activity. The CAT assay was performed in 0.25M Tris-HCl pH 7.5, containing 80μg acetyl-coenzymeA (Sigma, enzymatically produced) and 0.125μCi of ^{14}C-labelled chloramphenicol (60mCi/mmol, Amersham Radiochemicals) in a final volume of 120μl, incubated at 37°C for 2-6 hours (the time required varied depending upon the cell line used). The reaction was stopped by adding 1ml of ethyl acetate, with thorough mixing by vortexing for 1 minute. The aqueous and organic phases were separated by brief centrifugation and 900μl of the upper organic phase removed into a fresh eppendorf tube. The ethyl acetate was removed by evaporation in a Savant SpeedVac and the reaction products of the assay redissolved in 20μl of ethyl acetate. The reaction products, mono- and di-acetylated chloramphenicol, were separated from any unreacted chloramphenicol by thin layer chromatography using a 95:5 mix of chloroform:methanol. The dried film was then exposed to X-ray film for 18-72 hours, following which the radioactive spots corresponding to the reaction products and unreacted chloramphenicol were cut out of the TLC plate and the radioactivity quantified using a scintillation counter.

2.4.4 Preparation of depleted serum ('charcoal-stripped serum')
(Method provided by Dr. Jacques Ghysdael, INSERM, Lisle).
To 500ml of heat-inactivated FCS were added 5g ion exchange resin AG1X8, 20-50 mesh chloride ion form (Biorad) and 2g of activated charcoal, the mixture then incubated at room temperature for 4 hours on a roller platform. The suspension was clarified by centrifugation (4000g for 30 minutes) and the depleation process repeated a further three times. After the final low speed spin, the FCS was fully clarified by high speed centrifugation (20k g for 30 minutes), filter sterilized and stored in 20ml aliquots at -20°C.

2.5 Polyacrylamide gel electrophoresis.

2.5.1 Non-denaturing gels.

As required for EMSA and Ferguson plots.

A stock 30% solution of 29:1 acrylamide:NN'-methylene bisacrylamide (wt/vol) (Gibco-BRL) was prepared in HPLC-grade water, filtered through a disposable 0.45 micron filter and stored in the dark at 4°C. The stock solution was used within 2 weeks. Details of gel preparation are given in the section describing the EMSA (section 2.2.3).

2.5.2 SDS denaturing polyacrylamide gels.

As required for protein molecular weight determination. The stock 30% acrylamide:bis-acrylamide solution was prepared as described in section 2.5.1 above. Gels were prepared according to the method of Laemmli as modified and described in Maniatis (Sambrook *et al* 1989).

A: The running gel consisted of 0.375M Tris-HCl pH 8.8, 0.1% SDS and polymerised after filtration and degassing by the addition of N,N,N',N'-tetramethyletthylenediamine (TEMED) and 10% (wt/vol) ammonium persulphate (APS). For a gel volume of 30ml, 20μl TEMED and 60μl APS were required. The running gel mix was poured to a level 2cm below the slot comb, then overlaid with isobutanol while polymerising. The isobutanol was removed, the gel washed with HPLC water, then stacking gel buffer.

B: The stacking gel was a 4.5% gel mix containing 0.125M Tris-HCl pH 6.9, 0.1% SDS. For 10ml, 30μl of 10% APS and 10μl TEMED were used to polymerise the gel.

C: Reservoir buffer; 25mM Tris, 0.192m glycine, 0.1% SDS.

D: x2 loading buffer; 100mM Tris-HCl pH 6.9, 4% SDS, 20% glycerol, 0.1% bromophenol blue. Equal volumes of x2 loading buffer and protein sample were combined, 1M DTT added to a final concentration of 100mM, the preparation was then heated to 95°C for 5 minutes, cooled briefly and centrifuged to bring down the condensation. The samples could then be loaded onto the gel. Gels were routinely run at a constant current of 30mA.

2.5.3 Sequencing gels.

Gels were 0.1mm thick, prepared either as a uniform x1 TBE or with an ionic gradient. A stock 40% solution of 25:1 acrylamide:bisacrylamide was freshly prepared and filtered as described in section 2.5.1.

A: For a 6% gel with x1 TBE the following was prepared: 15ml acrylamide stock; 10ml x10TBE; 45g urea, HPLC-grade H_2O to 100ml. The urea was dissolved

in the acrylamide the solution filtered and degassed and 400μl of 10% APS and 40μl TEMED were added with mixing and the gel mix poured between the prepared glass plates. The plates had been carefully cleaned, with a final wash with ethanol before being taped together. The smaller of the plates was treated with dichlorodimethyl silane to aid separation of the plates. The gel was allowed to polymerise for 1 hour, pre-run for 30 minutes at 150 volts before loading samples.

B: Ionic gradient gels: (Biggin et al 1983) gradient from 0.5x TBE to 5x TBE with 0.5x TBE in the upper reservoir and 1x TBE in the lower. Separate gel mixes were prepared as above but with the TBE concentrations given, 0.05% bromophenol blue was added to the 5x TBE mix to aid pouring the gradient. Polymerization was initiated with 2μl 25% APS and 2μl TEMED per ml of gel, 10ml of 0.5x TBE then drawn into a 25ml pipette, followed by 15ml of 5x TBE, a single air bubble then drawn in to mix the interface, then the gel mix poured into the prepared plates. A further 60ml of 0.5x TBE gel mix was poured to complete the gel. The gel was not pre-run.

Both types of sequencing gel were electrophoresed at 40W constant power for the desired time. After electrophoresis, the plates were separated, leaving the gel attached to the large plate. The gel was immersed in 20% methanol/10% acetic acid to leach out the urea, dried under vacuum then autoradiographed.

2.5.4   Preparation of RNA.

Total cellular RNA was prepared using the acid-guanidinium-phenol-chloroform (AGPC) method as described by Chomcznski and Sacchi (1987). Briefly, the denaturing solution (solution D) was prepared by dissolving 250g of guanidinium isothiocyanate (Fluka Chemicals) in 293ml HPLC-grade water in the manufactures bottle (to avoid unnecessary handling) together with 17.6ml 0.75 M sodium citrate pH 7.0, 24.4 ml 10% sarcosyl (BDH chemicals) and heated to 65°C. Just before use, 0.36ml of β-mercaptoethanol was added to 50ml of the above solution. Cells were harvested from exponentially growing cultures, pelleted by centrifugation (1000g for 10minutes at 4°C), then resuspended in solution D, using 100ml per 10^6 cells. Sequentially, 10ml of 2M sodium acetate pH 4.0, 100ml of water-saturated phenol and 20ml of chloroform:iso-amylalcohol (49:1) were added for every 100ml of solution D, with mixing by inversion after the addition of each reagent. The final mixture was vigorously shaken for 10 seconds, cooled on ice for 15 minutes then centrifuged at 10,000g for 25 minutes at 4°C. The upper aqueous phase containing RNA was removed, avoiding the interface which contained DNA,

then mixed with an equal volume of isopropyl-alcohol and left at -20°C for 2 hours to precipitate the RNA. The RNA was pelleted by centrifugation at 10,000g for 20 minutes at 4°C. The pellet was then redissolved in a small volume of solution D and re-precipitated with x3 volume of ethanol at -20°C. After centrifugation, the RNA pellet was washed x3 with 75% ethanol and was stored in ethanol at -20°C. When required, an aliquot was removed, pelleted by centrifugation (microfuge, 17,000g, 10') the pellet dried and the RNA dissolved in HPLC-grade water that had been treated with diethyl-pyrocarbonate (DEPC, Sigma). The concentration of RNA was determined from the OD_{260} assuming that 40mg/ml of RNA has an OD_{260} of 1.0.

2.5.5 Northern blot analysis of RNA:

A: Formaldehyde-agarose gel electrophoresis:

X5 gel buffer: 0.1M sodium morpholinopropanesulphonic acid (MOPS)

40mM sodium acetate

5mM EDTA (pH 8.0)

The pH was adjusted to 7.0 with 2M NaOH. The buffer was stored in the dark at 4°C. All buffers were prepared using DEPC-treated HPLC-grade water.

Sample loading buffer: 50% glycerol, 1mM EDTA, 0.25% bromophenol blue

A 1.5% agarose/2.2M formaldehyde/x1 MOPS buffer gel was prepared with DEPC-treated water. The RNA samples were prepared for electrophoresis by freezing-drying aliquots containing 15µg (in aqueous solution) under vacuum in dry-ice. The RNA pellets were redissolved in 4.5µl DEPC-treated water, 2.0µl x5 MOPS, 3.5µl formaldehyde, 10µl de-ionised formamide (Gibco-BRL), heated to 65°C for 15 minutes and cooled on ice. Just before loading onto the gel, 2µl of loading buffer was added to each sample. The gel was electrophoresed at 5V/cm with recirculation of the buffer.

RNA was visualized after electrophoresis using acridine orange (Wilkinson, Doskow, and Lindsey 1991). The gel was gently agitated in a solution of acridine orange (AO, Sigma) 15µg/ml in 10mM sodium phosphate pH 6.5, 3% formaldehyde for 3 minutes then destained in the same solution, without AO, for 20-30 minutes. The RNA bands were visualized with medium wave ultra-violet. RNA was more visible after transfer to either nylon (Gene-Screen Plus, Dupont) or nitrocellulose (Schleicher and Schuell).

Transfer of RNA to membranes for probing was achieved using capillary transfer (Thomas 1980). The transfer was confirmed by viewing the blot with a hand-held ultra-violet lamp. The blot was air dried, prehybridized at 60°C in 15ml 1M sodium chloride, 10% dextran sulphate, 1% SDS and 150µg sheared

salmon sperm DNA for 1 hour using a rotissary oven (Hybaid). Radioactive probe (final concentration approximately 10ng/ml, specific activity $1\text{-}5\times10^5$cpm/ml) was heated at 100°C for 5 minutes to denature then added to the prehydridization mixture. Hybridization normally proceeded for 15-20 hours. The blot was washed twice with 100ml of x2 SSC at room temperature for 10 minutes, twice with x2 SSC/1% SDS at 60°C for 20 minutes and finally twice with x0.2 SSC at 60°C for 10 minutes. The blot was wrapped in Saranwrap while still wet and the bands visualized by autoradiography.

2.5.6 Western blot analysis of proteins:

A: Transfer of proteins to membrane.
Reagents: *10x CAPS buffer*
(100mM 3-[cyclohexlamino]-1-propanesulphonic acid pH 11.0); electroblotting buffer - 1x CAPS and 5% methanol.
After electrophoresis on a SDS-polyacrylamide gel, the gel was removed from the glass plates and soaked for 10 minutes in electroblotting buffer. Proteins were electroblotted onto an Immobilon-P membrane (polyvinyl difluoride, Millipore) that had been presoaked in methanol, using a Pharmacia semi-dry transfer apparatus, using a current of 0.8mA per cm^2 of membrane for 2 hours. The membrane was then briefly rinsed with transfer buffer then HPLC-grade water and transfer of protein confirmed by staining the membrane with Ponceau red (0.2% solution), followed with destaining by briefly rinsing in water.

B: Detection of specific protein bands by immunostaining:
After confirming the protein transfer, lanes containing molecular weight markers were separated and the remainder of the membrane soaked in TN buffer (25mM Tris-HCl pH 7.5; 500mM NaCl) containing 5% non-fat dried milk powder, for 1 hour at room temperature. The primary polyclonal rabbit antibody was added at an appropriate dilution, and further incubated for 4-6 hours at room temperature. The blot was washed x3 with TN buffer then incubated with a 1 in 10,000 dilution of goat anti-rabbit antibody conjugated to alkaline phosphatase (Promega) for 2 hours at room temperature. The blot was washed x3 with TN and positive bands developed using a chromogenic reaction for alkaline phosphatase with 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium (BCIP/NBT) as substrate (Maniatis).

2.5.7 Radio-labelling DNA.

A: Probes for Southern and northern blots.

Probes were prepared using the method of random hexanucleotide priming as described by Feinberg and Vogelstein (1984).

50ng of double stranded DNA fragment was denatured by heating at 100°C for 5 minutes, cooled on ice, then labelled in 50μl containing 50mM Tris-HCl pH 8.0, 5mM MgCl$_2$, 10mM β-mercaptoethanol, 4mM each of dATP, dGTP and TTP, 110μg/ml mixed hexanucleotides, 0.4mg/ml bovine serum albumin and 1.85x10^6 Bq (1.1x10^{14} Bq/mmol) of [α-^{32}P]dCTP (Amersham). 10 units of Klenow fragment of DNA polymerase (Northumbrian Biotech) were added and the reaction mix incubated at 37°C for 1 hour. The reaction was terminated by adding 50μl TE. Labelled DNA was separated from unincorporated radionucleotide on a G-50 Sephadex spin column. Immediately before adding the probe to a hybridization solution it was denatured by heating at 100°C for 5 minutes then chilled on ice.

B. 5' End-labelling of DNA fragment (for Dnase I footprinting and methylation interference assays).

50μg of plasmid DNA containing the fragment was cleaved with the first appropriate restriction enzyme to generate linear DNA with a 5' overhang and dephosphorylated using calf intestinal alkaline phosphatase (Boehringer). The dephosphorylated linear DNA was extracted twice with 100μl of phenol/0.1% hydroxyquinolone, once with chloroform:iso-amyl alcohol and precipitated with ethanol. The pellet was redissolved in TE and the concentration checked on an agarose gel. This could be stored frozen at -20°Cuntil required.

Labelling reaction: 2pmol (approximately 2μg) of DNA fragment in 2μl TE; 5μl of x10 kinase buffer; 10μl γ-^{32}P-ATP; 32ml H$_2$O; 10 units polynucleotide kinase. Final volume 50μl. Incubated at 37°C for 45 minutes. Phenol/chloroform extraction and ethanol precipitation. The DNA was digested with the second restriction enzyme to release linear DNA fragment, labelled on one 5' end only. Fragment then isolated by electrophoresis through a 5% polyacrylamide gel, band excised after brief exposure to X-ray film to locate the band. Fragment extracted from the gel in low salt buffer (LS) and contaminating gel removed using an Elutip (Schleicher and Schull) after washing with LS and eluting fragment with high salt buffer. The 5' labelled fragment was then precipitated with ethanol, washed with 70% ethanol and finally redissolved in TE o give a sloution of approximately 20,000cps per μl.

Materials and Methods: Appendix 1

DNase I footprint assay.

Principle of assay: This assay exploits the endonuclease activity of DNase I. A binding assay is prepared, using more nuclear extract than in the EMSA in order to ensure that all potential specific protein binding sites are occupied. The assay also requires poly(dI-dC) to remove non-specific DNA-binding proteins, the amount required must be determined empirically. Limited DNase I digestion of the end-labelled fragment is performed, with the aim of cleaving each molecule in one site only. Analysis of the digested mixture on a sequencing gel produces bands representing digested fragments of different sizes. At sites of protein binding, the DNA fragment is relatively protected and the sequencing gel will show a region with bands of reduced intensity, a footprint. Specificity of binding can be demonstrated by the inclusion of excess unlabelled DNA either as the intact fragment or as specific oligonucleotides with sequences corresponding to particular sites.

1. Binding reaction: 20mM Hepes buffer pH 7.6, 40mM NaCl, 5mM $MgCl_2$, 2mM CaCl2, 1mM dithiothreitol, 10% glycerol, 5fmol end-labelled DNA fragment (see section 2.5.7 B), poly(dI-dC), nuclear extract were mixed together in a final volume of 100μl. Several binding reactions were prepared varying the amounts of nuclear extract between 0 - 50μl (0 - 10mg/ml protein) and poly(dI-dC) present, each in duplicate. Binding reaction then incubated on ice for 30 minutes.

2. Digestion with DNase I: Following binding, the DNA fragment was subjected to limited DNase I digestion using 1μl of (20ng/μl in Hepes buffer as used in binding reaction) for 15 - 30 seconds at room temperature. Reaction was stopped by addition of stop buffer (2% SDS, 10mM EDTA, pH 8.0, 0.1mg/ml tRNA).
Sample then extracted x2 with phenol/chloroform and precipitated with ethanol. The pellet was washed with 75% ethanol then re-dissolved in 10ml of sequencing solution (95% v/v deionised formamide, 0.025% w/v xylene cyanol FF, 0.025% w/v bromophenol blue). The sample was heated immediately at 90°C for 3 minutes then rapidly frozen at -80°C, then thawed immediately before loading onto a polyacrylamide sequencing gel.

Appendix 2

Materials and Methods: Methylation interference assay.

Principle of assay: Partial methylation at the N7 position of guanine residues is achieved using dimethyl sulphate (DMS). Ideally, 50% of the DNA molecules should be methylated at one site only. The length of the reaction required to achieve this is determined in preliminary experiments in which the labelled DNA fragment is methylated with DMS for varying lengths of time and then cleaved with piperidine and run on a sequencing gel. The methylated DNA fragment is used in an EMSA binding reaction, the presence of methylated G residues will block protein binding at these sites. Comparison of the DNA fragments in the retarded and the non-retarded bands reveals G residues at sites critical for protein-DNA interaction.

Limited methylation of guanine residues with DMS:

1. 200µl of 50mM sodium cacodylate (Sigma) pH 8.0, 1mM EDTA and 200fmol of end-labelled DNA fragment were mixed in a 1.5ml Eppendorf screw-capped tube.

2. The mixture was then chilled on ice and 1µl of DMS added (BDH, HPLC grade) added. Incubated at 20°C for 3 minutes.

3. 25µl 3M sodium acetate, pH 7.0 and 600µl ice-cold ethanol were added. The tube was chilled at -70°C for 4-5 minutes then centrifuged at 12,000g for 45 minutes in a refrigerated bench-top microfuge.

4. The supernatant was removed and transferred to a waste bottle containing 5M NaOH.

5. The DNA pellet was washed with 70% cold ethanol, air dried then re-dissolved in TE.

NB The procedure was performed in a fume hood as DMS is a carcinogen.

Methylation interference assay:

1. The following were combined in 120-200µl of EMSA binding buffer;

      50-100fmol of end-labelled methylated DNA fragment (approximately
      4×10^5cpm)

      poly(dI-dC) and poly(dA-dT)

      nuclear extract

Approximately x10 quantities of all components as in a EMSA were used.

The binding assay was incubated on ice for 1 hour then loaded onto 3-5 adjacent wells on a 3mm thick 5% polyacrylamide gel, running buffer 0.25% TBE.

2. After electrophoresis, one plate was removed, the wet gel was wrapped in cling-film and exposed to X-ray film for 4-6hrs.

3. The retarded and non-retarded bands were localised on the gel using the exposed film and excised. The DNA was extracted from the gel as previously described, using Elutips.

Piperidine cleavage of DNA:

4. The purified DNA fragments were dissolved in 10μl freshly prepared 1.0M piperidine (BDH, HPLC grade) in a screw-topped Eppendorf tube and heated to 90°C for 30 minutes.

5. The tube was chilled on ice, centrifuged briefly and holes made in the top with a hypodermic needle.

6. The DNA was dried for 2-3 hours in a Speedivac centrifugal dessicator.

7. The DNA was re-dissolved in 50μl of water and dried again in the dessicator for 1-2 hour.

8. Repeat procedure 7.

9. DNA re-dissolved in 5μl of sequencing sample solvent (as above), heated to 90°C for 3 minutes then snap frozen at -70°C until ready to load and run on a sequencing gel.

Chapter 3

Identification of binding sites in the NRE for cellular DNA-binding proteins.

Chapter 3

Identification of binding sites in the NRE for cellular DNA-binding proteins.

3.1   Introduction.

Although the existence of a negative regulatory element (NRE) in the LTR of HIV-1 has been questioned (Berkhout et al 1992; Dinter et al. 1987; Zeichner et al. 1991) overwhelming evidence from several groups indicates the presence of sequences upstream from the core promoter that exert a net inhibition on LTR-driven transcription under a wide range of conditions and in different cell types (Garcia et al. 1987; Giacca et al. 1992; Kundu et al. 1995; Lu et al. 1989; Lu et al. 1990; Okamoto et al. 1990; Rosen et al. 1985; Siekevitz et al. 1987; Smith and Greene 1989; Toyama et al. 1992). The limits of the NRE had initially been determined by 5' deletion analysis of the LTR, the size and site of deletions varying between different studies according to the restriction sites chosen. Although convenient as a working model this arbitrary segmentation of the LTR does not completely correlate with the experimental evidence as both positive and negative regulatory elements have been demonstrated in what has been defined as the NRE. Consequently, the alternative term 'modulatory element' has been suggested for the region -454 to -78 (Gaynor 1992). For convenience the term negative regulatory element (NRE) will be used.

Prior to the publication of the results presented in this thesis, no protein binding sites had been identified and characterised within the LTR upstream of -270 (Garcia et al. 1987). Potential sites for AP1 binding had been proposed because of sequence homologies in the NRE with the consensus binding site for AP1. Franza et al, using DNA-affinity precipitation followed by Western blot analysis and in separate experiments using 2-D gel electrophoresis, were able to show a very weak association of AP1 with an oligonucleotide containing the putative AP1 site in the LTR (Franza et al. 1988). However, this study can be criticised in several ways. Firstly, the HIV-1 strain from which the sequence was derived was not given, indeed comparison of the sequence used for the experiments with the Los Alamos HIV-1 sequence data-base produced only one strain of HIV-1 that came nearest to a complete match (HIVANT70); secondly, significant differences can be found between an AP1 consensus sequence and the LTR consensus sequences in the Los Alamos data-base, this would explain the weak association of AP1 with the sites; third, methods that would directly demonstrate the interaction of protein with DNA were not attempted e.g. DNAse 1 footprinting and gel retardation assays, super-shift assay; finally, the functional significance of the sites was not determined.

This chapter details the experiments performed to identify and characterise sites within the NRE that specifically bind cellular DNA-binding proteins. The following techniques were employed; electrophoretic mobility shift assay (EMSA), DNAse I footprinting and methylation interference. Two sites were identified, site A and site B, neither of which had been previously described. Because the DNA-protein complex of site B consistently produced a stronger footprint and mobility shift, experimental work concentrated on characterising this site and the cellular proteins binding to it. In addition, site B will be demonstrated to act as a repressor of LTR-driven gene expression in J6.

3.2 Results.

3.2.1

Delineation of protein-binding sites in the *Kpn1-Rsa1* '220' fragment.

In order to identify sites within the NRE with which T-lymphocyte nuclear proteins may interact, DNAse I footprint analysis was performed using the '220' Kpn1-Rsa1 fragment from the U3 region of the HIV-1 LTR and nuclear extracts from Jurkat T-lymphocyte cell line (J6). In figure 3.1. (lane B) two regions were footprinted by the J6 nuclear extract, located at -379 to -366 (site A) and -356 to -324 (site B) when the lower strand of the fragment was labelled. Footprinting with an upper strand labelled fragment produced protected regions that were complementary to those of the lower strand. These two regions are shown diagramatically in figure 3.2. Each of these footprints was specifically eliminated when a 100-fold excess of a 30 base pair oligonucleotide spanning the protected sequence was included in the assay (fig. 3.1, lanes C and D). When both oligonucleotides were included (lane E) a DNAse I cleavage pattern identical to that in the absence of nuclear extract was obtained. An oligonucleotide with an unrelated sequence had no effect on either footprint. Identical patterns were obtained by nuclear extracts prepared from either non-activated J6 cells or J6 cells stimulated by a combination of phorbol ester and phytohaemaglutinin. In addition to the two footprints, DNAse I hypersensitive sites were revealed, one within site B (lane D arrowhead) when an excess of unlabelled site B oligonucleotide was present, another at the edge of site B when an excess of site A oligonucleotide was present (arrowhead lane C) which also resulted in the enhancement of the site B footprint. Since DNAse I hypersensitive sites are induced by conformational changes to the DNA double helix, the presence of a hypersensitive site within site B indicates that not only are these internal sequences exposed to DNAse I cleavage but the DNA may undergo conformational changes induced by cellular protein(s) binding to site B e.g. DNA bending.

Fig 3.1

**DNase I footprint analysis of the negative
regulatory element of the HIV-I LTR.**

Lane A, no nuclear extract; lanes B-E labelled NRE fragment incubated with Jurkat
nuclear extract. In lane B the assay mixture did not contain any unlabelled competitor
oligonucleotide; lane C a 100-fold excess of unlabelled site A oligonucleotide was
included; in lane D the assay contained an excess of unlabelled site B oligonucleotide;
lane E the assay contained excess unlabelled oligonucleotides for both site A and B. A
Maxam and Gilbert G reaction (Maxam and Gilbert 1980) of the NRE fragment used
for DNase footprinting is shown as sequence marker (lane G). The extent of footprints
over the two sites is indicated.

Fig 3.2   **Region of NRE from HIV-1 LTR of strain HXB2 showing site A and site B footprints.**
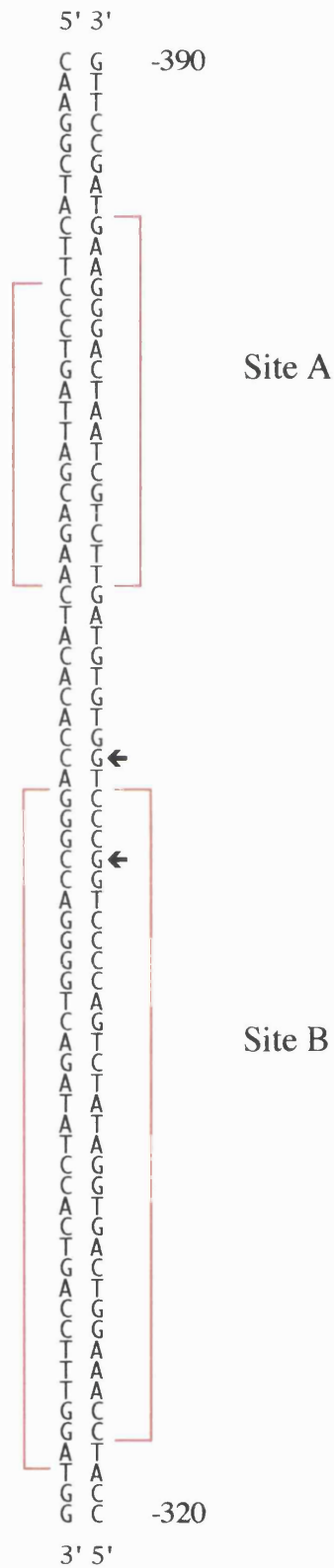
5' 3'

```
C  G   -390
A  T
A  T
G  C
G  C
C  G
T  A
A  T
C  G
T  A
T  A
C  G
C  G
C  G
T  A
G  C
A  T
T  A
T  A
A  T
G  C
C  G
A  T
G  C
A  T
A  T
C  G
T  C      Site A
A  T
C  G
A  T
C  G
A  T
C  G
C  G
A  T
G  G  ←
G  T
G  C
C  C
C  G  ←
A  G
G  T
G  C
G  C
G  C
T  C
C  A
A  G
G  T
A  C
G  T
A  A
T  T
A  A
T  G
C  G      Site B
C  T
A  G
C  A
T  C
G  T
A  G
C  G
C  A
T  A
T  A
T  C
G  C
G  T
A  A
T  C
G  C
G  C   -320
```

3' 5'

Figures indicate the number of nucleotides upstream of the start site for transcription.
Brackets delineate the extent of site A and site B footprints.
The positions of the DNase I hypersensitive sites are shown by arrows.

136

Figure 3.3 shows the result of a gel retardation assay using the Kpn1-Rsa1 fragment I ('220') of the NRE as the end-labelled probe and a nuclear extract prepared from non-stimulated T-lymphocytes. Two low mobility bands were seen, representing protein-DNA complexes (lane A). One of these bands was specifically eliminated by competition with an unlabelled site A oligonucleotide (lane B), the other by competition with a site B oligonucleotide (lane C). In the same experiment, two bands of higher mobility were also specifically competed by the site B oligonucleotide. These complexes may have been derived by proteolytic cleavage of the same protein(s) involved in the abundant slower mobility complex An alternative explanation would be that they are formed by the binding of smaller protein factors, possibly sub-units of the proteins involved in the larger complex. Fig 3.3 are the results of competitive EMSA using a number of unrelated cold oligonucleotides. None of the bands competed by either the site B or site A oligonucleotides were eliminated by competition with an unrelated DNA sequence, a 100-fold molar excess of pBR322 (lane F). Inspection of the sequence of site A revealed homology with the octamer element to which binds the ubiquitous transcription factor Oct-1 (core consensus ATGCAAATNA), however, an oligonucleotide containing a consensus octamer sequence failed to compete either complex (lane E). Site B contains sequences previously proposed as potential AP-1 binding sites (Franza et al. 1988), but an oligonucleotide containing an AP-1 consensus sequence failed to compete the complex in which site B was involved (lane D).

Fig 3.3

EMSA of the NRE '220' fragment (-456 to -300)
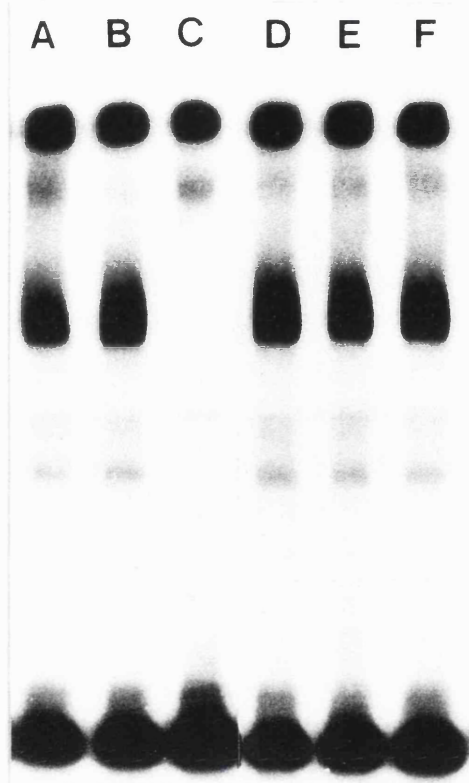
NRE '220' fragment as labelled probe; lane A, assay performed without unlabelled competitor DNA; lanes B to F assay performed with 100-fold excess of unlabelled oligonucleotide competitor as follows: lane B unlabelled site A oligonucleotide; lane C unlabelled site B oligonucleotide; lane D unlabelled oligonucleotide containing a consensus AP1 binding site from the SV40 enhancer; lane E unlabelled oligonucleotide containing a consensus octamer binding site; lane F 100 molar excess of plasmid pBR322.

3.2.2 Electrophoretic mobility shift assays (EMSA) with site B oligonucleotide as probe.

The strong footprint and mobility shift associated with site B suggested that the Jurkat nuclear extracts contained relatively abundant protein(s) specifically interacting with the site B sequence. To further investigate the DNA-protein complex at site B, gel retardation assays using a site B oligonucleotide probe were performed.

Figures 3.4a and 3.4b shows the results of these experiments. With nuclear extracts from either non-stimulated or stimulated Jurkat cells a single major specific DNA-protein complex was formed (open arrow). Specificity was proven by the elimination of the complex with an excess of unlabelled site B oligonucleotide (lane C) but not by an excess of the site A oligonucleotide (lane E) or pBR322 plasmid DNA (lane H). Two more complexes, with faster mobility, were specifically associated with the site B oligonucleotide (faint bands, indicated by closed arrows). These complexes were much less abundant and appeared to correspond to the minor bands when the '220' fragment was used as a probe.

To determine whether the strong complex seen with the site B oligonucleotide was formed by *fos-jun* heterodimers binding to the weak AP1 sequence competition with an oligonucleotide containing the simian virus 40 enhancer AP1 was performed (lane F). As can be seen, no competition for any of the complexes occurred. The presence of fos-jun heterodimers in the Jurkat nuclear extracts was confirmed by EMSA with the consensus AP1 oligonucleotide as probe, a single DNA-protein complex was formed, which substantially increased in quantity when a nuclear extract from Jurkat cells stimulated by PHA/PMA was used (N. Perkins, personal communication). During experiments to determine the binding characteristics of the Jurkat protein to site B an unusual but reproducible effect was seen. This is demonstrated in fig. 3.4b. Treatment of protein extracts with detergents disrupts protein-protein and protein-DNA interactions leading to either complexes of faster mobility in the EMSA or complete loss of the retarded band. Surprisingly, a marked decrease in the mobility of the complex was produced. Specificity of binding was confirmed by competition with excess site B oligonucleotide (fig. 3.4b, lane C) but neither by the AP1 nor the octamer oligonucleotides (lanes D and E). The detergent induced 'supershift' was also specific for the major protein-DNA complex since the minor bands were not affected.

Fig. 3.4

EMSA with labelled site B oligonucleotide as probe.

a)

b)
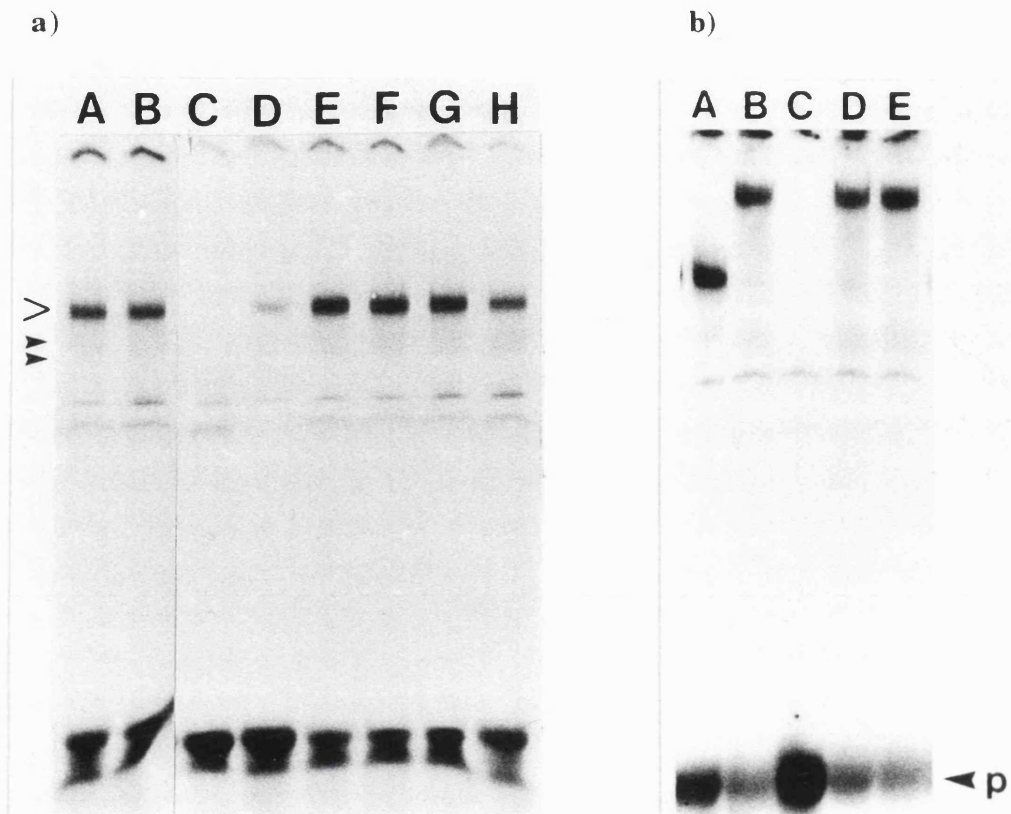
a) EMSA with labelled site B oligonucleotide as probe and Jurkat nuclear extract. Open arrow showing major specific DNA-protein complex, small closed arrows indicate the position of two faint specific bands of higher mobility. Lanes A and B no unlabelled competitor DNA; lanes C to H contained 100-fold excess of unlabelled competitor as follows: lane C site B oligonucleotide; lane D an oligonucleotide containing the thyroid hormone response element (TRE) from rat growth hormone promoter; lane E site A oligonucleotide; lane F on oligonucleotide containing SV40 enhancer AP1 site; lane G oligonucleotide containing the SV40 AP2 site; lane H 100-fold molar excess of plasmid pBR322 DNA.

b) EMSA demonstrating the effect of the detergent on the Jurkat nuclear factor complex; the EMSA mixture was incubated on ice for 30 minutes with 0.8% deoxycholate and 0.5% Brij (lanes B to D).Specificity of the retarded complexes shown by the effect of 100-fold excess of unlabelled oligonucleotide: lanes A and B no competitor; lane C excess site B oligonucleotide; lane D excess API oligonucleotidde; lane E excess octamer oligonucleotide.

3.2.3 Sequence analysis of site B element.

DNAse I footprinting delineated the extent of the putative transcription factor binding sites A and B within the NRE, but in order to investigate in greater detail the nature of protein-DNA interaction at these sites the following techniques were required;

i) EMSA with competition using unlabelled oligonucleotides containing defined consensus sequences for known transcription factors,

ii) Methylation interference assay.

iii) EMSA with oligonucleotides containing specific mutations

3.2.3.1 Electrophoretic mobility shift assays (EMSA).

In the initial competition experiments, site B binding was partially competed by a 100-fold excess of an oligonucleotide containing the thyroid hormone response element (TRE) from the rat growth hormone promoter. (fig 3.4a lane D). In the TRE oligonucleotide the core binding site is formed by the heptameric sequence GGTCA as an imperfect inverted repeat, separated by a 6 base-pair spacer (Samuels et al. 1989)

Examination of the nucleotide sequence of site B as delineated by DNase I footprinting revealed a palindromic arrangement of bases with the motif GGTCA repeated with a nine base-pair spacer. This sequence has strong homology not only to the thyroid hormone response element (TRE) but also with hormone response elements (HREs) for several members of the superfamily of steroid and thyroid hormone receptors (fig. 3.12) with the half-site consensus sequence (A/G)GGTCA (Evans 1988; Yamamoto 1985). Further EMSAs were performed to determine the affinity of the site B protein with TRE- or ERE containing oligonucleotides. Figure 3.5 shows the result of a titration with increasing molar excesses of unlabelled oligonucleotides containing either the TRE from the rat growth hormone promoter (Flug et al. 1987) or the ERE from the *Xenopus* vitellogenin A2 promoter (Klein-Hitpab et al. 1986) have been used in competition with a fixed concentration of labelled site B oligonucleotide. The ERE was recognised with an approximately three-fold lower affinity than the site B oligonucleotide, the TRE with approximately twenty-fold lower affinity. An oligonucleotide containing a consensus glucocorticoid response element (GRE) failed to compete for binding.

Fig 3.5

Competition EMSA with site B oligonucleotide as probe and titrated against unlabelled oligonucleotides containing an ERE or TRE as competitor.
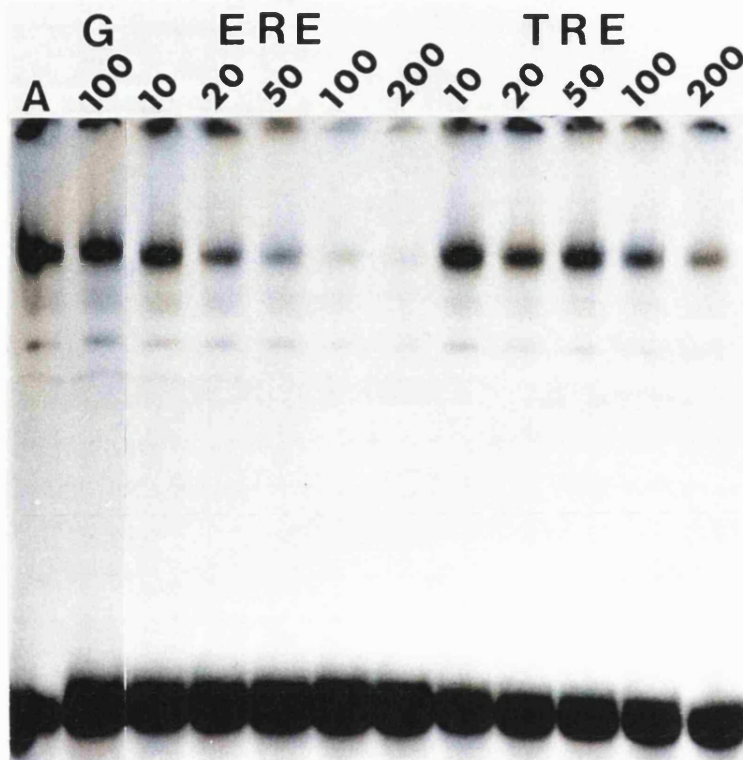
EMSA performed with site B oligonucleotide as probe in competition with molar excess of unlabelled oligonucleotides containing the ERE from the *Xenopus* vitellogenin A2 promoter or the TRE from the rat growth hormone promoter as indicated. Lane 'A': no competitor; lane 'G': 100-fold excess of an oligonucleotide containing a consensus GRE.

Oligonucleotide sequences:

ERE: 5' CTAGAAAGTCA**GGTCA**CAG**TGACC**TGATCAAT 3'

GRE: 5' CTAGAAAGTCA**GAACA**CAG**TGTTC**TGATCAAT 3'

TRE: 5' CGATCGTGC**GGTCA**CGTCCC**TGATC**TTACCTTTCCA

Site B: 5' TCGACAGG**GGTCA**GATATCCAC**TGACC**TTC 3'

142

3.2.3.2 Methylation interference assay (Siebenlist and Gilbert 1980)

The chemical methylation of specific guanine residues in the target DNA sequence will interfere with protein binding if the methylated guanine residue is an important component of the binding site.

Figure 3.6 shows the result of a methylation interference assay using an end-labelled NRE *Kpn1-Rsa1* '220' fragment as probe. A symmetrical pattern of interference was revealed (arrows) and the corresponding nucleotide sequence shown alongside the MIA.

Methylation of guanine residues, indicated by asterisks, interrupted protein binding suggesting that these residues were involved in the close contact of protein with site B. The pattern of methylation of guanine residues seen in the HREs showed that guanines in the palindromic elements were important for protein binding but that guanine bases in the 'spacer' region were not involved in protein binding.

Fig. 3.6

Methylation interference on the NRE '220' fragment.
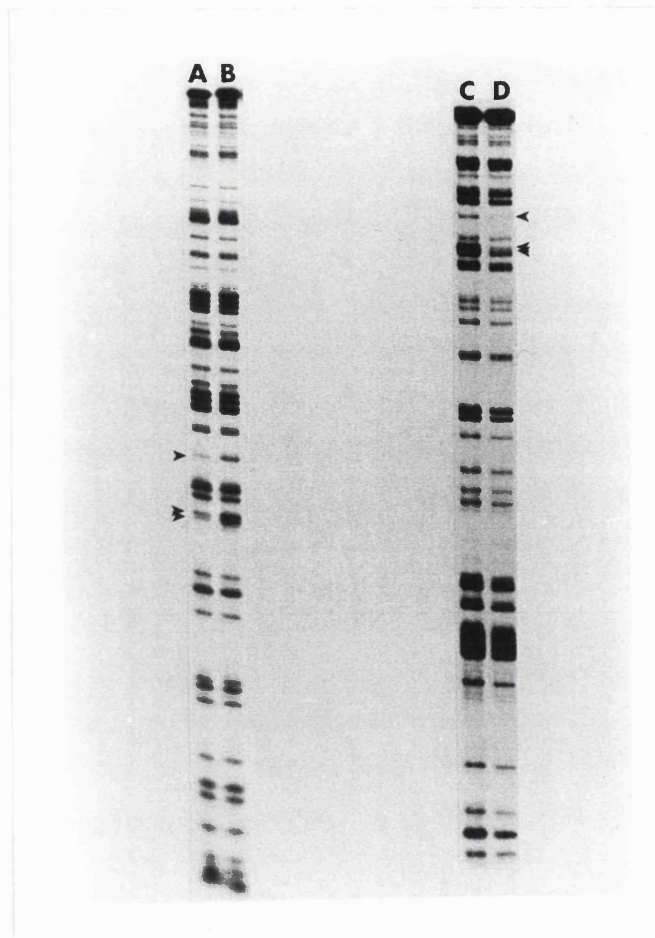
Methylation interference on the NRE complex eliminated by competition with site B oligonucleotide (in fig3.3). Lanes A and B, NRE '220' frgament labelled on the anti-sense strand; lanes C and D, NRE '220' fragment labelled on the sense strand; either unbound DNA (lanes B and C) or DNA recovered from the protein-DNA complex. Arrows indicate reduction of intensity of a band associated with the complex corresponding to G-residues intimately associated with protein in the complex. (performed by N. Perkins)

3.2.3.3   **Palindrome mutant EMSAs.**

The results of methylation interference assays highlighted the importance of guanine residues within the site B footprint. These residues are present as part of the sequence TGACC which is repeated but inverted with a nine base pair spacer sequence.

Site B: GGGCCAGGGGTCAGATATCCACTGACCTTTG

The guanine residues involved in protein contact (or the complementary cytosines) are highlighted in red.

As can be seen from figure 3.12 a general characteristic of the HREs is the presence of the consensus half-site duplicated as an inverted repeat, an everted repeat or direct repeat. To determine the importance of the component palindromes of site B, mutations were introduced changing either the 5' or 3' half of the palindrome (oligonucleotides Bm5 and Bm3 respectively) and also the spacer sequence (oligonucleotide BmS).

Site B:	5'	tcgaCAGG**GGTCA**GATATCCAC**TGACC**TTc
Bm5:	5'	tcgaCACACTAGCGATATCCAC**TGACC**TTc
Bm3:	5'	tcgaCAGG**GGTCA**GATATCCACGCTAGTGc
BmS:	5'	tcgaCAGG**GGTCA**CTAGCACTA**TGACC**TTc

Results of EMSAs using either labelled site B, Bm5 or Bm3 oligonucleotides as probes are shown in figure 3.7. Results of the complementary experiment, competition EMSA with labelled site B oligonucleotide as probe competing with increasing concentrations of unlabelled Bm5 or Bm3 are shown in fig. 3.8. This set of experiments clearly demonstrated the importance of both halves of the palindrome for maximal binding of the T-cell protein to site B, with only a very weak complex forming when the Bm3 oligonucleotide was used as probe and virtually no complex formation with Bm5. From the competition titration (fig 3.8), the relative affinity of the T-cell protein for the mutated sequences could be determined. Removal of the 5' half of the palindrome decreased binding by approximately 100-fold while loss of the 3' half resulted in a forty-fold decrease.

Fig. 3.7

EMSA with mutations in 5' or 3' palindrome.
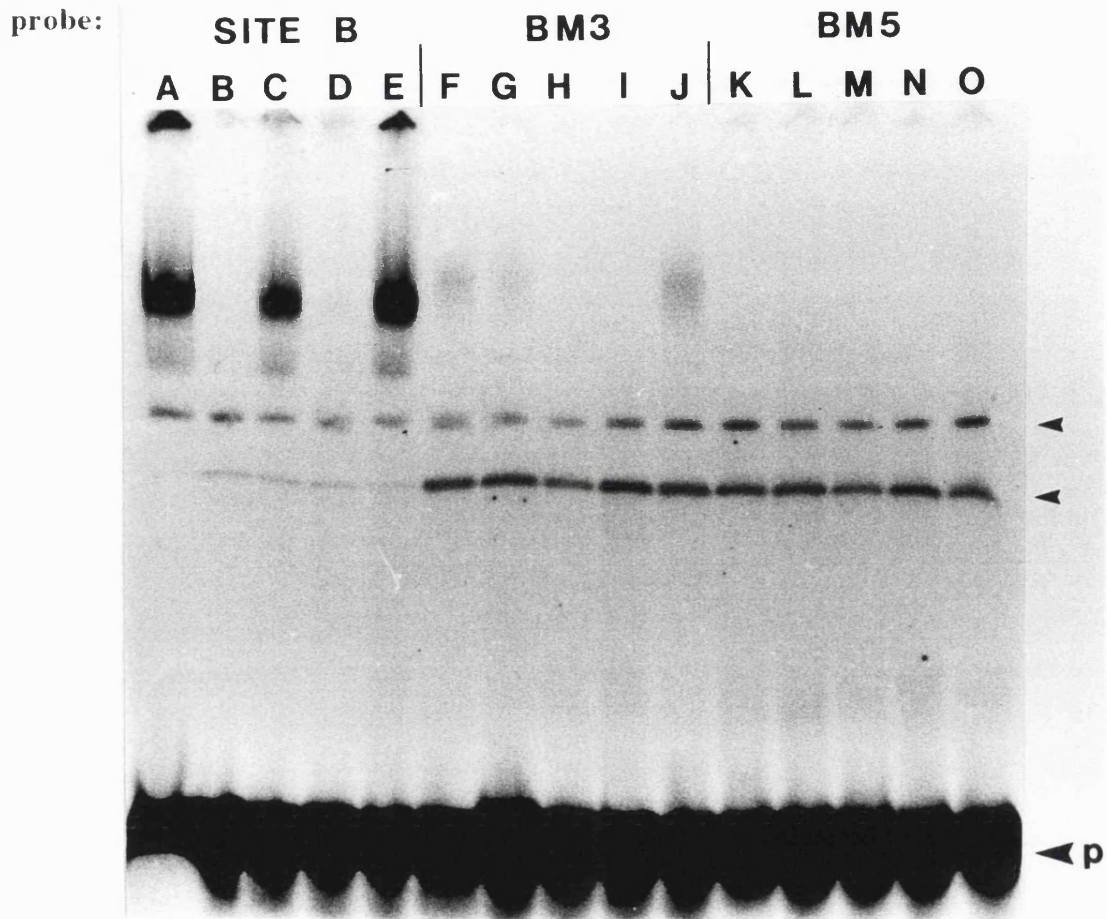
EMSA using site B oligonucleotide as probe (lanes A to E); BM3, an oligonucleotide with the 3'-TGACC sequence changed (lanes F to J); BM5, an oligonucleotide in which the 5'-TGACC has been changed. Additions to the EMSA were as follows: no competitor DNA (lanes A, F anf K); unlabelled site B oligonucleotide (lanes B, G and L); BM5 oligonucleotide (lanes C, H and M); BM3 oligonucleotide (lanes D, I and N) or AP1 oligonucleotide (lanes E, J and O); all competition EMSAs were performed with a 100-fold excess of unlabelled oligonucleotide.

Fig. 3.8

EMSA competition titration with BM3 and BM5.
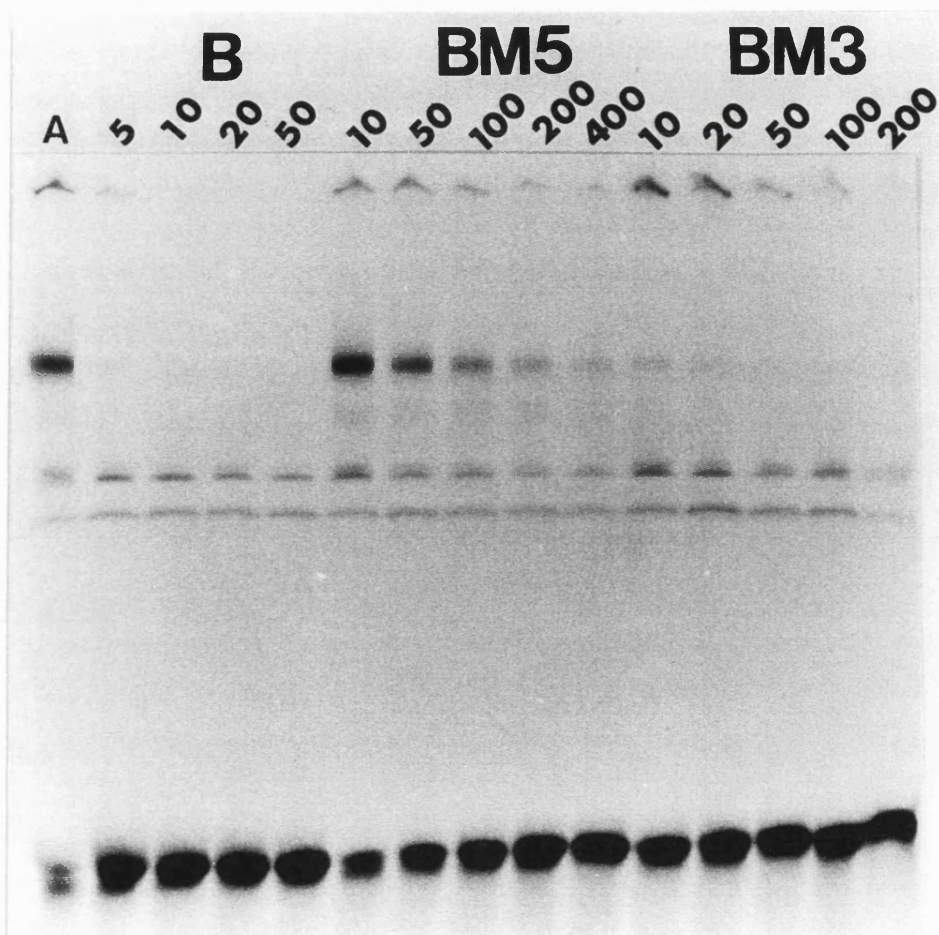
EMSA using labelled site B oligonucleotide as probe and Jurkat nuclear extract. Competitions with titrated excess of unlabelled site B, BM5 or BM3 oligonucleotide as indicated above lane. Lane A, no unlabeled DNA added.

3.2.3.4   Effect on EMSA of spacer region changes.

Fig. 3.9 shows the result of EMSA with oligonucleotide BmS used as probe and as an unlabelled competitor. There was no difference in the appearance of the major DNA-protein complex formed by Jurkat nuclear extract with either site B or BmS oligonucleotides (lanes E to G and A to D respectively). There appeared to be a very small change in the mobility of the minor bands with BmS. Unlabelled site B or BmS competed the major band equally but a faint band remained in the region of the minor bands when unlabelled site B oigonucleotide was used to compete labelled BmS (lane C). The results shown in fig. 3.9 clearly indicate that the sequence of nucleotides within the spacer between the inverted repeat TGACC were not required for the interaction of the majorJurkat nuclear protein with site B. However bases within the spacer region of site B may have been involved in the interaction of other less abundant proteins present in Jurkat nuclear extracts. The effect of the changes to the site B sequences in oligonucleotide BmS appeared to alter the mobility of these minor bands, but as no titration EMSAs were performed the effect of the sequence changes on the affinity of the binding of proteins represented in the minor bands could not be determined.

3.2.4   Functional significance of site B sequences.

Identification that the seven base-pair mutation present in Bm5 oligonucleotide abolished protein binding, allowed the effect of the protein/site B interaction on HIV-1 LTR transcription to be specifically analysed. The same base-pair changes present in the Bm5 oligonucleotide were introduced into the plasmid LTR-CAT by site-directed mutagenesis (Amersham). Because of the potential negative effect of site B on transcription, the experiment was conducted in Jurkat-tat cells. These are a derivative of the Jurkat T-lymphocyte cell line which constitutively expresses the HIV-1 transactivating protein tat, which allows a basal level of LTR-driven expression could be detected. Plasmid constructs were transfected into Jurkat-TAT cells with DEAE-dextran and extracts prepared for promoter driven CAT assay 48 hours later.

The results from a series of transfections using either LTR-CAT or Bm5.LTR-CAT are shown in fig. 3.10. An approximately two-fold increase in CAT expression from Bm5.LTR-CAT over wild-type LTR-CAT was seen over a range of transfected plasmid concentrations. Similar results were obtained with three separate preparations of Bm5.LTR-CAT and LTR-CAT plasmids. (CAT assays performed by Dr. M. Collins).

Fig. 3.9

EMSA with oligonucleotide BmS.
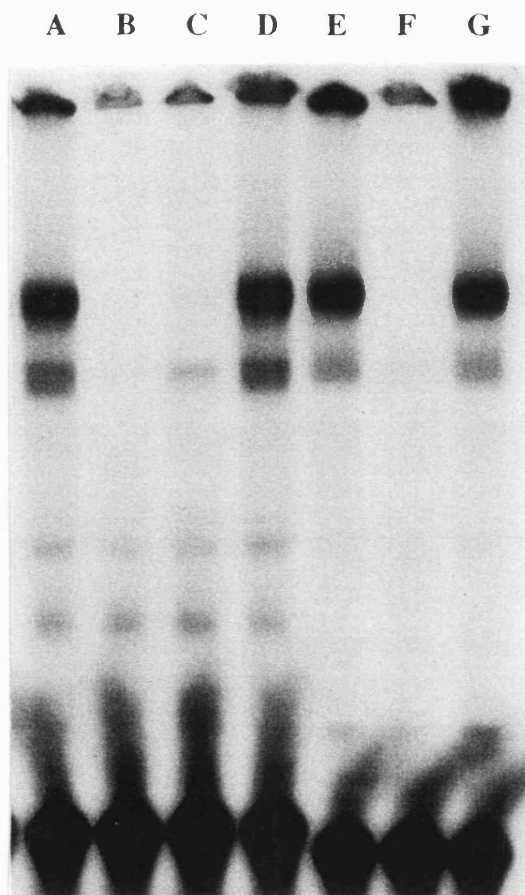
EMSA with labelled oligonucleotide BmS (lanes A to D) or site B (lanes E to G). For competition unlabelled site B (lanes C and F) or BmS (lane B) was included in the incubation mix at a 100-fold molar excess. The effect of an unlabelled non-specific competitor oligonucleotide containing site A sequences is shown in lanes D and G.

Fig. 3.10

CAT activity in Jurkat.TAT cells transfected with LTR.CAT or Bm5LTR.CAT reporter constructs

TRANSFECTION OF LTR-CAT OR
Bm5 LTR-CAT INTO JURKAT.TAT CELLS
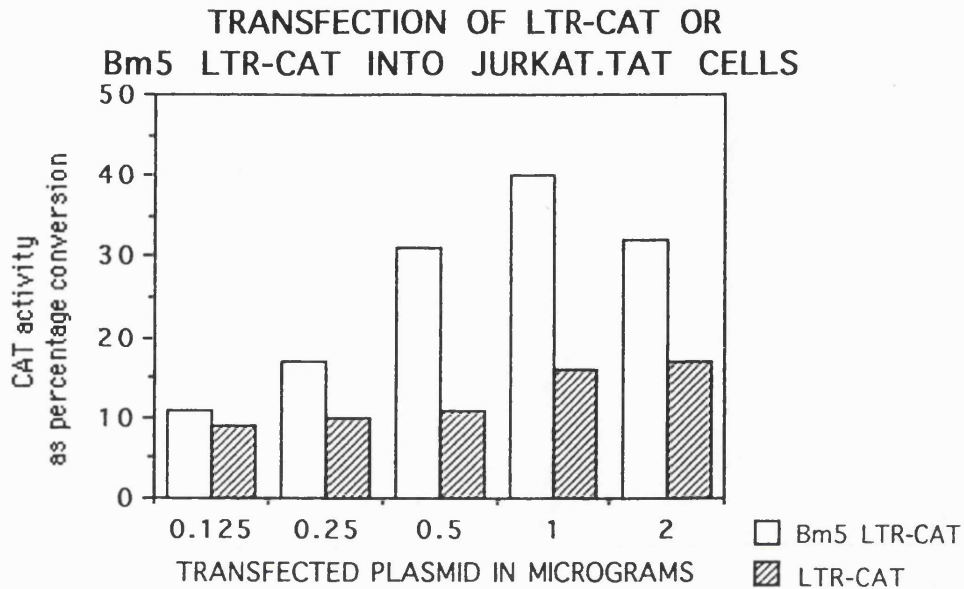
Histogram representing the chloramphenicol acetyltransferase (CAT) activity in Jurkat-TAT cells. LTR.CAT or BM5-LTR.CAT reporter constructs were introduced into Jurkat-TAT cells by DEAE-dextran-mediated transfection. Forty-eight hours after transfection cells were harvested and extracts prepared for CAT activity. Extracts were standardised using β-galactosidase activity. Results are expressed as the percentage of total ^{14}C-labelled chloramphenicol converted into the acetylated form.
(Transfection and assays performed by Dr. Mary Collins).

Fig 3.11 **Region of NRE from HIV-1 LTR of strain HXB2 summarising
DNase I footprinting and methylation interference results.**
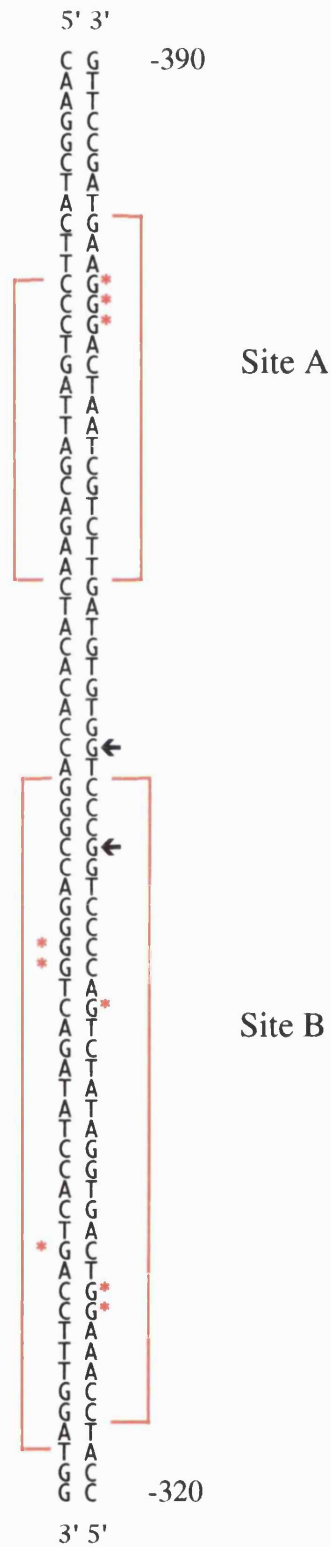
5' 3'

-390

Site A

Site B

-320

3' 5'

Figures indicate the number of nucleotides upstream of the start site for transcription.
Brackets delineate the extent of site A and site B footprints. Asterisks indicate guanine
residues critical for protein binding. The positions of the DNase I hypersensitive sites are
shown by arrows.

3.3 Discussion.

3.3.1. Identification of specific binding sites in the NRE for Jurkat nuclear factors.

The results described in this chapter are summarised in fig 3.11. The position of the two protein binding sites in the NRE are shown, the guanine residues important for protein contact indicated by asterisks.

Initial EMSA with the '220' fragment revealed two major specific protein-DNA complexes, A and B. In addition, two minor complexes were resolved. These had a higher mobility than complex A or B but represented less than 1% of bound probe. The competition pattern of these minor bands paralleled that of the site B complex. Two minor bands were also produced when site B oligonucleotide was used as probe; these may represent the binding of degradation products from the larger site B factor or binding of unrelated, less abundant factors present in the crude extracts. A consistent characteristic of these minor bands that differentiated them from the site B protein-DNA complex was the effect of the combination of detergents, deoxycholate/Brij, which had no effect upon their mobility in EMSA (fig 3.4b lane B). The mobility of a protein-DNA complex through a non-denaturing gel depends upon three characteristics; i) the molecular size of the complex; ii) the conformation; iii) the net charge. The presence of detergents may have altered any one or more of these characteristics for the site B complex, or allowed additional factors present in the crude nuclear extract to interact with the initial complex. Although the exact cause of the detergent effect was not determined it did provide some evidence for the potential of the site B sequence to interact with more than one factor present in the Jurkat nuclear extracts.

DNase I footprinting confirmed the specificity of the interaction and located the sequences within the NRE fragment; complex A formed by protein binding to sequences between -379 to -366 (site A) and complex B by protein binding to sequences -356 to -324 (site B). Results of methylation interference assays identified guanosine bases within each footprint site that were important for protein-DNA interaction. These lay within the palindromic TGACC sequence of site B whereas guanosine residues lying between the palindromic half-sites were not important for protein binding.

Both halves of the site B palindrome were required for maximal binding of the Jurkat protein although mutation of the 5' half affected binding more severely suggesting that this part of the element is more important in complex formation. Similar hierarchical binding to the half-sites of a palindromic

element has been described for the glucocorticoid receptor response element (Tsai et al. 1988) and may reflect the importance of flanking sequences that influence the optimal formation of the protein-DNA complex. The sequences in the spacer between the two halves of the palindrome were not involved in the binding of the Jurkat factor.

3.3.2 Site B does not bind AP1 (fos-jun).

Neither site A nor B had previously been described in the literature, although as discussed in the introduction to this section, two potential AP1 sites with some homology with the consensus AP1 sequence had been identified (Franza et al. 1988). Indeed the site B footprint contained both of these putative AP1 sites and the results of methylation interference were consistent with AP1 binding. However, EMSA in the presence of excess unlabelled AP1 oligonucleotide (fig 3.3 '220' shift lane D and fig 3.4a oligo shift lane F) failed to compete either site A or site B complexes. The presence of AP1 (fos-jun) in the Jurkat nuclear extract was confirmed by using the same AP1 oligonucleotide as the labelled probe in EMSA. A single specific protein-DNA complex was generated, which was markedly increased upon activation of the cells with PMA/PHA. This complex was removed by a 10-fold excess of unlabelled AP1 oligonucleotide but not by a 100 fold excess of either site A or site B oligonucleotides (N. Perkins, personal communication). In contrast the EMSAs obtained with the '220' fragment, site A and site B oligonucleotides were not affected by the activation state of the cells. Antiserum to human fos, which decreased the mobility of the AP1 complex had no effect upon the site B complex (N. Perkins, personal communication). The discrepancy between these results and those published by Franza et al may be due to greater sensitivity of their method, DNA-affinity precipitation, compared with EMSA, indeed the authors point out that substantially less AP1 was recovered with the HIV-1 sequences compared with the other putative AP1 sequences examined, implying that the HIV-1 site only weakly interacted with fos-jun dimers. Further work from this group using in vitro transcription assays with 'wild-type' HIV-1 LTR (strain of HIV-1 not stated) and LTRs with mutations of the 'AP1' sequences did not show any contribution from these sequences to PMA activation of the LTR (Li et al. 1994). In agreement with the results shown in this chapter at least one other independent group has also failed to demonstrate any interaction of AP1 present in nuclear extracts prepared from a variety of cell lines of CNS origin with site B (Canonne Hergaux et al. 1995). These results demonstrate that the sequences in the NRE described as putative AP1 sites do not interact with native AP1 present in nuclear extracts.

3.3.3 Site B sequence analysis.

Comparison of the site B sequence and the methylation interference pattern revealed close similarity to the response elements for several members of the steroid hormone receptor super-family of transcription factors (fig 3.12).

Fig. 3.12

Hormone response elements (HREs):

SITE B:	5'-	CCAGGGGTCAGATATCCACTGACCTTTGG
TRE	5'-	CGTGCGGTCACGTCCCTGATCTTACC
ERE	5'-	AGTCAGGTCACAGTGACCTGAT
COUP	5'-	AAGTTTGACCTTTGACCATAGA
RAREβ	5'-	CCAGGGTTCACCGAAAGTTCACT
RARE-mCRBP	5'-	GTAGGTCAAAAGGTCAGA
RARE apo-A1	5'-	ACTGACCCCTGACCCCTGCCT
RARE γF-crystallin	5'-	AGTGACCCTTTTAACCAGGTCAGTGA
SITE B:	5'-	CCAGGGGTCAGATATCCACTGACCTTTG

Competition EMSA using oligonucleotides containing consensus ERE or TRE confirmed that the Jurkat factor was able to interact with these sites. These results raised the possibility that site B could function as an HRE, with closest homology to the steroid/thyroid/retinoic acid receptor response elements but with a unique spacing of the palindromic half-sites. The corollary of this finding is that the Jurkat nuclear protein binding to site B may be a member of this group of transcription factors, either one of those previously identified (e.g. ER, TR, RAR, RXR, COUP-TF) or a new, uncharacterised member of this group. The results of experiments to discriminate between these possibilities are given in the subsequent chapters.

Since the initial description and characterisation of the two protein binding sites A and B in the region of the HIV-1 LTR from -327 to -379 (Orchard et al 1990) several groups have confirmed the original findings using a variety of techniques. Using *in vivo* footprinting, DeMarchi *et al* delineated protein-binding sites on the LTR of HIV$_{HTLVIIIB}$ in H9 cells (DeMarchi et al. 1992) and demonstrated protection of guanine residues within the site B sequences. They did not however show protection over site A, but this may have been due to their choice of primers which extended from -390 to +20 on the coding strand and only -343 to +20 on the noncoding strand. Interestingly, when they repeated the footprinting in U1 cell line, which was derived from the promonocytic cell line U937 chronically infected with HIV (Folks et al. 1988), no footprints were obtained over sites A or B. Feng *et al* demonstrated the binding of factors from U937 nuclear extracts to a site B oligonucleotide resolving four to six proteins that specifically bound to the palindrome (Feng et al. 1993). Swingler et al used nuclear extracts from neural cell lines for foot-print analysis of the LTR from HIV$_{LAI}$ (Swingler et al. 1994). They demonstrated DNAse footprints over sites A and B with nuclear extracts from two neuroblastoma cell lines and an astrocytoma line, the site A protected region was slightly smaller than that produced with a Jurket extract while site B was evident by two discreet regions of protection, from -355 to -346 and -337 to -326, separated by a site of strong DNAse I sensitivity. This pattern of protection agrees with that obtained with Jurkat nuclear extract and confirms the importance of the palindromic elements and the lack of protein interaction with the spacer region. They also reported the importance of the intact two halves of the palindrome in EMSA using neuronal cell extracts. Canonne-Hergaux et al obtained similar footprints in the regions -363 to -380 (site A) and -352 to -324 (site B) with an Aval-Kpn1 fragment of LTR from either the lymphotrophic HIV$_{LAI}$ or the neurotropic strain HIV$_{JR-FL}$ using nuclear extracts from Jurkat, HeLa and neuronal cell lines SK-N-MC (neuroblastoma), TC-620 (oligodendroglioma), and U373-MG (astrocytoma). The attempts to characterise the cellular factors binding to site B by these and other groups will be discussed in detail in chapters 4 and 5.

Chapter 4

Characterisation of factors interacting with site B.

Chapter 4:

Characterisation of factors interacting with site B.

4.1.1 Introduction.

Initial investigations on site B indicated the importance of the five base-pair sequence TGACC, repeated and inverted with a nine base-pair spacer sequence. The spacer sequence was not required for maximal binding of the major protein present in Jurkat nuclear extract. The palindromic sequence shows homology to the known thyroid/steroid hormone response elements (HREs), the consensus binding sites for the thyroid and oestrogen receptors were effective in competing with site B for the binding of the major Jurkat nuclear protein. This chapter describes the experiments designed to determine some of the physical characteristics of the Jurkat nuclear protein involved in this interaction, allowing comparison with known transcription factors. Ultimately, the nature of the Jurkat protein could only be fully determined by identifying and cloning the corresponding gene. Characterisation of the physical properties of the protein was also important in providing information helpful in large scale purification as a prelude to isolating the cDNA encoding the Jurkat protein.

The molecular size of the protein was determined by three separate methods;
1. UV-cross-linking;
2. an adaptation of the Ferguson plot (Bryan 1977);
3. gel renaturation (performed by Dr. G. Lang).
The tissue distribution of factors interacting with site B was determined using nuclear extracts from HeLa (a cervical carcinoma cell line), U937 (a myelo-monocytic cell line) and IL-2 expanded peripheral blood T-lymphocytes.

4.1.2 COUP-TF and site B.

After the initial description and publication of site A and B footprints (Orchard et al 1990), an apparent homology of site B to a transcriptional control element contained within the chicken ovalbumin upstream promoter was described (Cooney et al. 1991). This element had previously been used to isolate and identify a new subgroup of the thyroid/steroid hormone superfamily of transcription factors, termed COUP-TFs (chicken ovalbumin upstream promoter transcription factors) (Sagami et al. 1986; Wang et al. 1989)

In addition, partially purified COUP-TF derived from HeLa cells was shown to bind specifically to the site B sequence in a manner identical to the binding to

ovalbumin COUP element. Human COUP-TF had already been independently cloned from a human embryo fibroblast library via homology with human erb A and termed Ear-3 (Miyajima et al. 1988). A closely related transcription factor, COUP-TF II has been cloned from a placental library and was initially termed ARP-1 (apoprotein A1 regulatory protein) (Ladias and Karathanasis 1991).

COUP-TF was first described as a positive regulator of the chicken ovalbumin gene (Wang et al. 1989) and requires the co-factor S300-II to transactivate in this context (Sagami et al. 1986; Tsai et al. 1987). S300-II lacks specificity for binding to the ovalbumin promoter but it is required for efficient activation of transcription by COUP-TF. Although initially shown to act as a positive effector, subsequent work has revealed that the members of the COUP-TF sub-family function primarily as transcriptional repressors (Cooney et al. 1993; Cooney et al. 1992; Kliewer et al. 1992; Tran et al. 1992; reviewed by Tsai and Tsai 1997).

COUP-TF has been shown to be present in HeLa cells (Pastorcic et al. 1986) and a number of other cell lines and tissues (reviewed by Richie et al. 1990) but not reported to be present in haemopoietic or lymphoid cell lines or tissues. COUP-TFs could be the mediators of the repression exerted by site B as demonstrated in Jurkat cells but this hypothesis needs to be supported by the demonstration of COUP-TF proteins in these cells. Cooney et al suggested that the major DNA-protein complex found in EMSA using Jurkat nuclear extracts and site B oligonucleotides involved a high molecular weight form of COUP-TF with a relative molecular size of 66-74 kD. However this conclusion was based upon evidence from EMSA and super-shift experiments using 'anti-COUP-TF' rabbit antisera, evidence of this type is circumstantial and does not prove that the EMSA complexes seen using Jurkat nuclear extract are due to COUP-TFs.

4.1.3 Polymorphism of site B between different HIV-1 strains.

Comparison of the -350 to -320 sequences of the known HIV isolates on the Los Alamos Data Base (Myers 1991) revealed significant sequence conservation between HIV strains (figure 4.1). HIV-1 displays an unusually high rate of variation across the majority of the genome, any region of relative sequence conservation implies a selective pressure to maintain the sequence within functionally permissible limits. Part of the selective pressure may be due to the fact that site B lies within the coding sequence for Nef, however Nef proteins from various HIV-1 isolates show up to 17% variation of amino-acid sequence, yet there is third base conservation of Nef codons spanning site B.

Furthermore, the Nef sequence is prematurely terminated upstream of site B in several laboratory strains of HIV-1, e.g. HXB2, which would remove the selective pressure for conservation of the downstream sequences unless a further functionally important role for this region existed. A comparison of the palindromic site B sequence with HIV-1 LTR sequences of primary isolates on a sequence data base (December 1999) showed that the perfect palindrome was highly conserved in a large number of isolates from clades A-J.

Fig. 4.1

Sequence comparison across site B for different HIV-1 isolates.

Variant 1:	CCAGG**GGTCA**GATATCCAC**TGACC**TTT
Variant 2:	CCAGGG**A**TCAGATATCCAC**TGACC**TTT
Variant 3:	CCAGG**A**GTCAGATATCCAC**TGACC**TTT

The '220' fragment of the HIV-1 5' LTR was used in the initial EMSAs and DNase I footprint analysis. This region of the LTR was derived from the cloned HIV-1 isolate HXB2 provided by Baltimore et al. However, subsequent sequencing of the cloned LTR provided revealed a single base-pair polymorphism within site B that differed from the published sequence for HXB2. This sequence is shown as variant 1 in figure 4.1, a sequence that is present in HIV-1 isolates HIVBRV, HIVLAI and HIVHAN.

Variant 2 is present in HIVHXB2, HIVMN, HIVSF2 and HIVSF162.

Variant 3 is present in HIV isolates HIVBH10, HIVMNCG and HIVNL43.

In order to determine the significance of site B polymorphism in the 5' LTR, EMSAs were performed using oligonucleotides derived from the sequences of these HIV variants.

4.1.4 Influence of the spacer sequence size on binding of the Jurkat nuclear factor.

The nucleotide residues comprising the spacer between TGACC repeats were not involved in the interaction of the Jurkat nuclear factor with site B. This is consistent with the binding constraints of thyroid/steroid/retinoic acid receptors to HREs. However, as described in chapter 1, the size of the spacer has been shown to be important in determining the specificity of binding favouring different members of the thyroid/steroid/retinoic acid superfamily of transcription factors. The number of base-pairs within the spacer and can change a positive site to a negative site. To investigate the influence of the spacer region size of the binding of the major factor in Jurkat nuclear

extracts, EMSA were performed with labelled oligonucleotides containing the TGACC inverted palindrome separated by a spacer region of varying size, from zero to nineteen nucleotides.

The data presented in this chapter show that the major Jurkat nuclear factor has a relative molecular size (M_r) of 100-110 kD, unlike any previously described member of the thyroid/steroid hormone family of transcription factors. It bound to site B as a dimer $M_r \approx 220$ kD. No evidence was obtained in support of claims for a significant role of COUP-TF in Jurkat cells - Northern blot analysis failed to detect COUP-TF mRNA in Jurkat cells and Western blot analysis failed to detect COUP-TF protein in Jurkat nuclear extracts. The size of the spacer region was shown to influence the binding efficiency of the major Jurkat nuclear factor.

4.2. Results.

4.2.1 Tissue distribution of factors interacting with site B.

In order to determine the tissue distribution of factors interacting with site B, nuclear extracts were prepared from cells lines representing the major CD4-positive cell targets of HIV-1, i.e. T-lymphocytes and cells of the monocyte-macrophage lineage, In addition, nuclear extracts were prepared from HeLa cells, originally derived from a cervical carcinoma, which have been shown to express COUP-TF.

Figure (4.2) shows the results of EMSA using nuclear extracts from the cell lines described above. A similar pattern of bands emerged for all of the cell types tested but there was variation in the relative proportion of bands between cell types. A low mobility band ('A'), which represents the predominant DNA-protein complex formed with nuclear extracts from Jurkat, peripheral blood lymphocytes and U937 cells. Two higher mobility bands ('B' and 'C') representing DNA-protein complexes were present in relatively low levels in Jurkat and U937 extracts, undetectable in PBL nuclear extracts but formed the predominant bands with the HeLa extract. Occasionally in the EMSA with HeLa cell nuclear extract a further high mobility band could be resolved ('D'), this presumably representing a complex formed with a low abundance protein present in HeLa nuclear extracts or derived from one of the more abundant proteins by proteolysis. The U937 nuclear extracts formed a single predominant complex with a mobility identical to band 'C' with lower abundance of complexes corresponding to bands 'A' and 'B'.

159

Fig. 4.2 EMSA with site B oligonucleotide and nuclear extracts from Jurkat, HeLa, U937 and peripheral blood lymphocytes
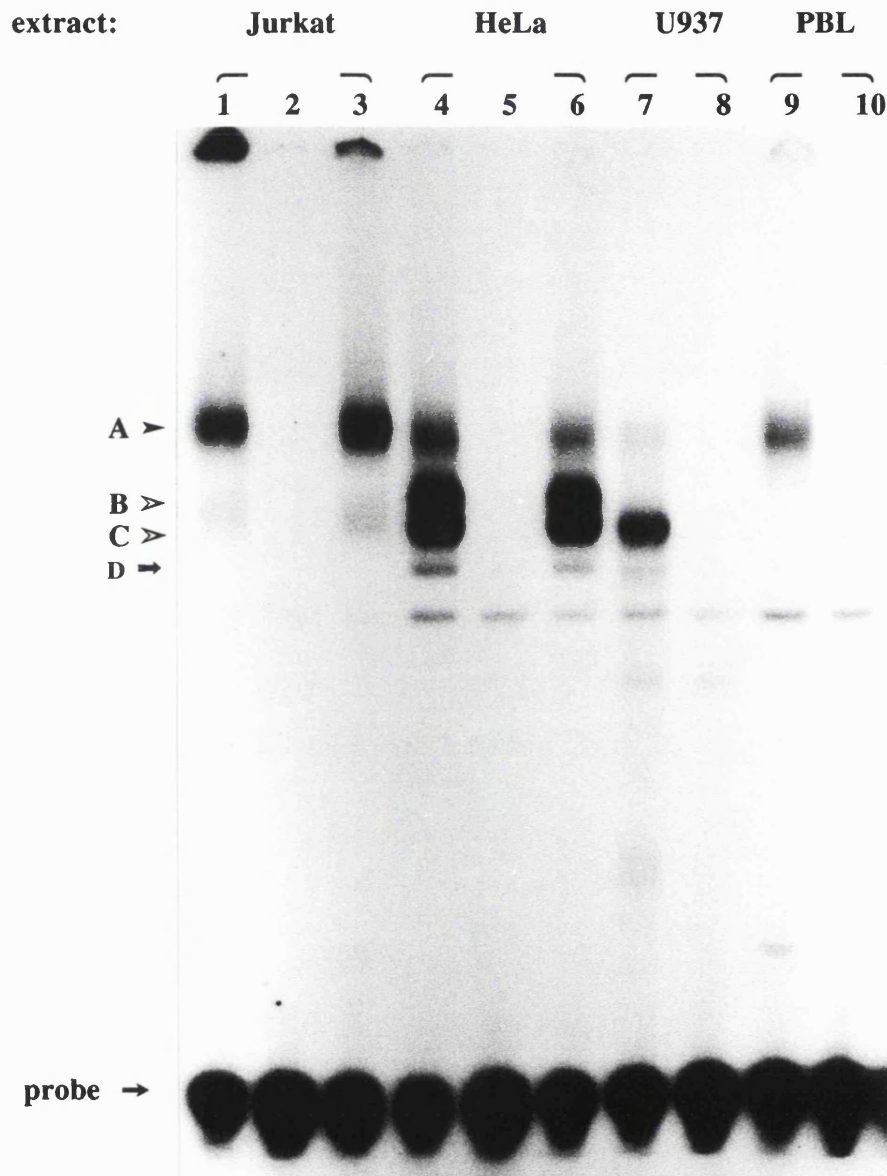
EMSA with labelled site B oligonucleotide as probe and nuclear extracts from Jurkat (lanes 1-3), HeLa (lanes 4-6), U937 (lanes 7, 8) and peripheral blood lmphocytes (lanes 9, 10) as indicated. The major DNA-protein complex formed with the site B oligonucleotide and nuclear extract from Jurkat cells and peripheral blood lymphocytes is indicated by a closed arrow 'A'. The two major complexes formed with HeLa nuclear extract are indicated by open arrows 'B' and 'C'.

Specificity of bands was proven by competition with 50-fold excess of either unlabelled site B oligonucliotide (lanes 2, 5, 8, 10) or an oligonucleotide containing an unrelated sequence (site A oligo in lanes 3, 6). Non-specific band indicated as 'D'.

4.2.2 COUP-TF and site B:

In view of the suggestion that COUP-TF proteins could interact with site B (Cooney et al. 1991) a comparison was made of the binding of Jurkat and HeLa nuclear extracts to site B and the COUP-TF binding site taken from the chicken ovalbumin promoter (Pastorcic et al, 1986 and Sagami et al. 1986).

Site B: CAGG<u>GGTCA</u>GATATCCAC<u>TGACC</u>TTT

COUP-TF: ATGG<u>TGTCA</u>A<u>GGTCA</u>AACTTCT

The COUP-TF binding site from the ovalbumin promoter consists of an imperfect direct repeat of TGACC/A and was used as the basis of a 30 base pair oligonucleotide for use in EMSA.

Figure (4.3) shows the results of EMSA using either site B (lanes 9 to 15) or the COUP-TF binding site (lanes 1 to 8) as a probe with the Jurkat nuclear extract (lanes 5 to 8 and 12 to 15) or HeLa cell nuclear extract (lanes 1 to 4 and 9 to 11). The pattern of complexes formed with Jurkat nuclear extract was identical with both probes, a major low mobility band 'A' and two higher mobility complexes 'B' and 'C'. However the HeLa nuclear extracts produced a different pattern of complexes with the two oligonucleotide probes compared to that using Jurkat nuclear extracts with a faint band of similar low mobility at 'A' but strong bands corresponding to 'B' and 'C'. These data suggest that similar or identical Jurkat and HeLa cell nuclear proteins interact with both site B and a COUP-TF binding site but the relative abundance of these proteins differs between the cell types.

4.2.3 Supershift with 'anti-COUP-TF'.

EMSA can only provide indirect evidence for the comparison of proteins binding to any particular sequence of DNA, there is a danger that too great an emphasis is placed on such results and data over interpreted. A frequently used adjunct to EMSA is the supershift of a DNA-protein complex induced by the addition of antisera specific for one of the proteins involved in forming the complex. However the antibody must be of defined specificity, a suitable antigen to raise a specific antiserum would be a synthetic peptide or recombinant protein expressed in bacteria, yeast or in a Baculoviral system.

Fig. 4.3 EMSA with site B and COUP-TF oligonucleotides
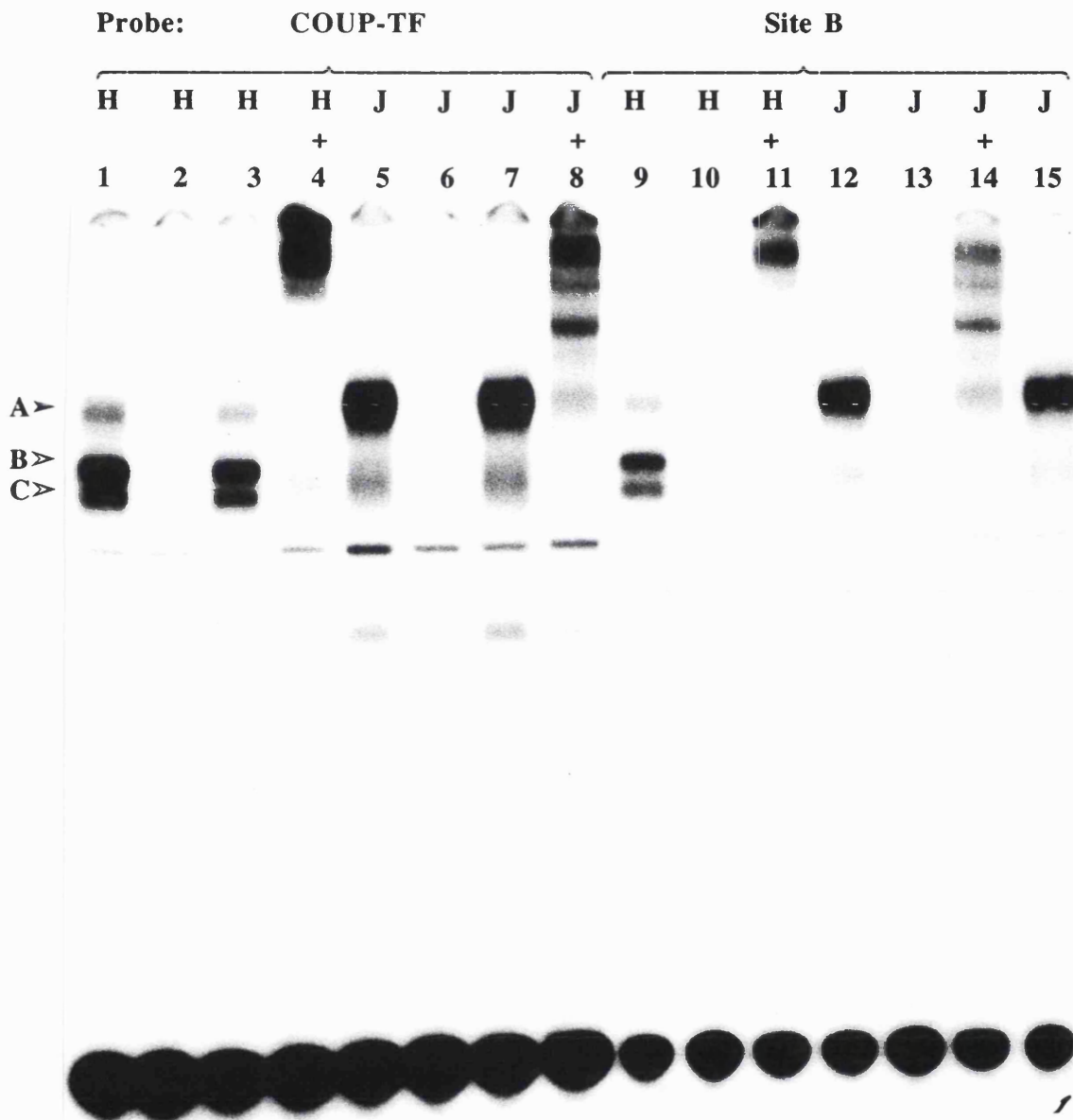
EMSA with labelled site B oligonucleotide (lanes 9-15) or an oligonucleotide containing a consensus COUP-TF binding site ('COUP-TF' lanes 1-8) as probe and nuclear extracts from Jurkat (lanes 5-8 and 12-15) and HeLa (lanes 1-4 and 9-11) as indicated. 1µl of rabbit 'anti-COUP-TF' anti-serum was included in the binding reactions indicated by '+' sign above the lane. Competition with 50-fold excess site B (lanes 2, 6, 10, 13) or an unrelated oligonucleotide (site A in lanes 3, 7, 12, 15).

Details of oligonucleotide probes:
        site B: 5'TCGACAGGGGTCAGATATCCACTGACCTTC3'
        COUP-TF: 5'TCGAATGGTGTCAAGGTCAAACTTCTTCGA3'

162

The results of using an 'anti-COUP-TF' antiserum (gift of Dr. M-J. Tsai) on the Jurkat and HeLa cell complexes is shown in fig 4.3 (lanes 4, 8, 11 and 14), all the proteins binding to site B or the COUP-TF site were supershifted. On first inspection this could be interpreted as strong evidence for the major proteins binding to site B to be COUP-TF and other proteins sharing similar antigenic determinants. However, the 'anti-COUP-TF' antiserum was raised against the proteins retained on a column containing a concatermerised COUP-TF DNA binding site (Wang et al. 1989) and will thus react with all proteins capable of binding to such a site regardless of any relationship at the level of protein sequence. To identify a DNA-binding factor by the similarity of the binding characteristics with another factor is inaccurate and misleading. Consequently the interpretation of the antiserum induced supershift result were unfortunately limited.

4.2.4 Western and Northern blots.

Having established the presence of DNA-binding proteins in Jurkat nuclear extract binding in a similar manner to both site B and a COUP-TF binding site it was important to determine whether Jurkat cells expressed COUP-TF protein. Fig. (4.4) shows the result of Western blots using Jurkat and HeLa nuclear extracts. The same 'anti-COUP-TF' antiserum was used as in the supershift EMSA, detecting a major 45 kD protein in HeLa nuclear extract in agreement with the published size of COUP-TF and a series of minor bands of up to 103 kD in size. However the 45 kD was not detected in Jurkat nuclear extract although faintly staining bands up to 103 kD were detected as in HeLa extract.

For Northern blots, total RNA was prepared from Jurkat, HL60, U937, THP-1 and, as a control, HeLa. A 350 base pair probe was generated by PCR from the DNA-binding domain of ear-3 (100% homology to COUP-TF 1) and used to probe a Northern blot of the prepared RNA, results shown in fig. (4.5a). Two strong hybridisation bands were detected in HeLa RNA at 4.5 kb and 2.5 kb, similar bands detected in U937 RNA and considerably fainter bands in THP-1 and HL60 RNAs. The 2.5 kb transcript corresponds to COUP-TF I, the 4.5 kb transcript with the published size of ARP-1 (Ladias and Karathanasis 1991) a member of the same steroid/thyroid/COUP-TF family of transcription factors which binds to the regulatory region of the apolipoprotein AI (ApoAI) gene and subsequently shown to be identical to a second COUP-TF factor, COUP-TF II (Wang et al. 1991). No hybridisation band was detected in Jurkat RNA even after long exposure of the filter indicating an absence of COUP-TF expression in these cells. A cDNA probe for the constitutively expressed calcium binding protein p68 was used as a control to confirm RNA loading in each track (Fig. 4.5b).

Fig. 4.4

Western blots with Jurkat and HeLa nuclear extracts
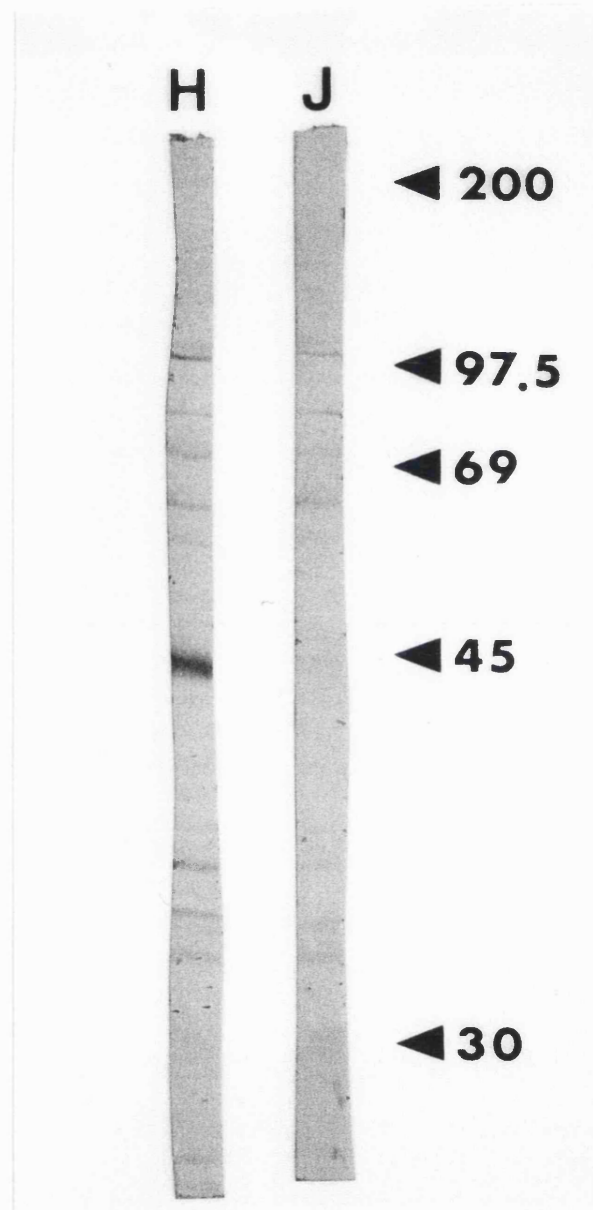
Western blot of Jurkat ('J') and HeLa ('H') nuclear extracts. Extracts were run on a denaturing SDS polyacrylamide gel and electroblotted onto a Immobilon-P membrane as described in section 2.5.6. The resulting blot was probed using the rabbit polyclonal anti-serum raised against a mixture of proteins binding to a COUP-TF site (a generous gift of Dr. M-J Tsai). Bound antibody was visualised with a goat anti-rabbit IgG antibody conjugated to alkaline phosphatase (Promega) followed by incubation with the chromogenic reagent BCIP/NBT. Each blot track represents the equivalent of 10μl of original nuclear extract. Protein size markers as run on the SDS gel and transfered with the nuclear extracts are indicated.

Fig. 4.5

Northern Blots of total RNA from Jurkat, HeLa, HL60 and U937 cells

4.5a
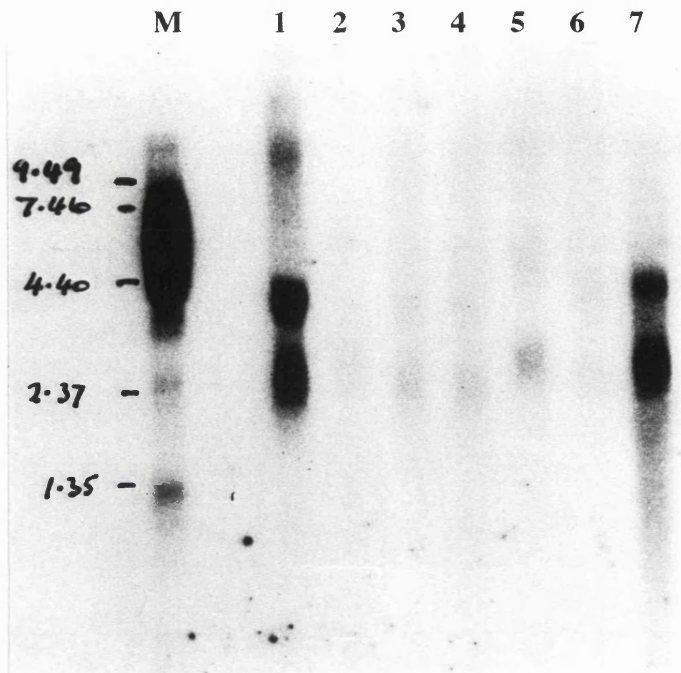
4.5b
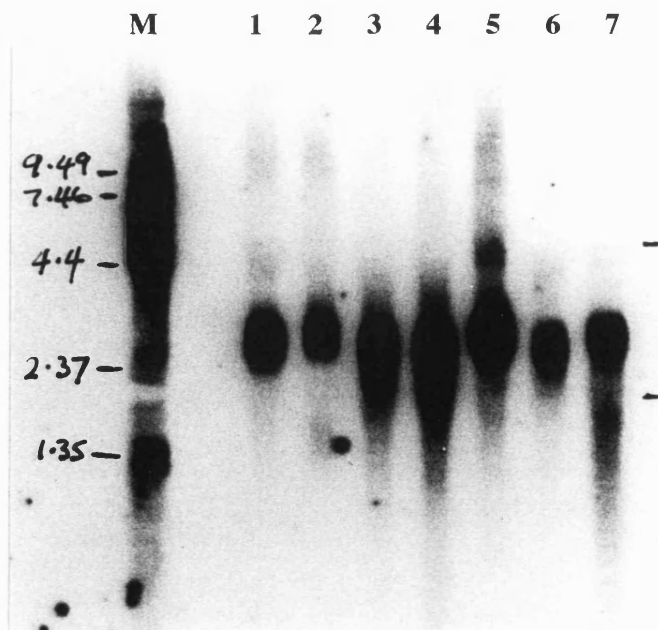
Northern blots probed with either a labelled probe generated from the DNA-binding domain of a COUP-TF cDNA clone (panel A) or a control probe generated from a cDNA for the constitutively expressed calcium binding protein p68 (panel B, a gift of Dr. Mark Lawley). Total RNA was extracted from HeLa (lane 1), Jurkat (lane 2), and the monocyte lineage cell lines HL60 (lanes 3 and 4), THP-1 (lanes 5 and 6) and U937 (lane 7). For HL60 and THP-1 RNA was prepared before and after treatment with all-trans retinoic acid (lanes 3, 5 and 4, 6 respectively). RNA molecular size markers are shown in lane 'M'.

165

4.2.5 UV-Cross-linking.

Cross-linking of Jurkat nuclear protein to bromo-deoxyuridine substituted site B oligonucleotide was performed as described in section 2.2.8. Two methods were employed to determine the molecular weight of the complex under denaturing conditions. In the first, the UV-exposed EMSA mix was heated at 98^0 C for 5' in the presence of 20% SDS, then loaded onto an SDS-polyacrylamide gel and subjected to electrophoretic separation. Following electrophoresis the gel was fixed, dried and exposed to X-ray film.

In the second method, the UV-exposed EMSA mix was initially separated on a low-ionic strength non-denaturing gel, under conditions identical to those used for the standard EMSA The wet gel was then exposed to X-ray film for 6 hours at 4^0 C, the band corresponding to site B protein excised and analysed by electrophoresis on a denaturing SDS-polyacrylamide gel. Following electrophoresis, the gel was fixed, dried and exposed to X-ray film.

Fig. (4.6) shows the results of direct SDS-gel analysis of the UV-exposed EMSA mix with BrdU-substituted site B as probe. A single major complex was formed, the specificity was proven by competition with unlabelled site B oligonucleotide. The relative molecular size of the site B complex was determined from a standard curve using the molecular weight standards. A relative molecular weight of 130-135 kD was obtained and after subtracting 22 kD for the contribution of the oligonucleotide to the total molecular weight, a M_r of 108-113 kD was obtained. Also visible are two fainter bands which were removed by competition with unlabelled site B, with M_r of 80 and 84 kD, subtraction for the contribution from the oligonucleotide probe producing a result of 58 and 62 kD for these minor bands.

Fig. 4.6

UV-cross-linking 1
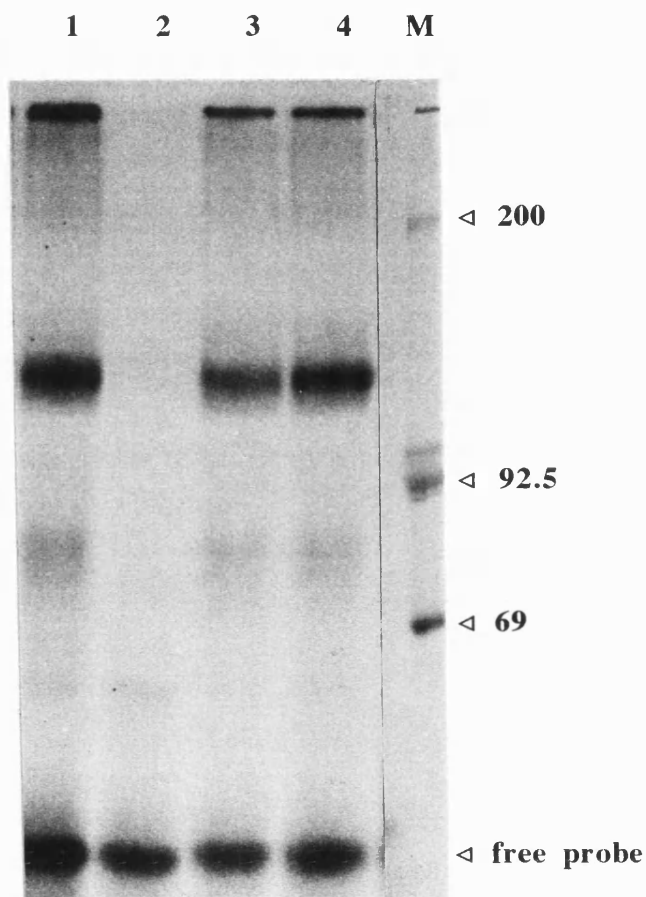
UV-cross-linking of Jurkat nuclear extract with labelled BrdU-substituted site B oligonucleotide as cross-linking probe. Standard EMSA were performed with the BrdU probe, either without a competing unlabelled oligo. (lane 1), or with a 100-fold excess of unlabelled site B (lane 2), site A (lane 3) or AP1 (lane 4) oligonucleotides. then exposed to UV illumination as described in section 2.2.7. The UV-exposed EMSA mix was then run on a denaturing SDS polyacrylamide gel together with ^{14}C-labelled protein molecular weight markers (lane 'M').

Fig. (4.7) shows the results of the second method. Gel bands were excised from EMSA in which either Jurkat nuclear extract (lane 'J') or HeLa nuclear extract (lane 'H') had been used in the binding reaction. The major protein/DNA complex formed between the site B oligonucleotide and Jurkat nuclear extract had an estimated relative molecular size of 95-100 kD which was close to that determined by the direct method. The HeLa nuclear extract formed two predominant protein/DNA complexes with the site B probe, of approximately 55 and 45 kD in size. Similar results were also obtained by gel renaturation experiments in which HeLa and Jurkat nuclear extracts were initially fractionated on a denaturing polyacrylamide gel, proteins from gel slices corresponding to specific molecular weight ranges were extracted, renatured and used in EMSA. The size range 97-160 kD contained the protein(s) which formed the characteristic low mobility complex with site B (G. Lang, personal communication and Orchard and Lang '93). In contrast, the high mobility complex characteristic of the HeLa cell nuclear extract was detectable when protein below 83 kD was used in the assay. Thus, by three separate techniques the predominant nuclear protein in T-lymphocytes that specifically interacts with a site B oligonucleotide was shown to be approximately 100-110 kD in size while the predominant HeLa nuclear proteins were much smaller, in the range of 45-62 kD, in agreement with previous published estimates for the size of COUP (Wang et al. 1989).

4.2.6 M_r of the DNA-protein complex: the Ferguson plot.

UV-induced cross linking between a BrdU-oligonucleotide and protein occur only when the protein and DNA are in intimate contact, with on average a single cross-link forming in each complex. Thus the calculated molecular size of 100-110 kD corresponds to a single polypeptide or monomer binding to site B. Transcription factors can interact with DNA as homo- and/or hetero-dimers, particularly those in the steroid/thyroid/retinoic acid/COUP superfamily. In order to determine whether the Jurkat protein bound as a monomer or part of a larger complex the mobility of the site B complex was compared with that of standard proteins of known molecular weight on a series of polyacrylamide gels of increasing concentration. The results obtained for each protein standard are shown in the series of graphs in fig (4.8). The Ferguson plot of the resulting data is shown in fig. (4.9), from which a molecular weight of 226 kD for the major Jurkat complex and 107.5 and 112.5 kD for the two HeLa cell site B complexes were derived. A similar size of 90 kD has been reported for HeLa cell COUP-TF as determined by either gel filtration or glycerol gradient centrifugation to determine the native molecular size of the protein complex in the absence of DNA (Wang et al, 1987 and 1991, Sagami et al. 1986).

Fig. 4.7

UV-cross-linking 2
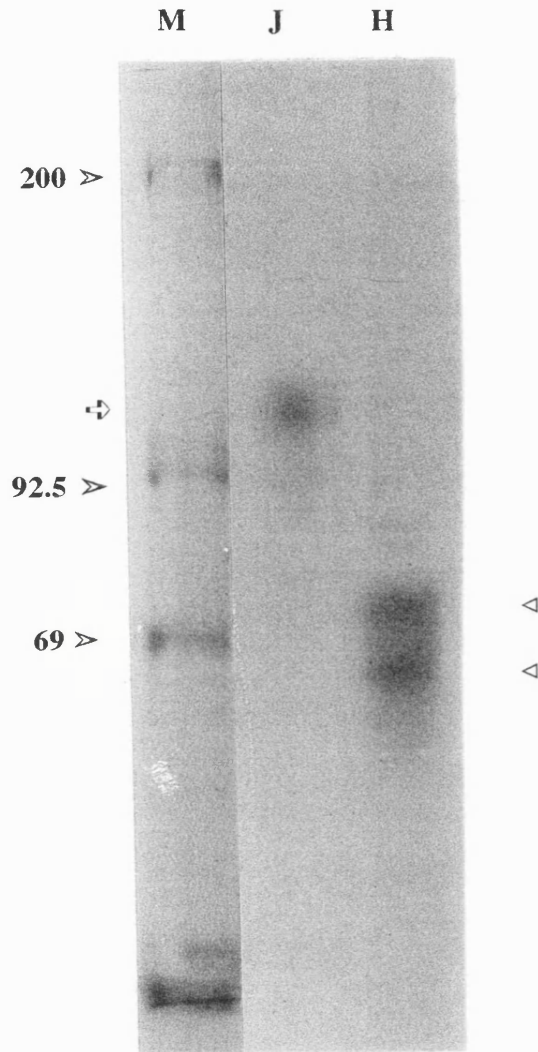
UV-cross-linking with labelled BrdU-substituted site B oligonucleotide as probe using Jurkat ('J') or HeLa ('H') nuclear extracts. UV-cross-linking was performed on a completed EMSA polyacrylamide gel, labelled DNA-protein retarded bands were identified by brief exposure of the wet gel to X-ray film and then excised from the gel. Gel slices containing the crossed-linked DNA-protein complexes were heated in SDS-loading buffer then placed into slots in a denaturing SDS gel taking care to maintain the correct orientation of the slice. ^{14}C-labelled protein molecular weight markers (lane 'M') were run in adjacent tracks.

Figure 4.8

CARBONIC ANHYDRASE

$y = 140.59 - 1.7120x$ $R^2 = 0.782$

CHICKEN EGG ALBUMIN

$y = 197.16 - 2.7760x$ $R^2 = 0.968$

BSA DIMER

$y = 202.08 - 7.6440x$ $R^2 = 0.982$

BSA MONOMER

$y = 195.39 - 4.1160x$ $R^2 = 0.932$

UREASE TRIMER

$y = 238.92 - 20.312x$ $R^2 = 0.980$
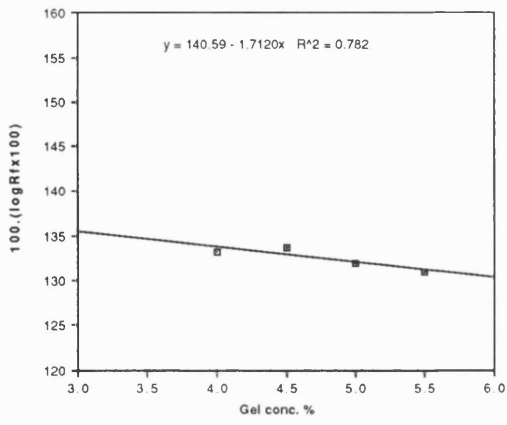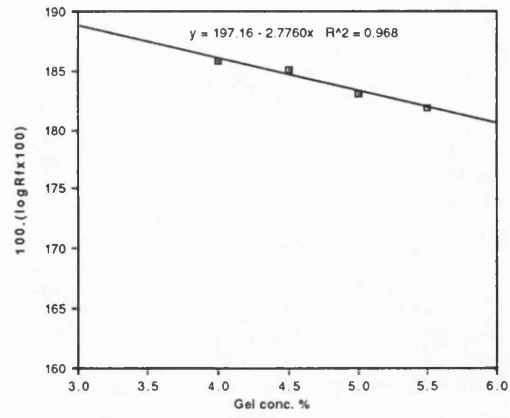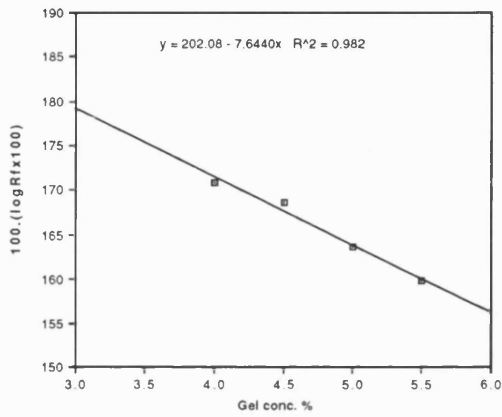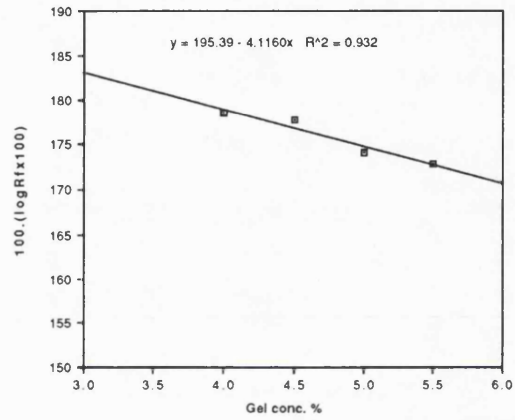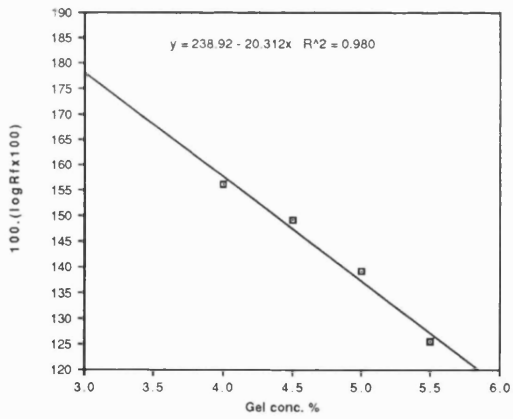
JURKAT SITE B PROTEIN BAND

$y = 217.18 - 16.964x$ $R^2 = 0.994$

HeLa BAND 1

$y = 191.95 - 8.4148x$ $R^2 = 0.998$

HeLa BAND 2

$y = 197.43 - 8.7626x$ $R^2 = 0.998$

170
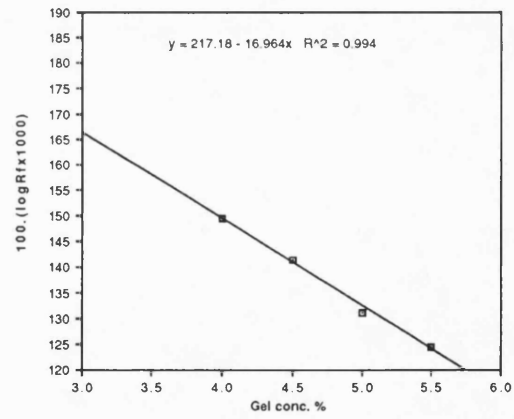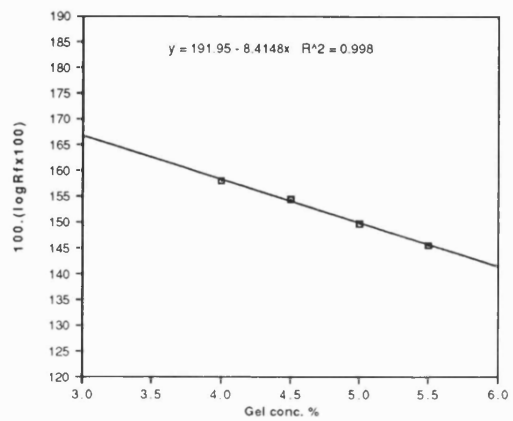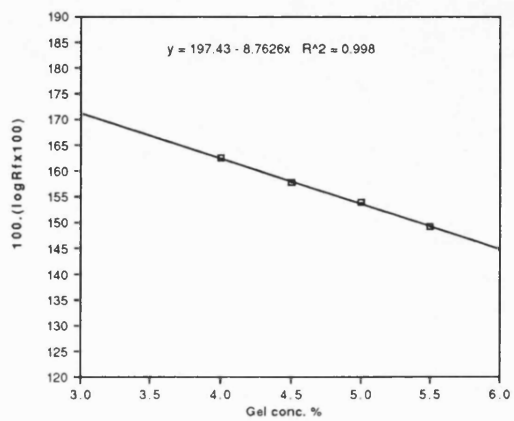
Legend to fig. 4.8

Graphs of standard curves for the Ferguson plot, axes 100.(logRf x 100) vs gel concentration (as percentage of polyacrylamide). Protein standards were run on non-denaturing polyacrylamide gels of varying concentrations. The relative mobility (Rf) of the protein as compared with the running dye was determined for each protein on each gel and the data plotted as shown. Graphs were plotted using Cricket Graph.

The slopes of the resulting straight-line plots were taken from the equation for the graph as generated by Cricket Graph (results shown in figure 4.9).

The Rfs for Jurkat and HeLa DNA-protein complexes were determined on EMSA run under standard conditions, but at varying gel strengths. The position of DNA-protein complexes were found by exposure to X-ray film.

Fig. 4.9   **Ferguson plot data**

Protein standards	Molecular Wt. x1000	Negative slope
carbonic anhydrase	*29*	*1.7*
chicken egg albumin	*45*	*2.78*
BSA monomer	*66*	*4.12*
BSA dimer	*132*	*7.64*
urease	*272*	*20.31*
Band shift results:		
Jurkat	*J1*	*16.96*
HeLa band 1	*H1*	*8.41*
HeLa band 2	*H2*	*8.76*

Panel A: The molecular weight standards with the negative slope of the graphs for 100(logRfx100) vs gel concentration for each of the protein standards (from data in section 2.3.2 table 2.1 and fig. 4.8) and the negative slope of the EMSA DNA-protein bands for Jurkat and HeLa nuclear extracts.

B.   **Derived Ferguson plot**
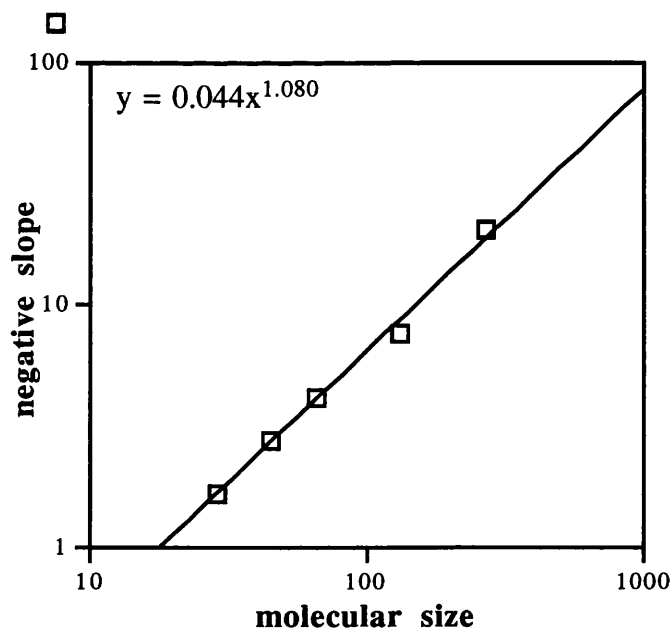
Panel B shows the logarithmic plot of the negative slope vs molecular size x 10^3.
The molecular size of the DNA-protein complexes are found by applying the formula for the curve. The values obtained in this way are 248 for the Jurkat protein and 134.5 for the HeLa H2 and 129.5 for HI. Following subtraction of 22 in order to correct for the contribution from the 32 mer oligonucleotide produces a Mr of 226 for the Jurkat protein and 112.5 and 107.5 for the HeLa proteins H2 and H1.

Formula for the curve: $log\ y = log\ 0.044 + 1.080\ .log\ x$

4.2.7 Effect of site B polymorphism on EMSA.

Figure 4.10 shows the results of EMSA using labelled variant 1 (lanes 9-12) which contains the intact TGACC 5' and 3'-palindromes; variant 2 (lanes 5-8) or variant 3 both of which contain single G → A change in the 5'-palindrome (fig 4.1). Competition EMSA were performed using excess unlabelled variant 1 site B oligonucleotide as indicated. The results show a marked difference in the binding of the Jurkat nuclear protein to site B. In agreement with the results shown in chapter 3, changes to the 5'-palindrome reduced binding of the Jurkat protein. The G→A polymorphism of variant 2 resulted in a marked reduction in the formation of protein-DNA complex while the variant 3 polymorphism produced a complex of slightly lower mobility and approximately 2-fold weaker than that of variant 1.

Figure 4.11 shows the complementary set of EMSAs using variant 1 as labelled probe and competition with unlabelled variant 1, 2 or 3 site B oligonucleotide. These results confirm the relative binding affinities of Jurkat nuclear protein for the three site B sequences:

variant 1 > variant 3 > variant 2.

4.2.8

Effect of spacer sequence size on Jurkat nuclear protein binding.
Figures 4.12a and 4.12b show the results of EMSA using labelled oligonucleotides based on variant 1 site B but with a spacer region between palindrome half sites varying from zero nucleotides, 1 through to 8 and finally 14 and 19.
The efficiency of binding of the Jurkat nuclear protein was similar using oligonucleotides with zero, 1 or 2 spacer region but then reduced for the 3-spacer and markedly reduced with the 4-spacer oligonucleotide. Interestingly, binding was restored with a single base-pair insertion to the 5-spacer oligonucleotide and with 6 or 8 remained similar to the wild-type 9-spacer. Interestingly, binding was maintained even when the spacer region was increased to 14 and 19 base-pairs. For comparison EMSA were performed with oligonucleotides M3 and M5 as described in chapter 3.
In order to quantify the binding, dried gels were exposed to X-ray film and the bands identified and excised, including the free probe. The percentage of activity in a specific band was expressed as a fraction of the total activity per track. These results are displayed as a histogram in figure 4.13, taking the activity in the band formed with the wild type site B spacer of 9 as 100%.

173

These results suggest that the binding of the Jurkat nuclear factor is dependent on the configuration of the two palindrome half-sites. The DNA double helix undergoes a complete turn approximately every nine base pairs. By introducing a four base pair spacer, the orientation of the palindromes is in opposition to that shown by the nine-spacer wild-type site B. If binding of the Jurkat factor was dependent on the spatial orientation of the two half-sites it could only bind at one half-site (probably the 5' half-site) producing a complex on EMSA similar to that formed with the half-site mutation oligonucleotides M3 and M5. This interpretation is consistent with the experimental evidence. The importance of the orientation rather than separation of the half-sites was demonstrated by the efficient binding of the Jurkat nuclear factor to oligonucleotides with zero, 1, 2, 5, 6 and 8 base pair spacer regions. Efficient binding was possible even when the half-sites were further separated, by 14 or 19 base pairs. These would have in theory placed the half-sites in the same orientation. It is interesting to speculate that the next sizes of spacer to place the half-sites in opposition, 12 and 22, would have resulted in a weakened interaction. Unfortunately these experiments were not performed.

Fig. 4.10

Competition EMSA with site B sequence variants
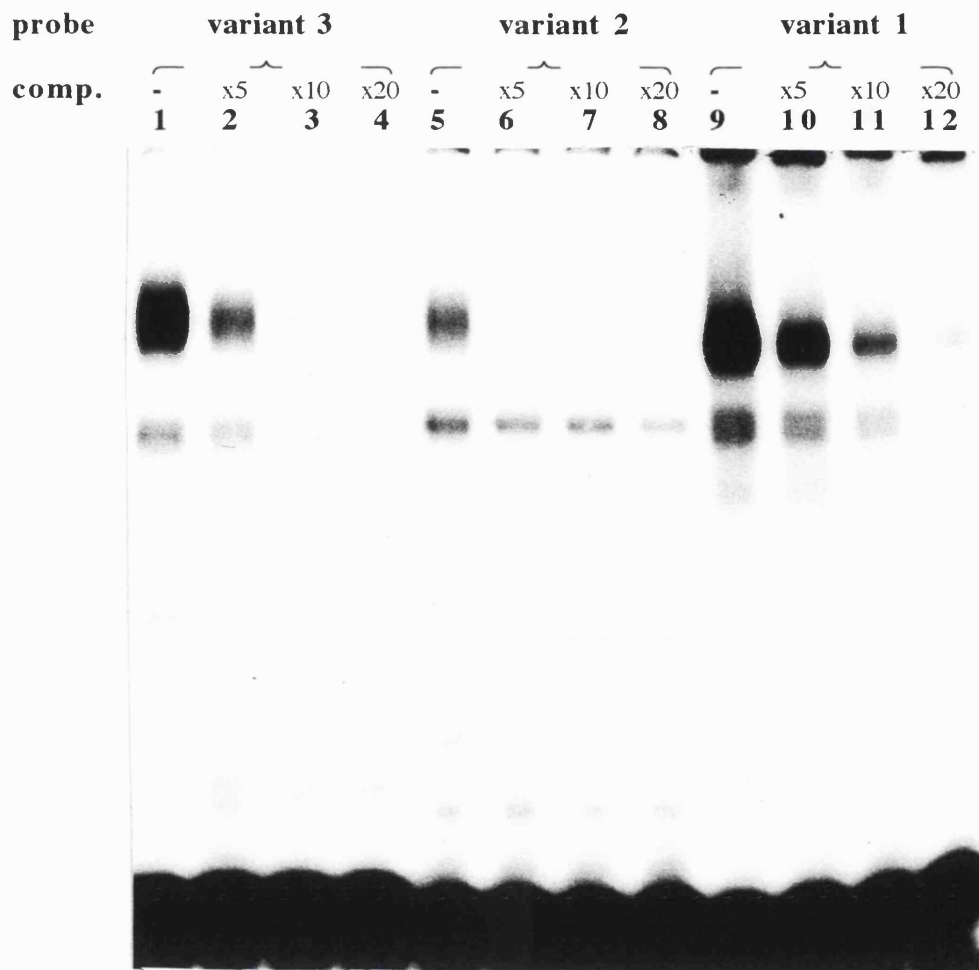
EMSA using labelled site B oligonucleotides with polymorphisms in the 5'-TGACC palindrome. Competitions performed with excess unlabelled site B variant 1 (perfect TGACC palindromes) as indicated above the lanes.

Details of oligonucleotides:

```
variant 1:   5'-TCGACAGGGGTCAGATATCCACTGACCTTC-3'
variant 2:   5'-TCGACAGGGATCAGATATCCACTGACCTTC-3'
variant 3:   5'-TCGACAGGAGTCAGATATCCACTGACCTTC-3'
```

Fig. 4.11

Competition EMSA with site B sequence variants
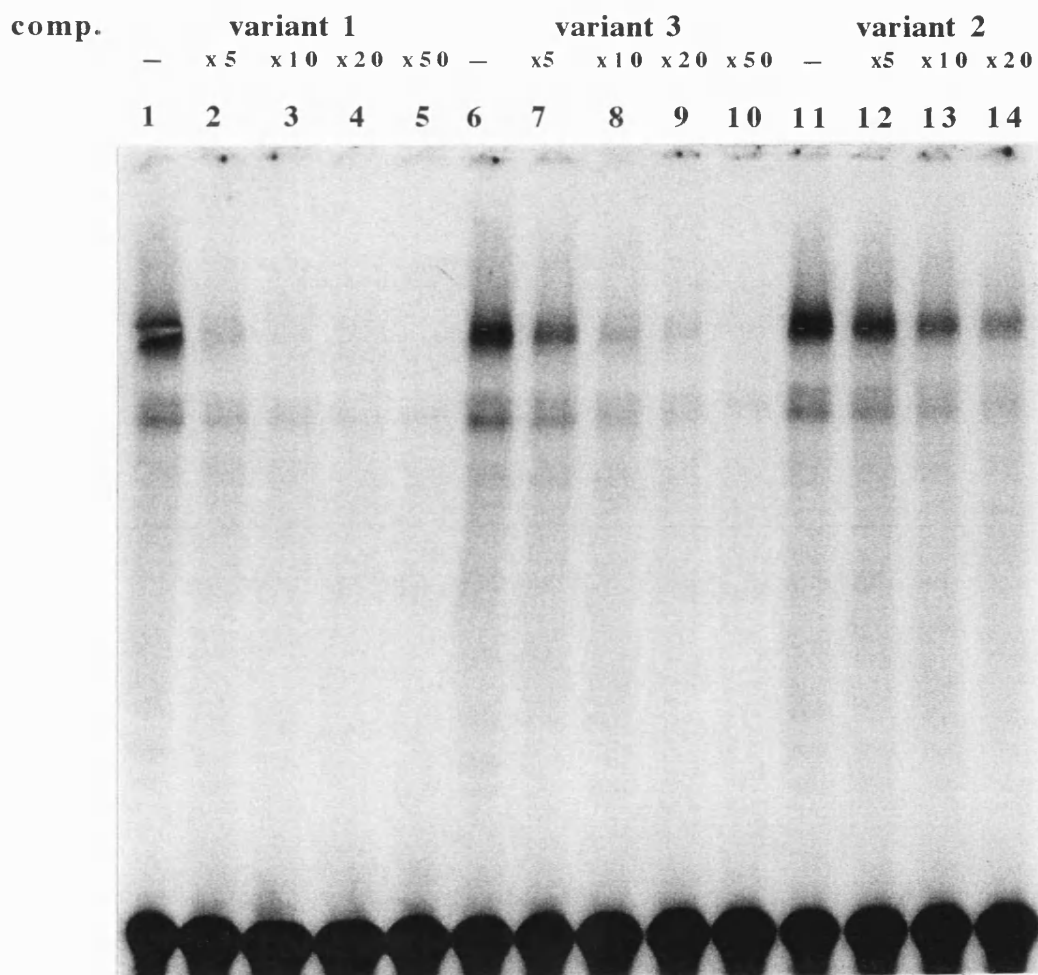
Competition titration with labelled site B (perfect palindrome variant 1) as probe in all lanes with increasing molar excess of unlabelled oligonucleotides for variant 1 (lanes 2-5), variant 3 (lane 7-10) or variant 2 (lanes 12-14). EMSA in lanes 1, 6 and 11 were performed without competing oligonucleotide.

Fig. 4.12a

EMSA with TGACC half site spacer changes
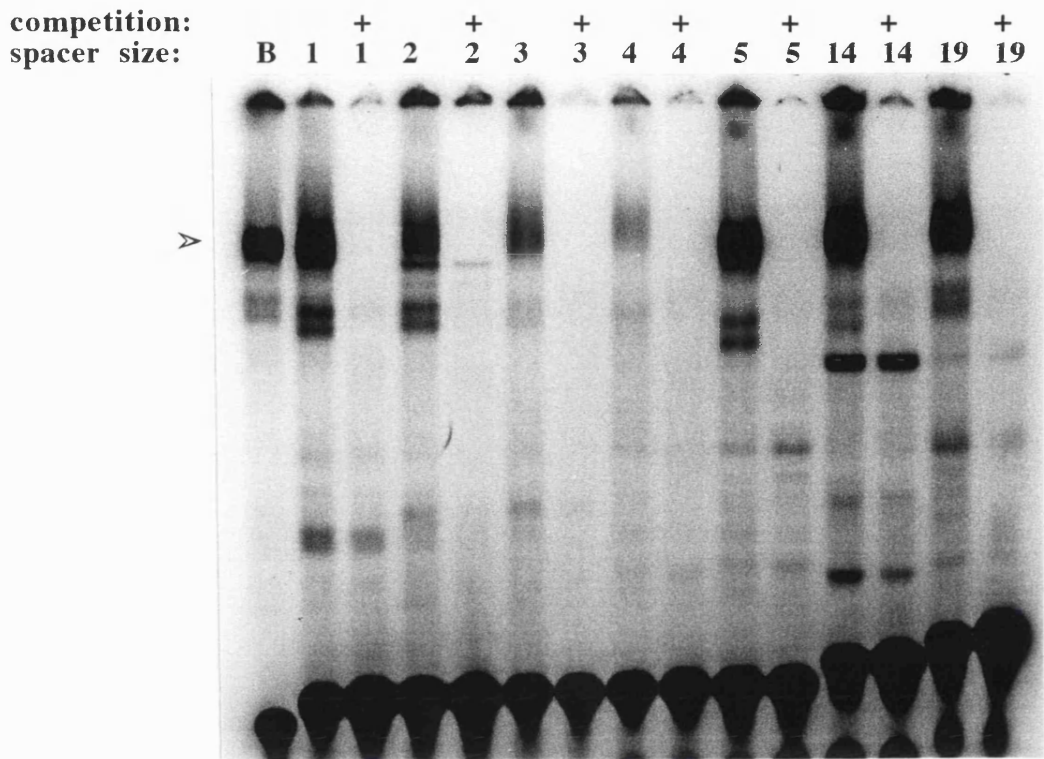
EMSA with TGACC palindrome spacer size changes as indicated above each lane. The wild type site B oligonucleotide (lane 'B') contained a nine nucleotide spacer. Binding reactions were performed in the presence or absence of a fifty-fold excess of unlabelled homologous oligonucleotide as indicated by a '+' above the lane. The arrow indicates the position of the major DNA-protein complex formed with Jurkat nuclear extract.

Fig. 4.12b

EMSA with TGACC half site spacer changes and oligonucleotides m3 and m5.
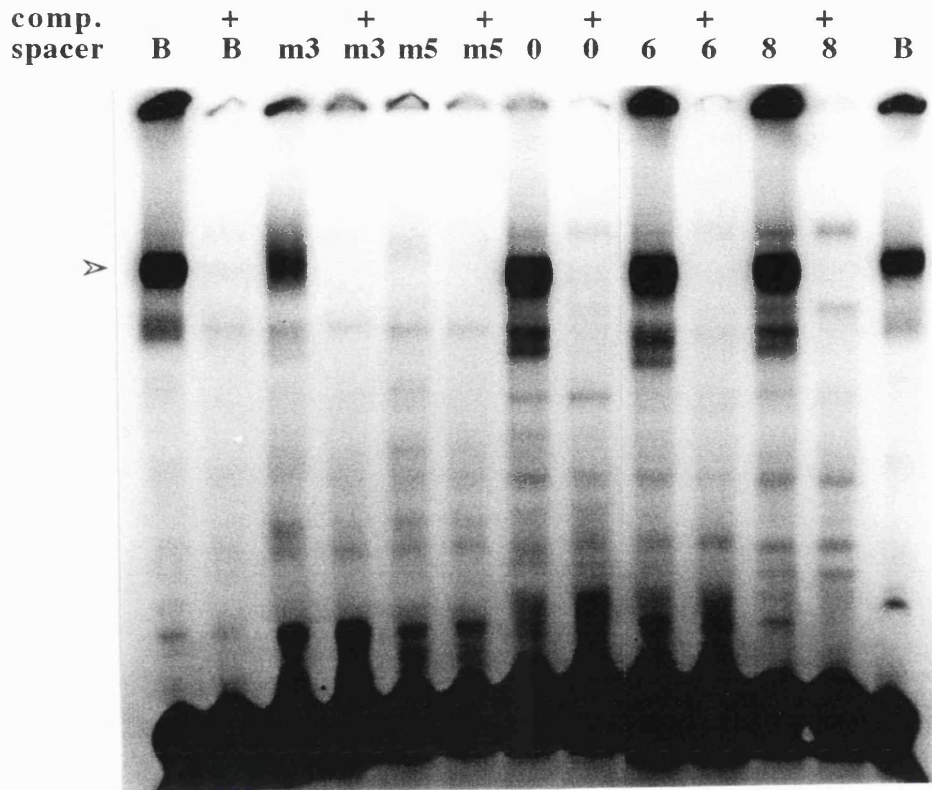
EMSA with TGACC palindrome spacer size changes or half site mutated oligonucleotides m3 and m5 as indicated above each lane. The wild type site B oligonucleotide (lanes indicated as 'B') contained a nine nucleotide spacer. Binding reactions were performed in the presence or absence of a fifty-fold excess of unlabelled homologous oligonucleotide as indicated by a '+' above the lane. The arrow indicates the position of the major DNA-protein complex formed with Jurkat nuclear extract.

Details of oligonucleotides:

Spacer		
	0:	5'-TCGACACACGG<u>GGTCA</u>T<u>GACC</u>TTCATTAGT-3'
	1:	5'-TCGACACACGG<u>GGTCA</u>G<u>TGACC</u>TTCATTAG-3'
	2:	5'-TCGACACAGG<u>GGTCA</u>GA<u>TGACC</u>TTCATTAG-3'
	3:	5'-TCGACACGG<u>GGTCA</u>GAT<u>TGACC</u>TTCATTAG-3'
	4:	5'-TCGACAGG<u>GGTCA</u>GATA<u>TGACC</u>TTCATTAG-3'
	5:	5'-TCGACAGG<u>GGTCA</u>GATAT<u>TGACC</u>TTCATTA-3'
	6:	5'-TCGACAGG<u>GGTCA</u>GATATC<u>TGACC</u>TTCATT-3'
	8:	5'-TCGACAGG<u>GGTCA</u>GATATCCA<u>TGACC</u>TTCA-3'
site B	9:	5'-TCGACAGG<u>GGTCA</u>GATATCCACT<u>GACC</u>TTC-3'
	14:	5'-TCGACAGG<u>GGTCA</u>CCGATATCCACTCTT<u>GACC</u>TTC-3'
	19:	5'-TCGACAGG<u>GGTCA</u>CCGATATCCACTCTCCTGC<u>TGACC</u>TTC-3'

178

Fig. 4.13

EFFECT OF SPACER SIZE ON BINDING
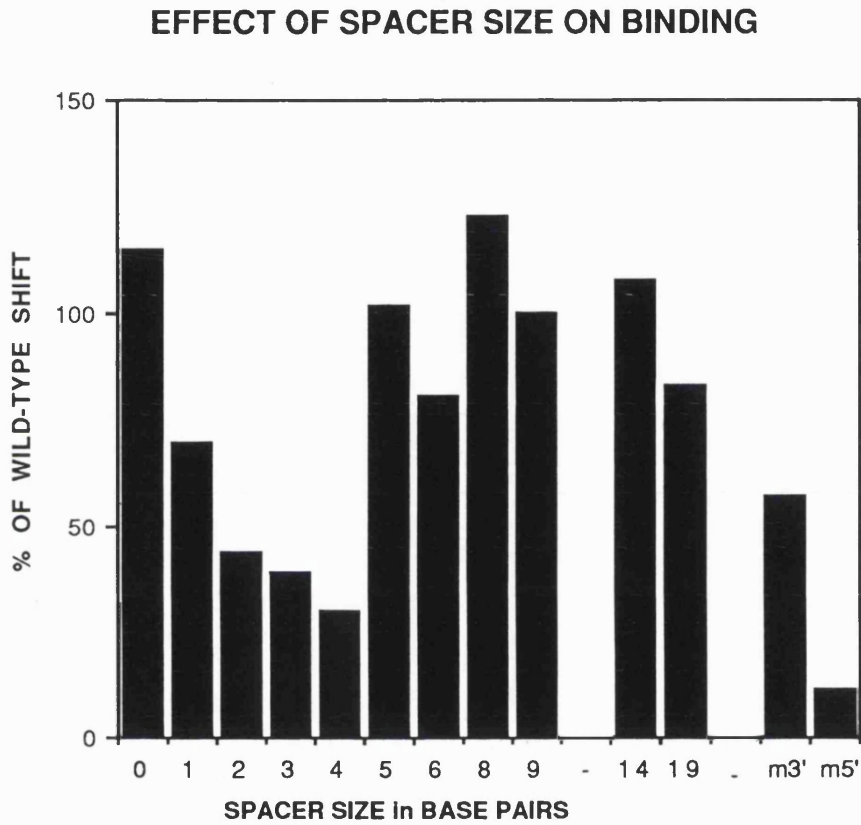
Histogram showing the relative binding efficiencies of oligonucleotides containing different spacer size between the TGACC halves of 'wild-type' site B (n=9). Binding to each oligonuleotide was calculated taking the P-32 activity in the DNA-protein complex formed with wild-type site B and Jurkat nuclear extract as 100%. Results relate to the EMSA shown in figures 4.12a and 4.12b.

4.3.1 Discussion and conclusions

The following conclusions can be drawn from these results.

1. At least two different proteins were shown to interact with site B; COUP-TF, abundant in HeLa cells and U937, less abundant in HL60 and THP-1 cells but absent in Jurkat cells. A protein forming a complex with site B but of lower mobility than COUP-TF was abundant in Jurkat nuclear extract.

2. The Jurkat nuclear protein had a M_r of approximately 110 kD, it bound to site B as a dimer with a total M_r of 220 kD. Based on EMSA this factor was the predominant protein present in Jurkat nuclear extracts able to specifically interact with site B. A complex of identical mobility was shown to be formed using nuclear extract prepared from normal peripheral blood T lymphocytes, confirming the relevance of Jurkat cells in studying the T-cell protein. Using the same conditions, the major proteins present in HeLa nuclear extracts were between 45-62 kD consistent with the published Mr of COUP-TF proteins (Tsai, S.Y and M-J Tsai. 1997).

3. The observation that COUP-TF proteins present in HeLa nuclear extract bind to site B was confirmed (Cooney et al. 1991) with similar EMSA complexes formed when either site B or a COUP-TF binding site was used as probe.

4. It was not possible however to demonstrate a protein in Jurkat nuclear extracts of similar Mr size as COUP-TF that formed detectable complexes with site B, nor was mRNA homologous to the highly conserved DNA-binding domain of COUP-TF detected in Jurkat RNA. Either COUP-TF proteins are expressed at levels below the sensitivity of the assays used or they are not expressed at all in Jurkat cells.

5. The binding of the Jurkat nuclear protein to site B was dependent on the precise sequence of the two halves of the palindrome, particularly that of the 5'-half, and the spacing between the two halves. As shown in chapter 3, the sequences within the spacer region itself were not involved in the binding of the Jurkat nuclear factor.

4.3.2 COUP-TF in T-lymphocytes.

The conclusion of Cooney et al that the Jurkat nuclear protein was related to the COUP-TF family of transcription factors was derived from three pieces of experimental evidence; 1) the similarity of the nucleotide sequence in the DNase I footprint over site B with the consensus binding site for COUP-TF;

2) the formation of DNA-protein complexes using Jurkat or HeLa nuclear extracts that were competed by an excess of unlabelled COUP-TF oligonucleotide and 3) the results of supershift EMSA. The 'anti-COUP-TF' used in these experiments was not shown to be specific for COUP-TF protein and the results are open to only limited interpretation. To define all proteins present in nuclear extracts as 'COUP-TFs' based on their ability to bind to both a COUP-TF consensus oligonucleotide and a site B oligonucleotide is clearly flawed as both sites are able to interact with multiple transcription factors (see chapter 5), indeed, the same group consistently co-purified a 53 kDa protein along with COUP-TF using affinity chromatography. This protein was identified as the human oestrogen related protein 1 (hERR1) a protein originally cloned and named because of close homology with the oestrogen receptor (Giguere et al. 1988).

4.3.3 COUP-TF in other haematopoietic cell lines.

The results of the Northern blot using a probe derived from the sequence for the DNA-binding domain of COUP-TF confirmed the presence of mRNAs for both COUP-TF I and COUP-TF II (ARP-1) in HeLa cells. The probe also hybridised with RNA of similar abundance from U937 cells indicating the presence of COUP-TF I and II in these cells. Weaker hybridisation bands were seen in RNA from THP-1 and HL60 cells suggesting that the mRNA s for the two forms of COUP-TF were present but far less abundant in these cells. Treatment of the cells with retinoic acid, which induces differentiation into monocytes, did not result in a change in the abundance of COUP-TF RNAs. The comparison of EMSA (fig. 4.2) and Northern blot (fig. 4.5a) for U937 and HeLa leads to the conclusion that band 'c' in fig. 4.2 was formed by the interaction of COUP-TF I/II with sequences in site B. EMSA using nuclear extracts from THP-1 or HL60 produced a pattern of bands similar to those produced by Jurkat nuclear extract (data not shown) indicating again a correlation between the detection of COUP-TF RNAs and specific bands in EMSA.

The results obtained with oligonucleotides corresponding to site B polymorphisms present in clinical isolates of HIV-1 have implications for the potential role of site B *in vivo*. Several proteins present in nuclear extracts from cells that can be infected by HIV-1 can interact with site B. The effect of site B on HIV-1 transcription may be influenced by not only the relative abundance of these proteins in specific cell types but also by the sequences contained in the palindromic elements of site B present in the LTR of the integrated provirus. Ladias performed an extensive analysis of site B

polymorphisms and the relative binding of several members of the steroid hormone receptor super-family, using *in vitro* expressed proteins in the EMSAs (Ladias 1994). By using expressed proteins for the EMSAs Ladias was able to precisely analyse the relative binding efficiency of each factor to the various site B sequences. Ladias examined the binding of eight members of the steroid hormone receptor super-family, COUP-TF I, COUP-TF II, ear-2, HNF-4 (hepatocyte nuclear factor 4), peroxisome proliferator-activated receptor (PPAR), nerve growth factor-inducible protein B (NGFI-B), RARα and RXRα. Ladias performed EMSAs using the three site B variants already described and extended the analysis with a further fifteen oligonucleotides corresponding to polymorphic sequences contained within or adjacent to site B. Some factors demonstrated a tolerance to sequence changes in site B forming complexes with most of the probes, for example COUP-TF I and II, ear-2 and HNF-4. Other factors were more stringent in binding to sequences in or adjacent to site B, for example RXRα homodimers (binding to only three of the sequences), RXRα/RARα and RXRα/PPAR heterodimers (four of the sequences gave strong binding with a further six showing weaker interactions). Ladias concluded that this region of the LTR was a complex hormone response element, extending 3' and 5' from the original description of site B and contained five distinct elements. He suggested that the region acted as a functional unit in the LTR and was the convergence site for multiple signal pathways operating through the diverse members of the steroid hormone receptor super-family. Interestingly, the sequences that most profoundly disrupted binding of all the steroid receptors were present in the 5'-half of the original site B palindrome.

Canonne-Hergaux et al examined the interaction of nuclear extracts from cells of neuronal origin with the same three site B variants described here (Canonne-Hergaux et al, 1995). They reported almost identical EMSA results to those described in this thesis using Jurkat nuclear extract e.g. a major slow mobility band (complex 'C1'). They also described a less abundant higher mobility complex ('C4') but from the data shown, complex 'C4' may be a non-specific DNA-binding protein as it was poorly competed using unlabelled oligonucleotides. Using HeLa nuclear extract, they described three complexes, 'C1', 'C3' and 'C4', with barely detectable 'C1' and an abundant 'C4' complex. Using nuclear extracts from neuronal cell lines, four complexes were formed, 'C1' to 'C4'. Unfortunately, they used the polyclonal 'anti-COUP' provided by Tsai to define the proteins in the complexes 'C1', 'C2' and 'C3' as COUP-TFs, an error of interpretation already discussed. They did however confirm the effect of site B polymorphism on the relative binding of the various proteins present in

Jurkat and HeLa nuclear extracts and extended the analysis to cells of neuronal origin.

4.3.4 Spacer sequence size and Jurkat nuclear factor binding

Figures 14.14a and 14.14b show a model for the interaction of the Jurkat nuclear protein with site B TGACC half-site elements. Results of the EMSAs with varying spacer size indicated that some constraint operated on the binding of the Jurkat protein to site B, for maximal binding of the dimer both half-sites had to be accessible to both monomer units. By reducing the spacer size to four nucleotides or less the relative orientation of the half-sites was significantly changed (fig. 14.14b). An alternative explanation would be that important bases were removed in the construction of the n1-4 spacer oligonucleotides and it was the loss of these that resulted in the observed reduction of binding. However, an earlier EMSA in which the entire sequence of the n9 spacer region was changed resulted in no discernible reduction in binding (chapter 3, fig. 3.9).

The model shows interaction via the major groove, this is merely a suggested model, if the sequence is moved by two residues interaction between the Jurkat nuclear protein and the half-site sequences is possible at the minor groove (not shown). NMR and X-ray crystallographic evidence indicates that, for those transcription factors for which the techniques have been applied, sequence specific binding occurs in the major groove of the DNA double helix, for example the ER (Schwabe et al 1993), NF-κB p50 homodimers (Ghosh et al 1995), p50/p65 heterodimers (Chen F, et al 1999) and STAT-1 (Chen X, et al 1998). The polarised heterodimer RXR/TR, although predominantly binding through contacts in the major groove, also interacts with important bases in the minor groove in a sequence specific manner (Rastinejad et al 1995). Binding of the Jurkat nuclear protein was virtually unaffected when the spacer size was zero between 5 and 9 (fig. 4.13) indicating that the structural constraints were not solely due to physical blocking of a half site by one of the monomer subunits. Members of the steroid-thyroid receptor family of transcription factors display remarkable tolerance to the separation and orientation of half-site elements. As discussed in section 1.4.5.2 the nuclear hormone response elements (NHREs) contain the canonical core sequence PuGGTCA repeated as a palindrome (as in site B), everted repeats and direct repeats. The orientation of half-sites and the size of the spacer region determine which particular NHRs are able to bind and thus determine selectivity.

Examples of NHREs and their orientation:

Palindromes (PALx)	Direct repeats (DRx)	Everted repeats (ERx)

PAL3
 	
- oestrogen receptor
 	
- glucocorticoid
 	
  receptor
 	
PAL0
 	
- Thyroid hormone
 	
  receptor
 	
PAL9
 	
- Jurkat nuclear protein
 	
- COUP-TF
	DR1: RAR/RXR; COUP-TF
	DR2: COUP-TF
	DR3: RXR/VDR, TR/VDR
	DR4: RXR/TR
	DR5: RXR/RAR
		Complex response
		elements
		
		COUP-TF

(From Gronemeyer and Moras 1995)

For RAR and RXR heterodimers, the favoured binding site consists of a DR, the spacing between half sites in part determining which heterodimer partner is involved. COUP-TF sites also contain at least one GGTCA element half-site, the other half can display some deviation from this consensus. The majority of natural COUP-TF sites are DR2 but COUP-TF can bind to a wide range of sites, including PAL0, PAL9 and DR6. Binding of homodimers to the symmetrical PAL elements occurs by the interaction of the C I zinc finger. Sequences in the P-box determine the specificity of binding, an α-helical structure is involved in sequence specific contact with the half site element, binding in the major DNA grove. Dimerisation occurs through the D-box in the C II zinc finger. Binding of heterodimers or COUP-TF to the polarised direct repeat elements is possible by conformational changes that are induced in both halves in the complex (Gronemeyer and Moras 1995; Rastinejad et al 1995 and Cooney et al 1992).

The Jurkat nuclear protein binds to a PAL9 element (site B), a finding consistent with the observation that homodimers tend to bind to PAL elements. Further characterisation of the interaction of the Jurkat protein with site B would require the use of recombinant expressed protein. The potential for the Jurkat protein to interact directly with RAR, RXR and COUP-TF could thus be explored. The ultimate analysis of protein-DNA interaction requires NMR and X-ray crystallography, as has been used for a number of transcription factors interacting with their response elements.

Figure 4.14a

**Diagram indicating the possible interaction of the
Jurkat nuclear protein with site B nucleotides.**
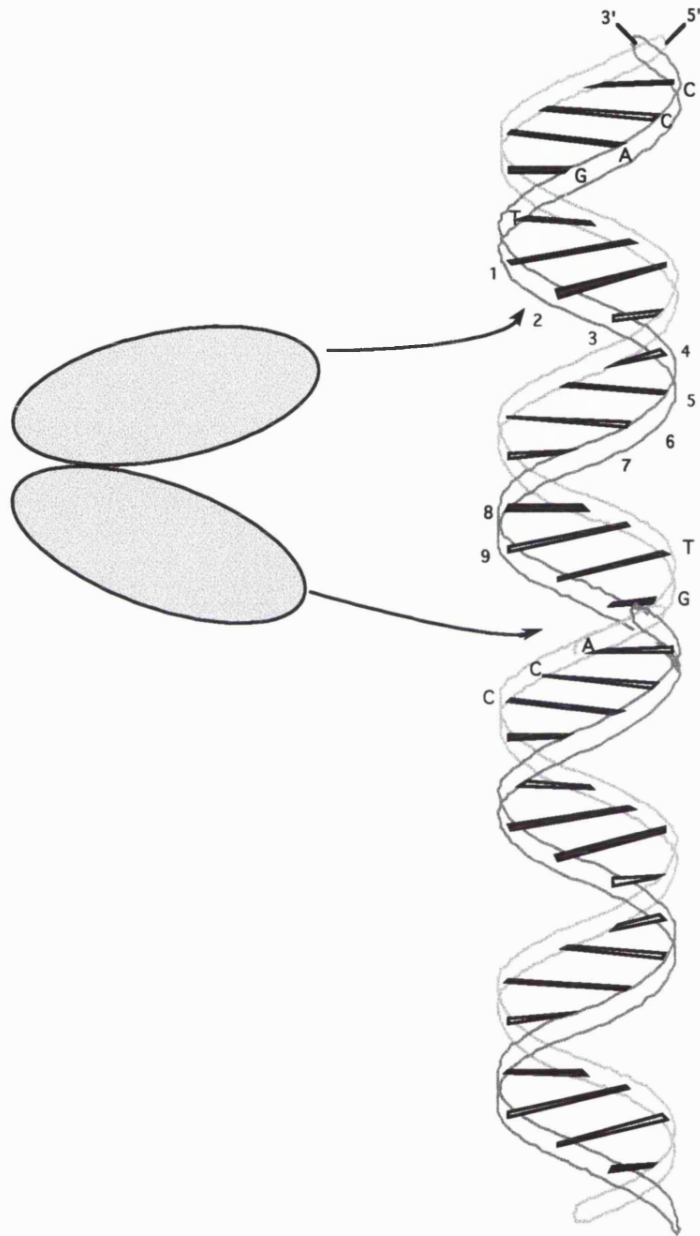
The Jurkat nuclear protein binds to site B as a dimer, interacting via the TGACC half-site sequences. Interaction is shown to occur in the major groove of the double helix.

Figure 4.14b

Diagram representing the binding of the Jurkat protein to site B sequences containing a four nucleotide spacer between TGACC half-site

A four nucleotide spacer region between TGACC half-sites presents one half-site across the minor groove in a position inaccessible to one of the Jurkat nuclear protein monomers. Binding of the Jurkat protein dimer to site B is only possible through one TGACC sequence only.

Chapter 5

Site B as a Hormone Responsive Element.

Chapter 5

Site B as a Hormone Responsive Element.

5.1: Introduction

The preceding chapters describe the delineation and characterisation of a palindromic element within the NRE of HIV-1 and of the major protein factor present in Jurkat nuclear extracts that interacts with the site. The sequences within site B show a marked similarity to the transcriptional elements that are recognised by members of the steroid/thyroid/retinoic acid/vitamin D3 superfamily of transcription factors, the hormone responsive elements (HREs) (fig. 5.1). As discussed in the introduction, the hormone receptor (HR) super-family is large and contains members which bind to HREs and function as ligand activated transcription factors but also contains members that have no known ligands, the so-called orphan receptors. These transcription factors bind to the appropriate response elements as dimers, some can only form homodimers (e.g. the glucocorticoid receptor) but others, particularly the retinoic acid receptors RARs and RXRs, the thyroid hormone receptor, vit D3 receptor and several orphan receptors are also able to form heterodimers. Further levels of complexity of functional significance are created by the influence of ligands for these hormone receptors (e.g. all-*trans* retinoic acid and 9-*cis* retinoic acid) which influence the formation of heterodimers and their binding to RAREs. A further level of complexity resides in the HREs themselves, the orientation of the repeated elements, the size of the spacer sequence, the precise sequence of bases within the repeat elements and flanking sequences all have profound influence upon the function of the site as a response element, the same site can function as a positive element under one set of conditions, but as a negative element under a different set of conditions.

Because of the similarity of site B to the known HREs it was important to explore the possibility that site B could function as an HRE. Early competition EMSA had shown that the HREs for the thyroid hormone receptor and oestrogen receptor could compete with site B for binding of a Jurkat nuclear factor (chapter 3), these results provided the initial indication that the sequences contained within the region defined as site B could interact with members of the hormone receptor family of transcription factors. A number of retroviruses have been shown to contain hormone response elements within their LTRs and viral gene expression can be modulated by the appropriate hormones. For example the glucocorticoid hormone response

elements in the LTR of the mouse mammary tumour virus and Moloney murine sarcoma virus (von der Ahe et al. 1985; DeFranco and Yamamato 1986) and the thyroid hormone response elements in the Moloney murine leukaemia virus (Sap et al. 1989).

Although we were able to demonstrate the binding of recombinant thyroid hormone receptor and oestrogen receptor to site B (data not shown) these receptors are not abundant in cells infected by HIV-1. However retinoids have profound effects upon cells of the monocyte/macrophage lineage influencing cell proliferation, differentiation, attachment to substrate, effector cell responses, expression of membrane determinants, cytokine production and prostaglandin and glycolipid synthesis (Goodman 1990; Mehta et al 1986; Turpin et al 1990; Dillehay et al 1988; Trechsel et al 1988; Murray et al 1988; Goldman 1984) indicating indirectly that these cells contain intact response pathways to retinoids. Monocytic/macrophage lineage cells are important HIV target cells, macrophages in the CNS, lung, lymph nodes and follicular dendritic cells are infected chronically and are reservoirs of virus (reviewed by Balter et al 1996; Salahuddin et al 1986; Gendelman et al 1988). These cells have been shown to be a source of circulating virus in the later stages of infection while viral replication in lymph nodes is dramatically increased by concurrent active infection by opportunistic organisms such as *Mycobacterium avium* and *Pneumocystis carinii* (Orenstein, et al 1997). Furthermore, chronically infected monocyte/macrophage lineage cells produce large amounts of progeny virus without the cytopathic effect active viral replication has on lymphocytes (Nicholson et al. 1986; Orenstein, et al 1997).

Retinoids have been shown to affect the dynamics of HIV-1 infection of monocytic cells in several ways. The pre-treatment of immature monocytic cell lines HL-60 and U937 with all-trans retinoic acid made them more susceptible to productive infection with HIV-1 (Kitano et al 1990; Turpin et al 1992; Poli et al. 1992). Pre-treatment of the monocytic cell line THP-1, which has a more mature phenotype, or peripheral blood monocytes with all-trans retinoic acid resulted in an increase in virus production rather than enhanced infectivity (Kitano et al 1990; Turpin et al 1992). In primary monocytes the level of viral replication was highest in the cells exposed to ATRA pre and post infection, indeed brief treatment prior to infection with HIV-1 and then withdrawal of ATRA from the cultures resulted in a reduction in viral replication (Turpin et al 1992).

The results obtained from these experiments, although demonstrating important effects of retinoids on the replication of HIV-1 in monocytic cell lines and primary cultures, did not define the level at which retinoids act on HIV-1 replication. There are two possible ways in which retinoids affect HIV-1 replication, indirect or direct. Indirectly retinoids, by altering the general cellular milieu, increase the permissivity of monocytic cells for HIV-1 replication. This may be effective at any or several points in the replication cycle of the virus, for example integration of provirus, expression of viral mRNA due to NF-kB binding to the LTR, viral protein synthesis and assembly of progeny viral particles. Retinoids could however be acting directly on the viral LTR, through specific retinoic acid response elements (RAREs), increasing LTR-driven expression. Because of the clear homology of site B with known RAREs (fig. 5.1) the effects of retinoids on the expression of HIV-1 in macrophage-lineage cells could operate by a direct mechanism through this site. We therefore decided to determine whether the site B sequences could function as a retinoic acid response element. This was explored by using EMSA, with labelled site B oligonucleotide as probe and cell extracts containing recombinant expressed retinoic acid receptors. The functional significance of site B as a RARE was tested by transient transfection assays using multimerised or single copies of site B to drive a reporter gene.

Fig. 5.1　　　　　**Hormone response elements (HREs)**

SITE B:	5' CCAGG<u>GGTCA</u>GATATCCAC<u>TGACC</u>TTTGG GGTCCCCAGTCTATAGGTGACTGGAAACC
TRE:	5' CGTGC<u>GGTCA</u>CGTCCC<u>TGATC</u>TTACC GCACGCCAGTGCAGGGACTAGAATGG
ERE:	5' AGTC<u>AGGTCA</u>CAG<u>TGACC</u>TGAT TCAGTCCAGTGTCACTGGACTA
COUP:	5' AAGT<u>TTGACCTT</u>T<u>GACCA</u>TAGA TTCAAACTGGAAACTGGTATCT
RAREβ:	5' CCAGG<u>GTTCA</u>CCGAAA<u>GTTCA</u>CT GGTCCCAAGTGGCTTTCAAGTGA
RARE-mCRBP:	5' GT<u>AGGTCA</u>AAA<u>GGTCA</u>GA CATCCAGTTTTCCAGTCT
apo-A1	5' AC<u>TGACCCCTGACCC</u>CTGCCT TGACTGGGGACTGGGGACGGA
γF-crystallin	5' AG<u>TGACCC</u>TTTTAACC<u>AGGTCA</u>GTGA TCACTGGGAAAATTGGTCCAGTCACT
mouse H2Kᵇ	5' GAG<u>TGACCCC</u>GGGTCGG<u>AGGTCA</u>CGA CTCACTGGGGCCCAGCCTCCAGTGCT
Human H1°	5' AT<u>TGACCCC</u>CGAGAGG<u>TGACC</u>G TAACTGGGGGCTCTCCACTGGC
SITE B:	5' CCAGG<u>GGTCA</u>GATATCCAC<u>TGACC</u>TTTGG GGTCCCCAGTCTATAGGTGACTGGAAACC

Comparison of HREs with site B showing the sequence homologies of the core binding sites (sequences underlined). TRE: thyroid hormone receptor response element from the rat growth hormone promoter (Flug F. *et al* 1987)

ERE: oestrogen receptor response element from the *Xenopus* vitellogenin A2 promoter (Klein-Hitpab, L. *et al* 1986). COUP: response element for COUP-TF (ear-3) from the chicken ovalbumin upstream promoter (Sagami, I. *et al* 1986).

RAREs: RAR-β promoter (de Thé, H. *et al* 1990); mCRBP1 (mouse cellular retinol binding protein 1) promoter (Smith, W.C. *et al* 1991); apo-A1 apolipoprotein promoter (Rottman, J.N. *et al* 1991); γF-crystallin gene promoter (Tini, M. *et al* 1993); mouse H2Kb (mouse MHC class 1) promoter (Jansa *et al* 1996); human H1° (Histone H1°) promoter (Bouterfa *et al* 1995)

Half-site sequences as reported in the corresponding publication have been underlined.

5.2: Results.

5.2.1 Retinoic acid receptors RARα, β, γ and RXRα bind to site B.

The ability of recombinant human retinoic acid receptors RARα, β and γ and RXRα (hRXRα) to bind to site B was assessed in gel retardation assay using extracts of HeLa cells infected with recombinant vaccinia viruses expressing the appropriate receptor. Figure 5.2 shows the results of EMSA with HeLa cell nuclear extracts prepared from cells infected with recombinant vaccinia virus expressing hRARα, β or γ, compared with nuclear extract from HeLa cells infected with the vaccinia virus vector as control. A specific complex (solid arrow) formed with the site B probe when extracts containing hRARα (lane 4), hRARβ (lane 6) or hRARγ (lane 8), which was competed by a 100-fold excess of unlabeled site B oligonucleotide (lanes 5, 7, 9). A specific band of lower mobility (open arrow) was formed with all extracts including the control (lane 10) suggesting that this was due to a constitutively expressed protein present in HeLa cells. No comparable complex was formed with HeLa extract prepared from cells infected with wild-type virus using either labelled site B (lane 10) or RARE (lane 12) as probe. The relative mobility of the site B-hRAR complexes are compared with that of the Jurkat protein and site B (lane 1). A 100-fold excess of unlabeled site B oligonucleotide specifically competed the major and minor DNA-protein complexes (lane 2), similarly a 100-fold excess of an unlabeled RARE from the RARβ-promoter (de The et al. 1990) efficiently competed the major complex although the two complexes of higher mobility are faintly visible. Figure 5.3 shows the results of EMSA obtained with an oligonucleotide containing the RARE from the RARβ-promoter as probe (lanes 1-5) or site B oligonucleotide (lanes 6-7) with Jurkat nuclear extract (lanes 1 and 6-7) or the HeLa cell extract containing the RAR as indicated. These results confirm the presence of the appropriate RAR in the HeLa extracts and show the ability of a factor in Jurkat nuclear extract to bind to the RARE oligonucleotide (lane 1). Comparison of the EMSA band patterns formed with Jurkat nuclear extract and the RARE or site B oligonucleotide (lanes 1 and 6) shows a predominant DNA-protein complex of similar mobility but the two minor high mobility bands formed characteristically with site B are absent when the RARE oligonucleotide was used as probe.

Fig. 5.2   EMSA with recombinant retinoic acid receptors   1

extract:

wt HeLa											+		+
hRARα				+	+							+	
hRARβ						+	+						
hRARγ								+	+				
J6	+	+	+										
competition:				+		+		+				
probe:	B	B	B	B	B	B	B	B	B	B	R	R
	1	2	3	4	5	6	7	8	9	10	11	12
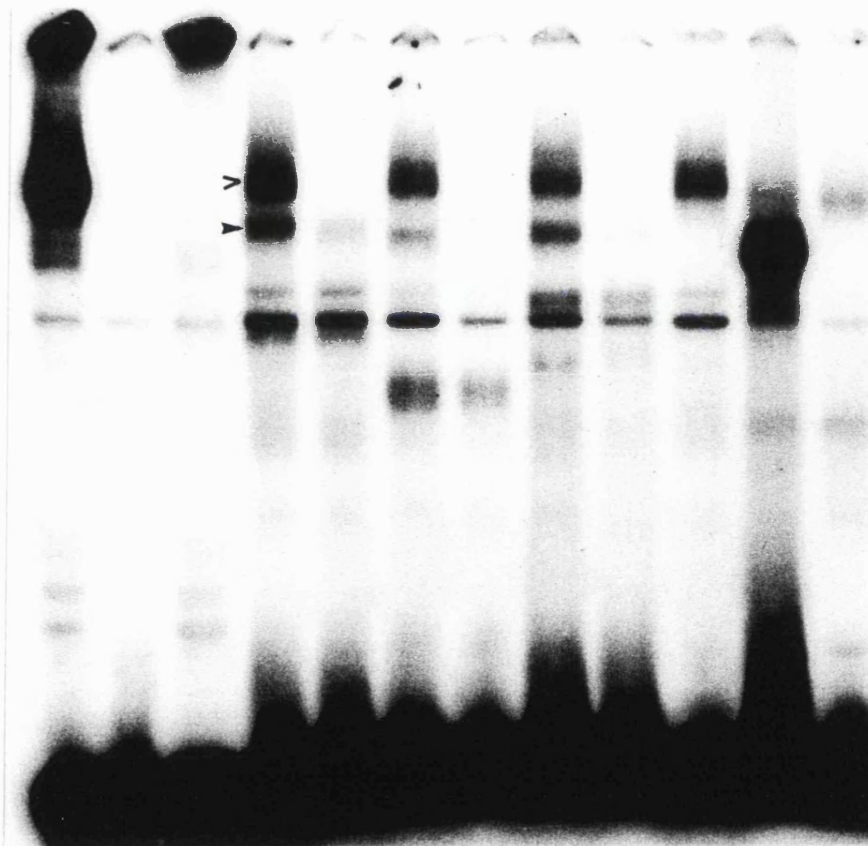
EMSA with site B probe (B in lanes 1-10) or the RARE from RARβ promoter (R in lanes 11, 12) as probe. Extracts incubated with probe as indicated; J6 Jurkat (1μl, lanes 1, 2 and 3); hRARα (lanes 4,5), hRARβ (lanes 6,7), or hRARγ (lanes 8,9) indicates 2μl HeLa cell nuclear extract infected with recombinant vaccinia virus expressing corresponding retinoic acid receptor. Control HeLa nuclear extract infected with wild-type viral vector (lanes 10, 12). Competitions were performed with 100-fold excess of unlabeled oligonucleotide as indicated.

competition:
    B
    RARE
extract:
    wt HeLa
    hRARα
    hRARβ
    hRARγ
    J6
probe:
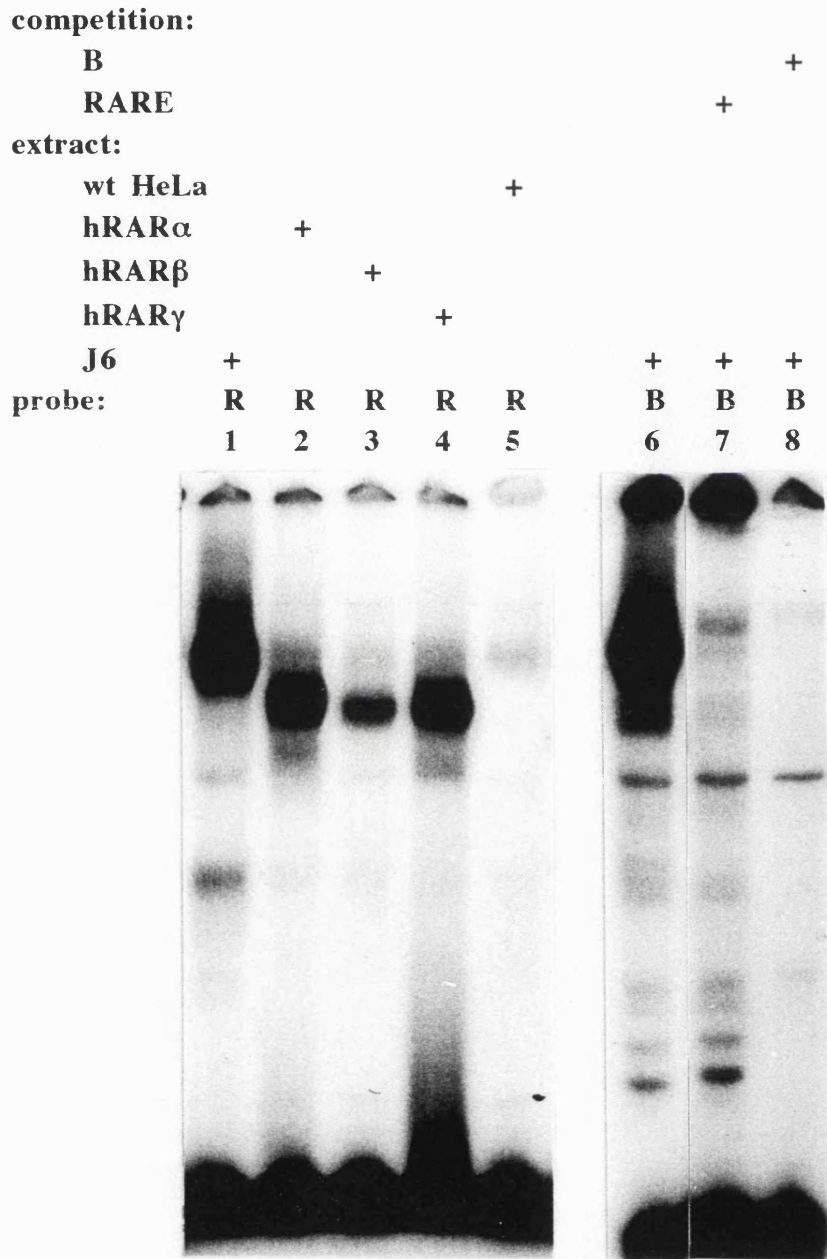
EMSA with either an oligonucleotide containing the RARE from the RARβ promoter (R in lanes 1-5) or site B (B in lanes 6-8) as probe. Probes were incubated with 2ml of nuclear extract prepared from HeLa cells infected with recombinant vaccinia virus expressing either hRARα (lane 2), hRARβ (lane 3) or RARγ (lane 4). The EMSA obtained with J6 and site B are shown for comparison (lanes 6-8).

Supershift EMSAs were performed using either anti-RARα or anti-RARβ in the binding reaction. The results of the supershift EMSA are shown in figure 5.4. Neither anti-RARα nor anti-RARβ produced a supershift of the major band formed with Jurkat nuclear extract. The activity of the anti-RARα was confirmed by the formation of a supershifted band with vaccinia expressed RARα (lane 5).

RARs bind only weakly to retinoic acid response elements (RAREs) as homodimers, however RAR/RXR heterodimers bind to RAREs with much higher affinity. To explore whether the presence of hRXRα would result in binding of a heterodimer to site B EMSAs were performed using a combination of HeLa nuclear extracts from cells infected with recombinant vaccinia virus expressing hRARα together with nuclear extract from cells infected with recombinant vaccinia virus expressing hRXRα. In these experiments (fig. 5.5), much stronger binding was observed when a combination of extracts containing hRARα and hRXRα were used (lanes 4 and 5). No binding was shown by hRXRα extract alone (lane 7), however these EMSA were performed in the absence of the natural ligand for RXR and the RXR/RAR heterodimer, 9-cis retinoic acid (Lehmann et al 1992; Zhang et al. 1992).

The specificity of binding confirmed by competition with unlabeled site B oligonucleotide (fig 5.6 lanes 3 and 4) but not by an unrelated oligonucleotide (lane 2). The rabbit polyclonal anti-RARα antibody resulted in a supershift of the complex (lane 5) the specificity of which was confirmed by competition with excess unlabelled site B oligonucleotide (lane 4). The addition of anti-RARγ to the binding reaction did not cause a supershift (lane 6).

195

Fig. 5.4   EMSA with recombinant retinoic acid receptors   3

antibody:

anti-RARα   +   +   +

anti-RARβ   +

unlabeled oligo:

RARE   +

extract: J6   J6   α   α   α   J6   J6

probe: B   R   R   R   R   R   R

             1   2   3   4   5   6   7
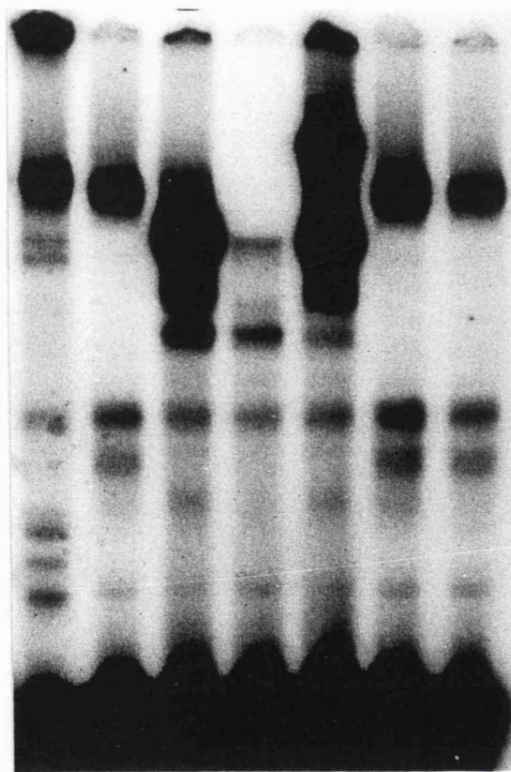
Supershift EMSA with probes and nuclear extracts as shown. Jurkat (J6) (lanes 1, 2, 6 and 7) or 2μl of a nuclear extract from HeLa cells infected with vaccinia virus expressing hRARα (α) (lanes 3-5); site B oligo. (lane 1) or the RARE from RARβ promoter (lanes 2-7). Competition with 100-fold excess RARE demonstrating specific bands (lane 4). Rabbit anti-RARα (lanes 1, 5 and 7) or anti-RARβ (lane 2) was included in the incubation mix.

Fig. 5.5 EMSA with recombinant retinoic acid receptors 4

extract:

hRARα	+	+	+	+	+	+	
hRXRα				+	+	+	+
probe:	**B**	**B**	**B**	**B**	**B**	**B**	**B**
	1	**2**	**3**	**4**	**5**	**6**	**7**
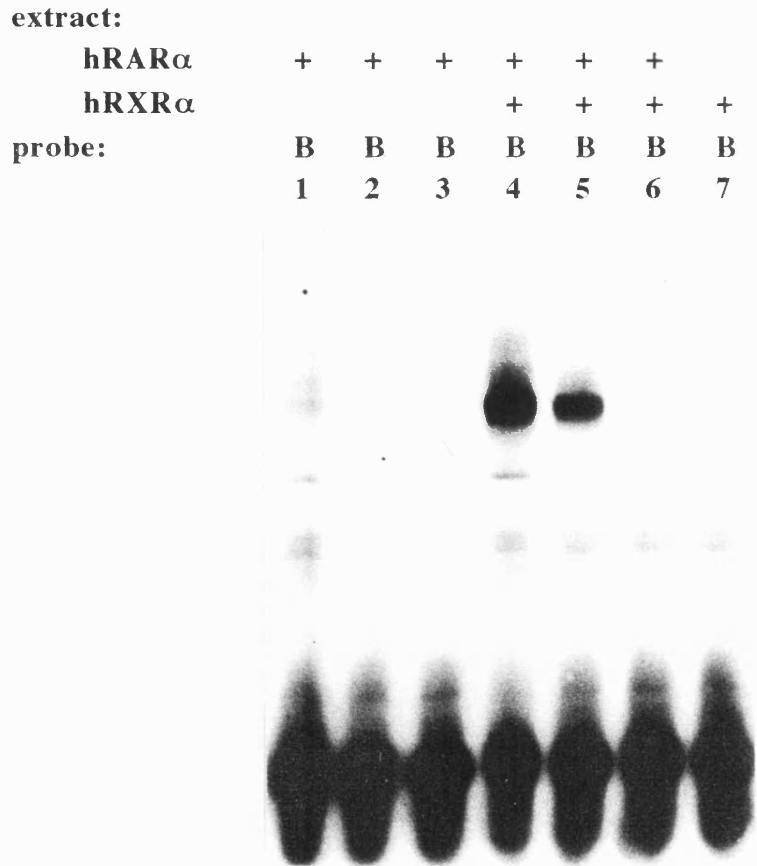
EMSA using site B oligonucleotide as probe and 1µl (lanes 1 and4), 0.1µl (lanes 2 and 5) or 0.01µl (lanes 3 and 6) of a nuclear extract prepared from HeLa cells infected with a recombinant vaccinia virus expressing hRARα. In lanes 4-6, 0.5µl of an extract prepared from HeLa cells infected with vaccinia virus expressing hRXRα has been added. Lane 7 shows the result of HeLa extract containing hRXRα and site B alone.

Fig. 5.6   EMSA with RARα and RXRα

antibody:

anti-hRARγ						+
anti-hRARα				+	+	

extract:

hRARα	+	+	+	+	+	+
hRXRα	+	+	+	+	+	+
competition:	-	A	B	B	-	-
probe:	B	B	B	B	B	B
	1	2	3	4	5	6
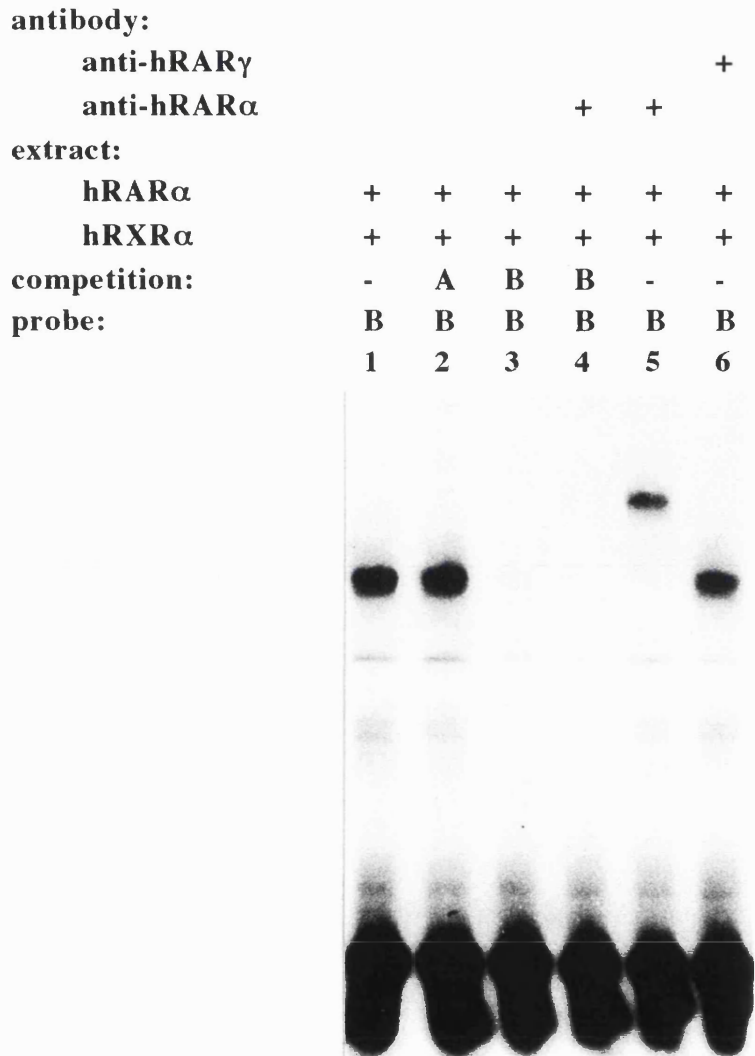
EMSA using site B oligonucleotide as probe and 0.5ml of the hRARα expressing HeLa cell extract and 0.2μl of hRXRα expressing extract together in the incubation mix. The assays were performed in the absence of competitor oligonucleotide (lanes 1, 5 and 6) or with a 200-fold excess of unlabeled oligonucleotide present in the incubation, in lane 2 an unrelated oligonucleotide (site A from HIV-1 LTR) or site B (lanes 3 and 4). In lanes 4 and 5, 0.5μl of a murine anti-hRARα monoclonal antibody was added, while lane 6 shows the result of adding 0.5μl of a murine anti-hRARγ monoclonal antibody.

The following conclusions can be drawn from these results;

1) The recombinant RARs present in HeLa extracts were shown to bind, albeit weakly, to the site B sequences.

2) A protein present in Jurkat nuclear extract formed a complex with the RARE from the RARβ promoter of similar mobility to the complex formed with site B.

3) The protein factors present in Jurkat nuclear extracts that constitute the minor, high mobility bands bound more weakly to the RARE.

4) The major Jurkat nuclear protein was not supershifted by anti-RARα or anti-RARβ anti-sera indicating that the Jurkat protein was not one of these two members of the retinoic acid family of transcription factors.

5) RARα/RXRα heterodimers were shown to bind more effectively to site B than RARα homodimers (compare figures 5.2 and 5.5). RXRα homodimers did not bind to site B under the conditions used for these binding reactions.

5.2.2 Site B functions as a RARE:

Although specific binding of RARs and RAR/RXR heterodimers to site B could be demonstrated, it was important to determine the functional significance of such interaction. Functionality was determined using the chloramphenicol acetyl-transferase reporter gene system. Either a single copy or multimers of site B oligonucleotide were cloned into the *Sal* I site of the pSP65.tk.CAT vector, the orientation of the cloned oligonucleotide or multimer determined by sequencing across the cloning site. Clones were selected where the orientation of site B was found to be the same as in the HIV-1 LTR. Transfections were performed in Cos cells and F9 cells. Initial transfections performed in Cos cells produced variable and weak all-*trans* retinoic acid (ATRA) stimulation of the reporter vector pBL2.tk.CAT, which contained three copies of a thyroid hormone response element and has been shown to function as a RARE. The weak response of the control reporter vector to ATRA was thought to be due to low endogenous levels of retinoic acid receptors in Cos. This possibility was addressed in two ways; in Cos cells an expression vector for RARa was co-transfected with the reporter plasmid; secondly transfections were performed in F9 cells, which are naturally responsive to retinoic acid and therefore contain functionally significant levels of retinoic acid receptors (Strickland and Mahdavi 1978).

5.2.3 Details of plasmids used in transfections:

pSP65tk.CAT: based on the expression vector pSP65 and contained the Herpes simplex thymidine kinase promoter (tk) driving the bacterial chloramphenicol acetyl transferase gene (CAT). A single copy of site B oligonucleotide was cloned into the *Sal* 1 site of pSP65tk.CAT in plasmid $(B)_1$tk.CAT. Multimerised site B or mutated site B oligonucleotides were cloned into the *Sal* 1 site upstream of the tk promoter to create the reporter vectors $(B)_4$tk.CAT and $(B)_3$tk.CAT which contained either four or three head-to-tail copies of site B in the same orientation as in the HIV-1 LTR. Vector $(mB2)_4$tk.CAT contained four head-to-tail copies of an oligonucleotide in which both halves of the site B palindrome were mutated. All constructs were sequenced in both directions across the insertion site to confirm the orientation and number of inserted multimers

pBLCAT2: based on the vector pBL2 which contained a truncated tk promoter (with a single octamer binding site deleted) driving the CAT gene (Luckow and Schutz 1987). This was used to derive the control vector $(TRE)_3$CAT containing three copies of the consensus thyroid hormone receptor element (TRE) (5'AGCTTAGGTCAGGGACGTGACCTT3') cloned upstream of the tk promoter (a generous gift of Dr. Jacques Ghysdale, Institute Pasteur, Paris). This has been shown to function as an efficient RARE (Zelent et al. 1989).

LTR.CAT: equivalent to the vector pU3R-III (described by Rosen et al 1985) and contained the 5' LTR of HIV-1 (site B sequence with perfect TGACC palindromes) from -453 to +80 (cap site +1) cloned upstream of the CAT gene.

p2.1.CAT: derived from LTR.CAT and contained the HIV-1 LTR with mutated 5' TGACC half of the site B palindrome.

The RARα eukaryotic expression plasmid contained the human RARα receptor cloned into the *Sma* I site of pJ3Ω placing it under the control of the Moloney murine leukaemia virus LTR. This plasmid also contained the SV40 large T origin of replication (a generous gift of Professor P. Chambon, INSRM, Strasbourg). Where indicated, co-transfection experiments were performed with a Tat expression vector under the control of the promoter from Rous sarcoma virus LTR.

Transfections into Cos cells were performed by $CaCl_2$ precipitation as described in chapter 2. Transfections were performed with 2.5µg of test reporter plasmid,

2.5 μg of RARα expression vector, or the parental vector pJ3Ω and 0.5μg of a β-galactosidase expression plasmid (used to standardise cell extracts for CAT assay). All DNA samples were made up to a total of 20μg with pUC1813 plasmid before transfection. Transfections into murine embryonal carcinoma F9 cells were performed without the RARα expression vector.

Figure 5.7 shows the results of CAT assays in Cos cells transfected with the reporter vectors (B)₄tk.CAT or (mB2)₄tk.CAT. The expression vector for RARα or the parental eukaryotic expression vector, pJ3Ω, were co-transfected as indicated. Transfected Cos cells were incubated for 24 hours with or without the addition of all-*trans* retinoic acid at a final concentration of 10^{-6}M. The vector (B)₄.tk.CAT showed ATRA inducible expression of CAT activity, this ATRA induction was not shown by the reporter construct (mB2)₄tk.CAT containing alterations to the TGACC palindromes of site B. The control reporter vector (TRE)₃.CAT demonstrated appropriate retinoic acid inducible CAT activity. The reporter construct (B)₃tk.CAT also demonstrated retinoic acid induction of CAT activity in Cos cells (Figure 5.8).

Figure 5.9 shows the results of transfections in F9 cells, without exogenous RARα. The (B)₃tk.CAT reporter construct demonstrated retinoic acid inducible expression of CAT activity (lanes 1 and 2), there was no induction of CAT activity with the parental pS56tk.CAT vector (lanes 3 and 4). The control plasmid (TRE)₃tk.CAT also demonstrated retinoic acid induction of CAT activity in F9 cells (lanes 5 and 6).

This series of experiments confirmed that multimerised site B oligonucleotide, in the context of a heterologous reporter construct with the tk promoter, was activated in the presence of all-*trans* retinoic acid. The ATRA responsiveness of the heterologous promoter was removed by base pair changes in the TGACC palindrome elements. A single site B oligonucleotide adjacent to the tk promoter in the reporter construct (B)₃tk.CAT was inconsistent in demonstrating ATRA induction of CAT activity in Cos cells (data not shown). However, when transfected into F9 cells, (B)₁tk.CAT clearly showed ATRA induced induction of CAT activity (Figs. 5.9 and 5.11).

Fig. 5.7 CAT assays in COS cells with multimerised site B
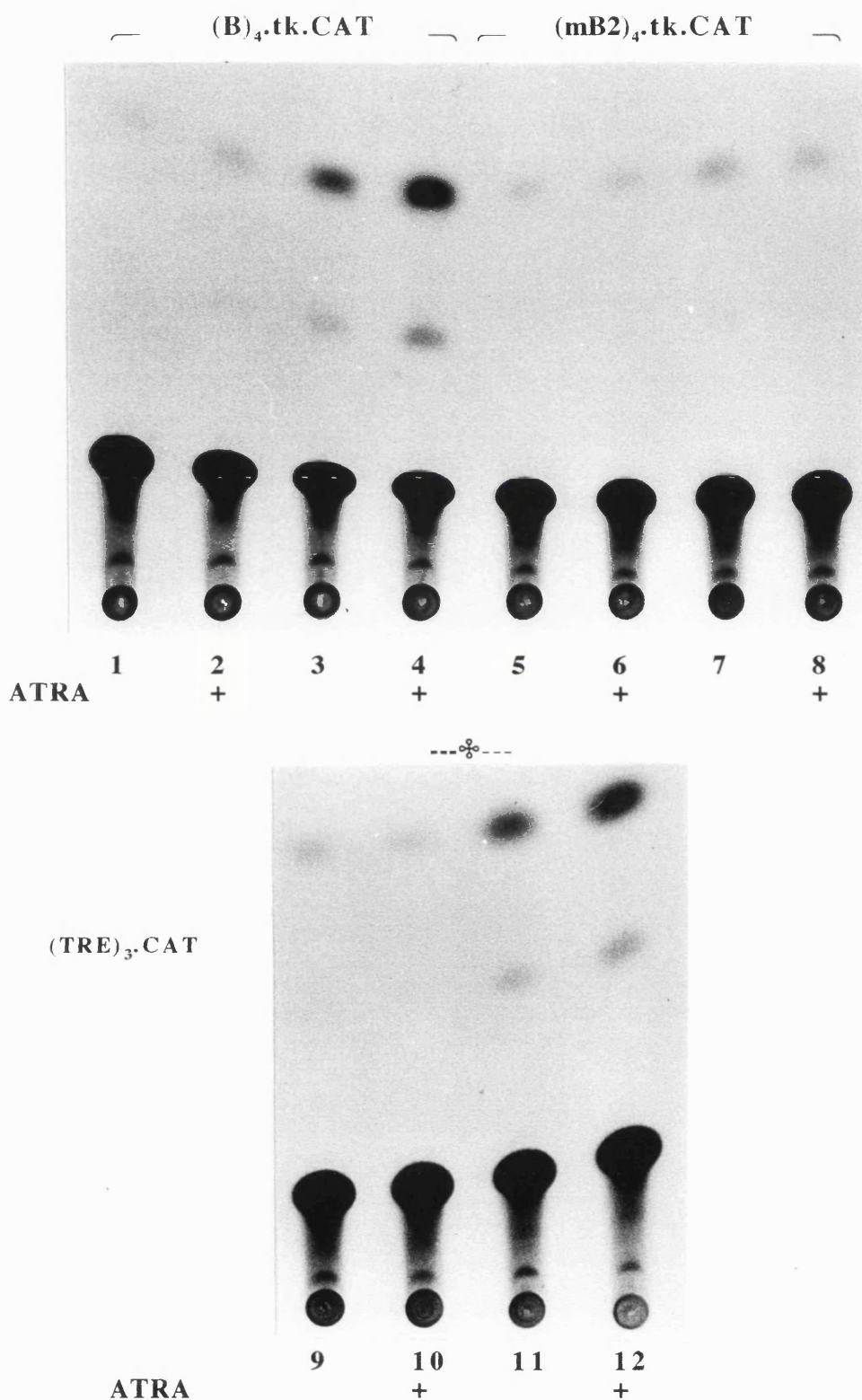
$(B)_4.tk.CAT$ $(mB2)_4.tk.CAT$

	1	2	3	4	5	6	7	8
ATRA		+		+		+		+

$(TRE)_3.CAT$

	9	10	11	12
ATRA		+		+

Assay of chloramphenicol acetyltransferase activity following the transfection of Cos cells with the reporter vectors $(B)_4.tk.CAT$ (lanes 1-4), $(mB2)_4.tk.CAT$ containing four copies of the mutated site B oligo. (lanes 5-8) or the positive control reporter vector $(TRE)_3.CAT$ (lanes 9-12). Co-transfection of the eukaryotic expression vector pJ3Ω was performed in all transfections, either the parental vector alone (lanes 1, 2, 5, 6, 9, 10) or with the vector containing the RARα gene (lanes 3, 4, 7, 8, 11, 12). 10^6 M all-trans retinoic acid (ATRA) incubated with cells following transfections as indicated

Fig. 5.8 CAT assays in COS cells with multimerised site B
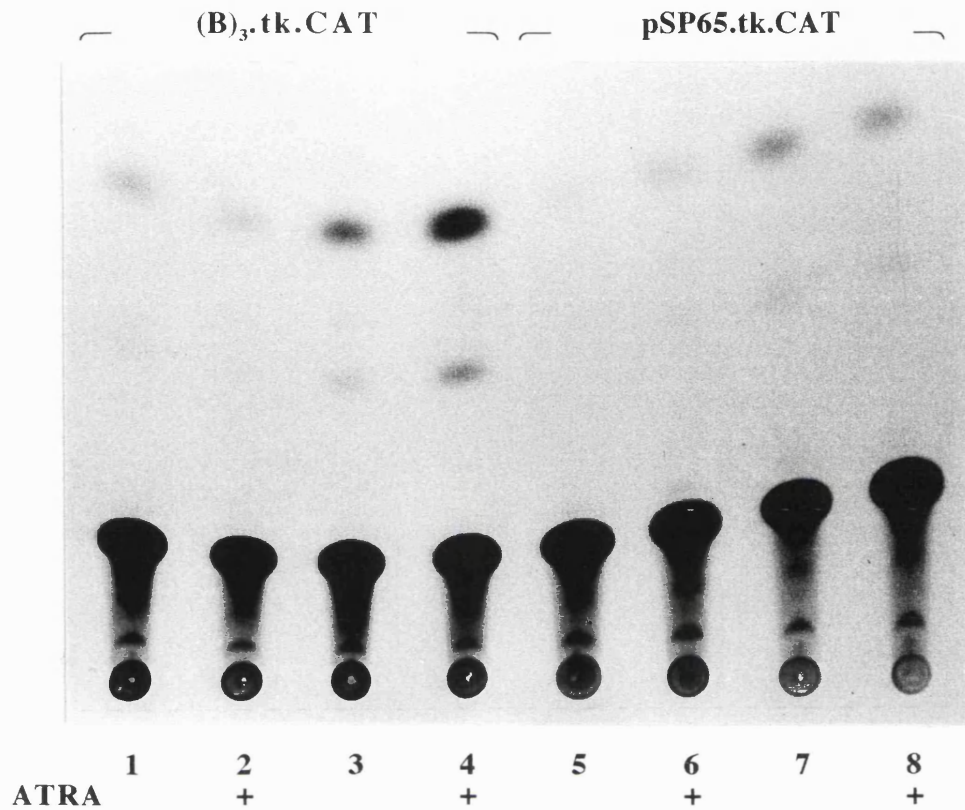
CAT activity following the transfection of Cos cells with reporter plasmid $(B)_3$.tk.CAT containing three multimerised copies of site B oligonucleotide (lanes 1-4) or the parental vector pSP65.tk.CAT (lanes 5-8). Co-transfection with the eukaryotic expression vector pJ3Ω containing the RARα gene (lanes 3, 4, 7, 8) or the parental vector (lanes 1, 2, 5, 6) was performed and transfected cells incubated in the presence (lanes 3, 4, 6, 8) or absence (1, 3, 5, 7) of 10^{-6} M all-*trans* retinoic acid (ATRA).

Fig. 5.9

Chloramphenicol acetyltransferase activities in F9 cell transfections
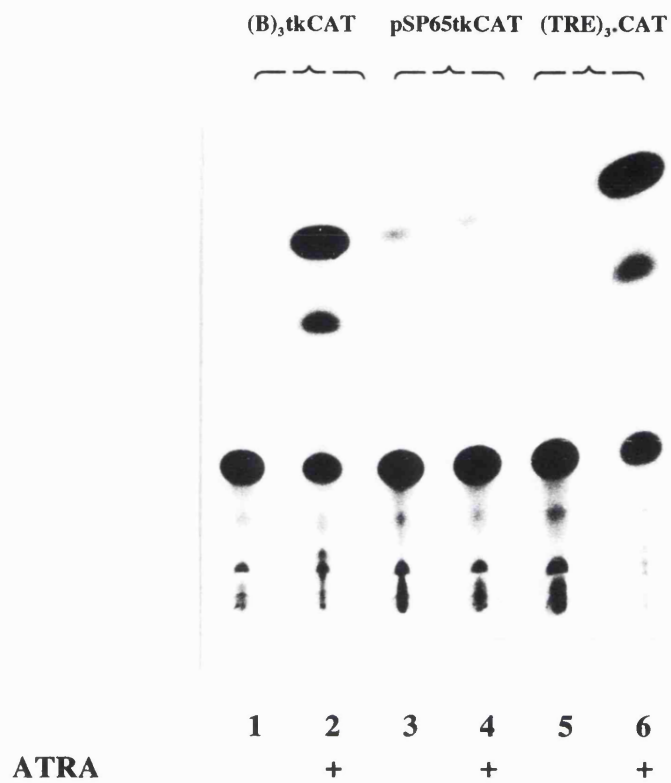
Assay of chloramphenicol acetyltransferase activity following the transfection of F9 cells with the reporter construct $(B)_3$.tk.CAT containing three copies of site B (lanes 1 and 2), the pSP65.tk.CAT expression vector alone (lanes 3 and 4) or the control reporter plasmid $(TRE)_3$.CAT (lanes 5 and 6). Following transfection, cells were incubated in the absence (lanes 1, 3 and 5) or presence of 10^{-6} M all-trans retinoic acid (ATRA) (lanes 2, 4,and 6).

Fig. 5.10

CAT assays in COS cells with the LTR.CAT reporter construct
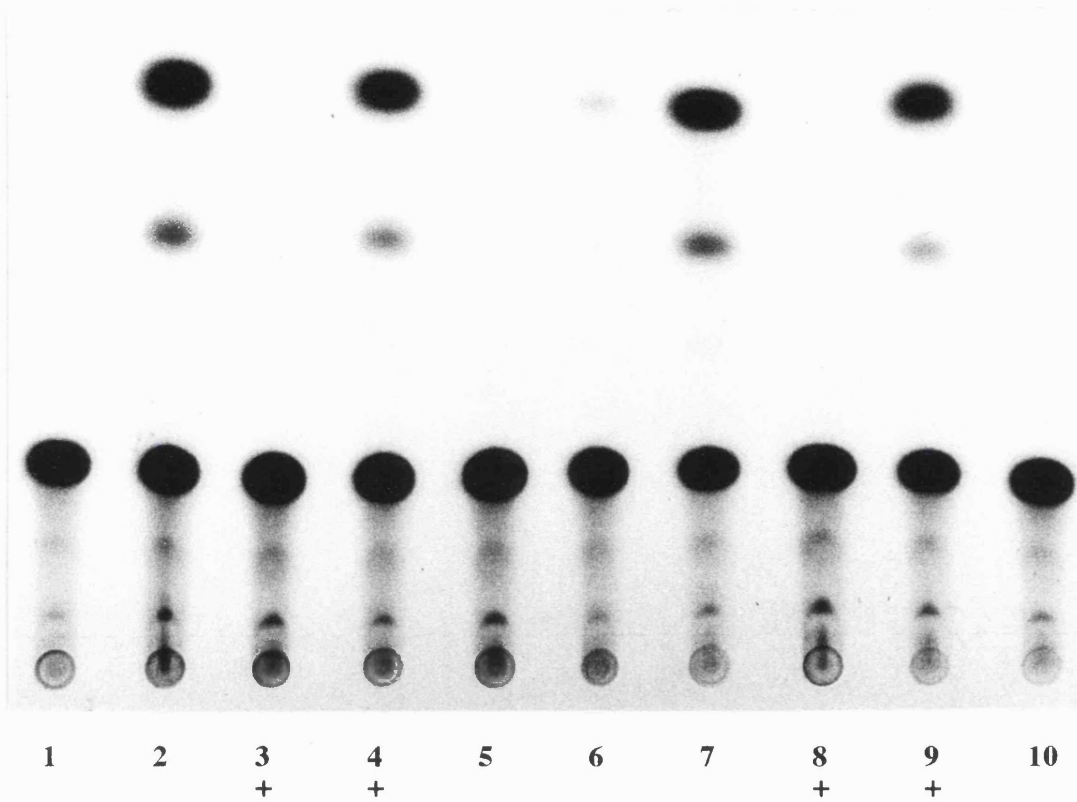
CAT activity in Cos cell extracts following transfection with the reporter construct LTR.CAT (lanes 1-5) or p2.1.CAT which contained mutations in the site B palindrome of the HIV-1 LTR (lanes 6-10). The following expression vectors were co-transfected: RARα expression vector (lanes 1-4 and 6-9); expression vector containing the Tat gene from HIV-1 (lanes 2, 4, 7 and 9). The addition of all-*trans* retinoic acid following transfection has been indicated by a '+' below the lane.

Fig. 5.11

Chloramphenicol acetyltransferase activities in F9 cell transfections
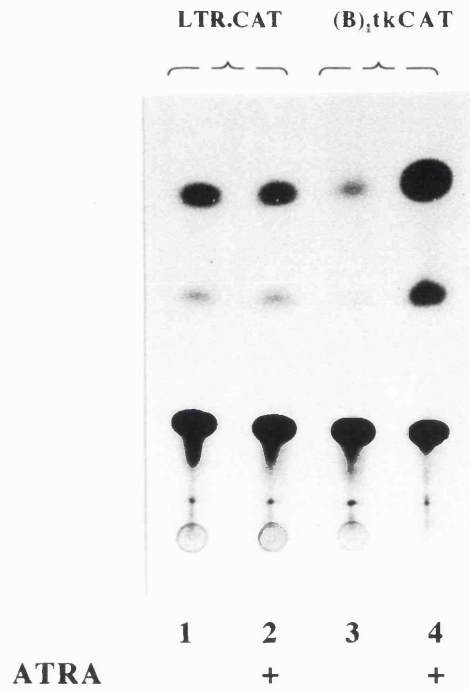
Assay of CAT activity following the transfection of F9 cells with the reporter construct LTR.CAT in which the human immunodeficiency virus type-1 long terminal repeat drives expression of the CAT gene (lanes 1 and 2) or with the construct (B)$_1$.tk.CAT containing a single copy of site B (lanes 3 and 4). Samples in lanes 2 and 4 were from cells incubated with 10^{-6} M all-*trans* retinoic acid (ATRA).

It was important to determine whether the intact HIV-1 LTR could show ATRA induced activation of the reporter gene. Figure 5.10 shows the results of transfections performed with the reporter constructs LTR.CAT and p2.1.CAT in Cos cells, the expression vectors for RARα and Tat were co-transfected. Although, as has been described, Tat could be shown to increase LTR-driven transcription in these experiments, no ATRA induction of CAT activity could be demonstrated. In F9 cells, although a single site B in a heterologous promoter was activated by ATRA (fig. 5.11, lanes 3 and 4), the LTR alone demonstrated a significant level of basal activity which was not further stimulated by the addition of ATRA (lanes 1 and 2). The basal activity of this construct (as compared with the basal activity of pSP65.tk.CAT and $(TRE)_3$.CAT in these cells) may have been due to the presence of constitutively expressed transcription factors or the activation of transcription factors by the transfection process, for example NF-κB.

5.3   Discussion:

As discussed in section 1.4.4.1, the binding of RARs and RXRs to HREs is complex. RXRs bind as heterodimers with THR, VD3R and RARs but can also bind as homodimers in the presence of their natural ligand 9-*cis* retinoic acid (Lehmann et al. 1992; Zhang et al. 1992). The effect upon transcription is also a function of the type of dimer (hetero vs homo), the presence of ligand, the configuration of the HRE involved and the presence of certain regulatory co-factors that exert either a stimulatory effect (co-activators) or repress transcription (co-repressors) (Chen and Evans 1995; Horlein et al. 1995; Kurokawa et al. 1993).

The results of the EMSA described above are consistent with these characteristics of RARs and RXR. In EMSA using recombinant RARs, all three subtypes of RAR (α, β and γ) were shown to bind weakly to site B sequences. Increased binding was demonstrated by RAR/RXR heterodimers (figure 5.5) although RXR homodimers could not form a complex with site B. This pattern has been described for other RAREs, RXR heterodimers only efficiently bind to RAREs in the presence of the natural ligand for RXR, 9-*cis* retinoic acid (9-*cis* RA) while RXR/RAR heterodimers can efficiently bind even in the absence of ATRA or 9-*cis* RA (Lehmann et al. 1992; Zhang et al. 1992). The EMSA described were performed without the addition of either ATRA or 9-*cis* RA which may account for the results observed.

Although site B was be shown to function as an RARE in the presence of ATRA, as a single site B oligonucleotide or as a multimer, the intact LTR was not inducible by ATRA. These observations may be due to one, or a combination, of the following points;

1)   site B may be a relatively weak RARE, in the context of the intact LTR, RARs binding to site B are unable to efficiently interact with the RNA pol II complex. When placed adjacent to the transcription initiation site, as in the heterologous promoter constructs, site B was able to influence more effectively the rate of transcription via RAR binding. This effect has been demonstrated for other RAREs, an effect now ascribed to the requirement for co-activators (Berkenstam et al. 1992).

2)   The co-transfection of an RXR expression vector would have allowed the binding of RAR/RXR heterodimers to site B, the addition of 9-*cis* RA as the natural ligand would have been more relevant. Unfortunately neither an RXR

expression vector nor 9-*cis* RA were available for inclusion in the transfections.

Following the description of site B as a potential RARE in the HIV-1 LTR (Orchard et al. 1993) several other groups confirmed these findings. Lee et al used recombinant RARs and RXR produced in reticulocyte lysate and were able to demonstrate binding of RAR/RXR heterodimers but not RAR homodimers to site B. RXR homodimers only bound to site B in the presence of 9-*cis* RA (Lee et al. 1994). They went on to demonstrate the importance of the TGACC half sites for RAR/RXR binding and the ability of recombinant COUP-TF protein to bind to the site. Functional analysis of site B was performed with transient transfections in CV-1 cells, a reporter construct containing two head-to-tail copies of a site B oligonucleotide adjacent to the HSVtk promoter and co-transfection of expression vectors for RARα, RARβ, RARγ and RXRα. The recombinant RARs alone were able to show a small increase of CAT activity with the addition of ATRA or 9-*cis* RA, however RXRα alone and RAR/RXRα combinations produced marked increases in CAT activity in cells treated with 9-*cis* RA. RXR alone was not significantly activated by ATRA, RARα/RXRα and RARβ/RXRα heterodimers were responsive to ATRA, but only a weak effect was shown with RARβ/RXRα heterodimers. They also assessed the ability for the two site B variant sequences (as described in chapter 4) to function as RAREs, showing that both variants functioned as efficient RAREs, indeed with slightly greater activation by ATRA and 9-*cis* RA than shown by the perfect site B palindrome. Using an HIV-1 LTR.CAT reporter construct they were able to show considerable induction of CAT activity by ATRA and 9-*cis* RA in cells co-transfected with RARα/RXRα, but only by 9-*cis* RA in cells co-transfected with RXRα alone. These results support the second point discussed above. The same group also demonstrated the ability of COUP-TF to repress RAR/RXR heterodimer activation of site B in the presence of ATRA or 9-*cis* RA. Their conclusions were that site B was a functional RARE which displayed unusual properties when compared with other RAREs.

As discussed in chapter 4, Ladias (Ladias 1994) described the ability for several members of the nuclear hormone receptor super family to bind to site B and concluded that the site formed part of a larger nuclear hormone response element (NRRE) although evidence for the importance and functionality of sequences adjacent to site B itself was lacking.

In contrast to the results described in this thesis and confirmed by Lee et al, Sawaya et al found that RARs and RXR caused an inhibition of LTR-driven transcription in a human oligodendroglioma cell line TC-620. They also

demonstrated a positive effect on LTR-driven CAT activity with a co-transfected COUP-TF expression vector, an effect inhibited by RAR or RXR (Sawaya et al. 1996a). Thus, in a different cell model, the functional roles of RAR, RXR and COUP-TF appear to be reversed.

The results described in this chapter confirm the ability for recombinant retinoic acid receptors to bind to site B. Functional analysis of site B proved that a single copy of the palindromic element was able to function as an RARE, in the context of a heterologous promoter. The lack of a more complete repertoire of reagents to analyse the importance of site B, especially 9-*cis* retinoic acid and an expression vector for RXR to include in transfection assays, meant that the optimal conditions for analysis of site B were not used. In particular, it was not possible to demonstrate the ability of site B to function as an RARE in the context of the intact HIV-1 LTR.

Chapter 6

General discussion and future directions.

Chapter 6
General discussion and future directions.

**6.1   Delineation of transcription factor binding sites in the
        negative regulatory element of the 5'LTR of HIV-1.**

The aims of the experimental work described in this thesis were to identify
binding sites for cellular transcription factors in the NRE of HIV-1, to
determine their functional significance and to characterise the cellular
factors interacting with them. At the time the project was initiated the NRE
was defined by deletion mapping of the LTR in transient transfection assays,
deletion of the NRE also resulted in an increase of viral replication using
infectious viral constructs. Although the exact boundaries of the NRE varied
between studies, it was apparent that deletion of the LTR upstream of
nucleotide -180 from the initiation site for transcription resulted in an
increase in LTR driven expression (Rosen et al '85; '88; Lu et al '89; '90;
Siekevitz et al '87; Garcia et al '87). The observed negative effect of the NRE
seemed likely to be due to specific *cis*-acting sequences binding cellular
factors.

Two new transcription factor binding sites were identified during this study,
site A from -379 to -366 and site B from -356 to -324 upstream of the
transcription start site. Although identified by DNase I footprint analysis of
the '220' fragment of the LTR using Jurkat nuclear extract, the cellular factors
that interacted with site A were neither characterised nor was the functional
significance of the site determined. Two independent groups have obtained
similar footprints over site A using nuclear extracts from Jurkat cells and
from several cell lines of neuronal and glial origin (Canonne Hergaux et al.
1995; Swingler et al. 1994). Schwartz et al identified three distinct proteins
present in nuclear extracts from Jurkat and HeLa cell lines binding to the core
sequence TGATTGGC forming four distinct DNA-protein complexes on EMSA
(Schwartz et al. 1997). Some characterisation of these proteins was reported; a
43 kDa protein was purified and sequenced, this appeared to represent a novel
factor as no homologous proteins were found on a data base search. A second
factor showed binding characteristics similar to that of NF1. Functional
analysis of site A suggested a small inhibitory effect of the site in the context
of the entire LTR, with a two-fold increase in reporter gene transcription with
four nucleotide substitutions in the core sequence.

6.2   Site B (-350 to -327).

The DNA-protein complex formed with site B and Jurkat nuclear extracts produced a consistently stronger footprint than that of site A and the predominant band on EMSA using the '220' LTR-fragment. Subsequent work was directed at the characterisation of site B and the identification of the factor(s) present in Jurkat nuclear extract interacting with the site.
Site B is a complex binding site, consisting of the heptanucleotide sequence TGACC repeated as a palindrome with a nine base-pair spacer region. For optimal binding both halves of the palindrome were required with some preferential binding to the 5'heptanucleotide sequence. Mutation of the spacer sequence did not affect formation of the complex using Jurkat nuclear extracts. Further evidence for the intimate interaction of protein with the palindrome sequences was provided by methylation interference experiments. Methylation of guanine bases in the palindrome disrupted protein binding, whereas methylation of those within the spacer region did not affect binding.

The organisation of site B showed a similarity to the hormone response elements, these are short DNA sequences that are recognised by members of the nuclear hormone receptor/retinoic acid receptor/COUP-TF family of transcription factors (the abbreviation 'NHRs' will be used to indicate this large transcription factor super-family). These elements are characterised by the presence of short sequences repeated as palindromes, direct repeats or inverted repeats. In particular, naturally occurring response elements for the oestrogen receptor (ER), thyroid hormone receptor (ThR), retinoic acid (RAR) and RXR receptors and COUP-TF which contain TGACC or closely related sequences (fig 6.1). I was able to demonstrate that the consensus elements for the ThR, ER RAR/RXR and COUP-TF could compete with site B for binding of the Jurkat nuclear factor.

Fig. 6.1   Sequence comparison of natural HREs and site B

Hormone response elements (HREs)

SITE B: 5'-CCAGGGGTCAGATATCCACTGACCTTTGG

TRE   5'-CGTGCGGTCACGTCCCTGATCTTACC

ERE   5'-AGTCAGGTCACAGTGACCTGAT

COUP   5'-AAGTTTGACCTTTGACCATAGA

RAREβ   5'-CCAGGGTTCACCGAAAGTTCACT

RARE-mCRBP   5'-GTAGGTCAAAAGGTCAGA

RARE apo-A1   5'-ACTGACCCCTGACCCCTGCCT

RARE (γF-crystallin)   5'-AGTGACCCTTTTAACCAGGTCAGTGA

SITE B: 5'-CCAGGGGTCAGATATCCACTGACCTTTG

Nucleotides forming the defined transcription factors binding sites have been highlighted in red.

These findings generated two related areas for investigation:

1) could site B function as a hormone response element?

2) was the protein present in Jurkat nuclear extracts a known nuclear hormone receptor, a new member of the NHR super-family or an unrelated protein that could interact with the consensus binding site?

6.3.1 Functional significance of site B sequences.

I had shown that site B could act as a modest repressor of HIV-1 LTR driven transcription. Site-directed mutagenesis of site B sequences in the palindrome resulted in a two-fold increase of expression of a reporter gene in Jurkat cells using a transient transfection assay model. The selection of the Jurkat cell line, in retrospect, may not have been ideal to demonstrate the potential for site B as a transcriptional repressor, as will be discussed later. Hormone response elements have been shown to function both as transcriptional activators and repressors, the overall influence of an element depending upon several factors. These include the organisation of the DNA-binding site (i.e. arrangement of repeated elements, presence of additional elements and flanking sequences), promoter structure and the cell context (cell type, activation state and differentiation stage).

The cell context perhaps provides the most complex, and hence most difficult to define, influence of a particular hormone response element on transcription. The effect of NHRs are dependent on the presence of the appropriate ligand, co-activators (steroid-receptor co-activators such as p/CIP, SCR-1) and co-repressors (N-CoR/RIP13 and SMRT/TAAC-2). Heterodimer formation, particularly with RXR, adds further complexity (reviewed by Beato et al. 1995; Mangelsdorf et al. 1995a; Montminy 1997). Of particular interest was the potential for site B to act as a retinoic acid response element. As discussed in chapter 5, retinoids have profound effects on HIV-1 replication and LTR activity in both T-cells and cells of the monocyte/macrophage lineage. I was able to demonstrate that recombinant retinoic acid receptors could specifically bind to site B in EMSA, binding was significantly enhanced in the presence of recombinant RXR (chapter 5). These findings were consistent with the enhanced binding of RXR/RAR heterodimers to RAREs compared to RAR or RXR homodimers (reviewed by Mangelsdorf and Evans 1995b). A single site B also functioned as a RARE in the context of a heterologous promoter construct when transfected into F9 cells. However, in the context of the intact LTR, I was unable to consistently demonstrate an induction of transcription with retinoic acid in the same cells.

The explanation for these results may lie in the complexity of the LTR in comparison to the constructed heterologous promoter, based on the pSP65 thymidine kinase expression vector. This is supported by the functional activity of a RARE present in the promoter of the RAR-β gene. Retinoic acid induction of this RARE in the context of the natural promoter requires the presence of specific cellular co-factors. In a heterologous promoter (the Herpes simplex thymidine kinase promoter) ligand activated transcription was not dependent on the presence of co-factors (Berkenstam et al. 1992). These co-factors have been subsequently identified and their role in enhancing the activating function of NHRs appears to be as bridging factors between the NHR and components of the transcription apparatus including the CREB-binding protein (CBP) and P300, proteins that are essential for the activation of transcription by a large number of transcription factors including NHRs (Chakravarti and al 1996; reviewed by Janknecht and Hunter 1996a and Janknecht, 1996b) and themselves interact with components of the basal transcription apparatus such as TFIIB. The function of these co-receptors and the facilitation of interaction of the NHRs with the basal transcription apparatus through them has yet to be fully determined. Possibly, by placing a RARE adjacent to the initiation site of transcription (the TATA box) the RXR/RAR heterodimer can directly interact with the transcription complex, i.e. the function of the cofactors may be to increase the effectiveness of distal and/or relatively 'weak' RARE sites. Site B in the context of the HIV-1 LTR may represent a relatively weak RARE. The consensus response element for the non-steroid members of the NHR superfamily consists of a minimal core hexamer AGGTCA usually occurring as a direct repeat (DR), inverted repeat (IR) or everted repeat (ER). The most potent of these HREs are direct repeats, in RAREs the two halves of the DR are separated by between 1 to 5 nucleotides (DR1 to DR5). However RAREs have been described that differ from this 'ideal' RARE, for example the RARE from the γF-crystalline promoter consists of an everted (A/G)GGTCA repeat with an 8 nucleotide spacer (Tini et al. 1993) i.e. TGACCC(8)AGGTCA. A RARE with an 8 nucleotide spacer separating direct repeat TGACC elements has been identified in the promoter of the H1° gene (Bouterfa et al. 1995). Site B does show some similarity to the γF-crystalline RARE, the site B half-site element can be considered as (A/G)GGTCA although the spacing remains as 9 nucleotides.

6.3.2 Subsequent studies by independent groups.

Following the initial description of site B several independent groups have confirmed and extended the characterisation of this complex binding site. Site B has been shown to be a functional COUP-TF site (Cooney et al. 1991). Cooney et al described the binding of COUP-TF from HeLa cell nuclear extracts to site B and suggested a major role for COUP-TF in the regulation of HIV-1 expression but, as will be discussed later, I could find no evidence for COUP-TF protein or mRNA in T cells. COUP-TF mRNA was present in some monocytic/macrophage cell lines and may play a significant role in the modulation of HIV-1 expression in these cells. Swingler et al (Swingler et al. 1994) used nuclear extracts from neural cell lines to produce a DNase I footprint over site B but did not characterise the proteins present in these cell types, rather they assumed that it was COUP-TF. Canonne-Hergaux et al (Canonne-Hergaux et al. 1995) extensively characterised the nuclear factors present in several cell lines of neural origin that interacted with site B. In EMSA the number of specific DNA-protein complexes formed varied between cell types, with four complexes formed using U373-MG (astrocytoma) or SK-N-MC (neuroblastoma), three with TC-620 (oligodendroma) or HeLa cells and two using extracts from Jurkat cells. All cell types produced DNase I footprints over site B. Their conclusion was that the proteins present in these nuclear extracts belonged to the NHR family of transcription factors and that COUP-TF was present in two of the complexes identified. However, that conclusion was based on supershift analysis using the same polyclonal 'anti-COUP-TF' as was used by Cooney et al which, as discussed in chapter 3, was not convincingly specific for COUP-TF. The same group have further investigated the role of site B in neuronal cells showing that RXRα, RARα and COUP-TF all interacted with the site (Sawaya et al. 1996a). Functional analysis of site B was performed in neuronal cell lines using co-transfection of RXRα or RARα expression vectors and an HIV-1 LTR CAT reporter construct, in the presence or absence of the natural ligands all-*trans* retinoic acid (ATRA) or 9-*cis* retinoic acid (9-*cis*-RA). The results revealed a complex interaction of these transcription factor and site B. In the absence of ligand both RARα and RXRα induced LTR-driven expression in TC-620 cells (oligodendroglioma) while the addition of ligand antagonised activation. Co-transfection of a COUP-TF expression vector and an LTR-CAT construct increased reporter gene expression 9-fold, the addition of RXRα or RARα resulted in a decrease in COUP-TF induction of expression. Furthermore, Sawaya et al examined the role of the COUP-TF/site B interaction in neuronal cell lines (Sawaya et al. 1996b). In TC-620 cells the co-transfection of a COUP-TF expression vector increased LTR-driven reporter gene expression 9-10 fold, no induction was seen with SK-N-MC or U373-MG cells. This COUP-TF induction of

HIV-1 LTR driven transcription in oligodendroglioma cells appears, in part, to be due to COUP-TF binding to site B and cooperativity of COUP-TF and Sp1 binding to the core promoter (Rohr et al. 1997).

Further evidence for site B as a functional RARE has been produced by Lee et al (Lee et al. 1994). Using EMSA they were able to confirm binding of RAR/RXR heterodimers and RXR homodimers to a site B oligonucleotide. RXRα homodimers were only able to interact with site B in the presence of 9-cis retinoic acid, a characteristic of RXR interaction with other naturally occurring RAREs (Lehmann et al. 1992; Zhang et al. 1992). Single base-pair mutations in either half of the palindrome abolished RAR/RXR and RXR binding. They also demonstrated the interaction of COUP-TF I and II with site B, possibly involving sequences in the spacer region as well as the palindrome, unlike the Jurkat protein. In a series of elegant transient transfection assays in CV-1 cells they were able to demonstrate that site B sequences could function as a RARE in a heterologous reporter construct (pBL-CAT2) when co-transfected with expression vectors for RARα, RARβ RXRα. The most effective combination for induction of expression was RXRα and 9-cis retinoic acid with 100-fold induction of reporter gene expression, all-trans retinoic acid was less effective. The expression of an LTR-driven reporter gene was enhanced by the co-transfection of RXRα and a combination of RXRα/RARα but only in the presence of retinoids, again 9-cis retinoic acid being the most active. Examination of the functionality of three natural LTR variants that contained single base pair changes in the palindromic elements of site B and demonstrated a difference in the relative level of induction by retinoids.

COUP-TF in these assay conditions was shown to act as a competitive inhibitor of RXRα and RARα/RXRα plus 9-cis retinoic acid induced activation of LTR-driven transcription. A synthetic retinoid antagonist, SR11335, effectively inhibited the induction of LTR-driven transcription by both all-trans and 9-cis retinoic acid. The conclusions from this work were as follows: 1) The sequences identified as site B could function as a RARE both in a heterologous promoter and in three natural LTRs; 2) COUP-TF I and II could interact competitively with the retinoic acid receptors for binding to site B, inhibiting the induction in expression by retinoids.

An extensive analysis of site B was performed by Ladias (Ladias 1994). Using recombinant proteins expressed in Cos-1 cells he demonstrated the interaction of several members of the NHR superfamily of transcription factors with site B, including COUP-TF I, COUP-TF II, ear-2, HNF-4 (hepatocyte nuclear factor 4),

peroxisome proliferator-activated receptor (PPAR), nerve growth factor-inducible protein B (NGFI-B), RARα and RXRα. He termed the site 'NRRE', for nuclear receptor response element. The 'NRRE' also interacted with heterodimers formed between RXRα and RARα, ear-2, and PPAR. The functional significance of these NHRs binding to site B was analysed in the human choriocarcinoma cell line JEG-3. HNF-1, NGFI-B and RXRα induced an approximately 2-fold expression of LTR-driven expression, the addition of 9*cis* or all-*trans* retinoic acid with RXRα produced a 3-4 fold induction while the combination of RXRα and RARα with ligand resulted in a 5-6 fold induction. Binding of NHRs was also dependant upon the strain of HIV-1 from which the LTR sequence was derived. The conclusions from this extensive analysis were that site B was functionally highly conserved and it could function as a hormone response element through which signals from a wide variety of pathways could converge. Similar complex NRRE sites present in other promoters have been described (Carter et al. 1994).

6.4.1 Site B as an enhancer and repressor element: an apparent paradox.

The existence of site B as a functional regulatory element in the LTR of HIV-1 has been confirmed by several independent groups and demonstrated to function not only as a potential retinoic acid response element but also as a complex site interacting with a large number of NHRs. The effect of site B on LTR-driven expression appears to be either positive or negative depending upon the precise context in which the analysis is performed, in particular the cell type used, the activation state of the cell and the presence of specific ligands/signals. This effect can in part be explained by the ability of several transcription factors, particularly those of the NHR-superfamily, to interact with the site. Many hormone response elements from naturally occurring promoters demonstrate a similar context-dependent function, operating either as transcription enhancers or inhibitors (reviewed by Leng et al 1995; Renkawitz 1990; Tsai and O'Malley 1994). Elements that interact with COUP-TF factors also demonstrate this duality of function,

COUP-TF has been shown to act as an important repressor of transcription in a wide range of promoters by competing with positively acting factors such as RXR, thyroid hormone receptor, HNF-4, oestrogen receptor or VD3 receptor (Carter et al. 1994; Cooney et al. 1992; Barger, 1997 and reviewed by Qui et al 1994). The repressor effect of COUP-TF factors has been shown to be due to several mechanisms; 1) passive repression, in which COUP-TF protein occupies an element recognised by (or overlapping/adjacent) an activating factor e.g.

RXR/RAR heterodimer. 2) active repression, COUP-TF binds competitively to specific elements and acts in cis to inhibit the formation of the transcription complex. Finally 3) transrepression, COUP-TF removes either activating factors (such as RXR) or co-factors by binding with them, either with or without DNA-binding (Achatz et al. 1997; Leng et al. 1996; reviewed by Tsai and Tsai 1997).

6.4.2 COUP-TFs as transcriptional activators.

Although the predominant function of COUP-TFs are as transcriptional silencers from shared transcription elements, COUP-TFs also behave as transcriptional activators. The original description of COUP-TF was as an activator of transcription from the chicken ovalbumin promoter where the response element is a DR-1 (Sagami et al. 1986). Other examples of COUP-TF as activators of transcription include the intestinal fatty acid binding protein (FABPI) promoter (DR-1 element) (Rottman and Gordon 1993), the rat insulin II gene promoter (a DR-6 element (Hwung et al. 1988), the arrestin gene promoter (a DR-7 element) (Lu, X. et al. 1994) in neuronal cells. COUP-TF interacts with a DR-7 element in the transferrin promoter to enhance transcription in Hep3B cell but not in Sertoli cells (Schaeffer et al. 1993) indicating that the cell context is important in determining the functional effect of COUP-TF. The co-transfection of a COUP-TF expression vector and site B in the context of an intact LTR, significantly increased transcription in JEG-3 (choriocarcinoma) and TC-620 (oligodendroglioma) cells (Ladias 1994; Sawaya et al. 1996b). In U373-MG cells COUP-TF had no effect on LTR-driven expression while in SK-N-MC cells activation of transcription only occurred if dopamine was present (Sawaya et al. 1996b).

The mechanisms that determine whether COUP-TF behaves as a transcriptional activator or repressor are not fully understood but parameters such as promoter context, presence of co-factors and co-repressors appear to be important. The orientation of the COUP-TF element may induce allosteric changes to the bound protein revealing activator or repressor domains involved in active repression and transrepression (Achatz et al. 1997; Leng et al. 1996).

6.5. Expression of potential site B binding factors in HIV-1 target cells.

Clearly, with the potential for numerous transcription factors to interact with site B, it is of some importance to be able to define more precisely the intracellular environment. Thus, one would need to identify which of these transcription factors are expressed in the cell type used to analyse the functionality of the element, and as these may interact in a competitive manner, their relative abundance. Furthermore, as members of the NHR superfamily require co-factors for optimal activity, and specific inhibitory factors can decrease the activity, the relative abundance of these accessory proteins should also be determined. With the increasing understanding of the complexity of these NHRs the naiveté of simple transient transfection experiments, even with the co-transfection of expression vectors for particular NHRs of interest, becomes apparent.

Table 6.1 shows the expression of specific NHR transcription factors, in HIV-1 target cells, that have been shown to interact with site B.

Table 6.1   Nuclear hormone receptor pattern in cell lines

Key:

Italics are used to indicate indirect evidence of a specific factor

e.g. from EMSA pattern or supershift EMSA

Direct evidence of a specific factor indicated by "Yes" or "low" in normal type, lack of evidence for specific factor by "negative" as determined by direct analysis using Western or Northern blot.

Cell line   Transcription   factor

T-cell: site B factor   COUP-TF   RARα   RARβ   RXRα

Jurkat	*Yes*	negative	?	?	?
H9	?	?	Yes	?	low
CEM	?	?	Negative	?	Negative
SLB-1	?	?	Yes	?	?

Monocytic:

HL60	*Yes*	Yes (low)	Yes	?	?
U937	*Yes*	Yes	Yes	?	Yes
THP-1	?	Yes (low)	Yes	?	Yes

CNS derived:

SK-N-MC	*Yes*	*Yes*	Yes	Yes	Yes
U373-MG	*Yes*	*Yes*	low	Negative	Yes
TC-620	*low*	*Yes*	Yes	Yes	Yes
Microglia 1	?	*Yes*	?	?	?

Cervical carcinoma:

HeLa	*low*	Yes	Negative	Negative	?

References:

(Geisen et al. 1997; Kizaki et al. 1990; Orchard et al. 1992; Rohr et al. 1997; Yamaguchie et al. 1994)

Although incomplete, the table does indicate the variation in the expression of the transcription factors that have been shown to interact with site B.

6.6.1 Characterisation of Jurkat protein interacting with site B.

Using three independent methods the protein present in Jurkat nuclear extracts that formed the most abundant DNA-protein complex with site B was shown to have a Mr of approximately 100-115 kDa (chapter 4). Using the same methods, the predominant proteins from HeLa cell nuclear extracts interacting with site B had Mr of 45-62 kDa, in agreement with the published estimates for COUP-TF I and II (Wang et al. 1991; Wang et al. 1989). Initially, Cooney et al (Cooney et al. 1991) suggested that the Jurkat protein was a variant high Mr form of COUP-TF. However no direct evidence for the presence of COUP-TF in the major Jurkat DNA-protein complex was given, except that of a supershift using 'anti-COUP-TF' antisera. I was not able to detect the presence of COUP-TF protein in nuclear extracts from Jurkat cells by Western blot. A 350 base pair DNA probe from the highly conserved DNA-binding domain of COUP-TF failed to detect COUP-TF specific mRNA in Jurkat cells. The same group later confirmed these findings and suggested that the 'high COUP-TFs' seen in Jurkat nuclear extracts were produced by 'another gene' (Wang et al. 1991). Given the affinity of site B for numerous transcription factors, particularly those in the NHR superfamily, this later statement may be more correct. The evidence presented in this thesis shows that the Jurkat protein is not sufficiently related to COUP-TF either in size or sequence homology (in the DNA-binding domain) to be considered a member of the COUP-TF branch of the NHR superfamily.

6.6.2 Physical characteristics of site B protein/DNA interaction:

The Jurkat protein bound to site B as a dimer and the size of the DNA-protein complex, under the non-denaturing conditions used in the Fergusson assay, indicated that no other proteins were associated with the site B binding protein. It seems likely that each monomeric unit of the Jurkat protein interacted with a TGACC half-site of the palindromic unit. The spacer sequences between palindromic elements were not involved in the interaction with the Jurkat protein. There appeared to be steric constraints to the binding of the Jurkat protein to site B, as demonstrated by altering the number of base pairs in the spacer between half-site elements. Removing the spacer totally (zero base pairs) had no effect on binding but the introduction of single base pairs resulted in a step-wise reduction with 4 base pairs spacer reducing the interaction of the Jurkat protein with the element to 30% of the natural nine base-pair palindrome. Further increase of the spacer size restored binding. The effect on binding of the Jurkat protein with the 4-bp spacer oligonucleotide was similar to that seen with oligonucleotides containing half-site mutations. The possible explanation for these observations are that

the Jurkat protein was able to effectively interact with only a single half-site at a time in the 4-bp spacer oligonucleotide, the other half-site being now positioned on the opposite side of the double-helix and unavailable for binding. This is in contrast to the binding of RAR and RXR to DR-1 up to DR-5 elements, either the RAR and RXR dimers are interacting with half-sites only in the DR-4 elements or are less sterically constrained than the Jurkat protein.

6.6.3 Attempts to isolate the Jurkat site B binding protein.

The complete characterisation of a transcription factor requires the isolation of the complementary DNA encoding the protein allowing comparison with known transcription factors. Three approaches to achieve this aim were attempted;

1) screening a Jurkat cDNA library cloned into a λgt11 expression vector using a concatenated site B oligonucleotide probe (from the method described by Singh et al. 1988 and Vinson, 1988);

2) screening a Jurkat cDNA library with a degenerative oligonucleotide probe based on sequences in the conserved first zinc-finger of the DNA-binding domain of COUP-TF/ThR /hER/RAR/hVD3R.

3) screening a Jurkat library with a 350 base pair DNA probe derived from the DNA-binding domain of COUP-TF I

None of these methods were successful.

The predominant protein present in nuclear extracts from Jurkat cells and T lymphocytes that interacts with site B in the 5'LTR of HIV-1 still awaits complete identification.

6.7 Model for the role of site B in transcription driven by the HIV-1 LTR in T lymphocytes and other HIV-1 target cell types.

Figure 6.2 shows a possible model for the role of site B and cellular transcription factors shown to interact with this element. In Jurkat cells mutation of site B sequences resulted in a small increase in LTR-driven transcription. This suggests that the predominant factor binding to the site had a net inhibitory influence on transcription in these cells. This may have been due to the exclusion of positive transactivating factors from site B or by a direct effect of the unknown protein on the transcriptional apparatus. In other cell types, in which the relative abundance of factors capable of interacting with site B was different to that in T lymphocytes, site B may interact with factors such as RAR/RXR heterodimers or COUP-TF which can act as transcriptional activators, operating through co-activator, or even transcriptional repressors, via co-repressors. For NHRs which have ligands the presence or absence of the ligand may also influence transcription.

Fig. 6.2

Proposed model for role of site B and cellular factors on
LTR-driven transcription

1. In T-cells, undefined site B protein predominates
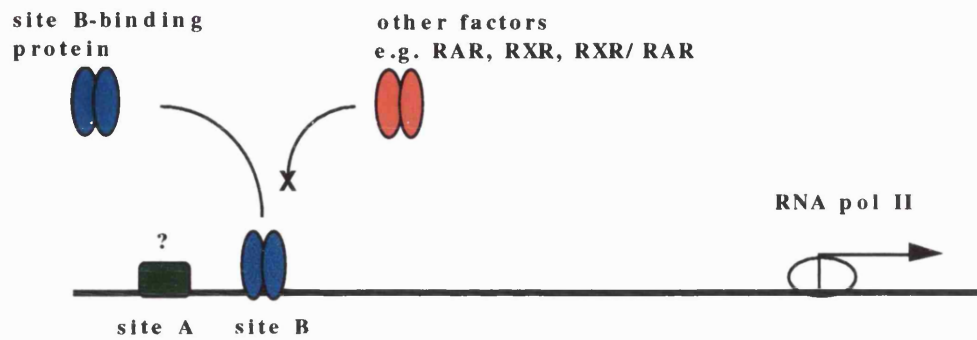
2a. monocytic and neuronal cells
less site B protein relative to other factors
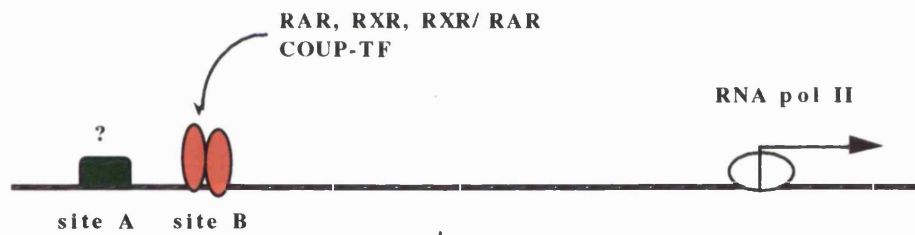
2b.
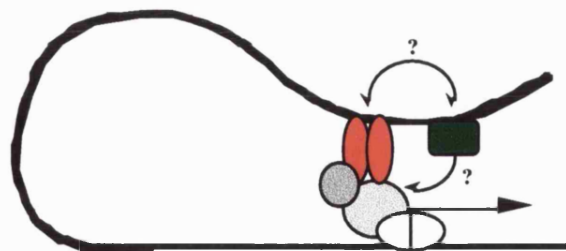
CBP/ p300

co-activator or co-repressor

6.8   Future directions.

Since my original publication of the two sites A and B much of what would have been proposed for the future investigation of the two sites has now been performed, including confirmation of the sites by other groups and the identification of cellular factors capable of interacting with them.

6.8.1 Factors interacting with sites A and B.

Clearly in order to fully understand the function of the sites A and B in the natural target host cells of HIV-1 all of the factors present in these cells able to interact with the sites need to be identified. It has already been demonstrated that many members of the steroid hormone/thyroid hormone/retinoic acid superfamily of transcription factors interact with the sequences of site B. Several of these are naturally expressed in HIV target cells. However the predominant factor present in nuclear extracts from Jurkat cells and IL-2 expanded peripheral blood T lymphocytes has not yet been identified despite attempts to do so. The isolation of cDNA representing this protein is required in order to investigate its' role not only in the regulation of HIV-1 expression but also of cellular genes. With the identification of the cDNA and site B protein, reagents can be generated to investigate the expression of the mRNA and protein in HIV-1 target cells, in cell lines and more importantly in primary cells and tissues.

The expression of the members of the NHRs superfamily of transcription factors that have been shown to interact with site B should be determined in a wider range of cell types and in primary cells and tissues. The expression of NHR co-factors and co-repressors in these cells should similarly be defined. In this way the cellular environments that influence LTR activity via site B can be defined and provide some insight into the role of site B *in vivo*.

6.8.2 Functional analysis of sites A and B.

Transient transfection assays, although providing useful data, are too crude to elucidate the functional significance of relatively weak regulatory elements. Plasmid DNA will not be incorporated into the nuclear matrix or chromatin structure and this may have relevance for particular regulatory elements. In addition, many transient transfection methods activate NF-κB particularly in lymphocytic and monocytic cells, swamping the subtle influences of other elements. Cell lines containing stably integrated LTR-reporter gene constructs may help to avoid the potential experimental artefacts of transient transfection assay, the activity of the integrated LTR-reporter gene should be examined in a wide range of T lymphocyte, monocytic and neuronal cell lines. Defined minimal mutations can be introduced to sites A and B in a parallel set

of stable lines. The influence of the various ligands for the NHRs, for example all-*trans* and 9-*cis* retinoic acid, can be examined in these experimental models.

The functional significance of sites A and B need to be explored in more detail and in the context of the integrated provirus in appropriate host cell environments. There is increasing evidence for the interaction of transcription elements within a single promoter, by the influence of adjacent or distant sequences, through the proteins binding to them, hence the need to study these sites not only in isolation but in the context of the natural promoter. The complexity of action of the NHRs and the proven affinity of site B for members of this family of transcription factors raises important questions for the role of these widely expressed transcription factors in the regulation of HIV-1 transcription. The complex way in which members of the NHR super-family regulate gene expression is slowly being elucidated, in so doing an explanation for the apparently conflicting results from experiments with NHR elements and NHRs has become apparent.

Certainly the major influences on HIV-1 expression are mediated through the core promoter, enhancer elements and the TAR region which provide the essential basal transcription elements, cell activation-responsive elements (the NF-κb sites) and a viral-specific enhancer (TAR). However, a considerable degree of redundancy appears to exist in the LTR with potential binding sites for numerous cellular transcription factors now identified. The role of these sites in the global regulation of HIV-1 expression is far from clear. They may allow the integrated provirus to regulate expression over a wider range of cellular environments, for example where the potent enhancer factors such as NF-κb are down-regulated, or to respond when these signals are not activated. The message for those who continue to study the expression of HIV-1 is that, as has been shown for other aspects of HIV-1, this is a highly sophisticated virus with a subtle and complex life cycle and there are certain to be important details of its' interaction with the host still to be discovered.

Bibliography.

Aboulafia, D. M. 1994. Human immunodeficiency virus-associated neoplasms: epidemiology, pathogenesis, and review of current therapy. *Cancer Pract* 2:297-306.

Achatz, G., B. Holzl, R. Speckmeyer, C. Hauser, F. Sandhofer, and B. Weber. 1997. Functional domains of the human orphan receptor ARP-I/COUP-TF II involved in active repression and transrepression. *Mol Cell Biol* 17:4914-4932.

Agostini, I., J.M. Navarro, F. Rey, M. Bouhamdan, B. Spire, R. Vigne, and J. Sire. 1998. The human immunodeficiency virus type 1 vpr transactivator: cooperation with promoter-bound activator domains and binding to TFIIB. *J Mol Biol* 261:599-606.

Ahmad, N., and S. Venkatesan. 1988. Nef protein of HIV is a transcriptional repressor of HIV-1 LTR. *Science* 241:1481-1485.

Aiken, C., J. Konner, N.R. Landau, M.E. Lenburg, and D. Trono. 1994. Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell* 76:853-864.

Alcami, J., T. Lain de Lera, L. Folgueira, M. A. Pedraza, J. M. Jacque, F. Bachelerie, A. R. Noriega, R. T. Hay, D. Harrich, R. B. Gaynor, and et al. 1995. Absolute dependence on kappa B responsive elements for initiation and Tat-mediated amplification of HIV transcription in blood CD4 T lymphocytes. *EMBO J* 14:1552-1560.

Alkhatib, G., C. Combadiere, C.C. Broder, Y. Feng, P.E. Kennedy, P.M. Murphy, and E.A. Berger. 1996. CC-CKR5: a RANTES, MIP-1 beta receptor as a fusion cofactor for macrophage tropic HIV-1. *Science* 272:1955-1958.

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* 66:4617-4621.

Andreasson, P. A., F. Dias, A. Naucler, S. Andersson, and G. Biberfeld. 1993. A prospective study of vertical transmission of HIV-2 in Bissau, Guinea-Bissau. *Aids* 7:989-993.

Antoni, B. A., S. B. Stein, and A. B. Rabson. 1993. Regulation of Human Immunodeficiency Virus Infection: Implications for Pathogenesis. *Advances in Virus Research* 43:53-145.

Arya, S. K., C. Guo, S. F. Josephs, and F. Wong-Staal. 1985. Transactivation gene of human T lymphotropic virus type III (HTLV-III). *Science* 229:69-73.

Ashorn, P.A., E.A. Berger, and P. Moss. 1990. Human immunodeficiency virus envelope glycoprotein/CD4 mediates fusion of non-primate cells with human cells. *Journal of Virology* 64:2149-2156.

Bachelerie, F., J. Alcami, F. Arenzana-Seisdedos, and J. L. Virelizier. 1991. HIV enhancer activity perpetuated by NF-kappa B induction on infection of monocytes. *Nature* 350:709-712.

Baeuerle, P. A., and T. Henkel. 1994. Function and activation of NF-κB in the immune system. *Annu. Rev. Immunol.* 12:141-179.

Bagasra, O., E. Lavi, L. Bobroski, K. Khalil, JP. Pestaner, R. Tawadros, and RJ. Pomerantz. 1996. Cellular reservoirs of HIV-1 in the central nervous system of infected individuals: identification by the combination of *in situ* polymerase chain reaction and immunohistochemistry. *AIDS* 10:573-585.

Baldwin, A. S., K. P. Leclair, H. Singh, and P. A. Sharp. 1990. A large protein containing zinc finger domains binds to related sequence elements in the enhancers of the class I major histocompatibility complex and kappa immunoglobulin genes. *Mol Cell Biol* 10:1406-1414.

Balter, M. 1996. HIV's other immune-system targets: macrophages. *Science* 274:1464-65.

Bangham, C., S. Daenke, R.E. Phillips, J.K. Cruickshank, and J.I. Ball. 1988. Enzymatic amplification of exogenous retroviral sequences from DNA of patients with tropical spastic paraparesis. *EMBO J* 7:4179-4184.

Baniahmad, A., C. Steiner, A. C. Kohne, and R. Renkawitz. 1990. Modular Structure of a Chicken Lysozyme Silencer: Involvement of an Unusual Hormone Receptor Binding Site. *Cell* 61:505-514.

Barger, P.M., and D.P. Kelly. 1997. Identification of a retinoid/chicken ovalbumin upstream promoter transcription factor response element in the human retinoid X receptor γ2 gene promoter. *J Biol Chem* 272:2722-2728.

Barre-Sinoussi, F., J.C. Chermann, F. Rey, M.T. Nugeyre, S. Chamaret, J. Gruest, C. Dauget, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for AIDS. *Science* 220:868-870.

Bartel, D.P., M.L. Zapp, M.R. Green, and J.W. Sodroski. 1991. HIV-1 rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell* 67:529-536.

Bassuk, A.G., R.T. Anadappa, J.M. Leiden. 1997. Physical interactions between ets and NF-kappaB/NFAT proteins play an important role in their cooperative activation of the human immunodeficiency virus in T cells. *J Virol* 71:3563-3573.

Baur, A.S., E.T. Sawai, P. Dazin, W.J. Fantl, C. Cheng-Meyer, and B.M. Peterlin. 1994. HIV-1 Nef leads to inhibition or activation of T-cells depending on its intracellular localisation. *Immunity* 1:373-384.

Beato, M., P. Herrlich, and G. Schutz. 1995. Steroid hormone receptors: many actors in search of a plot. *Cell* 83:851-857.

Bednarik, D. P., J. A. Cook, and P. M. Pitha. 1990. Inactivation of the HIV LTR by DNA CpG methylation:evidence for a role in latency. *EMBO J* 9:1157-1164.

Berkenstam, A., M. D. M. Vivanco-Ruiz, D. Barettino, M. Horikoshi, and H. G. Stunnenberg. 1992. Cooperativity in transactivation between retinoic acid receptor and TFIID requires an activity analogous to E1A. *Cell* 69:401-412.

Berkhout, B., and K-T. Jeang. 1992. Functional roles for the TATA promoter and enhancers in basal and Tat-induced expression of the human immunodeficiency virus type i long terminal repeat. *J. Virol* 66:139-149.

Biberfield, P., K.J. Chayt, L.M. Marselle, G. Biberfield, R.C. Gallo, and M.E. Harper. 1986. HTLV-III expression in infected lymph nodes and relevance to pathogenesis of lymphadenopathy. *Am J Pathol* 125:436-442.

Biggar, R.J. 1990. AIDS incubation in 1891 HIV seroconverters from different exposure groups. International Registry of Seroconverters. *AIDS* 4:1059-1066.

Biggin, MD., TJ. Gibson, and GF. Hong. 1983. Buffer gradient gels and S-35 label as an aid to rapid DNA sequence determination. *PNAS* 80:3963-65.

Bjorklund, S., and Y-J. Kim. 1996. Mediator of transcriptional regulation. *TIBS* 21:335-337.

Blanchard, A., and L. Montagnier. 1994. AIDS-associated mycoplasmas. *Annu Rev Microbiol* 48:687-712.

Blattner, W.A. 1991. HIV epidemiology: past, present and future.
FASEB J 5:2340-2348.

Bleul, C.C., M. Farzan, H. Choe, C. Parolin, I. Clark-lewis, J. Sodroski, and T.A. Springer. 1996. The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature* 382:829-833.

Bour, S., U. Schubert, and K. Strebel. 1995. The human immunodeficiency virus type 1 vpu protein specifically binds to the cytoplasmic domain of CD4: implication for the mechanism of degradation. *J Virol* 69:1510-1520.

Bouterfa, H.L., F.J. Piedrafita, D. Doenecke, and M. Pfahl. 1995. Regulation of H1o gene expression by nuclear receptors through an unusual response element:implications for regulation of cell proliferation.
DNA and Cell Biology 14:909-919.

Bryan, J.K. 1977. Molecular weights of protein multimers from polyacrylamide gel electrophoresis. *Analytical Biochem* 78:513-519.

Bryant, M.L., R.O. Henkeroth, J.T. Kimata, L. Ratner, and J. Gordon. 1989. Replication of Human Immunodeficiency Virus 1 and Moloney Murine Leukaemia Virus is inhibited by different hetero-atom containing analogues of myristic acid. *PNAS* 86:8655-8659.

Burley, S.K., and R.G. Roeder. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annual Review of Biochemistry* 6:769-799.

Canonne Hergaux, F., D. Aunis, and E. Schaeffer. 1995. Interactions of the transcription factor AP-1 with the long terminal repeat of different human immunodeficiency virus type 1 strains in Jurkat, glial, and neuronal cells.
J Virol 69:6634-42.

Cao, Y., L. Qin, L. Zhang, J. Safrit, and D.D. Ho. 1995. Virologic and immunologic characterisation of long term survivors of human immunodeficiency virus type 1 infection. *N Eng J Med* 332:201-206.

Carreer, R., H. Groux, J. C. Ameisen, and A. Capron. 1994. Role of HIV-1 Nef expression in activation pathways in CD4+ T cells.
AIDS Res Hum Retroviruses 10:523-527.

Carter, M.E., T. Gulick, D.D. Moore, and D.P. Kelly. 1994. A pleiotropic element in the medium chain acyl-coenzyme A dehydrogenase gene promoter mediates

transcriptional regulation by multiple nuclear receptor transcription factors and defines novel receptor-DNA binding motifs. *Mol Cell Biol* 14:4360-4372.

Catovsky, D., M.F. Greaves, M. Rose, D.A.G. Galton, A.W.G. Goolden, D.R. McCluskey, J.M. White, I. Lampert, G. Bourikas, R. Ireland, J.M. Bridges, W.A. Blattner, and R.C. Gallo. 1982. Adult T-cell lymphoma-leukaemia in Blacks from the West Indies. *Lancet* 1:639-643.

Centers for Disease Control. 1981. Pneumocystis pneumonia - Los Angeles. *Morbidity and Mortality Wkly Rep* 30:250-252.

Centers for Disease Control. 1993.
Morbidity and Mortality Wkly Rep 42:869-71.

Centers for Disease Control. 1995.
HIV/AIDS Surveillance Report 7:1-39.

Centres for Disease Control. 1997. From the Centers for Disease Control. Transmission of HIV possibly associated with exposure of mucous membrane to contaminated blood. *JAMA* 278:539-41.

Chakravarti, D., and et al. 1996. Role of CBP/P300 in nuclear receptor signalling. *Nature* 383:99-103.

Chang, P. Y., K. Stellrecht, S. Melana, and B. G. Pogo. 1994. Elements in the long terminal repeat of HIV-1 that interact with nuclear extracts from Jurkat cells persistently infected with vaccinia virus. *Virus Res* 34:127-38.

Cheingsong-Popov, R., C. Panagiotid, M. Ali, S. Bowcock, P. Watkins, A. Aronstam, M. Wassef, and J. Weber. 1990. Antibodies to HIV-1 nef (p27): prevalence, significance and relationship to seroconversion.
AIDS Res and Human Retroviruses 6:1099-1125.

Chen, C., and H. Okayama. 1987. High Efficiency Transformation of Mammalian Cells by Plasmid DNA. *Molecular and Cellular Biology.* 7:2745-2752.

Chen, F.E., D.B. Huang, Y.Q. Chen, and G. Ghosh. 1998
Crystal structure of p50/p65 heterodimer of transcription factor NF-kappaB bound to DNA. *Nature* 391:410-413.

Chen, J.D., and R.M. Evans. 1995. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* 377:454-457.

Chen X., U Vinkemeier, Y. Zhoa, D. Jeruzalmi, JE. Darnell Jr., J. Kuriyan. 1998. Crystal structure of a tyrosine phosphorylated STST-1 dimer bound to DNA. *Cell* 93:827-839

Chen, Z , P Telfer, A Gettie, P Reed, L Zhang, D Ho, and P Marx. 1996. Genetic characterisation of new West African Simian Immunodeficiency Virus SIVsm: Geographic clustering of household-derived SIV strains with Human Immunodeficiency Virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *Journal of Virology* 70:3617-3627.

Cheng-Meyer, C., C. Weiss, D. Seto, and Levy JA. 1989. Isolates of human immunodeficiency virus type 1 from the brain may constitute a special group of the AIDS virus. *Proc Natl Acad Sci USA* 86:8575-8579.

Chiang, C-M., and R. G. Roeder. 1995. Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. *Science* 267:531-536.

Chiang, C. M., H. Ge, Z. Wang, A. Hoffmann, and R. G. Roeder. 1993. Unique TATA-binding protein-containing complexes and cofactors involved in transcription by RNA polymerases II and III. *EMBO J* 12:2749-62.

Choe, H., M. Farzan, Y. Sun, N. Sullivan, B. Rollins, P.D. Ponath, L. Wu, C.R. Mackay, G. La Rosa, W. Newman, N. Gerard, C. Gerard, and J. Sodroski. 1996. The beta-chemokine receptors CCR-3 and CCR-5 facilitate infection by primary HIV-1 isolates. *Cell* 85:1135-1148.

Chomcznski, P. and N. Sacchi. 1987. Single step method of RNA isolation by acid guanidinium-phenol-chloroform extraction. *Analytical Biochem.* 162:156-159.

Chou, C.S., O. Ramilo, and E.S. Vitetta. 1997. Highly purified CD25 negative T cells cannot be infected de novo with HIV-1. *PNAS* 94:1361-1365.

Clapham, PR., JN. Weber, D. Whitby, K. McIntosh, AG. Dalgleish, PJ. Maddon, KC. Deen, RW. Sweet, and RA. Weiss. 1986. Soluble CD4 blocks the infectivity for diverse strains of HIV and SIV for T-cells and monocytes but not for brain and muscle cells. *Nature* 337:368-370.

Clapham, P.R., and R.A. Weiss. 1997. Spoilt for choice of co-receptors. *Nature* 388:230-231.

Clouse, K.A., D. Powell, I. Washington, G. Poli, K. Strebel, W. Farrar, B. Barstad, J. Kovacs, A.S. Fauci, and T.M. Folks. 1989. Monokine regulation of HIV-1 expression in a chronically-infected human T cell clone.
J Immunol 142:431-438.

Coffin, J. 1991. Retroviruses and their replication, chapter 27. In *Fundamental Virology*, edited by B. N. Fields and D. M. Knipe. Raven Press Ltd: New York.

Coffin, J., A. Haase, J. A. Levy, L. Montagnier, S. Oroszlan, N. Teich, H. Temin, K. Toyoshima, H. Varmus, P. Vogt, and et al. 1986. Human immunodeficiency viruses [letter]. *Science* 232:697.

Coffin, J.M. 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis and therapy. *Science* 267:483-489.

Coffin, J.M. 1996. HIV viral dynamics. *AIDS* 10 (suppl 3):S75-S84.

Cohen, E.A., E.F. Terwilliger, Y. Jalinoos, J. Proulx, J.G. Sodroski, and W.A. Haseltine. 1990. Identification of HIV-1 *vpr* product and function.
Journal of Acquired Immune Deficiency Syndrome 3:11-18.

Colebunders, R.L., and A.S. Latif. 1991. Natural history and presentation of HIV-1 infection in adults. *AIDS* 5 (suppl):s103-s112.

Colgan, J., H. Ashali, and J.L. Manley. 1995. A direct interaction between a glutamine-rich activator and the N-terminus of TFIIB can mediate transcriptional activation in vivo. *Mol Cell Biol* 15:2311-2320.

Collins, K.L., B.K. Chen, S.A. Kalams, B.D. Walker, and D. Baltimore. 1998. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* 391:397-401.

Cooney, A. J., S. Y. Tsai, B. W. O'Malley, and M-J. Tsai. 1991. Chicken ovalbumin upstream promoter transcription factor binds to a negative regulatory region in the human immunodeficiency virus type 1 long terminal repeat.
J Virol 65:2853-2860.

Cooney, A.J., S-Y. Tsai, B.W. O'Malley, and M-J. Tsai. 1992. Chicken ovalbumin upstream promoter transcription factor (COUP-TF) dimers bind to different GGTCA response elements, allowing COUP-TF to repress hormonal induction of the vitamin D3, thyroid hormone and retinoic acid receptors.
Mol Cell Biol 12:4153-4163.

Cooney, A.J., X. Leng, S.Y. Tsai, B.W. O'Malley, and M-J. Tsai. 1993. Multiple mechanisms of chicken ovalbumin upstream promoter transcription factor-dependent repression of transactivation by the vitamin D3, thyroid hormone and retinoic acid receptors. *J Biol Chem* 268:4152-4160.

Cornelissen, M., G. Mulder-Kampinga, J. Veenstra, F. Zorgdrager, C. Kuiken, S. Hartman, J. Dekker, L. Van der Hoek, C. Sol, R. Cautinho. 1995. Syncytium-inducing (SI) phenotype suppression at seroconversion after intramuscular inocculation of a non-syncytium inducing/SI phenotypically mixed human immunodeficiency virus type 1 population. *J Virol* 69:1810-1818.

Crabtree, G. R. 1989. Contingent genetic regulatory events in T-lymphocyte activation. *Science* 243:355-361.

Cujec, T. P., H. Cho, E. Maldonado, J. Meyer, D. Reinberg, and B. M. Peterlin. 1997a. The human immunodeficiency virus transactivator Tat interacts with the RNA polymerase II holoenzyme. *Mol Cell Biol* 17:1817-1823.

Cujec, T. P., H. Okamoto, K. Fujinaga, J. Meyer, H. Chamberlin, D. O. Morgan, and B. M. Peterlin. 1997b. The HIV transactivator Tat binds to the CDK-activating kinase and activates the phosphorylation of the carboxy-terminal domain of RNA polymerase II. *Genes and Development* 11:2645-57.

Cullen, B.R. 1986. Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. *Cell* 46:973-982.

Cullen, B.R., J. Hauber, K. Cambell, J.G. Sodroski, W.A. Haseltine, and C.A. Rosen. 1988. Subcellular localisation of the human immunodeficiency virus *trans*-acting *art* gene product. *J Virol* 62:2489-2501.

Cullen, B. R. 1991. Regulation of HIV-1 gene expression.
FASEB J 5:2361-2368.

Cullen, B. R. 1992. Mechanism of action of regulatory proteins encoded by complex retroviruses. *Microbiol Rev* 56:375-394.

Cullen, B. R. 1994. RNA-sequence-mediated gene regulation in HIV-1.
Infect Agents Dis 3:68-76.

Cullen, B. R. 1996. Virology. New trick from an old foe. *Nature* 379:208-209.

Dabis, F., P. Msellati, D. Dunn, P. Lepage, M. L. Newell, C. Peckham, and P. Van de Perre. 1993. Estimating the rate of mother-to-child transmission of HIV. Report

of a workshop on methodological issues Ghent (Belgium), 17-20 February 1992. The Working Group on Mother-to-Child Transmission of HIV. *AIDS* 7:1139-1148.

Dale, B. A. 1989. Origins of the AIDS viruses, HIV-1 and HIV-2: fact or fiction. *J R Soc Med* 82:243-4.

Dalgleish, A.G., P.C.L. Beverley, P.R. Clapham, D.H. Crawford, M.F. Greaves, and R.A. Weiss. 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS virus. *Nature* 312:763-767.

Davie, J.R. 1997. Nuclear matrix, dynamic histone acetylation and transcriptionally active chromatin. *Mol Biol Reports* 24:197-207.

Davis, LE., BL. Hjelle, VE. Miller, DL. Palmer, AL Llewellyn, and et al. 1992. Early viral brain invasion in iatrogenic human immunodeficiency virus infection. *Neurology* 42:1736-1739.

Dayton, A.I., J.G. Sodroski, C.A. Rosen, W.C. Goh, and W.A. Haseltine. 1986. The transactivator of the human T cell lymphotrophic virus type III is required for replication. *Cell* 44:941-947.

Dayton, E.T., D.A. Konings, D.M. Powell, B.A. Shapiro, L. Butini, J.V. Maizel, and A.I. Dayton. 1992. Extensive sequence-specific information through the CAR/RRE, the target sequence of the human immunodeficiency virus type 1 Rev protein. *J. Virol* 66:1139-1151.

De Cock, K. M., G. Adjorlolo, E. Ekpini, T. Sibailly, J. Kouadio, M. Maran, K. Brattegaard, K. M. Vetter, R. Doorly, and H. D. Gayle. 1993. Epidemiology and transmission of HIV-2. Why there is no HIV-2 pandemic. *JAMA* 270:2083-2086.

de The, H., M.D.M. Vivanco-Ruiz, P. Tiollais, H. Stunnenberg, and A. Dejean. 1990. Identification of a retinoic acid responsive element in the retinoic acid receptor β gene. *Nature* 343:177-180.

de Vincenzi, I. 1994. A longitudinal study of human immunodeficiency virus transmission by heterosexual partners. European Study Group on Heterosexual Transmission of HIV. *N Engl J Med* 331:341-346.

Deacon, N.J., A. Tsykin, A. Solomon, K. Smith, M. Ludford-Menting, D.J. Hooker, D.A. McPhee, A.L. Greenway, A. Ellett, C. Chatfield, and et al. 1995. Genomic structure of an attenuated quasi-species of HIV-1 from a blood transfusion donor and recipients. *Science* 270:988-991.

Debouck, C, J.G. Gorniak, J.E. Strickler, T.D. Meek, B.W. Metcalf, and M. Rosenberg. 1988. Human Immunodeficiency Virus protease expressed in E. coli exhibits autoprocessing and specific maturation of the gag precursor. *PNAS* 85:2449-2453.

DeFranco, D, and KR Yamamato. 1986. Two different factors act separately or together to specify functionally distinct activities at a single transcriptional enhancer. *Mol Cell Biol* 6:993.

Demarchi, F., P. D'Agaro, A. Falasci, and M. Giacca. 1992. Probing protein-DNA interactions at the long terminal repeat of human immunodeficiency virus type 1 by in vivo footprinting. *Journal of Virology* 66:2514-2518.

Demarchi, F., P. D'Agaro, A. Falaschi, and M. Giacca. 1993. In vivo footprinting analysis of constitutive and inducible protein-DNA interactions at the long terminal repeat of human immunodeficiency virus type 1. *J Virol* 67:7450-7460.

Des Jarlais, D. C., S. R. Friedman, D. M. Novick, J. L. Sotheran, P. Thomas, S. R. Yancovitz, D. Mildvan, J. Weber, M. J. Kreek, and R. Maslansky. 1989. HIV-1 infection among intravenous drug users in Manhattan, New York City, from 1977 through 1987. *JAMA* 261:1008-12.

Des Jarlais, D. C., S. R. Friedman, K. Choopanya, S. Vanichseni, and T. P. Ward. 1992. International epidemiology of HIV and AIDS among injecting drug users. *AIDS* 6:1053-68.

Desai-Yajnik, V., and H. H. Samuels. 1993. Regulation of the human immunodeficiency virus type 1 long terminal repeat: interactions of thyroid hormone receptor with retinoid-X receptor, nuclear factor kappa B, Sp1, and Tat. *Trans Assoc Am Physicians* 106:13-32.

Desai-Yajik, V., and H. H. Samuels. 1993. The NF-kB and Sp1 motifs of the human immunodeficiency virus type 1 long terminal repeat function as novel thyroid hormone response elements. *Mol Cell Biol* 13:5057-5069.

Desai-Yajnik, V., E. Hadzic, P. Modlinger, S. Malhotra, G. Gechlik, and H. H. Samuels. 1995. Interactions of thyroid hormone receptor with the human immunodeficiency virus type 1 (HIV-1) long terminal repeat and the HIV-1 Tat transactivator. *J Virol* 69:5103-12.

di Fagagna, F. D., G. Marzio, M. I. Gutierrez, L. Y. Kang, A. Falaschi, and M. Giacca. 1995. Molecular and functional interactions of transcription factor USF with the long terminal repeat of HIV-1. *J Virol* 69:2765-2775.

Dignam, J.D., R.M. Lebovitz, and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* 11:1475-1489.

Dillehay, DL, AS Walia, and EW Lamon. 1988. Effects of retinoids on macrophage function and IL-1 activity. *J. Leukocyte Biol.* 44:353.

Dingwall, C., I. Ernberg, M. J. Gait, S. M. Green, S. Heaphy, J. Karn, A. D. Lowe, M. A. Singh, and R. Valerio. 1989. Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *PNAS* 86:6925-6929.

Dinter, H., R. Chiu, M. Imagawa, M. Karin, and K.A. Jones. 1987. In vitro activation of the HIV-1 enhancer in extracts from cells treated with a phorbol ester tumour promoter. *EMBO J* 6:4067-4071.

Dodd, R. Y. 1992. The risk of transfusion-transmitted infection. *N Engl J Med* 327:419-421.

Donnelly, C., W. Leisenring, P. Kanki, T. Awerbuch, and S. Sandberg. 1993. Comparison of transmission rates of HIV-1 and HIV-2 in a cohort of prostitutes in Senegal. *Bull Math Biol* 55:731-43.

Dragic, T., V. Litwin, G.P. Allaway, S.R. Martin, Y. Huang, K.A. Nagashima, C. Cayanan, P.J. Maddon, R.A. Koup, J.P. Moore, and W.A. Paxton. 1996. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* 381:667-673.

Drysdale, C. M., and G. N. Pavlakis. 1991. Rapid activation and subsequent down-regulation of the human immunodeficiency virus type 1 promoter in the presence of Tat: possible mechanisms contributing to latency. *J Virol* 65:3044-3051.

Du, H., A. L. Roy, and R. G. Roeder. 1993. Human transcription factor USF stimulates transcription through the iitiator elements of the HIV-1 and the Ad-ML promoters. *EMBO J.* 12:501-511.

Duesberg, P.H. 1989. Does HIV cause AIDS? *J Acq Immunodef Synd.* 2:514-517.

Duesberg, P.H. 1989. Human immunodeficiency virus and acquired immunodeficiency syndrome: correlation but not causation. *PNAS* 86:755-764.

Dynan, W.S. 1989. Modularity in promoters and enhancers. *Cell* 58:1-4.

Ehresmann, B., R. Marquet, F. Baudin, J.L. Darlix, N. Tounekti, J. Paoletti, and C. Ehresmann. 1991. The dimerisation of HIV-1 RNA. Mechanism, effect on the structure of the 5' untranslated terminal region and comparison with another retrovirus MoMuLV. RNA Tumor Viruses, Cold Spring Harbour, N.Y.

el Kharroubi, A., and E. Verdin. 1994. Protein-DNA interactions within DNase I-hypersensitive sites located downstream of the HIV-1 promoter.
J Biol Chem 269:19916-24.

Ellerman, V, and O Bang. 1908. *Fizentrabl. Bakteriol.* (46):595.

Embretson, J., M. Zupancic, J.L. Ribas, A. Burke, P. Racz, K. Tenner-Racz, and A.T. Haase. 1993. Massive covert infection of helper T lymphocytes and macrophages by HIV during the incubation period of AIDS. *Nature* 362:359-362.

Emiliani, S., W. Fischle, M. Ott, C. van Lint, CA. Amella, and E. Verdin. 1998. Mutations in the tat gene are responsible for human immunodeiciency type 1 post-integration latency in the U1 cell line. *J Virol* 72:1666-1670.

Enmark, E., and J-A. Gustafsson. 1996. Orphan nuclear receptors - the first eight years. *Molecular Endocrinology* 10:1293-1307.

Epstein, LG., LR. Sharer, VV. Joshi, M. Fojas, MR. Koenigsberger, and JM. Oleske. 1984. Progressive encephalopathy in children with acquired immune deficiency syndrome: Clinical and neuropathological findings.
Ann Neurol 16:414-228.

Epstein, LG., LR. Sharer, VV. Joshi, M. Fojas, MR. Koenigsberger, and JM. Oleske. 1985. Progressive encephalopathy in children with acquired immune deficiency syndrome. *Ann Neurol* 17:488-496.

Epstein, LG., C. Kuiken, BM. Blumberg, S. Hartman, LR. Sharer, and et al. 1991. HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. *Virology* 180:583-590.

Essex, M., M.F. Mclane, T.H. Lee, L. Falk, C.W.S. Howe, J.I. Mullins, C. Cabradilla, and D.P. Francis. 1983. Antibodies to cell membrane antigens associated with human T-cell leukaemia virus in patients with AIDS. *Science* 220:859-862.

European Study Group on Heterosexual Transmission of HIV. 1992. Comparison of female-to-male and male-to-female transmission of HIV in 563 stable couples. *British Medical Journal* 304:809-813.

Evans, R M. 1988. The steroid and thyroid hormone receptor superfamily. *Science* 240:889-895.

Fan, C. M., and T. Maniatis. 1990. Generation of p50 subunit of NF-kappa B by processing of p105 through an ATP-dependent pathway. *Genes Dev* 4:29-42.

Fauci, A.S. 1993. Multifactorial nature of human immunodeficiency virus diseases: implications for therapy. *Science* 262:1011-1018.

Fauci, A.S. 1996. Host factors and the pathogenesis of HIV-induced disease. *Nature* 384:529-534.

Feinberg, A.P., and B. Vogelstein. 1984. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity.
Analytical Biochem. 132:6-13 (addendum in *Analytical Biochem.* 137:266-267).

Feinberg, M. B., R.F. Jarrett, A. Aldovini, R.C. Gallo, and F. Wong-Staal. 1986. HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. *Cell* 46:806-817.

Feinberg, M.B. 1996. Changing the natural history of HIV disease.
Lancet 348:239-246.

Felsenfeld, G. 1992. Chromtin as an essential part of the transcriptional mechanism. *Nature* 355:219-224.

Felzien, L.K., C. Woffendin, M.O. Hottiger, R.A. Subbramanian, E.A. Cohen, and G.J. Nabel. 1998. HIV transcriptional activation by the accessory protein vpr is mediated by the p300 co-activator. *PNAS* 95:5281-5286.

Feng, C-P., M. Kulka, and L. Aurelian. 1993. The GGTCA palindrome and cognate cellular factors in trans-regulation of human immunodeficiency virus long terminal repeat by herpes simplex virus. *J Gen Virol.* 74:715-723.

Feng, S., and E. C. Holland. 1988. HIV-1 tat transactivation requires the loop sequences within the TAR. *Nature* 334:165-167.

Feng, Y., C.C. Broder, P.E. Kennedy, and E.A. Berger. 1996. HIV-1 entry co-factor: functional cDNA cloning of a seven transmembrane, G-protein coupled receptor. *Science* 272:872-877.

Finnegan, A., K. A. Roebuck, B. E. Nakai, D. S. Gu, M. F. Rabbi, S. Song, and A. L. Landay. 1996. IL-10 cooperates with TNF-alpha to activate HIV-1 from latently

and acutely infected cells of monocyte/macrophage lineage. *J Immunol* 156:841-51.

Fischer, U., J. Huber, W.C. Boelens, I.W. Mattaj, and R. Luhrmann. 1995. The HIV-1 rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* 82:475-483.

Fisher, A. G., E. Collalti, L. Ratner, R. C. Gallo, and F. Wong Staal. 1985. A molecular clone of HTLV-III with biological activity.
Nature 316:262-265.

Fisher, A.G., L. Ensoli, M. Ivanoff, S. Chamberlain, L. Petteway, R. Gallo, and F. Wong-Staal. 1987. The *sor* gene of HIV-1 is required for efficient virus transmission *in vitro. Science* 237:888-893.

Fisher, A. G., M. B. Feinberg, S. F. Josephs, M. E. Harper, L. M. Marselle, G. Reyes, M. A. Gonda, A. Aldovini, C. Debouk, R. C. Gallo, and F. Wong-Staal. 1986. The trans-activator gene of HTLV III is essential for virus replication.
Nature 320:367-371.

Fletcher IIIrd, T.M. , B. Brichacek, N. Sharova, M.A. Newman, G. Stivahtis, P.M. Sharp, M. Emerman, HahnB.H., and M. Stevenson. 1996. Nuclear import and cell cycle arrest functions of the HIV-1 vpr protein are encoded by two separate genes in HIV-2/SIVsm. *EMBO J* 15:6155-6165.

Flug, F. , RP. Copp, J. Casanova, ZD. Horowitz, L. Janocko, M. Plotnick, and HH. Samuels. 1987. *cis*-acting elements of the rat growth hormone gene which mediate basal and regulated expression by thyroid hormone.
J Biol Chem 262:6373-6382.

Folgueira, L., J. A. McElhinny, G. D. Bren, W. S. MacMorran, M. T. Diaz Meco, J. Moscat, and C. V. Paya. 1996. Protein kinase C-zeta mediates NF-kappa B activation in human immunodeficiency virus-infected monocytes.
J Virol 70:223-31.

Folks, T. M., J. Justement, A. Kinter, J. Schnittman, J. Orenstein, G. Poli, and A. S. Fauci. 1988. Characterisation of a promonocytic clone chronically infected with HIV and inducible by 13-phorbol-12-myristate acetate.
J Immunol 140:1117-1122.

Franza, B. R., F. J. Raucher, S. F. Josephs, and T. Curran. 1988. The fos complex and fos-related antigens recognise sequence elements that contain AP-1 binding sites. *Science* 239:1150-1153.

Freeman, H. J., T. K. Shnitka, J. R. Piercey, and W. M. Weinstein. 1977. Cytomegalovirus infection of the gastrointestinal tract in a patient with late onset immunodeficiency syndrome. *Gastroenterology* 73:1397-1403.

Froland, S. S., P. Jenum, C. F. Lindboe, K. W. Wefring, P. J. Linnestad, and T. Bohmer. 1988. HIV-1 infection in Norwegian family before 1970 [letter]. *Lancet* 1:1344-5.

Fujita, T., G. P. Nolan, S. Ghosh, and D. Baltimore. 1992. Independent modes of transcriptional activation by the p50 and p65 sub-units of NF-kappa B. *Genes Dev* 6:775-787.

Gallo, R.C., P.S. Sarin, E.P. Gelmann, M. Robert-Guroff, E. Richardson, V.S. Kalyanaraman, D. Mann, G.D. Sidhu, R.E. Stahl, S. Zolla-Pazner, J. Liebowitch, and M. Popovic. 1983. Isolation of Human T-cell leukaemia virus in acquired immune deficiency syndrome. *Science* 220:865-867.

Gallo, R. C., S.Z. Salahuddin, M. Popovic, G.M. Shearer, M. Kaplan, B.F. Haynes, T.J. Palker, R. Redfield, J. Oleske, B. Safai, G. White, P. Foster, and P.D. Markham. 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* 224:500-504.

Garcia, J. A., R. Wu, R. Mitsuyasu, and R. B. Gaynor. 1987. Interactions of cellular proteins involved in the transcriptional regulation of the human immunodeficiency virus. *EMBO J.* 6:3761-3770.

Garcia, J.V., J. Alfano, and A.D. Miller. 1993. The negative effect of human immunodeficiency virus type 1 Nef on cell surface CD4 expression is not species specific and requires the cytoplasmic domain of CD4. *J Virol* 67:1511-1516.

Garcia-Martinez, L. F., G. Mavankal, P. Peters, F. Wu Baer, and R. B. Gaynor. 1995. Tat functions to stimulate the elongation properties of transcription complexes paused by the duplicated TAR RNA element of human immunodeficiency virus 2. *J Mol Biol* 254:350-363.

Gaynor, R. 1992. Cellular transcription factors involved in the regulation of HIV-1 gene expression. *AIDS* 6:347-363.

Geisen, C., C. Denk, B. Gremm, C. Baust, A. Karger, W. Bollay, E. Schwarz. 1997. High level expression of the retinoic acid receptor β gene in normal cells of the uterine cervix is regulated by the retinoic acid receptor α and is abnormally down-regulated in cervical carcinoma cells. *Canc Res* 57:1460-67.

Gelderblom, H.R., E.H.S. Hausmann, M Ozel, G Pauli, and M.A. Kock. 1987. Fine structure of human immunodeficiency virus (HIV) and immunolocalisation of structural proteins. *Virology* 156:171-176.

Gelmann, E.P., M. Popovic, D. Blayney, H. Masur, G. Sidhu, R.E. Stahl, and R.C. Gallo. 1983. Proviral DNA of a retrovirus, human T-cell leukaemia virus, in two patients with AIDS. *Science* 220:862-865.

Gendelman, HE., JM. Orenstein, MA. Martin, C. Ferrua, R. Mitra, T. Phipps, LA. Wahl, HC. Lane, AS. Fauci, DS Burke, D. Skillman, and MS. Meltzer. 1988. Efficient isolation and propogation of human immunodeficiency virus on recombinant factor 1-treated monocytes. *J Exp Med* 167:1428-33.

Gessain, A., F. Barin, J.C. Vernant, O. Gout, L. Maurs, A. Calender, and G. de The. 1985. Antibodies to the human T-lymphotropic virus type-1 in patients with tropical spastic paraparesis. *Lancet* 2:407-410.

Gheysen, D., E. Jacobs, F. de Foresta, C. Thiriart, M. Francotte, D. Thines, and H. De Wilde. 1989. Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus infected insect cells. *Cell* 59:103-112.

Ghosh, S., L. R. Gifford, L. R. Riviere, P. Tempest, G. P. Nolan, and D. Baltimore. 1990. Cloning of the p50 DNA binding subunit of NF-κB - homology to Rel and Dorsal. *Cell* 62:1019-1029.

Ghosh, G., G. van Duyna, S. Ghosh, and P.B. Sigler, 1995
Structure of NF-kappaB p50 homodimer bound to a kappa B site.
Nature 373:303-310.

Ghys, P. D., K. Fransen, M. O. Diallo, V. Ettiegne Traore, I. M. Coulibaly, K. M. Yeboue, M. L. Kalish, C. Maurice, J. P. Whitaker, A. E. Greenberg, and M. Laga. 1997. The associations between cervicovaginal HIV shedding, sexually transmitted diseases and immunosuppression in female sex workers in Abidjan, Cote d'Ivoire. *AIDS* 11:F85-93.

Giacca, M., M. I. Gutierrez, S. Menzo, F. D. Di Fagagna, and A. Falaschi. 1992. A human binding site for transcription factor USF/MLTF mimics the negative regulatory element of human immunodeficiency virus type 1.
Virology 186:133-147.

Giguere, V., N. Yang, P. Segui, and R.M. Evans. 1988. Identification of a new class of steroid hormone receptor. *Nature* 331:91-94.

Glass, JD., and RT. Johnson. 1996. Human Immunodeficiency Virus and the Brain. *Annual Review of Neuroscience* 19:1-26.

Goldman, R. 1984. Effect of retinoic acid on the proliferentiation and phagocytic capability of murine macrophage-like cell lines. *J Leukocyte Biol.* 120:91.

Golub, E. I., G. R. Li, and D. J. Volsky. 1990. Differences in the basal activation of the long terminal repeat determine different replicative capacities of two closely related human immunodeficiency virus type 1 isolates. *J Virol* 64:3654-3660.

Gonzalez-Scarano, F., M.N. Waxman, A.M. Ross, and J.A. Hoxie. 1987. Sequence similarities between the human immunodeficiency virus gp41 and paramyxovirus fusion proteins. *AIDS Res and Human Retroviruses* 3:245-252.

Goodman, R. 1990. Control of lipoprotein lipase secretion by macrophages: effect of macrophage differentiating agents. *J Leukocyte Biol* 47:79.

Gorelick, R.J., S.M. Jr. Nigida, J.W. Jr. Bess, L.O. Arthur, L.E. Henderson, and A. Rein. 1990. Non-infectious human imunodeficiency virus type-1 mutants deficient in genomic RNA. *J Virol* 64:3207-3211.

Gottlieb, M. S., R. Schroff, H. M. Schanker, J. D. Weisman, P. T. Fan, R. A. Wolf, and A. Saxon. 1981. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med* 305:1425-31.

Gottlinger, H.G., T. Dorfman, E. Cohen, and W.A. Haseltine. 1993. Vpu protein of human immunodeficiency virus type 1 enhances the release of capsids produced by *gag* gene constructs of widely divergent retroviruses. *PNAS* 90:7381-7385.

Gottlinger, H.G., J.G. Sodroski, and W.A. Haseltine. 1989. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of Human Immunodeficiency Virus type 1. *PNAS* 86:5781-5785.

Gragerov, A., T. Kino, G. Ilyina-Gragerova, G.P. Chrousos, and G.N. Pavlakis. 1998. HHR23A, the human homologue of the yeast DNA-repair protein RAD23, interacts specifically with vpr protein and prevents cell cycle arrest but not the transcriptional effects of vpr. *Virology* 245:323-330.

Gray, G., J.A. McIntyre, and S.F. Lyons. 1996. The effect of breast feeding on verticle transmission of HIV-1 in Soweto, South Africa. Paper read at XI International Conference on AIDS., Vancouver.

Greene, W. C. 1990. Regulation of HIV-1 gene expression. *Annu Rev Immunol* 8:453-475.

Green, S. 1993. Promiscuous liasons. *Nature* 361:590-591.

Griffin, G. E., L. Kwanyee, T. M. Folks, S. Kunkel, and G. J. Nabel. 1989. Activation of HIV gene expression during moncyte differentiation by induction of NF-κB. *Nature* 339:70-73.

Gronemeyer H and D Moras. 1995. How to finger DNA. *Nature* 375:190-191

Gross, L. 1951. "Spontaneous" leukaemia developing in C3H mice following inoculation, in infancy, with AK-leukaemic extracts or AK-embryos. *Proc. Soc. Exp. Biol. Med.* 76:27-32.

Grosschedl, R., and M.L. Birnsteil. 1980. Spacer DNA sequences up-stream of the TATAAATA sequence are essential for promotion of histone H2A transcription *in vivo. PNAS* 77:7102-7106.

Gualberto, A., and A. S. Baldwin. 1995. p53 and Sp1 interact and c0-operate in the tumor necrosis factor-induced transcriptional activation of the HIV-1 long terminal repeat. *J Biol Chem* 270:19680-19683.

Guimaraes, M.D.C., D. Vlahov, and E.A. Castilho. 1997. Post-coital vaginal bleeding as a risk factor for transmission of human immunodeficiency virus in a heterosexual partner study in Brazil. *Arch of Int Medicine* 157:1362-1368.

Guo, B., P.R. Odgren, A.J. van Wijnen, T.J. Last, J. Nickerson, S. Penman, J.B. Lian, J.L. Stein, and G.S. Stein. 1995. The nuclear matrix protein NMP-1 is the transcription factor YY1. *PNAS* 92:10526-10530.

Gurtler, L. G., P. H. Hauser, J. Eberle, A. von Brunn, S. Knapp, L. Zekeng, J. M. Tsague, and L. Kaptue. 1994. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *J Virol* 68:1581-5.

Hagen, G., S. Muller, M. Beato, and G. Suske. 1994. Sp1-mediated transcriptional activation is repressed by Sp3. *EMBO J* 13:3843-3851.

Hahn, B. H., G.M. Shaw, M.E. Taylor, R.R. Redfield, P.D. Markham, S.Z. Salahuddin, F. Wong-Staal, R.C. Gallo, E.S. Parks, and W.P. Parks. 1986. Genetic variation in HTLV III/LAV over time in patients with AIDS or at risk for AIDS. *Science* 232:1548-1553.

Hammes, S.R., E.P. Dixon, M.H. Malim, B.R. Cullen, and W.C. Greene. 1989. Nef protein of human immunodeficiency virus type 1: evidence against its role as a transcriptional inhibitor. *PNAS* 86:8549-8553.

Hampsey, M, and D Reinberg. 1997. Transcription: Why are TAFs essential? *Current Biology* 7:R44-R46.

Hamy, F., U. Asseline, J. Grasby, I. Shigenori, C. Pritchard, G. Slim, P. J. G. Butler, J. Karn, and M. J. Gait. 1993. Hydrogen-bonding contacts in the major grove are required for human immunodeficiency virus type-1 tat protein recognition of TAR RNA. *J. Mol. Biol.* 230:11-123.

Hanenberg, R. S., W. Rojanapithayakorn, P. Kunasol, and D. C. Sokal. 1994. Impact of Thailand's HIV-control programme as indicated by the decline of sexually transmitted diseases. *Lancet* 344:243-5.

Harada, S., Y. Koyanagi, H. Nakashima, N. Kobayashi, and N. Yamamoto. 1986. Tumor promoter, TPA, enhances replication of HTLV III/LAV. *Virology* 154:249-258.

Harouse, JM., S. Bhat, SL. Spitalnik, and et al. 1991. Inhibition of entry of HIV-1 in neural cell lines by antibodies against galactosyl ceramide. *Science* 253:320-324.

Harper, M.E., L.M. Marselle, R.C. Gallo, and F. Wong-Staal. 1986. Detection of lymphocytes expressing human T-lymphotrophic virus type III in lymph nodes and peripheral blood from infected individuals by in situ hybridisation. *PNAS* 83:772-776.

Harrich, D., J. Garcia, R. Mitsuyasu, J. Gonzalez, and R. Gaynor. 1989. Role of Sp1-binding domains in *in vitro* transcriptional regulation of the human immunodeficiency virus type 1 long terminal repeat. *J Virol* 63:2585-2591.

Harris, M., and K. Coates. 1993. Identification of cellular proteins that bind to the human immunodeficiency virus type 1 nef gene product in vitro: a role for myristylation. *J Gen Virol* 74:1581-1589.

Hart, C. E., C. Y. Ou, J. C. Galphin, J. Moore, L. T. Bacheler, J. J. Wasmuth, S. R. J. Petteway, G. Schochetman. 1989. Human chromosome 12 is required for elevated HIV-1 expression in human-hamster hybrid cells. *Science* 246:488-491.

Hayes, J.J., and A. P. Wolffe. 1992. The interaction of transcription factors with nucleosomal DNA. *Bioassays* 14:597-603.

Haynes, B.F., G. Pantaleo, and A.S. Fauci. 1996. Towards an understanding of the correlates of protective immunity to HIV infection. *Science* 271:324-328.

Heath, S.L., J.G. Tew, A.K. Szakal, and G.F. Burton. 1995. Follicular dendritic cells and human immunodeficiency virus. *Nature* 377:740-744.

Heinzinger, N.K., M.I. Bukrinsky, S.A. Haggerty, A.M. Ragland, V. Kewalramani, M-A. Lee, H. Gendelman, L. Ratner, M. Stevenson, and M. Emerman. 1994. The vpr protein of human immunodeficiency virus type 1 influences nuclear localisation of viral nucleic acids in non-dividing cells. *PNAS* 91:7311-7315.

Helseth, E., U. Olshevsky, C. Furman, and J. Sodroski. 1991. Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *J Virol* 65:2119-2123.

Henrard, D.R., J.F. Phillips, L.R. Muenz, W.A. Blattner, D. Wiesner, M.E. Eyster, and J.J. Goedert. 1995. Natural history of HIV-1 cell-free viraemia. *JAMA* 274:554-558.

Ho, DD., TR. Rota, Schooley RT., and et al. 1985. Isolation of HTLV-III from CSF and neural tissues of patients with AIDS related neurological syndromes. *N Engl J Med* 313:1493-1497.

Ho, D.D., I.T. Moudgi, and M. Alain. 1989. Quantitation of human immunodeficiency virus type 1 in the blood of infected persons. *N Engl J Med* 321:1621-1625.

Ho, D.D., A.U. Neumann, A.S. Perelson, W. Chen, J.M. Leonard, and M. Markowitz. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373:123-126.

Höglund, S., Å. Öhagen, K. Lawrence, and D. Gabuzda. 1994. Role of *vif* during packing of the core of HIV-1. *Virology* 201:349-355.

Holland, S.M., M. Chavez, S. Gersberger, and S. Venkatesan. 1992. A specific sequence with a bulged guanosine residue(s) in a stem-bulge-stem structure of Rev-responsive element RNA is required for transactivation by human immunodeficiency virus type 1 Rev. *J Virol* 66:3699-3706.

Holzmeister, J., B. Ludewig, G. Pauli, and D. Simon. 1993. Sequence specific binding of the transcription factor c-Ets 1 to the human immunodeficiency virus type 1 long terminal repeat. *BBRC* 197:1229-1233.

Horlein, A.J., A.M. Naar, T. Heinzel, J. Torchia, B. Gloss, R. Kurokawa, A. Ryan, Y. Kamei, M. Soderstrom, C.K. Glass, and M.G. Rosenfeld. 1995. Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* 377:397-403.

Howcroft, T., K. Strebel, M. Martin, and D. Singer. 1993. Repression of MHC class I gene promoter activity by two-exon Tat of HIV. *Science* 260:1320-1322.

Huang, L., and K. T. Jeang. 1993. Increased spacing between Sp1 and TATAA renders human immunodeficiency virus type 1 replication defective: implications for Tat function. *J Virol* 67:6937-6944.

Huet, T., R. Cheynier, A. Meyerhans, G. Roelanti, and S. Wain-Hobson. 1990. Genetic organisation of a chimpanzee lentivirus related to HIV-1. *Nature* 345:356-359.

Huminer, D., J. B. Rosenfeld, and S. D. Pitlik. 1987. AIDS in the pre-AIDS era. *Rev Infect Dis* 9:1102-1108.

Hwang, S.R., T.J. Boyle, K. Lyerly, and B.R. Culler. 1991. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* 253:71-74.

Hwung, Y.P., D.T. Crowe, L.H. Wang, S.Y. Tsai, and M-J. Tsai. 1988. The COUP transcription factor binds to an upstream promoter element of rat insulin II gene. *Mol Cell Biol* 8:2072-2077.

Imberti, L., A. Sottini, A. Bettinardi, M. Puoti, and D. Primi. 1991. Selective depletion in HIV infection of T-cells that bear specific T-cell receptor Vβ sequences. *Science* 254:860-863.

Israel, A. 1997. IkB kinase all zipped up. *Nature* 388:519-521.

Ivey-Hoyle, M., R.K. Clarke, and M. Rosenberg. 1991. The N-terminal 31 amino acids of human immunodeficiency virus type 1envelope protein gp120 contains a potential gp41 contact site. *J Virol* 65:2682-2685.

Jacks, T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, and H. E. Varmus. 1988. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* 331:280-283.

Janknecht, R., and T. Hunter. 1996a. Transcription: A growing coactivator network. *Nature* 383:22-23.

Janknecht, R., and T. Hunter. 1996b. Versatile molecular glue: Transcriptional control. *Curr Biol* 6:951-954.

Jansa, P. and J. Forejt, 1996. A novel type of retinoic acid response element in the second intron of the mouse H2Kb gene is activated by the RAR/RXR heterodimer. *Nucleic Acids Res.* 24:694-701

Janssens, W., K. Fransen, M. Peeters, L. Heyndrickx, J. Motte, L. Bedjabaga, E. Delaporte, P. Piot, and G. van der Groen. 1994. Phylogenetic analysis of a new chimpanzee lentivirus SIVcpz-gab2 from a wild-captured chimpanzee from Gabon. *AIDS Res Hum Retroviruses* 10:1191-1192.

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* 67:6224-6233.

Jones, K. A., J. T. Kadonaga, P. A. Luciw, and R. Tijian. 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor Sp1. *Science* 232:755-759.

Jones, K. A., P. A. Luciw, and N. Duchange. 1988. Structural arrangements of transcriptional control domains within the 5'-untranslated leader regions of the HIV-1 and HIV-2 promoters. *Genes Dev.* 2:1101-1114.

Kadonaga, J. T. 1998. Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* 92:307-313.

Kaiser, K., and M. Meisterernst. 1996. The human general co-factors. *TIBS* 21:342-345.

Kaiser, K., G. Stelzer, and M. Meisterernst. 1995. The coactivator p15 (PC4) initiates transcriptional activation during TFIIA-TFIID-promoter complex formation. *EMBO J* 14:3520-3527.

Kamine, J., and G. Chinnadurai. 1992. Synergistic activation of the human immunodeficiency virus type 1 promoter by the viral Tat protein and cellular transcription factor Sp1. *J Virol* 66:3932-3936.

Kanki, P. J. 1991. Biologic features of HIV-2. An update. *AIDS Clin Rev*.1045-77.

Kanki, P. J., K. U. Travers, S. Mboup, C. C. Hsieh, R. G. Marlink, NDiaye A. Gueye, T. Siby, I. Thior, M. Hernandez Avila, J. L. Sankale, and et al. 1994. Slower heterosexual spread of HIV-2 than HIV-1. *Lancet* 343:943-6.

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* 330:489-493.

Kato, H., M. Horikoshi, and R. G. Roeder. 1991. Repression of HIV-1 transcription by a cellular protein. *Science* 251:1476-1479.

Kaufmann, J., and S.T. Smale. 1994. Direct recognition of initiator elements by a component of the transcription factor IID complex.
Genes and Development 8:821-829.

Keen, N.J., M.J. Churcher, and J. Karn. 1997. Transfer of Tat and release of the human immunodeficiency virus type 1 transcription elongation complex.
EMBO 16:5260-5272.

Kerkau, T., R. Schmitt-Landgraf, A. Schimpl, and E. Wecker. 1989. Downregulation of HLA class I antigens in HIV-1 infected cells.
AIDS Res. Hum. Retroviruses 5:613-620.

Kerkau, T. 1997. The human immunodeficiency viru type1 Vpu protein interfers with an earlier step in the biosynthesis of major histocompatibility complex (MHC) class I molecules. *J Exp Med*185:1295-1305.

Kestler, H.W., D.J. Ringler, K. Mori, D.L. Panicali, P.K. Sehgal, D. Daniel, and R.C. Desrosiers. 1991. Imortance of the nef gene for maintenance of high virus loads and for development of AIDS. *Cell* 65:651-662.

Kieran, M., V. Blank, F. Logeat, J. Vandekerckhove, F. Lottspeich, O. Lebail, M. B. Urban, P. Kourilsky, P. A. Baeuerle, and A. Israel. 1990. The DNA binding subunit of NF-κB is identical to factor κBF1and homologous to the Rel oncogene product. *Cell* 62:1007-1018.

Kim, C. H., S. Gollapudi, A. Kim, T. Lee, and S. Gupta. 1996. Role of protein kinase C-beta isozyme in the activation of latent human immunodeficiency virus type 1 in promonocytic U1 cells by phorbol-12-myristate acetate. *AIDS Res. Hum. Retroviruses* 12:1361-1366.

Kim, S., K. Ikeuchi, R. Byrn, J. Groopman, and D. Baltimore. 1989a. Lack of negative influence on viral growth by the nef gene of human immunodeficiency virus type 1. *PNAS* 86:9544-9548.

Kim, S., R. Byrn, J.E. Groopman, and D. Baltimore. 1989b. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *J Virol* 63:3708-3713.

Kingsman, S. M., and A. J. Kingsman. 1996. The regulation of human immunodeficiency virus type-1 gene expression. *Eur. J. Biochem.* 240:491-507.

Kirchoff, F., T.C. Greenough, D.B. Brettler, J.L. Sullivan and R.C. Desrosiers, 1995 Absense of intact nef sequences in a long term survivors with non-progression HIV-1 infection. *N Eng J Med* 332:228-232.

Kitano, K, GC Baldwin, MA Raines, and DW Golde. 1990. Differentiating agents facilitate infection of myeloid leukaemia cell lines by monocytotropic HIV-1 strains. *Blood* 76:1980-1988.

Kizaki, M., H.P. Koeffler, C-W Lin, and C.W. Miller. 1990. Expression of retinoic acid receptor mRNA in haematopoietic cells. *Leuk Res* 14:645-655.

Klatzmann, D., F. Barre-Sinoussi, M.T. Nugeyre, C. Dauguet, E. Vilmer, C. Griscelli, F. Bran-Veziret, C. Rouzioux, J.C. Gluckman, and J.C. Chermann. 1984a. Selective tropism of lymphadenopathy associated virus (LAV) for helper inducer T lymphocytes. *Science* 225:59-63.

Klatzmann, D., E. Champagne, S. Chamaret, J. Gruest, and D. Guetard. 1984b. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* 312:767-768.

Klein-Hitpab, L., M. Schorpp, U. Wagner, and G. Ryffel. 1986. An estrogen-responsive element derived frm the 5' flanking region of the Xenopus vitellogenin A2 gene functions in transfected human cells. *Cell* 46:1053-1061.

Kliewer, S.A., K. Umesono, R.A. Heyman, D.J. Mangelsdorf, J.A. Dyke, and R.M. Evans. 1992. Retinoid X receptor-COUP-TF interactions modulate retinoic acid signalling. *PNAS* 89:1448-1452.

Klimkait, T., K. Strebel, M.D. Hoggan, M.A. Martin, and J.M. Orenstein. 1990. The human immunodeficiency virus type 1 specific protein vpu is required for efficient virus maturation and release. *J Virol* 64:621-629.

Koenig, S., HE. Gendelman, JM Orenstein, and et al. 1986. Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. *Science* 233:1089-1093.

Koup, R.A., J.T. Safrit, Y. Cao, C.A. Andrews, and G. McLeod. 1994. Temporal association of cellular immune responses with the initial control of viraemia in primary human immunodeficiency virus type 1 syndrome. *J Virol* 68:4650-4655.

Kowalski, M.L., J. Potz, L. Basiripour, T. Dorfman, W.C. Goh, E. Terwilliger, A. Dayton, C. Rosen, W. Haseltine, and J. Sodroski. 1987. Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. *Science* 237:1351-1355.

Koyanagi, Y., W.A. O'Brien, J.Q. Zhao, D.W. Golde, J.C. Gasson, and I.S. Chen. 1988. Cytokines alter production of HIV-1 from primary mononuclear phagocytes. *Science* 241:1673-1675.

Kundu, M., A. Srinivasan, R.J. Pomerantz, and A.S. Khalili. 1995. Evidence that a cell cycle regulator E2F1 down-regulates transcriptional activity of the human immunodeficiency type 1 promoter. *J Virol* 69:6940-6946.

Kurokawa, R., V.C. Yu, A. Naar, S. Kyakumoto, Z. Han, S. Silverman, M.G. Rosenfeld, and C.K. Glass. 1993. Differential orientations of the DNA-binding domain and carboxyterminal dimerisation interface regulate binding site selection by nuclear receptor heterodimers. *Genes and Development* 7:1423-35.

Ladias, J. A. A. 1994. Convergence of multiple nuclear receptor signaling pathways onto the long terminal repeat of human immunodeficiency virus-1. *J Biol Chem* 269:5944-5951.

Ladias, J.A.A., and S.K. Karathanasis. 1991. Regulation of the apolipoprotein A1 gene by ARP-1, a novel member of the steroid receptor suoer-family. *Science* 251:561-565.

Laga, M., N. Nzila, and J. Goeman. 1991. The interrelationship of sexually transmitted diseases and HIV infection: implications for the control of both epidemics in Africa. *AIDS* 5 Suppl 1:S55-63.

Laga, M., A. Manoka, M. Kivuvu, B. Malele, M. Tuliza, N. Nzila, J. Goeman, F. Behets, V. Batter, M. Alary, and et al. 1993. Non-ulcerative sexually transmitted diseases as risk factors for HIV-1 transmission in women: results from a cohort study. *AIDS* 7:95-102.

Laga, M., M. Alary, N. Nzila, and et al. 1994. Condom promotion, sexually transmitted diseases treatment and declining incidence of HIV-1 infection in female Zairian sex workers. *Lancet* 344:246-248.

Laspia, M. F., A. P. Rice, and M. B. Mathews. 1989. HIV-1 Tat protein increases transcriptional initiation and stabilises elongation. *Cell* 59:283-292.

Latchman, D. 1995. *Gene Regulation.* 2nd ed: Chapman and Hall.

Lee, C. A. 1996. Transfusion-transmitted disease. *Baill Clin Haematol* 9:369-94.

Lee, F., K. M. Galvin, and Y. Shi. 1993. Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *PNAS* 90:6145-6149.

Lee, M-O., P.D. Hobbs, X-K. Zhang, M.I. Dawson, and M. Pfahl. 1994. A synthetic retinoid antagonist inhibits the human immunodeficiency virus type 1 promoter. *PNAS* 91:5632-5636.

Lefrere, J.J., V. Marotti, D. Hecoq, B. Noel, A.M. Courouce, P. Lambin, C. Salmon, and P. Rouger. 1991. No evidence of frequent human immunodeficiency virus type 1 infection in seronegative at-risk individuals. *Transfusion* 31:205-211.

Lehmann, J.M., L. Jong, A. Fanjul, J.F. Cameron, X-P. Lu, P. Haefner, M.I. Dawson, and M. Pfahl. 1992. Retinoids selective for retinoid X receptor response pathways. *Science* 258:1944-1946.

Leis, J, D Baltimore, J. M. Bishop, J. Coffin, E. Fleissner, S.P. Goff, S. Oroszlan, H. Robinson, A. M. Skalka, and H. M. Temin. 1988. Standardised and simplified nomenclature for proteins common to all retroviruses. *J Virol* 62:1808-1809.

Leng, X., S-Y. Tsai, and M-J. Tsai. 1995. The nuclear hormone eceptor superfamily:structure and function. In *Hormones and Cancer,* edited by W. V. Vedeckis. Boston: Birkhauser Publishing Co.

Leng, X., A.J. Cooney, S.Y. Tsai, and M-J. Tsai. 1996. Molecular mechanisms of COUP-TF-mediated transcriptional repression: evidence for transrepression and active repression. *Mol Cell Biol* 16:2332-2340.

Leonard, J., C. Parrott, A. J. Buckler-White, W. Turner, E. K. Ross, M. A. Martin, and A. B. Rabson. 1989. The NF-kappa B binding sites in the human immunodeficiency virus type 1 long terminal repeat are not required for virus infectivity. *J Virol* 63:4919-4924.

Lever, A., H. Gottinger, W. Haseltine, and J. Sodroski. 1989. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J Virol* 63:4085-4087.

Levy, D.N., L.S. Fernandes, W.V. Williams, and D.B. Weiner. 1993. Induction of cell differentiation by human immunodeficiency virus 1 vpr. *Cell* 72:541-550.

Levy, D.N., Y. Refaeli, and D.B. Weiner. 1995. Extracellular vpr protein increases cellular permissiveness to human immunodeficiency virus replication and reactivates virus from latency. *J Virol* 69:1243-1252.

Levy, J. A., A. D. Hoffman, S. M. Kramer, J. A. Landis, J. M. Shimabukuro, and L. S. Oshiro. 1984. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* 225:840-2.

Levy, JA., J. Shimabukuro, H. Hollander, J. Mills, and L. Kaminsky. 1985. Isolation of AIDS associated retrovirus from cerebrospinal fluid and brain of patients with neurological symptoms. *Lancet* 2:586-588.

Li, C., C.F. Lai, and D.S. Sigman. 1991. Cloning of a cellular interleukin binding factor that binds to NFAT-like motifs in the human immunodeficiency virus long terminal repeat. *PNAS* 88:7739-7743.

Li, Y., G. Mak, and B.R. Franza. 1994. In vitro study of functional involvement of Sp1, NF-kB/Rel and AP1 in phorbol 12-myristate 13-acetate-mediated HIV-1 long terminal repeat activation. *J Biol Chem* 269:30616-30619.

Linney, E. 1992. Retinoic acid receptors: transcription factors modulating gene regulation, development and differentiation. *Curr Top Dev Biol* 27:309-351.

Liu, J., N. D. Perkins, R. M. Schmid, and G. J. Nabel. 1992. Specific NF-κB subunits act in concert with Tat to stimulate human immunodeficiency virus type 1 transcription. *J Virol* 66:3883-3887.

Liu, R., W. A. Paxton, S. Choe, D. Ceradini, S. R. Martin, R. Horuk, M. E. MacDonald, H. Stuhlmann, R. A. Koup, and N. R. Landau. 1996. Homozygous defect in HIV-1 co-receptor accounts for resistance of some multiply exposed individuals to HIV-1 infection. *Cell* 86:367-377.

Lu, X., T. M. Welsh, and M. Peterlin. 1993. The human immunodeficiency virus type 1 long terminal repeat specifies two different transcription complexes, only one of which is regulated by Tat. *J. Virol.* 67:1752-1760.

Lu, X.P., G. Salbert, and M. Pfahl. 1994. An evolutionary conserved COUP-TF binding element in a neural-specific gene and COUP-TF expression patterns support a major role for COUP-TF in neural development. *Mol Endocrinol* 8:1774-1788.

Lu, Y., M. Stenzel, J. G. Sodroski, and W. A. Haseltine. 1989. Effects of long terminal repeat mutations on human immunodeficiency virus type 1 replication. *J. Virol.* 63:4115-4119.

Lu, Y. C., N. Touzjian, M. Stenzel, T. Dorfman, J. G. Sodroski, and W. A. Haseltine. 1990. Identification of cis-acting repressive sequences within the negative regulatory element of human immunodeficiency virus type 1. *J Virol* 64:5226-5229.

Lu, Y-L., P. Spearman, and L. Ratner. 1993. Human immunodeficiency virus type 1 viral protein R localisation in infected cells and virions. *J Virol* 67:6542-6550.

Luciw, P.A., C. Cheng-Mayer, and J.A. Levy. 1987. Mutational analysis of the human immunodeficiency virus: the orf-B region down-regulates virus replication. *PNAS* 84:1934-1938.

Luckow, B., and G. Schutz. 1987. CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. *Nucleic Acids Research* 15:5490.

Maddon, P.J., A.G. Dalgleish, J.S. McDougal, P.R. Clapham, R.A. Weiss, and R. Axel. 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* 47:333-348.

Maekawa, T., T. Sudo, M. Kurimoto, and S. Ishii. 1991. USF-related transcription factor, HIVTF1 stimulates transcription of human immunodeficiency virus 1. *Nucleic Acids Res* 19:4689-4694.

Mahalingam, S., B. MacDonald, K.E. Ugen, V. Ayyavoo, M.G. Agadjanyan, W.V. Williams, and D.B. Weiner. 1997. In vitro and in vivo tumor growth suppression by HIV-1 Vpr. *DNA and Cellular Biology* 16:137-143.

Majello, B., P. De Luca, G. Hagen, G. Suske, and L. Lania. 1994. Different members of the Sp1 multigene family exerts opposite transcriptional regulation of the long terminal repeat of HIV-1.
Nucleic Acids Res. 22:4914-4921.

Malim, M.H. , and B.R. Cullen. 1991. HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE; implications for HIV-1 latency. *Cell* 65:241-248.

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 the nuclear export of unspliced viral mRNA. *Nature* 338:254-257.

Malim, M.H., L.S. Tiley, D.F. McCarn, J.R. Rusche, J. Hauber, and B.R. Cullen. 1990. HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell* 60:675-683.

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 and Development 11:2633-2644.

Mangelsdorf, D.J., K. Umesono, and R.M. Evans. 1994. The retinoid receptors. In *The Retinoids: Biology, Chemistry and Medicine.*, edited by M. B. Sporn, A. B. Roberts and D. S. Goodman. New York: Raven Press.

Mangelsdorf, D. J., C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, and R. Evans. 1995a. The nuclear receptor superfamily: the second decade. *Cell* 83:835-839.

Mangelsdorf, D. J., and R. M. Evans. 1995b. The RXR heterodimers and orphan receptors. *Cell* 83:841-850.

Mansky, L. M. 1996. Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. *AIDS Res and Hum Retroviruses.* 12:307-314.

Marciniak, R. A., M. A. Garcia-Blanco, and P. A. Sharp. 1990. Identification and characterisation of a HeLa nuclear protein that specifically binds to the *trans-*activation TAR) element of human immunodeficiency virus.
PNAS 87:3624-3628.

Margolis, D. M., M. Somasundaran, and M. R. Green. 1994. Human transcription factor YY1 represses human immunodeficiency virus type 1 transcription and virus production. *J. Virology* 68:905-910.

Markovitz, D. M. 1993. Infection with the human immunodeficiency virus type 2. *Ann Intern Med* 118:211-218.

Marlink, R. G., D. Ricard, S. M'Boup, P. J. Kanki, J. L. Romet Lemonne, I. N'Doye, K. Diop, M. A. Simpson, F. Greco, M. J. Chou, and et al. 1988. Clinical, hematologic, and immunologic cross-sectional evaluation of individuals exposed to human immunodeficiency virus type-2 (HIV-2). *AIDS Res Hum Retroviruses* 4:137-48.

Marx, P.A., R.J. Munn, and K.I. Joy. 1988. Emulation of thin section electron microscopy predicts an envelope-associated icosahedral capsid for HIV. *Laboratory Investigation* 58:112-118.

Masur, H., M.A. Michelis, J.B. Greene, I. Onorato, R.A. Vande Stouwe, R.S. Holzman, G. Wormser, L. Brettman, M. Lange, H.W. Murray, and S. Cunningham-Rundles. 1981. An outbreak of community-acquired *Pneumocystis carinii* pneumonia. *N Eng J Med* 305:1431-1438.

May, M. J., and Ghosh S. 1998. Signal transduction through NF-κB. *Immunology Today* 19:80-88.

McClelland, D. B., J. J. McMenamin, H. M. Moores, and J. A. Barbara. 1996. Reducing risks in blood transfusion: process and outcome. *Transfus Med* 6:1-10.

McCune, J.M., L.B. Rabin, M.B. Feinberg, M Lieberman, J.C. Kosek, G.R. Reyes, and I.L. Weissman. 1988. Endoproteolytic cleavage of gp160 is required for the activation of Human Immunodeficiency Virus. *Cell* 53:55-67.

McKnight, S., and R. Tjian. 1986. Transcriptional selectivity of viral genes in mammalian cells. *Cell* 46:795-805.

McKnight, S.L. 1996. Transcription revisited: A commentary on the 1995 Cold Spring Harbor Laboratory meeting, "Mechanisms of eukaryptic transcription". *Genes and Development* 10:367-381.

Mehta, K, and G Lopez-Berenstein. 1986. Expression of tissue transglutaminase during differentiation. *Cancer Res* 46:1388.

Mercurio, F., J. DiDonato, C. Rosette, and M. Karin. 1993. p105 and p98 precursor proteins play an active role in NF-κB-mediated signal transduction. *Genes and Dev* 7:705-718.

Miller, M.D., M.T. Warmerdam, I. Gaston, W.C. Greene, and M.B. Feinberg. 1994. The human immunodeficiency virus-1 nef gene product: a positive factor for viral infection and replication in primary lymphocytes and macrophages. *J Exp Med* 179:101-113.

Miyajima, N., Y. Kadowaki, S.I. Fukushige, S.I. Shimizu, K. Semba, Y. Yamanashi, K.I. Matsubara, K. Toyoshima, and T. Yamamoto. 1988. Identification of two novel members of erbA superfamily by molecular cloning: the gene products of the two are highly related to each other. *Nucleic Acids Res.* 16:11057-11066.

Miyamoto, S., and I. M. Verma. 1995. Rel/NF-kB/IkB story. *Adv. Cancer Res* 66:255-292.

Montminy, M. 1997. Trancriptional activation: Something new to hang your HAT on. *Nature* 387:654-655.

Moore, J.P., Y. Cao, D.D. Ho, and R.A. Koup. 1994. Development of the anti-gp-120 antibody responses during seroconversion to human immunodeficiency virus type 1. *J Virol* 68:5142-5155.

Muchardt, C., J-S. Seeler, A. Nirula, D-L. Shurland, and R. B. Gaynor. 1992. Regulation of human immunodeficiency virus enhancer function by PRDII-BF-1 and c-*rel* gene products. *J Virol* 66:244-250.

Muller, M. M., T. Gerster, and W. Schaffner. 1988. Enhancer sequences and the regulation of gene transcription. *Euro J Biochem* 176:485-495.

Murray, JL, K Mehta, and G Lopez-Berenstein. 1988. Induction of adenosine deaminase and 5' nucleotidase activity in cultured human blood monocytes and moncytic leukaemia (THP-1) cells by differentiating agents. *J. Leukocyte Biol.* 44:205.

Murre, C., P. McCaw, and D. Baltimore. 1989. The new DNA binding and dimerisation motif in immunoglobulin enhancer binding daughterless, Myo D and myc proteins. *Cell* 56:777-783.

Myers, G. 1991. *Database of Human Retroviruses and AIDS.* Los Alamos, New Mexico: Los Alamos National Laboratory.

Myers, G., K. MacInnes, and B. Korber. 1992. The emergence of simian/human immunodeficiency viruses. *AIDS Res Hum Retroviruses* 8:373-386.

Nabel, G., and D. Baltimore. 1987. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature* 326:711-713.

Nagelkerke, N. J., F. A. Plummer, D. Holton, A. O. Anzala, F. Manji, E. N. Ngugi, and S. Moses. 1990. Transition dynamics of HIV disease in a cohort of African prostitutes: a marker model approach. *AIDS* 4:743-747.

Najera, I., A. Holguin, M. E. Quinones-Mateu, M. A. Munoz-Fernandez, R. Najera, C. Lopez-Galindez, and E. Domingo. 1995. Pol gene quasi-species of human immunodeficiency virus: mutations associated with drug resistance in virus from patients undergoing no drug therapy. *J Virol* 69:23-26.

Nakanishi, Y., Y. Masamme, and N. Kabayashi. 1991. A novel cis-acting element that controls transcription of human immunodeficiency virus type 1 DNA depending on cell type. *J Virol* 65:6334-6338.

Narayan, N., JE. Clements, JD. Stranberg, LC. Cork, and DE. Griffin. 1980. Biological characterization of the virus causing leukoencephalitis and arthritis in goats. *J Gen Virol* 50:69-79.

Nathanson, N., G. Georgsson, PA. Palsson, JA. Najjar, R. Lutley, and G. Petursson. 1985. Experimental visna in Icelandic sheep: the prototype lentivirus infection. *Reviews in Infectious Diseases.* 7:75-82.

Newstein, m., E. J. Stanbridge, G. Casey, and P. R. Shank. 1990. Human chromosome 12 encodes a species-specific factor which increases human immunodeficiency virus type 1 tat-mediated transactivation in rodent cells. *J Virol* 64:4565-4567.

Nicholson, JKA., GD. Cross, CS. Calaway, and JS. McDougal. 1986. *In vitro* infection of human monocytes with human T lymphotropic virus type III/lymphadenopathy-associated virus (HTLV-III/LAV). *J Immunol* 137:323-28.

Niederman, T.M., B.J. Thielan, and L. Ratner. 1989. Human immunodeficiency virus type 1 negative factor is a transcriptional silencer. *PNAS* 86:1728-1732.

Niederman, T.M.J., W.R. Hastings, and L. Ratner. 1993. Myristoylation-enhanced binding of the HIV-1 Nef protein to T-cell skeletal matrix. *Virol* 197:420-425.

Nolan, G. P., and D. Baltimore. 1992. The inhibitory ankyrin and activator rel proteins. *Curr. Opin. Genet. Dev.* 2:211-220.

Nuovo, GJ., F. Gallery, P. MacConnell, and A. Braun. 1994. In situ detection of polymease chain reaction amplified HIV-1 nucleic acids and tumour necrosis factor-alpha RNA in the central nervous system. *Am J Pathol* 144:659-666.

O'Brien, W.A., Y. Koyanagi, A. Nanazie, J-Q, Zhao, A. Diagne, K. Idler, J.A. Zack, and I.S.Y. Chen. 1990. HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. *Nature* 348:69-73.

Oberlin, E., A. Amara, F. Bachelerie, C. Bessia, J-L. Virelizier, F. Arenzana-Seisdedos, O. Schwartz, J-M. Heard, I. Clark-Lewis, D.F. Legler, M. Loetscher, M. Baggiolini, and B. Moser. 1996. The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature* 382:833-835.

Ogawa, K., R. Shibata, T. Kiyomasu, I. Higuchi, Y. Kishida, A. Ishimoto, and A. Adachi. 1989. Mutational analysis of the human immunodeficiency virus *vpr* open reading frame. *J Virol* 63:4110-4114.

Ohana, B., P. A. Moore, S. M. Ruben, C. D. Southgate, M. R. Green, and C. A. Rosen. 1993. The type - human immunodeficiency virus Tat binding protein is a transcriptional activator belonging to an additional family of evolutionarily conserved genes. *PNAS* 90:138-142.

Okamoto, T., T. Benter, S. F. Josephs, M. R. Sadaie, and F. Wong-Staal. 1990. Transcriptional activation from the long-terminal repeat of human immunodeficiency virus in vitro. *Virology* 177:606-614.

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.* 66:5594-5597.

Orchard, K. H. , N. Perkins, C. Chapman, J. Harris, V. Emery, G. Goodwin, D. Latchman, and M. Collins. 1990. A Novel T-Cell Protein Which Recognizes a Palindromic Sequence in the Negative Regulatory Element of the Human Immunodeficiency Virus Long Terminal Repeat. *J Virol* 64:3234-3239.

Orchard, K.H., G. Lang, M. Collins, and D. Latchman. 1992. Characterisation of a novel T lymphocyte protein which binds to a site related to steroid/thyroid

hormone receptor response elements in the negative regulatory sequence of the human immunodeficiency virus long terminal repeat. *Nucleic Acids Res.* 20:5429-5434.

Orchard, K. H., G. Lang, J. Harris, M. Collins, and D. Latchman. 1993. A palindromic element in the human immunodeficiency virus long terminal repeat binds retinoic acid and can confer retinoic acid responsiveness on a heterologous promoter. *J Acquired Immun Def Syn* 6:440-445.

Orenstein, JM., C. Fox, and SM. Wahl. 1997. Macrophages as a source of HIV during opportunistic infections. *Science* 276:1857-61.

Ou, S-H. I., L. F. Garcia-Martinez, E. J. Paulssen, and R. B. Gaynor. 1994. Role of flanking E box motifs in human immunodeficiency virus type 1 TATA element function. *J. Virol.* 68:7188-7199.

Ou, S-H. I., F. Wu, D. Harrich, L.F. Garcia-Martinez, and R. B. Gaynor. 1995. Cloning and characterisation of a novel cellular protein TDP-43 that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J Virol* 69:3584-3596.

Palombella, V. J., O. J. Rando, A. L. Goldberg, and T. Maniatis. 1994. The ubiquitin-proteosome pathway is required for processing the NF-κBI precursor protein and the activation of NF-κB. *Cell* 78:773-785.

Pantaleo, G., C. Graziosi, L. Butini, P.A. Pizzo, S.M. Schnittman, D.P. Kotler, and A.S. Fauci. 1991. Lymphoid organs function as major reservoirs for human immunodeficiency virus. *PNAS* 88:9838-9842.

Pantaleo, G., C. Graziosi, J.F. Demarest, L. Butini, M. Montroni, C.H. Fox, J.M. Orenstein, D.P. Kotler, and A.S. Fauci. 1993. HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. *Nature* 362:355-359.

Pantaleo, G., J.F. Demarest, H. Soudeyns, C. Graziosi, F. Denis, and et al. 1994. Major expansion of CD8+ T cells with a predominant V usage during the primary immune response to HIV. *Nature* 370:463-467.

Pantaleo, G., S. Menzo, M. Vaccaraezza, C. Graziosi, O.J. Cohen, J.F. Demarest, D. Montefiori, J.M. Orenstein, C. Fox, L.K. Schrager, J.B. Margolick, S. Buchbinder, J.V. Giorgi, and A.S. Fauci. 1995. Studies in subjects with long-term nonprogressive human immunodeficiency virus infection. *N Eng J Med* 332:209-211.

Pantaleo, G., and A.S. Fauci. 1996. Immunopathogenesis of HIV infection. *Annu Rev Microbiol* 50:825-854.

Pape, J.W., S. S. Jeun, J. L. Ho, A. Hafner, and W. D. Jr. Johnson. 1993. Effect of ioniazid prophylaxis on incidence of active tuberculosis and progression of HIV infection. *Lancet* 342:268-27

Parrott, C., T. Seidner, E. Duh, J. Leonard, T. S. Theodore, A. Buckler-White, M. A. Martin, and A. B. Rabson. 1991. Variable role of the long terminal repeat Sp1-binding sites in human immunodeficiency virus replication in T lymphocytes. *J Virol* 65:1414-1419.

Pastorcic, M., H. Wang, A. Elbrecht, S.Y. Tsai, M-J. Tsai, and B.W. O'Malley. 1986. Control of transcription initiation in vitro requires binding of a transcription factor to the distal promoter of the ovalbumin gene. *Mol Cell Biol* 6:2784-2791.

Paxton, W. A., S. R. Martin, D. Tse, T. R. O'Brien, J. Skurnick, N. L. VanDevanter, N. Padian, J. F. Braun, D. P. Kotler, S. M. Wolinsky, and R. A. Koup. 1996. Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high risk sexual exposure. *Nat Med* 4:412-417.

Peckham, C., and D. Gibb. 1995. Mother-to-child transmission of the human immunodeficiency virus. *N Engl J Med* 333:298-302.
Peeters, M., C. Honore, T. Huet, L. Bedjabaga, S. Ossari, P. Bussi, R. W. Cooper, and E. Delaporte. 1989. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS* 3:625-30.

Perelson, A.S., A.U. Neumann, M. Markowitz, J.M. Leonard, and D.D. Ho. 1995. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span and viral generation time. *Science* 271:1582-1586.

Perkins, N. D., N. L. Edwards, C. S. Duckett, A. B. Agranoff, R. M. Schmid, and G. J. Nabel. 1993. A co-operative interaction between NF-kappa B and Sp1 is required for HIV-1 enhancer activation. *EMBO J* 12:3551-3558.

Perkins, N. D., A. B. Agranoff, C. S. Duckett, and G. J. Nabel. 1994. Transcription factor AP-2 regulates human immunodeficiency virus type 1 gene expression. *J Virol.* 68:6820-6823.

Phillips, R. E., S. Rowland-Jones, D. F. Nixon, F. M. Gotch, J. P. Edwards, A. O. Oganlesi, J. G. Elvin, J. A. Rothbard, C. R. Bangham, and C. R. Rizza. 1991. Human immunodeficiency virus genetic variations that can escape cytotoxic T cell recognition. *Nature* 345:453-459.

Piatak, M. Jr., M. S. Saag, L. C. Yang, S. J. Clark, J. C. Kapper, K. C. Luk, B. H. Hahn, G. M. Shaw, and S. D. Lifson. 1993. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* 259:1749-1754.

Pinter, A., W.J. Hounen, S.A. Tilley, C. Bona, H. Zaghouani, M.K. Gorny, and S. Zolca-Pazner. 1989. Oligomeric structure of gp41, the transmembrane protein of human immunodeficiency virus type 1. *J Virol* 63:2674-2679.

Pleskoff, O., C. Trebouke, A. Brelot, N. Heveker, M. Seman, and M. Alizon. 1997 Identification of a chemokine receptor encoded by human cytomegalovirus as a co-factor for HIV-1 entry. *Science 276:1874-1878*

Poiesz, B.J., F.W. Ruscetti, J.W. Mier, A.M. Woods, and R.C. Gallo. 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *PNAS* 77:7415-7419.

Poli, G., A. Kinter, J. S. Justement, J. H. Kehrl, P. Bressler, S. Stanley, and A. S. Fauci. 1990. Tumour necrosis factor alpha functions in an autocrine manner in the induction of human immunodeficiency virus expression. *PNAS* 87:782-785.

Poli, G, AL Kinter, JS Justement, P Bressler, JH Kehrl, and AS Fauci. 1992. Retinoic acid mimics transforming growth factor in the regulation of human immunodeficiency virus expression in monocytic cells. *PNAS* 89:2689-2693.

Pomerantz, R.J., T. Seshamma, and D. Trono. 1992. Efficient replication of human immunodeficiency virus type 1 requires a threshold level of rev: potential implications for latency. *J Virol* 66:1809-1813.

Power, C., JC. McArthur, RT. Johnson, DE. Griffin, JD. Glass, and et al. 1994. Demented and non-demented patients with AODS differ in brain derived human immunodeficiency virus type 1 envelope sequences. *J Virol* 68:4643-4649.

Prada, C. A., J-B. Yoon, and R. B. Roeder. 1995. A novel LBP-1 mediated restriction o HIV-1 transcription at the level of elongation in vitro. *J. Biol. Chem.* 270:2274-2283.

Price, RW., B. Brew, J. Sidtis, M. Rosenblum, AC. Scheck, and P. Cleary. 1988. The brain in AIDS: central nervous system HIV-1 infection and AIDS dementia complex. *Science* 239:586-592.

Ptashne, M. 1988. How eukaryotic transcription factors work. *Nature* 335:683-689.

Ptashne, M., and A. Gann. 1997. Transcriptional activation by recruitment. *Nature* 386:569-577.

Pugh, B.F., and R. Tjian. 1990. Mechanism of transcriptional activation by Sp1: evidence for co-activators. *Cell* 61:1187-1197.

Qui, Y.H., S-Y. Tsai, and M-J. Tsai. 1994. COUP-TF: an orphan member of the steroid/thyroid hormone receptor superfamily. *Trends Endocrinol Metabol* 5:234-239.

Qui, Y., V. Krishnan, FA. Pereira, SY. Tsai and M-J Tsai. 1996. Chicken ovalbumin upstream promoter-transcription factors and their regulation. *J Steroid Biochem. Mol. Biol.* 56:81-85.

Quinn, T.C., P. Piot, J.B. McCormick, F.M. Feinsold, H. Taelman, B. Kapita, W. Stevens, and A.S. Fauci. 1987. Serological and immunologic studies in patients with AIDS in North America and Africa. The potential role of infectious agents as cofactors in human immunodeficiency virus infections. *JAMA* 257:2617-2621.

Quinn, T. C. 1994. Population migration and the spread of types 1 and 2 human immunodeficiency viruses. *PNAS* 91:2407-2414.

Quinn, TC. 1996. Global burden of the HIV pandemic. *Lancet.* 348:99-106.

Rastinejad F., T. Perlmann, R. Evans and PB Sigler. 1995 Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* 375:203-211

Ratner, L., R. C. Gallo, and F. Wong Staal. 1985. HTLV-III, LAV, ARV are variants of same AIDS virus *Nature* 313:636-7.

Reines, D., J. Conaway, and R Conaway. 1996. The RNA polymerase II general elongation factors. *Trends in Biochemical Science* 21:351-355.

Renkawitz, R. 1990. Transcriptional repression in eukaryotes. *Trends Genet.* 6:192-197.

Rey, M-A., A.G. Laurent, J. McClure, B. Knot, L. Montagnier, and A.G. Hovanessian. 1990. Transmembrane envelope glycoproteins of Human Immunodeficiency Virus type 2 and simian immunodefficiency virusSIVmac exist as homodimers. *J Virol* 64:922-926.

Rice, N. R., M. L. Mackichan, and A. Israel. 1992. The precursor of NF-κB p50 has IkB-like functions. *Cell* 71:243-253.

Richie, H.H., L-H. Wang, S. Tsai, B.W. O'Malley, and M-J. Tsai. 1990. COUP-TF gene: a structure unique for the steroid/thyroid receptor superfamily. *Nucleic Acids Res* 18:6857-6862.

Roberts, S.G., E. Maldonado, D. Reinberg, and M.R. Green. 1993. Interaction between an acidic activator and transcription factor TFIIB is required for transcriptional activation. *Nature* 363:741-744.

Roeder, R.G. 1991. The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly.
Trends in Biochemical Science 16:402-408.

Roeder, R.G. 1996. The role of general initiation factors in transcription by RNA polymeras II. *Trends in Biochemical Science* 21:327-335.

Rohr, O., D. Aunis, and E. Schaeffer. 1997. COUP-TF and Sp1 interact and cooperate in the transcriptional activation of the human immunodeficiency virus type 1 long terminal repeat in human microglial cells.
J Biol Chem 272:31149-31155.

Romerio, F., M.N. Gabriel, and D. M. Margolis. 1997. Repression of human immunodeficiency virus type-1 through the novel cooperation of human factors YY1 and LSF. *J. Virology* 71:9375-9382.

Romieu, I., R. Marlink, P. Kanki, S. M'Boup, and M. Essex. 1990. HIV-2 link to AIDS in West Africa. *J Acq Immune Defic Syndr.* 3:220-230.

Roos, M. T., Lange J.M., R.E. de Goede, Coutinho R.A., P.T. Schellekens, F. Miedema, and M. Termatte. 1992. Viral phenotype and immune response in primary human immunodeficiency virus type I infection.
J Infect Dis 165:427-432.Rosen, C. A. 1991. Tat and rev: positive modulation of human immunodeficiency virus gene expression. *Gene Expression* 1:85-90.

Rosen, C. A., J. G. Sodroski, and W. A. Haseltine. 1985. The location of *cis*-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat. *Cell* 41:813-823.

Rottman, J.N., RL. Widom, B. Nadal-Ginard, V. Mahdavi, and SK. Karathanasis. 1991. A retinoic acid-responsive element in the apolipoprotein A1 gene distinguishes between two different retinoic acid response pathways. *Molecular and Cellular Biology* 11:3814-3820.

Rottman, J.N., and J.I. Gordon. 1993. Comparison of the patterns of expression of rat intestinal fatty acid binding protein/human growth hormone fusion genes in cultured intestinal epithelial cell lines and in the gut epithelium of transgenic mice. *J Biol Chem* 268:11994-12002.

Rous, P. 1911. A sarcoma of the fowl transmissible by an agent separable from the tumour cells. *J. Exp. Med.* 13:397-411.

Roy, S., U. Delling, C-H. Chen, C. A. Rosen, and N. Sonnenberg. 1990. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat mediated trans-activation. *Genes and Development* 4:1365-1373.

Ruben, S., A. Perkins, R. Purcell, K. Joung, R. Sia, R. Burghoff, W.A. Haseltine, and C.A. Rosen. 1989. Structural and functional characterisation of human immunodeficiency virus tat protein. *J Virol* 63:1-8.

Ruben, S. M., P. J. Dillon, R. Schreck, T. Henkel, C. H. Chen, M. Maher, P. A. Baeuerle, and C. A. Rosen. 1991. Isolation of a *rel*-related human cDNA that potentially encodes the 65-kD subunit of NF-κB. *Science* 251:1490-1493.

Saag, M.S., B.H. Hahn, J. Gibbins, Y. Li, E.S. Parks, and G.M. Shaw. 1988. Extensive variation of human immunodeficiency virus type 1 *in vitro*. *Nature* 334:440-444.

Saag, M.S., M. Holodniy, D.R. Kuritzkes, W.A. O'Brien, R. Coombs, M.E. Poscher, D.M. Jacobsen, G.M. Shaw, D.D. Richman, and P.A. Volberding. 1996. HIV vial load markers in clinical practice. *Nature Med* 2:625-629.

Sadaie, M.R., J. Rappaport, T. Benter, S.F. Josephs, R. Willis, and F. Wong-Staal. 1988. Missense mutations in an infectious human immunodeficiency viral genome: functional mapping of tat and identification ot the rev splice acceptor. *PNAS* 85:9224-9228.

Sagami, I., SY. Tsai, H. Wang, M-J. Tsai, and BW. O'Malley. 1986. Identification of two factors required for transcription of the ovalbumin gene. *Molecular and Cellular Biology* 6:4259-4267.

Saito, Y., LR. Sharer, LG. Epstein, J. Michaels, M. Mintz, and et al. 1994. Over-expression of nef as a marker for restricted HIV-1 infection of astrocytes in post-mortem pediatric central nervous tissues. *Neurology* 44:474-481.

Saksela, K., G. Cheng, and D. Baltimore. 1995. Proline-rich (PxxP) motifs in HIV-1 nef bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of nef+ viruses but not for down-regulation of CD4. *EMBO* 14:484-491.

Salahuddin, SZ., RM. Rose, JE. Groopman, PD. Markham, and RC. Gallo. 1986. Human T lymphotropic virus type III infection of human alveolar macrophages. *Blood* 65:281-287.

Sambrook, J., E.F. Fritsch, T., T. Maniatis. 1989.
Molecular Cloning: A Laboratory Manual. 2nd edition.
Cold Spring Harbor Press.

Samson, M., F. Libert, B.J. Doranz, and et al. 1996. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 382:722-725.

Samuels, H. H., B. M. Forman, Z. D. Horowitz, and Z-S. Ye. 1989. Regulation of gene expression by thyroid hormone. *Ann Rev Physiol* 51:623-639.

Sanchez-Pescador, R, M.D. Power, P.J. Barr, K.S. Steinier, M.M. Stempien, S.L. Brown-Shimer, W.W. Gee, A. Renard, A. Randolph, J.A. Levy, D. Dina, and P.A. Luciw. 1985. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). *Science* 227:484-492.

Sap, J, A Munoz, J Schmitt, H Stunnenberg, and B Venstrom. 1989. Repression of transcription mediated at a thyroid hormone response element by the v-erb-A oncogene product. *Nature* 340:242-244.

Sawai, E.T., A. Baur, H. Struble, B.M. Peterlin, and J.A. Levy. 1994. Human immunodeficiency virus type 1 nef associates with a cellular serine kinase in T-lymphocytes. *PNAS* 91:1539-1543.

Sawaya, BE., O. Rohr, D. Aunis, and E. Schaeffer. 1996a. Regulation of human immunodeficiency virus type 1 gene transcription by nuclear receptors in human brain cells. *J Biol Chem.* 271:22895-22900.

Sawaya, B.E., O. Rohr, D. Aunis, and E. Schaeffer. 1996b. Chicken ovalbumin upstream promoter transcription factor, a transcriptional activator of HIV-1 gene expression in human brain cells. *J Biol Chem* 271:23572-23576.

Schaeffer, E., F. Guillou, D. Part, and M.M. Zakin. 1993. A different combination of transcription factors modulates the expression of the human transferrin promoter in liver and sertoli cells. *J Biol Chem* 268:23399-23408.

Schawaller, M., C.E. Smith, J.J. Skehel, and D.C. Wiley. 1989. Studies with cross-linking reagents on the oligomeric structure of the env-glycoprotein of HIV. *Virology* 172:367-369.

Schechter, M. T., and M. Laga. 1995. AIDS 1995. Epidemiology: overview. *AIDS* 9 Suppl A:S55-7.

Schennach, H., D. Schonitzer, H. Wachter, and D. Fuchs. 1997. Blood donations and viruses. *Lancet* 349:1327-8.

Scheppler, J., J. Nicholson, D. Swan, A. Ahmed-Ansari, and J. McDougal. 1989. Down-modulation of MHC-1 in a CD4 + cell line, CEM-E5, after HIV-1 infection. *J Immunol* 143:2858-2866.

Schmid, R. M., N. D. Perkins, C. S. Duckett, P. C. Andrews, and G. J. Nabel. 1991. Cloning of an NF-κB subunit which stimulates HIV transcription in synergy with P65. *Nature* 352:733-736.

Schnittman, S.M., M.C. Psallidopoulos, H.C. Lane, L. Thompson, M. Baselar, F. Massain, C.H. Fox, N.P. Salzman, and A.S. Fauci. 1989. The reservoir for HIV-1 in human peripheral blood is a T cell that maintains expression of CD4. *Science* 245:305-308.

Schreck, R., P. Rieber, and P. A. Baeuerle. 1991. Reactive oxygen intermediates as apparently widely used messengers in the activation of the NF-κB transcription factor and HIV-1. *EMBO J.* 10:2247-2258.

Schwabe, J.W., L. Chapman, J.T. Finch, and D. Rhodes. 1993.
The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements.
Cell 75: 567-578

Schwartz, C., F. Canonne-Hergaux, D. Aunis, and E. Schaeffer. 1997.
Characterisation of nuclear proteins that bind to the regulatory TGATTGGC motif in the HIV-1 long terminal repeat. *Nuc Acids Res* 25:1177-1184.

Schwartz, O., V. Marechal, O. Danos, and J-M. Heard. 1995. Human immunodeficiency virus type 1 nef increases the efficiency of reverse transcription in the infected cell. *J Virol* 69:4053-4059.

Schwartz, O., V. Marechal, S. Le Gall, F. Lemonnier, and J. Heard. 1996. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nature Medicine* 2:338-342.

Schwartz, S., B.K. Felber, D.M. Benko, E.M. Fenyo, and G.N. Pavlakis. 1990. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J Virol* 64:2519-2529.

Seale, J. 1989. Origins of the AIDS viruses: HIV-1 and HIV-2, fact or fiction? *J R Soc Med* 82:508.

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 and Development* 3:547-558.

Seth, A., D.R. Hodge, D.M. Thompson, L. Robinson, A. Panayiotakis, D.K. Watson, and T.S. Papas. 1993. Ets family proteins activate transcription from HIV-1 long terminal repeat. *AIDS Res Human Retroviruses* 9:1017-1023.

Sharer, LR. 1992. Pathology of HIV-1 infection of the central nervous system. *J Neuropath Exp Neurology.* 51:3-11.

Sharma, DP., MC. Zink, M. Anderson, R. Adams, JE. Clements, and et al. 1992. Derivation of neurotropic simian immunodeficiency virus from exclusively lymphotropic parenteral virus: pathogenesis of infection in macaques. *J Virol* 66:3550-3556.

Sharp. 1994. *AIDS* 8(suppl):S27-42.

Shaw, J. P., P. J. Utz, D. B. Durand, J. J. Toole, E. A. Emmel, and G. R. Crabtree. 1988. Identification of a putative regulator of early T cell activation genes. *Science* 241:202-205.

Sheline, C. T., L. H. Milocco, and K. A. Jones. 1991. Two distinct nuclear transcription factors recognise loop and bulge residues of the HIV-1 TAR RNA hairpin. *Genes and Dev* 5:2508-2520.

Shi, Y., E. Seto, L-S. Chang, and T. Shenk. 1991. Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* 67:377-388

Shioda, T., J.A. Levy, and C. Cheng-Meyer. 1991. Macrophage and T-cell line tropisms of HIV-1 are determined by specific regions of the envelope gp120 gene. *Nature* 349:167-169.

Siebenlist, U., and W. Gilbert. 1980. Contacts between Escherichia col RNA polymerase and an early promoter of phage T7. *PNAS* 77:122-126.

Siebenlist, U., G. Franzoso, and K. Brown. 1994. Structure, regulation and function of NF-κB. *Annu. Rev. Cell Biol.* 10:405-455.

Siegal, F.P., C. Lopez, G.S. Hammer, A.E. Brown, S.J. Kornfeld, J. Gold, J. Hassett, S.Z. Hirschman, C. Cunningham-Rundles, B.R. Adelberg, D.M. Parham, M. Siegal, S. Cunningham-Rundles, and D. Armstrong. 1981. The acquired immunodeficiency in male homosexual, manifested by chronic perianal ulcerative herpes simplex lesions. *N Engl J Med* 305:1439-1444.

Siekevitz, M., S.F. Josephs, M. Dukovich, N. Peffer, F. Wong-Staal, and W. Greene. 1987. Activation of the HIV-1 LTR by T-cell mitogens and the trans-activator protein of HTLV-1. *Science* 238:1575-1578.

Siena, Consensus Workshop. 1995. Strategies for prevention of perinatal transmission of HIV infection. Siena Consensus Workshop II.
J Acquir Immune Defic Syndr Hum Retrovirol 8:161-75.

Simon, F., S. Matheron, C. Tamalet, I. Loussert Ajaka, S. Bartczak, J. M. Pepin, C. Dhiver, E. Gamba, C. Elbim, J. A. Gastaut, and et al. 1993. Cellular and plasma viral load in patients infected with HIV-2. *AIDS* 7:1411-7.

Simpson, R. T. 1990. Nucleosome position can affect the function of a *cis* - acting DNA element *in vivo*. *Nature* 343:387-389.

Singh, H., J.H. LeBowitz, A.S. Baldwin, and P.A. Sharp. 1988. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. *Cell* 52:415-423.

Smith, M. R., and W. C. Greene. 1989. The same 50-kDa cellular protein binds to the negative regulatory element of the interleukin 2 receptor alpha chain gene and the human immunodeficiency virus type 1 long terminal repeat. *PNAS* 86:8526-8530.

Smith, W.C., H. Nakshatri, P. Leroy, J. Rees, and P. Chambon. 1991. A retinoic acid response element is present in the mouse cellular retinol binding protein I (mCRBPI) promoter. *EMBO* 10:2223-2230.

Smith, C.L., S.A. Onate, M-J. Tsai, and B.W. O'Malley. 1996. CREB binding protein acts synergistically with steroid receptor coactivator-1 to enhance steroid receptor-dependent transcription. *PNAS* 93:8884-8888.

Snider, WD., DM. Simpson, SL. Nielson, JWM. Gold, CE. Metroka, and JB. Posner. 1983. Neurological complications of acquired immunodeficiency syndrome: analysis of 50 patients. *Ann Neurol* 14:403-418.

Sodroski, J. G., R. Patarca, C. Rosen, F. Wong-Staal, and W. A. Haseltine. 1985. Location of the *trans* -activating region on the genome of the human T cell lymphotropic virus type III. *Science* 229:74-79.

Sova, P., and D.J. Volsky. 1993. Efficiency of viral DNA synthesis during permissive and non-permissive cells with *vif*-negative human immunodeficiency virus type 1. *J Virol* 67:6322-6326.

Spina, C.A., T.J. Kwoh, M.Y. Chowers, J.C. Guatelli, and D.D. Richman. 1994. The importance of nef in the induction of human immunodeficiency virus type 1 replication from primary quiescent CD4 lymphocytes. *J Exp Med* 179:115-123.

Spina, C.A., J.C. Guatelli, and D.D. Richman. 1995. Establishment of a stable, inducuble form of human immunodeficiency virus type 1 DNA in quiescent CD4 lymphocytes *in vitro. J Virol* 69:2977-2988.

Stanley, S.K., M.A. Ostrowski, J.S. Justement, K. Gantt, S. Hedayati, M. Mannix, K. Roche, D.J. Schwartzentruber, C.H. Fox, and A.S. Fauci. 1996. Effect of immunisation with a common recall antigen on viral expression in patients infected with human immunodeficiency virus type 1.
N Eng J Med 334:1222-1230.

Staprans, S.I., B.C. Hamilton, S.E. Folanskee, T. Elbeik, P. Barbosa, R.M. Grant, and M.B. Feinberg. 1995. Activation of virus replication after vaccintion of HIV-1 infected individuals. *J Exp Med* 182:1727-1737.

Starcich, B., B.H. Hahn, G.M. Shaw, and etal. 1986. Identification and characterisation of HTLVIII/LAV the retrovirus of AIDS. *Cell* 45:637-648.

Stein, B.S., S.D. Gowda, J.D. Lifson, R.C. Penhallow, K.G. Bensch, and E.G. Engleman. 1987. pH-independent HIV entry into CD4 positive T cells via virus envelope fusion to the plasma membrane. *Cell* 49:659-668.

Stein, D.S., J.A. Korvick, and S.H. Vermundi. 1992. CD4 positive lymphocyte cell enumeration for prediction of clinical course of human immunodeficiency virus disease: a review. *J Infect Diseases* 165:352-363.

Stein, D.S., N.M. Graham, L.P. Park, D.R. Hoover, J.P. Phair, R. Detels, M. Ho, and A.J. Saah. 1994. The effect of the interaction of aciclovir with zidovudine on progression to AIDS and survival. Analysis of data in the multicentre AIDS cohort study. *Ann Int Med* 121:100-108.

Stevenson, M, T.L. Stanwick, M.P. Dempsey, and C.A. Lamonica. 1990. HIV-1 replication is controlled at the level of T cell activation and proviral integration. *EMBO J* 9:1551-1560.

Strebel, K., K. Daughtery, D. Clouse, T. Cohen, T. Folks, and M.A. Martin. 1987. The HIV 'A' (*sor*) gene product is essential for virus infectivity. *Science* 328:728-730.

Strickland, S., and V. Mahdavi. 1978. The induction of differentiation in teratocarcinoma stem cells by treatment with retinoic acid. *Cell* 15:393-403. Subbramanian, R.A., and E.A. Cohen. 1994. Molecular biology of the human immunodeficiency virus accessory proteins. *J Virol* 68:6831-6835.

Subbramanian, R.A., A. Kessous-Elbaz, R. Lodge, J. Forget, X.J. Yao, D. Bergeron, and E.A. Cohen. 1998. Human immunodeficiency virus type 1 vpr is a positive regulator of viral transcription and infectivity in human macrophages. *Exper Med* 187:1103-1111.

Suzan, M., D. Salaun, C. Neuveut, B. Spire, I. Hirsch, P. Le Bouteiller, G. Querat, and J. Sire. 1991. Induction of NF-κB during monocyte differentiation by HIV type 1 infection. *J Immunol* 146:377-376.

Svejstrup, J.Q., P. Vichi, and J-M. Egly. 1996. The multiple roles of transcription/repair factor TFIIH. *Trends in Biochemical Science* 21:346-350.

Swingler, S., A. Morris, and A. Easton. 1994. Tumour necrosis factor α and interleukin-1β induce specific sub units of NFKB to bind the HIV-1 enhancer: characterisation of transcription factors controlling human immunodeficiency virus type 1 gene expression in neural cells. *Biochemical and Biophysical Research Communications.* 203:623-630.

Syu, W-J., W-R Lee, Q-C Du, M. Essex, and T-H Lee. 1991. Role of concerved gp41 cysteine residues in the processing of Human Immunodeficiency Virus envelope precursor and viral infectivity. *J Virol* 65:6349-6352.

Teich, N. 1985. Taxonomy of Retroviruses. In *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*, edited by R. Weiss, N. Teich, H. Varmus and C. J: Cold Spring Harbour Laboratories, Cold Spring Harbour, N.Y.

Terwilliger, E., J.G. Sodroski, C.A. Rosen, and W.A. Haseltine. 1986. Effects of mutations within the 3' *orf* open reading frame region of human T-cell virus type III (HTLV-III/LAV) on replication and cytopathogenicity. *J Virol* 60:754-760.

Tesmer, V. M., A. Rajadhaksha, J. Babin, and M. Bina. 1993. NF-IL6-mediated transcriptional activation of the long terminal repeat of the human immunodeficiency virus type 1. *PNAS* 90:7298-7302.

Thomas, D.J., J.S. Wall, J.F. Hainfield, M. Kaczorek, F.P. Booy, B.L. Trus, F.A. Eiserling, and A.C. Steven. 1991. gp160, the envelope glycoprotein of human immunodeficiency virus type 1, is a dimer of 125 kilodalton subunits stabilised throughinteractions between their gp41 domains. *J Virol* 65:3797-3803.

Thomas, PS. 1980. Hybridisation of denatured RNA and small DNA fragments transferred to nitrocellulose. *PNAS* 77:5201-5205.

Thompson, C.C., and McKnight S.L. 1992. Anatomy of an enhancer. *Trends in Genetics* 8:232-236.

Tindall, B., and D.A. Cooper. 1991. Primary HIV infection: host responses and intervention strategies. *AIDS* 5:1-14.

Tini, M., G. Otulakowski, M.L. Breitman, L-C. Tsui, and V. Giguere. 1993. An everted repeat mediates retinoic acid induction of the gF-crystalline gene: evidence of a direct role for retinoids in lens development. *Genes and Development* 7:295-307.

Toggas, S.M., E. Masliah, E.M. Rockenstein, G.F. Rall, and C.R. Abraham. 1994. Central nervous system damage produced by expression of the HIV-1 coat protein gp120 in transgenic mice. *Nature* 367:188-193.

Tominaga, K., S. Kato, M. Negishi, and T. Takano. 1997. A high frequency of defective *vif* genes in peripheral blood mononuclear cells from HIV type 1 infected individuals. *AIDS Research and Human Retroviruses.* 12:1543-1549.

Tornatore, C., R. Chandra, JR. Berger, and EO. Major. 1994. HIV-1 infection of subcortical astrocytes in the pediatric central nervous system. *Neurology* 44:481-487.

Toyama, R., S.M. Bende, and R. Dhar. 1992. Transcriptional activity of the human immunodeficiency virus type 1 LTR promoter in fission yeast Schizosaccharomyces pombe. *Nucleic Acids Res* 20:2591-2596.

Tran, P.B.V., X.K. Zhang, G. Salbert, T. Hermann, J.M. Lehmann, and M. Pfahl. 1992. COUP orphan receptors are negative regulators of retinoic acid response pathways. *Mol Cell Biol* 12:4666-4676.

Trechsel, U, V Evequoz, and H Fleisch. 1988. Stimulation of interleukin 1 and 3 production by retinoic acid *in vitro*.
Biochem J. 230:339.

Trono, D. 1995. HIV accessory proteins: leading roles for the supporting cast. *Cell* 82:189-192.

Tsai, M-J., and B.W. O'Malley. 1994. Molecular mechanisms of action of steroid/thyroid receptor superfamily. *Ann Rev. Biochem.* 63:451-486.

Tsai, S-Y., I. Sagami, L.H. Wang, M-J. Tsai, and B.W. O'Malley. 1987. Interactions between a DNA-binding transcription factor (COUP) and a non-DNA-binding factor (S300-II). *Cell* 50:701-709.

Tsai, S-Y., D.J. Carlstedt, N.L. Weigel, K. Dahlman, J.A. Gustafsson, M-J. Tsai, and B.W. O'Malley. 1988. Molecular interactions of steroid hormone receptor with its enhancer element: evidence for receptor dimer formation. *Cell* 55:361-369.

Tsai, S-Y., and M-J Tsai. 1997. Chick ovalbumin upstream promoter-transcription factors (COUP-TFs): coming of age. *Endocrine Rev* 18:229-240.

Turpin, J, M Blink, K Mehta, and G Lopez-Bernstein. 1990. Effect of retinoic acid and retinol on TNF release and *c-fos* and TNF mRNA expression during human monocyte *in vitro* differentiation. *J. Leukocyte Biol* 48:205.

Turpin, JA, M Vargo, and MS Meltzer. 1992. Enhanced HIV-1 replication in retinoid-treated monocytes. *J Immunol* 148:2539-2546.

Vaishnav, Y. N., and F. Wong-Staal. 1991. The Biochemistry of AIDS.
Annual Rev. Biochem. 60:577-630.

van Lint, C., A. Burny, and E. Verdin. 1991. The intragenic enhancer of human immunodeficiency virus type 1 contains functional AP-1 binding sites.
J Virol 65:7066-7072.

van Lint, C., J. Ghysdael, P. Paras, A. Burny, and E. Verdin. 1994. A transcriptional regulatory element is associated with a nuclease-hypersensitive site in the *pol* gene of human immunodeficiency virus type-1. *J Virol* 65:6790-6799.

van Lint, C., S. Emiliani, M. Ott, and E. Verdin. 1996. Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. *EMBO J.* 15:1112-1120.

van Lint, C., C. A. Amella, S. Emiliani, M. John, T. Jie, and E. Verdin. 1997. Transcription binding sites downstream of the human immunodeficiency virus type 1 transcription start site are important for virus infectivity. *J Virol* 71:6113-6127.

Vanden Haesevelde, M., J. L. Decourt, R. J. De Leys, B. Vanderborght, G. van der Groen, H. van Heuverswijn, and E. Saman. 1994. Genomic cloning and complete sequence analysis of a highly divergent African human immunodeficiency virus isolate. *J Virol* 68:1586-96.

Vanden Haesevelde, M. M., M. Peeters, G. Jannes, W. Janssens, G. van der Groen, P. M. Sharp, and E. Saman. 1996. Sequence analysis of a highly divergent HIV-1-related lentivirus isolated from a wild captured chimpanzee. *Virology* 221:346-50.

Vanley, G. T., R. Huberman, and R. B. Lufkin. 1982. Atypical Pneumocystis carinii pneumonia in homosexual men with unusual immunodeficiency. *Am J Roentgenol* 138:1037-41.

Vasios, G., S. Mader, J. D. Gold, M. Leid, Y. Lutz, M-P. Gaub, P. Chambon, and L. Gudas. 1991. The Late Retinoic Acid Induction of Laminin B1 Gene Transcription Involves RAR Binding to the Responsive Element. *EMBO J.* 10:1149-1158.

Verdin, E. 1991. DNase I-hypersensitive sites are associated with both long terminal repeats and with the intragenic enhancer of integrated human immunodeficiency virus type 1. *J Virol* 65:6790-6799.

Verdin, E., P. Paras, and C. van Lint. 1993. Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. *EMBO J* 12:3249-3259.

Verma, I. M., J. K. Stevenson, E. M. Schwarz, D. Van Antwerp, and S. Miyamoto. 1995. Rel/NF-κB/IκB family: intimate tales of association and dissociation. *Genes and Devel* 9:2723-2735.

Verrijzer, C.P., and R. Tjian. 1996. TAFs mediate transcriptional activation and promoter selectivity. *TIBS* 21:338-341.

Vinson, C.R., K.L. LaMarco, P.F. Johnson, W.H. Landschulz, and S.L. McKnight. 1988. In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes Devel* 2:801-806.

Vittecoq, D., S. Chevret, L. Morand-Joubert, F. Heshmati, F. Audat, M. Bary, T. Dusantoir, A. Bismuth, J.P. Vivard, and F. Barre-Sinoussi. 1992. Passive immunotherapy in AIDS: a double-blind randomised study based on transfusions of plasma rich anti-human immunodeficiency virus 1 antibodies vs transfusions of seronegative plasma. *PNAS* 92:1195-1199.

Vivanco-Ruiz, M., T. H. Bugge, P. Hirschmann, and H. G. Stunnenberg. 1991. Funtional Characterization of a Natural Retinoic Acid Responsive Element. *EMBO J* 10:3829-3838.

von der Ahe, D, DS Janich, C Scheidereit, R Renkawitz, G Schultz, and M Beato. 1985. Glucocorticoid and progesterone receptors bind to the same sites in two hormonally regulated promoters. *Nature* 313:706.

von Schwedler, U., C. Song, C. Aiken, and D. Trono. 1993. Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. *J Virol* 67:4945-4955.

Vuylsteke, B., R. Bastos, J. Barreto, T. Crucitti, E. Folgosa, J. Mondlane, T. Dusauchoit, P. Piot, and M. Laga. 1993. High prevalence of sexually transmitted diseases in a rural area in Mozambique. *Genitourin Med* 69:427-430.

Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Allzon. 1985. Nucleotide sequence of the AIDS virus, LAV. *Cell* 40:9-17.

Wang, L.H., S.Y. Tsai, I. Sagami, M-J. Tsai, and B.W. O'Malley. 1987. Purification and characterisation of chicken ovalbumin upstream promoter transcription factor from HeLa cells. *J Biol Chem* 262:16080-16086.

Wang, L.H., S.Y. Tsai, R.G. Cook, W.G. Beattie, M-J. Tsai, and B.W. O'Malley. 1989. COUP transcription factor is a member of the steroid receptor superfamily. *Nature* 340:163-166.

Wang, L-H., N.H. Ing, S.Y. Tsai, B.W. O'Malley, and M-J. Tsai. 1991. The COUP-TFs compose a family of functionally related transcription factors. *Gene Expression* 1:207-216.

Wang, L., S. Mukherjee, F. Jia, O. Narayan, and L.J. Zhao. 1998. Interaction of virion protein vpr of human immunodeficiency virus type 1 with cellular transcription factor Sp1 and transactivation of viral long terminal repeat. *J Biol Chem* 270:25564-25569.

Waterman, M. L., and K. A. Jones. 1990. Purification of TCF-1 alpha, a T-cell specific transcription factor that activates the T cell receptor C alpha gene enhancer in a context-dependent manner. *New Biologist* 2:621-636.

Waterman, M. L., W. H. Fischer, and K. A. Jones. 1991. A thymus specific member of the HMG protein family regulates the human T cell receptor C alpha enhancer. *Genes Dev* 5:656-669.

Wei, X., S.K. Ghosh, M.E. Taylor, V.A. Johnson, E.A. Emini, P. Deutsch, J.D. Lifson, S. Bonhoeffer, M.A. Nowak, B.H. Hahn, M.S. Saag, and G.M. Shaw. 1995. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 373:117-122.

Weniger, B. G., K. Limpakarnjanarat, K. Ungchusak, S. Thanprasertsuk, K. Choopanya, S. Vanichseni, T. Uneklabh, P. Thongcharoen, and C. Wasi. 1991. The epidemiology of HIV infection and AIDS in Thailand. *AIDS* 5 Suppl 2:S71-85.

WHO. 1995. Global Programme on AIDS. The current global situation of the HIV-AIDS pandemic. *Weekly Epidemiological Record.* 70:353-355.

Wieland, U., A. Seelhoff, A. Hofmann, J.E. Kuhn, H.J. EGGERS, P. Mugyenyi, and S. Schwander. 1997. Diversity of the *vif* gene of human immunodeficiency virus type 1 in Uganda. *J Gen Virol* 78:393-400.

Wiley, CA., RD. Schrier, JA. Nelson, PW. Lampert, and MBA. Oldstone. 1986. Cellular localisation of human immunodeficiency virus infection within the brains of acquired immune deficiency syndrome patients. *PNAS* 83:7089-7093.

Wilkinson, M, J. Doskow, and S. Lindsey. 1991. RNA blots: staining procedures and optimization of conditions. *Nucleic Acids Res.* 19:679.

Willey, RL., JS. Bonifacino, BJ. Potts, MA. Martin, and RD. Klausner. 1988. Biosynthesis, cleavage, and degredation of the human immune deficiency virus type 1 envelope glycoprotein gp160. *PNAS* 85:9580-9584.

Willey, R.L., F. Maldarelli, M.A. Martin, and K. Strebel. 1992a. Human immunodeficiency virus type 1 vpu protein regulates the formation of intracellular gp160-CD4 complexes. *J Virol* 66:226-234.

Willey, R.L., F. Maldarelli, M.A. Martin, and K. Strebel. 1992b. Human immunodeficiency virus type 1 vpu protein induces rapid degradation of CD4. *J Virol* 66:7193-7200.

Wu, F. K., J. A. Garcia, D. Harrich, and R. B. Gaynor. 1988a. Purification of the human immunodeficiency virus type 1 enhancer and TAR-binding proteins EBP-1 and UBP-1. *EMBO J.* 7:2117-2129.

Wu, F. K., J. Garcia, R. Mitsuyasu, and R. Gaynor. 1988b. Alterations in binding characteristics of the human immunodeficiency virus enhancer factor. *J Virol* 62:218-225.

Wu, F., J. Garcia, and D. Sigman. 1991. Tat regulates binding of the human immunodeficiency virus *trans*-activating region RNA loop-binding protein TRP-185. *Genes and Development* 5:2128-2140.

Yamaguchi, K., J.E. Groopman, and R.A. Byrn. 1994. The regulation of HIV by retinoic acid correlates with cellular expression of the retinoic acid receptors. *AIDS* 8:1675-1682.

Yamamoto, K., S. Mori, T. Okamoto, K. Shimotohno, and Y. Kyogoku. 1991. Identification of transcriptional suppressor proteins that bind to the negative regulatory element of the human immunodeficiency virus type 1. *Nucleic Acids Res* 19:6107-12.

Yamamoto, K. R. 1985. Steroid receptor regulated transcription of specific genes and gene networks. *Annual Review of Genetics* 19:209-252.

Yamamoto, T., and M. Horikoshi. 1997. Novel substrate specificity of the histone acetyltransferase activity of HIV-1-Tat interactive protein Tip60. *J Biological Chem* 272:30595-30598.

Yang, Z., and J. D. Engel. 1993. Human T cell transcription factor GATA-3 stimulates HIV-1 expression. *Nucleic Acids Res* 21:2831-2836.

Yao, T.P., G. Ku, N. Zhou, R. Scully, and D.M. Livingstone. 1996. The nuclear hormone receptor coactivator SRC-1 is a specific target of p300. *PNAS* 93:10626-10631.

Yoon, J-B., G. Li, and R. B. Roeder. 1994. Characterisation of a family of related cellular transcription factors which can modulate human immunodeficiency virus type 1 transcription in vitro. *Mol. Cell Biol.* 14:1776-1785.

Yu, L., Z. Zhang, P. M. Loewenstein, K. Desai, Q. Tang, D. Mao, J. S. Symington, and M. Green. 1995. Molecular cloning and characterisation of a cellular protein that interacts with the human immunodeficiency virus type 1 Tat transactivator and encodes a strong transcriptional activation domain. *J Virol* 69:3007-3016.

Zack, J.A., S.J. Arrigo, S.R. Weitsman, A.S. Go, A. Haislip, and I.S. Chen. 1990. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. *Cell* 61:213-222.

Zack, J.A., A.M. Haislip, P. Krogstad, and I.S. Chen. 1992. Incompletely reverse transcribed human immunodeficiency virus type 1 genomes in quiescent cells can function as intermediates in the retroviral life-cycle. *J Virol* 66:1717-1725.

Zagury, D., J. Bernard, R. Leonard, R. Cheynier, M. Feldman, P. S. Sarin, and R. C. Gallo. 1986. Long term culture of HTLV-III-infected T cells: a model of cytopathology of T-cell depletion in AIDS. *Science* 231:850-853.

Zawel, L., and D. Reinberg. 1995. Common themes in assembly and function of eukaryotic transcription complexes. *Ann Rev Bioch* 64:522-561.

Zeichner, S. L., J. Y. H. Kim, and J. C. Alwine. 1991. Linker-scanning mutational analysis of the transcriptional activity of the human immunodeficiency virus type 1 long terminal repeat. *J Virol* 65:2436-2444.

Zelent, A., A. Krust, M. Petrovich, P. Kastner, and P. Chambon. 1989. Cloning of murine alpha and beta retinoic acid receptors and a novel gamma predominantly expressed in skin. *Nature* 339:714-717.

Zhang, L., Y. Huang, H. Yuan, B. K. Chen, J. Ip, and D. D. Ho. 1997. Identification of a replication-competent pathogenic human immunodeficiency virus type 1 with duplication in the TCF-1α region but lacking NF-κB binding sites. *J Virol* 71:1651-1656.

Zhang, X-K., J. Lehmann, B. Hoffman, M.I. Dawson, J. Cameron, G. Graupner, T. Hermann, and M. Pfahl. 1992. Homodimer formation of retinoid X receptor induced by 9-cis retinoic acid. *Nature* 358:587-591.

Zhang, Y., K. Doyle, and M. Bina. 1992. Interactions of HTF4 with E-box motifs in the long terminal repeat of human immunodeficiency virus type 1.
J. Virol. 66:5631-5634.

Zhou, Q., R. W. Gedrich, and D. A. Engel. 1995. Transcriptional repression of the c-fos gene by YY1 is mediated bya direct interaction with ATF/CREB.
J Virology 69:4323-4330.

Zhu, T., H. Mo, N. Wang, D.S. Nam, Y. Cao, R.A. Koup, and D.D. Ho. 1993. Genotypic and phenotypic characterisation of HIV-1 patients with primary infection.
Science 261:1179-1181.

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

Additional references.

Farzan M, Choe H, Desjardins E, Sun Y, Kuhn J, Cao S, Archambault D, Kolchinsky P, Koch M, Wyatt R, Sodroski J. Stabilisation of Human Immunodeficiency Virus Type 1 envelope glycoprotein trimers by disulphide bonds introduced into gp41 glycoprotein. *J Virol* 1998 72:7620-7625

Martinson JJ, Chapman NH, Lin YT, Clegg JB. Global distribution of the CCR5 gene 32 base-pair deletion. *Nat Genet* 1997 16:100-103

Wimmer J, Fujinaga K, Taube R, Cujec TP, Zhu Y, Peng J, Price DH, Peterlin BM. Interactions between TAT and TAR and human immunodeficiency virus replication are facilitated by human cyclin T1 but not cyclins T2a or T2b. *Virology* 1999 255:182-189

Ramanathan Y, Reza SM, Young TM, Mathews MB, Pe'ery T. Human and rodent transcription elongation factor P-TEFb: interactions with human immunodeficiency virus type 1 tat and carboxy-terminal domain substrate. *J Virol* 1999 73:5448-5458

Gao F, Bailes, E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes. Nature* 1999 397:436-441