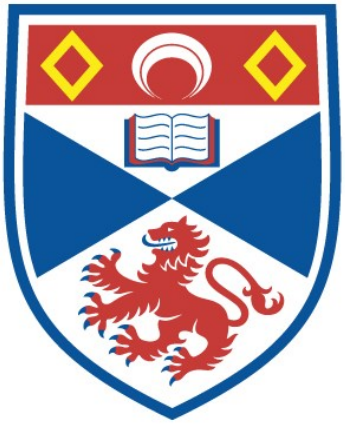
INTERACTION OF THE HELA CELL PROTEIN EBP1 WITH VIRAL AND CELLULAR ENHANCERS

Lilian Clark

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1989

Full metadata for this item is available in
St Andrews Research Repository
at:
http://research-repository.st-andrews.ac.uk/

Please use this identifier to cite or link to this item:
http://hdl.handle.net/10023/13971

This item is protected by original copyright

Interaction of the HeLa Cell Protein, EBP1, with Viral and Cellular Enhancers

by

Lilian Clark

A Thesis Presented for the Degree of

Doctor of Philosophy

in

The Faculty of Science

at

The University of St. Andrews

Department of Biochemistry and Microbiology
Irvine Building
North Street
St. Andrews
Fife
KY16 9AL

December 1988

ProQuest Number: 10170752

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 10170752

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

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

TL A900

DECLARATION FOR THE DEGREE OF Ph.D.

(a) I Lilian Clark hereby certify that this thesis has been composed by myself, that it is a record of my own work, and that it has not been accepted in partial or complete fulfilment of any other degree of professional qualification.

Signed ... **Date** 20th December 1988

(b) I was admitted to the Faculty of Science of the University of St. Andrews under Ordinance General No 12 on 1st October 1985 and as a candidate for the degree of Ph.D. on 1st October 1985.

Signed... **Date** 20th December 1988

(c) I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate to the Degree of Ph.D.

Signature of Supervisor... .... **Date** 20 December 1988

(d) In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am deeply indebted to Dr. Ron Hay for his supervision, guidance and many long and fruitful discussions. Thanks are also due to all members of the "Hay" group for their help over the past three years, especially Iain Nicholl for provision of HeLa cells and synthetic oligonucleotides, and to John Nicholson for help in the computer graphics analysis. I am extremely grateful to Bill Blythe for his excellent and prompt photographic work, to Mike Burdon for painstakingly editing this thesis and to Alexander for putting things in perspective.

Many thanks are due to my parents for all their help and encouragement (not forgetting the "red cross parcels"), and to my sister, Josephine, for her support and wonderful "Taxi service".

Throughout the course of this study the author was the recipient of a Science and Engineering Research Council Studentship for postgraduate training.

SUMMARY

A cellular protein, EBP1, was identified, which bound to the "core" region of the SV40 enhancer. The DNA sequences requirements for specific EBP1 binding distinguished this protein from a number of previously reported enhancer binding proteins. Multiple copies of the EBP1 binding site could serve, *in vivo,* to enhance transcription. The *in vivo* transcriptional activity of wild type and mutant SV40 enhancers correlated with EBP1 binding. Mutations that abolish EBP1 binding also severely reduced transcription, indicating that this protein may be important for SV40 enhancer activity.

Nuclease protection and chemical probing studies identified purine bases and backbone phosphate groups which participate in the formation of a specific EBP1/DNA complex. "Footprinting" with hydroxyl radicals revealed deoxyribose residues in the binding site which are protected from cleavage by EBP1. Computer graphics were used to display this information indicating that EBP1 made specific base and backbone contacts over one complete turn of the DNA double helix, supporting a model whereby the protein makes sequence-specific contacts in the major groove, although binding may also be influenced by interactions in the minor groove.

Competition and contact point analyses revealed that EBP1 bound similarly to sites present in several other viral and cellular enhancers. Mutational analysis of EBP1 binding sites identified base pairs important for specific EBP1/DNA complex formation. All high affinity binding sites contained the sequence 5'–GG$(N)_6$CC–3'. Although single base pair changes in the region between the conserved guanines and cytosines can generally be tolerated, it is clear that the central six base pairs contribute to binding affinity. Mutations in the recognition site which could lead to gross structural changes in the DNA abolish EPB1 binding.

ABBREVIATIONS

A	adenine
APS	ammonium persulphate
ATP	adenosine triphosphate
bp	base pair
Br.dUTP	5-bromo-2' deoxyuridine triphosphate
BSA	bovine serum albumin
C	cytosine
cDNA	complemetary deoxyribonucleic acid
CNBr	cyanogen bromide
c.p.m.	counts per minute
dATP	2-deoxyadenosine-5' triphosphate
dCTP	2-deoxycytidine-5' triphosphate
dGTP	2-deoxyguanosine-5' triphosphate
dNTP	deoxynucleotide triphosphate
dTTP	2-deoxythymidine-5' triphosphate
DEAE	diethylaminoethyl
DMS	dimethylsulphate
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetra-acetic acid
EGTA	ethylene glycol-bis(ß-aminoethylether) N,N,N',N-tetra-acetic acid
EtBr	ethidium bromide
G	guanine
HEPES	*N*-2-Hydroxyethylpiperazine *N'*-2-ethanesulphonic acid
KOAc	potassium acetate

mRNA	messenger ribonucleic acid
NaOAc	sodium acetate
NP40	nonidet P40
OD	optical density
PMSF	Phenylmethylsulphonyl fluoride
RNA	ribonucleic acid
r.p.m.	revolutions per minute
SDS	sodium dodecyl sulphate
SV40	simian virus 40
T	thymine
TCA	trichloroacetic acid
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol, (tris)
UV	ultra violet
v/v	volume to volume ratio
w/v	weight to volume ratio

INTRODUCTION

Transcription is the process whereby genes, .

are selectively located, recognised and transcribed into RNA by DNA-dependent RNA polymerases. At each step in RNA synthesis, transcription is regulated via a complex series of interactions of the RNA polymerase, DNA, nascent RNA, a variety of regulatory proteins (transcription factors) and small molecule ligands: this control of gene expression is critical to cell function and differentiation throughout the lifetime of the cell (reviewed by von Hippel *et al.*, 1984).

Genes determine the structures of the molecules that constitute living cells. At any one time, each cell uses only a subset of its genes to direct the synthesis of other molecules - these expressed genes are "switched on" whereas those not expressed are "switched off" (reviewed by Ptashne, 1986b). Although genes are transmitted essentially unchanged from parental to daughter cells, their expression may vary in different cell types. How are these genes "switched on" and "off" and what are the molecular mechanisms controlling gene expression?

In an attempt to answer these questions gene expression in the bacterium *E. coli,* its bacteriophage and various animal viruses has been extensively studied, and what emerges is that essentially the same molecular principles are used in prokaryotic and eukaryotic systems.

At a glance, it would appear that DNA plays a somewhat passive role in these control mechanisms, and, that it is only through the interactions of proteins with specific sequences in the DNA that the messages encoded in DNA are read and acted upon. Consequently, of prime importance in the control of gene expression, is the study of DNA–protein complexes. Identifying proteins which bind to specific sequences in DNA will provide us with a better understanding of the mechanisms controlling gene expression; understanding the chemistry of how genetic "switches" are controlled will require detailed structural information on these DNA–protein complexes.

The aim of this introduction is to provide a background to the initiation of transcription in both prokaryotes and eukaryotes, and to provide some insight into to the various mechanisms controlling gene expression, with particular emphasis on the interactions of regulatory proteins with transcriptional control elements.

1. *Control of Gene Expression in Prokaryotes and their Bacteriophages*

Gene expression involves the sequential flow of information from DNA to protein. The first stage of this process, transcription, involves the synthesis of an RNA chain representing one strand of the DNA duplex, the template strand. In contrast to DNA replication, the genome is not copied in its entirety but rather as defined units of genetic information which are transcribed into RNA molecules. These messenger RNAs (mRNAs) are then used in the process known as translation to direct the synthesis of proteins.

Control of initiation of RNA synthesis by DNA-dependent RNA polymerase appears to occur at two distinct levels: the frequency and the site of initiation. Initiation of transcription involves the binding of RNA polymerase to double stranded DNA which then unwinds to make available the template strand. Local unwinding of the DNA occurs in the region of bound RNA polymerase. Initiation is generally regarded as including the primary recognition of double stranded DNA, unwinding of the DNA helix and incorporation of the first few nucleotides in the RNA chain. The entire nucleotide sequence necessary to facilitate such an interaction is called the promoter. The site of incorporation of the first nucleotide is known as the mRNA start site (reviewed by Lewin, 1985).

The sites of control of transcriptional initiation, the promoters, are segments of conserved DNA sequence that bind RNA polymerase, and which are located upstream (5') of the initiation or mRNA start site (Rosenberg and Court, 1979; Siebenlist *et al.*, 1980). These features in the DNA, necessary for RNA polymerase binding, were identified by comparing the sequences of different promoters. If a particular nucleotide is essential for promoter function then theoretically it should be conserved in all promoters. However, this base need not be conserved at every single position; some variation may be permitted. Putative DNA recognition sites can be defined with regard to an optimum sequence that represents the base most often present at each position. A consensus sequence is defined by aligning all known examples so as to maximise their homology. Such conserved sequences have been located approximately 10 bp and 35 bp upstream of the initiation sites of *E. coli* promoters (reviewed by Lewin, 1985). In addition to these constitutive promoter elements, sequences governing negative and/or positive regulation have been identified.

Negative regulation of transcription involves the binding of a repressor protein to operator DNA (which often overlaps with the RNA polymerase binding site). Positive regulation, on the other hand, involves binding of an activator protein to a specific DNA sequence upstream of the promoter and stimulating transcription, presumably by making contact with RNA polymerase (Hochschild *et al.*, 1983). The structure of a typical prokaryotic promoter is shown in Figure 1.

RNA chain initiation by bacterial RNA polymerase is dependent on the ability of the enzyme to act as a sequence-specific DNA binding protein (often in concert with other proteins), to then effect unwinding of DNA and strand separation over a limited region, and finally to act as an RNA synthetase. In addition, initiation is regulated by both general and promoter-specific means facilitated by the direct interaction and response of RNA polymerase to low molecular weight effectors and other regulatory proteins (reviewed by Travers, 1987a).

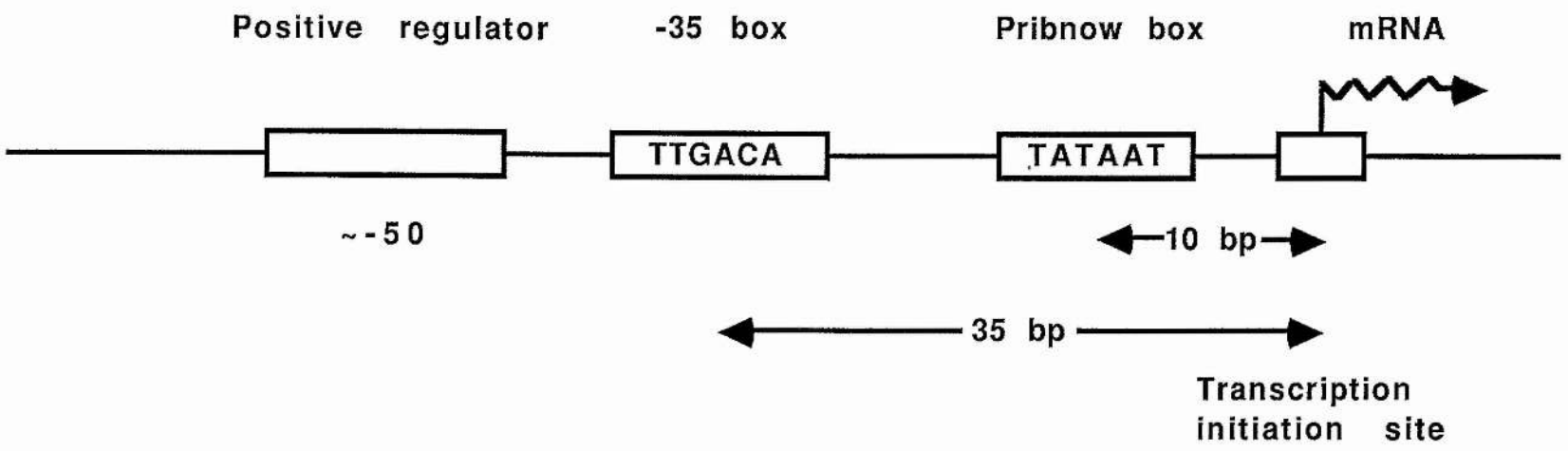
The relationship of structure to function is central to understanding the biological processes governing transcription initiation. Detailed structural information on promoters is now widely available but for RNA polymerase structural information is not quite so abundant. However, a rather simplified model involving only three overall steps, has for a number of years been invaluable in the investigations of transcription initiation (reviewed by McClure, 1985). This scheme, outlined in Figure 2, was suggested as a result of observations by Walter and coworkers (1967) and involves the initial binding of RNA polymerase and the promoter to form an inactive intermediate referred to as the closed complex (RP_C). This closed complex subsequently isomerises to form the transcriptionally active open complex (RP_O). Both binding and isomerisation involve noncovalent interactions. The energy available in nucleoside triphosphates is not utilised until the subsequent binding of the template-directed nucleoside triphosphates occurs and RNA chain elongation begins, eventually leading to promoter clearance (reviewed by McClure, 1985).

1.1 RNA Polymerase

The eubacterial DNA-dependent RNA polymerase is a complex multisubunit enzyme which is structurally and evolutionarily related to the RNA polymerases of archaebacteria and eukaryotes (reviewed by Travers, 1987a). The best studied prokaryotic RNA polymerase is that isolated from *E.*

Figure I. Structure of Prokaryotic (eubacterial) promoters

The -35 box and TATA (or Pribnow) box interact with RNA polymerase. Positive regulation can be mediated via neighbouring upstream elements binding activator proteins which in turn contact RNA polymerase. Spacing of these control elements is quite strict – the distances of the various elements from the transcription initiation site are indicated. In addition, prokaryotic promoters are often regulated by a repressor protein binding to an operator sequence near the transcription start site (not shown on this diagram).


Positive regulator
-35 box
Pribnow box
mRNA
TTGACA
TATAAT
~-50
10 bp
35 bp
Transcription initiation site

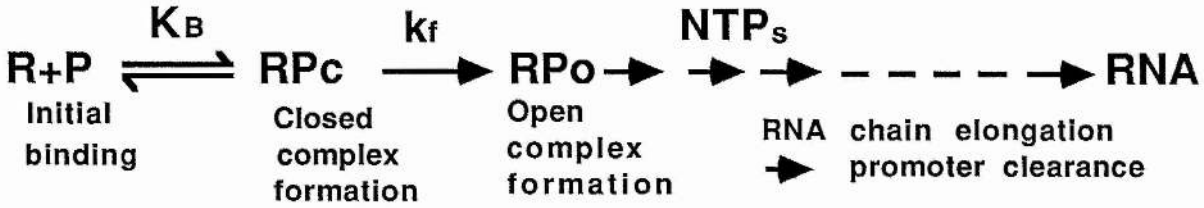


Figure 2. Mechanism of transcription initiation

This scheme involves the initial binding of RNA polymerase (R) to promoter DNA (P) with a binding constant, K_B, to form an inactive intermediate known as the closed complex (RP_C). This complex subsequently isomerises, with rate constant k_f, to form the transcriptionally active open complex (RP_O). Binding of template-directed nucleoside triphosphates (NTPs) occurs and RNA chain elongation commences.

coli . The enzyme can be functionally separated into two macromolecular components, the core polymerase and a sigma (σ) factor (Burgess *et al.*, 1969). The structure of the enzyme is understood only at relatively low resolution from neutron and X-ray small angle studies (Meisenberger, 1980). However, it is well established that the core enzyme, which has the subunit composition $\alpha_2\beta\beta'\omega_2$ with individual molecular weights of 155.6, 150.6, 13.5 and ~10 Kdaltons respectively (Chamberlin, 1982), contains the catalytic machinery necessary for the synthesis of an RNA chain. The genes encoding the α, ß and ß' subunits (*rpoA, rpoB,* and *rpoC* respectively) have been cloned and shown to be essential for polymerase function; the involvement of the ω subunits still remains to be determined. The sigma factor associates with the core polymerase to form the holoenzyme which can bind specifically to promoter sites to effect correct initiation of RNA chains at the majority of *E.coli* promoters. The major sigma factor in *E. coli* is the vegetative sigma of molecular weight 70.2 Kdaltons encoded by the *rpoD* gene (Burton *et al.*, 1981) and which is present in the form of RNA polymerase active in the selection of the majority of *E. coli* promoters. Transcription of other sets of genes, for example those activated by heat shock or genes required for nitrogen fixation, require sigma factors encoded by the *rpoH* and *rpoN* (*ntrA*) genes respectively (Grossman *et al.*, 1984; Kustu *et al.*, 1986). However, the structures of the different sigma factors are similar and all contain two regions of "helix-turn-helix" motif, characteristic of a variety of sequence-specific DNA binding proteins, as well as a conserved region postulated to be a site of contact between the core polymerase and the sigma factor (Stragier*et al.*, 1985, Doi and Wang, 1986).

Investigation of RNA synthesis in *Bacillus subtilis* (*B. subtilis*) has demonstrated the presence of at least five sigma factors one of which has been shown to be involved in transcriptional specificity during the early stages of sporulation development (Losick and Youngman, 1984). In addition, SP01, a *B. subtilis* bacteriophage (phage), encodes two factors that are responsible for temporal expression of its genes (Losick and Pero, 1981). Phage-induced RNA polymerases display a very high degree of specificity for their promoters on the corresponding phage DNA. Moreover, these polymerases appear to recognise a contiguous sequence of DNA near the transcription start site (reviewed by McClure, 1985). Diverse strategies are exhibited by the various phage which is

reflected in the finding that some phage use the host enzyme early and then use a phage encoded enzyme for the intermediate and late transcription units. Other phage use a modified form of the host RNA polymerase throughout their development.

Very little is known about the detailed three dimensional structure of the RNA polymerase holoenzyme. However X-ray scattering and electron microscope studies have provided some insight into the overall shape of the molecule in solution (Hawley and McClure, 1983) – details on the precise geometry of the DNA around the enzyme in the polymerase-promoter complex remain to be elucidated.

1.2 Promoter Structure

The nucleotide sequence of a promoter site should reflect not only its capacity to act as a specific recognition site for DNA binding proteins, but also to facilitate strand separation. The DNA sequences of more than 200 promoters utilised by RNA polymerase from *E. coli* and the closely related *Salmonella typhimurium* have now been determined (Hawley and McClure, 1983 for sequences up to 1982). Comparison of these sequences suggest that the characteristics of the different classes of promoter allows differential recognition by DNA binding proteins and the subsequent enzymatic manipulation which results in strand separation. However, when functional parameters of an individual promoter are to be derived from structural information alone, such DNA sequences are of little predictive value. A better understanding of how the complex functional make-up of a promoter is stored in the nucleotide sequence requires quantitative information on the function of defined promoter sequences.

Differential recognition of promoters by RNA polymerase containing different sigma subunits is the principal determinant for the separation of promoter sequences into discrete classes (reviewed by Travers, 1987a). In general, a promoter can be defined as the segment of DNA that contains signals which direct the proper binding of the RNA polymerase holoenzyme and its subsequent activation to a form capable of specific RNA transcription initiation. Additional DNA sequences controlling the specific binding of accessory repressor or activator proteins often lie within or immediately adjacent to the promoter (Pabo and Sauer, 1984)

The efficacy with which a particular promoter competes for and activates RNA polymerase will determine how often that promoter is used. The relative transcriptional effectiveness of different promoters is often referred to as "promoter strength"; a parameter which is determined by the intrinsic DNA sequence of the promoter itself and by accessory proteins which may modulate certain promoter–RNA polymerase interactions (von Hippel *et al.*, 1984).

The most abundant class of promoters is that recognised by the vegetative holoenzyme $E\sigma^{70}$. Within each promoter class two regions of strong DNA sequence homology are normally apparent. The two most highly conserved regions, with consensus sequences TTGACA and TATAAT, are centred approximately 33 and 10 bp, respectively, upstream of the mRNA start site and are usually termed the "-35" and "-10" boxes (Hawley and McClure, 1983). In each motif, three bases are highly conserved: the TTG at -35 and the TA---T near -10. However, the -35 and -10 hexamers are generally referred to as the principal features in each region, even though conserved sequence homologies exist on either side of the hexamers. The structural features of prokaryotic promoters are shown in Figure 1. Surprisingly, no wild type *E. coli* promoter has been identified that completely matches the consensus promoter, even in highly conserved hexamer regions, suggesting that promoter function is ordinarily optimised *in vivo* and not maximised (reviewed by McClure, 1985). The distance between the two regions can vary but for the majority of these promoters it is 16 to 18 bp, with 17 bp occurring most frequently. Similarly, the separation between the -10 region and the start site varies between 4 and 8 bp (reviewed by Travers, 1987a). However, it should be stressed that the sequence arrangements of other promoter classes recognised by different holoenzymes show significant deviation from this pattern. The promoters used by *E. coli* $E\sigma^{32}$ during heat shock response have similar spacing between the two conserved major regions but the -10 region sequences differ substantially from those recognised by $E\sigma^{70}$ (Cowing *et al.*, 1985). In addition, the RpoN protein, which acts as a sigma factor for genes involved in nitrogen metabolism, has two conserved regions separated by one turn of the DNA double helix at -25 and -16 (Ausubel, 1984; Benyon *et al.*, 1983). In contrast, promoters utilised by the T4 bacteriophage gp55 sigma factor contain only a single region of conserved sequence similar in position to the -10

box of *E. coli* $E\sigma^{-70}$ promoters (Elliot and Geiduschek, 1984). In spite of this diversity of promoter structure, one structural motif appears to be common to all classes of promoters in eukaryotes, eubacteria and archaebacteria – the TATA sequence, which is normally located at ~-10 in eubacterial promoters.

Since most genetically defined promoter mutations fall within the conserved regions at -10 and -35, it is likely that these sequences will define the strongest interactions between the promoter DNA and RNA polymerase. Recent experimental data has shown, however, that promoters which exactly match or fit very closely to the consensus sequence in these conserved regions, demonstrate only moderate activity with respect to the most active promoters (Deuschle *et al.*, 1986). It would seem then that additional sequence constraints determine promoter strength. Deuschle and coworkers (1986) were able to measure accurately the *in vivo* strength of well defined promoter sequences and to endeavour to interpret this sequence data based on functional information. A system was developed in which promoter activity was measured at the RNA level, with respect to an internal standard, independently of translation effects or gene dosage. In all, fourteen promoters were characterised and up to 75-fold differences in promoter strength were observed. The most efficient signals were seen using promoters from the coliphages T7 and T4. The activity of these promoters was very close to the strength of fully induced promoters of rRNA operons, which may approach the functional optimum of a single sequence. This contrasts the results obtained using a synthetic "consensus" promoter which generated less efficient signals. It appears then that optimal promoter strength can be achieved by alternate structures, strongly suggesting that information outside the "classically" defined promoter region contributes to promoter activity.

Functional dissection of *E. coli* promoters by Kammerer and coworkers (1986) led to the definition of a promoter as a 70 bp sequence containing a "core" of 35 bp (from positions +1 to -35), an upstream region of 15 bp and a downstream region of 20 bp. They could demonstrate that the downstream region of a promoter (P_{N25} of T5) contained information which could increase the strength of at least some promoters. In addition, the "core region" of the promoter contained the essential elements necessary for the early steps of the interaction between a promoter and RNA

polymerase including recognition of a nucleotide sequence by the enzyme and isomerisation of the initial complex into a state capable of initiating RNA synthesis. This in turn would suggest that the downstream regions contain signal functions required at a later stage of the overall process. The data of Kammerer *et al.*, (1986) show that the strength of a given promoter can be limited at different levels of the overall process and as a consequence, promoters of identical strength can exhibit different structures due to an alternate optimisation of their pathway to transcription initiation.

The actual location *per se* of a promoter on the DNA may play a regulatory role in initiation of transcription. Hawley and McClure (1983) compiled the sequences of 112 *E. coli* promoters and found that more than one third were situated close to other promoters. The geometry associated with these closely spaced promoters suggests three different classes. In the first class two or more promoters are orientated in the same direction and transcribe the same gene or operon as is the case with the *gal* and *lac* genes. The second class comprises the divergently orientated promoters which are capable of binding two RNA polymerase molecules with a common region and which transcribe separate genes or operons in opposite directions, for example the *araBAD* operon and *araC* gene. Thirdly, a less common class are the convergent promoters where RNA polymerase molecules actually oppose one another and transcribe both strands of the DNA over a common interval, for example in phage lambda promoters P_R and P_{RM}. In many of the aforementioned cases, the close proximity of the promoters has been shown to be an important component of the regulatory function displayed by these promoters (reviewed by McClure, 1985). It is possible that promoter juxtaposition could also result in an enhancement of transcription from one or both sites – possibly mediated through protein–protein interactions.

1.3 DNA Bending and Bendability

DNA–protein interactions often involve imposing a curvature on the DNA molecule along a precisely defined route on the surface of a protein complex, for example in the interaction between *E. coli* DNA gyrase and DNA (Kirchhausen *et al.*, 1985). This curvature is dependent on two factors; bendability and intrinsic bending of the DNA.

The bendability of a DNA molecule is its ability to adopt a preferred direction of curvature when

constrained in a tight loop on the surface of a protein or in the form of a small circle. This bending of the DNA double helix means that the external circumference of the circle or loop will be significantly greater than the internal circumference. As a consequence of this, the average widths of the major and minor grooves must be greater on the outside than on the inside such that particular sequences will adopt a preferred orientation with respect to the direction of curvature (Drew and Travers, 1984). Yet another parameter to be considered is that the direction of the helical axis must change. This change is correlated with the angle between the long axis of the base pair planes termed "roll". Rotational positioning of DNA about the histone octamer in nucleosomes appears to be determined by such sequence- dependent modulations of DNA structure (Drew and Travers, 1985). In nucleosomal DNA the variation in groove width is correlated with the appearance of particular short sequences such that, e.g., the trinucleotide AAA or TTT is more frequently found where the minor groove points inwards towards the histone octamer, whereas GGC or GCC prefers to be situated in an outward facing minor groove (reviewed by Travers and Klug. 1987). Any preferred direction of curvature of promoter DNA in the polymerase-promoter complex should subsequently be reflected in the preferred occurrence of sequences correlated with rotational orientation at particular positions in the DNA. This has indeed been shown to be the case by analysis of available promoter DNA whereby sequences associated with an outward facing minor groove are enriched at positions separated by ~10 bp – a value resembling the helical twist of DNA free in solution (reviewed by Travers, 1987a).

Intrinsic bending is the selection of a highly preferred direction of bending of the DNA double helix. This unusual property has been attributed to the presence of homopolymer (dA).(dT) stretches usually greater then 4 bp in length (Koo *et al.*, 1986). Such stretches tend to assume, in solution, a highly characteristic structure with a helical screw of 10 bp per turn and a relatively narrow minor groove (Nelson *et al.*, 1987). However, the intrinsic bending associated with such sequences is probably due to the stacking of immediately flanking base pairs on highly propeller-twisted (dA).(dT) base pairs which themselves are maintained in a relatively inflexible stack (reviewed by Travers, 1987a). Base stacking, in particular purine– purine stacking, is a very significant force in stabilising homopolymer structures. This is maximised by poly(dA).poly(dT)

through the high degree of propeller twist which rotates the bases around their longitudinal axis thereby facilitating more interactions between neighbouring bases (Nelson *et al.*, 1987). The propeller twist causes the major groove to point towards the 3'-end of its DNA strand which in turn pushes the N-6 atom of adenine, which lies in the major groove, towards the O-4 atom of thymine on the 3'-side of it. This allows the formation of a three-centred bifurcated H-bond consisting of a non-Watson-Crick H-bond diagonally across the major groove as well as the two normal Watson-Crick H-bonds. The distinctive features of poly(dA).poly(dT), that is additional H-bonds and good base stacking, are such as would confer a rigidity greater that that expected from base pairs with two H-bonds. Runs of about 5 bp of the homopolymer phased every helical repeat effect an overall bend to the DNA configuration (reviewed by Nelson *et al.*, 1987). The basis of this bending may be due, in part, to the contrast between the special geometry of the oligo(dA). oligo(dT) tract and abrupt change at the DNA sequence junctions, or a gradual change over the intervening region, or perhaps a combination of both (Nelson *et al.*, 1987). So, it would appear that the distribution of dA:dT blocks within promoters can define, in part, by the intrinsic bend, the potential pathway of DNA on the polymerase and also determine the limits (by their rigidity) of the regions which can be bent.

1.4 RNA polymerase–promoter interactions

Classically, the initiation of transcription by RNA polymerase has been regarded as essentially a two step process, described by the greatly oversimplified equation shown in Figure 2, which involves binding, isomerisation and promoter clearance. This model is based on the assumption that the first step in the process is the formation of a complex between RNA polymerase and the double stranded promoter DNA. This "closed" complex (RP_C) is in rapid equilibrium with free polymerase. A subsequent step which requires DNA strand separation results in the formation of a more stable "open" complex (RP_O) in which one to one and a half turns of the DNA double helix in the vicinity of the initiation site are no longer base paired (reviewed by Travers, 1987a).

Extensive investigations of this bipartite model have been carried out. Abortive initiation assays, template competition assays, filter binding assays and various "footprinting" techniques

have now been employed to quantitate a particular species of the pathway (reviewed by Buc, 1987).

The initial step in the formation of a closed complex is the location of the specific promoter site by RNA polymerase. RNA polymerase, like other DNA binding proteins, has a significant non-specific affinity for DNA (Strauss *et al.,* 1980). This means that when DNA and RNA polymerase are mixed in solution, an individual RNA polymerase molecule will be in close association with the DNA for a substantial proportion of the time. This effectively increases the target size for an initial interaction due to the number of non-specific sites on a given DNA molecule greatly exceeding the number of specific sites. Once bound non-specifically, a polymerase molecule could then migrate along the DNA in an essentially one-dimensional, diffusion-limited manner. Ricchetti *et al.,* (1988) have reported such a mechanism to facilitate promoter location. Four A1 promoters of phage T7 were polymerised head to tail and the length of the DNA upstream from the first promoter, or downstream from the last promoter, varied. The relative occupancy of the promoters was measured by quantitating the relative amounts of the corresponding transcripts. In this way they could demonstrate that increasing the length of the downstream region led to a positive gradient of occupancy from sites one to four, indicating a polarity in the process of promoter search. Since RNA polymerase and the promoter are asymmetric structures, it is possible that only those molecules bound in the correct orientation with respect to the direction of transcription are able to successfully initiate transcription. In addition, one has to consider that only those polymerase molecules that approach the promoter from downstream are successful. A possible explanation for these findings is that a potential barrier, at the DNA level, could be created by a set of base pairs upstream of the promoter preventing RNA polymerase from sliding through – this idea requires further investigation.

Another view of promoter location (or indeed the recognition of any sequence specific protein binding site on the DNA) is dependent on the observation that the external topography of the DNA double helix is not uniform but is a reflection of the internal relationships between the bases (Drew and Travers, 1984). This means that the protein must first find the appropriate external structure characteristic of its binding site, for example, local variations in groove width or overall curvature of the double helix. Irrespective of the precise structures comprising such a recognition site, promoter

location requires one-dimensional sliding along the DNA until the protein encounters a feature which is in close proximity to, or coincides with, a sequence- specific binding site perhaps allowing the protein to adopt a sequence-specific binding conformation (Drew and Travers, 1985b). Details of the recognition of RNA polymerase and promoter DNA are unknown but it is presumed to involve an interaction of the "helix-turn-helix" domain of the polymerase sigma subunit with either the -10 or -35 regions – strongly favoured is the major groove of the -35 region (reviewed by Travers, 1987a).

It is possible that a common multistep sequential process operates at all promoters during initiation of transcription – the rate-limiting steps being different and quite sequence-specific. Several competing mechanisms may be at work and, depending on the sequence of the promoter, one particular pathway may be more efficient than the others. Different sets of sequences could lead to the optimisation of different mechanisms for open complex formation. In this case, hybrid promoters from two optimal sequences would prove to be weaker – such context-dependent effects have been described by Bujard *et al.*, (1986). Two general features of the DNA structure have been implicated in these multistep processes.

The first of these is the spontaneous tendency of free promoter DNA to unwind slightly upstream of the Pribnow box. When certain promoters are inserted into negatively supercoiled DNA templates they become particularly sensitive to single stranded nucleases in this region, others do not. In one particular case, the loss of sensitivity to nuclease SI in this region, as seen when one base pair of the Pribnow box is mutated, correlates with the inability of the altered promoter to form an open complex with RNA polymerase (Drew *et al.*, 1985; Travers,1986b). In addition, the remarkable sensitivity of strong lactose promoters, not complexed with RNA polymerase, to the nuclease action of the OP/Cu^{+} complex, could also "signal" to the RNA polymerase the region where nucleation should be induced (Sigman *et al.*, 1985; Spassky and Sigman, 1985).

The second feature of the DNA structure, resulting from detailed studies performed at the *lac* UV5 promoter, suggest that the bendability of the whole promoter region could assist the nucleation process. Data from several laboratories (reviewed by Buc, 1987) indicate that upstream elements having an intrinsic curvature, increase the promoter strength when correctly positioned

with respect to the -35 region.

Another important point of note is that promoter strength does not rely solely on the ability of RNA polymerase to rapidly lead to formation of an open complex, but also on its ability to exit from this stable complex and to carry out the first translocation steps. This promoter clearance step has been found to be rate-limiting, and, in some cases, suggested to be the target of regulation (reviewed by Buc, 1987). Bujard *et al.*, (1986) have provided data which support the idea that improvement of the rate of open complex formation could impair the overall process, including the first translocation and the transition from an initiating to an elongating complex. They conclude that regions upstream and downstream of the promoter are involved in optimising the last steps.

1.5 Structural Transitions In Promoter DNA

Prior to transcription initiation, the product of the interaction of RNA polymerase with promoter DNA is the separation of the DNA strands over ~12 bp in the region of the transcription startpoint. The extent of the region over which this occurs has been estimated by chemical modification of the open complex on the *lac* UV5 promoter at 37°C. Bases in the region from ~-9 to +3 were shown to have the reactivity characteristic of single stranded DNA. This melted region does not include the TATA box which has already been shown in both the *tyrT* and *lac* UV5 promoters to be the region of the promoter at which unstacking is most readily induced by negative superhelical stress or thermal changes (reviewed by Travers, 1987a). This suggests that the TATA region serves as the nucleation point for untwisting and that the untwisted region is subsequently propagated to the transcription start site (Figure 3). For this to happen would require that untwisting occurs prior to open complex formation and that any mutation which blocks untwisting may trap the polymerase-promoter complex at an early step in the initiation pathway. A$_{-8}$T to A transversion converting the *tyrT* -10 region from TATGAT to TATGAA resulted in at least a 50-fold decrease in the rate of initiation both *in vivo* (Berman and Landy, 1979) and *in vitro* (Travers *et al.*, 1983). It is likely that the T to A transversion blocks polymerase-mediated chain extension and the polymerase promoter complex remains in the closed state. This block may be due to the lack of a crucial contact with the most highly conserved position of the -10 region or, alternatively, from failure to untwist the DNA in

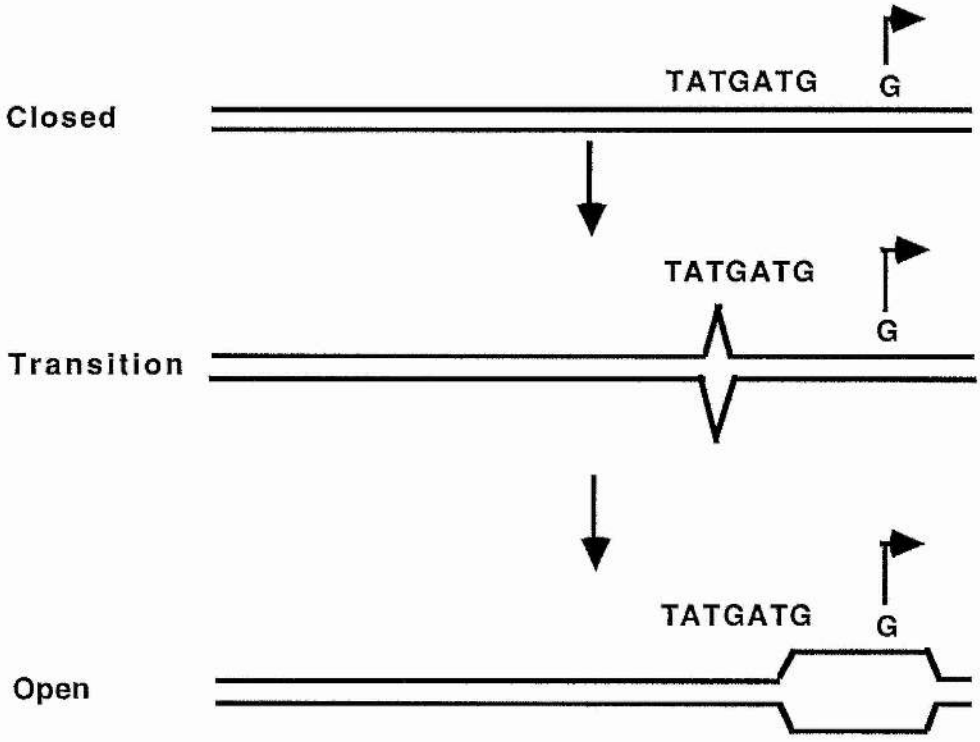


Figure 3. Nucleation of unwinding at the *tyrT* promoter
Arrow depicts startpoint and direction of transcription. The sequence of the TATA region is indicated.

(redrawn from Travers, 1987)

this region. It is possible that untwisting of the DNA to form additional or altered contacts with the polymerase which in turn triggers a conformational change resulting in polymerase extension (reviewed by Travers, 1987a). Since untwisting by itself does not necessarily involve strand separation, at what stage of the initiation process does strand separation first become detectable? To address this point, studies on cytosine reactivity to DMS were carried out (Spassky *et al.*, 1985b). Results from these studies showed that cytosine residues close to the initiation site were methylated at 37°C (that is, in the open complex) but not at 17°C (in the intermediate complex where promoter DNA is double stranded) suggesting that strand separation occurs between 17°C and 37°C on formation of the open complex. In both the intermediate and open complexes the distal limit of protection by RNA polymerase was positioned at +20. In a similar set of experiments, Straney and Crothers (1985) separated two "open" transcription complexes at the *lac* UV5 promoter (by native gel electrophoresis) which were interconvertible, one predominating at 16°C, the other at higher temperatures. In both cases DNA/RNApolymerase complexes were protected from DNase I digestion to position +20 and both were transcriptionally active *in vitro*. The relationship of the low temperature "open" complex of Straney and Crothers and the 17°C complex of Spassky *et al.*, is unclear at the present time – it is possible that it represents an additional type of complex.

Topological unwinding during closed complex formation at the *lac* UV5 promoter might take the form of "wrapping" the DNA in a tight left-handed supercoil about the polymerase molecule. At the other extreme, the unwinding of 1.3 turns of DNA could in theory be accommodated by untwisting, distributed either across the whole polymerase binding site or locally in the -10 region. For either of these alternatives to occur would require a reduction in the local twist to 18° in the case of the former, or in the formation of a 6–8 bp stretch of left-handed DNA in the latter case (reviewed by Travers, 1987a).

1.6 Models for Transcription Initiation in Prokaryotes

The pathway to open complex formation involves structural transitions in both the DNA template and RNA polymerase occurring, to a certain extent, independently of any direct interaction between the enzyme and promoter DNA (Travers, 1987)a. Kinetic and thermodynamic data

suggest that at least two mechanistically distinguishable pathways exist leading to open complex formation. These pathways are exemplified by the *lac* UV5 and lambda P_R promoters, respectively (reviewed by Travers, 1987a). In the case of the *lac* UV5 promoter, RNA polymerase would interact with partially unwound promoter DNA in the -10 region to form an initial complex. This state is then followed by an intermediate complex where the DNA remains in the unwound state but with no detectable strand separation. At this stage RNA polymerase covers ~60–70 bp of the promoter from positions +20 to -40 or -50. The final step in initiation involves strand separation in the region of the start site, that is there has been migration of the untwisted region from -10 with the RNA polymerase remaining extended – the polymerase now becomes competent to initiate RNA synthesis. In the lambda P_R promoter a crucial difference is that the initial interaction of RNA polymerase is with unstressed DNA. In the initial complex RNA polymerase remains compact with no unwinding of the DNA. This is followed by a concerted transformation to an intermediate complex whereby DNA unwinding and RNA polymerase extension take place resulting in the formation of an equivalent complex to that observed on the *lac* UV5 pathway. Transition to the open complex would then be common in both pathways (Travers, 1987a).

To what extent the observed differences leading to open complex formation on different promoters relate to an intrinsic variability in promoter function, or simply reflect assay conditions which are not fundamentally comparable, remains to be elucidated. However, if as suggested by Travers (1987a) that the structure of free RNA polymerase is a determinant of the pathway leading to open complex formation, then defined variation in reaction conditions should uncover alternative pathways to open complex formation from the same promoter.

1.7 Regulation of Transcription in *E. coli*

The rate of transcription initiation at the majority of bacterial promoters is regulated in concert with the cellular requirement for the expression of particular genes. In general, this regulation occurs at two levels, the first of which is a direct modulation of the specificity and overall affinity of RNA polymerase for a particular class of promoter mediated by sigma factor substitution and by interaction with regulatory nucleotides. The second level of control is governed by regulatory

proteins which interact primarily with additional sequence- specific DNA binding sites, usually in close proximity to the polymerase binding site, which can act either positively or negatively to control transcription on particular transcription units (Hochschild *et al.*, 1983).

Repressors of transcription initiation bind to specific DNA sites (operators) in the vicinity of the promoter. These operator sequences have been found in regions as far upstream as -60 and downstream as far as +12. and have also been shown to overlap physically the promoter in their DNA sequence (reviewed by McClure, 1985). In addition, there appears to be a wide spectrum of repressor-operator affinities ranging from a dissociation constant of ~10^{-11} M for *E. coli lac* (Barklay and Bourgeois, 1978) to 10^{-7} M for the *Arc* repressor of phage P22 (A.D. Vershon and R.T. Sauer, unpublished observations). Thus the most tightly binding repressors could result in total shutoff of transcription initiation (on a physiological time scale). More weakly binding repressors, on the other hand, could contribute simply to a down modulation of transcription – this might be advantageous to the cell in that reduced expression from a particular promoter could then be affected by other cellular or phage products.

1.7.1. The Lactose Operon

One of the best studied models for negative control of gene expression is the *lac* operon of *E. coli* (Jacob and Monod, 1961: Miller and Reznikoff, 1980). The expression of *lac* appears to be dependent on two factors – the presence in cells of an inducer and cAMP (reviewed by Beckwith, 1988). In this system, the lactose repressor protein (the product of the *lacI* gene) binds as a tetramer to the operator DNA, physically blocking RNA polymerase/promoter DNA interactions and hence transcription initiation of the lactose metabolic enzymes. Induction of transcription by binding of inducer sugars to the repressor causes a conformational change in the repressor resulting in a lowered operator DNA binding affinity. RNA polymerase can now bind to the *lac* promoter region and transcription of the *lac* gene cluster occurs.

From studies involving chemical and enzymatic probing techniques, proteolysis, as well as mutagenesis of both the repressor and its operator site (Straney and Crothers, 1987 and references therein), a great deal is now known about the structure of the *lac* repressor, and how it

binds DNA. Kinetic and equilibrium studies of repressor-operator interactions have also been carried out (Fried and Crothers, 1981) and a detailed picture of the binding of *lac* repressor to its operator site is now emerging. The main features resulting from these studies are a high equilibrium binding constant (K_{eq} ~10^{-13} M; Winter and von Hippel, 1981) and two regions of the protein which contribute to specific binding. The first of these regions, the N-terminal region, contains a putative "helix-turn-helix" motif (Weber *et al.*, 1982) and the remaining "core" region which is required for binding in a specific inducer-sensitive manner (Manley *et al.*, 1983).

Experiments to determine the stoichiometry of operator sequences/repressor tetramers suggest that the repressor protein occupies two operator sites on the DNA simultaneously (reviewed by Hsieh *et al.*, 1987). Fried and Crothers (1984) have also demonstrated, by electrophoretic analysis of repressor/DNA complexes in direct transfer experiments of repressor protein between operator and non-specific DNA, results consistent with the binding of repressor at two operator sites. In addition, Barbier *et al.*, (1984) found two non-specific DNA binding sites on the *lac* repressor by cross-linking small DNA fragments to the protein.

In the *lac* operon two regions of DNA sequence (pseudooperators), in addition to the primary operator, have been found which bind the repressor, albeit with a lower affinity to that of the primary sequence (reviewed by Hsieh *et al.*, 1987). The specific function of these pseudooperators is unclear at present but Winter and von Hippel (1981) have pointed out increased binding affinities when pseudooperators accompany the primary operator sequence on DNA fragments. A role for these sequences in stabilising binding via an intramolecular ternary complex of operator-repressor-pseudooperator, forming a looped double stranded structure, has recently been suggested (Whitson *et al.*, 1986). Hsieh and coworkers (1987) examined the influence of sequence and distance between two operators on *lac* repressor binding and obtained results consistent with the formation of operator-repressor-pseudooperator ternary complex to generate a looped structure – complex formation was not observed when the two operators were too distant (>400 bp) or too proximal (<100 bp). Ternary complex formation was found to be dependent on the energy required for DNA binding to form the complex, phasing of the sequences along the DNA , local concentration of the second site and the relative affinity of the repressor for the DNA

sequence at the second site. The mechanism by which the protein searches for the second binding site once it is bound at the first site has not been determined but it will clearly be of great interest.

The three-dimensional structure of DNA under the influence of negative supercoiling has been found to be critical in the modulation of gene expression in both prokaryotes and eukaryotes (reviewed by Gellert, 1981). Whitson and coworkers (1987), using a variety of operator/pseudooperator-containing, negatively supercoiled plasmid DNAs, examined the dissociation of the repressor-operator complex as a function of the sequence context, orientation and spacing. Their finding s were consistent with formation of an intramolecular complex with a looped DNA segment stabilised by the combination of increased local concentration of binding sites and torsional stress on the DNA which favour binding in supercoiled DNA. Comparison of these results with observations in more complex systems, where effects of enhancer sequences on transcription occur at large distances from the transcription initiation site, should prove most useful.

Straney and Crothers (1987) demonstrated that the *lac* repressor can bind to the L8UV5*lac* promoter at the same time as RNA polymerase and that the contacts with the DNA in such a ternary complex were different from those made by either protein alone. Another surprising finding, which contrasts with most models of prokaryotic repression for *lac* repressor as well as others (Majors, 1975; Ptashne *et al.*, 1980), was that the repressor exerted its effect at the isomerisation step and not at the initial binding step. It should be stressed however that just as individual activators such as CAP and lambda repressor have been shown to act at different steps (Hawley and McClure, 1982) there is no reason why different repressors cannot repress at different steps. It is likely that repressors inhibit RNA synthesis by a variety of mechanisms and that it cannot be presumed that the overlap of polymerase and repressor binding sites will simply exclude polymerase binding in the presence of repressor.

Yet another major finding of Straney and Crothers (1987) is that the *lac* repressor can act as a transient gene activator. They could show that at low polymerase concentrations the presence of *lac* repressor (and then the gratuitous inducer IPTG) increased the rate of the first round of productive transcription thereby allowing the system to respond rapidly to the release of repression

and making the repressor a more efficient regulator in the cell. A somewhat analogous situation has been described in eukaryotes (Gilmour and Lis, 1986) in which RNA polymerase II has been found to bind to a heat shock promoter in the absence of activating factors – transcription does not occur but polymerase is stably bound waiting to be activated.

The catabolite gene activator protein (CAP), or cAMP receptor protein (CRP), together with its cofactor cAMP, regulates the expression of several catabolite-sensitive operons in *E. coli* (Adhya and Garges, 1982). This regulation is brought about by the specific binding of CAP-cAMP to sites within catabolite sensitive promoters. The sequence of these sites is somewhat variable, as is their location relative to the transcription start site among different promoters, suggesting that the influence of the CAP/DNA complex on the RNA polymerase open promoter complex is variable (DeGrazia *et al.*, 1985). However, the ability of CAP-cAMP to act as a repressor of transcription for some promoters (Aiba, 1983) must also be taken into account when formulating any mechanism of CAP-cAMPs enhancement of transcription.

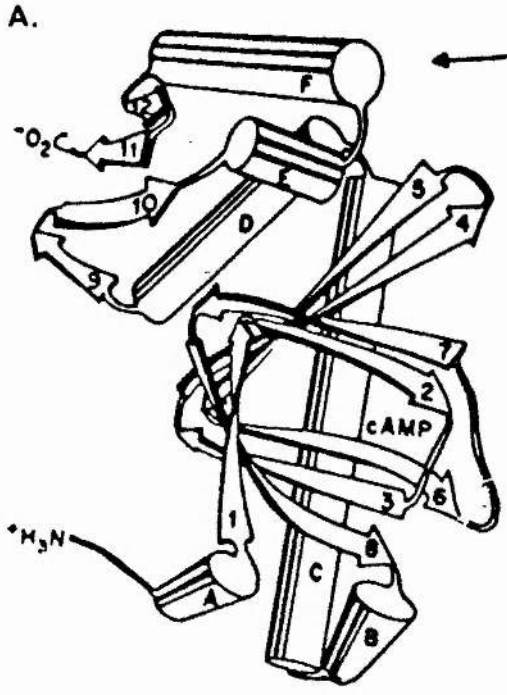
An initial model, put forward to explain the mechanism of activation of CAP-cAMP, proposed that CAP-cAMP, when bound to the DNA, interacts directly with RNA polymerase (Gilbert, 1976). A more recent model to elucidate CAP activation, indicates that CAP-cAMP stimulates mRNA expression of the *gal* and *lac* operons by blocking RNA polymerase from binding to a strong binding but weak transcription site, thereby allowing RNA polymerase to bind a nearby productive initiation site (reviewed by DeGrazia *et al.*, 1985). DeGrazia and coworkers (1985) examined the effect of CAP on the thermal stability of DNA and results from their studies indicate that site-specific CAP-cAMP binding stabilises the promoter site. This physical behaviour is consistent with CAP functioning as a repressor by blocking RNA polymerase interaction at strong initiation sites, and as an activator by displacing RNA polymerase from poor initiation regions to good initiation regions. It is also possible that the influence of CAP-cAMP on RNA polymerase may involve other interactions as a consequence of DNA bending and/or protein-protein contacts – the idea of a transmitted change in the DNA has still to be established.

The CAP protein contains 209 amino acid residues and can be divided into two distinct domains (Figure 4). The C-terminal domain binds DNA while the N-terminal domain binds cAMP and

Figure 4. The CAP monomer and the proposed interaction between the C-terminal domain of CAP and its binding site in the *lac* operon

A digramatic representation of the CAP monomer is shown in **A**. The position of each domain is indicated, as well as their primary functions (McKay *et al.*, 1982). The proposed interaction between the C-terminal domains of a CAP dimer and the CAP binding site in the *lac* operon is outlined in **B**. Filled dots indicate phosphate groups in the DNA backbone contacted by the protein (Steitz and Weber, 1983).

A.



C TERMINAL DOMAIN

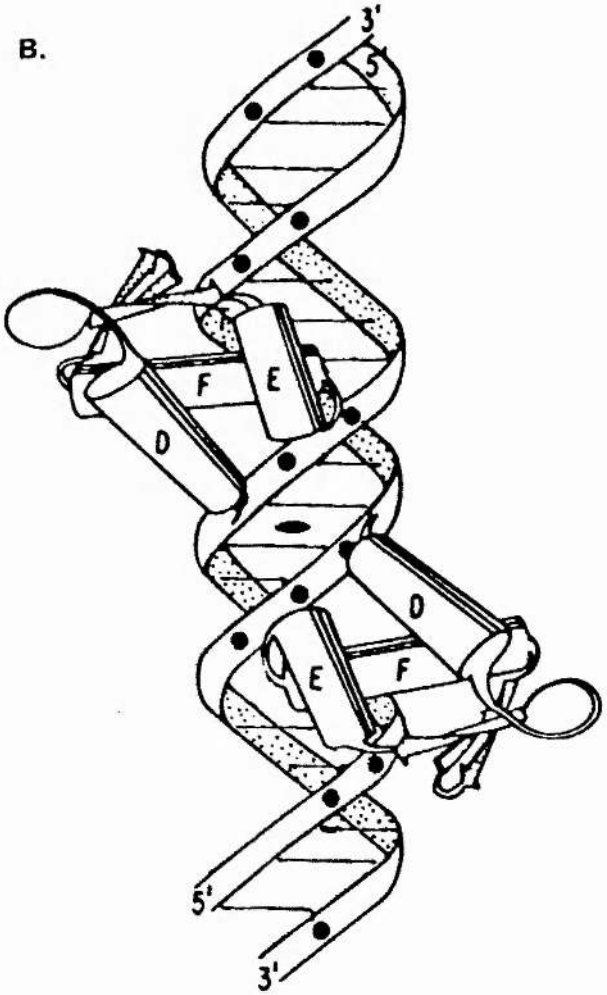
DNA BINDING

N TERMINAL DOMAIN

DIMER CONTACTS

cAMP BINDING

B.



provides a surface for most of the CAP dimer contacts. In solution, CAP forms a stable dimer and it is this dimer which is the active DNA binding species (reviewed by Pabo and Sauer, 1984). The N-terminal domain comprises 135 amino acids which consist of two short helices (A and B), an eight-stranded antiparallel ß-roll and a long alpha-helix (C). The C-terminal domain includes amino acid residues 136-209 and contains three alpha helices (D, E, and F) and two pairs of short antiparallel ß-strands as shown in Figure 4A. The CAP dimer contacts, which are in the N-terminal domain, involve a pairing of the C-helices as well as some additional contacts between the C-helix of one subunit and the ß-roll of the other subunit. Part of this dimer interface is also occupied by the cAMP which is completely buried within the interior of the N-terminal domain but it forms H-bonds which span the dimer interface (McKay *et al.*, 1982).

The sequence of the CAP binding site in the *lac* promoter exhibits marked two-fold symmetry which is mirrored by the pattern of CAP-induced protection against chemical attack indicating that the dimeric CAP molecule binds symmetrically (reviewed by Busby, 1986). Current data suggest that CAP binds to right-handed B-DNA using the N-terminal portions of its F helices to contact the major groove. Each of the DNA binding regions in the two CAP subunits appear to bind to one of the symmetric sequence elements of the CAP binding site and based on these findings a plausible molecular model has been proposed for CAP-DNA interactions (Weber and Steitz, 1984). In this model, residues in the "helix-turn-helix" motif on the two subunits bind to bases that appear in two adjacent major grooves (Figure 4.B). The centre of symmetry in the DNA binding site corresponds approximately to the centre of symmetry in the protein. Mutational analysis confirms that DNA binding is determined by the correct juxtaposition of two "helix-turn-helix" motifs and that the specificity of binding depends on the exact amino acid sequence in this zone.

1.7.2 The Tryptophan Operon

The synthesis of L-tryptophan in *E. coli* is regulated in part by repression of the biosynthetic operon (*trpEDCBA*) for L-tryptophan, and *aroH* whose gene product catalyses the first step in amino acid biosynthesis (Zurawski *et al.*, 1981; Kelley *et al.*, 1982). Repression of these operons is regulated by a straight forward negative feedback loop -- as the concentration of L-tryptophan

increases two molecules of L-tryptophan bind to the inactive *trp* aporepressor (a dimer of two identical 107 amino acid subunits) forming the active *trp* repressor which binds as a dimer in a sequence-specific manner to operator DNA of these operons (Joachimiak *et al.*, 1983a). In the same way, the *trp* repressor is also autoregulatory since it decreases transcription of the *trpR* gene which encodes the *trp* aporepressor polypeptide.

In addition to regulation by repression, a second control mechanism – transcription attenuation – is used to regulate the structural genes of the *trp* operon. The transcription start site in the *trp* operon is separated from *trpE* (the first major structural gene) by a leader region of ~160 bp (Yanofsky *et al.*, 1984). This region is responsible for attenuation control, which provides the option for transcription termination. This leader RNA segment contains a coding region for a 14 amino acid peptide, the most significant feature of which is the presence of two Trp codons, in tandem, near its distal end (Das *et al.*, 1983). The inability to translate these Trp codons results in the relief of termination at the attenuator, leading to continuation of transcription of the structural genes. Transcript segments that can form a stable hairpin loop, followed by a run of Us, are located just downstream of the leader coding region (Kuroda and Yanofsky, 1984). When transcription terminates in the *trp* operon leader region, the transcript ends with a run of Us. There is a growing body of evidence (reviewed by Yanofsky, 1988) to suggest that the transcribing RNA polymerase molecule recognises this structure as a Rho-independent termination signal.

The leader transcript can form two additional hairpin structures, the pause and antiterminator structures, each of which plays a crucial role in attenuation. The promoter proximal hairpin causes transcription pausing which may be responsible for synchronising transcription of the leader region with translation of the peptide coding region (Kuroda and Yanofsky, 1984; Winkler and Yanofsky, 1981). It may be that polymerase pausing introduces a transcription delay allowing time for a ribosome to load onto the leader transcript and begin synthesis of the leader peptide. The moving ribosome may in turn disrupt the pause complex thus restoring active transcription. When the translating ribosome completes synthesis of the leader peptide, the terminator structure presumably forms and transcription termination occurs. If however the translating ribosome stalls at either of the Trp codons (as a consequence of inadequate levels of charged $tRNA^{Trp}$) then the

antiterminator hairpin forms (Kuroda and Yanofsky, 1984). Thus the availability of charged $tRNA^{Trp}$ determines which RNA hairpin will form in the leader transcript which in turn controls transcription termination in the *trp* operon.

The determination of the three-dimensional structure of DNA binding proteins (Pabo and Lewis, 1982: McKay and Steitz, 1981) has provided a general framework with which to understand the interactions of proteins with regulatory sequences in DNA. Of special interest is the structural basis of *trp* repressor's function because its regulatory role is modulated by binding of the small molecule, L-tryptophan. Altered patterns of crystallisation (Joachimiak *et al.*, 1983b) and proteolysis (Tsapakos *et al.*, 1985) have suggested that the binding of L-tryptophan produces a conformational change that may be responsible for the activation of the aporepressor to the repressor.

The crystal structure of the *E. coli trp* repressor has been solved to atomic resolution by Schevitz and coworkers (1985). They deduced that 77 of the 107 amino acid residues form six alpha helices connected by short runs and all but one of the helices of one subunit make contact with some part of the opposing subunit. This makes the dyad-related subunits so completely interlinked that it would be difficult to disengage them without significantly altering their tertiary structures. The crystal structure described by Schevitz *et al.*, (1985) is of the active repressor and contains two equivalent molecules of L-tryptophan per dimer. Both subunits of the repressor contribute to each of the two symmetrical tryptophan binding sites which are located in a natural cavity near the surface of the *trp* repressor. The DNA surface which interacts with the repressor was indicated by several observations. First, it contains as a prominent surface feature the "helix-turn helix" motif that has both sequence and structural homology to the DNA binding proteins cro (Anderson *et al.*, 1981), lambda repressor (Pabo and Lewis, 1982) and CAP (McKay and Steitz, 1981). This presumed interactive surface is separated from its dyad-related counterpart by one full turn of the DNA double helix. Second, this interactive surface holds the pocket containing the L-tryptophan site. Figure 5A shows the overall appearance of the modelled repressor/operator complex. In addition to the interaction of the "helix-turn-helix" elements (helices D and E) with the major groove, there is a more central contact surface in which the DNA backbones on either side of the minor groove approach the repressor surface close to the common dyad of the complex, but,

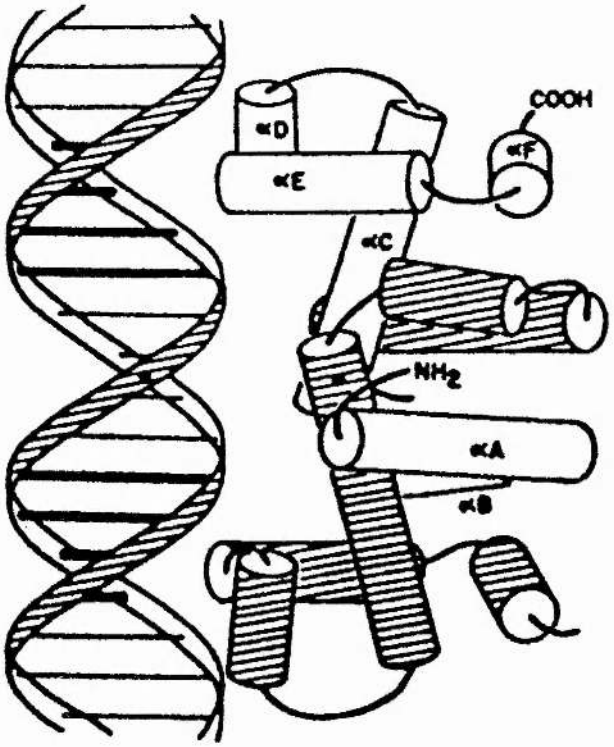
Figure 5. Models for the interaction of trp repressor with operator DNA

A. Schematic representation of the docked Trp repressor dimer with B-form DNA. The ∝helices of trp are represented by cylinders and the operator DNA is shown as a ladder with the rungs representing base pairs. The 4 bp in each operator half site critical for binding are indicated by the bold rungs; dots in the centre of each molecule define dyad axes (Bass *et al.*, 1988).

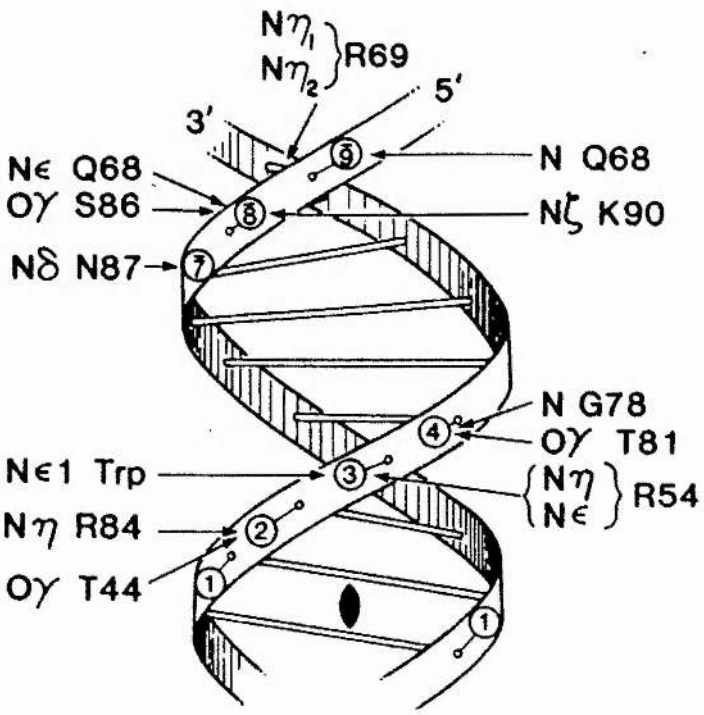
B. Schematic view of the upper half of trp operator as "viewed" by trp repressor showing phosphates (circles with numbers) and repressor's functional groups (shown by arrows) that form direct H-bonds (Otwinowski *et al.*, 1988).

C. Schematic view of the amino acid positions (circled amino acid numbers) in the "helix-turn-helix" substructure that make direct H-bonds to the operator (shown by arrows). The Figure stresses that the second helix, or E-helix, of the bi-helical motif is "pointed into" the major groove with its helical axis almost perpendicular to the global DNA axis. The E-helix does not lie in the major groove with its axis parallel to the groove (Otwinowski *et al.*, 1988).

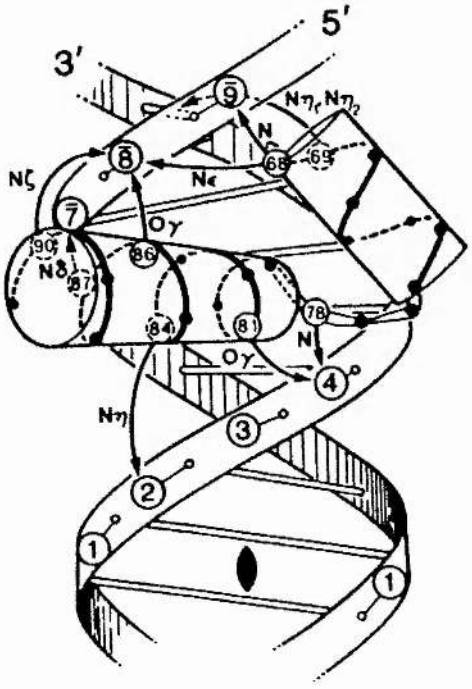
A


COOH
αD
αF
αE
αC
NH2
αA
αB

B


Nη1
Nη2
R69
3′
5′
Nε Q68
Oγ S86
Nδ N87
N Q68
Nζ K90
N G78
Oγ T81
Nε1 Trp
Nη
Nε
R54
Nη R84
Oγ T44

C


5′
3′
Nη1 Nη2
N
Nζ
Nε
Oγ
Nδ
Nη

the bases in the minor groove do not appear to be in contact with the repressor (Schevitz *et al.*, 1985). It would seem that the "helix-turn-helix" motif of the *trp* repressor is related to its operator in a similar manner to that of lambda repressor (Lewis *et al.*, 1983) rather than that of cro (Ohlendorf *et al.*, 1982). The much shorter D helix of *trp* repressor fits across the major groove between the sugar-phosphate backbone.

The roles of bound tryptophan appear to be at least two-fold (Schevitz *et al.*, 1985). Since L-tryptophan is wedged against the side of helix E, the ligand stabilises the functionally critical orientation of the "helix-turn-helix" substructure that is responsible for the operator-specific contacts. The second role is more local and results from interactions of the bound tryptophan's carboxyl and amino groups with functional groups of the repressor near its DNA binding surface.

Mutant *trp* repressors with new DNA binding specificities have been identified (Bass *et al.*, 1988). Thirty-six mutant repressors (each with single amino acid changes) were analysed for their ability to bind to each member of a set of 28 different operators closely related to the *trp* operator consensus sequence. Mutations that create new binding specificities create new contacts between the repressor and operator. Or, alternatively, they could eliminate negative interactions between the wild type amino acid and wild type operator. *In vivo* interaction of wild type Trp repressor with the mutant operators revealed that four adjacent base pairs of a half-site are critical for binding (Bass *et al.*, 1987). The new specificities indicate that the hydroxyl group of Thr^{81} makes a specific contact with one of the four critical base pairs and the methyl group of Thr^{81} determines the specificity at a second critical base pair. Unlike lambda repressor, lambda Cro protein and coliphage 434 repressor, the Trp repressor does not appear to use the first two hydrophobic amino acids of its "recognition alpha-helix", Ile^{79} and Ala^{80}, to make sequence specific contacts with the DNA. Instead, specific contacts between Trp repressor and its operator occur via amino acid side chains from both the first alpha-helix (D) and the second alpha-helix (E) of the "helix-turn-helix" substructure (Bass *et al.*, 1988).

Zhang and coworkers (1987) have reported the crystal structure determination and refinement of *trp* aporepressor to a nominal resolution of 1.8Å and have contrasted this with the partially refined model of the trigonal crystal form of the repressor described above (Schevitz *et al.*, 1985). Their

findings show that binding of L- tryptophan activates the repressor dimer 1000-fold by moving two symmetrically-disposed flexible bihelical motifs. These flexible "DNA reading heads" flank a highly inflexible core domain formed by an unusual arrangement of interlocking alpha-helices from both repressor subunits. It would appear that the appropriate shape and orientation of the repressor's reading head is a crucial feature for operator-specific binding and that it is this conformation that is modulated in the aporepressor–repressor transition. The structural changes in the flexible reading heads involve two alternative arrangements of buried residues in the hydrophobic brace (a cluster of hydrophobic residues in the repressor). In the active conformation of the repressor, the indole ring of bound tryptophan is a fundamental part of the brace that stabilises what would be an otherwise unstable conformation. However, in the inactive conformation of the aporepressor, an alternative arrangement of buried hydrophobic side chains forms a stable compact brace in the absence of tryptophan. This conformation of the DNA reading heads in the aporepressor produces a surface non-complementary to the recognition surface of the operator.

The crystal structure of the *trp* repressor/operator complex, refined to 2.4Å resolution, has recently been determined (Otwinowski *et al.*, 1988). The co-crystal shows an extensive contact surface between the Tpr repressor and its operator which includes 24 direct and six solvent-mediated H-bonds to the phosphate groups of the DNA. These water-mediated H-bonds to six bases have been demonstrated, *in vivo* , to be important in specifying the operator as the target sequence for the *trp* repressor (Bass *et al.*, 1987). There are 14 direct H-bonds between the half-repressor/half-operator and all but two of these are involved in the unesterified oxygens of the six phosphate groups. The exceptions are the two direct H-bonds between Arg^{69} and the functional groups of G-9 (Otwinowski *et al.*, 1988). Of the 14 direct contacts to the half-operator, four are made by amino acid residues which do not form part of the "helix-turn-helix" motif implying that specificity in this instance is not due to direct H-bonded contact of the major groove's polar groups by the H-bonding groups of the protein (von Hippel and Berg, 1986). It is believed that in the *trp* repressor/operator system, a primary mode of sequence-specific recognition is that the operator sequence enables the DNA to adopt a characteristic conformation that will make 24 direct H-bonds to the protein through its phosphate groups and increase the solvent- excluded surface

(Otwinowski *et al.*, 1988). Since four of the H-bonded phosphates in the promoter accept more that one H-bond and two of these are caged by three direct H-bonds, means that the DNA backbone must adopt a very precise geometry. This in turn permits the formation of a stable interface. Figure 5.B and C show residues involved in direct contacts between *trp* repressor and operator. Specificity in the *trp* repressor/operator system appears to be independent of a simple complementarity between the repressor amino acid side chains and operator bases. Rather, it is dependent on a combination of solvent mediated interactions with crucial bases and as yet poorly understood principles governing the way nucleotide sequence restrains the conformation of B-form DNA (Otwinowski *et al.*, 1988).

1.7.3 The LexA Repressor

E. coli has developed a system of ~20 genes involved mainly in the repair of DNA damage and enhanced mutagenesis known as the "SOS system". The "SOS" genes are expressed upon induction of the SOS machinery by chemical and/or physical carcinogens (reviewed by Little and Mount, 1982). These DNA modifications *in vivo* lead to the production of an "SOS signal" which in turn activates the existing cellular pool of RecA protein which catalyses the specific cleavage of the LexA protein, the common repressor of all the SOS genes including the *recA* and *lexA* genes.

LexA can essentially be viewed as a two domain protein with a flexible hinge region linking the two structural domains (Little and Hill, 1985). At alkaline pH, LexA undergoes a self-cleavage reaction at the Ala-84–Gly-85 bond within the hinge region, generating the same two protein fragments as in the RecA-mediated cleavage reaction. The two fragments have been purified and the N-terminal fragment shown to harbour the DNA binding site of the protein (Hurstel *et al.*, 1986). It has also been shown that the C-terminal domain dimerises in the same way as the entire protein (S. Hurstel, M. Granger-Schnarr and M. Schnarr, unpublished observations). No high resolution structural information is available as yet for LexA, and it remains to be determined whether the protein contains the common "helix-turn-helix" motif found in many other regulatory proteins (reviewed by Pabo and Sauer, 1984). However, there is some evidence that the N-terminal fragment of LexA contains a substantial amount of alpha-helical structure(Hurstel *et al.*, 1986). To

compare the overall positioning of LexA on the DNA double helix with that of other repressors with a known three-dimensional structure, Hurstel and coworkers (1988) carried out hydroxyl radical footprinting and ethylation interference experiments to determine the backbone contacts of the entire LexA repressor and its N-terminal fragment, with *recA* operator DNA. Both these methods are complementary in that hydroxyl radicals primarily attack the deoxyribose residues (Tullius and Dombroski, 1986) whereas ethylnitrosourea reacts mainly with phosphate groups of the DNA backbone (Siebenlist and Gilbert, 1980), enabling non-identical but closely related parts of the DNA backbone to be probed. These techniques revealed essentially the same contacts between both proteins and one side of the DNA double helix (assuming that the DNA stays in the normal B-conformation). This result was somewhat unexpected since methylation protection studies (Hurstel *et al.*, 1986) suggested a somewhat twisted recognition surface. In an attempt to reconcile the two sets of data (that is, backbone and major groove contacts), Hurstel *et al.*, (1988) propose a model whereby the bulk of the protein body sticks to one face of the DNA helix through contacts with the DNA backbone, and a protruding "reading head" probes the bottom of the major groove (where recognition takes place) slightly sideways. The backbone contacts are symmetrically disposed with respect to the centre of the operator indicating that the operator binds two LexA monomers. A comparable, but non-identical, relative disposition between backbone contacts and recognised base pairs has been demonstrated in the case of lambda repressor (Ptashne *et al.*, 1980; Tullius and Dombroski, 1986). Obviously this similarity is insufficient to conclude that LexA contains the structural "helix-turn helix" motif, but it is certainly an argument in favour of this possibility.

Another slight difference between the backbone contacts of LexA and lambda repressor concerns the relative "weight" of the protected regions in that LexA has a greater number of contacts in the outer part of each operator half site whereas lambda repressor shows the opposite behaviour (Hurstel *et al.*, 1988). One might speculate at this stage that if LexA contained the common "helix-turn-helix" substructure, the roughly inverted backbone contact pattern could arise from an ~180° rotation of the "helix-turn-helix" motif upon insertion into the major groove – such a rotation has recently been suggested by Boelens *et al.*, (1987), in the case of the *lac* repressor.

1.7.4 The Galactose Operon

Three of the enzymes involved in the galactose metabolism pathway in *E. coli* are encoded by the *galK, galI* and *galE* genes which constitute the galactose (*gal*) operon. The induction of the *gal* genes by galactose or fucose is inhibited by the presence of glucose in the growth medium (Dreyfus *et al.*, 1985; Monod, 1947). The *gal* operon contains two promoters, P_{G1} and P_{G2}, the first of which requires cAMP and CRP, and the second which is normally inhibited by cAMP and CRP (this promoter functions in cells devoid of cAMP or functional CRP, under which conditions the original promoter becomes defective) (reviewed by Adhya, 1988). The presence of two promoters and their opposing controls, mediated by cAMP, ensures that the enzymes are available in both normal physiological conditions and catabolic repression.

The two *gal* promoters overlap slightly residing within the -52 to +19 region. Both promoters have their own -35 and -10 consensus sequence regions and it is likely that RNA polymerase has the option to form a closed complex and then an open complex at P_{G1} or P_{G2} in a mutually exclusive way (reviewed by Adhya, 1988). It is believed that binding of a single molecule of cAMP.CRP at the -37 region inhibits P_{G2} activity by steric hindrance of the corresponding closed complex formation by RNA polymerase at the -40 to +10 region (McClure, 1980). The mechanism of cAMP.CRP activation of P_{G1} remains to be elucidated.

Regulation of the two *gal* promoters is mediated through the Gal repressor, the product of the unlinked *galR* gene (Buttin, 1963). Inactivation of the repressor by the addition of galactose or fucose derepresses the operon. Genetic and biochemical studies have shown that the *gal* operon contains two operator loci, O_E and O_I, which show strong sequence homology and are located at -60 and +55 respectively (reviewed by Adhya, 1988). Each homologous sequence contains a dyad symmetry. Mutation at either site may cause derepression of P_{G1} and P_{G2}. The Gal repressor has been shown to bind specifically to the two operators (Majumdar and Adhya, 1984; Adhya and Majumdar, 1987). Active Gal repressor is a dimer in solution and as such each symmetrical sequence in O_E and O_I is probably recognised by the dimer (Nakanishi *et al.*, 1973). The N-terminal domain of Gal repressor binds DNA and makes specific contacts, through the second alpha helix of

a "helix-turn-helix" motif, in the major groove of the DNA; the C-terminal domain is responsible for dimerisation and inducer contact (von Wilcken-Bergmann *et al.*, 1982). Binding of the inducer to the C-terminal domain inactivates specific DNA binding ability of the N-terminal domain by allosteric alteration of the protein.

A striking feature of the organisation of the *gal* operon is the existence of the two essential operators separated by more than 100 bp which flank, but do notoverlap, the promoters (reviewed by Adhya, 1988). Several models have been proposed to explain the mechanism by which the Gal repressor acts, neither of which are exclusive, nor mutually exclusive (Irani, *et al.*, 1983; Majumdar and Adhya, 1984; Adhya, 1988). Repressor action is thought to be brought about by steric hindrance of RNA polymerase. Gal repressor bound to O_E, because of its close proximity to the promoter, may hinder RNA polymerase or CRP entry to, or activity at, the promoters, or both (this is assuming that repressor binding to O_E is weak and thus that repressor bound to O_I interacts with repressor at O_E and prevents dissociation of repressor from O_E). Alternatively, RNA polymerase could enter the *gal* promoters from outside the O_E–O_I DNA segment and gain access to the promoters. Repressor occupation of both O_E and O_I (which do not overlap the promoters) could block rapid access of RNA polymerase to the promoters. In a third model, repressor bound to O_E and O_I could form a complex via protein-protein interactions generating a DNA loop with a topologically independent DNA domain containing the promoter region. This DNA domain could have an altered promoter conformation which does not support RNA polymerase activity. Finally, binding of repressor to O_E and O_I could facilitate further binding of more repressor to the intervening DNA (which includes the promoters). This would "cover up" the intervening DNA, thereby preventing RNA polymerase binding in this region. Alternatively, repressor occupation of O_E and O_I may signal the condensation of another, say histone-like, protein in the promoters (reviewed by Adhya, 1988).

Although the overall genetic circuitry of the *gal* and other operons may be known, many questions still remain to be answered. For instance, how does the switch from cAMP-inhibited to cAMP-activated promoter occur? How are the genes in the operon coordinated from a

non-coordinated model? And how does repressor occupation at two distal operator sites modulate two intervening promoters? Answers to these questions will be invaluable in determining the mechanisms of transcriptional control at the various steps.

1.8 Control of Gene Expression in Bacteriophage

That phage lambda (λ) is able to grow in two different modes has provided us with an invaluable insight into the regulation of gene expression. In a λ-infected bacterium, the phage chromosome can enter the lytic cycle whereby various sets of phage genes are switched on ,or off, as directed by a precisely regulated series of events. The λchromosome is then extensively replicated, phage proteins synthesised, new phage particles packaged within the bacterium, followed by bacterial lysis releasing ~100 phage some 45 minutes post-infection. Alternatively, the phage chromosome can lysogenise its host cell. All but one of the phage genes, i.e. that coding for the λrepressor, are switched off. Integration of the phage chromosome into the host chromosome then occurs where it is passively replicated and distributed to the progeny bacteria . If unperturbed, this process can carry on indefinitely, with phage production occurring only very rarely. The phage chromosome can utilise bacterial enzymes to "sense" an impending demise of its host and can then abandon its passive replication pathway (reviewed by Ptashne, 1986b).

In very broad terms the switch from the lysogenic to the lytic state works in the following way. In a lysogen, a single phage gene, encoding the λrepressor, is on. The protein acts both positively and negatively to regulate gene expression. Binding of λrepressor to six sites, within two operator regions, on the phage DNA can turn off all other phage genes at the same time as switching on its own gene. There are ~100 active molecules of repressor per lysogen. The excess repressor is free to bind to an additional λchromosome which may be injected into that cell thus preventing λ from growing lytically on a λ-lysogen. Ultra violet irradiation of lysogens inactivates the repressor resulting in the synthesis of a second phage regulatory protein, Cro, which promotes, and is required for lytic growth. This protein binds to the same operator sites on the phage DNA as does repressor, but with opposite physiological effects. Lambda Cro and repressor, together with RNA polymerase, and their promoter and operator sites on the DNA, constitute the "genetic switch", where, in a lysogen,

the repressor gene is on and the Cro gene off, and where, during lytic growth, the Cro gene is on whereas the repressor gene is off (reviewed by Ptashne, 1986b).

Despite differences in size, domain organisation and tertiary structure, λrepressor and Cro interact with DNA in essentially the same way. Each binds to operator DNA as a dimer and uses α-helices to contact adjacent major grooves along one face of the DNA double helix (reviewed by Pabo and Sauer, 1984).

The *cI* and *cro* genes which encode λrepressor and Cro are adjacent and divergently transcribed from the λchromosome (Figure 6). The mRNA start sites of these two genes are separated by ~80 bp and within this region lie promoter and operator sequences which are directly involved in the binding of Cro and repressor (reviewed by Ptashne, 1986b). The intergenic region contains two promoters. One of these, P_R, directs rightward transcription of a set of early genes, including *cro*. The other promoter, P_{RM}, directs leftward transcription of *cI* in a lysogen (Ptashne *et al.*, 1980). A tripartite operator, O_R, overlaps these two promoters. Three adjacent sites, O_R1, O_R2 and O_R3 comprise O_R, the right operator of λ. Each of the operator sites overlaps one or the other promoter, or, in the case of O_R2, both of the promoters (Figure 6) (reviewed by Ptashne, 1986b). Each of the three operator sites is a 17 bp, the sequences of which are similar, but non-identical, and each of which is approximately symmetric. These sites are recognised specifically by both repressor and Cro; each protein having different affinities for the individual sites (Ptashne *et al.*, 1980).

1.8.1 The Action of Lambda Repressor

Lambda repressor acts as both a negative and a positive regulator of transcription. Repressor dimers normally bind cooperatively to adjacent operator sites O_R1 and O_R2, the dimer at O_R1 helping a second to bind to O_R2, and block transcription of the early lytic genes when bound at these sites (reviewed by Hochschild and Ptashne, 1988). Stimulation of repressor transcription from P_{RM} appears to involve contact between repressor bound at O_R2 and RNA polymerase bound at P_{RM}. The repressor protein contains 226 amino acids and can fold into two nearly equal sized

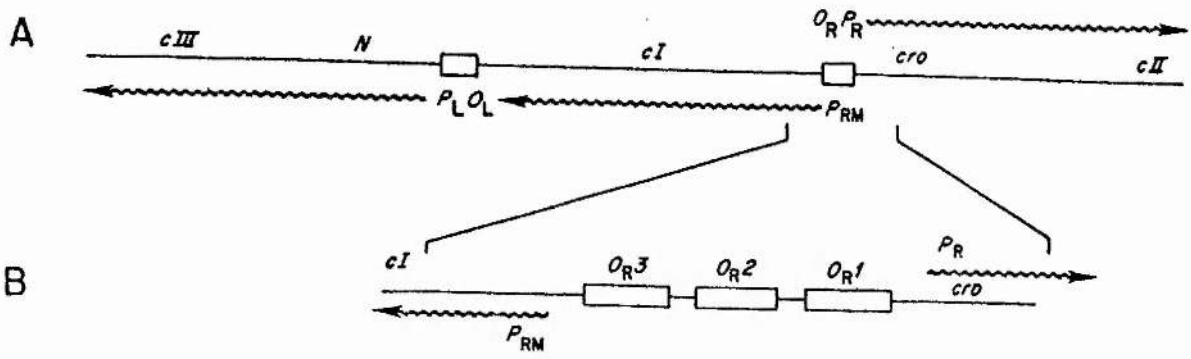


Figure 6. λ genes and regulatory elements

A. A portion of the λ genome. The arrows indicate the directions and start sites of transcription of various genes. $O_L P_L$ and $O_R P_R$ are the left and right operator and promoter regions, respectively. P_{RM} is the *cI* promoter active in a lysogen.

B. An expanded diagram of the λO_R region. $O_R 1$, $O_R 2$ and $O_R 3$ are the 17 bp repressor and Cro binding sites. The start sites of transcription from P_R and P_{RM}, which are located outside the operators, are indicated (Ptashne *et al.*, 1980).

domains connected by a string of 40 amino acids. The amino terminal domain contacts the DNA and RNA polymerase (Pabo and Lewis, 1982), and the carboxyl domain which facilitates dimer formation and mediates interaction between dimers (Sauer *et al.*, 1979). Each of the three sites in O_R can bind one repressor dimer. The sole role of O_R1 (a high affinity binding site for repressor) is to help ensure the occupancy of O_R2 (a weak repressor binding site), which means that at higher concentrations repressor can bind to O_R2 and activate transcription in the absence of repressor bound at O_R1 (Meyer *et al.*, 1980).

Positive control of transcription by repressor is mediated by the action of repressor at O_R2 helping RNA polymerase to bind and begin transcription at P_{RM} (which governs transcription of *cI* in a lysogen). This increases the rate of transcription ~tenfold. The repressor dimer bound to O_R2 increases the affinity of P_{RM} for RNA polymerase since protein–protein contacts as well as RNA polymerase/DNA contacts are required to position the polymerase. A repressor dimer bound at O_R2 therefore represses P_R by excluding the binding of RNA polymerase to that promoter (Meyer *et al.*, 1980). This in turn prevents transcription to the right but facilitates transcription to the left (a difference which is attributed to O_R2 being slightly closer to P_R than to P_{RM}). A repressor dimer bound at O_R1 would block binding of RNA polymerase to P_R but would be too far away to affect polymerase at P_{RM} which functions only at a very low level since no repressor is bound at O_R2. Binding of a single repressor dimer at O_R3 would block binding of RNA polymerase to P_{RM} and have no effect on transcription from P_R. In summary, when λrepressor acts as a positive regulator of transcription, repressor dimers are nearly always bound at O_R1 and O_R2 with O_R3 being generally free of repressor. This arrangement of repressor at O_R turns off transcription from P_R but turns on P_{RM}. In this way repressor, but not Cro, is synthesised (reviewed by Ptashne, 1986b).

Interaction of repressor dimers with the three operator sites is dependent on two factors. The first is the affinity of repressor for each isolated site in O_R – this is known as the "intrinsic" affinity of a site and can be measured using O_R DNA in which neighbouring sites are mutated such that they no

longer bind repressor. The second factor affecting the order of binding is the interaction of repressor dimers at adjacent binding sites (Ptashne *et al.*, 1980). Just as repressor at O_R2 assists binding of RNA polymerase to promoter P_{RM}, so repressors can interact with each other to facilitate binding.

The cooperative binding of repressor to O_R was demonstrated in *In vitro* studies of the binding of repressor at various concentrations to wild type and mutant O_R templates (Ptashne *et al.*, 1980). The intrinsic affinities of O_R2 and O_R3 for repressor are approximately equal and are more than an order of magnitude lower than that of O_R1. However, over a broad concentration range (including that found in a lysogen) O_R1 and O_R2 are filled and O_R3 tends to be vacant. This is due to a favourable interaction between repressor dimers bound at O_R1 and O_R2 which ensures that these sites are filled coordinately. Yet another favourable interaction between dimers bound at O_R2 and O_R3 is apparent if no repressor is bound to O_R1. This cooperativity of repressor binding has been ascribed by Ptashne and coworkers (1980) to protein–protein interactions as opposed to conformational changes in the DNA transmitted through the double helix (isolated amino domain of repressor binds to all three sites in O_R non-cooperatively). It is believed that the flexibility in the region connecting the amino and carboxyl domains could orient the carboxyl portion of a repressor bound at O_R2 such that it could contact one neighbouring dimer but not another. This contrasts the situation found with Cro/operator interactions in which Cro binds non-cooperatively. As with the repressor amino domain, mutations of any site in O_R affects the affinity of that site alone for Cro (Johnson *et al.*, 1979).

In a lysogen, repressor bound at O_R1 and O_R2 maintain the *cro* gene in the off position while stimulating transcription of its own gene, *cI*. As cells grow and divide, repressor is constantly being synthesised. However, if the repressor concentration increases, perhaps due to cell division being temporarily inhibited, then repressor would bind also to O_R3, thus turning off the *cI* gene. Once cell division had been resumed, and repressor concentration dropped to the proper level, the *cI* gene would then function to provide more repressor. In this way a constant level of repressor is

maintained in the cell despite changes in growth rate (reviewed by Ptashne, 1986b).

The concentration of repressor in a cell is high enough to ensure that O_R1 and O_R2 are likely to be filled at any one time. This means that in the absence of an inducing agent the lysogenic state is stable and can thus be maintained more-or-less indefinitely. If the switch from the lysogenic to the lytic state is induced then the action of Cro comes into play.

1.8.2 The Action of Lambda Cro

On induction of lysogens with UV light or other inducing agents, the primary effect is to damage DNA which leads to quite a marked change in the behaviour of the bacterial protein RecA (Roberts and Devoret, 1983). Under normal conditions, RecA catalyses recombination of DNA, but when the DNA is damaged, RecA then becomes a highly specific protease that cleaves repressor monomers at a site in the region connecting the amino and carboxyl domains. It is this physical separation of the two domains that is responsible for repressor inactivation because the separated amino domains can no longer dimerise. The affinity of the monomeric domains for the operator is too low to result in efficient binding at the concentrations found in a lysogen. This means that as repressor dimers are released from the operator, there are too few dimers to replace them (reviewed by Ptashne, 1986b).

As repressor vacated O_R1 and O_R2, the rate of repressor synthesis drops; RNA polymerase then binds to P_R and transcription of *cro* is initiated (Ptashne, 1986b). The sequence of events which "activate the switch" is now determined by Cro.

Cro dimer binds non-cooperatively to the same three 17 bp sites in O_R as repressor (Ptashne *et al.*, 1980; Johnson *et al.*, 1981). However, unlike repressor, Cro acts solely as a negative regulator of transcription. A salient feature of Cro's action is the order of its affinity for the three sites in O_R which is opposite to that of repressor (Ptashne *et al.*, 1980). Cro first binds to O_R3 preventing RNA polymerase from binding to P_{RM} which in turn prevents further synthesis of repressor. At this stage the "switch" from the lysogenic to the lytic state is activated and lytic growth ensues. As P_R continues to operate and more *cro* is transcribed, so to are the genes to the right of *cro* (whose

products are required for the early stages of lytic growth). Production of Cro continues until a level is reached whereby O_R1 and O_R2 are also filled and RNA polymerase is prohibited from binding to P_R. In this way Cro functions firstly to turn off repressor synthesis and then, at a slightly later stage, to turn off expression of its own gene, as well as that of other early lytic genes (reviewed by Ptashne, 1986b).

So it would appear that the ability of λ to grow in two alternate modes is dependent on the interplay of two regulatory proteins with specific sites on the phage DNA. Repressor and Cro bind to the same three sites in O_R and it is the order of site occupancy by repressor or Cro which activates the switch from the lysogenic to the lytic state. The effects on transcription of the proteins binding to O_R can be summarised as follows. Occupation of O_R1 and O_R2 (or both) by either protein represses P_R; occupation of O_R3 by either protein represses synthesis from P_{RM}. Finally, repressor bound to O_R2 stimulates P_{RM}.

1.8.3 Specific Recognition of Operator Sites by Repressor and Cro

Lambda repressor and Cro, in common with many other repressors and activators, use a common structural motif to recognise their binding sites in DNA (reviewed by Pabo and Sauer, 1984). This consists of a pair of α-helices (the "helix-turn-helix" substucture), one of which , the recognition helix, protrudes from the surface of the protein and fits into the major groove of the DNA. Solvent-exposed residues lie along the external surface of this helix contacting DNA functional groups in the major groove (W.F. Anderson *et al.*, 1981; J.E. Anderson *et al.*, 1981; Wharton and Ptashne, 1985). Hochschild and coworkers (1986) carried out a detailed study to explain how repressor and Cro distinguish between O_R1 and O_R3 in terms of specific amino acid/base pair interactions. Base pair changes were made in the operator sequences as well as amino acid substitutions in the proteins to analyse the basis of repressor binding more tightly to O_R1 than to O_R3, and *vice versa* for Cro. Solvent exposed amino acids were exchanged between the recognition helices of repressor and Cro. The sensitivities of these mutant proteins to the base pair changes that distinguish O_R1 and O_R3 were then tested.

Models of the proposed contacts made to specific positions in the operator are depicted in Figure 7. One half site of both O_R1 and O_R3 is the so called "consensus sequence" with the two operators differing at positions 3, 5 and 8 in the other half site (Figure 7A). Provided that flanking DNA sequences have no effect on binding, it would seem feasible that Cro and repressor must each distinguish between O_R1 and O_R3 on the basis of one or more of these three base pair differences. If we examine the model describing the interaction of the recognition α-helix of repressor and Cro with a consensus operator half site, we see that the first two residues in the recognition helices (Gln and Ser) contact base pairs at positions 2 and 4, respectively (Figure 7B). The base pairs at these two positions are invariant in all 12 lambda operator half sites. This would suggest that repressor and Cro make identical contacts to invariant positions in the operator but that other residues in the recognition helices (and amino- terminal arm in the case of repressor) are used to distinguish between O_R1 and O_R3 (Hochschild and Ptashne, 1986). This was indeed shown to be the case by Hochschild and coworkers (1986) where they demonstrated that the preferences for the two operator sites are determined by amino acid residues 5 and 6 of the recognition helices of both repressor and Cro, and by the amino terminal arm in the case of repressor. In addition, the most important base pairs in the operator which enable repressor and Cro to discriminate between O_R1 and O_R3 are position 3 (for Cro) and positions 5 and 8 (for repressor). If the recognition helices of repressor and Cro are similarly disposed and the amino acids in the helix function somewhat independently, then a particular amino acid in one of the two recognition helices should function analogously in the other recognition helix. Results from studies in which homologously positioned amino acids were exchanged found this to be the case (Hochschild and Ptashne, 1986). This suggests that the alignment between amino acid position in the operator is approximately the same for the two proteins. So it would appear that both repressor and Cro make an optimal set of contacts to the consensus half site and that the two proteins are differentially sensitive to alterations from the consensus as seen in the non-consensus half sites of O_R1 and O_R3.

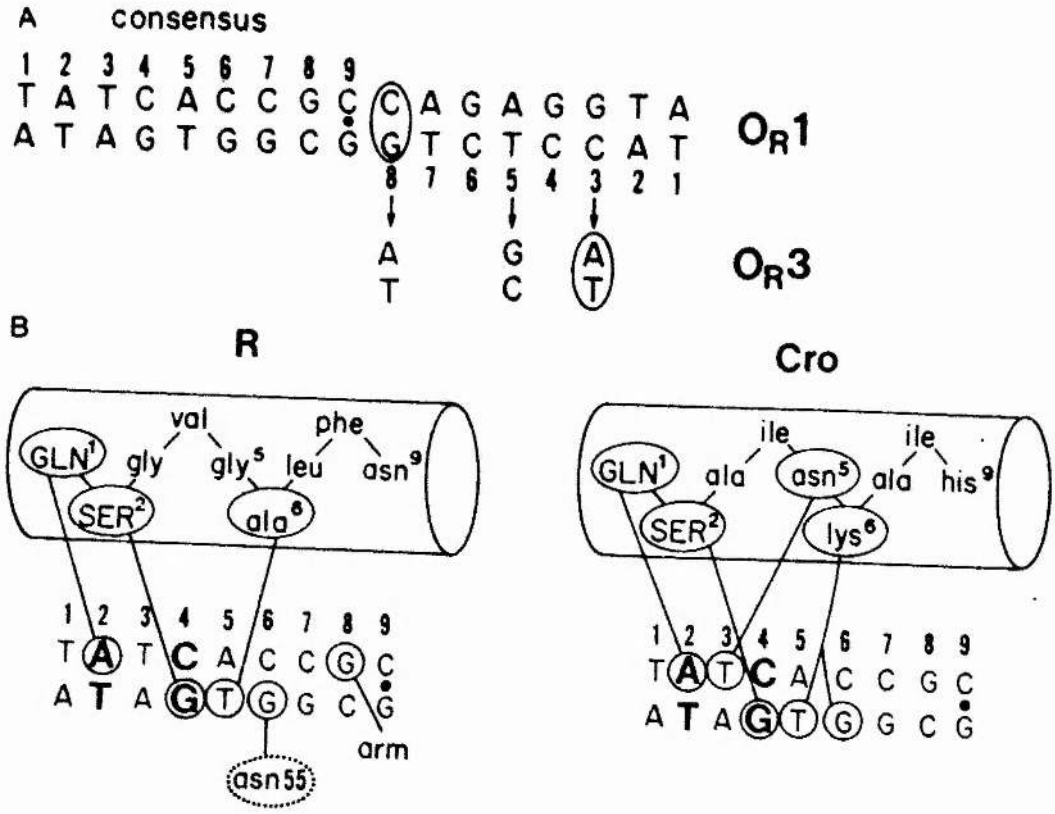


Figure 7. Operator sequences and proposed interactions with recognition helices

A. Comparison of the sequences of O_R1 and O_R3 are identical in the consensus half site. In their non-consensus half sites, O_R1 and O_R3 differ at positions 3,5 and 8. The circled bp each match the bp found at the corresponding position in the consensus half site. At position 5, neither O_R1 nor O_R3 had the consensus bp in the non-consensus half site.

B. Recognition α-helices of Cro and repressor and models of their proposed contacts. For repressor, the contacts are those proposed by Lewis *et al.* (1982) and for Cro, the contacts are those proposed by Ohlendorf *et al.* (1982) and Hochschild *et al.* (1986). Two contacts are shown involving amino acids outside the recognition helices. One involves the repressor's N-terminal arm and the other makes a proposed contact with repressor's Asn^{55} (Hochschild *et al.*, 1986).

1.8.4 Phage DNA–Protein Interactions

In addition to phage λ, repressor and Cro proteins are also encoded by the related phages 434 and P22. The amino acid sequences of the proteins are homologous especially in a region in which repressor and λCro have a similar "helix-turn-helix" secondary structure (Sauer *et al.*, 1982). Model-building studies indicated that this structure is important in specific binding to DNA and indeed this "helix-turn-helix" motif has been shown to be a common feature of many DNA binding proteins (reviewed by Pabo and Sauer, 1984). Even though a known structural motif can be assigned to a given DNA binding protein, this still does not allow us to predict how the protein actually contacts its binding site, or which residues, if changed, would alter the specificity of the interaction. Detailed structural information about site-specific recognition for several DNA–protein complexes is needed to ascertain whether or not any general principles arise. Such detailed information is now emerging and indeed X-ray crystallographic data on the repressor/operator, 434 repressor/operator and the 434 Cro/operator complexes is now available (Jordan and Pabo, 1988; Anderson *et al.*, 1987; Wolberger *et al.*, 1988).

The crystal structure of the amino terminal domain of λ repressor has been reported and a model for the repressor/operator interactions proposed (Pabo and Lewis, 1982; Lewis *et al.*, 1983). More recently the crystal structure of a complex containing the DNA binding domain of λrepressor and a λoperator site has been determined (Jordan and Pabo, 1988) which has revealed important side- chain/side-chain interactions previously unpredicted from modelling or genetic and biochemical studies.

The amino-terminal domain of λ repressor was co-crystalised with a 20 bp fragment that contains the operator site O_L1 (Jordan and Pabo, 1988). The overall structure of the complex reveals that the protein dimer binds by inserting one α helix from each subunit into the major groove of the DNA. The complex is roughly symmetric with the twofold axis of the dimer coinciding approximately with the twofold axis of the operator site. A similar arrangement was also observed in co-crystals that contain the 434 repressor (Anderson *et al.*, 1987). As previously observed in the protein crystal structure (Pabo and Lewis, 1982), the repressor monomer contains five α helices the first four of which form a compact globular domain, with helices 1, 2 and 3 on the face that is closest to the

operator DNA (Jordan and Pabo, 1988). Helices 2 and 3 form the conserved "helix-turn-helix" substructure. Helix 5 extends out from the globular domain and contacts helix 5' of the other protein subunit forming the dimer interface. The amino terminal arm wraps around the DNA and makes sequence-specific contacts in the major groove. The operator DNA is primarily B-form DNA with the largest deviations from this form occurring at the ends of the operator site, which bend slightly towards the repressor.

The repressor contacts sites in the major groove and phosphate oxygens of the DNA backbone (Jordan and Pabo, 1988). Gln^{44}, the first amino acid of helix 3, forms two H-bonds with the adenine of bp 2 (Figure 8A). The contact made by Gln^{44} is stabilised by a H-bond from the side-chain of Gln^{33} (the first residue in helix 2); this residue also H-bonds to the oxygen of phosphate P_B. This gives an extended H-bonding network whereby two protein side-chains connect a phosphate with a base in the major groove. A similar type of extended interaction involving a glutamine at the start of helix 2 and at the start of helix 3 has been demonstrated in a high resolution study of 434 repressor co-crystals (Aggarwal *et al.*, 1988). which may help explain why the conserved "helix-turn-helix" substructure generally begins with a glutamine (Pabo and Sauer, 1984). Ser^{45} (the second residue in helix 3) H-bonds to the N-7 of guanine 4 (a weaker interaction with the O-6 of this guanine may also occur) (Jordan and Pabo, 1988). Asn^{55}, which is in a loop three residues after the end of helix 3, H-bonds to the N-7 of guanine in bp 6. The co-crystal structure also demonstrates that hydrophobic interactions involving thymine methyl groups in bps 3 and 5 of the consensus half site contribute to the specificity of the repressor/operator interaction.

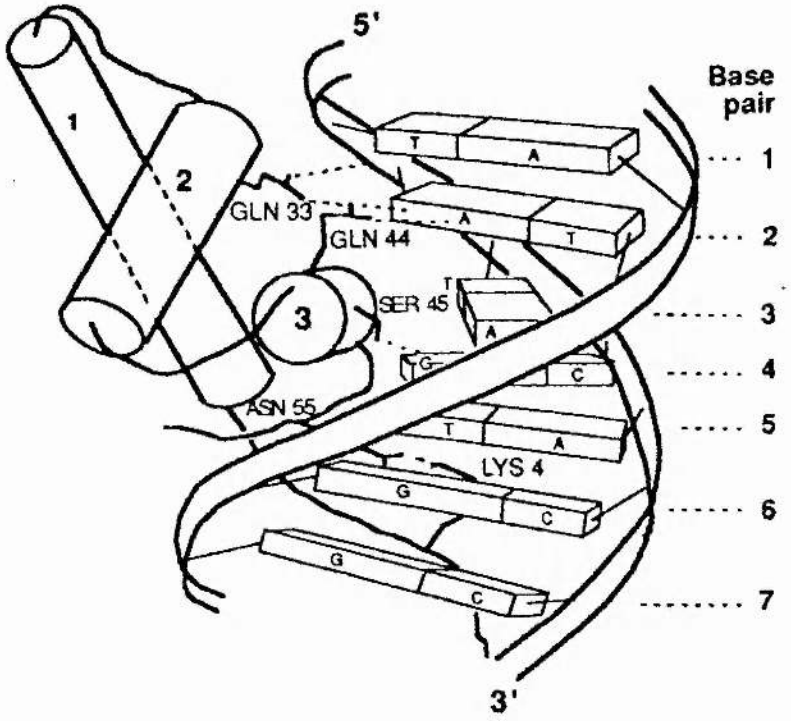
The repressor makes H-bonds with oxygens on each of five phosphates in both operator half sites (Jordan and Pabo, 1988). Many of these backbone contacts are to residues flanking the helix 2/helix 3 recognition unit and it is presumed that these contacts help to orient and position the residues that contact the edges of the base pairs (Figure 8B). It has been suggested that sequence-dependent variations in the structure of the DNA backbone may be important for site-specific recognition (Dickerson, 1983). However, in the λ repressor/operator complex no dramatic kinks or distortions in the DNA are observed (Jordan and Pabo, 1988) which would point towards those backbone contacts enhancing specificity by positioning the residues that contact the

Figure 8. Details of the λ repressor/operator interactions

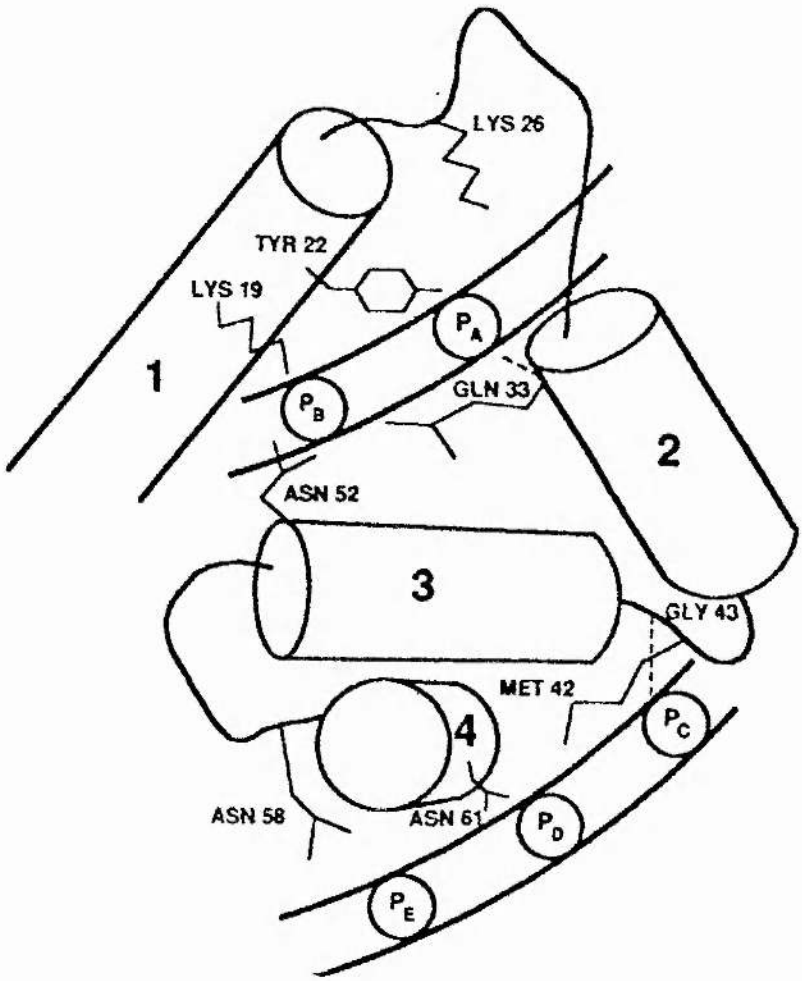
A. Schematic representation of the amino acid side chains that interact with base pairs in the major groove. H-bonds are indicated by the dotted lines. In the upper half of the figure, the major groove is readily visible for base pairs 1–3; in the lower half, the minor groove is readily visible.

B. Diagram showing the amino acid side chains that interact with the sugar-phosphate backbone in the consensus half site. Contacted phosphates (P_A–P_E) are indicated (Jordan and Pabo, 1988).

A.


5'
Base pair
1
2
3
GLN 33
GLN 44
SER 45
ASN 55
LYS 4
T
A
G
C
3'

B.


LYS 26
TYR 22
LYS 19
1
P_A
P_B
GLN 33
2
ASN 52
3
GLY 43
MET 42
4
P_C
ASN 58
ASN 61
P_D
P_E

bases. Jordan and Pabo (1988) have demonstrated that several amino acids from different sections of repressor may cooperate to recognise a particular bp, and it appears that recognition, like protein folding, involves cooperative interactions.

The Cro and repressor proteins of phage 434 are much more similar in structure and sequence than the corresponding proteins of λ which structurally agree only within the "helix-turn-helix" motif (Ohlendorf *et al.*, 1983). Transcriptional regulation in phage 434 is dependent on the ability of repressor and Cro to bind with different orders of affinity to the same set of six, partially symmetric, 14 bp operator sites. However, studies using systematic base substitutions in a "standard" operator sequence have shown that these two remarkably similar proteins are affected differently by a single substitution in otherwise identical operator sequences. To compare how 434 repressor and Cro bound to operator DNA, Wolberger and coworkers (1988) determined the crystal structure of 434 Cro bound to the same 14 bp operator site previously used in determining the co-crystal structure of the DNA binding domain of 434 repressor (R1-69) and its binding site in the operator DNA. It was presumed that since Cro and repressor are so similar in sequence and three dimensional structure they would form very similar complexes when bound to the same DNA fragment. This was found not to be the case: the 14 bp operator adopted a different conformation depending on whether Cro or R1-69 was bound.

The model of the Cro/DNA complex is shown in Figure 9 (Wolberger *et al.*, 1988). Each monomer consists of 5α-helixes. In each Cro monomer the recognition helix, α-3, lies in the major groove. The N-terminus of α-2 lies against the sugar–phosphate backbone on one side of the major groove. Helix 4 lies with its N-terminus near a phosphate and its C-terminus pointing away from the DNA. Helix 5 lies between the end of helix 4 and the N-terminus of helix 1. There appears to be no significant conformational change in the Cro monomers on binding to DNA.

If we superimpose models of Cro and R1-69 (Anderson *et al.*, 1987) we see a basic similarity (Figure 9A and B). However, there are also several important differences between the two structures in the conformations of both DNA and protein dimer (Wolberger *et al.*, 1988). Firstly the relative disposition of the two subunits are different as is the overall conformation of the two operator half sites. Also, the relation between protein and DNA in each operator is different. If we

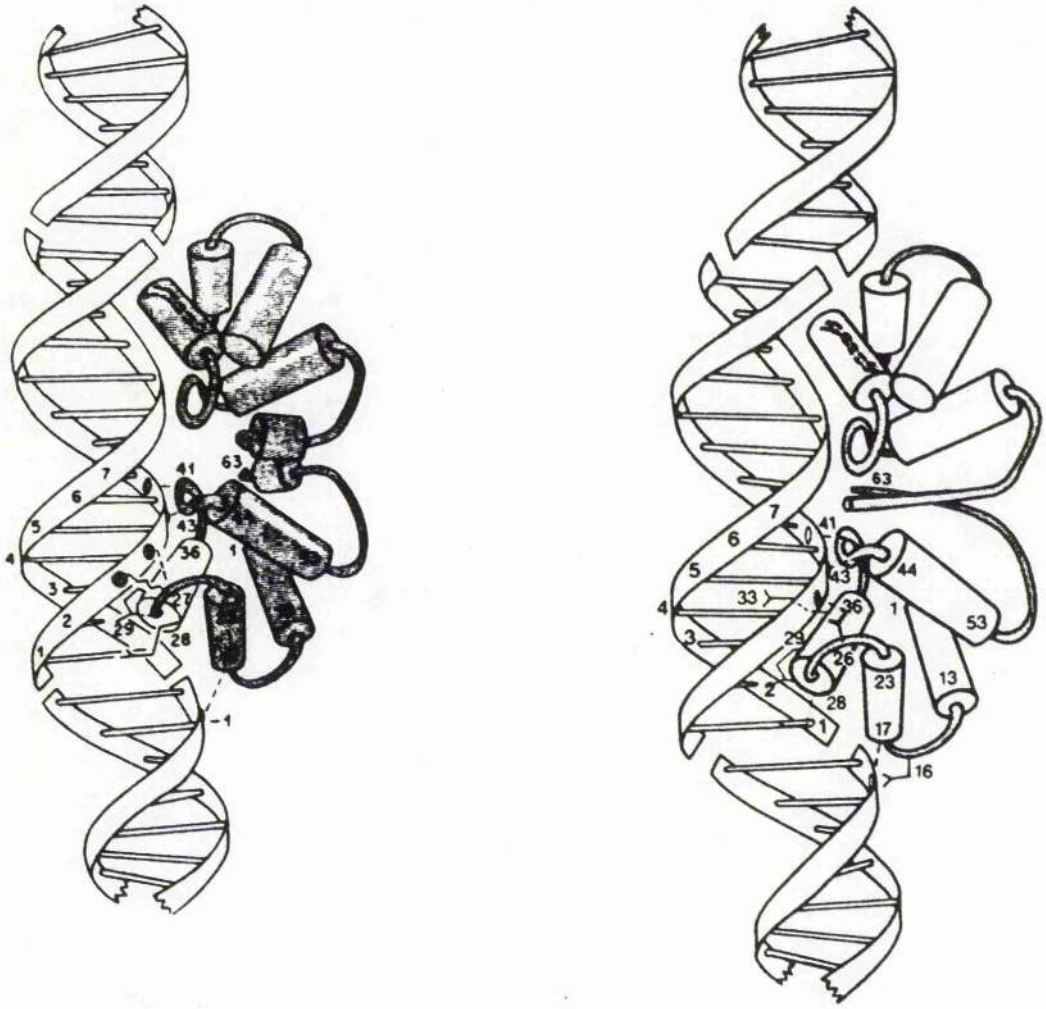


Figure 9. Models of the 434 Cro/operator and 434 repressor/operator interactions

A. General view of 434 Cro/14-mer complex. Parts of adjacent 14-mers are shown, stacked as in the crystal. Amino acid residue numbers are indicated at the beginning and end of the 5α-helices. In addition, certain residues important for the interaction with operator are indicated. Base pairs in one half site are numbered (1–7) along the backbone ribbon in a 5' to 3' direction (Wolberger *et al.*, 1988).

B. Schematic representation of the elements of the 434 repressor/operator complex. R1-69 α-helices are represented as cylinders and non-identical polypeptide chains as tubes. Key residues in the lower R1-69 monomer are shown as are the numbers of the first and last amino acid residues in each α-helix. Numbers along the ribbon indicate DNA base pairs in the operator half site (Anderson *et al.*, 1987).

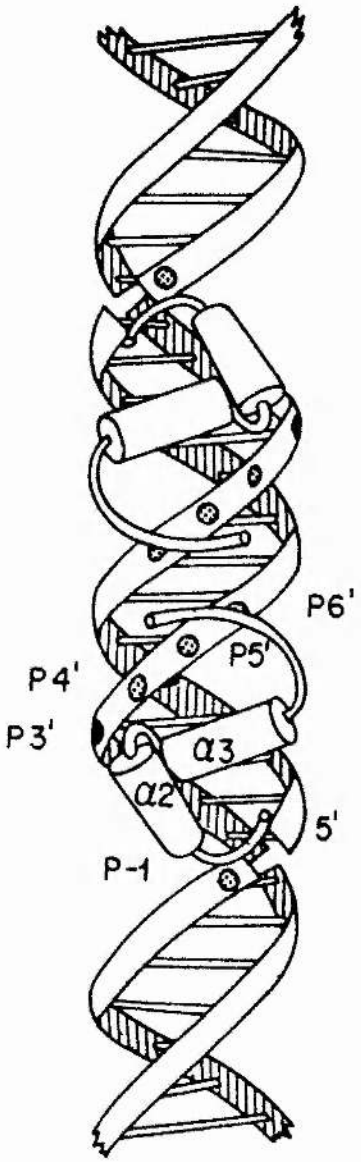


Figure 10. Contacts between 434 Cro and operator DNA backbone phosphates

Schematic representation of the phosphate residues contacted by 434 Cro. Phosphates 1 and 1' are not shown. The positions of the recognition helices, α2 and α3, are indicated. Phosphate 6' appears to have H-bonds from peptide -NH groups 40 and 41; phosphate 5', from -NH 43; phosphate 4', from -NH 27; phosphate 3', from Lys 27; and phosphate -1', from -NH 17 (Wolberger *et al.*, 1988).

examine the Cro monomer with respect to R1-69 positions, each appears to be rotated about the axis of helix 3 in such a way as to approximately preserve the spacing between the two alpha3 helices and, at the same time, widening the gap across the dimer interface. The difference in relative disposition of protein monomers in the Cro and R1-69 complexes stem from different protein–protein contacts across the dimer interface.

In the Cro/DNA complex, only a slight bending of the DNA is observed (Wolberger *et al.*, 1988) whereas DNA in the R1-69/14-mer complex is bent quite substantially (Anderson *et al.*, 1987). This difference in bending has been confirmed by results from gel electrophoresis DNA binding assays which indicate that binding of 434 repressor, or of R1-69 to the 14 bp operator, bends the DNA whereas binding of Cro does not (Koudelka *et al.*, 1988).

Each Cro monomer interacts in its half site with at least five phosphates (Figure 10) which correspond to those previously identified in ethylation interference studies (Wolberger *et al.*, 1988). With the exception of one phosphate group (P3'), ethylation of the same phosphates interferes with repressor binding (Bushman *et al.*, 1985). The contacts to P6', P5' and P1 are similar to those described in the R1-69 complex. Four of the five contacted phosphates can form H-bonds with the peptide -NH groups (Wolberger *et al.*, 1988). In two cases, the -NH groups lie at the N-terminus of an α-helix. This is consistent with protein/DNA backbone interactions also observed in the R1-69/14-mer and 20-mer complexes (Anderson *et al.*, 1987; Aggarwal *et al.*, 1988).

The affinity of 434 Cro and repressor for operator DNA is influenced by bp in the centre of the binding site which are not contacted by either protein (Koudelka *et al.*, 1987). This must therefore be due to effects of the central bp on DNA structure. This has indeed been shown to be the case with 434 repressor/operator complexes in which the non-contacted central bp influence repressor binding affinity by affecting the ability of the operator DNA to twist along its helical axis (Koudelka *et al.*, 1988). Unlike 434 repressor, Cro does not significantly bend the DNA but nevertheless exhibits the same sensitivity to central bp changes, as does R1-69, suggesting that the central bp likewise influence Cro affinity by affecting the ability of the operator to twist (Wolberger *et al.*, 1988). A network of bifurcated H-Bonds is seen in the R1-69/20-mer complex (Aggarwal *et al.*, 1988). It is possible that such a network is also found in a similar complex with Cro to the extent that the array of

non-coplanar bp which participate in such a network could represent a response to helical twist at the central bp rather than to bending and minor groove compression (Wolberger *et al.*, 1988).

Although it may be possible to design a set of rules governing the specific recognition for a given DNA binding protein with its cognate site, these rules need not necessarily be true, even for proteins such as the 434 repressor and Cro that recognise DNA by virtue of the same structural motif (Anderson *et al.*, 1987; Wolberger *et al.*, 1988).

2.Control of Gene Expression in Eukaryotes

2.1 Transcriptionally Active Chromatin

To date, most detailed information available on how genes function has come from studies of prokaryotes whose genetic material is distributed throughout the cell. A major difference between prokaryotes and eukaryotes is the presence in the latter of a group of basic proteins, histones, which are strongly complexed with DNA in chromosomes (reviewed by Bradbury *et al.,* 1981). Chromatin represents the state of DNA in the living eukaryotic cell in which linear DNA molecules are compacted up to 10,000 fold in association with histones and other proteins (reviewed by Gross and Garrard, 1987). The basic unit of chromatin is the nucleosome, a repeating subunit constructed by wrapping 166 to 168 bp of DNA into two left-handed superhelical turns around an octameric histone core (H2A, H2B, H3, $H4)_2$. One molecule of histone H1 is bound to DNA entering and leaving the nucleosome particle thus stabilising this region. A variable length of linker DNA, up to 70 bp, is found between neighbouring particles (reviewed by Yaniv, 1986). Most chromatin within the eukaryotic nucleus is then packed into a higher order structure, a 30 nm solenoid, that results from the ordered folding of polynucleosome strings (Felsenfeld and McGhee, 1986).

Cell free extracts capable of specific gene transcription *in vitro* have been established but they generally employ non-physiological templates (Gross and Garrard, 1987). Indeed, nucleosome formation *in vitro* has been shown to block initiation of transcription (Gottesfeld and Bloomer, 1982) yet most genes that are being expressed *in vivo* (as well as quiescent genes) are organised into nucleosomes (Weintraub, 1985). How then is the transcriptional machinery able to transcribe through a nucleosomal template?

When nuclei are digested with DNase I, an interesting difference between transcriptionally active and inactive genes emerges. Active genes show increased sensitivity to nuclease treatment when compared to their inactive counterparts (Weintraub and Groudine, 1976). They are also preferentially solubilised by brief micrococcalnuclease treatment (Bellard *et al.,* 1977). These differences mat be attributable to chromatin- containing active genes being packed less tightly in relation to the more compact chromatin-containing inactive genes. In addition, modification of the primary nucleosome structure of active genes may also result in a modified higher order structure.

Transcriptional control in eukaryotes is dependent on the recognition of specific *cis*- acting

DNA sequences with a variety of *trans*- acting regulatory proteins. In view of this, an important question then is whether proteins which recognise sequences in B-form DNA in solution can recognise the same sequence features when the DNA is incorporated into nucleosomes. Such recognition is likely to depend not only on the conformation of nucleosomal DNA , i.e. on parameters such as helical twist, bp opening angles, groove width etc., but also with its configuration, i.e. the long-range bending of the DNA and its proximity to the histone proteins (reviewed by Morse and Simpson, 1988).

A fundamental clue towards understanding the mechanisms that initiate the formation of transcriptionally active chromatin, was provided by the discovery of nuclease hypersensitive sites in viral and cellular chromatin (reviewed by Gross and Garrard, 1987). Sites hypersensitive to DNase I often reside within *cis*- acting DNA sequences that are crucial for gene expression, such as upstream promoter regions. These sites often appear before the onset of transcription and are developmentally programmed (Weintraub, 1985). The occurrence of DNase I hypersensitive sites has been frequently observed, in many different types of genes, i.e. constitutively expressed, inducible, developmentally regulated and viral genes. For example, in the genes coding for the *Drosophila* 70 Kda heat shock protein, rearranged genes coding for the kappa chains of immunoglobulins, mouse alpha and ß- globin, c-myc, histone genes of *Drosophila* or sea urchins and SV40 and polyoma viruses. Nuclease hypersensitive sites have also been mapped around promoter and UASs of several yeast genes (reviewed by Yaniv, 1986).

The SV40 genome is transcribed into two primary transcripts originating from opposite strands. The 5' ends of the transcripts are located near the origin of DNA replication. SV40 mini chromosomes contain a nuclease sensitive region which lies between the origin of replication and the major start site of late mRNA (Varshavsky, *et al.*, 1979; Sargosti *et al.*, 1980). Treatment of nuclei, prepared from infected cells or of isolated mini chromosomes, with DNase I or a variety of restriction endonucleases, resulted in double stranded cleavage of the viral DNA in this region of the genome. In addition, electron microscopy studies of extracted SV40 mini chromosomes revealed a "gap", or nucleosome-free segment of DNA, in the circular nucleo-protein complexes (reviewed by Yaniv, 1986). This 400 bp segment was mapped by restriction analysis to the region between the origin of DNA replication and the 5' end of the late mRNA (Jakobovits *et al.*, 1980).

The presence of this nucleosome free region was not, however, as a result of the dissociation of one or two histone octamers since the total number of nucleosomes counted in gapped or ungapped molecules was the same (Saragosti *et al.*, 1980). It is believed that a segment of DNA can be excluded from nucleosomes thus becoming more sensitive to nuclease digestion than nucleosomal DNA. This region then acts like an "open window" in the chromosome structure but that the sensitivity of this window is modulated by the interaction of certain non-histone proteins, in turn creating as series of protected and hypersensitive DNA sequences. Indeed, a variety of *trans*-acting cellular factors have been identified which bind to the gapped region of SV40 containg the viral enhancer element and early promoter (Zenke *et al.*, 1986; for review see N.C. Jones *et al.*, 1988).

Many DNase I hypersensitive sites are observed in sequences upstream of the transcription initiation site of a gene in tissues that express such a gene, suggesting that these sites are intimately associated with gene activity arising as a consequence of the transcription process. Alternatively, these sites could play a crucial role in permitting this transcription. The acquisition of at least some of the hypersensitive sites appears to be a prerequisite for transcription activation (reviewed by Yaniv, 1986).

Brown and coworkers (1988) have recently described a hypersensitive site in chromatin formed over sequences 5' to the human *hsp70B* gene transcription unit. Determinants sufficient to form the hypersensitive site were contained in a 280 bp sequence which corresponded approximately to the region that is hypersensitive. Deletion of sequences from either end of this region resulted in reduced hypersensitivity which suggested that multiple genetic elements contributed to the formation of this chromatin structure. This structure is present prior to heat treatment (Wu, 1980; Costlow and Lis, 1984; Brown *et al.*, 1988) and does not change in intensity or position after heat shock, even though *hsp70* gene expression is entirely dependent on heat induction. Heat shock gene expression shows a sequence-dependence that is best explained by a requirement for several heat shock regulatory elements (Schiller *et al.*, 1988). However, binding of heat shock transcription factor (HSTF) or heat shock activator protein, appear not to be responsible for the hypersensitive chromatin structure (reviewed by Brown *et al.*, 1988). In addition to heat shock element consensus sequences, there are a number of potential binding sites for the transcription

factor Sp1 (Briggs *et al.*, 1986) within the hypersensitive region of *hsp70B* (Brown *et al.*, 1988). The DNA segment 5' to the *hsp70B* gene also contains several close matches to a sequence resembling the core of the human metallothionein II metal response element (Karin *et al.*, 1984). Such an element is also present upstream from the human *hsp70A* transcription unit (Wu *et al.*, 1985). So it may be that chromatin structure is dependent on the interactions between proteins bound to DNA segments that are distant from one another and the spacing between these binding sites. Their phasing on the DNA double helix with respect to one another, may also be crucial. Since determinants for the hypersensitive site described by Brown *et al.* (1988) cannot be resolved to a segment of DNA shorter that about 280 bp, it suggests that multiple protein-binding sites are involved. It has been speculated (Ptashne 1986a) that hypersensitive sites may be regions of chromatin distorted due to interactions between proteins bound at sites separated by long stretches of DNA.

An extensive study of the tissue-specific hypersensitive site within the 200 bp promoter region immediately upstream of the chicken adult ß-globin gene has been carried out (Emerson *et al.*, 1985) and results have shown this element to be nucleosome free but containing several regions of protection-associated specific *trans*- acting factors. What is of particular interest is that the hypersensitive site can be reconstituted *in vitro* on cloned DNA templates provided that red blood cell protein factors are present prior to, or during the assembly of nucleosomes. The site cannot be induced by these *trans*- acting factors *in vitro* once a nucleosome has formed within the promoter region. In a similar manner, the ability of transcription factor TFIIIA to generate a transcriptionally active 5S RNA gene template is dependent on the addition of non-histone chromosomal proteins prior to, or during nucleosome formation, but not after (Gottesfeld and Bloomer, 1982; Kmiec and Worcel, 1985).

It has recently been observed that nucleosomes are frequently positioned at preferred, or even fixed locations, along specific DNA sequences. The determinants governing this "phasing" appear to depend, in part, on the bendability of the DNA molecule itself (Travers, 1987b). Sequence-dependent variation in the mechanical properties of the DNA seem to define the rotational orientation of the major and minor grooves with respect to the histone octamer, i.e., which helical face of the DNA is in contact with the octamer. However, it is believed that additional

sequences are required to bring about a precise translational setting for the nucleosome, i.e., the nucleotide sequences entering and leaving the nucleosome or the nucleosome phasing frame (reviewed by Gross and Garrard, 1987). So far phasing frames appear to be constant irrespective of whether or not a gene is destined for expression in a given tissue. For example, the mouse ß-globin gene has nucleosomes phased along the sequence in an identical manner in erythroid and non-erythroid cells (Benezra *et al.*, 1986). In erythroleukamia cells, four nucleosomes appear to be removed, or modified, from the promoter region of the gene before, and after, induction of globin gene expression. The adjacent phasing frames are not disturbed – a phenomenon also observed in fibroblasts (reviewed by Gross and Garrard, 1987).

It is easy to visualise how "open" regions within promoters could act as entry sites for RNA polymerase molecules. Enhancer elements also possess DNase I hypersensitive sites but these transcriptional control elements can be located several kilobases away from promoters. It is possible that enhancers and promoters come into intimate contact via interaction between *trans*- acting factors bound at their hypersensitive sites. This could then lead to a looping out of the intervening DNA (Ptashne, 1986a). Such a multiprotein complex as the base of this loop would afford a strong attraction site for RNA polymerase to initiate transcription . A specific class of DNA sequences, matrix associated regions (MARs) have been identified, which sometimes lie close to enhancer elements, and which contain consensus sequences for the interaction with topoisomerase II, known to be localised at or near the base of chromosomal loops in mitotic chromosomes (reviewed by Gross and Garrard, 1987). It is possible that the base of a chromosomal loop could regulate the level of DNA supercoiling in the entire loop (or mini-loop) through the action of topoisomerase II and *trans*- acting factors with MARs and enhancers.

Of current interest is whether chromosomal loops that contain transcriptionally active chromatin are under torsional stress, i.e., do they contain unrestrained DNA supercoils? *E. coli* chromosomal DNA is organised into loops, or domains. The DNA is negatively supercoiled. The torsional strain is introduced into the bacterial DNA by an ATP-dependent action of DNA gyrase. Topoisomerase I has the reverse action, it relaxes negatively supercoiled DNA. So the final extent of supercoiling of bacterial DNA is achieved as a result of the balance between the action of these two enzymes. In bacterial DNA, changes in torsional strain correlate with changes in gene expression (reviewed by

Yaniv, 1986). Recent evidence for torsional stress in eukaryotic chromatin has been provided in that amphibian oocyte 5S RNA genes are assembled, both *in vivo* and *in vitro,* into chromatin that contains topoisomerase I-relaxable, supercoiled DNA (Kmiec and Worcel, 1985). These stressed molecules are believed to be transcriptionally active in that they are preferentially sensitive to nucleases, they are insoluble, they have a disrupted nucleosomal structure and their product is stimulated by the 5S-specific transcription factor TFIIIA (Kmiec and Worcel, 1985). However, yeast 2μm minichromosomes, which are moderately transcribed, are not under torsional stress *in vivo* and do not appear to act as substrates for the putative gyrase-like activity of topoisomerase II (reviewed by Gross and Garrard, 1987). It appears then that torsional stress may play a role in transcriptional activation of some eukaryotic genes, but not of others.

Trans- acting factors may only be required to establish transcriptionally active chromatin and not for its maintenance or propagation. Indeed many of the biochemical properties of transcriptionally active chromatin are maintained and/or transmitted to progeny long after transcription has ceased. For example, the promoter region of a previously transcribed immunoglobulin light chain can spontaneously delete with a concomitant reduction in transcription to 0.7 % of the normal level, but, without any loss of the transcriptionally active chromatin phenotype (reviewed by Gross and Garrard, 1987). In addition, preferential nuclease sensitivity and DNase I hypersensitive sites persist within the chicken ß-globin domain and vitellogenin gene long after the removal of an inducing stimulus or transcription has come to an end. This would suggest that the transient binding of *trans*- acting factors to enhancers and/or other *cis*- acting regulatory elements directs chromatin to generate and memorise a stable heritable complex. The mechanisms that determine "memory " are probably due to multiple phenomena such as alternative DNA conformations, dynamic equilibria between protein–DNA and protein–protein interactions, timing of DNA replication, accessibility to post-translational machinery and selective undermethylation of DNA (reviewed by Gross and Garrard, 1987)

It is clear that future research has to centre on characterising the proteins that bind to particular sites in active chromatin and on understanding their function at the domain level. It is necessary to pinpoint the mechanisms controlling synthesis of these proteins and on their binding to target DNA sequences in certain cell types and not in others.

2.2 Yeast Promoter Elements

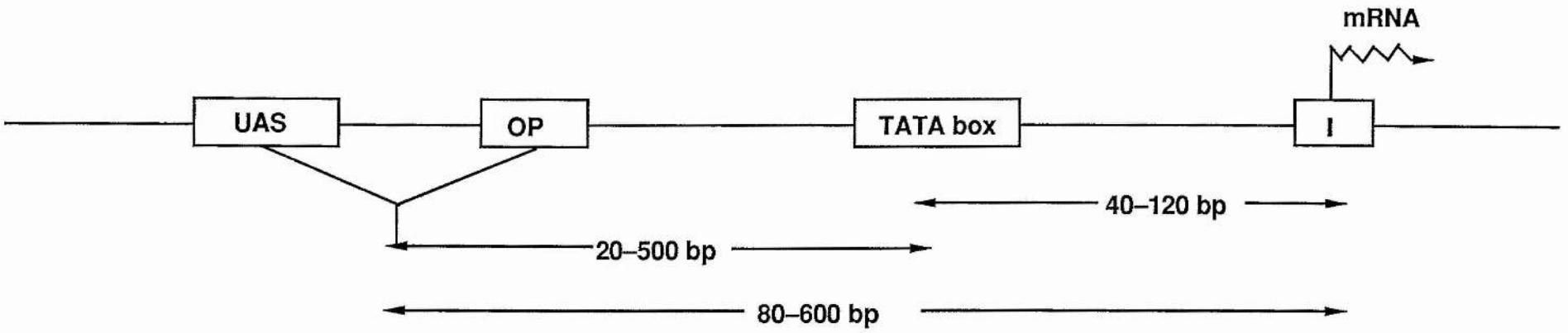
Analysis of RNA polymerase II transcribed genes in the yeast *Saccharomyces cerevisiae* (*S.cerevisiae*) has identified three essential promoter elements (Figure 1), the TATA box, the upstream activation site, and the initiation site, all of which have been shown to be necessary for the control, accuracy and frequency of transcriptional initiation (Olesen *et al.*, 1987 and references therein). In addition, some yeast promoters contain operator (OP) elements that are involved in transcriptional repression (reviewed by Brent, 1985).

The initiator element (I) is located near the mRNA start site and is the principal determinant governing where transcription begins. In contrast to the situation found in the promoters of higher eukaryotes, in which initiation sites are generally determined by their distance from the TATA element, accurate initiation in yeast is still observed even when the distance between the TATA element and the initiator element is varied (Chen and Struhl, 1985). The close proximity of the initiation site to the initiator element may represent specific DNA sequences preferred by RNA polymerase II itself, or by a closely associated factor.

TATA elements (consensus sequence TATAAA) are present in nearly all yeast promoters and are a necessary component of the promoter. However, in isolation they are not sufficient for transcriptional initiation of most yeast genes (Struhl, 1987). The distance between the TATA element and the transcriptional start site ranges between 40 and 120 bp. This differs from the situation in higher eukaryotes where TATA sequences are frequently located 25 to 30 bp upstream of the initiation site. Because they are present in a variety of promoters, TATA elements are presumed to play a somewhat general role in the transcription process, perhaps mediated by the binding of a general transcription factor. However, two functionally distinct classes of TATA elements, involved in constitutive or inducible transcription of genes, have been described in *S. cerevisiae* (Struhl, 1986). *his* 3 and *pet* 56 are adjacent genes that are expressed at similar levels under normal growth conditions, and are transcribed in opposite directions from initiation sites that are separated by 200 bp. Although each gene has its own TATA element, a 17 bp poly (dA-dT) sequence, located between the genes, serves as the upstream promoter element for both (Struhl, 1985). This constitutive element thus acts bidirectionally to activate transcription of two unrelated

Figure 11. Yeast promoter elements

The regions of a yeast promoter including the transcription start site or initiator element (I), the TATA box, operator (OP) and upstream activating sequence (UAS) are shown. The approximate distances between the various promoter elements and the transcription initiation site are indicated.


mRNA
UAS
OP
TATA box
I
40–120 bp
20–500 bp
80–600 bp

genes.

During conditions of amino acid starvation, *his* 3 but not *pet* 56 transcription is induced over the basal level, even though the critical regulatory site is upstream of both respective TATA regions. In addition, only one of the two normal *his* 3 initiation sites is subject to this induction. It has been suggested that the *his* 3-*pet* 56 intergenic region contains constitutive and inducible promoter elements with different properties (Struhl, 1986). In particular, two classes of TATA elements, constitutive (T_C) and regulatory (T_R) that are involved in constitutive and inducible *his* 3 transcription. These elements are distinguishable on the basis of several criteria; by their ability to respond to upstream regulatory elements, by their effects on the selection of particular initiation sites, and by their physical structure in nuclear chromatin (presumably due to the unusual structure of poly (dA-dT)) (Struhl, 1986). Recent genetic evidence (Chen and Struhl, 1988) from studies using saturation mutagenesis *in vitro* and functional analysis *in vivo,* demonstrated a strict DNA sequence requirement for transcriptional induction from T_R, strongly suggesting that the T_R element is a target site for a sequence-specific DNA binding protein. Since most yeast promoters contain the sequence TATAAA or TATATA near the mRNA start sites – it is extremely likely that the protein which binds to T_R will be involved in transcription initiation. However, not all yeast promoters contain sequences compatible with T_R function, despite the fact that they may contain A/T-rich sequences (Chen and Struhl, 1988). If we assume that these promoters require "downstream elements" for transcription, it follows that such elements will not be recognised by the putative T_R binding protein but, instead, are targets for other DNA binding proteins. It is possible then that yeast cells contain multiple proteins with distinct sequence specificities that function in a similar way to that of the T_R binding protein and that yeast promoters can be divided into different classes on the basis of their downstream promoter elements. This is suggestive of the situation involving the different bacterial sigma factors that interact with the core RNA polymerase to generate distinct holoenzymes recognising different -10 and -35 promoter sequences (Losick and Pero, 1981). By this analogy, the protein which binds to T_R may be similar to the 70 Kda sigma factor in that both proteins are general transcription factors important in the expression of the majority of the organism's genes. In view of this, one could speculate that proteins which recognise T_C, or other

sequences, would parallel the minor sigma factors (Chen and Struhl, 1988). However, the yeast T_R binding protein differs from the bacterial sigma factor in that it binds directy to DNA and not to RNA polymerase.

The phosphoglycerate kinase (PGK) gene of *S. cerevisiae* encodes one of the most abundant mRNA and protein species in the cell (Holland and Holland, 1978). Deletion analysis of the 5' non-coding region of the PGK gene revealed that a UAS located some 450 bp upstream of the mRNA start site was required for full transcriptional activity, but expression of this gene did not appear to require any TATA sequences. Deletion of one or both of the potential TATA sequences did not affect either the levels of PGK mRNA, or the site of transcription initiation (Ogden, *et al.*, 1986). This finding that PGK gene expression does not appear to be dependent on TATA sequences is rather novel amongst yeast genes but is similar to the situation observed in the expression of several higher eukaryotic genes, such as those coding for papovavirus late functions and adenovirus-2 DNA binding protein, which both lack TATA boxes (reviewed by Benoist and Chambon, 1981). Similarly, there is no TATA box requirement for expression of the sea urchin histone H2A gene when injected into *Xenopus* oocytes, although its absence leads to the use of several new cap sites (Hentschel *et al.*, 1980). Ogden and coworkers suggest that, in the case of PGK gene expression, the mRNA initiation site may be all that is required to position the transcriptional start site.

Ruden *et al.*, (1988) have demonstrated GAL4-activated transcription at nearly identical levels when the GAL4 responsive element was located at nine different positions, 33 to 67 bp upstream of the 5'TATAAA3' sequence in the *GAL1* promoter (which contains the *GAL1* TATA box). This activation was not dependent on GAL4 binding to the same side of the DNA double helix with respect to the TATAAA sequence. This apparent lack of requirement for a strict alignment between an activator (GAL4) and its target (for example the TATA-binding protein or RNA polymerase II), assuming that they contact each other through the formation of DNA loops, might be explained if, when bound to DNA, the interacting regions of one or both the proteins protrude laterally from the opposite (or nearly opposite) sides of the DNA helix, but this would make detection of any helical periodicity very difficult (Ruden *et al.*, 1988). Alternatively, the interaction between the activator and its target site may be strong enough to supply the energy required to twist and bend the DNA;

DNA supercoiling might decrease the energy required for this loop formation (Kramer *et al.*, 1988).

The third class of RNA polymerase II transcribed promoter elements are the upstream activating sequences (UASs) which normally determine the particular regulatory properties of a given yeast promoter (reviewed by Guarente, 1984). These upstream elements resemble mammalian enhancer sequences in that they can function at long distances (up to 600 bp) in an orientation-independent manner with respect to the responding promoter. But, unlike most other enhancer elements, they do not appear to function downstream from the responding promoter (Guarente and Hoar, 1984; Struhl, 1984). Regulatory proteins have been isolated, such as the yeast *trans*- activators GAL4 and GCN4, which bind in a sequence-specific manner to the upstream elements (Giniger *et al.*, 1985; Hope and Struhl, 1986). In addition, UASs for several constitutively expressed yeast genes have been shown to consist of poly (dA-dT) homopolymer sequences which may act by altering chromatin structure rather than binding specific factors (Struhl, 1985).

Chen and coworkers (1987) carried out a series of experiments designed to distinguish between chromatin accessibility and protein-protein contact models for the mechanism of transcription activation by eukaryotic upstream elements. Two types of recombinant plasmids were introduced into yeast cells; those which expressed T7 RNA polymerase and a second set of target molecules in which the yeast UAS_G or the poly(dA-dT) element from the *ded1* gene were fused upstream of the T7 promoter which in turn was fused directly to the *his3* mRNA coding sequences, effectively replacing the yeast (TATA-dependent) transcription machinery with T7 RNA polymerase. The chromatin accessibility model predicts that activation should be independent of the particular transcription machinery. This means that upstream elements that operate by this mechanism should also stimulate T7 RNA polymerase transcription. In contrast, elements that work by protein-protein contact should be unable to stimulate transcription by T7 RNA polymerase since it is unlikely that yeast activator proteins could make the required specific contacts with this prokaryotic enzyme. By comparing the levels of *his3* RNA initiated from target or control molecules, they were able to determine whether transcription by T7 RNA polymerase is enhanced by yeast upstream elements. Under conditions where the UAS_G would normally be either activating or non-activating, T7 RNA polymerase- stimulated transcription, above the level observed in the absence of any upstream element, was not detected. They could also show that the yeast *ded1* poly(dA-dT) upstream

element enhances transcription by T7 RNA polymerase in yeast cells but not *in vitro* on purified DNA indicating that the observed enhancement reflects properties of the transcriptional activation process *in vivo*, as opposed to some unusual property of the RNA polymerase. So it would appear that activation by the *ded1* poly(dA-dT) element may involve effects on the chromosome template, thereby facilitating entry of the transcription machinery, whereas activation by the UAS_G may involve specific contacts between GAL4 and the transcriptional machinery.

Repression of transcription in yeast has been shown to be dependent on the binding of regulatory proteins to the DNA at sites other than the UAS (Brent, 1985). The best studied case of transcriptional repression in yeast depends on the product of the *MATα2* gene (α2). *S. cerevisiae* can exist as one of three distinct cell types; as a non-mating diploid, or as a haploid of either the **a** or the α-mating type (Brand *et al.*, 1987). The α2 protein is a cell-type specific repressor which turns off transcription of its target genes – the **a**-specific genes – by binding to an operator located upstream of each **a**-specific gene. The mating type of each cell is determined by the allele at the *MAT* locus which encodes the α or **a** gene products. At almost every cell division, haploid cells are able to "switch" their mating type by replacing the allele at the *MAT* locus with an allele of the opposite mating type copied from either the *HMR* (**a**) or *HML* (α) loci. Because alpha2 is present in α cells but not in **a** cells, the **a**-specific genes are turned off in α cells but are expressed in **a** cells (reviewed by Herskowitz and Oshima, 1981). The repressor protein(s), in the presence or absence of cofactors, might interfere sterically with the binding of positive regulatory proteins to the UAS, or the TATA element, or even hinder the movement of an RNA polymerase molecule from its binding site in the promoter to the mRNA start site.

2.3 Yeast Activator Proteins

Genetic evidence has shown that the expression of several well characterised yeast genes occurs by positive activation of transcription (reviewed by Guarente, 1984). Attempts to identify yeast activator proteins have generally begun by the isolation of *trans*- acting mutations that are defective in the induction of specific genes. However, biochemical evidence is necessary to prove that a particular mutation directly affects transcription, for example, by altering a specific DNA–protein interaction. Such specific binding to upstream promoter elements has been

demonstrated for the yeast GAL4, HAP1 and GCN4 proteins (reviewed by Struhl, 1987). In addition, mutational analysis of yeast activator proteins indicates that DNA binding and transcriptional activation are independent functions carried out by distinct regions of the protein (Brent and Ptashne, 1986; Hope and Struhl, 1986; Keegan *et al.*, 1986).

2.3.1 The GAL4 *Trans*- Activator

When yeast is grown on medium containing galactose, transcription of the galactose metabolic genes, GAL1, GAL7 and GAL10, is induced by more than 5000-fold (Hopper *et al.*, 1978; St. John and Davis, 1981). In the case of GAL1 and GAL10, induction of these divergently transcribed genes is dependent on a short DNA sequence located between them, the GAL upstream activation sequence (UAS_G), as well as the product of the constitutively expressed *GAL4* gene.

GAL4 recognises and binds to four related, 17 bp dyad symmetric sequences within UAS_G, both *in vivo* and *in vitro* (Giniger and Ptashne, 1988). Moreover, a double stranded, synthetic oligonucleotide, the 17-mer, which is a near consensus of the GAL4 binding sites and has a higher affinity for GAL4 *in vitro* than any of the binding sites from UAS_G, was suficient to confer nearly wild type levels of regulated, GAL4-mediated inducibility in yeast (Giniger *et al.*, 1985). A single copy of the synthetic 17-mer was placed 130–300 bp upstream of transcription start sites of separate constructs which fused the GAL1 or CYC1 genes to the coding sequence of the *E. coli lacZ* gene. In yeast cells grown on galactose-containing media, these plasmids produced 50–70% of the ß-galactosidase activity observed in analagous constructs that carried a complete UAS_G. Moreover, the 17-mer conferred no activity when plasmids containing this sequence were transformed into a yeast *GAL4* - strain. The yeast GAL4 protein can also activate transcription in mammalian cells. Kakidani and Ptashne (1988) demonstrated that GAL4 could activate the mouse mammary tumour virus (MMTV) promoter in mammalian cells when the glucocorticoid-response elements (GREs) were replaced by UAS_G. However, insertion of a GAL4 binding site in the MMTV promoter containing the GREs abolished the promoter's ability to respond to glucocorticoid receptor (GR) in the presence of dexamethasone. This may be a consequence of displacing the GREs from their normal position by insertion of the UAS_G, thus impairing the response to GR. Another possibility is that a protein

present in the transfected cells binds to the GAL4 regulatory sequences and blocks the effect of an upstream activator. Considering the findings of Giniger and Ptashne (1988), it is feasible that GR and GAL4 could bind cooperatively and this may reflect a direct interaction between these proteins or an indirect interaction mediated by contact with a common, third factor of the transcriptional machinery.

Giniger and Ptashne (1988) demonstrated that two of the naturally occurring GAL4 binding sites for UAS_G, sites 3 and 4, work synergistically to activate transcription (neither site alone had any appreciable UAS activity but together formed a UAS nearly 20% as active as the wild type UAS_G). This synergism was attributed to the cooperative binding of GAL4 – *in vivo* footprinting experiments showed that neither site 3 nor site 4 was occupied to any great extent by GAL4 when the sites were isolated, but, when paired, both sites were significantly (though not fully) occupied. Studies carried out using the synthetic 17-mer suggested that there is no mechanistic requirement for multiple sites for each bound GAL4 to activate transcription efficiently, rather the activity of a GAL UAS is roughly proportional to the number of GAL4 molecules bound.

The cooperative binding of GAL4 to UAS_G is very similar to the situation observed in phage lambda repressor binding to its operator sites (O_R1, O_R2 and O_R3) on the phage DNA. The intrinsic affinities of O_R2 and O_R3 for lambda repressor are approximately equal but are more that 10-fold lower than that of O_R1 (reviewed by Ptashne *et al.*, 1980). However, over a broad concentration of repressor, O_R1 and O_R2 are filled and O_R3 is usually vacant. It would appear that repressor-bound O_R1 facilitates repressor binding at O_R2 much more readily than repressor binding at O_R2 facilitates binding at O_R1. In addition, an alternative favourable interaction between repressor dimers bound at O_R2 and O_R3 is evident only if no repressor is bound at O_R1. The observed cooperativity was ascribed to protein-protein interactions and not to DNA conformational changes transmitted through the doublehelix.

The DNA binding and transcriptional activation domains of GAL4 are clearly separable functional domains. The 74 N-terminal amino acids of GAL4 are sufficient for DNA binding but fail to activate transcription (Keegan *et al.*, 1986; Hope and Struhl, 1986). The transcriptional activation function of GAL4 has been ascribed to two relatively short regions (region I, residues 148–196, and region II,

residues 768–881) both of which have an acidic character (Ma and Ptashne, 1987a).

Webster and coworkers (1988) have reported that the yeast UAS_G exhibited properties indistinguishable from those of higher eukaryotic enhancer elements in HeLa cells producing GAL4. In addition, a UAS_G 17-mer and the SV40 enhancer could synergistically stimulate transcription from the ß-globin promoter in HeLa cells synthesising GAL4 (a phenomenon also observed in studies of the MMTV promoter by Kakidani and Ptashne, 1988). A chimeric protein containing the human oestrogen receptor DNA binding domain and GAL4 activating region II could stimulate oestrogen response element-controlled transcription with approximately the same efficiency as the wild type oestrogen receptor. Takahashi *et al.*, (1986) have suggested that protein factors bound to enhancer elements stimulate transcription by interacting with factors bound to other promoter elements – such a mechanism, involving DNA-bound GAL4 has also been suggested in yeast (Ptashne, 1986). This in turn implies that the same type of contacts exist in yeast and HeLa cells between GAL4 and other proteins. Accordingly, it would be of interest to establish whether there is any stereoalignment required for GAL4 to stimulate transcription from higher eukaryotic promoter elements. Further studies with higher eukaryotic transcription factors are necessary to investigate the possible presence and function of acidic regions similar to those present in yeast activators.

Mutational analysis of the activating regions of GAL4 indicate that the acidic nature of these regions is important for activity (Gill and Ptashne, 1987). Giniger and Ptashne (1987) constructed a synthetic protein bearing a novel 15 amino acid peptide linked to a DNA binding fragment of GAL4. The synthetic peptide was acidic and should therefore form an amphipathic α-helix, having one hydrophilic face bearing the acidic residues, and one hydrophobic face. When expressed in yeast, this artificial protein could efficiently activate transcription of the GAL1 gene, but a similar protein with the novel 15 amino acids in a scrambled order, which was unable to form an amphipathic structure, was unable to activate transcription of GAL1. This lends support to the idea that amphipathicity of putative α-helices in GAL4 is necessary for its transcriptional activity. Furthermore, the regions of bacteriophage lambda and 434 repressors involved in transcriptional activation are known from X-ray crystallography to be negatively charged, amphipathic -helices, and it has been proposed that these function by contacting RNA polymerase (Pabo and Lewis, 1982; Hochschild *et*

al., 1983; Anderson *et al.*, 1985; Bushman and Ptashne, 1986).

In the absence of galactose, the stimulatory action of GAL4 is abrogated by the specific negative regulatory protein GAL80. Thus, yeast strains which lack GAL80, but not GAL4, transcribe the GAL genes constitutively (reviewed by Ma and Ptashne, 1987b). GAL4 binds specifically to DNA sequences in UAS_G by its N-terminus. A surface of GAL4, not involved in binding to DNA, appears to activate transcription via interaction with another component of the transcriptional machinery. In the absence of galactose, GAL80 binds to GAL4 and blocks gene activation . The region recognised by GAL80 is the 30 amino acids at the C-terminus of GAL4, an essential element of activating region II (Ma and Ptashne, 1987a). Johnston *et al.*, (1987) suggest the following model for negative regulation by GAL80. Under non-inducing conditions, GAL4 is bound to UAS regions of GAL1 and other galactose regulated genes. Transcriptional activation is blocked by GAL80 binding to the C-terminal region of GAL4, thus preventing this region from interacting with a transcription factor(s). Galactose (or a derivative) activates transcription by binding to GAL80 and inducing a conformational change in the protein, thereby releasing GAL80 from GAL4 which is now capable of interacting with other transcription factors. Under non-inducing repressing conditions (i.e. glucose medium), other regulatory circuits are probably involved (Oshima, 1982).

In addition to activating transcription in yeast, GAL4 can also function as a transcriptional repressor in *E. coli* (Paulmier *et al.*, 1987). The chromosomal *lac* operator was replaced by a yeast GAL4 binding site and this was found to repress ß-galactosidase synthesis by at least 30-fold in GAL4-containing bacteria. Paulmier and coworkers (1987) have suggested that binding of a GAL4 protein dimer to the target DNA overlapping the transcriptional initiation site will inhibit transcription by a similar mechanism to that of the *lac* repressor. These findings provide us with yet another means for the study of eukaryotic DNA binding proteins using the tools of the classical bacterial geneticists. The DNA target sites of various eukaryotic transcription regulatory proteins could be substituted for the *lac* operator, expression cDNA libraries in plasmid or bacteriophage introduced, followed by screening for the few *lac*$^-$ descendants that can be expected. It is feasible that even short cDNA sequences or yeast DNA fragments carrying only the DNA binding domain of a protein will allow repression and hence easier identification of genes encoding transcription factors.

Gill and Ptashne (1988) reported that various GAL4 derivatives, when expressed at high levels

in yeast, could inhibit transcription of certain genes lacking GAL4 binding sites. This inhibitory effect was localised to the activating region of GAL4 and that derivatives bearing stronger activating regions had a greater inhibitory effect. Two closely related transcripts were analysed and only the one which was inducible by GAL4 when the UAS_G was present could be inhibited by high levels of GAL4. They suggest that this inhibition, termed "squelching", reflects a titration of a transcription factor by the activating region of GAL4. This inhibition could be relieved by simultaneous expression at high levels of the negative regulator GAL80. The stimulatory effect of GAL80 may be a consequence of GAL80 masking the activating region of GAL4, thereby relieving the "squelching". It is possible that interaction between an activating region on a protein and a transcription factor, for example the TATA-binding protein, although facilitated by DNA binding, can also occur free in solution. At high concentrations of activator, this reaction would appropriate the factor thereby inhibiting transcription. (Gill and Ptashne, 1988).

2.3.2 The HAP Activator Proteins

The products of the HAP1, HAP2 and HAP3 regulatory loci in yeast activate transcription of two cytochrome c genes, CYC1 and CYC7 (Pfeifer *et al.*, 1987a and 1987b). There are two upstream activation sites adjacent to the CYC1 gene; UAS1 which is activated by HAP1, in a reaction dependent on haem, by HAP1, and UAS2 which is activated by the combined action of HAP2 and HAP3 (Guarente *et al.*, 1984; Pinkham and Guarente, 1985).

Within UAS1 there are two distinct subsites, A and B, both of which are required for activity. Region B forms part of a recognition site that binds to a yeast factor, RC2, *in vitro* (Arcangioli and Lescure, 1985). When cells are grown under haem-defficient conditions, extracts do not demonstrate RC2 binding activity – neither is activity restored by addition of haem to the extracts.

Pfeifer and coworkers (1987a) investigated the activity of the HAP1 gene product *in vitro* and found that the protein bound in region B of UAS1 to the same DNA sequences as RC2, but that binding of these factors was mutually exclusive. Moreover, haem was shown to stimulate the binding of HAP1 to UAS1 but not RC2, even though haem was required for the *in vivo* synthesis of RC2. This would suggest that haem is directly involved in the activation of UAS1. Haem may function as a ligand that binds to HAP1 or, alternatively, the effect of haem could be indirect,

occurring via a second protein that modifies HAP1 in some way.

The role of RC2 in activation is not known. However, the regulation of the activity of RC2 and HAP1 suggests that RC2 is not a repressor that is inactivated by the inducer haem. It is possible that RC2 is a negative regulator that modulates UAS1 activity by competing with HAP1 for binding. Another possibility is that RC2 could be a positive regulatory factor involved in the activation of the adjacent UAS2. An additional factor, RAF (Region A Factor), distinct from HAP1, has been identified, and it has been suggested that it is the complex of RAF and HAP1 that is required for transcriptional activation.

The CYC7 gene which encodes the minor form of cytochrome c in yeast is regulated coordinately with CYC1 (Pfeifer *et al.*, 1987b). In addition, loss of function mutations in the HAP1 gene abolish the activity of the CYC7 UAS as well as UAS1. HAP1 binds to a site in the CYC7 UAS that bears no obvious sequence similarity to the HAP1 binding site in UAS1 of the CYC1 gene. Furthermore, the binding site of HAP1 to the CYC7 site, like its binding to UAS1, is stimulated by haem (Pfeifer *et al.*, 1987a and b).

In general, genes coordinated by the same regulator are marked by a consensus sequence in their 5'-flanking DNA. The case of CYC1 and CYC7 is an obvious exception to this rule with the best alignment of the binding site sequences yielding only a match of seven nucleotides that are scattered across the 23 bp bound (Pfeifer *et al.*, 1987b). How then does HAP1 recognise two different DNA sequences?

HAP1 could contain physically distinct DNA binding regions, one of which recognises UAS1 and the other CYC7 UAS. However, dramatically increasing the dosage of HAP1 results only in a two-fold increase in CYC7 expression (K. Pfeifer, unpublished data), suggesting that the supply of HAP1 does not limit CYC7 expression to any great extent. Another possibility is that a single DNA binding domain of HAP1 could recognise both UAS1 and CYC7 UAS. This idea is consistent with the findings of Pfeifer *et al.* (1987b), namely that truncated HAP1 proteins do not separate the binding domains for the two sites. Secondly, the specific contacts made between HAP1 and the two DNA sites are similar, and finally, UAS1 and CYC7 sequences compete with each other, with comparable affinity, for binding to HAP1. How this single DNA binding domain recognises these two divergent sequences remains to be elucidated but the answer may in fact lie in the DNA itself in

that the DNA sequences of the UAS1 and CYC7 sites could create some common structure that contributes to recognition by HAP1 (Pfeifer *et al.*, 1987b; Rhodes and Klug, 1986).

Activation of UAS2 of the CYC1 gene is highly regulated by catabolite repression and is repressed about 30-fold by a shift of cells from lactose to glucose containing medium. The activity of UAS2 depends on the products of both the HAP2 and HAP3 genes (Guarente *et al.*, 1984). Mutations in either HAP2 or HAP3 result in the inability to activate UAS2 and a global defect in the expression of genes involved in respiratory metabolism.

Analysis of the UASs of the HEM1 gene (Keng and Guarente, 1987), COX4 (Schneider and Guarente, unpublished data) and UAS2 reveals that all these HAP2-HAP3 dependent UASs contain the sequence TN(A/G)TTGGT. Wild type UAS2 contains the sequence TGGTTGGT but a G to A transition in UAS2 (termed UAS2UP1) alters this to TGATTGGT, thus increasing the homology to the HEM1 and COX4 UASs and resulting in a 10-fold increase in UAS2 activity *in vivo* (Guarente *et al.*, 1984). Linker scanning analysis of UAS2 has confirmed the critical nature of the consensus element for activity of the site (Forsburg and Guarente, 1988). In addition, this element is homologous to the CCAAT box of higher eukaryotes. Moreover, major groove contacts made by HAP2-HAP3 in UAS2UP1, as determined by methylation interference experiments, and the CCAAT box binding factor are identical (Olesen *et al.*, 1987).

Olesen and coworkers (1987) have shown that both HAP2 and HAP3 in yeast extracts bind to UAS2UP1 and give rise to a single DNA–protein complex in native polyacrylamide gels. The binding of either HAP2 or HAP3 is abolished if either of these two proteins is synthesised in a strain mutant in the complementary HAP gene, implying that the binding of HAP2 and HAP3 is interdependent. It may be that the two proteins bind in very close proximity to one another and stabilise each other's binding cooperatively, or they exist as a heterodimer.

A series of experiments were carried out by Hahn and Guarente (1988) in an attempt to distinguish between the aforementioned possibilities. HAP2 and HAP3 were tagged by gene fusion to *E. coli LexA* and *lac Z* genes, respectively, and the products purified through four chromatographic steps. The co-purification of LexA-HAP2, HAP3-ß-galactosidase, and UAS2 binding activity demonstrated that HAP2 and HAP3 associate in the absence of DNA to form a multisubunit activation complex – a complex which must be capable of at least three functions. First,

the complex must bind specifically to UAS2 and related sequences in the yeast genome. It must activate transcription once bound at UAS2, and finally it must be able to respond to signals generated under inducing conditions in media containing a non-fermentable carbon source. To date, it would appear that these functions are not distributed evenly between the two proteins. Although both HAP2 and HAP3 are required for binding to UAS2UP1, either HAP2 and HAP3 both make contacts with the DNA, or one protein makes all the contacts and is held in the proper conformation by the other. In addition, activation of transcription by the bound complex appears to require both proteins (Olesen and Guarente, unpublished data).

Recently Chodosh and coworkers (1988) demonstrated that the subunits of the human CCAAT-binding protein, CBP1, and yeast HAP2/HAP3 heterodimer are functionally interchangeable in that the yeast/human hybrid complexes which are formed retain the ability to specifically recognise eukaryotic CCAAT-containing transcription elements. A CCAAT motif has also been found in the UASs of several yeast genes regulated by HAP2 and HAP3 (Keng and Guarente, 1987), suggesting that the HAP2/HAP3 complex is a global activator of respiratory genes, functioning specifically by recognising CCAAT elements located in the UASs of this family.

2.3.3 The Yeast GCN4 Activator Protein

The product of the yeast *GCN4* gene is required for the coordinate induction of amino acid biosynthetic genes (reviewed by Jones and Fink, 1982). GCN4 protein, synthesised *in vitro* from the cloned gene, has been shown to bind specifically to the promoter regions of *his* 3 and three other coordinately regulated genes, but not to similar regions regulated by general control (Hope and Struhl, 1985). Deletion analysis of the *his* 3 promoter region, coupled with DNAse I footprinting, revealed that GCN4 binds specifically to a 12 bp region that is critical for the induction of transcription *in vivo*. It appears that the sequence requirements for protein recognition and for transcriptional activation are indistinguishable. Mutations which abolish *his* 3 induction also fail to bind GCN4. Similarly, derivatives which confer *his* 3 inducibility are able to bind GCN4, strongly suggesting that GCN4 binding to the *his* 3 regulatory site directly mediates induction *in vivo* (Hill *et al.*, 1986).

Hope and Struhl (1986) generated a series of deletions of the *GCN4* gene and examined the

properties of the resulting peptides in three different ways. First the mutant proteins, after transcription and translation *in vitro,* were tested for their ability to bind specifically to *his* 3 DNA. Second, the truncated derivatives were introduced into yeast cells and assayed for their ability to induce *his* 3 expression under conditions of amino acid starvation. Finally, by analogy with experiments of Brent and Ptashne (1985), deleted derivatives of LexA-GCN4 hybrid proteins were examined for their ability to induce transcription of a promoter containing the *lexA* operator as an upstream element *in vivo.* Results of these studies demonstrated that the 221 N-terminal amino acids of GCN4 can be deleted with no apparent effect on specific DNA binding *in vitro.* This indicates that the remaining 60 C-terminal amino acids can fold independently of the rest of the protein into a discrete, fully functional DNA binding domain (Hope and Struhl, 1986).

When Chou and Fasman (1978) rules for predicting protein secondary structure were applied to GCN4, three alpha-helices and a stretch of ß-turn were predicted within the DNA binding domain. However, these sequences fail to show any of the homology typically observed in the "helix-turn-helix" motif of prokaryotic repressors and activators (reviewed by Pabo and Sauer, 1984). Further, the "zinc finger" motif, exemplified by the eukaryotic transcription factor TFIIIA (Miller *et al.,* 1985), does not apply to GCN4 since there are no cysteine residues in the protein. Clearly then, X-ray crystallographic analysis or 2-dimensional nuclear magnetic resonance studies will be required to determine the molecular basis for this particular DNA–protein interaction.

Hope and Struhl (1988) demonstrated that a 19 amino acid segment of GCN4 (amino acids 107 to 125) was critically involved in transcriptional activation suggesting that either this 19 amino acid region was sufficient for activation, or that GCN4 contains redundant activities. The apparent small size of the activating region would suggest that activation, *per se,* does not involve a catalytic function such as a topoisomerase, methylase or nuclease. It is more likely that activation involves protein–protein contacts with no strict sequence requirements.

Hope and coworkers (1988) investigated the structure and function of the GCN4 activation region using high resolution deletion analysis and proteolytic mapping. The primary observation from their studies was that the 19 amino acid segment was insufficient for transcriptional activation – many LexA-GCN4 derivatives removing part or all of this segment showed full or partial transcriptional activation. Analysis of internal deletions revealed that this 19 residue segment was

neither sufficient nor essential for transcriptional activation activity. Activation appeared to correlate fairly well with the amount of the acidic region of GCN4 (amino acid residues 88–147) remaining, but not so well with the particular regions that are retained. Perhaps the most striking feature resulting from their studies is that progressive deletion of the activation region causes a series of small, step-wise reductions in activity as opposed to defining a position which results in an abrupt loss of activity. This would indicate that transcriptional activating regions do not have a defined tertiary structure analogous to that found in active sites or domains in a protein. Results from proteolysis experiments (Hope *et al.*, 1988) lend support to this idea in that a protease resistant C-terminal domain was readily released by treatment with proteases which preferentially cleave unstructured protein regions. However, under the same conditions, the N-terminal portion of GCN4 was completely cleaved, indicating that GCN4 contains an independently structured DNA binding (C-terminal) domain with the remainder of the protein being relatively unstructured. In contrast to this, cleavage by chymotrypsin generates two equally stable intermediates suggesting an apparently contradictory view in which the N-terminal region is structured. However, deletion analysis showed that large N-terminal segments of GCN4 were resistant to protease digestion just as if they were part of a structured domain but that these segments could be removed without destroying the integrity of that domain. This unusual chymotrypsin cleavage pattern correlates well with the presence of a functional transcriptional activation region. One might interpret this result by speculating that the activation region has a local structure that inhibits chymotrypsin cleavage of an otherwise unstructured N-terminal region of GCN4.

Saturation mutagenesis of a target sequence within the *his* 3 promoter defined a 9 bp region as being the major determinant for GCN4 binding – with optimal binding to DNA containing the palindromic sequence ATGA(C/G)TCAT (Hill *et al.*, 1986). In the course of studies involving truncated GCN4 molecules, it was observed that the electrophoretic mobility of specific DNA–protein complexes was inversely related to the molecular weight of the GCN4 derivative (Hope and Struhl, 1986). Hope and Struhl (1987) devised a novel means of examining the subunit structure of GCN4. It was based on the principal that if GCN4 were to bind DNA as a dimer, then a protein–DNA complex involving a heterodimer of GCN4 monomers having different molecular weights should migrate electrophoretically between the two complexes involving the homodimers.

When both monomeric units are equally radiolabelled and all three possible dimeric species are capable of binding DNA, then the band corresponding to the heterodimer complex should be twice as intense as each of the bands representing the homodimer complexes. Using this strategy they observed three bands whose relative intensities were in a 1:2:1 ratio – completely consistent with GCN4 binding as a dimer to the single site in the *his* 3 target DNA, and not as a monomer or as a multimer with a larger number of subunits. Deletion analysis revealed that GCN4 could dimerise through the C-terminal 60 amino acids, suggesting that dimerisation requires an intact DNA binding domain. In addition, although dimer formation appears to be critical for DNA binding, stable dimers can form even in the absence of target DNA.

Although GCN4 binds as a dimer to a 9 bp palindrome, its binding appears to involve non-identical contacts between GCN4 monomers and overlapping half-sites in the target DNA. It is feasible that when GCN4 is bound to a regulatory element, it would be symmetrically disposed with respect to the DNA template and hence capable of activating transcription bidirectionally, even though the protein–DNA contacts cannot be identical for the half-sites of a single promoter. Thus bidirectional activation of eukaryotic transcription may reflect the DNA binding properties of activator proteins rather than the actual mechanism of transcriptional activation.

On the basis of several criteria, Hope and coworkers (1988) propose that the activation region of GCN4 is a dimer of intertwined alpha-helices, one helix from each GCN4 monomer. Firstly the unusual chymotryptic pattern reveals a structure in GCN4 which is dependent on the acidic activation region. Second, the stability of the dimeric DNA binding domain ,even in the absence of target DNA (Hope and Struhl, 1987), should allow the formation of and/or stabilise the proposed interactions between helices of different monomer subunits. Finally, since the LexA DNA binding domain binds very poorly to its operator because of weak dimerisation (Little and Hill, 1985; Hurstel *et al.,* 1988), it would seem likely that transcriptional stimulation via the LexA domain also requires that the activation region should permit dimerisation. This model would also explain why amphipathic helices should form functional transcriptional activation regions since this would readily permit a structure involving interacting hydrophobic residues which are protected from solvent and acidic residues.

A new oncoprotein, Jun, which causes sarcomas in chickens, has recently been identified

(Maki *et al.*, 1987). Jun shows significant homology to the DNA binding domain of GCN4 (Vogt *et al.*, 1987), even though the DNA binding domains share only 45% identity in amino acid sequence. This homology suggests that Jun might bind to DNA in a sequence-specific manner and perhaps recognise the same DNA sequences as GCN4. Additionally, the consensus sequence for binding of the mammalian transcription factor AP-1, which interacts with phorbol-ester inducible promoter elements (Angel *et al.*, 1987; Lee *et al.*, 1987), is very similar to the GCN4 recognition site. To determine if the Jun oncoprotein would bind to the same DNA sequences as GCN4, molecules encoding LexA-GCN4-Jun (LGJ) hybrid proteins, which contained the LexA DNA binding domain, the GCN4 transcriptional activation region and the putative Jun DNA binding domain, were constructed and the DNA binding properties of the hybrids analysed in yeast (Struhl, 1987b). The main finding of these studies was that the C-terminus of the Jun oncoprotein could replace the GCN4 DNA binding domain *in vivo*. In addition, the LGJ hybrid could induce the expression of *his* 3 and other amino acid biosynthetic genes, whereas Jun-dependent transcriptional activation appeared to require the LexA DNA binding domain. The most likely explanation for this observation is that the LexA domain facilitates dimerisation of the Jun domain – given the palindromic nature of the optimal *his* 3 target site, Jun, like GCN4, probably forms dimers in order to bind DNA.

To determine whether or not the Jun protein itself was capable of transcriptional activation in yeast, LexA-Jun hybrids were constructed in which the entire JUN coding region was fused directly to the LexA DNA binding domain (Struhl, 1988). This construct was introduced into yeast strains lacking the *GCN4* gene but harbouring the *E.coli* LexA operator fused upstream of a yeast TATA element and a *lacZ* structural gene. Transcriptional activation through the LexA DNA binding domain could then be quantified by ß-galactosidase activity, and activation via the Jun oncoprotein could be assessed by complementation of the *GCN4* deletion. LexA-Jun was found to activate transcription, thus indicating that the JUN oncoprotein contains a sequence (or sequences) which can function as a transcriptional activator in yeast. LexA-Jun could functionally replace GCN4 even though this hybrid protein lacked any GCN4 coding sequences. This also suggests that Jun and GCN4 contain transcriptional activation regions that are equally functional when fused to a heterologous DNA binding domain. It may be that JUN activation in yeast reflects both an evolutionary and mechanistic similarity between yeast and vertebrate transcription factors.

The similarity between the DNA recognition sites for GCN4, Jun and AP-1 suggests that the Jun oncoprotein may be derived from the normal AP-1 transcription factor. If this is indeed the case, it is possible that signals such as phorbol-esters, are received at the cell membrane and transmitted to intracellular messengers such as protein kinase C. Activated protein kinase C could modify AP-1, thereby resulting in altered gene expression. It may be that the Jun oncoprotein differs from AP-1 in such a way as to escape the normal regulatory pathway (Struhl, 1987b).

The AP-1 recognition element (ARE) consists of the following consensus sequence: (C/G)TGACT(C/A)A (Angel *et al.*, 1987; Lee *et al.*, 1987). The ARE has the same central pentameric sequence as the GCN4 recognition element (GCRE): TGACT (Hill *et al.*, 1986). During analysis of DNA binding proteins from yeast nuclear extracts, Harshman and coworkers (1988) identified a factor that binds to the SV40 ARE. This factor was present in wild type and *GCN4* deletion mutant strains, demonstrating that this protein was not GCN4. The ARE found in the SV40 enhancer was also able to activate transcription from an heterologous promoter in *S. cerevisiae* strains lacking GCN4. This activation was attributed to a previously unknown yeast transcription factor. Moreover, this factor and AP-1 have nearly identical DNase I, methidiumpropyl-EDTA-iron(II) and DMS "footprints", suggesting that it is the yeast homologue of the mammalian AP-1 and has therefore been designated yAP-1 (Harshman *et al.*, 1988; R.H. Jones *et al.*, 1988). An interesting feature of the DNA binding properties of yAP-1 is that it binds very poorly to the GCRE, despite the fact that the GCRE is almost identical to the ARE. GCN4, on the other hand, can bind to both the GCRE and the ARE with near equal affinities. This would suggest that yAP-1 can discriminate between the GCRE and the ARE *in vivo*, whereas GCN4 cannot. Determination of the amino acid sequence of the gene encoding yAP-1 and comparison of the DNA binding domain with those of others in this class of transcriptional activators, that is GCN4, Jun and AP-1, might shed some light as to which amino acids are responsible for the ability of yAP-1 to discriminate between the GCRE and the ARE.

The human proto-oncogene c-*jun* has recently been isolated (Bohmann *et al.*, 1987) and the DNA sequence of c-*jun* has revealed significant homology to the GCN4 DNA binding domain. Indeed, c-*jun* expressed in *E.coli* produces a protein that will bind to the ARE, strongly suggesting that c-*jun* is the *AP-1* gene. In contrast to AP-1, the structural homology between GCN4 and Jun is limited with virtually no identity in the N-terminal 250 amino acids and only small regions of homology

in the C-terminal 60 amino acid DNA binding domain (Bohnman *et al.*, 1987). Furthermore, the highly acidic amino acid sequence of the transcriptional activation domain of GCN4 is not conserved in JUN. This would suggest that GCN4 and Jun are involved in transcriptional regulation via different control pathways in the cell.

2.4 Repression of Transcription in Yeast

Saccharomyces cerevisiae can exist as one of three distinct cell types: as a non-mating diploid (**a**/α), or as a haploid of either the α or the **a** mating type (Brand *et al.*, 1987). A cell-type specific repressor of *S. cerevisiae*, α2, turns off transcription of the **a** genes by binding to an operator located upstream of each **a**-specific gene. Alpha2 is present in α cells but not **a** cells which means that the **a**-specific genes are turned off in α cells but are expressed in **a** cells.

Several observations suggested to Keleher and coworkers (1988) that α2 is not the only protein involved in repression by the α2 operator. The operator is a 32 bp region of highly conserved sequence (Johnson and Herskowitz, 1985) which contains much more information than would be expected to form the recognition site for a single protein. In addition, the operator appears to contain two different axes of partial 2-fold symmetry which may function as recognition sequences for two different proteins. Another interesting feature is that purified α2 binds to two 13 bp sites at the ends of the operator leaving the centre of the operator accessible to DNase I cleavage (Sauer *et al.*, 1988). The centres of these protected regions are separated by 25 bp, thus positioning them on opposite faces of the DNA double helix. Finally, mutant α2 proteins with small deletions in the N-terminal domain, but which retain the ability to bind to the operator *in vitro,* fail to bring about repression *in vivo* (Hall and Johnson, 1987). These findings suggested that one of the functions of the N-terminal domain of α2 might be to contact a second regulatory protein, bound to the centre of the operator, which is also required for repression. A non cell- type- specific yeast protein, GRM (general regulator of mating type), which recognises the central region of the operator, has been identified (Keleher *et al.*, 1988). GRM binds cooperatively with α2 to the operator in a manner presumably brought about by protein–protein interaction between GRM and the N-terminal domain of α2. (The isolated C-terminal domain of α2 binds the α2 operator but shows no cooperativity with GRM). The arrangement of GRM and α2 on the operator is somewhat unusual

in that GRM binds to the centre of the operator and is flanked on either side by an α2 monomer – a feature which is attributable to the flexible and elongated structure of α2 (Sauer *et al.*, 1988).

In addition to a role in α2 repression, GRM may also function as a transcriptional activator in α cells (Keleher *et al.*, 1988). A yeast protein, PRTF, has been identified by Bender and Sprague (1987). This protein binds, together with the α cell-type-specific activator protein α1, to the α2 operator to turn on the α- specific genes. The binding sites for GRM and PRTF share 10 out of 12 bp identity and like GRM, PRTF is found in all three yeast cell types. In addition, a mutant which has decreased GRM binding activity and α2 mediated repression, also shows decreased expression of an α-specific gene (S. Passmore, unpublished data), a phenotype which would be expected of a PRTF mutant. So it would seem likely that GRM and PRTF are identical, suggesting that in α cells the same protein serves as a co-repressor with α2 and a co-activator with α1.

The GRM may also function as a transcriptional activator in **a** cells. Kronstad *et al.*, (1987) have shown that the α2 operator located upstream of an **a**-specific gene was required for expression of that gene in **a** cells. Similarly, Keleher *et al.*, (1988) have shown that the α2 operator located upstream of the *STE6* gene supplies weak UAS activity and that this activity resides within the central region of the operator. Bender and Sprague (1987) have also arrived at this conclusion based on studies of PRTF in **a** cells. Accordingly it would appear that a single non-cell-type- specific protein, GRM/PRTF, has different functions (that is, activation or repression) depending on the cell-type-specific protein with which it acts in concert.

The two yeast mating type loci *HML* and *HMR*, although they contain complete copies of the mating type genes together with the appropriate regulatory sequences, are transcriptionally inactive or "silent" (Brand *et al.*, 1987). This repression of the two silent mating type loci is brought about by the action of at least four *trans*- acting gene products, SIR1, 2, 3, and 4, as well as *cis*- acting sites located a kilobase or more from the mating type gene promoters (reviewed by Klar *et al.*, 1984). One of the regulatory sites at the *HMR* locus, a "silencer" sequence, exhibits eukaryotic enhancer-like distance- and orientation-independence, but it represses rather than stimulates transcription (Brand *et al.*, 1985). Brand *et al.*, (1987) have defined three short DNA sequences within the silencer that are required for its ability to repress transcription: A, E and B. Mutations in the individual elements do not seriously affect transcriptional control but mutating two out of three of

these elements completely inactivates the silencer suggesting that the silencer is composed of sequences that are able to substitute functionally for one another. In addition to their role in repressing transcription, these sequences are also responsible for the autonomous replication sequence (ARS) activity or the centromere-like segregation functions of the silencer. This implies that DNA replication and segregation, or the proteins involved in these processes, may play pivotal roles in silencing transcription.

The E and B sequences of the silencer contain the binding sites for two proteins which may be involved in silencer function; RAP1 and SBF-B (Shore and Nasmyth, 1987; Brand *et al.*, 1987), but neither of these proteins is the product of an *SIR* gene (Shore *et al.*, 1987). Element E exhibits sequence homology to the UASs of several ribosomal protein genes. Moreover, both the E and B sites can activate transcription from the CYC1 gene (Brand *et al.*, 1987). However, the ability of silencer sequences to activate transcription may be unrelated to their role in transcriptional repression in that activation and repression may be brought about by different proteins which act via the same DNA sequences. On the other hand, factors which bind to the silencer could have roles in several different cellular processes, or be modified in some way by other factors so as to carry out the various functions.

Shore and Nasmyth (1987) have studied the function of RAP1 by affinity purifying the protein and cloning its gene. They were able to demonstrate quite clearly that RAP1 bound to sequences involved in silencer function at *HMR*. They were also able to show that RAP1 bound to the UASs of *MAT* and ribosomal protein genes *in vitro*. Could RAP1 be involved *in vivo* in both activation and repression? The *RAP1* gene has a null phenotype which indicates that the protein is required for cell growth. This finding is therefore consistent with RAP1 having an essential role in activation of ribosomal protein genes, ruling out the possibility that its only involvement is in silencer function. In addition, similarities were revealed between the eukaryotic CTF/NF-1 group of cellular proteins, implicated in both DNA replication and transcription, and RAP1 (Santoro *et al.*, 1988). (A 30 amino acid domain present in the N-terminus of CTF shows 30% identity and 60% sequence similarity with the C-terminal portion of RAP1). If RAP1 has a role in transcriptional activation then its silencer-associated binding sites should, under certain conditions, have UAS activity, This has been demonstrated for both the *HMR* and *HML* binding sites (Brand *et al.*, 1987; D. Shore,

unpublished data) which are both able to activate partially a UAS-deleted promoter. That RAP1 is involved in silencer function is suggested by the observation that the effect of deletion of its normal binding site at the *HMR* silencer can be overcome by the addition of UAS-associated binding sites from ribosomal protein genes (Shore and Nasmyth, 1987). To confirm that RAP1 does have a role in both activator and silencer functions will require the isolation and characterisation of conditional *RAP1* mutants. That SIR proteins bind directly to silencers is a possibility which at the present time cannot be ruled out. However, it is possible that RAP1 acts, not as a transcriptional activator at the silencer, but as a component of a protein–DNA complex unique to silencers which in turn targets them to be acted upon by the SIR proteins (Shore and Nasmyth, 1987). Examination of the possible interactions between RAP1, other silencer binding factors and the SIR proteins should prove to be most interesting.

2.5 Control of gene expression in higher eukaryotes

Eukaryotic transcriptional control elements are generally grouped into two distinct classes of *cis*- acting regulatory sequences, promoters and enhancers, which can overlap both physically and functionally (Serfling *et al.*, 1985). These sequences govern eukaryotic gene expression. Efficient transcription is generally dependent on the correct spatial arrangement of the promoter elements with respect to the mRNA start site (Cochran and Weismann, 1984; McKnight and Kignsbury, 1982; Everett *et al.*, 1983). However, considerably more flexibility is shown by enhancer sequences in that they are able to activate transcription independently of position or orientation from the responding promoter – an effect first observed in DNA sequences present in the small animal DNA tumour virus, SV40 (Banerji *et al.*, 1981; Moreau *et al.*, 1981).

The double stranded DNA viruses that infect animal cells rely heavily on the host cell's transcriptional machinery for expression of their genes, as well as encoding regulatory proteins which have the potential to modify transcriptional specificity. Much is known about the genetics and molecular biology of animal viruses and as such a detailed description of their interaction with the cellular transcriptional apparatus will provide us with an invaluable insight into how gene expression is controlled in animal cells. In addition to the genetic approaches which led to the identification of viral regulatory genes, two strategies have been adopted in studying the control of viral gene expression. In the first of these, DNA sequences which act in *cis* to regulate expression *in vivo* have been identified and analysed using *in vitro* mutagenesis. In the second, soluble cellular extracts capable of accurate transcription *in vitro* have been established and fractionated. Combination of these approaches has begun to shed some light on the molecular interactions between host cell transcription factors and viral regulatory information which account for viral selectivity (reviewed by McKnight and Tjian, 1986).

Gene expression in higher eukaryotes is controlled at many steps, including the accessibility of genes to transcription factors in exposed active chromatin as opposed to condensed inactive chromatin; the rate of transcription initiation on open chromatin; and various post-transcriptional steps (reviewed by Wasylyk, 1988). There are three types of eukaryotic RNA polymerases, I, II and III, which are known to transcribe ribosomal RNA genes, protein coding and small U RNA genes, and 5S and transfer RNA genes, respectively (Sentenac, 1985). The protein coding genes are the

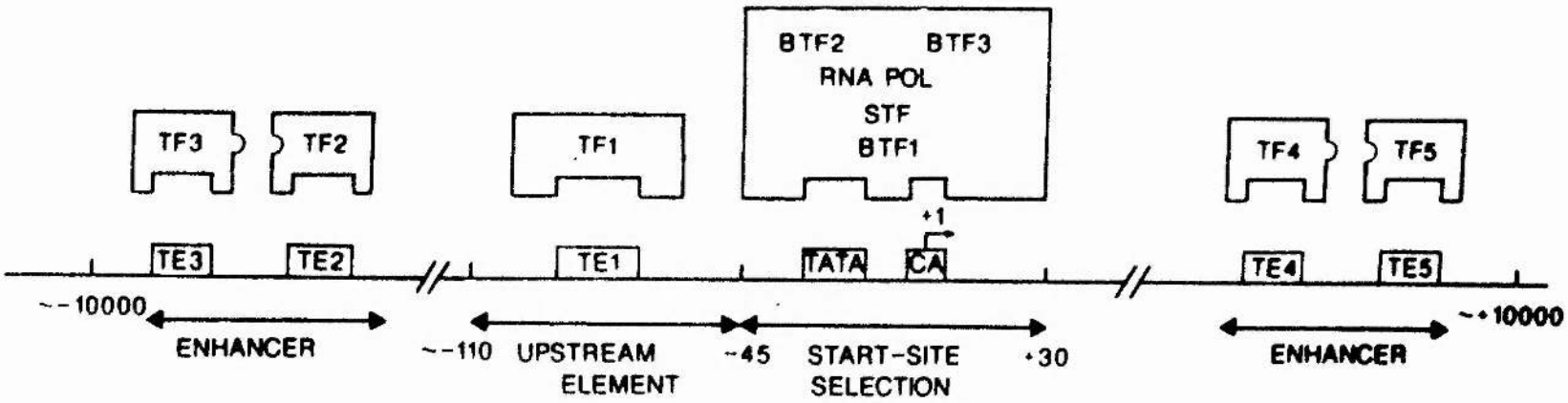
most heterogeneous class and are especially important in that they code for the regulatory proteins e.g., those that control development. RNA polymerase II has been the most extensively studied of the three polymerases and signals controlling the transcription of this class of genes have been defined. To this end, this review will focus on transcriptional regulation of the RNA polymerase II transcribed genes.

In general, promoter sequences in RNA polymerase II transcribed genes can be subdivided into those which select the site of initiation and those which modulate the specific rate of initiation e.g., upstream elements and enhancers. The mRNA start site selection region extends from about -45 to +30 and contains two highly conserved DNA sequence motifs, the TATAAA sequence at around -30 and the CA at +1 (see Figure 12). General transcription factors bind to this region and direct RNA polymerase II to the specific start site (reviewed by Wasylyk, 1988). The same general factors are probably involved in specific initiation from most protein coding genes. Promoter mapping studies of several eukaryotic genes have disclosed the importance of a GGGCGG hexanucleotide (GC box) and a CCAAT sequence (CAAT box) which are often found 40–100 bp upstream of the initiation site. These upstream elements appear to be critical in directing efficient transcription from certain mammalian promoters, e.g., the SV40 early promoter, mammalian ß-globin promoters and the herpes simplex virus (HSV) thymidine kinase (TK) promoter (reviewed by Kadonaga *et al.,* 1986).

In most protein coding genes the short DNA sequences of ~10 to 20 bp (transcription elements), which lie upstream of the mRNA start site, serve as the recognition sites for a variety of *trans* -acting factors. Promoters can be constructed from a large number of different transcription elements and transcription factors. The factors can vary in activity in response to a variety of stimuli, e.g., hormones, heat, phorbol esters, etc., or from cell type to cell type (reviewed by Wasylyk, 1988). This combinatorial organisation of eukaryotic promoters could therefore lead to quite convoluted regulatory patterns involving only a small number of transcription factors. How a given combination of these factors leads to a particular pattern of gene regulation is only now beginning to be understood.

Figure 12. Model of a eukaryotic RNA polymerase II transcribed gene promoter

BTF1 (TFIID) binds specifically to DNA in the mRNA start site selection region, possibly in association with STA (TFIIA). This protein–promoter complex is recognised by RNA polymerase II perhaps in combination with BTF2 (TFIIE) or BTF3 (TFIIB). Various other transcription factors (TFs) bind to specific DNA sequences or transcription elements (TEs), which can be located up to 10,000 bp upstream or downstream from the initiation site (+1). It is common to find at least one TE immediately upstream from the start site selection region (from -45 to -110). Efficient distal activation by enhancer elements appears to require the association of several (TFs) (Wasylyk, 1988).


BTF2
BTF3
RNA POL
STF
BTF1
TF3
TF2
TF1
TF4
TF5
+1
TE3
TE2
TE1
TATA
CA
TE4
TE5
~-10000
~+10000
ENHANCER
~-110
UPSTREAM
ELEMENT
-45
START-SITE
SELECTION
+30
ENHANCER

2.5.1 The Eukaryotic Promoter and Transcription Initiation

Two types of eukaryotic promoter can be distinguished: those in which the TATA and mRNA start site both make significant contributions to promoter strength; and those in which either region will have a predominant effect. The former situation is illustrated by the adenovirus major late promoter (MLP) in which point mutations in either the TA motif, or the mRNA start site, decrease transcription, whereas mutation of both these sequences has a synergistic effect resulting in the elimination of transcription (Concino *et al.*, 1984). In the SV40 early promoter, which has no TATA homology sequence, the start site regions are the primary determinants of transcription initiation. This has been demonstrated by the effects of small deletions or insertions upstream from the start sites which did not alter the locations of the 5'-ends of the RNA upon infection with the mutant viral strains *in vivo* (Somasekhar and Mertz, 1985). In addition, it is now evident that many promoters, particularly those of house-keeping genes, lack TATA boxes and are instead composed of GC-rich elements that are often located within methylation-free islands (Bird, 1986). Transcription from these promoters often initiates at multiple sites. It has been demonstrated that TATA-less promoters, such as the adenovirus E2a and SV40 late promoters, form complexes with the same purified transcription factors as other TATA box-containing promoters. However, the affinity of these factors is somewhat lower for the non-consensus promoters (reviewed by Wasylyk, 1988).

Purification of transcription factors from crude cell extracts has shown that at least four factors, as well as RNA polymerase II, are required for specific transcription initiation (Figure 12). Unfortunately these factors have been assigned different names. However, STF, BTF1, BTF2 and BTF3 (Moncollin *et al.*, 1986) are equivalent to TFIIA, TFIID, TFIIE and TFIIB, (Reinberg and Roeder, 1987; Reinberg *et al.*, 1987) respectively. Using highly purified preparations, it has been shown that two of these factors TFIIA and TFIID, act first to form a stable pre-initiation complex on the -45 to +30 region of the promoter which in turn commits the template to transcription. TFIIA does not bind directly to DNA but appears to act by stabilising the interaction between TFIID and the promoter, via protein–protein interactions (Fire *et al.*, 1984; Elgy *et al.*, 1984). At a later step in initiation, involving RNA polymerase II, TFIIE and TFIIB associate with the pre-initiation complex to help form a complex which mediates rapid initiation of RNA synthesis (Moncollin *et al.*, 1986; Reinberg and Roeder, 1987). TFIIE and TFIIB do not bind specifically to DNA. Several observations suggest that they

interact with other factors, including RNA polymerase II again via protein–protein interactions (reviewed by Wasylyk, 1988).

An increasing of evidence suggest that regulation of transcription initiation may be exerted directly on the general transcription factors. For example, the adenovirus Ela gene products, produced early in viral infection, can *trans*- activate transcription from a variety of promoters. The TA motif has been shown to be the target for Ela *trans*- activation in some cases (Berk, 1986). This motif also appears to be the target for *trans*- activation of the HSV TK gene by viral immediate early gene products (Coen *et al.,* 1986). The slow formation of the pre-initiation complex may also be a target for regulation of initiation. In the adenovirus MLP a specific factor binds to DNA sequences upstream of the TATA box and binding of this factor increases the rate of pre-initiation complex formation (Sawadogo and Roeder, 1985). *In vivo* experiments on the *Drosophila hsp70* gene have suggested that later steps in the initiation reaction may also be transcriptional control points. In the non-induced, non-transcribed state, the *hsp70* gene has protein bound to the -40 to +60 region of the promoter. RNA polymerase II is also associated with this promoter. Heat shock rapidly stimulates transcription through association of heat shock transcription factor (HSTF) with the upstream region. A simple interpretation of these results suggests that an inactive, pre-initiation complex is present on the promoter, and HSTF binding activates this complex (Wu, 1984; Parker and Topal, 1984).

Studies on the mechanism of basal transcription have mainly been carried out *in vitro.* However, the start site selection region generally requires additional elements for efficient transcription *in vivo.* These elements are probably the major determinants of the efficiency of transcription and may serve to activate a segment of DNA, thereby facilitating interaction of general transcription factors (reviewed by Wasylyk, 1988). The factors are most likely to be responsible for subsequent interactions which determine both the precise site(s) of initiation and the final efficiency of transcription.

Upstream promoter regions and enhancer elements can be distinguished functionally by the inability of the upstream elements to stimulate transcription when distanced from the start site selection region. Enhancers, on the other hand, are able to stimulate transcription at some distance from the responding promoter, raising the possibility that different transcription factors interact with

upstream, or enhancer, elements to activate transcription via different mechanisms. However, many transcription factors are capable of both short- and long-range activation suggesting that such a distinction may not actually exist (reviewed by Wasylyk, 1988).

Many factors which interact with elements in upstream regions are able to stimulate transcription from distantly located promoters. For example, in the SV40 early gene promoter (see "Results" Figure 1.1) there are binding sites, located in the 21 bp repeats, for the transcription factor Sp1. However, this upstream element is considerably less efficient at stimulating transcription when moved away from the transcription start site (Moreau *et al.*, 1981). The 21 bp repeats are also part of the divergently transcribed late promoter and it has been shown that Sp1 stimulates transcription from the late start sites located some 200 bp away (reviewed by Wasylyk, 1988). In addition, it has been reported that the Sp1 binding sites in this upstream element can behave like a transcriptional enhancer if the concentration of Sp1 is increased tenfold (unpublished, R. Tjian, 13th Harden Conference, 1988). Sp1 binding sites have been located up to 250 bp upstream from the start site of different eukaryotic promoters , including the HSV TK gene, HSV IE3 gene, HIV LTR promoter, mouse dihydrofolate reductase gene and the human metallothionein II_A gene (reviewed by Kadonaga *et al.*, 1986).

The immunoglobulin (Ig) octamer motif (ATGCAAAT) is present in both upstream elements (e.g., IgH and Igk promoters) and enhancers (e.g., SV40, IgH and *Xenopus* U2 gene; for review see Falkner *et al.*, 1986). Indeed B-cell-specific transcription of Ig genes is dependent on the presence of an intact octamer motif (Gerster *et al.*, 1987; Wirth *et al.*, 1987). These sequence elements are recognised by the B-cell specific octamer binding transcription factor OTF-2 (Landolfi *et al.*, 1986; Staudt *et al.*, 1986). This element is also an important functional component of many ubiquitously expressed genes where it is recognised by one or more ubiquitous octamer binding proteins, e.g., NFIII (Pruijn *et al.*, 1986), NF-AI (Singh *et al.*, 1986), OTF-I (Fletcher *et al.*, 1987) and OBP100 (Sturm *et al.*, 1987). An octamer motif is present in each 72 bp repeat of the SV40 enhancer and in its natural promoter context has been shown to be active only in lymphoid cells, but not in other non-lymphoid cells tested so far (Davidson *et al.*, 1986; Nomiyama *et al.*, 1987). This observation has been strengthened by the results of Schirm *et al.*, (1987), in which a multimerised domain A of the Sv40 enhancer (see "Results", Figure 1.2), which contains the octamer motif, was active only in

lymphoid cells. A single copy of this element did not appear to efficiently activate transcription of a linked gene at a distance. However, on multimerisation of this motif, transcription was enhanced (Gerster *et al.*, 1987). Generation of enhancer activity by duplication is now widely recognised (Herr and Gluzman, 1985; Weber *et al.*, 1984; Schirm *et al.*, 1987) suggesting that a large number of transcription elements have the potential to activate transcription at a distance.

2.5.2 Eukaryotic Promoter Organisation: The SV40 Early promoter

Many of the double stranded DNA viruses that infect animal cells express their genetic information in temporal cascades that are regulated by defined genetic loci. SV40 provides us with such an example (for a review see Tooze, 1982). The SV40 genome is a double stranded, covalently closed, circular molecule of 5243 bp. The genome can be functionally divided into two divergently transcribed regions: the early region which is transcribed throughout the virus lytic cycle and in transformed cells; and the late region which is transcribed only in permissive cell types after the onset of viral DNA replication (Figure 13). The primary transcripts from both regions are differentially spliced generating two early mRNAs which code for the early proteins, large T antigen (T antigen) and small t antigen (t antigen), and at least two mRNAs which code for the late capsid proteins VP1, VP2 and VP3 (reviewed by Hamer and Khoury, 1983). The transcriptional control elements of the virus lie within the 400 bp region between the early and late coding sequences (see expanded portion of Figure 13). Studies have centred on the role of this region in the initiation of early transcription which occurs primarily at the site marked "Early E" (nucleotides 5232–5236) in the early stage of lytic infection and in transformed cells. At late times during lytic infection, the major cap site shifts approximately 40 bp upstream to position "Early L" (nucleotides 28–34) (reviewed by Hamer and Khoury, 1983).

The SV40 early promoter has been the subject of intense investigation; deletions insertions, base substitutions and sequence rearrangements have been introduced into the SV40 chromosome and their effects analysed by a variety of assays, including virus propagation, morphological transformation, T antigen synthesis, as well as mRNA synthesis *in vivo* and *in vitro* (reviewed by McKnight and Tjian, 1986).

The 300 bp stretch lying between the early and late transcription units contains a number of

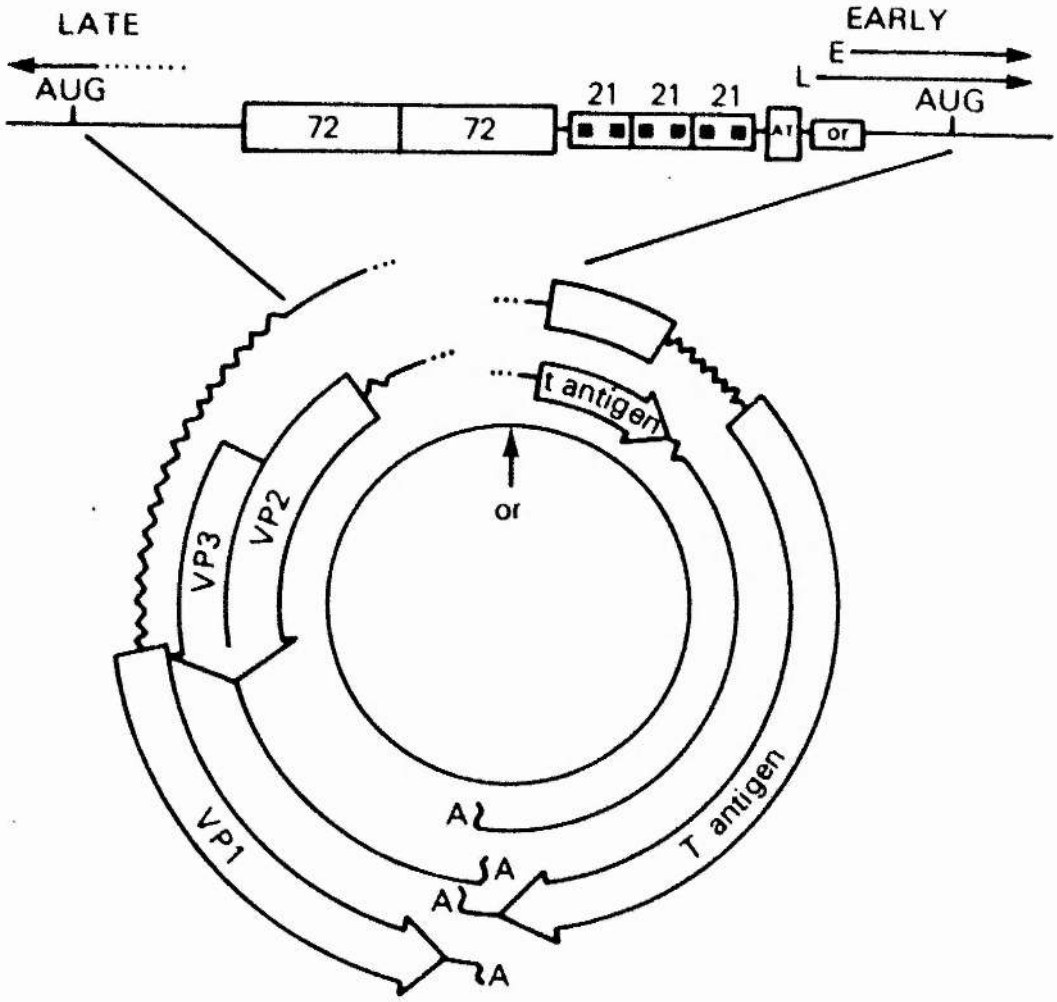


Figure 13. Transcription map and control region of SV40

The location of the origin of DNA replication (or) and the general structures of the predominant transcripts are indicated. The expanded top part shows the viral control sequences lying between the early and late AUG initiation codons. This region contains a number of early transcriptional regulatory elements, including an AT-rich sequence (the TATA box), three 21 bp repeats and two 72 bp repeat sequences (Hamer and Khoury, 1983).

regulatory DNA sequences, and a number of sequence-specific protein–DNA interactions are known to take place within this transcriptional control region. The origin of DNA replication is located within a 70 bp segment containing three binding sites for T antigen (Myers and Tjian, 1980), as well as the mRNA initiation sites from the early promoter (Khoury *et al.*, 1973). Perturbation of the DNA replication origin affects transcriptional control. Elimination of T antigen binding sites prevents initiation of early transcription. In addition, inactivation of the origin severely limits the number of templates available for late mRNA synthesis, resulting in a reduction of late gene expression (reviewed by McKnight and Tjian, 1986).

Directly adjacent to the replication origin is a stretch of DNA associated with defining the locations of the early mRNA start sites. Within the mRNA start site selection region there are several major and some minor start sites for transcription from the early promoter which are used both early and late in viral infection. About 30 bp upstream from the major EES there is a TATA-like motif within an A-T stretch of DNA. However, deletions in this region, and point mutations in the TATA motif, have only about a five-fold effect on transcription efficiency suggesting that this region contributes minimally to overall promoter efficiency (Benoist and Chambon, 1981; Mathis and Chambon, 1981; Wasylyk *et al.*, 1983). It may be that the A-T rich sequences in the early promoter can act as alternative TATA motif recognition sequences for general transcription factors resulting in initiation from the minor start sites (reviewed by Wasylyk, 1988).

Three direct repeats of a GC-rich 21 bp sequence lie in the upstream region of the early promoter. Within each repeat element are 2 copies of the hexanucleotide sequence 5'-GGGCGG-3' (GC motif). Deletion of all three 21 bp repeats severely inhibits promoter function, as well as virus viability (Myers *et al.*, 1981; Fromm and Berg, 1982; Everett *et al.*, 1983). However, it is interesting to note that viral mutants which lack up to four copies of the GC motif can still be propagated. This effect is probably due to the redundant nature of the 21 bp repeats (McKnight and Tjian, 1986). The GC motifs in the 21 bp sequences bind the transcription factor Sp1 (Barrera-Saldana *et al.*, 1985; Kadonaga *et al.*, 1986; McKnight and Tjian, 1986), which was isolated from HeLa cells and shown to activate transcription from the SV40 early promoter (Dynan and Tjian, 1983). Of the GC boxes that bind Sp1, boxes III and V have the strongest affinity for Sp1, II and VI have intermediate affinity, I is a weak binding site, with box IV being essentially inaccessible to Sp1 when GC box V is

occupied. Mutational analysis has indicated that interaction of Sp1 with three of the binding sites, GC boxes I, II and III, is largely responsible for mediating early transcription, whereas binding to boxes IV, V and VI predominantly affects late transcription (Gidoni *et al.*, 1985). Sp1 stimulates both early and late transcription whose start sites lie at different distances, and in opposite orientations, to one another. Moreover, Sp1 binding sites are found interspersed with other transcription elements, suggesting that it can interact with other transcription factors in addition to those found in the SV40 early promoter (reviewed by Wasylyk, 1988). However, Takahashi and coworkers (1986) have demonstrated that stereospecific alignment with the start site selection and enhancer regions is required for maximal transcription efficiency. They altered the spacing between the three SV40 early promoter elements by inserting either odd or even multiples of one half turn of the DNA helix and could show that marked differences in transcriptional activation *in vivo,* of these two types of insertions on initiation from the early promoter. This suggests that Sp1 probably mediates this effect via protein-protein interactions.

Finding six closely spaced, but apparently independent Sp1 binding sites has suggested how the bidirectionality of these promoter elements may function (Gidoni *et al.*, 1985). For example, the binding site closest to the transcription start site has the greatest influence on activation of RNA synthesis for a given direction of transcription, suggesting that the functional directionality of the 21 bp repeats could actually be due to multiple directional elements. Alternatively, each individual Sp1 binding sequence would have the ability to activate transcription in a position and orientation independent manner with the direction of transcription being determined by the position of the Sp1 binding site in the context of the other promoter elements. Bidirectional stimulation of transcription by Sp1 could occur by virtue of Sp1 having two symmetrically disposed interaction sites. The interaction could take place by having two functional domains linked by a flexible arm, one which binds DNA, and another which interacts with other transcription factors as seen in repressor proteins of *E. coli* phage lambda (reviewed by Ptashne, 1986b). Kadonaga and coworkers (1987) isolated a cDNA that encodes human Sp1 and by expression of truncated fragments of Sp1 in *E. coli* were able to localise the DNA binding activity to the C-terminal 168 amino acid residues. The DNA binding domain was found to harbour three contiguous "Zn finger" motifs (Miller *et al.*, 1985; Brown *et al.*, 1985) which are believed to be metalloprotein structures that interact with DNA. Indeed,

purified Sp1 required Zn(II) for sequence-specific binding to DNA suggesting that Sp1 interacts with DNA by binding of the "Zn fingers" (Kadonaga *et al.*, 1987).

Two additional tandem repeats are positioned towards the "late" side of the SV40 transcriptional control region (see "Results", Figure 1.1). Extensive mutagenesis, or removal of both these 72 bp repeats, impairs virus viability and reduces early transcription. However, deletion of only a single copy of the repeat does not, to a large extent, affect virus growth or early promoter function (reviewed by McKnight and Tjian, 1986). The 72 bp repeats are also able to potentiate early transcription regardless of their orientation. They can activate transcription when placed thousands of bp upstream from a promoter element, or downstream from a transcription unit. It should be stressed, however, that this enhancing activity is diminished at distant locations and that even a 5 bp displacement of the 72 bp sequence decreases promoter efficiency six- to ninefold (Takahashi *et al.*, 1986).

In addition to housing elements critical for early promoter function, the 72 bp repeats also contain determinants of the late promoter which appears to be the major contibuter to expression from the late promoter. A series of deletion mutants identified two domains within the enhancer element which contain sequences important for efficient late transcription. Both domains I and II function after DNA replication, and sequence motifs within these domains, shown to be important for late promoter activity, also corresponded to sequence motifs important for enhancer activity. May and coworkers (1987) carried out their deletion analysis in the absence of T antigen. They suggest that the change observed in the regulation of late genes, coincidental with initiation of DNA replication, may be due to T antigen binding to its three natural sites which could directly, or indirectly, modify some interactions between the different *trans*- acting factors bound to promoter elements. This in turn could lead to down regulation of the early promoter and to an up regulation of late gene expression.

2.5.3 Enhancer Elements

Enhancers are *cis*- acting DNA sequences which constitute an important part of the transcriptional control apparatus of eukaryotic cells. The main feature of an enhancer which distinguishes it from other *cis*- acting transcriptional control elements, e.g., promoter sequence

elements, is its ability to activate transcription of a linked gene in a position- and orientation-independent manner (Guarente, 1988). Enhancers have been identified in many different viruses that in experimental constructions fulfilled the following criteria:

1. strong transcriptional activation of a linked gene from its correct initiation (cap) site:

2. activation of transcription independently of the orientation of the enhancer element:

3. the ability to function over large distances (>1000 bp) from a position upstream or downstream relative to the initiation site: and

4. in cases examined, to preferentially stimulate transcription from the most proximal of two tandem promoters (reviewed by Schaffner, 1986 and Muller *et al.,* 1988).

The first cellular enhancer was discovered in the Ig heavy chain gene (deVilliers *et al.,* 1983; Gillies *et al.,* 1983; Neuberger, 1983). This was of great interest in the studies of eukaryotic gene regulation in that it showed that remote gene control by the enhancer effect was not simply a viral peculiarity; stimulation of transcription by the Ig enhancer was cell type-specific (the first genetic element described to have such a property); and, this enhancer element was found to be located within the Ig gene, downstream of the cap site (reviewed by Schaffner, 1986 and Muller *et al.,* 1988).

In general, DNA sequences with enhancer activity, are located -100 to -300 bp upstream of the transcription unit. They can however, be found in other positions, such as within introns (e.g., the Ig heavy chain gene enhancer) or between two genes. It is evident that upstream promoter components and enhancers overlap physically and functionally, implying that an enhancer is not so much a physical entity with clearly defined boundaries, but rather that there exists an "enhancer effect" (reviewed by Schaffner, 1986). This effect can be exerted by combinations of a variety of DNA sequence motifs.

The minimum enhancer activity of SV40 can be attributed to one copy of the 72 bp repeat plus an additional 30 bp of DNA located on the late side of the repeated element (Zenke *et al.,* 1986). It is possible to dissect this region into two large domains, A and B (see "Results". Figure 1.2), which in isolation have minimal enhancer activity, but which in combination act synergistically to generate the substantial enhancement of transcription typical of the complete enhancer (Zenke *et al.,* 1986). Deletion/duplication events within each large domain have identified small DNA sequence motifs

that are important for enhancer function (Swimmer and Shenk, 1984; Weber *et al.*, 1984; Herr and Gluzman, 1985; Herr and Clarke, 1986; Fromental *et al.*, 1988). In general, single copies of these motifs tend not to show enhancer activity but when multimerised, or combined with other motifs can function as *bona fide* enhancers (Ondek *et al.*, 1987; Schirm *et al.*, 1987). Each of these DNA sequence motifs appears to represent the binding site for at least one *trans*-acting factor (Wildeman *et al.*, 1986) and it is thought that formation of a nucleoprotein complex containing multiple *trans*-acting factors arrayed on the the enhancer is required for efficient stimulation of transcription.

It is now clear that individual motifs have the potential to bind multiple *trans*-acting factors and this is exemplified in the region of the enhancer "core" sequence (5'-GTGGA/TA/TA/TG-3') identified by comparison of DNA sequences from a number of different enhancers (Laimins *et al.*, 1982; Weiher *et al.*, 1983), where a variety of proteins bind to overlapping DNA recognition sites. Three proteins, one of which was cell type specific, were shown to bind to sites in the GT-I or "core" region of the SV40 enhancer (Xiao *et al.*, 1987) (see "Results". Figure 1.2). Although present in crude extracts these proteins could be distinguished from one another by their ability to bind to a variety of mutated templates and by their unique methylation interference patterns (Xiao *et al.*, 1987). Purification of factors which bind to this region has led to the identification of EBP20, a 20,000 molecular weight species present in rat liver (Johnson *et al.*, 1987) and AP-3, a protein derived from HeLa cell nuclei (Mitchell *et al.*, 1987). To date the relationships, if any, between these proteins has not been established.

Although the SV40 enhancer has been shown to be active in a variety of cell types, it also demonstrates distinct cell type specificities. Schirm *et al.*, (1987) demonstrated that the SV40 enhancer can be dissected into multiple segments, each with a different cell type specificity. They suggest that a particular cell type specificity is typical for the individual segments, and that enhancers of differing specificity could be assembled by combining the individual sequence motifs in different ways. Activation of transcription by individual DNA sequence motifs is cell type specific *in vivo* (Ondek *et al.*, 1987; Nomiyama *et al.*, 1987; Fromental *et al.*, 1988) and this is paralleled by cell type specificity in the binding of cellular factors to these motifs *in vivo* (Davidson *et al.*, 1986; Xiao *et al.*, 1987a).

Nomiyama and co-workers (1987) have proposed that the activity of an enhancer, in a given cell

type, results from the nature of its constituent motifs as well as the presence or absence of the appropriate specific *trans*- acting factors in that cell type. Thus in complex enhancers, such as that present in SV40, a situation exists whereby gene expression may be controlled by the binding of different combinations of interacting factors . Fromental and coworkers (1988) have investigated the cell-type specificity of individual SV40 enhancer motifs, the effects of spacing between motifs and the synergism between combinations of motifs. Their results showed the existence of three distinct classes of sequence motifs which interact with different classes of cell-type specific enhancer binding factors, but which in isolation do not demonstrate enhancer activity. This demonstrates three distinct levels of functional organisation the first of which is the DNA sequence motif itself, which represents the binding site for a particular class (A, B, C or D) of *trans*- acting factor. The second level of organisation comprises of the minimum enhancer element which when oligomerised, or combined with other enhancer elements, acts synergistically to generate enhancer activity. The third level exists when a single minimal enhancer element is oligomerised or combined with two or more similar elements – a situation which is apparent in the SV40 enhancer (Fromental *et al.,* 1987) and the Ig heavy chain gene enhancer (Leonardo *et al.,* 1987; Perez-Mutul *et al.,* 1988). In one category (functional level 1, class A), a HeLa cell factor, TEF-1 (Davidson *et al.,* 1988), binds specifically to two apparently unrelated sequence motifs (GT-IIC and Sph) (see "Results, Figure 1.2) in the SV40 enhancer. This factor binds cooperatively to DNA templates containing tandem repeats of its cognate motifs which correlates with the ability of the tandem repeats to generate enhancer activity *in vivo.*

Transcriptional activation in response to extracellular stimuli is now a well established phenomenon. The cellular transcription factor AP-1 interacts with DNA sequences common to both SV40 and the human metallothionein II_A ($hMTII_A$) gene enhancers (Lee *et al.,* 1987a), both of which are inducible in cells treated with the tumour promoter 12-O-tetradecanoyl-phorbol-14 acetate (TPA). It is thought that AP-1 is involved in mediating both the basal level of transcription and specific response to induction by TPA from these enhancer elements (Lee *et al.,* 1987b). In the case of the transcription factor NF-kB, which binds to a site in the Ig kappa light chain gene enhancer (Sen and Baltimore, 1986a) and whose activity is normally restricted to B-cells, stimulation of the factor occurs in other cell types after treatment with agents such as bacterial

lipopolysaccharides and phorbol esters (Sen and Baltimore, 1986b). In mammalian cells, transcription of several genes can be rapidly and transiently activated following serum stimulation. A short DNA sequence element, the serum response element (SRE), has been identified which binds the serum response factor (SRF) and mediates transcriptional activation of *c-fos* and cytoskeletal actin genes (Treisman, 1985; Gilman *et al.*, 1986). The precise role of the SRE in *c-fos* transcription has not been defined but it is probable that activation of *c-fos* gene expression involves an increase in the DNA binding properties of SRF and modification of its ability to activate transcription. In addition, exposure of cells to elevated temperatures leads to transcriptional activation of the heat shock genes, again mediated by a short DNA sequence, the heat shock element (HSE), which is the binding site for the heat shock transcription factor (HSTF) (Shuey and Parker, 1986; Sorger *et al.*, 1987; Weiderrecht *et al.*, 1987). HSTF has been shown to be regulated differently in yeast and HeLa cells. In HeLa cells, binding activity is detectable only after heat shock whereas in yeast cells, the HSTF appears to bind constitutively to DNA and is thought to activate transcription after heat-induced phosphorylation (Sorger *et al.*, 1987).

Transcriptional activity may also be modulated by the ability of bound proteins to exclude binding of other antagonistically acting factors. An example of this type is provided by the inducible enhancer element of the human ß-interferon gene. It is thought that prior to induction a repressor protein bound to the interferon gene regulatory element (IRE) excludes the binding of a transcriptional activator (Goodbourn *et al.*, 1985). Mutational analysis has indicated that the IRE contains a constitutive transcription element, PRDII, that is prevented from functioning in uninduced mouse C127 cells by an adjacent or overlapping negative regulatory element. The PRDII region is related to the SV40 enhancer "core" sequence and contributes significantly to the high basal level of gene expression when the negative regulatory element is deleted. The inability to observe "down" effects on transcription in uninduced cells when this region is mutagenised suggests that access of transcription factors to this element may be blocked by negative regulatory factors (Goodbourn and Maniatis, 1988).

In addition to their role in stimulation of transcription, enhancers have also been demonstrated to have a *cis*- acting involvement in Polyoma (Py) virus DNA replication (deVilliers *et al.*, 1984, and references therein). Only DNA molecules which have both the Py virus origin of DNA replication

and adjacent (but non-overlapping) sequences within the enhancer region can replicate in mouse COP-5 cells (Py transformed mouse fibroblasts), even though these cells constitutively express the viral T antigen (shown to be essential for virus DNA replication) (reviewed by deVilliers *et al.*, 1894). DeVilliers and coworkers (1984) constructed viral hybrid genomes in which the Py enhancer was replaced by other such DNA fragments in an attempt to identify heterologous sequences which might substitute for the replicatory activation component. What their results demonstrated was that Py virus DNA replication could be activated by enhancer sequences of both viral (Py and SV40) and cellular origin (Ig heavy chain gene). Deletion analysis of the Py enhancer region has identified a 23 bp DNA segment which exhibits minimal replicatory ability. However, tandem duplication of this fragment restores full activation of DNA replication while further oligomerisation generates an efficient transcriptional enhancer (Veldman *et al.*, 1985). If enhancers function by localising DNA at the nuclear matrix, or by altering the DNA conformation, it is possible that such molecular mechanisms could also be involved in DNA replication or transcriptional activation. In addition, transcription could be directly involved in DNA replication (reviewed by deVilliers *et al.*, 1984).

Studies on SV40 DNA replication have revealed that transcription elements do function as components of eukaryotic origins of DNA replication. Deletion of the 72 bp repeated elements does not affect DNA replication provided that at least three copies of the GC-boxes in the 21 bp repeats are present. Elimination of all six GC-boxes reduces DNA replication two- to threefold *iv vivo*. If the enhancer elements are also eliminated then a 10- to 100-fold reduction in DNA replication is observed (reviewed by DePamphilis, 1988). It has been suggested that the transcription elements facilitate binding of replication proteins to the core origin region but that this is not achieved by promoting transcription through the core origin. It may be that the role of transcription elements in DNA replication is to facilitate the interaction of replication proteins which bind to transcription elements. For, example, transcription factor AP-2, which binds to enhancer sequences of SV40 and stimulates RNA synthesis *in vitro*, also binds to T antigen (Mitchell *et al.*, 1987). AP-2 binds to sequences in the early promoter recognised by Sp1 and T antigen, and juxtaposed binding sites for AP-2 and SP1 in the 21 bp repeats may facilitate productive interaction between transcription factors (Mitchell *et al.*, 1987). Autoregulation of SV40 early transcription *in vitro* depends, to a large extent, on binding of T antigen to sites in the early promoter region. The

findings of Mitchell *et al.*, (1987) suggest that an additional component of this autoregulation could involve direct repression by T antigen of AP-2. Alternatively, proteins that recognise transcription elements may stabilise binding of DNA replication initiation proteins to the origin by promoting local unwinding of the DNA (DePamphilis, 1988)

In summary, it is likely that enhancer-mediated transcriptional control requires the formation a large nucleoprotein complex whereby a limited set of proteins interact with their recognition sites on the genome, and with the transcriptional machinery, to initiate RNA synthesis at the promoter, which may be some distance from the enhancer element. Understanding the mechanism of enhancer action will require the reconstruction of this process *in vitro*. Purification of individual transcription factors and detailed information on their interaction with DNA will clearly be required to attain this goal.

MATERIALS AND METHODS

1. Cells and tissue culture medium

HeLa S3 spinner cells (WS line) were grown in suspension in Earle's minimal essential medium (Gibco Europe Ltd., Paisley, Scotland) containing 50 units/ml of penicillin (Glaxo Labs. Ltd., Greenford, England), 50 µg/ml of streptomycin (Evans Medical Ltd., Greenford, England), 2.2 g/l sodium bicarbonate and 5% new born calf serum (Sera-lab Ltd., Sussex, England). HeLa S1 monolayer cells were grown in Glasgow modified minimal essential medium (Gibco)containing penicillin, streptomycin sodium bicarbonate as described above for HeLa S3 cells, plus 10% foetal calf serum (Sera-lab Ltd.).

2. Bacterial strains

All plasmids were grown in *E. coli* K12 strain DH1 (F^-, *rec* A1, *end* A1, *gyr* A96, *thi*-1, *hsd* R17 (r^-_K, m^+_K), *sup* E44, λ^-; Maniatis *et al.*, 1982).

Bacteriophage M13mp8 was propagated in *E. coli* K12 strain JM101 (*lacpro* ^{del}, *sup* E, *thi*, F' *tra* D36, *pro* AB, *lac* Iq, ZM15^{del} ; Messing, 1979).

3. Bacterial culture medium

Bacteria were propagated in Luria Broth (LB); 10 g/l bacto-tryptone (Difco Labs., Detroit, USA) and 5 g/l yeast extract (Difco) in 10 mM NaCl pH 7.5, sterilised by autoclaving. Agar plate cultures contained LB supplemented with 1.5% Agar (Difco) and $MgSO_4$ to a final concentration of 10 mM. Culture media were supplemented with 100 µg/ml ampicillin (Sigma) where appropriate.

4. Buffer solutions

The following buffers were routinely used in this study:

TE : 10 mM Tris.HCl, 1 mM EDTA, pH 8.0

TEN : 10 mM Tris.HCl, 1 mM EDTA, 100 mM NaCl, pH 8.0

TEN_{300} NaN_3 : 10 mM Tris.HCl, 1 mM EDTA, 300 mM NaCl, 0.02% NaN_3, pH 8.0

TBE : 100 mM Tris.Borate, 1 mM EDTA, pH 8.3

TGS : 25 mM Tris.HCl, 200 mM glycine, 0.1% SDS, pH 8.3

Hypotonic buffer : 25 mM HEPES.NaOH, 5 mM KCl, 0.5 mM $MgCl_2$, 0.5 mM DTT (Sigma), 0.5 mM PMSF (Sigma), pH 7.5

Isotonic buffer : 25 mM HEPES.NaOH, 5 mM KCl, 0.5 mM $MgCl_2$, 0.5 mM DTT, 0.2M sucrose, 1 mM PMSF, pH 7.5

Buffer A : 25 mM HEPES.NaOH, 1 mM EDTA, 1 mM DTT, 20% glycerol, 0.1% NP40, 1 mM PMSF, pH 7.5

Binding buffer : 25 mM HEPES.NaOH, 1 mM EDTA, 1 mM DTT, 10% glycerol, 0.05% NP40, 100 mM NaCl, 100 µg/ml BSA, pH 7.5

5. Reagents

Reagents were purchased from May & Baker Ltd., Dagenham, Essex, England and BDH Chemicals Ltd., Broom Road, Poole, Dorset, England (Analar grade), except where otherwise specified.

6. Restriction endonucleases, DNA and RNA modifying enzymes

Enzymes were obtained from Northumbria Biologicals Ltd., Cramlington, Northumberland, England; Boehringer Mannheim Ltd., Lewes, East Sussex, England; or New England Biolabs, Bishop's Stortford,Herts., England, and used as specified by the manufacturer.

7. Preparation of plasmid DNA

(i) Preparation of competent cells

Competent *E. coli* cells were prepared essentially as described by Hanahan (1983). Colonies were picked off a fresh streak of cells and resuspended in 6.5 ml SOB (2% bacto-tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$). The culture was maintained at 37°C in an orbital incubator (275 r.p.m.) until the cell density was ~4–7x10^7/ml (OD_{550} = 0.45–0.55), about 2–2.5 hours. Cells were placed on ice for 2 minutes, collected by centrifugation at 2,500 r.p.m., 4°C, for 10 minutes and resuspended in 2 ml FSB (10 mM KOAc, 100 mM KCl, 45 mM $MnCl_2.4H_2O$, 10 mM $CaCl_2.2H_2O$, 3 mM $HACoCl_3$), placed on ice for 10 minutes then pelleted at 2,500 r.p.m., 4°C, for 10 minutes. The pellet was resuspended in 0.5 ml FSB. Freshly prepared

DMSO was added to 0.5%, mixed, and the suspension allowed to stand on ice for 5 minutes. This DMSO treatment was repeated and incubation on ice continued for a further 5 minutes. Competent cells were dispensed into 200 µl aliquots, quick frozen in a solid CO_2/ethanol bath, and stored at -70°C until required.

(ii) Transformation of bacteria

Competent cells were transformed by the method of Hanahan (1983), **a)**, or the "one minute" transformation method of Golub (1988), **b)**.

a) Plasmid DNA (1–10 ng) was added to 200µl competent *E. coli* cells and incubated on ice for 30 minutes prior to heat-shock at 42°C for 90 seconds. After 1–2 minutes on ice, SOB (800µl), containing 20 mM glucose was added and the mixture incubated at 37°C for one hour. Aliquots of 50–100µl of the culture were removed, plated onto LB Agar plates supplemented with 100 µg/ml ampicillin and incubated at 37°C overnight to establish colonies.

b) Competent cells (3 µl) were mixed on ice with 1µl plasmid DNA (1–100 ng DNA in TE buffer, pH 8.0). The mixture was then transferred immediately to 44°C and, after one minute, 100µl of LB was added and the entire mixture directly plated onto selective media as described in **a)**.

(iii) Small scale preparation of plasmid DNA

For the initial characterisation of recombinant plasmids, small scale preparations of plasmid DNA were carried out essentially as described by Birnboim and Doly (1979). The colony of transformed bacterial cells to be tested was placed in 1.5 ml LB containing 100 µg/ml ampicillin and incubated at 37°C, 275 r.p.m., for 3 hours. Cells were collected by centrifugation in an MSE microfuge (12,000 x g for 10 seconds) then resuspended in 100 µl of 8% sucrose, 0.5% tritonX-100, 50 mM EDTA, 50 mM Tris.HCl, pH 8.0, by vortexing. Lysozyme (Sigma) was added (16µl of a 10 mg/ml solution), the sample vortexed then heated at 100°C for 2 minutes. Cell debris was sedimented in an MSE microfuge (12,000 x g) for 15 minutes. The gelatinous pellet was removed using a sterile toothpick and 100 µl isopropanol added to the supernatant. After incubation at room temperature for 10 minutes, precipitated DNA was collected by centrifugation for 5 minutes, the dried pellet taken up in 50 µl 3 mM EGTA then heated at 100°C for one minute. Appropriate restriction enzyme digests in the presence of RNAse A (50 µg/ml) were carried out on 10 µl of plasmid DNA/EGTA solution and reaction products analysed by electrophoresis in 6–8%

polyacrylamide gels. After electrophoresis, gels were stained in running buffer containing 1 μg/ml EtBr and DNA visualised under longwave UV.

(iv) Large scale preparation of plasmid DNA

Bacteria transformed by the plasmid of interest were grown to stationary phase in 500 ml cultures of LB containing 100 μg/ml ampicillin. Cells were harvested by centrifugation at 8,000 r.p.m. in a Beckman JA.14 rotor for 10 minutes. The pellet was resuspended in 9 ml of a buffer containing 50 mM glucose, 25 mM Tris.HCl, 10 mM EDTA, pH 8.0 and 1 ml of a freshly prepared solution of 20 mg/ml lysozyme in 0.2 M Tris.HCl, pH 8.0. After 5 minutes on ice, 20 ml of a freshly prepared 0.2 M NaOH– 1.0% SDS solution was added and the lysed cells swirled on ice for 5 minutes followed by the addition of 10 ml of ice cold 5 M KOAc, pH 4.8. Cell debris was mixed gently on ice for a further 15 minutes and the mixture clarified by centrifugation at 10,000 r.p.m. for 10 minutes. Nucleic acid was precipitated by the addition of 16.5 ml isopropanol. The precipitate was recovered by centrifugation at 10,000 r.p.m. for 15 minutes, the pellet washed with 70% ethanol and resuspended in 2 ml of TE buffer, pH 8.0. Caesium chloride and EtBr (Sigma), to final concentrations of 1.55 g/ml and 600 μg/ml respectively, were added to the DNA/TE solution, mixed well, allowed to stand at room temperature for 5 minutes and clarified by spinning in an MSE microfuge (12,000 X g) for 3 minutes. The supernatant was then loaded into 11X32 mm Beckman polyallomer bell-top Quick-Seal tubes, heat-sealed and plasmid DNA centrifuged to equilibrium at 100,000 r.p.m. for 4.5 hours at 20°C in a Beckman TLA100.2 rotor. Plasmid bands were visualised under longwave UV and removed by side puncture using a hypodermic needle. Caesium chloride was removed by passage over a 10 ml Sepharose CL-4B (Pharmacia Ltd., Milton Keynes, Buckinghamshire, England) column and the EtBr removed by passage over a 0.5 ml Dowex AG 50-WX8 (Bio-Rad, Richmond, California, USA) column in TEN buffer. DNA was then ethanol precipitated, treated with 50 μg/ml RNase A, extracted once with phenol and once with chloroform, ethanol precipitated and resuspended in H_2O. DNA concentrations were determined by scanning the absorbance from 220 to 320 nm in a Perkin-Elmer Lambda 5 UV/VIS spectrophotometer. One optical density unit (A_{260}) was taken to be equivalent to 40 μg of double stranded DNA. In addition, concentrations were checked by separating DNA samples on agarose gels containing 0.5 μg/ml EtBr and comparing the intensity of the fluorescence, under longwave UV, with that of known

standards.

8. Recombinant plasmids

Plasmid pUC1X72 contains the SV40 regulatory region inserted between the *Eco* RI and *Bam* HI sites of pUC13 (Figure 1M). The regulatory region was contained within an *Eco* RI/*Bam* HI fragment of pMKD 231 which contains the SV40 early promoter region, including both copies of the 72 bp repeat element (Everett *et al* ., 1983). Deletion of the DNA between the two *Sph* I sites of pMKD 231 generated a recombinant which contained only one copy of the 72 bp element. Digestion of this recombinant with *Eco* RI and *Bam* HI released a fragment containing the SV40 enhancer region, which was then transferred to pUC13, thus generating pUC1X72. Plasmid pMKD 231 was a kind gift from R.D. Everett, MRC Institute of Virology, Glasgow, Scotland.

The pA series of plasmids (Zenke *et al.,* 1986) contains clustered point mutations in the SV40 enhancer in which three neighbouring nucleotides were mutated simultaneously *in vitro* by generating non-complementary transversions (A ⇌ C, G ⇌ T). The mutated enhancer sequences were cloned into the expression vector pAO, yielding the recombinants pA11–pA16 (Figure 2M). These pA.. plasmids were generously provided by P. Chambon, CNRS, Strasbourg, France.

Plasmid pIF10 (Figure 3M) contains the human ß-interferon (ß-IFN) gene plus upstream regulatory sequences extending to the -77 ß-IFN nucleotide (Goodbourn *et al.,* 1985). The interferon gene regulatory element (IRE) is located between sequences -77 and -36. A series of pIF10 mutants were constructed containing single point mutations in the IRE. The nucleotide changes incorporated in these mutants are listed in Figure 3M.B. The pIF10.. series of plasmids was a kind gift from S. Goodbourn, ICRF, Lincoln's Inn Fields, London.

Plasmid pSPIRE was generated by replacing the small region between the *Bam* HI and *Bgl* II sites of pSP64 with the *Bam* HI/*Bgl* II IRE fragment (Goodbourn *et al.,* 1985) and was a kind gift from S. Goodbourn.

The recombinant plasmid pSV1 contains one copy of the SV1 double stranded synthetic oligonucleotide inserted into the *Bam* HI site of the pUC19 multicloning site. Digestion with *Eco* RI and *Pst* I, or *Hind* III and *Kpn* I, generated DNA restriction fragments of 66 bp and 58 bp respectively, both of which contained the double stranded SV1 oligonucleotide, as determined by Maxam and

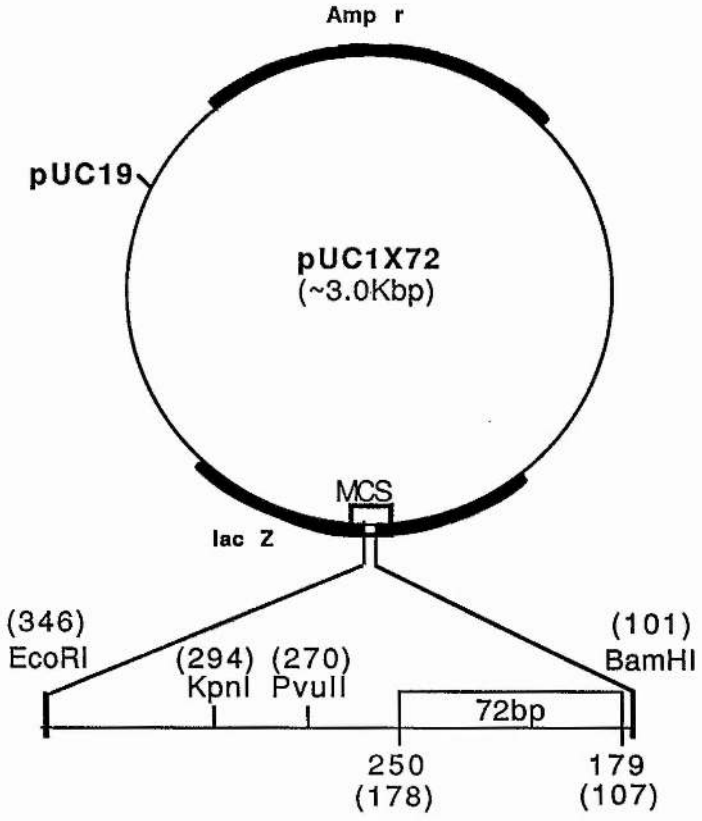


Figure 1M. Structure of plasmid pUC1x72

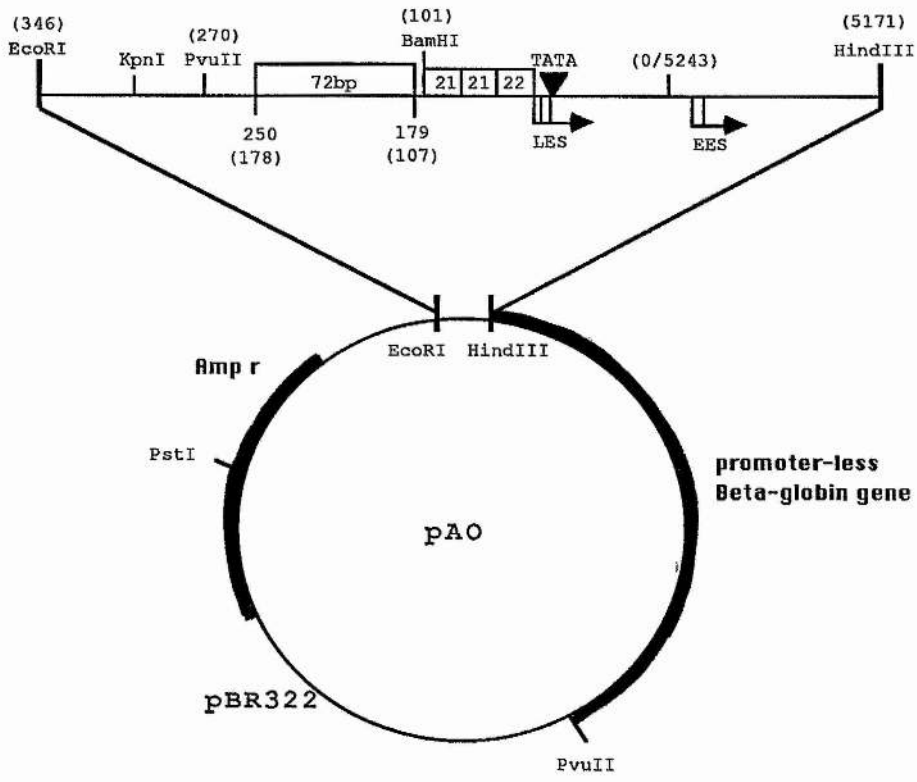
Plasmid pUC1X72 contains one copy of the SV40 72 bp repeat element plus upstream sequences extending to the Hpa II site (346). (In this construct, the HpaII site has been replaced with an EcoRI site). The EcoRI/BamHI fragment, containing the SV40 enhancer element, was inserted between the EcoRI and BamHI sites of the pUC19 plasmid multicloning site (MCS) within the lacZ gene.Relevant restriction enzyme cleavage sites are indicated and the positions of the ampicillin resistance and lacZ genes are represented by the black boxes. The BBB numbering system for SV40 nucleotides was employed (Tooze, 1982).

Figure 2M. Structure of pAO recombinants

A. Structure of the basic recombinant pAO which contains the SV40 early promoter region, from *Hpa* II (*Eco* RI in this contruct) to *Hind* III, cloned upstream of the coding sequences of the rabbit ß-globin gene, in a pBR322-based vector (Zenke *et al.,* 1986). Relevant SV40 DNA restriction sites, the 72 bp sequence and the 21 bp region, the TATA box, the early (EES) and late (LES) early transcription initiation sites, are indicated. All pA-type recombinants contain an engineered *Bam* HI site at the junction between the 72bp sequence and the 21 bp region. The BBB numbering system for SV40 nucleotides was employed (Tooze, 1982).

B. Sequence of the triple point mutations present in the pAO derivatives pA11–pA16, aligned with the wild-type SV40 DNA sequence.

A



B

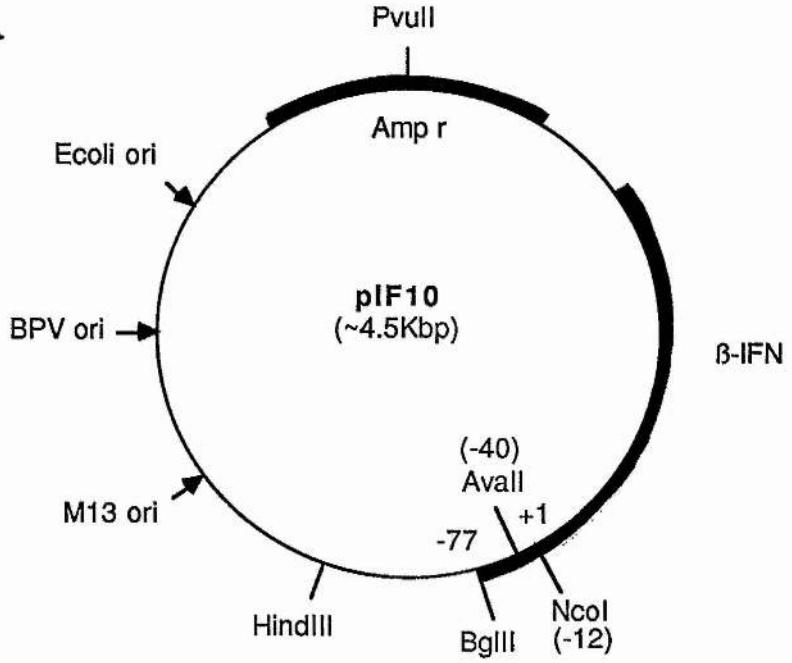
pA..	11 12 13 14 15 16
	TGT CCC AAA GTT TGA CTT
SV40	GTGTGGAAAGTCCCCAGG
	249 232

Figure 3M. Structure of the pIF10 recombinants

A. Structure of the basic recombinant pIF10 which contains cDNA sequences encoding the human ß-interferon (ß-IFN) gene plus its upstream regulatory sequences extending to the -77 ß-IFN nucleotide and encompassing the ß-IFN gene regulatory element (IRE), which spans nucleotides -77 to -36 (Goodbourn *et al.,* 1985). *E. coli,* Bovine Papilloma virus (BPV) and M13 origins of DNA replication, as well as relevant DNA restriction sites, are indicated.

B. Table of pIF10 mutants which contain single point mutations in the IRE. The nucleotide changes incorporated in these mutants are indicated.

A

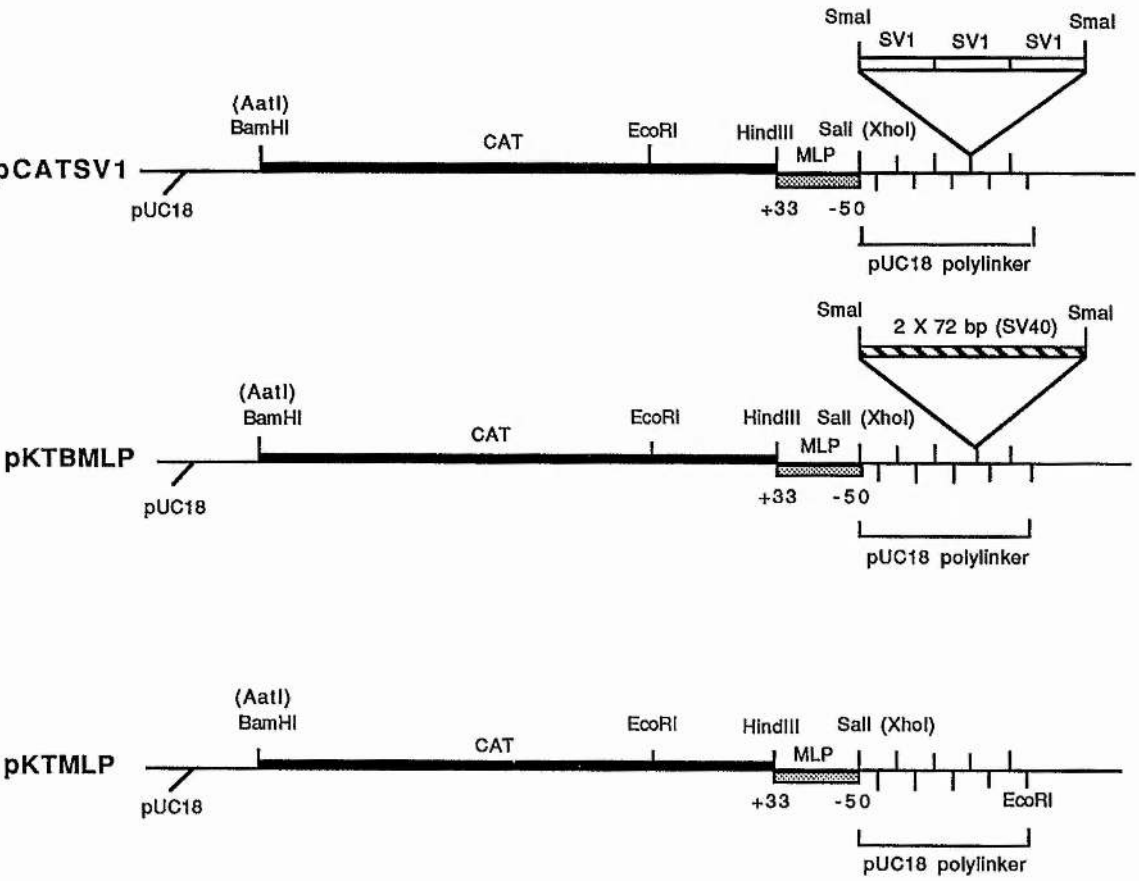


B

Mutant	Nucleotide change
pIF10-65	T → C
pIF10-64	G → A
pIF10-63	G → A
pIF10-62	G → A
pIF10-61T	A → T
pIF10-61G	A → G
pIF10-60	A → G
pIF10-57	T → C
pIF10-56	C → T
pIF10-55	C → T
pIF10-54	T → C

Figure 4M. Structure of the CAT expression vectors

Plasmid pKTMLP contains DNA sequences from the *E. coli* transposable element Tn9, which contains the coding sequences for the chloramphenicol acetyl transferase (CAT) gene. The CAT gene insert (depicted by the black box) consists of a 29 bp 5' untranslated segment, the CAT coding sequence and 86 bp 3' to the translation stop codon (Gorman *et al.,* 1982). Adjacent to the CAT gene are sequences derived from the Adenovirus type 2 major late promoter spanning nucleotides -50 to +33 (MLP), relative to the transcription start site at +1. Plasmid pKTBMLP also contains the *Bam* HI/*Pvu* II fragment from SV40 which contains two copies of the 72 bp element plus upstream flanking sequences (depicted by the striped box), cloned into the *Sma* I site of pKTMLP. Plasmids pCATSV1/1, 2 and 3 contain one, two or three copies, respectively, of the SV1 double stranded synthetic oligonucleotide inserted into the *Sma* I site of pKTMLP.


SmaI
SV1
SV1
SV1
SmaI
(AatI)
BamHI
CAT
EcoRI
HindIII
SalI (XhoI)
MLP
pCATSV1
pUC18
+33
-50
pUC18 polylinker
SmaI
2 X 72 bp (SV40)
SmaI
(AatI)
BamHI
CAT
EcoRI
HindIII
SalI (XhoI)
MLP
pKTBMLP
pUC18
+33
-50
pUC18 polylinker
(AatI)
BamHI
CAT
EcoRI
HindIII
SalI (XhoI)
MLP
pKTMLP
pUC18
+33
-50
EcoRI
pUC18 polylinker

Gilbert (1980) chemical sequencing of the DNA fragments.

Plasmid pHIVEn contains DNA sequences from the HIV enhancer from position -79 to -108, inserted into the *Bam* HI site of the pUC13 polylinker. Digestion with *Eco* RI and *Pst* I, or *Hind* III and *Sac* I, released DNA restriction fragments of 73 bp and 76 bp respectively, both of which contained the enhancer sequences. Plasmids pHIV-L and pHIV-R contain one copy of either the HIV-L or HIV-R double stranded oligonucleotide inserted into the *Bam* HI site of the pUC13 polylinker. Digestion with *Eco* RI and *Pst* I, or *Hind* III and *Sac* I, released DNA restriction fragments of 59 bp and 62 bp respectively, which contained the double stranded HIV-L (in the case of pHIV-L) or HIV-R (in the case of pHIV-R) oligonucleotides. The DNA sequences of the enhancer-containing restriction fragments were determined by Maxam and Gilbert (1980) chemical sequencing.

CAT expression vectors are derivatives of plasmid pKTMLP (Figure 4M) which contains the Adenovirus type 2 major late promoter (MLP) and the bacterial CAT gene inserted into the pUC18 polylinker. Plasmid pKTBMLP contains two tandem copies of the 72 bp repeat element of the SV40 enhancer inserted into the *Sma* I site of pKTMLP. Plasmids pKTMLP and pKTBMLP were a generous gift from C. Goding, Marie Curie Research Institute, Oxted, Surrey. Constructs containing SV1 double stranded oligonucleotides were prepared by first inserting *Bgl* II linkers into the *Sma* I site of pKTMLP. Double stranded SV1 oligonucleotides were then inserted into this engineered *Bgl* II site generating plasmids pCATSV1/1, pCATSV1/2 and pCATSV1/3 which contain one, two or three tandem copies of the SV1 oligonucleotide.

9. Radiolabelling of DNA

(i) Radiochemicals

[α-^{32}P]dATP (3,000 Ci/mmol), [α-^{32}P]dCTP (3,000 Ci/mmol), [γ-^{32}P]ATP (5,000 Ci/mmol) and [^{14}C] chloramphenicol (54 mCi/mmol) were all obtained from Amersham International Limited, Bucks., U.K.

(ii) Radiolabelling of DNA fragments

DNA fragments were 3'-end labelled using an end repair method (Drouin, 1980). The Klenow fragment of *E. coli* DNA polymerase I (Klenow *et al.*, 1971) was used to fill in from a recessed 3'-end using the corresponding 5'-extension as a template for DNA synthesis. A DNA

fragment to be labelled (10–1000 ng) was incubated with 20 μCi [α-^{32}P]dATP, 20 μCi [α-^{32}P]dCTP, unlabelled dTTP and dGTP to final concentrations of 100 μM, 0.5 μg gelatin and 10 units of the Klenow fragment in a 10 μl reaction containing 50 mM Tris.HCl, 5 mM $MgCl_2$, 1 mM DTT, pH 8.0. After 15 minutes at room temperature unlabelled dATP and dCTP were added to final concentrations of 100 μM and incubation at room temperature was allowed to proceed for a further 20 minutes. The reaction was either terminated by the addition of one fifth volume of 50% glycerol, 100 mM EDTA, bromophenol blue and xylene cyanol FF, or, labelled DNA was extracted once with phenol, once with chloroform, ethanol precipitated and taken up in H_2O prior to secondary cleavage with the appropriate restriction endonuclease. (The Klenow fragment of *E. coli* DNA polymerase I was obtained using a recombinant plasmid that directs overproduction of the polymerase fragment as described by Joyce and Grindley (1983) and was a kind gift of Celia Toogood).

DNA fragments with 5'-protruding termini were radiolabelled in a reaction catalyzed by T4 polynucleotide kinase (Maxam and Gilbert, 1977). The DNA was dephosphorylated at 5'-termini by treatment with alkaline phosphatase. One μg of DNA was incubated with 0.01 unit of bacterial alkaline phosphatase in 100 mM Tris.HCl, pH 8.5 for 30 minutes at 60°C. EDTA was added to 10 mM and the DNA purified by organic extraction followed by ethanol precipitation. The dry DNA pellet was redissolved in H_2O and incubated at 37°C for 30 minutes in the presence of T4 polynucleotide kinase (2 units/μg of DNA), 50 mM Tris.HCl, 10 mM $MgCl_2$, 5 mM DTT, pH 8.0 and 40 μCi [γ-^{32}P]ATP in a total reaction volume of 10 μl. Addition of EDTA to 50 mM terminated the reaction and the volume was adjusted to 100 μl by the addition of TEN buffer. Labelled DNA was purified as described previously and taken up in H_2O prior to secondary cleavage with the appropriate restriction endonuclease.

Double stranded synthetic oligonucleotides were 3'-end labelled as described for restriction fragments. In the case of all 3'- or 5'-end labelled fragments, unincorporated dNTPs were separated from labelled DNA in 8% polyacrylamide gels and the DNA electroeluted (Hay and DePamphilis, 1982).

When preparing end-labelled DNA fragments generated from the various recombinant plasmids, the following restriction digests and labelling procedures were carried out:–

a). Plasmid pUC1X72 was digested with *Bam* HI, 3'- or 5'-end labelled at the *Bam* HI site and the

labelled DNA cleaved with either *Pvu* II or *Eco* RI, generating fragments of 97 bp and 173 bp respectively.

b). End labelled restriction fragments containing mutant SV40 enhancer sequences were generated from the pA.. series of constructs by digesting with both *Bam* HI and *Pvu* II followed by 3'-end labelling at the *Bam* HI site. Fragments of 97 bp were generated in each case.

c). In generating labelled fragments from the pIF10 series of plasmids, double digests using *Bgl* II and *Nco* I were carried out and the fragments 3'-end labelled at the *Bgl* II site. Fragments of 65 bp were generated in each case.

d). Plasmid pSPIRE was digested with *Hind* III, 3'- or 5'-end labelled at the *Hind* III site and secondary cleavage of the labelled DNA with *Bgl* II carried out.

e). Radiolabelled fragments were generated from pSV1 by double restriction digests using either*Eco* RI and *Pst* I, or *Hind* III and *Kpn* I, followed by 3'-end labelling at either the *Eco* RI or *Hind* III sites.

f). Plasmids containing sequences from the HIV enhancer (pHIVEn, pHIV-L and pHIV-R) were digested with either *Eco* RI and *Pst* I, or *Hind* III and *Sac* I, followed by 3'-end labelling at either the *Eco* RI or *Hind* III sites.

10. Preparation of double stranded synthetic oligonucleotides

(I) Synthesis of single stranded oligonucleotides

Single stranded, synthetic oligonucleotides were synthesised on an Applied Biosystems Model 381A DNA-Synthesiser using ß-cyanoethyl phosphoramidites. This model of DNA synthesiser uses a solid phase synthesis chemistry in which the growing DNA chain remains covalently attached to an insoluble matrix, Controlled Pore Glass (CPG). CPG is easily derivatised with adenosine, guanosine, thymidine or cytosine phosphoramidites. An organic linker is attached to the surface of the CPG and the support is then derivatised by covalent attachment of the 3'-hydroxyl of a nucleoside to the linker via an ester bond. Following derivatisation, all free amino groups are capped.

To begin synthesis, a column is placed on the DNA synthesiser containing one of four support-bound nucleosides. This nucleoside is the 3'-terminus of the sequence and DNA synthesis

proceeds in a 3' to 5' direction.

The first step in the phosphoramidite chemistry of oligonucleotide synthesis is to remove the acid-labile, dimethoxytrityl (DMTO) group by treatment with TCA. This yields a reactive 5'-hydroxyl which can react with a phosphoramidite during the following coupling, or addition step. ß-cyanoethyl phosphoramidites are chemically modified nucleosides which contain a diisopropylamine group on a 3' trivalent phosphorous moiety. A ß-cyanoethyl protecting group is present on the 3' phosphorous group of the nucleoside with a dimethoxytrityl (DMTO) protecting group on the 5'-hydroxyl. The structure of guanosine ß-cyanoethyl phosphoramidite is shown in Figure 5M.

In the coupling step, tetrazole and the phosphoramidite are mixed as they enter the reaction chamber generating a highly reactive species which rapidly reacts with the free 5'-hydroxyl group of the support- bound nucleotide. Tetrazole protonates the nitrogen of the diisopropyl group on the 3'-phosphorous which converts the amine into a very good leaving group upon nucleophilic attack by the 5'-hydroxyl group. Thus an internucleotide linkage (5' to 3') is formed through a trivalent phosphorous group.

During the coupling step, a small percentage of support bound oligonucleotides may fail to undergo addition. These truncated, or failure sequences are chemically modified to prevent their participation in subsequent coupling steps. Acetic anhydride and dimethylaminopyridine are mixed and form a powerful acetylating agent which terminates, or "caps" any unreacted chains from the coupling step.

Immediately after "capping", the labile trivalent phosphorous linkage formed in the coupling step, is oxidised to the stable pentavalent phosphorous linkage of biologically active DNA using iodine as a mild oxidant. Following oxidation, one cycle of nucleotide addition and DNA synthesis is continued by removing the DMTO group at the 5'-terminus of the oligomer and repeating another cycle of base addition. This is done until DNA of the desired length and sequence has been synthesised.

After synthesis is complete, the chains are cleaved from the CPG support and the ß-cyanoethyl protecting groups on the phosphates removed by four 15 minute washes with concentrated ammonium hydroxide. The base protecting groups are then removed by heating at 55°C in

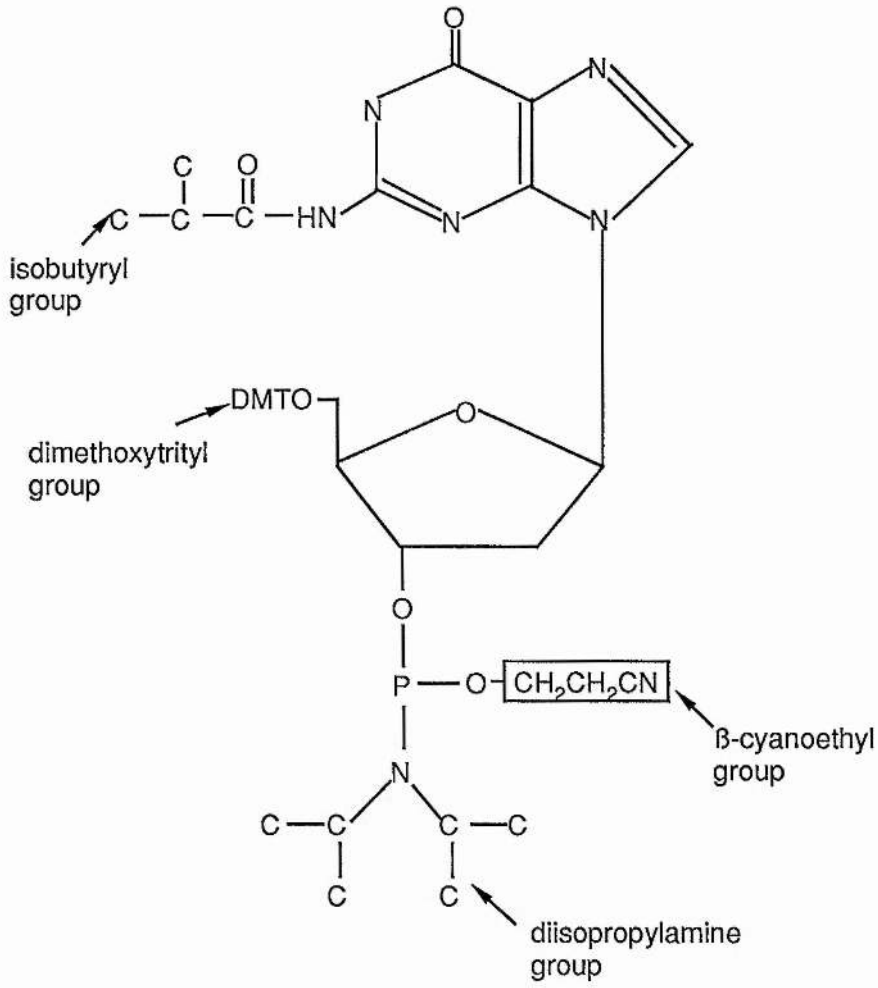


Figure 5M. Structure of guanosine ß-cyanoethyl phosphoramidite

The positions of the ß-cyanoethyl protecting group on the 3'-phosphorus moiety, the DMTO protecting group on the 5'-hydroxyl, the isobutyryl protecting group on the exocyclic amine and the diisopropylamine on the 3'-trivalent phosphorus moiety are indicated.

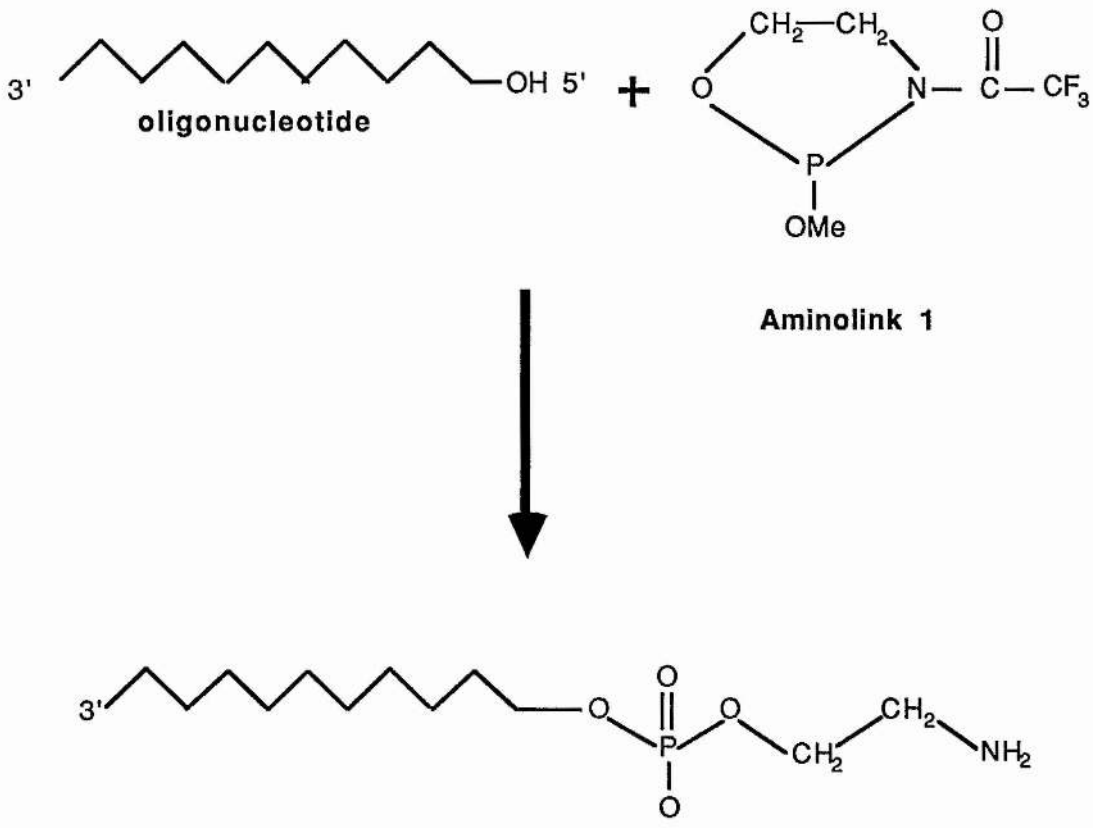


Figure 6M. Addition of the Aminolink 1 reagent to a synthetic single stranded oligonucleotide

The Aminolink 1 reagent is coupled to the 5'-end of the oligonucleotide generating a primary amine group which can then be linked to a variety of substrates..

ammonium hydroxide for 15 hours.

Following cleavage and deprotection, the ammonia is removed by evaporation under vacuum, DNA ethanol precipitated and the dry DNA pellet resuspended in H_2O. DNA concentrations were determined by scanning the absorbance from 220 to 320 nm in a Perkin-Elmer Lambda 5 UV/VIS spectrophotometer. One optical density unit (A_{260}) was taken to be equivalent to 40 µg of DNA. Oligonucleotides were then further purified by gel electrophoresis. Electrophoresis of the oligonucleotides was carried out in 20% acrylamide (19% acrylamide : 1% N,N'-methylene-bisacrylamide) gels containing 7 M urea, cast between glass plates separated by 1.5 mm spacers. After electrophoresis, gels were removed from the glass plates and placed on clear plastic film on top of a fluorescent thin layer chromatography plate. The UV absorbing oligonucleotide masks fluorescein emission in the plate, and appears as a shadow, when viewed under shortwave UV. Oligonucleotides corresponding to the appropriate size were excised from the gels and the DNA recovered from the acrylamide by electroelution (Hay and DePamphilis, 1982).

(ii) Synthesis of single stranded Aminolink-oligonucleotides

Single stranded oligonucleotides were synthesized with primary amine groups coupled to their 5'-ends. In the last step of the synthesis, Aminolink 1 reagent is coupled to the DNA in a reaction which generates an aminoethyl-phosphate linker containing a trifluoroacetyl amine-protecting group. Figure 6M depicts the addition of the Aminolink reagent to a single stranded oligonucleotide. Prior to deprotection, the Aminolink 1 phosphite triester is oxidised to the phosphate triester as described above. The trifluoroacetyl protecting group is removed during deprotection resulting in the functional amine.

All steps in the synthesis of oligonucleotides were carried out as described in the Applied Biosystems Model 381A DNA-Synthesiser User's Handbook. Chemicals for the synthetic reactions were obtained from Applied Biosystems Limited, Warrington, Cheshire, England.

(iii) Hybridisation of complementary single stranded oligonucleotides

Complementary, single stranded synthetic oligonucleotides were annealed by heating to 100°C in 0.1M NaCl, 10mM Tris.HCl, 1mM EDTA, pH 8.0, followed by slow cooling to 16°C. Table 1 shows the nucleotide sequences of the various oligonucleotides used in this study.

Table 1.

Nucleotide sequences of the double stranded synthetic oligonucleotides used in this study.

Oligonucleotide	Nucleotide sequence
SV1	5' GATCTAGGGTGTGGAAAGTCCCG 3' 3' ATCCCACACCTTTCAGGGCCTAG 5'
SV1.M1	5' GATCTAGGGTGTCCAAAGTCCCG 3' 3' ATCCCACAGGTTTCAGGGCCTAG 5'
SV1.M2	5' GATCTAGGGTGTGGAATGTCCCG 3' 3' ATCCCACACCTTACAGGGCCTAG 5'
SV1.M3	5' GATCTAGGGTGTGGAAAGTGGCCG 3' 3' ATCCCACACCTTTCACCGGCCTAG 5'
SV2	5' GATCTAAAGAACCAGCTGTGGAATGTGTG 3' 3' ATTTCTTGGTCGACACCTTACACACCTAG 5'
SV3	5' GATCCTGACTAATTGAGATGCATGCTTTGCATACTTCA 3' 3' GACTGATTAACTCTACGTACGAAACGTATGAAGTCTAG 5'
SVUP	5' GATCTGAGGCGGAAAGAACCAGCTG 3' 3' ACTCCGCCTTTCTTGGTCGACCTAG 5'
IRE	5' GATCAAAGTGGGAAATTCCTCTG 3' 3' TTTCACCCTTTAAGGAGACCTAG 5'
HIV	5' GATCAAGGGACTTTCCGCTGGGGACTTTCCAGG 3' 3' TTCCCTGAAAGGCGACCCCTGAAAGGTCCCTAG 5'
HIV-L	5' GATCTAGGGACTTTCCGCG 3' 3' ATCCCTGAAAGGCGCCTAG 5'
HIV-R	5' GATCTGGGGACTTTCCAGG 3' 3' ACCCCTGAAAGGTCCCTAG 5'
NF-kB	5' GATCTGGGGACTTTCCGAG 3' 3' ACCCCTGAAAGGCTCCTAG 5
H2TF1	5' GATCTGGGGATTCCCCAG 3' 3' ACCCCTAAGGGGTCCTAG 5'
H2KD	5' GATGGGGAATCCCCAGCCCTGGGCTTCCCCACCCCTGACCTCA 3' 3' ACCCCTTAGGGGTCGGGACCCGAAGGGGTGGGGACTGGAGTGA 5'
EnA	5' GATCCAAGCGCCGGATGTGGTAAAAGTGACA 5' 3' GTTCGCGGCCTACACCATTTTCACTGTCTAG 5'
EnB	5' GATCCAGAGGAAGTGAAATA 3' 3' GTCTCCTTCACTTTATCTAG 5'
SYM	5' GATCATGGGGAATTTCCCCAG 3' 3' TACCCCTTAAAGGGGTCCTAG 5'

11. In vivo transient expression assays

HeLa S1 monolayer cells were grown to ~30–40% confluency (approximately 8 X 10^5 cells/50 mm plate) in Glasgow modified minimal essential medium containing 10% foetal calf serum. Calcium phosphate-DNA coprecipitates were prepared essentially as described by Graham and van der Eb (1973). Fine precipitates were achieved by incubating the DNA (4 µg) with 113 µl TE and 125 µl 2 X HBS (560 mM NaCl, 100 mM HEPES, 3 mM $(Na)_2HPO_4$) then allowing the mixture to stand at room temperature for 2 hours. Calcium chloride (to a final concentration of 100 mM) was added to the DNA solution and air gently bubbled through the reaction mixture. Precipitates were allowed to stand at room temperature without agitation before being added dropwise to the tissue culture cells. The cells were then incubated at 37°C for 48 hours. Prior to harvesting, cells were washed twice with STE buffer (150 mM NaCl, 10 mM Tris.HCl, 1 mM EDTA, pH 7.8) and scraped into 1 ml of STE. Cells were collected by centrifugation, resuspended in 50 µl of 250 mM Tris.HCl, pH7.5 and subjected to three rapid cycles of freezing and thawing at -70°C and 37°C respectively. Cell debris was sedimented by centrifugation at high speed for 10 minutes in an MSE microfuge. 50 µl of cell supernatant was added to 100 µl of assay mixture containing 70 µl 1 M Tris.HCL, pH7.8, 2 µl 40 mM acetyl coenzyme A (Sigma), 1.6 µl [^{14}C] chloramphenicol (54 mci/mmol) (CAM) and the reaction mixture incubated at 37°C for 60 minutes (Gorman *et al.*, 1982). The reaction was stopped by the addition of 1 ml of cold ethyl acetate, which was also used to extract the CAM. Samples were dried under vacuum, taken up in 20 µl ethyl acetate, spotted onto silica-gel thin-layer plates and developed with chloroform-methanol (95:5, ascending). Plates were air dried and exposed to X-ray film at -70°C.

12. Gel electrophoresis DNA binding assay

Assays contained 0.1-1.0 ng labelled probe (~10,000 cpm), 0.5-1 µg of unlabelled DNA (equimolar amounts of poly [d(A-T)]:poly [d(C-G)]) (Pharmacia), 20 mM HEPES.NaOH pH 7.5, 1 mM DTT, 1 mM EDTA, 10% glycerol, 100 mM NaCl, 0.05% NP40 and 1-2 µl of protein fraction in a final reaction volume of 20 µl. Binding reactions were allowed to proceed to equilibrium for 20 minutes at 20°C after which time 5 µl of 50% glycerol, 25 mM HEPES.NaOH pH 7.5, 0.1% bromophenol blue was added and the entire reaction loaded onto 6% (80:1 acrylamide: N,N'-methylene-bisacrylamide)

polyacrylamide gels containing 0.5 X TBE buffer. Electrophoresis was carried out at 200V for one hour. Gels were fixed in 10% acetic acid, dried and exposed to Fuji RX X-ray film (Fuji Photo Film (U.K.) Ltd., Finchley, London, England) at -70°C with an intensifying screen (Du Pont (U.K.) Ltd., Stevenage, Herts., England). As a rapid and quantitative means of determining the yield of enhancer-binding activity at each step during the fractionation, bands corresponding to bound DNA were excised from the gels and their radioactivity determined by Cerenkov radiation counting. One unit of activity represents the retention of 15 fmol specific DNA bound in the DNA/protein complexes in the standard gel electrophoresis DNA-binding assay.

13. Ion-exchange and affinity chromatography

(i) Preparation of DEAE-Sepharose

Pre-swollen DEAE-Sepharose CL-6B was supplied by Pharmacia Fine Chemicals. The Sepharose was equilibrated in buffer A containing 0.3 M or 50 mM NaCl prior to use.

(ii) Preparation of DNA affinity matrices

Sequence-specific DNA affinity matrices were prepared essentially as described by Kadonaga and Tjian (1986). Single stranded synthetic oligonucleotides (250 μg of each containing the sequences

5'–GATCTAGGGTGTGGAAAGTCCCG–3' and 3'–ATCCCACACCTTTCAGGGCCTAG–5') were phosphorylated in a reaction containing T4 polynucleotide kinase (2 units/μg of DNA), 50 mM Tris.HCl, 10 mM $MgCl_2$, 5 mM DTT, pH 8.0, and 5 μCi [$^{-32}$P]ATP (5,000Ci/mMol). Reactions were incubated at 37°C for 5 minutes after which time ATP was added to a final concentration of 2 mM and incubation continued at 37°C for 2–3 hours. Complementary single stranded oligonucleotides were annealed and polymerisation of double stranded oligonucleotides carried out in the presence of 600 units of T4 DNA ligase, 15 mM DTT and 2 mM ATP, overnight at room temperature. After extraction with phenol and chloroform, and ethanol precipitation, the ligated duplex oligonucleotides were taken up in H_2O and coupled to CNBr-activated Sepharose 4B (Pharmacia).

The CNBr-activated Sepharose (2 g) was swollen in 10 ml 1mM HCl followed by extensive washing (400 ml) with the same solution on a sintered glass filter (the activated Sepharose was never allowed to suction filter into a dry cake). The Sepharose was then washed with 400 ml H_2O,

400 ml 10 mM $NaPO_4$ buffer pH 8.0, and ligated oligonucleotides added immediately to the activated Sepharose slurry to facilitate efficient coupling. The DNA was allowed to couple overnight at room temperature with gentle rotation. In general, the concentration of covalently bound DNA in the affinity resin was 30–40 μg of DNA per ml of resin, as determined by retention of radioactively labelled DNA on the CNBr-Sepharose. Unreacted sites on the CNBr-derivatised Sepharose were blocked by washing with 100 ml of 1 M ethanolamine HCl pH 8.0 and the ethanolamine-containing slurry allowed to rotate gently at room temperature for 4 hours, after which time the resin was washed with 100 ml each of 0.1 M NaCl, 10 mM $NaPO_4$ buffer pH 8.0, 1 M NaCl and $TEN_{300}NaN_3$. The affinity resin was stored as a slurry in $TEN_{300}NaN_3$ at 4°C until required. For the recognition site affinity chromatography steps, the resin was equilibrated in buffer A containing 0.25 M NaCl plus 3 mM n-octylglucopyranoside (Sigma).

As a means of generating an affinity resin with a high coupling efficiency of specific DNA sequences, single copy, duplex oligonucleotides containing primary amine groups at their 5'-ends, were covalently coupled to CNBr-activated Sepharose essentially as described above.

(iii) Pouring of columns

DEAE-Sepharose was packed in chromatography columns supplied by Amicon Limited, Gloucester, U.K. The Sepharose was poured as a slurry and packed using a peristaltic pump (Pharmacia) at a flow rate equivalent to two column volumes per hour.

The 1st round recognition site affinity matrices were packed in 5 ml plastic syringes containing two glass fibre discs (Whatman Limited, Maidstone, England) as a bed support. The 2nd and 3rd round affinity matrices were packed in 5 ml disposable chromatography columns containing 70 μm filters (lab m, Bury, Lancs., England). Affinity matrices were allowed to settle to form 5 ml, 1 ml and 0.5 ml packed volumes under the influence of gravity.

All column chromatography manipulations were carried out at 4°C.

14. Purification of EBP1 from HeLa nuclear extract

Uninfected HeLa cells (~2 X 10^{10}) were harvested from 40 litre spinner cultures by centrifugation at 3,000 r.p.m. for 10 minutes in an MSE Coolspin. The wet weight of the cell pellet was ~40 g. From this point onwards, all manipulations in the purification of EBP1 were carried out at

4°C. Cells were fractionated into nuclei and cytosol by swelling on ice in 80 ml of hypotonic buffer, lysed by 10 strokes with a tight fitting pestle in a Dounce homogenizer, and nuclei sedimented by centrifugation at 3,000 r.p.m. for 10 minutes (Challberg and Kelly, 1979). (The supernatant containing the cytoplasmic fraction did not contain any appreciable amounts of EBP1 and was stored at -70°C for other purposes).

Nuclei were washed three times with isotonic buffer, resuspended in the same buffer containing 0.35 M NaCl and stirred on ice for 30 minutes. Particulate material was removed by centrifugation at 100,000 X g for 60 minutes. The supernatant was adjusted to 0.3 M NaCl and applied to a 60 ml DEAE-Sepharose CL-6B (Pharmacia) column equilibrated with buffer A containing 0.3 M NaCl. The column was washed with 20 ml of the same buffer and the flow through dialysed against buffer A containing 50 mM NaCl for 5 hours. Insoluble material was removed by centrifugation at 100,000 X g for 30 min. The clarified extract was applied to a second DEAE-Sepharose column (25 ml) equilibrated with Buffer A containing 50 mM NaCl. After washing with two column volumes of equilibration buffer the column was developed with a linear gradient of 50 to 400 mM NaCl. Fractions containing EBP1 activity (as determined by gel electrophoresis DNA binding assays) were pooled,mixed with carrier DNA (poly [d(A-T)]: poly [d(C-G)]) and subjected to three rounds of recognition site affinity chromatography on 5 ml, 1 ml and 0.5 ml DNA affinity columns prepared with double stranded SV1 oligonucleotides which contain the sequence

```
5'-GATCTAGGGTGTGGAAAGTCCCG-3'
    3'-ATCCCACACCTTTCAGGGCCTAG-5'
```

essentially as described by Kadonaga and Tjian (1986) except that 400, 50, and 5 µg of carrier was added to the first, second, and third round affinity loads, respectively and 3 mM n-octyl glucopyranoside was included in all affinity chromatography buffers (Treisman,1987). The NaCl concentration of the affinity loads was adjusted to 0.25 M by dilution in buffer A. Single step elution in buffer A containing 1 M NaCl was carried out at each round of affinity purification. Active eluates were aliquoted into small amounts and stored at -70°C. Extracts were stable for up to 6 months at this temperature.

Protein extracts from various stages of the purification, which contained equivalent amounts of EBP1 binding activity, as determined in the standard gel electrophoresis DNA binding assay, were precipitated in 10% TCA, washed with 80% acetone, resuspended in SDS/ ß-mercaptoethanol and

analysed by SDS-polyacrylamide gel electrophoresis followed by silver staining (Heine *et al.*, 1974; Oakley *et al.*, 1980).

15. Gel electrophoresis

(i) Agarose gel electrophoresis

Slab gels (100 mm X 70 mm) containing 1% agarose (w/v), 1XTBE buffer and 0.5 µg/ml EtBr, were run at 100V at room temperature in 1XTBE containing 0.5 µg/ml EtBr. One fifth volume of DNA dye mix (50% glycerol, 100mM EDTA, bromophenol blue and xylene cyanol FF) was added to samples prior to electrophoresis. DNA was visualised under longwave UV.

(ii) Native polyacrylamide gel electrophoresis

Gels were made up to the required acrylamide concentration (generally between 6 and 10%) from a 30% acrylamide (29% acrylamide : 1% N,N'-methylene-bisacrylamide) stock and contained 0.5 or 1XTBE buffer. Polymerisation was initiated by the addition of 1/250 volume of 25% APS and 1/2500 volume of TEMED (Eastman Kodak Co., Rochester, New York, USA) stock. Electrophoresis was carried out at up to 250V.

(iii) SDS polyacrylamide gel electrophoresis and silver staining

10 to 15% resolving gels were prepared from an ~29% acrylamide stock (28% acrylamide : 0.735% diallyltartardiamide) and contained 0.375 M Tris.HCl pH 8.6 and 0.1% SDS. Polymerisation was carried out as described for native polyacrylamide gels. Gels were cast between glass plates separated by 0.35 mm thick spacers and sealed with Teflon tape. Resolving gels were poured to leave room for a 1 cm deep stacking gel and overlaid with 0.5 ml H_2O to facilitate polymerisation and to ensure an even surface. Stacking gels contained 5% acrylamide stock in 0.105 M Tris.HCl pH 7.0 and 0.1% SDS.

Samples were denatured prior to electrophoresis by heating to 100°C for 2 minutes in a solution containing 12.5% (v/v) glycerol, 2.5% (w/v) SDS, 62.5 mM Tris.HCl pH 7.0, 0.875 M ß-mercaptoethanol and 0.015% (w/v) bromophenol blue. Electrophoresis was carried out in 1XTGS buffer, at 50V through the stacking gel and 100V through the resolving gel, at room temperature, until the dye front was 0.5 cm from the bottom of the gel.

Gels were fixed and stained using the Bio-Rad Silver Staining Kit (total protein stain) by

following the manufacturers instructions.

16. UV cross linking

Double stranded SV1 oligonucleotide was inserted into the *Bam* HI site of bacteriophage M13mp8. Clear plaques (which contained inserts that inactivated the ß-galactosidase gene) were selected and bacteriophage grown in 1.5 ml cultures of *E. coli* JM101, in 2XTYE buffer (16 g/l Bacto- tryptone, 10 g/l yeast extract, 5 g/l NaCl, 8 mM NaOH) at 37°C, 275 r.m.p., overnight (*E. coli* cultures were prepared from a 1/100 dilution of strain JM101 which had been grown to saturation). Cells were collected by centrifugation for 5 minutes in an MSE microfuge and bacteriophage precipitated from the supernatant by the addition of 0.5 M NaCl and 2.5% polyethylene glycol. Bacteriophage were sedimented by centrifugation for 5 minutes in a microfuge and organic extraction of the DNA carried out after treatment with 0.5% SDS, 2 mM EDTA and proteinase K (50 µg/ml). DNA was ethanol precipitated, dried and resuspended in H_2O. The sequence of the cloned inserts was confirmed by dideoxy-sequencing (Sanger *et al.*, 1977). DNA synthesis, primed by the 17-mer sequencing primer, on single stranded M13 recombinant DNA was catalysed by the large (Klenow) fragment of *E. Coli* DNA polymerase I in the presence of dCTP, dATP, [α-^{32}P]dGTP (3,000 Ci/mMol) and 5-bromo- 2'deoxyuridine triphosphate (Pharmacia). DNA synthesis was allowed to proceed for 15 minutes at room temperature after which time the DNA was digested with *Eco* RI and *Pst* I, and the small fragment containing the EBPI binding site purified on 6% polyacrylamide gels and electroeluted (Hay and DePamphilis, 1982).

Binding reactions were prepared in the same way as for the gel electrophoresis DNA binding assays but in a total reaction volume of 50 µl. Once equilibrium had been reached samples were transferred to 100 µl glass capillary tubes and both ends of the tubes heat sealed. Samples were then irradiated on a TM-36 Transilluminator (Ultra Violet Products Ltd., Cambridge, England) - peak wavelength, 302nm; peak intensity, 7.0mW/cm^2 - the samples were packed on ice to maintain a constant temperature throughout irradiation. Reaction mixtures were then transferred to eppendorf tubes and digested with 0.17units/µl DNase I (Amersham) and/or 0.07U/µl micrococcal nuclease (Pharmacia) for 30 minutes at 37°C after addition of $MgCl_2$ and $CaCl_2$ to a final concentration of 3 mM . Reaction products were heated to 100°C for 2 minutes, in 12.5% (v/v) glycerol, 2.5% (w/v)

SDS, 62.5 mM Tris.HCl, pH 7.0, 0.875 M ß-mercaptoethanol and 0.015% (w/v) bromophenol blue, and fractionated by electrophoresis in 10% SDS-polyacrylamide gels and the dried gels exposed to X-ray film, in the presence of an intensifying screen, at -70°C.

17. DNase I protection

Reactions contained 0.1-1.0 ng labelled probe (~10,000 cpm), various amounts of unlabelled carrier DNA, poly [d(A-T)]:poly [d(C-G)], 25 mM HEPES.NaOH pH 7.5, 100 mM NaCl, and various amounts of protein fractions, containing 50% glycerol, 1 mM DTT, 1 mM EDTA, 0.1% NP40, in a total reaction volume of 100 µl. Binding was allowed to proceed for 20 minutes at 20°C then 0.5 units of DNase I (Amersham) plus $MgCl_2$ (to a final concentration of 4 mM) added and the reaction allowed to proceed for 60 seconds at 20°C after which time an equal volume of 0.6 M NaOAc, 20 mM EDTA was added. The DNA was then extracted once with phenol, once with chloroform, ethanol precipitated, washed with 70% ethanol, dried and redissolved in 10 µl formamide dyes (98% formamide, 20 mM NaOH, 1 mM EDTA with bromophenol blue and xylene cyanol FF). DNA was denatured at 100°C for 2 minutes, cooled on ice and reaction products fractionated by electrophoresis in 6% or 8% polyacrylamide gels containing 50% urea. Glass plates were treated with either SurfaSil siliconising agent (Pierce Chemical Company, Rockford, Illinois, USA) or Silane A-174 which covalently cross-links the acrylamide to the surface of the glass plate. After electrophoresis, gels were fixed in 10% acetic acid, baked dry onto the glass plates (Garoff and Ansgore, 1981) and exposed to X-ray film (Hyperfilm MP; Amersham) at -70°C in the presence of an intensifying screen. Sequence markers were prepared by subjecting end-labelled fragments to G+A or C+T specific cleavage reactions (Maxam and Gilbert, 1980).

18. Quantitation of protein

Protein concentrations were determined by the method of Bradford (1976). Protein samples, in a total volume of 10 µl, were mixed with 1 ml Bradford's reagent (100 mg Coomassie brilliant blue G250, 100 ml orthophosphoric acid and 50 ml ethanol made up to 1 litre with distilled H_2O), incubated at room temperature for 5 minutes and the absorbance at 595 nm measured in an LKB Biochrom Ultrospec II spectrophotometer. Concentrations were calculated by comparison with

a standard curve constructed using BSA (Sigma).

19. Methylation protection

Binding reactions were carried out as described for DNase I protection assays. After equilibrium had been reached, 0.5 µl of DMS was added and the mixture incubated at 20°C for 2 minutes. The reaction was stopped by the addition of two volumes of DMS "stop" buffer (0.3 M NaOAc, 1.5 M ß-mercaptoethanol, 50 µg/ml yeast tRNA). DNA was extracted once with phenol, once with chloroform, precipitated twice with ethanol, washed with 70% ethanol, dried and treated according to the "G greater than A" specific chemical cleavage method of Maxam and Gilbert (1980). DNA was denatured at 100°C for 2 minutes, cooled on ice and reaction products electrophoresed on 6% or 8% polyacrylamide gels containing 50% urea. Gels were fixed in 10% acetic acid, dried and exposed to X-ray film (Hyperfilm MP; Amersham) at -70°C in the presence of an intensifying screen (Du Pont). Sequence markers were prepared by subjecting end-labelled fragments to G+A or C+T specific reactions (Maxam and Gilbert, 1980).

20. Methylation interference

To ~20 ng end-labelled DNA and 1 µg unlabelled poly[d(A-T)]: poly[d(C-G)] in 200 µl of 50 mM sodium cacodylate, 1 µl DMS was added and incubation continued for 10 minutes at 20°C. The reaction was terminated by the addition of 50 µl 1.5 M sodium acetate, 1.0 M ß-mercaptoethanol containing 4 µg unlabelled poly[d(A-T)]: poly[d(C-G)] and the DNA isolated by two ethanol precipitations. Methylated DNA was resuspended in binding buffer and binding reactions carried out as described above. Specific DNA/protein complexes were separated from free DNA in 6% polyacrylamide gels containing 0.5 X TBE buffer. Electrophoresis was carried out at 200V for one hour after which time bands corresponding to bound or free DNA were excised from the gels and the DNA eluted overnight at 37°C in 2 ml 0.5 M ammonium acetate, 1 mM EDTA, 0.1% SDS, 10% methanol, 50 µg/ml proteinase K, 10 µg yeast tRNA and 1 µg sonicated salmon sperm DNA. The DNA was filtered through glass fibre, and, after concentration with 2-butanol and organic extraction with phenol/chloroform, DNA was ethanol precipitated, resuspended in 10 mM Tris-HCl pH8.0, 1 mM EDTA and reprecipitated by the addition of spermine to 1 mM. DNA was solubilised in 100 µl of

2mM EDTA pH8.0 and heated at 90°C for 15 minutes in a sealed tube. NaOH was added to a final concentration of 100 mM and incubation at 90°C continued for a further 15 minutes. The solution was neutralised by the addition of acetic acid to 100 mM, sodium acetate to 400 mM and the DNA precipitated with ethanol. Samples were processed for electrophoresis as described in "Methylation protection".

21. Ethylation Interference

~20 ng end-labelled DNA in 100 µl of 50 mM sodium cacodylate pH8.0 was mixed with 100 µl of ethanol saturated with ethylnitrosourea and incubated at 50°C for 60 minutes. After addition of 5 µg unlabelled poly [d(A-T)]: poly [d(C-G)] and ethanol precipitation, binding reactions containing the ethylated probe, 1.0 µg unlabelled poly [d(A-T)]: poly [d(C-G)] and 8 µl of affinity purified EBP1 were carried out as described for the standard "Gel electrophoresis DNA binding assay". Specific DNA/protein complexes were separated from free DNA and the DNA eluted and processed as described for "Methylation interference".

22. Orthophenanthroline/copper (OP/Cu+) chemical nuclease footprinting

Binding reactions containing ~10 ng of labelled probe, 1.0 µg unlabelled poly [d(A-T)]: poly [d(C-G)] and 8 µl of affinity purified EBP1 were carried out as described for the standard "Gel electrophoresis DNA binding assay". Specific DNA/protein complexes were separated from free DNA in 6% polyacrylamide gels containing 0.5 X TBE. Electrophoresis was carried out for one hour at 200V. Footprinting reactions with OP/Cu+ were then carried out within the acrylamide matrix by immersion of the gels in 50 mM Tris.HCl pH8.0 (200 ml) and gentle shaking for 15 minutes at room temperature after which time 20 ml of solution A (equal volumes of 40 mM 1,10-phenanthroline in ethanol and 9.0 mM $CuSO_4$ were mixed and diluted 1:10 with water and 20 ml solution B (3-mercaptopropionic acid was diluted 1:200 with water to give 58 mM 3-mercaptopropionic acid) were added and digestion allowed to proceed at room temperature for 10 minutes (Kuwabara and Sigman, 1987). The reaction was quenched by the addition of 20 ml of 28 mM2,9-dimethyl-1,10-phenanthroline (Sigma) and the resulting solution allowed to stand at room temperature for 10 minutes. Gels were washed with distilled water and bands corresponding to bound or free DNA

excised and DNA eluted as described above. Eluted DNA was ethanol precipitated, spermine precipitated, resuspended in formamide dyes, electrophoresed and visualised as described in "Methylation protection".

23. Computer graphics

The three dimensional projection of the EBP1 binding site was modelled on the crystal structure of the B-DNA dodecamer 5'-CGCGAATTGCGC-3'
3'-GCGCTTAACGCG-5'

(Wing *et al.*, 1980). This sequence is similar to the EBP1 binding site and base pair changes, using the CHEMX programme, CHEM model application for molecular modeling, were incorporated to accomodate the sequence

5'-GTGGAAAGTCCCCA-3'
3'-CACCTTTCAGGGGT-5'

which contains the EBP1 binding site that is present in the SV40 enhancer. Van der Waal's surfaces of the nitrogen atoms of bases which interfered with binding when modified by DMS, and phosphorus atoms in the phosphate groups of the DNA backbone whose oxygen atoms were modified by ethylnitrosourea, were displayed on the EBP1 binding site, using the CHEMX programme run on a MicroVax with a Sigmex S5688 terminal. For the purposes of clarity, the van der Waal's radius (which is the ratio of the size of the atoms) was set to 0.1. This was to ensure that the non-highlighted atoms were so small as to appear invisible thus generating the skeletal DNA sructure. Highlighted nitrogen and phosphorus atoms (coloured spheres) were assigned van der Waal's radii of 10 thus distinguishing them from non-highlighted atoms.

RESULTS

Chapter 1. Detection of a HeLa cell protein, EBP1, which interacts with the SV40 enhancer.

1.1 Detection of cellular proteins which interact with the SV40 enhancer

Control of transcription in eukaryotes has been shown to be modulated by the interaction of *trans*- acting, sequence-specific DNA binding proteins with their cognate *cis*- acting recognition sites on the DNA (for review see N.C. Jones *et al.*, 1988). Two distinct types of *cis*- acting regulatory sequences, promoters and enhancers, govern eukaryotic gene expression. Within eukaryotic promoters, DNA sequences responsible for defining the mRNA start site and the rate of transcription initiation have been identified (Grosscheldl and Birnsteil, 1980; Myers *et al.*, 1980; Dynan and Tjian, 1985) with efficient transcription being dependent on the correct spatial arrangement of the various promoter elements (Cochran and Weissman, 1984; McKnight and Kingsbury, 1982; Everett *et al.*, 1983). Considerably more flexibility is exhibited by enhancer sequences in that they are able to activate transcription of linked genes essentially independent of position or orientation with respect to the responding promoter. This enhancing effect was first identified in DNA sequences present in SV40 (Banerji *et al.*, 1981; Moreau *et al.*, 1981) and since then the SV40 enhancer, which is regarded as the prototype enhancer element, has been the subject of intense investigation. It is now clear that it represents a transcriptional element of great complexity.

The SV40 early promoter is ~300 bp long and can be subdivided into three distinct regions: the mRNA start site selection region, the upstream element containing the 21 bp repeats and the enhancer element containing the 72 bp repeats (Figure 1.1). The 21 bp repeats in the upstream region have been shown to contribute to the efficiency of the SV40 early promoter. Deletion of these elements decreases transcription by about 100-fold, with partial deletions having intermediate effects, suggesting that the repeated sequences can complement one another (Myers *et al.*, 1981; Everett *et al.*, 1983; Fromm and Berg, 1982). The GC motif, 5'-CCGCCC-3', is present as two copies in each 21 bp sequence, and is required for both promoter activity and binding of the transcription factor Sp1 (Barrera-Saldana *et al.*, 1985; Kadonaga *et al.*,1986; McKnight and Tjian, 1986).

Figure 1.1 The SV40 early promoter

The regions of the early promoter, including the transcription start site selection region, the upstream element (21 bp repeats) and the enhancer (72 bp repeats) are shown. The major early-early start sites (EES), late-early start sites (LES), A-T rich region (encompassing the TATA motif) and the coordinates of restriction sites are indicated.

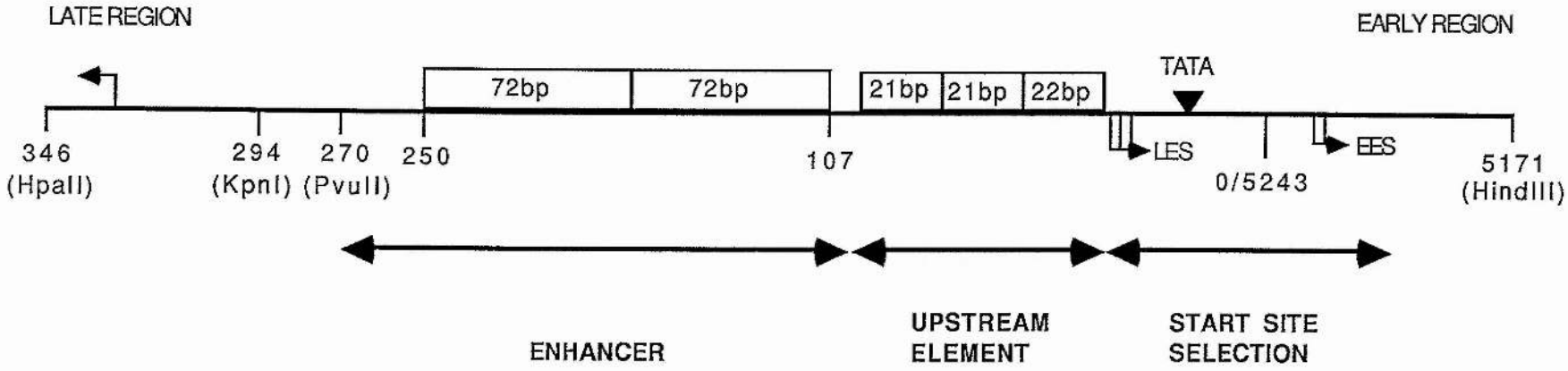
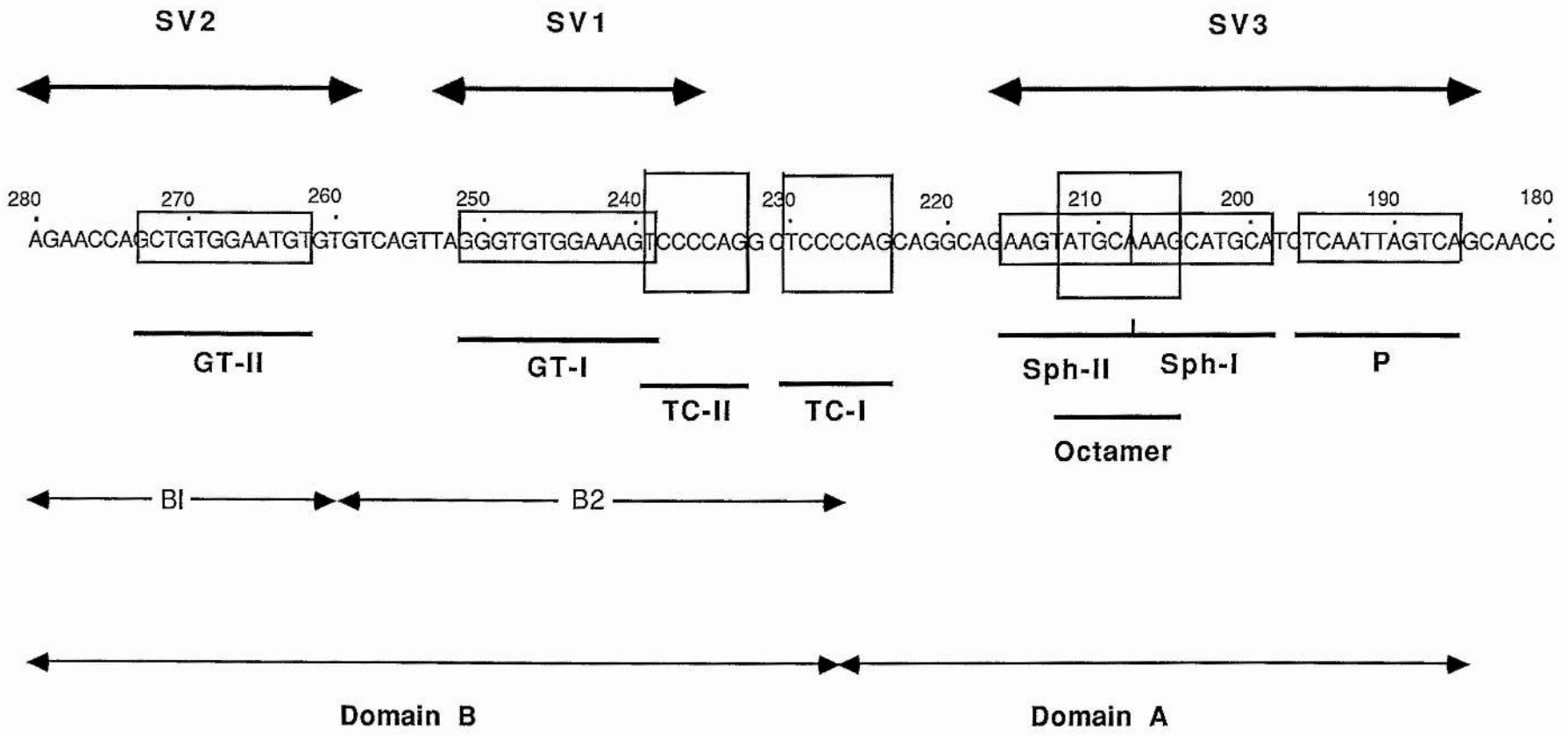

LATE REGION
EARLY REGION
72bp
72bp
21bp
21bp
22bp
TATA
LES
EBS
346
(HpaII)
294
(KpnI)
270
(PvuII)
250
107
0/5243
5171
(HindIII)
ENHANCER
UPSTREAM
ELEMENT
START SITE
SELECTION

Figure 1.2 Organisation of the DNA sequence motifs present in the SV40 enhancer

The DNA sequence of the top strand of the 72 bp repeat element and the 5' flanking sequence is shown. The location of the sequence motifs which have been identified in the enhancer region (Zenke *et al.*, 1986) is indicated by the boxes as well as the positions of the A, B1 and B2 enhancer domains. The SV40 nucleotide sequences present in the synthetic SV1, SV2 and SV3 oligonucleotides are indicated.


SV2
SV1
SV3
280
270
260
250
240
230
220
210
200
190
180
AGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACC
GT-II
GT-I
TC-II
TC-I
Sph-II
Sph-I
P
Octamer
BI
B2
Domain B
Domain A

The SV40 72 bp repeated sequence makes the largest contribution to the efficiency of the early promoter. Deletion of this element results in a 400-fold decrease in transcription (Zenke *et al.*, 1986). However, one copy of the 72 bp repeat can be deleted without having any substantial effect on virus viability (Gruss *et al.*, 1981). The minimum enhancer activity can thus be attributed to one copy of the 72 bp element plus an additional 30 bp of DNA located on the late side of the repeated element. Mutational analysis has identified short DNA segments, or motifs, within the SV40 enhancer (Figure 1.2) which on their own have little enhancing effect but when multimerised, or combined with other motifs, can act in concert to generate enhancer activity (Zenke *et al.*, 1986; Herr and Clarke, 1986; Ondek *et al.*, 1988; Fromental *et al.*, 1988). Each of these sequence motifs represents the binding site for at least one *trans* -acting factor (Wildeman *et al.*, 1986) and it appears to be the interaction of the various *trans* - acting factors with the transcriptional machinery that is responsible for enhancer function. The multiplicity of protein binding sites within the SV40 enhancer creates a situation whereby gene expression may be controlled by the binding of different combinations of interacting factors. In addition, cellular factors have been identified which bind to overlapping motifs within the SV40 enhancer (Xiao *et al.*, 1987b) suggesting that transcriptional activity may also be modulated by the ability of bound proteins to exclude binding of other antagonistically-acting factors.

Since enhancer-mediated transcriptional activation has been shown to involve the interaction of cellular *trans*- acting factors with short segments of DNA (Zenke *et al.*, 1986), the objectives of this research project were (i) to identify and purify a cellular factor that bound to the "core" region of viral and cellular enhancers (which mutational studies have shown is important in enhancement of transcription (Laimins *et al.*, 1982; Weiher *et al.*, 1983; Zenke *et al.*, 1986)); (ii) to then characterise the interactions of this "core"-binding protein with its sequence-specific binding site in DNA; (iii) to define the nucleotide sequence requirements necessary for formation of enhancer binding protein–DNA complexes, and (iv) to suggest possible models as to how interaction of this protein with its cognate site facilitates transcriptional activation.

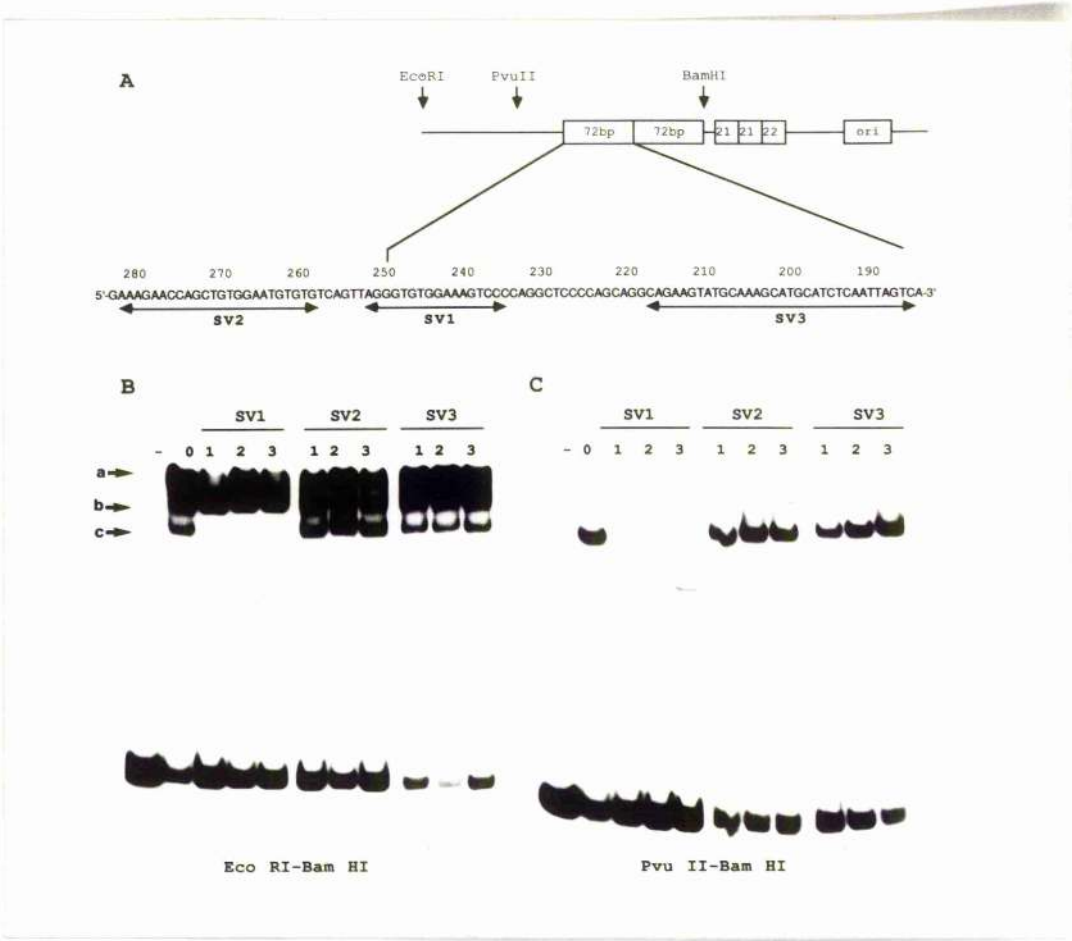
In this study a plasmid, pUC1X72 (Figure 1M), which has one copy of the 72 bp element plus additional upstream SV40 DNA sequences extending to the *Eco* RI site, was used as a source of

enhancer containing DNA , and ,as a means of detecting cellular proteins which bind specifically to the SV40 enhancer, a modification of the gel electrophoresis DNA binding assay originally described by Fried and Crothers (1981), and Garner and Revzin (1981), was employed. This is a rapid and sensitive procedure for the detection of DNA binding proteins which interact with specific and defined sites on DNA. It exploits the observation that complexes of proteins and DNA can jointly enter native polyacrylamide gels and remain stably associated during electrophoresis. The principal of this gel electrophoresis assay is that DNA–protein complexes have a lower electrophoretic mobility than free DNA through polyacrylamide gels. To eliminate interactions between the labelled DNA fragment, which contains the binding site of interest, and cellular non-specific DNA binding proteins, a large excess of unlabelled carrier DNA is included in the reaction mixture (Strauss and Varshavsky, 1984).

A 173 bp [^{32}P]-labelled, *Bam* HI to *Eco* RI, enhancer containing fragment was generated from pUC1X72 and incubated with nuclear extract from uninfected HeLa cells, which had been partially purified by ion-exchange chromatography over DEAE-Sepharose. Interactions between non-specific DNA binding proteins were minimised by the inclusion, in the binding reaction, of a large excess of an equimolar mixture of poly [d(C-G)] and poly [d(A-T)]. Native polyacrylamide gel electrophoresis of the binding reactions resolved free DNA from three distinct DNA-protein complexes labelled "a", "b" and "c" in Figure 1.3B. To determine the sequence specificity of the proteins responsible for the formation of the different DNA-protein complexes, double stranded oligonucleotides were synthesized which spanned a number of the motifs recognised in the SV40 enhancer. SV1 contains the GT-I and part of the TC-II motifs, while SV2 contains the GT-II motif, and SV3 the Sph-I, Sph-II, octamer and part of the P motif (Figure 1.2; Xiao *et al.*, 1987a; for review see N.C. Jones *et al.*, 1988). Competition analysis, in which an excess of the unlabelled double stranded oligonucleotide is included in the binding reaction, indicated that the formation of complex "a" was diminished in the presence of SV1 and to a lesser extent by SV2, whereas formation of complex "b" was unaffected by the presence of SV1, SV2 or SV3 (Figure 1.3B). Formation of complex "c" was abolished by inclusion of the SV1 oligonucleotide in the binding reaction, but was unaffected by the inclusion of the SV2 or SV3 oligonucleotides. The origin of complex "c" was investigated further by repeating these experiments using a shorter [^{32}P]- labelled DNA fragment of

Figure 1.3 Detection of enhancer binding proteins in HeLa cell nuclear extracts

(**A**) Organization of the SV40 transcriptional control region. Relevant restriction enzyme cleavage sites and important elements of the promoter and enhancer present in plasmid pMKD 231 are indicated. Shown below is the DNA sequence of the enhancer region. DNA sequences present in the double stranded synthetic oligonucleotides SV1, SV2 and SV3 are also indicated. [^{32}P]-labelled probes containing the 72bp enhancer element, *Eco* RI-*Bam* HI (**B**) or *Pvu* II-*Bam* HI (**C**) digested pUC1X72, were incubated with 2 µg nuclear extract (partially purified over DEAE-Sepharose) in the standard gel electrophoresis DNA binding assay as described in "Materials and Methods" (lanes (-) contained no nuclear extract). Also included in the reaction mixture was 25 (lanes 1), 50 (lanes 2) or 100 ng (lanes 3) of unlabelled SV1, SV2 or SV3 double stranded synthetic oligonucleotides.



97 bp which extends from the *Bam* HI to the *Pvu* II site. Only one specific DNA-protein complex was resolved by polyacrylamide gel electrophoresis after incubation of the labelled probe with the nuclear extract (Figure 1.3C). Equivalence of this single DNA-protein complex with complex "c" observed previously was suggested by a similar electrophoretic mobility, and by the fact that formation of the complex was competed for by inclusion of the SV1 oligonucleotide, but not by the SV2 or SV3 oligonucleotides (Figure 1.3C).

1.2 DNase I protection of the SV40 enhancer by partially purified HeLa cell factors

DNase I protection (or "footprinting") experiments were employed to define the binding sites of the partially purified proteins on the SV40 enhancer. This technique allows the identification of the nucleotide sequences covered by sequence-specific DNA binding proteins (Galas and Schmitz, 1978). DNase I footprinting involves incubating a DNA fragment, radioactively labelled at one end, with the protein, and digesting with limiting amounts of DNase I, sufficient to introduce, on average, one cut per DNA molecule. In the presence of a sequence-specific DNA binding protein, the DNA bound by the protein is less accessible to digestion with DNase I. Comparison of the DNase I cleavage products produced in the presence, and absence, of added DNA binding proteins after electrophoresis on a high resolution urea-polyacrylamide gel, identifies which phosphodiester bonds are protected from DNase I cleavage in the DNA binding protein (Figure 1.4). The presence of a sequence-specific DNA binding protein will alter DNase I cutting at the site of interaction, with respect to naked DNA, in that this region of the DNA will be protected from DNase I cleavage. This leads to weakening or disappearance of the corresponding bands detectable by gel elecrophoresis (Galas and Schmitz, 1978). The precise sequence can be identified by comparing the DNase I cleavage pattern with a DNA sequence ladder (Maxam and Gilbert, 1980). The identification of the limits of the "footprint", or protected region, is often facilitated by the observation that the phosphodiester bonds directly flanking the covered sequences are hypersensitive to DNase I, probably as a result of hydrophobic interactions between the enzyme and the DNA-bound protein.

A 5'-[^{32}P]-labelled DNA fragment containing SV40 enhancer sequences was incubated with the partially purified HeLa cell nuclear extracts , subjected to DNase I digestion, and the cleavage

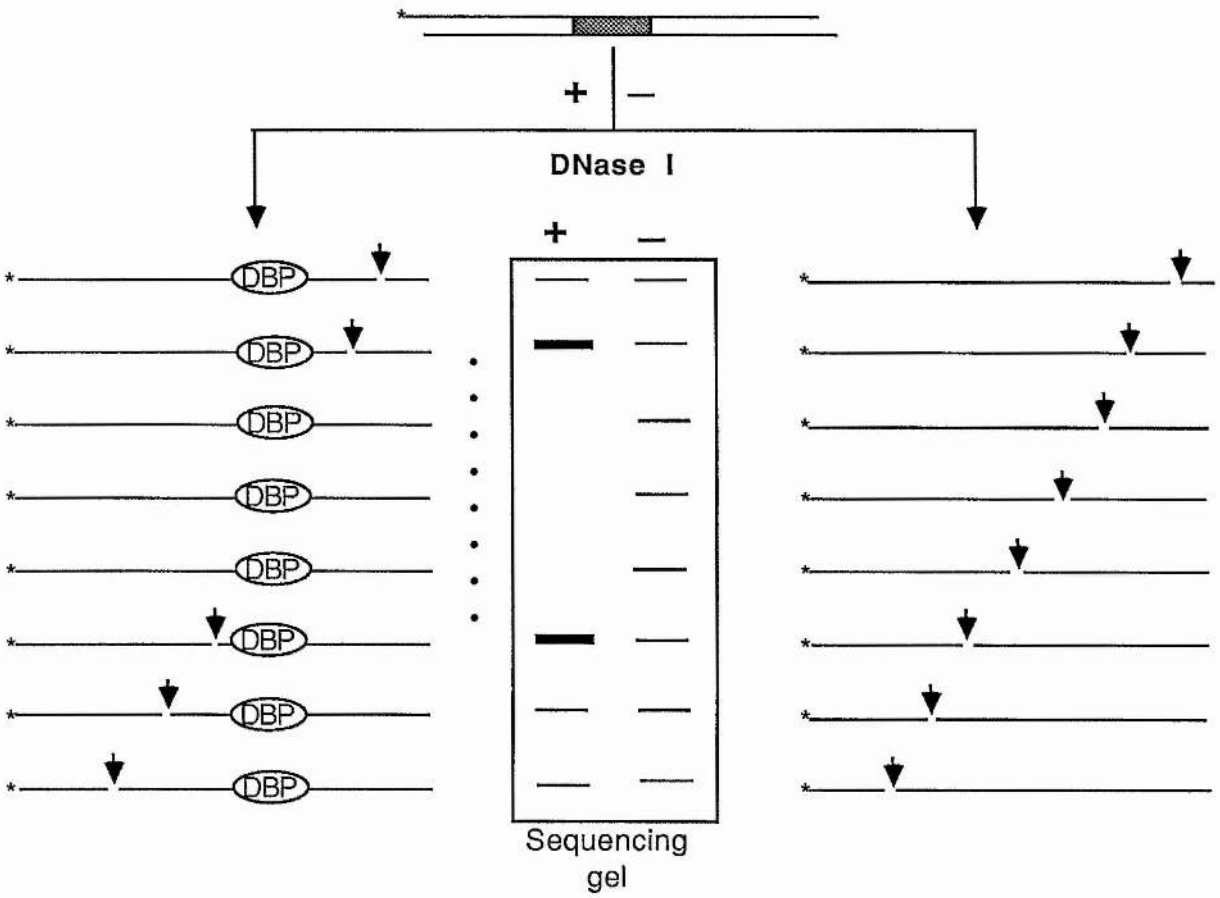
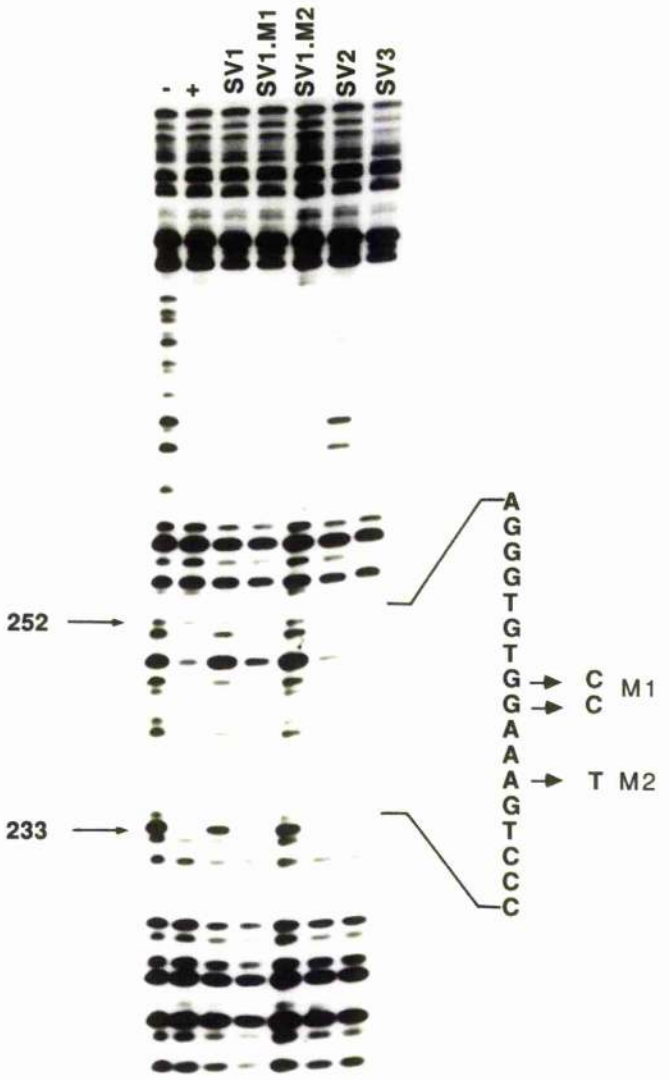


Figure 1.4 DNase I protection assay

The end-labelled DNA fragment is incubated with limiting amounts of DNase I (sufficient to introduce one cut per DNA molecule) in the presence (+) or absence (-) of a sequence-specific DNA binding protein (DBP). Cleavage by DNase I is indicated by the vertical arrows. (For purposes of clarity, the non-labelled lower strand has not been shown). The corresponding sequencing gel is shown with the protected region indicated by the dotted line. The DNase I hypersensitive sites flanking the footprint are also indicated by the bold horizontal bars.

Figure 1.5. DNase I protection of the SV40 enhancer by partially purified enhancer binding factors

A *Bam* HI to *Eco* RI fragment from pUC1X72, 5'-end labeled at the *Bam* HI site, was incubated with 40 ug of nuclear extract (DEAE-Sepharose fraction) and digested with DNase I as described in "Materials and Methods" (lane (-) contained no nuclear extract). Included in the reaction mixture was 100 ng of unlabeled double stranded oligonucleotides SV1, SV1.M1, SV1.M2, SV2 or SV3. After treatment with DNase I, reaction products were electrophoresed in an 6% denaturing polyacrylamide gel and the cleavage products visualized by autoradiography. Oligonucleotides SV1.M1 and SV1.M2 are derivatives of SV1 in which cytosines 244 and 245 have been altered to guanines (SV1.M1) and thymine 241 has been altered to adenine (SV1.M2).



products displayed (Figure 1.5). In the presence of HeLa cell protein, two areas of the SV40 enhancer were protected from digestion by DNase I (Figure 1.5, lane +). The protected regions extended from position 233 to 253, which contains the GT-I motif (Figure 1.2), and position 262 to 290, which contains the GT-II motif. Addition of a large excess of unlabelled SV1 double stranded oligonucleotide to the binding reaction abolished protection between positions 233 and 252, but had no effect in the region from 262 to 290 (Figure 1.5, lane SV1). Conversely, addition of an excess of unlabelled SV2 double stranded oligonucleotide reduced the protection over the GT-II motif, whereas the footprint spanning the GT-I motif was unchanged (Figure 1.5, lane SV2). The pattern of DNase I protection was not significantly altered by the inclusion of excess SV3 double stranded oligonucleotide in the reaction mixture (Figure 1.5, lane SV3).

To investigate the sequence requirements for specific binding in the GT-I region, competition assays in which double stranded oligonucleotides corresponding to SV1 with specific base pair alterations, were included in the binding reactions. Alteration of the two guanine residues at positions 244 and 245 to cytosines (SV1.M1) abolished the ability of the oligonucleotide to compete for binding of the nuclear factors (Figure 1.5, lane SV1.M1), whereas changing adenine 241 to thymidine (SV1.M2) did not alter the ability of the oligonucleotide to compete for binding of the nuclear factors (Figure 1.5, lane SV1.M2).

1.3 Effects of SV40 enhancer sequences on transient gene expression in HeLa cells

Genetic evidence has demonstrated that the SV40 transcriptional enhancer is composed of multiple DNA sequence motifs that can act synergistically to stimulate transcription from the SV40 early promoter (Zenke *et al.*, 1986; Herr and Clarke, 1986; Schirm *et al.*, 1987; Davidson *et al.*, 1986; Fromental *et al.*, 1988). In addition, it has been shown that multiple tandem copies of the "core" enhancer sequence function autonomously to activate transcription of a linked gene in a cell type-specific manner (Ondek *et al.*, 1987; Schirm *et al.*, 1987). Having identified a protein in HeLa cell nuclear extracts which binds in the GT-I region of the SV40 enhancer *in vitro*, we wished to determine whether or not the DNA sequences involved in the recognition event could activate transcription of a linked gene *in vivo*. To investigate the *in vivo* activity of the factor binding site, we

employed a strategy similar to that used by Veldman and coworkers (1985), in which tandem copies of DNA sequence elements were used to study redundant sequence elements of the polyomavirus enhancer that activate DNA replication and gene expression.

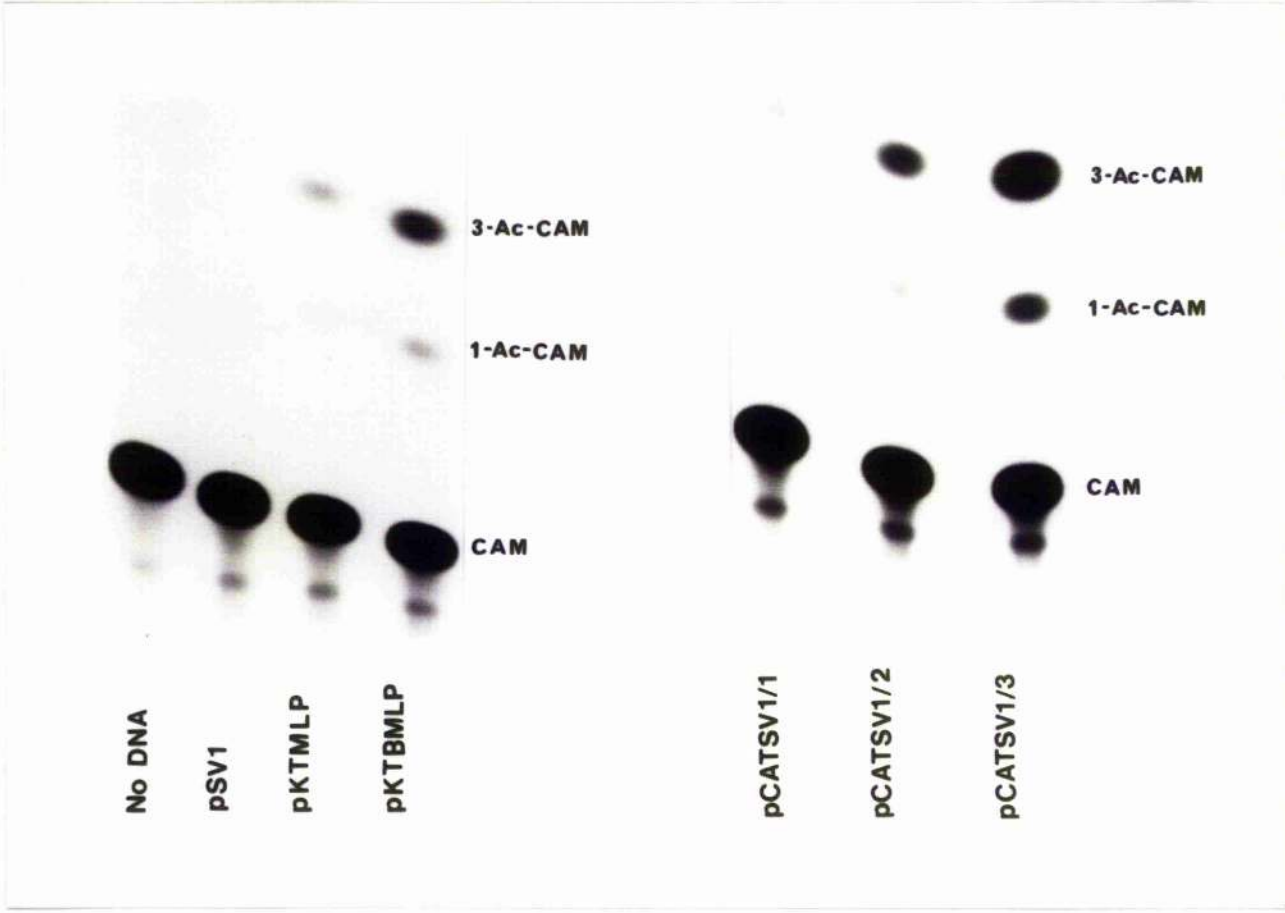
A powerful method for studying the role of a particular DNA sequence in transcription regulation is to combine the putative regulatory region with a "reporter" gene, to introduce the recombinant into appropriate tissue culture cells, and to measure expression of the reporter gene under various conditions. The reporter gene should encode an easily assayable protein readily distinguishable from endogenous cellular proteins. The bacterial gene for chloramphenicol 3-O-acetyltransferase (CAT) has proved to be extremely useful for this purpose because mammalian cells generally lack endogenous CAT activity (Gorman, *et al.*,1982), and sensitive enzymatic assays for CAT are available. We have used the customary thin-layer chromatography (TLC) -based assay (Shaw, 1975; Gorman *et al.*, 1982) in which [^{14}C]-labelled chloramphenicol (CAM) is used as a substrate, and TLC followed by autoradiography is used to separate and locate the acetylated (3-Ac-CAM and 1-Ac-CAM) and unacetylated CAM.

Recombinant expression vectors were constructed which contained 0, 1, 2 or 3 copies of the SV1 double stranded oligonucleotide, tandemly ligated in a head-to-tail orientation, upstream of the Adenovirus major late promoter (Ad MLP), linked to the bacterial CAT gene (Figure 4M). These recombinants were transfected into sub-confluent HeLa cell monolayers, harvested after 48 hours and CAT activity determined by TLC followed by autoradiography.

Transfection of the recombinant pSV1, which contains SV40 control sequences but no CAT gene, resulted in no detectable levels of CAT activity (Figure 1.6). In the case of plasmid pKTMPL (which does not contain any upstream enhancer sequences) a low basal level of CAT activity is observable which is probably a reflection of the promoter strength. When two copies of the SV40 72 bp repeat are present upstream of the Ad MLP, as in plasmid pKTBMLP, a substantial increase in CAT activity is observed. However, this increase in activity is considerably less than expected when two copies of the 72 bp element are present since deletion of this element has been shown to decrease transcription, from a responding promoter, by 400-fold (Zenke *et al.*, 1986). A similar increase in activity is also observed when two tandem copies of the SV1 oligonucleotide are inserted upstream of the Ad MLP (plasmid pCATSV1/2). That approximately the same levels of

Figure 1.6 CAT activity directed by SV40 enhancer-containing recombinants

HeLa cells were transfected with recombinant plasmids containing the Adenovirus major late promoter (Ad MLP) fused to the bacterial CAT gene (pKTMLP) and either two copies of the 72 bp repeat element (pKTBMLP) inserted upstream of the Ad MLP, or 1, 2 or 3 copies of the SV1 double stranded oligonucleotide (pCATSV1/1, 2 and 3 respectively). Plasmid pSV1 contains SV40 control sequences but no CAT gene. Cell extracts were prepared for CAT assay 48 hours after transfection and CAT activity determined by TLC followed by autoradiography. CAT activity was measured by determining the amount of acetylated chloramphenicol (3-Ac-CAM and 1-Ac-CAM) produced from [^{14}C] chloramphenicol (CAM).



CAT activity are observed for pKTBMLP and pCATSV1/2 may simply reflect differences in tissue culture. Insertion of only one copy of the SV1 oligonucleotide (plasmid pCATSV1/2) resulted in a similar basal level of CAT activity as observed using pKTMLP. However, when three tandem copies of SVI were present upstream of the MLP (plasmid pCATSV1/3) CAT activity was increased several-fold over that seen with pKTBMLP or pCATSV1/2. These results would therefore indicate that tandemly ligated copies of the SV1 oligonucleotide containing the GT-1 motif (Figure 1.2) can function as a transcriptional enhancer *in vivo* when inserted into an appropriate expression vector.

Chapter 2. Purification and Characterisation of EBP1

2.1 Purification of a cellular protein, EBP1, which binds to the "core" region of the SV40 enhancer

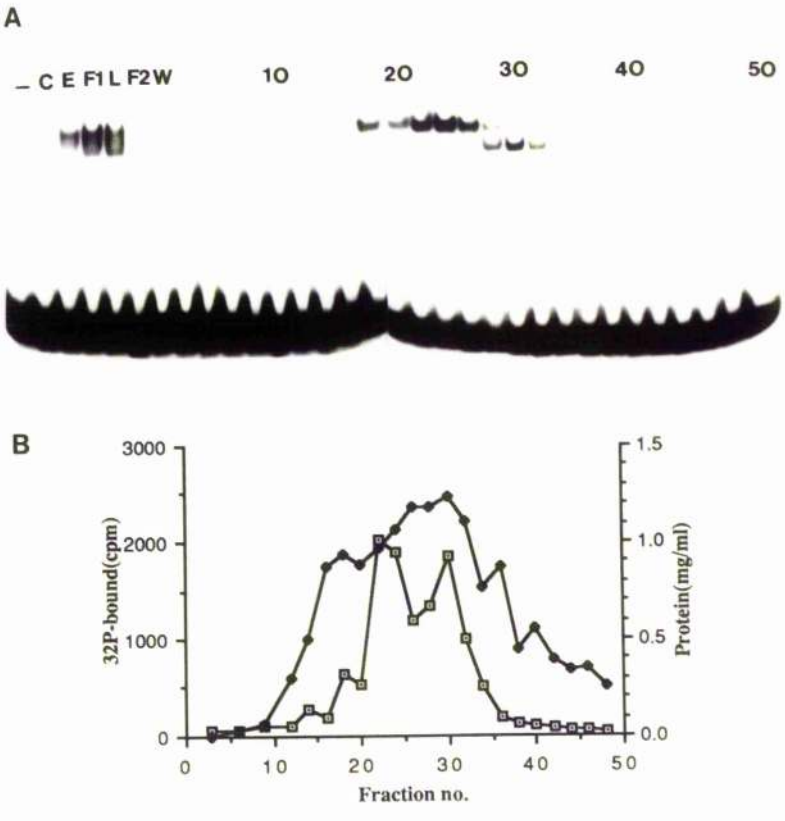
The factor which binds to the GT-I motif (Figure 1.2) of the SV40 enhancer, was also shown to bind to the SV1 double stranded oligonucleotide (Figures 1.3 and 1.5). Thus as a means of monitoring the fractionation, and purification, of the GT-I binding fraction, [^{32}P]-labelled SV1 double stranded oligonucleotide was used as probe in the gel electrophoresis DNA binding assay (Figures 2.1 and 2.2).

Uninfected HeLa cells (40 g wet weight) were fractionated into nuclei and cytosol and the nuclei extracted with 0.35 M NaCl. The crude nuclear extract, which contained the bulk of the specific DNA binding activity, was depleted of nucleic acids by passage over DEAE-Sepharose equilibrated with 0.3 M NaCl. After reduction of the NaCl concentration to 0.5 M, the flow through from the first DEAE fractionation was applied to a second DEAE-Sepharose column and bound proteins eluted with a linear gradient from 0.05 to 0.4 M NaCl. Individual fractions were monitored for DNA binding activity using the gel electrophoresis DNA binding assay. Using this procedure it was possible to resolve two distinct species which bound to the SV1 double stranded oligonucleotide (Figure 2.1). The two species appear to have identical sequence specificities and bromodeoxyuracil cross-linking experiments (Figure 2.4) indicated that the cross-linked polypeptide is of identical molecular weight in each case.

Further purification of the enhancer binding species by conventional ion-exchange chromatography was unsuccessful in that yields of active DNA binding protein were very poor, and

Figure 2.1 Fractionation of HeLa nuclear extract

Dialysed nuclear extract was loaded onto a DEAE-Sepharose CL-6B column and bound proteins eluted with a linear gradient of 0.05 to 0.4M NaCl. Fractions were incubated with [^{32}P] labelled SV1 double stranded oligonucleotide under standard conditions employed in the gel electrophoresis DNA binding assay. DNA/protein complexes were resolved on 6% polyacylamide gels. Gels were either dried down and autoradiographed (**A**) or enhancer protein binding activity (□) quantified by excising bound species from the gels and determining the radioactivity present by Cerenkov counting (**B**). Protein concentration (■) was determined by the method of Bradford (1976). Fractions sampled are indicated below each panel. C represents cytoplasmic extract, E nuclear extract, F1 material passing through the first DEAE- Sepharose CL-6B column and L, F2 and W representing column load, flow through and 0.05M NaCl wash from the second DEAE-Sepharose CL-6B column.



preparations were somewhat heterogeneous. Since the enhancer binding species observed in the DEAE-Sepharose column eluates bound the SV1 oligonucleotide, the next logical step to obtaining highly purified preparations of the protein would be to use the sequence-specific SV1 oligonucleotide as an affinity ligand. Indeed, a number of DNA binding proteins, including various RNA and DNA polymerases, hormone receptors and repressors have in the past been purified by non-specific DNA-Cellulose and DNA-Agarose affinity chromatography (Alberts and Herrick, 1971; Arndt-Jovin *et al.,* 1975). In this study, various types of affinity matrices were examined to enable the development of a rapid and sensitive method for purification of sequence-specific DNA binding proteins to near homogeneity. A DNA affinity resin was required which could maximise the recovery of the desired protein but which minimised contamination by other proteins. In addition, the DNA had to be accessible to the sequence-specific binding protein and to be stably bound to the protein.

Early attempts at generating an efficient affinity matrix were based on modifications of the schemes used to purify the HeLa cell factors Sp1 (Kadonaga and Tjian, 1986) and nuclear factor I (Rosenfeld and Kelly, 1986). Tandemly ligated double stranded SV1 oligonucleotides were covalently coupled to aminophenylthioether- cellulose (APT-cellulose) via a diazo-linkage between the double stranded oligonucleotides and the APT-cellulose (Seed, 1982). The second DEAE-Sepharose fractions, shown in the gel electrophoresis DNA binding assay (Figure2.1) to contain enhancer binding activity, were pooled and further purified over this affinity matrix. However, very poor yields of specific SV40 enhancer binding protein were recoverable on this recognition site affinity matrix. In view of this, a second type of sequence-specific affinity matrix was prepared. This was generated by incorporating biotin-11-dUMP into the ends of double stranded SV1 oligonucleotides and coupling them to Streptavidin- Agarose. Unfortunately, passage of the active DEAE-Sepharose fractions over this matrix did not improve the yields of specific enhancer binding activity above those obtained using APT-Cellulose. Successful recognition site affinity matrices were finally prepared by covalently coupling tandemly ligated, double stranded SV1 oligonucleotides, or double stranded SV1 oligonucleotides containing a primary amine group at their 5'-ends (Aminolink- oligonucleotides), to CNBr-activated Sepharose, essentially as described by Kadonaga and Tjian (1986).

Figure 2.2 Recognition site affinity chromatography of EBP1

A. Poly [d(G-C)] and poly [d(A-T)] were added to the DEAE-Sepharose fractions containing EBP1 activity and passed over an affinity matrix composed of tandemly ligated SV1 oligonucleotides covalently linked to CNBr-activated Sepharose-4B. Bound proteins were eluted with 1.0M NaCl and fractions assayed for EBP1 activity . L, load; FT, flow though; W, wash; 1-8, fractions from 1M NaCl elution.

B. Fractions from each stage of the EBP1 purification were precipitated with TCA and analysed by SDS-polyacrylamide gel electrophoresis and silver staining. Equivalent amounts (28 units) of each affinity column eluate were applied to each lane. The molecular weights of the marker proteins are indicated. D, DEAE-Sepharose fractions; 1, 2 and 3; first, second and third round affinity column eluates; M, protein molecular weight standards.

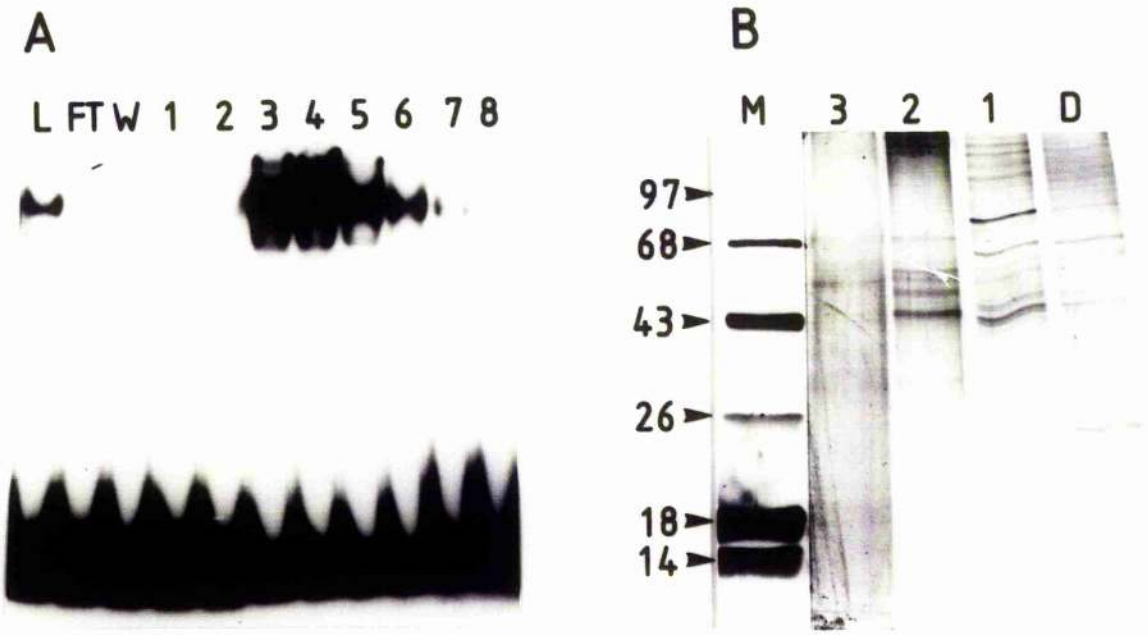


Table 2.1 Purification of EBP1 from HeLa cell nuclear extracts

Fraction	Protein (mg)	Units (*)	Yield (%)	Specific activity (units/mg protein)	Fold purification
Nuclear extract	861	84378	100	98	1
DEAE pool	77	36400	43	473	5
1st dsSV1 pool	1.3	25480	30	19600	210
2nd dsSV1 pool	0.08	19110	22.6	238875	2560
3rd dsSV1 pool	0.008	11275	13.4	1409375	15150

* One unit of activity represents the retention of 15 fmol specific DNA bound in the DNA/protein complexes in the gel electrophoresis DNA binding assay.

+ Estimate based on comparative silver staining.

Fractions from the second DEAE-Sepharose chromatography which contained the two enhancer binding species were pooled and further purified to near homogeneity by recognition site affinity chromatography (Rosenfeld and Kelly, 1986; Kadonaga and Tjian, 1986) on the SV1/CNBr-Sepharose resins. The pooled DEAE-Sepharose fractions were mixed with equimolar amounts of poly [d(A-T)] and poly [d(G-C)], as non-specific competitor DNAs, and applied to the DNA affinity resins. Whereas the bulk of the proteins in the DEAE fractions flowed through the affinity column, the SV1 binding activity was specifically retained on the column and could be eluted by raising the NaCl concentration to 1.0 M (Figure 2.2A). Two further rounds of affinity chromatography resulted in a preparation in which the specific binding activity had increased by 15,000– fold when compared to the crude nuclear extract (Table 2.1). (Yields of specific enhancer binding activity obtained by using tandemly ligated SV1 oligonucleotides or Aminolink- SV1 oligonucleotides were very similar).

This method of affinity chromatography exhibits some noteworthy differences to many previously published procedures. The most important feature is probably the addition of non-specific competitor DNA directly to the partially purified protein fraction prior to passage over the DNA-Sepharose resin (Kadonaga and Tjian, 1986). The inclusion of competitor DNA is to minimise the retention of non-specific DNA binding proteins on the affinity resin. In the past, plasmid DNA containing multiple binding sites has been used in generating sequence-specific resins (Rosenfeld and Kelly, 1986). Synthetic oligonucleotides were used in this study to maximise the specificity of the resins and also because they are more selective than plasmids containing multiple recognition sites. The oligonucleotides are quickly and easily prepared. In addition, affinity chromatography has generally been carried out using resins whereby DNA has been adsorbed to Cellulose (Alberts and Herrick, 1971; Rosenfeld and Kelly, 1986). In contrast, this method uses DNA-Sepharose resins prepared by covalent coupling of DNA to the solid support thus minimising loss of specific binding sites due to leaching of the DNA.

Another important feature of this method of protein purification is the inclusion of non-ionic detergents in the chromatography buffers. The high recovery of specific DNA binding activity was aided by the addition of 0.1% NP40 to both the loading and elution buffers. In addition, n-octyl-ß-D-glucopyranoside (to a final concentration of 3 mM) was included in the buffers used during the recognition site affinity chromatography steps (Sorger and Pelham, 1987; Treisman, 1987). It is

likely that these detergents inhibit the formation of protein aggregates which may accumulate during fractionation thereby releasing the sequence-specific DNA binding proteins for interaction with their binding sites.

To determine the purity at different stages during the course of fractionation, samples from various steps in the purification, which contained equivalent amounts of enhancer binding activity, were precipitated with TCA and analysed by SDS polyacrylamide gel electrophoresis followed by silver staining. The third round affinity fraction contained predominantly a 57,000 molecular weight polypeptide (Figure 2.2B). From 40 g of HeLa cells we obtained 8 μg of purified protein with 13.4% recovery of specific DNA binding activity (Table 2.1). Since the purified protein bound to the SV1 region of the SV40 enhancer, this protein has been designated EBP1.

2.2 Identification of the 57,000 molecular weight polypeptide (EBP1) as the specific DNA binding species

To confirm that the 57,000 molecular weight polypeptide (EBP1) contained the activity responsible for binding to the SV1 oligonucleotide, DNA-bound protein was cross linked to bromodeoxyuracil- substituted, [^{32}P]-labelled DNA (Ogata and Gilbert, 1977). When the methyl group of thymidine is substituted with bromine, by incorporation of bromodeoxyuridine into DNA, the DNA remains functional, both *in vivo* and *in vitro,* although some changes can occur in the binding of proteins as demonstrated by *lac* repressor binding tenfold more tightly to bromine-substituted operator DNA than to unsubstituted operator (Lin and Riggs, 1972). UV irradiation of substituted DNA displaces the bromine and generates a free radical at the 5-position of the deoxyuridine. A covalent cross link between a protein bound specifically to this DNA can then form if an appropriate amino acid side chain is close to the free radical. After nuclease treatment of the complex, the molecular weights of the cross linked proteins can be determined by electrophoresis on SDS polyacrylamide gels. The extensive nuclease treatment ensures that only short oligonucleotides from the probe remain covalently cross linked to the proteins during electrophoresis (Chodosh *et al.*, 1986).

The SV1 double stranded oligonucleotide was first inserted into the *Bam* HI site of M13mp8 and single stranded phage DNA prepared. A complementary strand was synthesized containing

Figure 2.3 The binding of EBP1 to bromodeoxyuridine-substituted DNA

A. A bromodeoxyuridine-substituted, [^{32}P]-labelled, double stranded DNA fragment containing the SV1 sequence was incubated in the presence (+), or absence(-) of partially purified EBP1 under conditions employed in the standard gel electrophoresis DNA binding assay, DNA–protein complexes resolved on 6% polyacrylamide gels and visualised by autoradiography. Arrows depict the positions of the DNA–protein complexes (B) or free DNA (F).

B. Competition analysis of EBP1 binding. Reactions were prepared as in **A** but with the inclusion of 10 (lane 1), 25 (lane 2), 50 (lane 3) or 100 ng (lane 4) of unlabelled, double stranded SV1 oligonucleotide as competitor, in the binding mixture. Complexes were resolved and visualised as described in **A**.

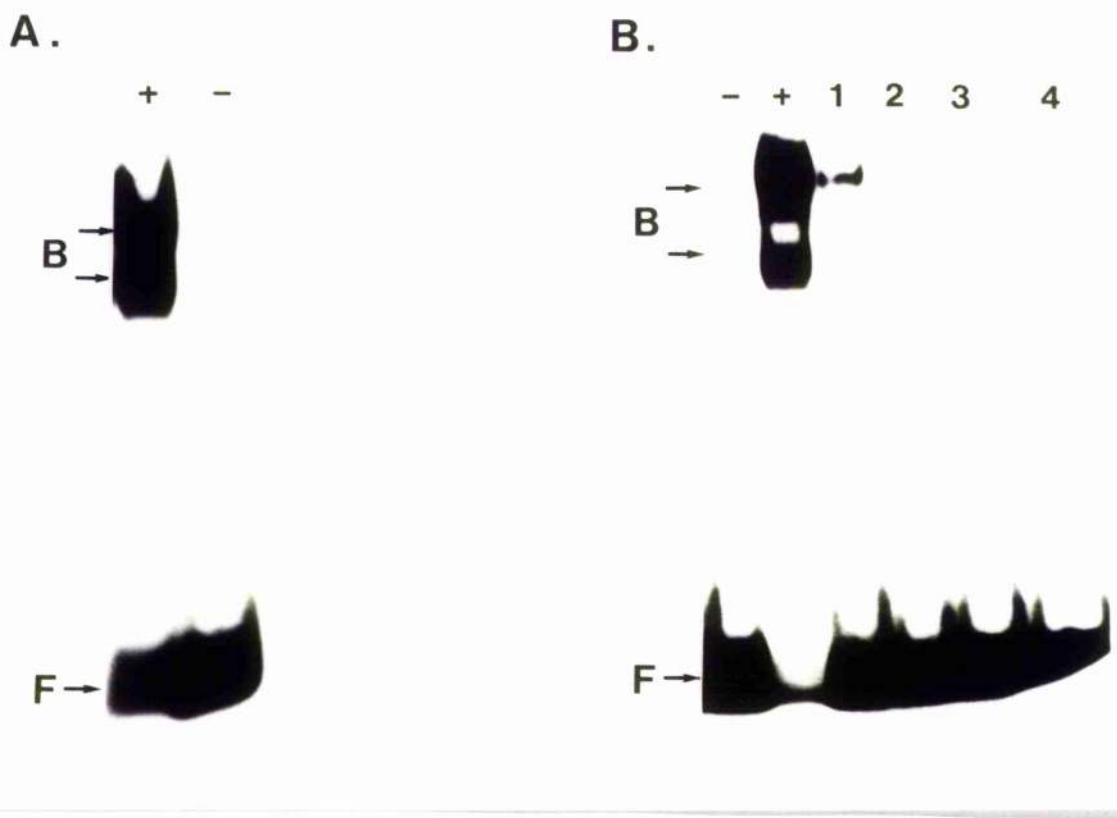
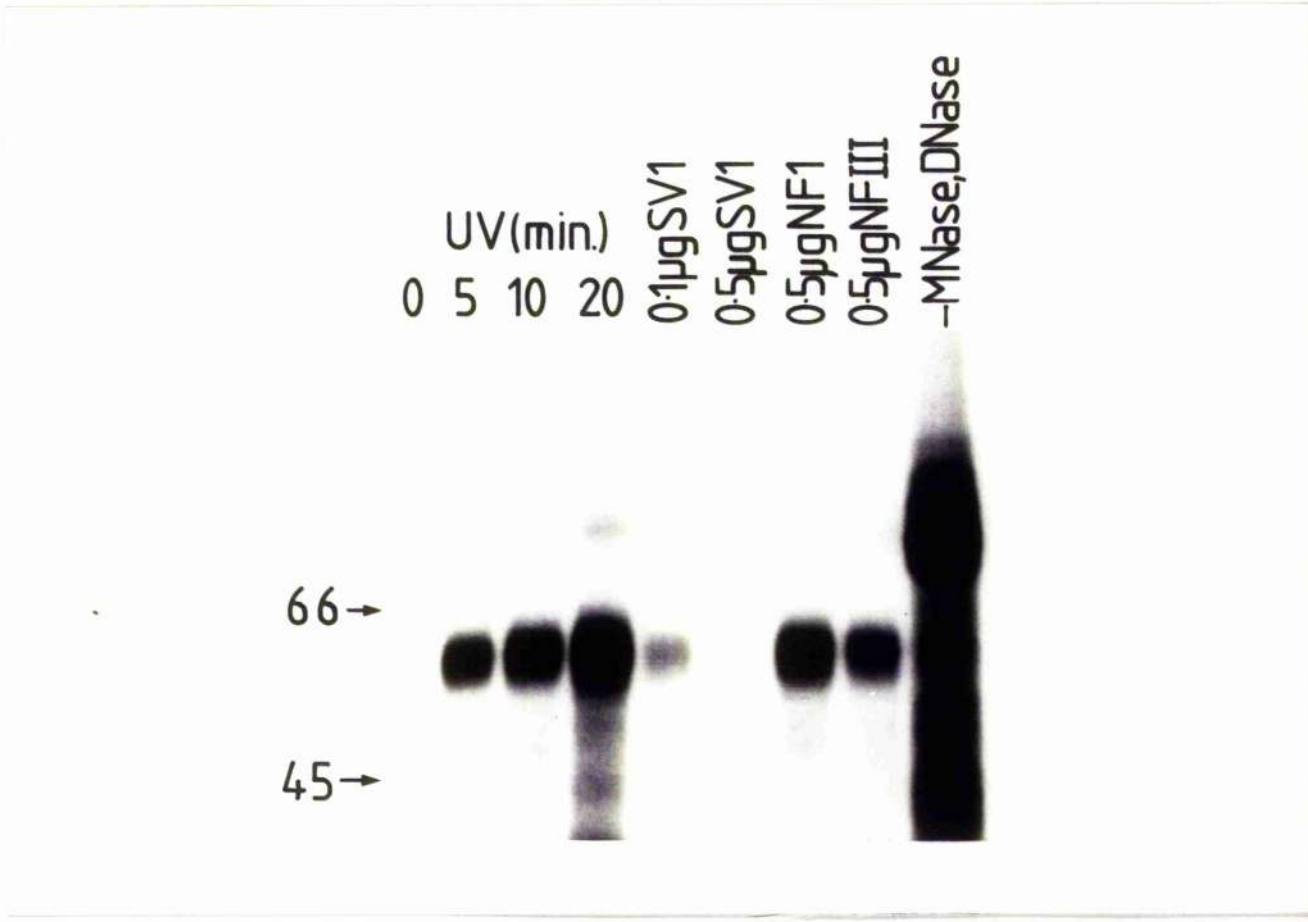


Figure2.4 Bromodeoxyuracil cross-linking of EBP1 to its binding site on the SV40 genome

A bromodeoxyuridine substituted, [^{32}P]-labelled double stranded DNA fragment containing the EBP1 binding site was prepared as described in "Materials and Methods". Probe DNA was incubated with affinity purified EBP1 under conditions employed in the gel electrophoresis DNA binding assay. Once equilibrium had been reached, samples were irradiated with UV light followed by treatment with DNase I and micrococcal nuclease as indicated. Reaction products were fractionated on 10% SDS polyacrylamide gels and visualised by autoradiography. The positions of the 66,000 and 45,000 molecular weight protein standards are indicated by the arrows. UV irradiation, in the absence of competitor was for 0, 5, 10 and 20 minutes. In the competition analysis EBP1 was allowed to bind in the presence of 0.1 µg SV1, 0.5 µg SV1, 0.5 µg NFI or 0.5 µg NFIII double stranded oligonucleotides prior to UV irradiation for 10 minutes. Reaction productes which were UV irradiated for 10 minutes without subsequent DNase I or mirococcal nuclease digestion are also shown (-MNase,DNase).



[^{32}P]dGMP and bromodeoxyuridine (as a substitute for thymidine). Restriction enzyme cleavage with *Eco* RI and *Pst* I released a small labelled fragment that contained the EBP1 binding site. Substitution of [α-^{32}P]deoxycytidine and bromodeoxyuridine for cytidine and thymidine did not affect the binding of EBP1 to the probe as determined in the standard gel electrophoresis DNA binding assay (Figure 2.3A). In addition, it was demonstrated that unlabelled SV1 double stranded oligonucleotide could successfully compete for EBP1 binding to the substituted DNA (Figure 2.3B).

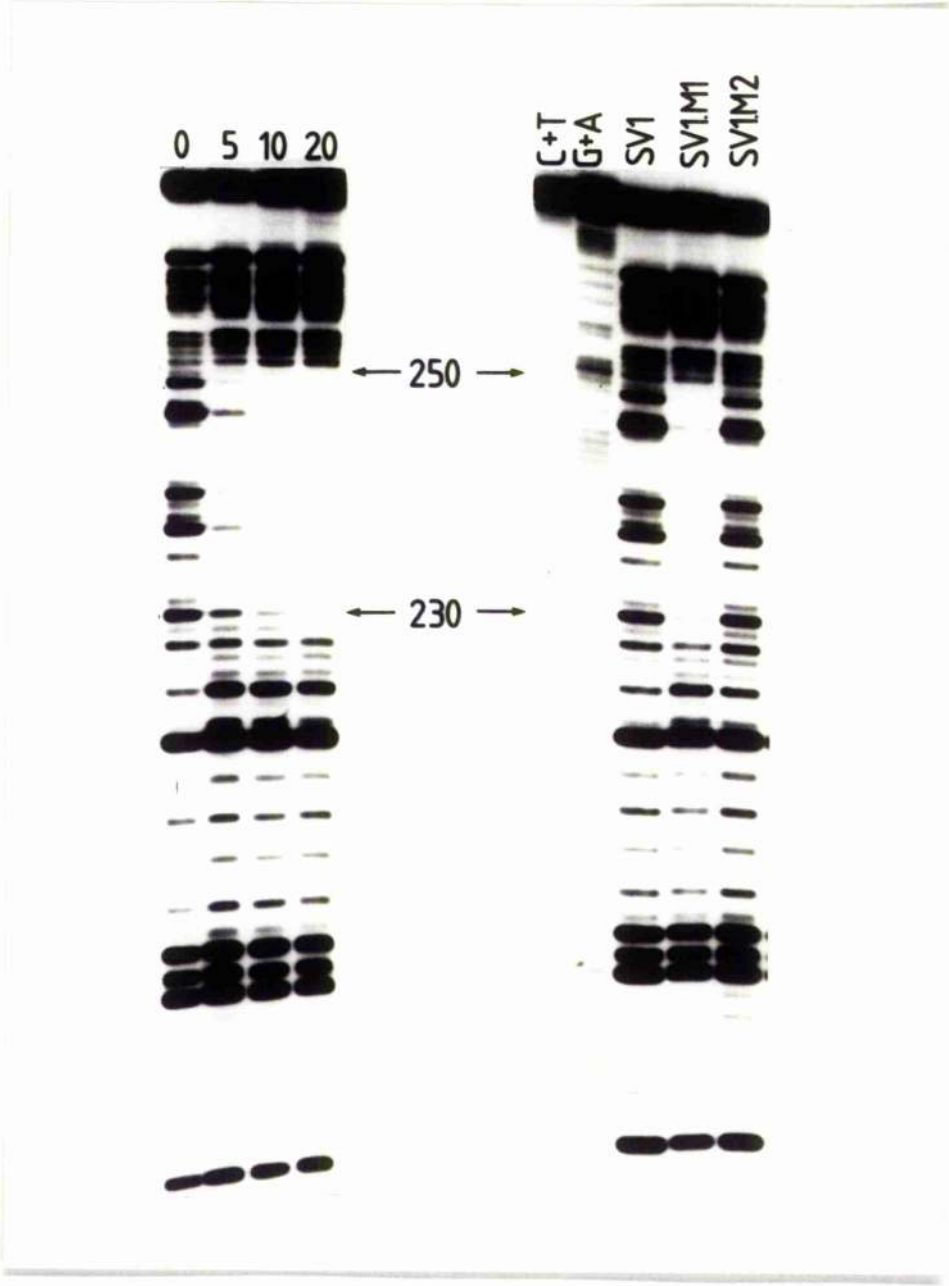
In the cross linking studies, probe DNA was incubated with affinity purified protein under conditions similar to those employed in the gel electrophoresis DNA binding assay. Once equilibration had been reached the samples were irradiated with UV light, digested with nucleases and analysed by autoradiography after SDS polyacrylamide gel electrophoresis. When the products of the cross-linking reactions were analysed without nuclease treatment it was evident that the labelled oligonucleotide had been cross-linked to a polypeptide present in the binding reaction (Figure 2.4). Treatment of the cross-linked products with DNase I and micrococcal nuclease, prior to gel electrophoresis, decreased the apparent molecular weight of the labelled species and identified a polypeptide with a molecular weight of ~60,000 (Figure 2.4). The labelled species accumulated with increased time of irradiation (Figure 2.4) but no cross-linking was observed if BSA was substituted for EBP1 in the DNA binding reaction (data not shown). Addition of an excess of unlabelled SV1 oligonucleotide to the binding reaction abolished cross-linking, whereas inclusion of double stranded oligonucleotides containing the binding sites for nuclear factor I (NFI) or nuclear factor III (NFIII) had little effect (Figure 2.4). These data indicate that a polypeptide of approximately 60,000 is responsible for specific binding to the GT-I region of the SV40 enhancer and is the major component present in our most highly purified preparations.

2.3 DNA protected by EBP1 from DNase I cleavage

To define more precisely the region of the SV40 enhancer that is bound by EBP1, DNase I protection experiments were employed. A 3'-[^{32}P]-labelled DNA fragment containing a single copy of the SV40 enhancer was incubated with 0, 5, 10 or 20 µl of affinity purified EBP1 (Figure 2.5), subjected to DNase I digestion and the cleaved DNA molecules fractionated by denaturing

Figure 2.5. DNase I protection of the SV40 enhancer by affinity-purified EBP1

The *Bam* HI to *Pvu* II fragment from pUC1X72, 3'-[^{32}P] labIled at the *Bam* HI site was incubated with 0, 5, 10 or 20 µl of affinity purified EBP1 as described in "Materials and Methods". Competition assays were carried out in which 10 µl of EBP1 was incubated with labelled probe in the presence of 500 ng unlabelled SV1, SV1.M1 or SV1.M2 double stranded oligonucleotide. After treatment with DNase I, reaction products were electrophoresed in a 6% denaturing polyacrylamide gel. C+T and A+G specific cleavage reactions of the labelled fragment were electrophoresed in parallel as markers.



polyacrylamide gel electrophoresis. The purified protein protects the region from 230 to 250 on the SV40 enhancer from cleavage by DNase I (Figure 2.5). A similar region of protection is also seen between positions 233 and 252 on the opposite strand (Figure 1.5) with partially purified extracts, indicating that the protein we have purified is probably responsible for the binding activity observed in the nuclear extracts (Figures 1.3 and 1.5). This conclusion was strengthened by the finding that the SV1 and SV1.M2 oligonucleotides can compete for binding of the enhancer binding factor, whereas the SV1.M1 oligonucleotide is ineffective at competing (Figure 2.5).

2.4 Characterisation of EBP1 by proteolysis

To determine if EBP1 could be dissected into distinct functional domains, we subjected the purified protein to mild digestion with a variety of proteolytic enzymes. To identify a distinct DNA binding domain, EBP1 was incubated with [^{32}P]-labelled SVI double stranded oligonucleotide under conditions similar to those employed in the gel electrophoresis DNA binding assay. Once equilibrium had been reached, protease was added to the binding reactions and digestion allowed to proceed for 15 minutes at 20°C. Products of the reaction were fractionated on native polyacrylamide gels and the labelled species visualised by autoradiography. Addition of protease XIV from *Streptomyces griseus* had a dramatic effect on the formation of the EBP1–specific DNA–protein complex. A large increase in the amount of labelled SV1 double stranded oligonucleotide present in the DNA–protein complex was observed after addition of as little as 5 ng of protease XIV (each reaction contains 1 µg of BSA as carrier) (Figure 2.6). Incubation with larger amounts of protease XIV resulted in the appearance of a DNA–protein complex with increased electrophoretic mobility (Figure 2.6). This effect was also observed, but to a lesser extent, with proteinase K. Treatment of the DNA–protein complex with elastase had only a marginal effect on the amount of complex formed (Figure 2.6). The small increase in the electrophoretic mobility of the DNA–protein complex after treatment with protease XIV suggests that the DNA binding domain of the protein constitutes the bulk of the EBP1 polypeptide, but that the activity of this domain can apparently be "activated" by mild proteolytic digestion. It should be noted that material trapped at the top of the gel is removed by treatment with all three proteases. Although EBP1 is purified in the presence of NP40, the material appears to result from the formation of aggregates of EBP1, since

Figure 2.6 Proteolytic cleavage of EBP1

Reactions contained 0.5 ng [^{32}P]-labelled, double stranded SV1 oligonucleotide, 1.0 μg of unlabelled poly [d(A-T)] poly [d(G-C)] and 2 μl of affinity purified EBP1; (–) indicates that EBP1 was omitted from these reactions. In addition, reactions contained 0 (lanes 0), 5 (lanes 1), 10 (lanes 2), 20 (Elastase and Type XIV lanes 3) or 5000 μg (Prot.K lane 3) of proteinase K, elastase or protease XIV. Proteolytic enzymes were added after the binding reactions had reached equilibrium and digestion was allowed to proceed for 15 minutes at 20°C. Reaction products were fractionated on native 6% polyacrylamide gels.

they are removed by inclusion of the detergent n-octylglucopyranoside in the column buffer during the affinity chromatography steps (Treisman, 1987).

Chapter 3. EBP1 binding to a variety of viral and cellular enhancers

3.1 Competition analysis of EBP1 binding to viral and cellular enhancers

DNA sequences similar to the recognition site of EBP1 in the SV40 enhancer are present in a variety of other viral and cellular transcriptional control regions, including the human immunodeficiency virus (HIV) enhancer (Nabel and Baltimore, 1987), human ß-interferon gene regulatory element (IRE) (Goodbourn *et al.*, 1985), mouse immunoglobulin kappa chain gene enhancer (Sen and Baltimore, 1986) and mouse class 1 H2 major histocompatibility gene enhancer (Baldwin and Sharp, 1987). An additional sequence is also present in a region of the SV40 genome located on the "late" side of the enhancer. To determine if these DNA sequences also represent binding sites for EBP1, double stranded synthetic oligonucleotides containing these sites were assessed for their ability to compete with a labelled probe for binding of EBP1. EBP1 was incubated with a [^{32}P]-labelled restriction enzyme fragment derived from the SV40 enhancer and binding monitored by DNase I footprinting. In the presence of EBP1 the region between 232 and 251 on the SV40 enhancer is protected from digestion by DNase I (Figure 3.1B).

Binding to the labelled probe is eliminated by inclusion of an excess of double stranded oligonucleotides containing the EBP1 binding site in the SV40 enhancer (SV1) but not by inclusion of double stranded oligonucleotides containing mutations within the EBP1 site in which guanines at positions 244 and 245 have been changed to cytosines (SV1.M1), or cytosines at positions 237 and 238 are changed to guanines (SV1.M3), thus identifying additional sequences lying outside the "core" region required to constitute a functional binding site for EBP1. This extends the EBP1 binding site on the SV40 enhancer past the GT-I motif and into the TC-II motif (Figure 1.2) differentiating this factor from a number of previously identified proteins that bind to the SV40 enhancer (Barrett *et al.*, 1987; Johnson *et al.*, 1987; Xiao *et al.*, 1987a).

Inclusion of an excess of double stranded oligonucleotides containing the sequences from the immunoglobulin kappa chain enhancer (NF-kB) or the mouse class 1 H2 major histocompatibility gene enhancer (H2TF1) competed for binding of the labelled SV40 probe (Figure 3.1B). Two

Figure 3.1 Competition analysis of EBP1 binding

A. Reactions contained 0.1 ng [^{32}P]-labelled double stranded SV1 oligonucleotide, 0.5 µg of unlabelled poly[d(A-T)]: poly[d(G-C)] and 2 µl of affinity purified EBP1 (dashes represent reactions which did not contain any EBP1). In addition reactions contained 0 (NC), 1, 5, 10, 25 or 100 ng of unlabelled SV1, SVUP, IRE, SYM, HIV, HIV-L, HIV-R, NF-kB or H2TFI synthetic oligonucleotides. Binding reactions were fractionated on 6% polyacrylamide gels and products visualised by autoradiography.

B. A *Bam* HI to *Pvu* II fragment from pUC1X72, was 5'-[^{32}P] labelled at the *Bam* HI site, incubated with 0.5 µg of unlabelled poly[d(A-T)]:poly[d(G-C)], 20 µl of affinity purified EBP1 (+) and digested with DNAse I as described in "Materials and Methods". Also included in the reactions were 100 ng of unlabelled double stranded SV1 (1), SV1.M1 (2), SV1.M3 (3), NF-kB (4), H2TF1 (5), HIV (6), HIV-L (7) and HIV-R (8) oligonucleotides. DNA was isolated, fractionated by electrophoresis in 8% denaturing polyacrylamide gels and the cleavage products visualised by autoradiography. G+A (Pu) specific cleavage reactions of the labelled fragment were electrophoresed in parallel as markers.

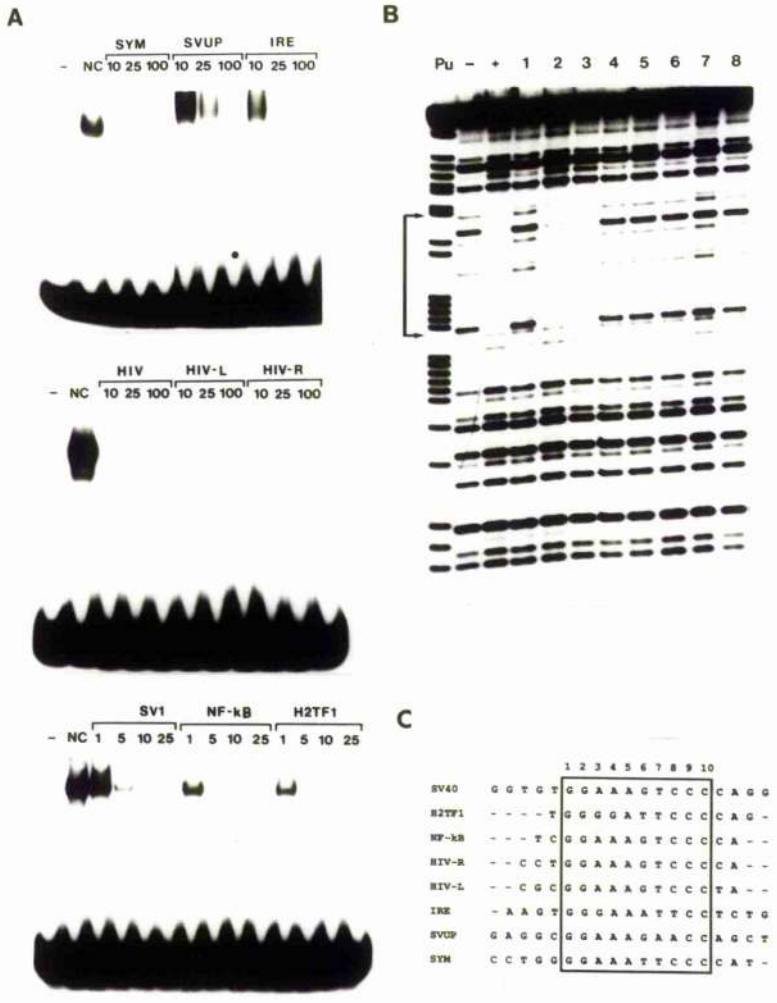
C. DNA sequences which compete for binding of EBP1. Sequences thought to be important for EBP1 recognition are boxed.

Figure 3.1 Competition analysis of EBP1 binding

A. Reactions contained 0.1 ng [^{32}P]-labelled double stranded SV1 oligonucleotide, 0.5 µg of unlabelled poly[d(A-T)]: poly[d(G-C)] and 2 µl of affinity purified EBP1 (dashes represent reactions which did not contain any EBP1). In addition reactions contained 0 (NC), 1, 5, 10, 25 or 100 ng of unlabelled SV1, SVUP, IRE, SYM, HIV, HIV-L, HIV-R, NF-kB or H2TFI synthetic oligonucleotides. Binding reactions were fractionated on 6% polyacrylamide gels and products visualised by autoradiography.

B. A *Bam* HI to *Pvu* II fragment from pUC1X72, was 5'-[^{32}P] labelled at the *Bam* HI site, incubated w 0.5 µg of unlabelled poly[d(A-T)]:poly[d(G-C)], 20 µl of affinity purified EBP1 (+) and digested with DNAse I as described in "Materials and Methods". Also included in the reactions were 100 ng of unlabelled double stranded SV1 (1), SV1.M1 (2), SV1.M3 (3), NF-kB (4), H2TF1 (5), HIV (6), HIV-L (and HIV-R (8) oligonucleotides. DNA was isolated, fractionated by electrophoresis in 8% denaturin polyacrylamide gels and the cleavage products visualised by autoradiography. G+A (Pu) specific cleavage reactions of the labelled fragment were electrophoresed in parallel as markers.

C. DNA sequences which compete for binding of EBP1. Sequences thought to be important for EBP1 recognition are boxed.


A
SYM
SVUP
IRE
- NC 10 25 100 10 25 100 10 25 100
HIV
HIV-L
HIV-R
- NC 10 25 100 10 25 100 10 25 100
SV1
NF-kB
H2TF1
- NC 1 5 10 25 1 5 10 25 1 5 10 25
B
Pu - + 1 2 3 4 5 6 7 8
C
1 2 3 4 5 6 7 8 9 10
SV40 G G T G T G G A A A G T C C C C A G G
H2TF1 - - - - T G G G G A T T C C C C A G -
NF-kB - - - T C G G A A A G T C C C C A - -
HIV-R - - C C T G G A A A G T C C C C A - -
HIV-L - - C G C G G A A A G T C C C C T A - -
IRE - A A G T G G G A A A T T C C T C T G
SVUP G A G G C G G A A A G A A C C A G C T
SYM C C T G G G G A A A T T C C C C A T -

tandem copies of potential EBP1 binding sites are present in the HIV enhancer and it is apparent that each sequence separately (HIV-R, HIV-L) or together (HIV) can compete for binding (Figure 3.1B). In addition, double stranded oligonucleotides containing the EBP1 binding site present in the IRE were also found to compete for binding of the labelled SV40 probe in DNase I footprint analysis (data not shown).

The relative affinity of the different sequences for EBP1 was determined by the addition of increasing amounts of each double stranded oligonucleotide to a binding reaction containing [^{32}P]-labelled SV1 oligonucleotide and purified EBP1 (Figure 3.1A). It is apparent that the EBP1 binding sequence present in the immunoglobulin kappa chain enhancer, class 1 H2 genes, the HIV enhancer and a synthetic symmetrical sequence , Sym (Table 1), all compete for binding of EBP1 with similar affinities (Figure 3.1A). The IRE sequence also competes for binding of EBP1, but a tenfold higher concentration of this binding site is required to give a similar level of competition to that obtained with SV1. Competition by the additional SV40 sequence on the late side of the enhancer (SVUP) only takes place at very high concentrations of oligonucleotide indicating that this sequence is a low affinity site for EBP1. The EBP1 recognition sites present in the aforementioned transcriptional control regions are shown in Figure 3.1C and comparison of the different EBP1 binding sites indicates that each contains a conserved 10 bp sequence (as depicted by the boxed region).

3.2 Binding of EBP1 to the human ß-Interferon response element (IRE)

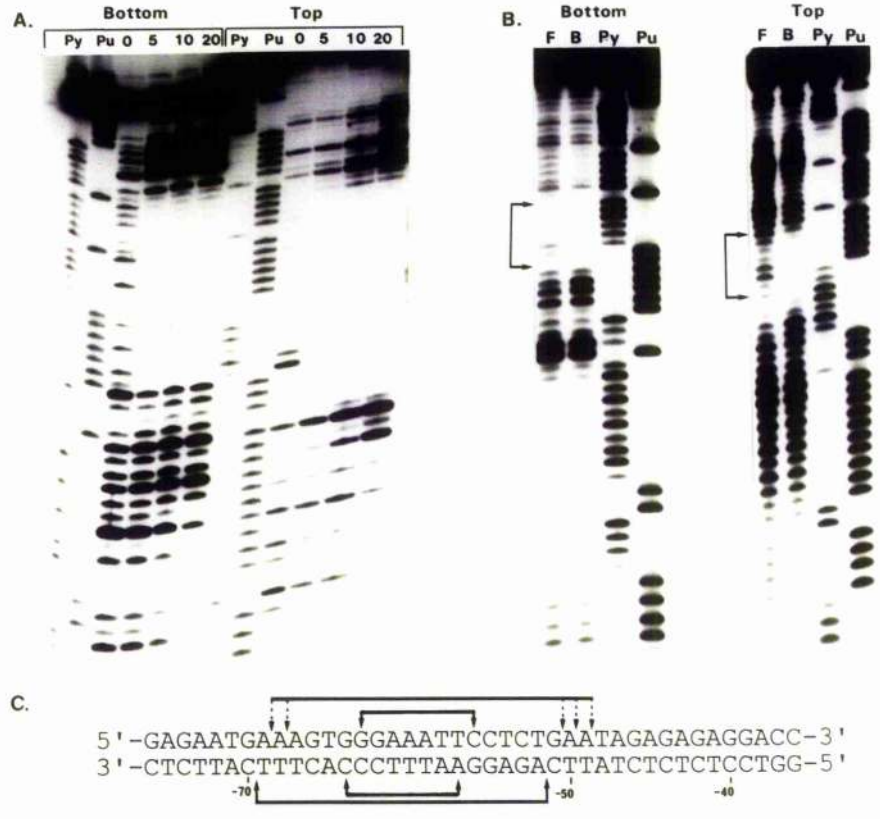
The *cis*- acting DNA sequences required for induction of the human ß-interferon gene have been identified as a 43 bp region located at position -77 to -35 with respect to the transcriptional start site. Within this region is a DNA sequence that is similar to the sequence present in the "core" region of the SV40 enhancer. In the previous section (FIgure 3.1A) it was been demonstrated,by competition analysis, that the IRE bound EBP1. The protein, purified by recognition site affinity chromatography, was therefore incubated with a restriction enzyme fragment [^{32}P]-labelled on either the top or bottom strand of the IRE, digested with DNase I and the cleavage products displayed on a denaturing polyacrylamide gel. On both the top and bottom strands a region of the IRE is protected by EBP1 from DNase I digestion (Figure 3.2A). On the bottom strand the region of

Figure 3.2 DNase I protection and OP/Cu+ cleavage of the IRE in the presence of EBP1

A. A *Hind* III to *Bgl* II fragment from pSPIRE, was 3'-[^{32}P]-labelled (top) or 5'-[^{32}P] labelled (bottom) at the *Hind* III site, incubated with 0, 5, 10 or 20 µl of affinity purified EBP1 and digested with DNase I as described in "Materials and Methods". DNA was isolated, fractionated by electrophoresis in 8% denaturing polyacrylamide gels and the cleavage products visualised by autoradiography. C+T (Py) and G+A (Pu) specific cleavage reactions of the labelled fragment were electrophoresed in parallel as markers.

B. The aforementioned labelled fragments were incubated with 20 µl of affinity purified EBP1 and, after binding reactions had reached equilibrium, free DNA (F) was separated from EBP1–DNA complexes (B) by electrophoresis in a native polyacrylamide gel. Cleavage with OP/Cu+ was carried out within the acrylamide matrix and DNA eluted as described in "Materials and Methods". DNA was fractionated, and products visualised as described in **A**. Deoxyribose residues within the bracketed region are protected from cleavage by OP/Cu+.

C. DNA sequence of the IRE containing the EBP1 binding site. Phosphate bonds within the large brackets are protected from DNase I cleavage broken lines depict potential protected bonds). Deoxyribose residues within the small bracketed regions are protected from OP/Cu+ cleavage in the presence of EBP1.



protection extends from the phosphate bond 3' of base -51 to the phosphate bond 3' of base -69. The boundary of the DNase I footprint on the top strand is not so well defined as DNase I cuts poorly in these regions. DNase I cleavage patterns are normally staggered by 2 bp which suggests that the protected region extends from the phosphate bond 3' of position -50 to the bond 3' of position -68.

To determine which deoxyribose residues in the DNA backbone are in the proximity of bound EBP1, we have subjected the EBP1–DNA complex to cleavage by hydroxyl radicals generated by the 1,10 orthophenanthroline/copper ion (OP/Cu+) complex. Restriction enzyme fragments [^{32}P]-labelled on either the top or bottom strand of the IRE were incubated with purified EBP1 and EBP1– DNA complexes resolved from free DNA by electrophoresis in a non-denaturing polyacrylamide gel. OP/Cu+ cleavage was conducted *in situ* (Kuwabara and Sigman, 1987), DNA extracted from the gel and the cleaved products fractionated by denaturing polyacrylamide gel electrophoresis. On each strand a total of eight deoxyribose moieties on the DNA backbone are protected from OP/Cu+ cleavage by EBP1(Figure 3.2B). The region of protection extends from the deoxyribose residue attached to base -57 to base -64 on the bottom strand, and from -56 to -63 on the top strand. A summary of the DNase I and OP-Cu+ protection data is presented in Figure 3.2C.

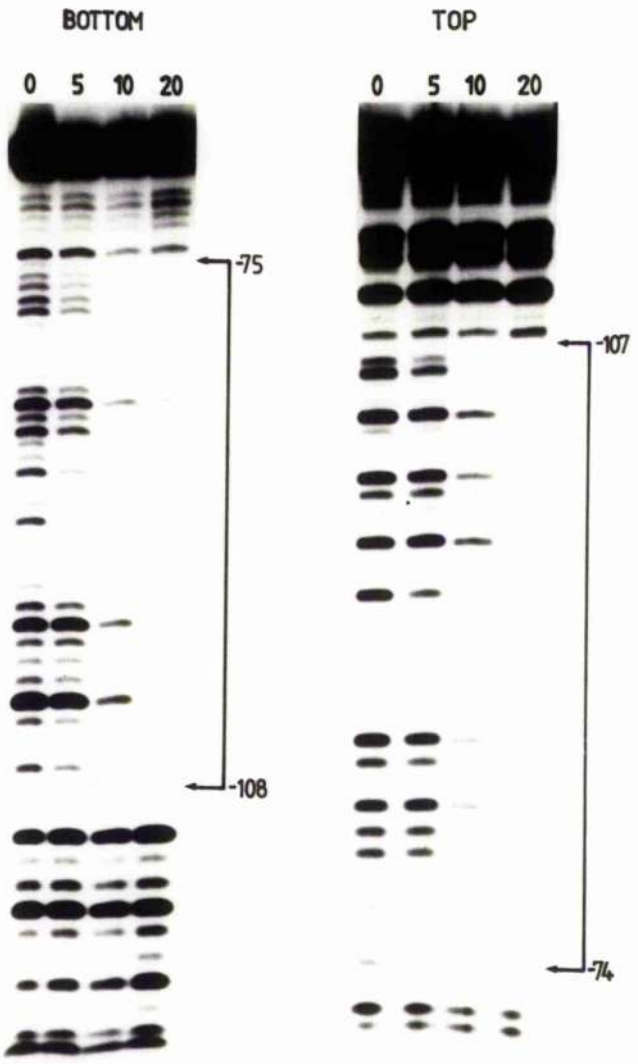
3.3 Binding of EBP1 to the human immunodeficiency virus (HIV) enhancer

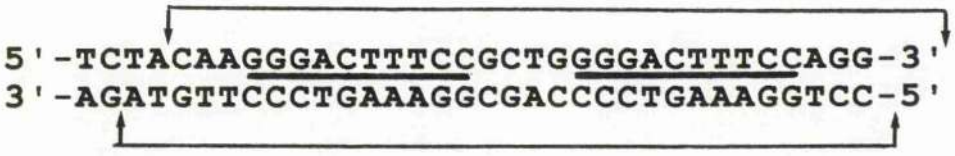
The human immunodeficiency virus (HIV) can express its gene products when transfected into a number of cell lines, including HeLa cells (Levy *et al.,* 1985), suggesting that the virus is able to use general cellular transcription factors for expression of its genes. Genetic analysis of the HIV long terminal repeat (LTR) has revealed several *cis* -acting regions important for viral gene expression including a negative regulatory region, an enhancer element, SP1 (Dynan and Tjian, 1983; Gidoni *et al.,* 1985; Briggs *et al.,* 1986) binding sites, TATA and untranslated regions (reviewed by Wu, *et al.,* 1988). Within the enhancer region are two direct repeats of a DNA sequence, -GGGACTTTCC-, similar to that present in the "core" region of the SV40 enhancer (Nabel and Baltimore, 1987). Competition analysis (Figure 3.1A and B) has demonstrated that EBP1 binds to both of these sequences present in the HIV enhancer. To determine how EBP1 bound to the HIV enhancer , the protein was incubated with a restriction enzyme fragment, [^{32}-P]-labelled on either the top or bottom strand of the HIV enhancer, digested with DNase I and

Figure 3.3 DNase I protection of the HIV enhancer in the presence of EBP1

A *Hind* III to *Sac* I fragment, or *Eco* RI to *Pst* I fragment, from pHIVEn was 3'-end labelled at the *Hind* III (Bottom; **A**) or *Eco* RI (Top; **B**) site, incubated in the absence (0) or presence of 5, 10 or 20 µl of affinity purified EBP1 and digested with DNase I as described in "Materials and Methods". DNA was isolated and fractionated by electrophoresis in 8 % denaturing polyacrylamide gels and the cleavage products visualised by autoradiography.

C. DNA sequence of the HIV enhancer containing the EBP1 binding sites (as depicted by the horizontal bars). Phosphate bonds within the bracketed regions are protected from DNase I cleavage in the presence of EBP1.





the cleavage products fractionated on a denaturing polyacrylamide gel. On both the top and bottom strands a region of 34 bp of the HIV enhancer (Figure 3.3A and B) is protected from DNase I digestion by EBP1. This region contains both the -GGGACTTTCC- direct repeats. On the bottom strand the region of protection extends from the phosphate bond 3' of base -75 to the phosphate bond 3' of base -108, and from the phosphate bond 3' of position -74 to the bond 3' of position -107 on the top strand. A summary of the DNase I data is presented in Figure 3.3C with the positions of the -GGGACTTTCC- repeats indicated by the horizontal bars.

Chapter 4. Mutational analysis of the EBP1 binding site

4.1 DNA binding specificity of purified EBP1

To determine the bases involved in sequence-specific recognition by EBP1 on the DNA, a detailed mutational analysis of sites present in the SV40 enhancer and IRE was carried out. Results of DNase I protection competition experiments revealed that guanines at positions 244 and 245, and cytosines at positions 236 and 237 (Figure 3.1B, lanes 2 and 3) were an important component in the recognition event. This was also found to be the case by analysing the interaction of the purified protein with [^{32}P]-labelled SV1 double stranded oligonucleotide in the standard gel electrophoresis DNA-binding assay (Figure 4.1). Whereas unlabelled SV1 oligonucleotide could efficiently compete for binding of the purified factor , the SV1.M1 oligonucleotide (with G to C changes at positions 244 and 245) was unable to do so. This was also true of the SV1.M3 oligonucleotide which contained C to G alterations at position 237 and 238 (Figure 4.1). Thus the purified protein requires interactions within the GT-I motif, at positions 244 and 245, and in the TC-II motif at positions 237 and 238 for sequence-specific recognition, as determined by DNase I footprinting (Figure 3.1B) and gel electrophoresis DNA binding assays (Figure 4.1)

4.2 Mutational analysis of the EBP1 binding site in the IRE

Comparison of the naturally occurring binding sites for EBP1 (Figure 3.1C) suggest which base pairs within the recognition site are important for EBP1 binding. To evaluate the contribution of individual base pairs to EBP1 binding we have made use of a series of point mutants (Figure 4.2B) which span its binding site in the IRE (Goodbourn and Maniatis, 1988). [^{32}P]-labelled restriction

Figure 4.1 DNA binding specificity of affinity purified EBP

Competition analysis of EBP1 binding. Reactions contained 0.5 ng [^{32}P]-labelled double stranded SV1 oligonucleotide, 1.0 μg of unlabeled poly [d(A-T)]: poly [d(C-G)] and 1.0 μl of affinity purified EBP1 (lane (-) contained no EBP1). In addition, reactions contained 25 (lanes 1), 50 (lanes 2) or 100 ng (lanes 3) of unlabelled double stranded SV1, SV1.M1 or SV1.M3 oligonucleotides. Reaction products were analysed on 6% polyacrylamide gels. The nucleotide sequence of the SV1 oligonucleotide is shown – the corresponding changes incorporated into SV1.M1 and SV1.M3 are indicated.

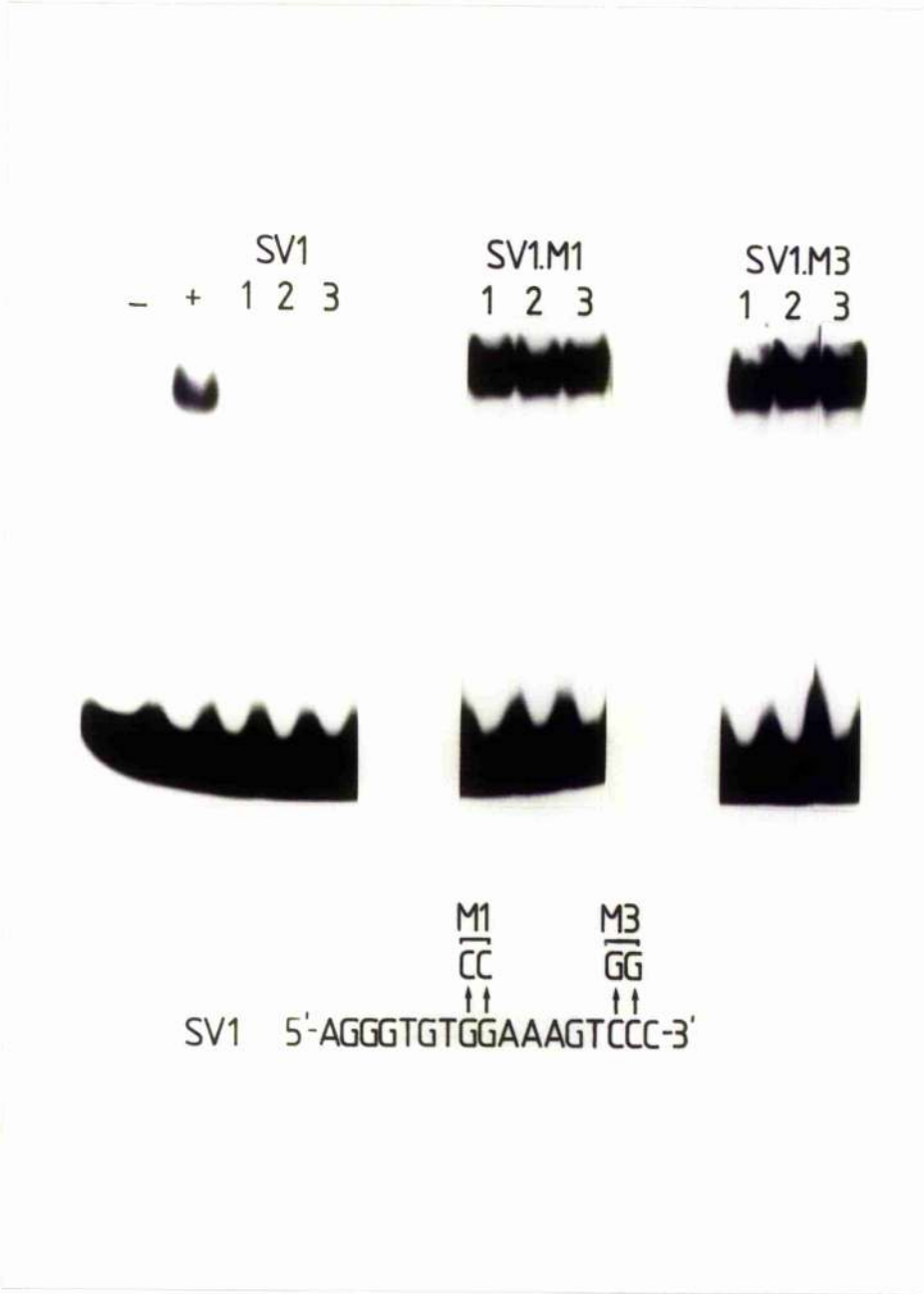
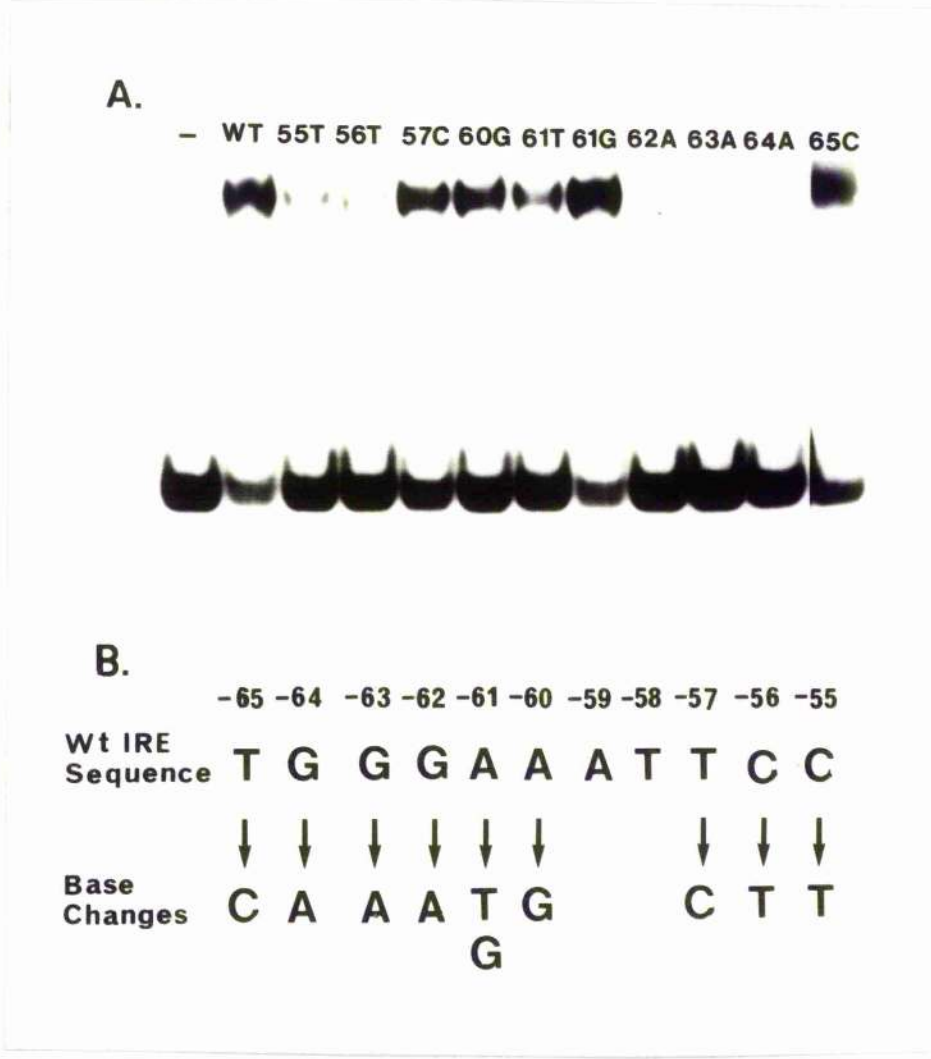


Figure 4.2 Mutational analysis of the EBP1 binding site present in the human ß-interferon gene regulatory element (IRE)

A. *Bgl* II to *Nco* I fragments were isolated from the pIF10 series of plasmids, which contain single point mutations throughout the IRE, and [^{32}P]-labelled at the *Bgl* II site. Labelled fragments (0.5 ng) were incubated with 0.5 µg of unlabelled poly[d(A-T)]: poly [d(G-C)] and 2 µl of affinity purified EBP1 in the standard gel electrophoresis DNA binding assay. Binding reactions were fractionated on 6% polyacrylamide gels and products visualised by autoradiography. The individual nucleotide changes incorporated into the mutant templates are indicated above the corresponding lanes. No EBP1 was present in the reaction represented by the dash.

B. Nucleotide changes incorporated in the pIF10 mutants used in this study. The top line represents the wild type (wt) IRE sequence together with the relevant nucleotide numbers. Arrows represent the bases which have been altered with the bottom line depicting the corresponding changes in these altered bases.



enzyme fragments containing the mutated bases were incubated with EBP1 and free DNA separated from DNA-protein complexes on a native polyacrylamide gel. It is apparent from the autoradiograph shown in Figure 4.2A that mutations within the EBP1 binding site have widely differing effects on EBP1 binding. Mutations -65C, -60G and -57C have a negligable effect on EBP1 binding, unlike mutations -64A, -63A, -62A, -56T and -55T which have a severe deleterious effect on EBP1 binding. Changes at position -61 are of interest in that an A to T change reduces binding whereas an A to G change increases binding above the wild type level. Another interesting observation occurs when the guanine at position -62 is changed to an adenine. In this situation EBP1 does not bind to the mutant IRE template. This is somewhat unexpected in that an adenine is permitted at this nucleotide position in the EBP1 binding site on the SV40 enhancer. In addition, this base lies in the sequence between the conserved bases at positions 1 and 2, and 9 and 10 (Figure 3.1C). These central base pairs were originally thought to be mutable in that changes could be incorporated without having any severe effect on EBP1 binding. Table 4.1 lists the DNA sequences which bind EBP1 (A) and those which abolish EBP1 binding (B).

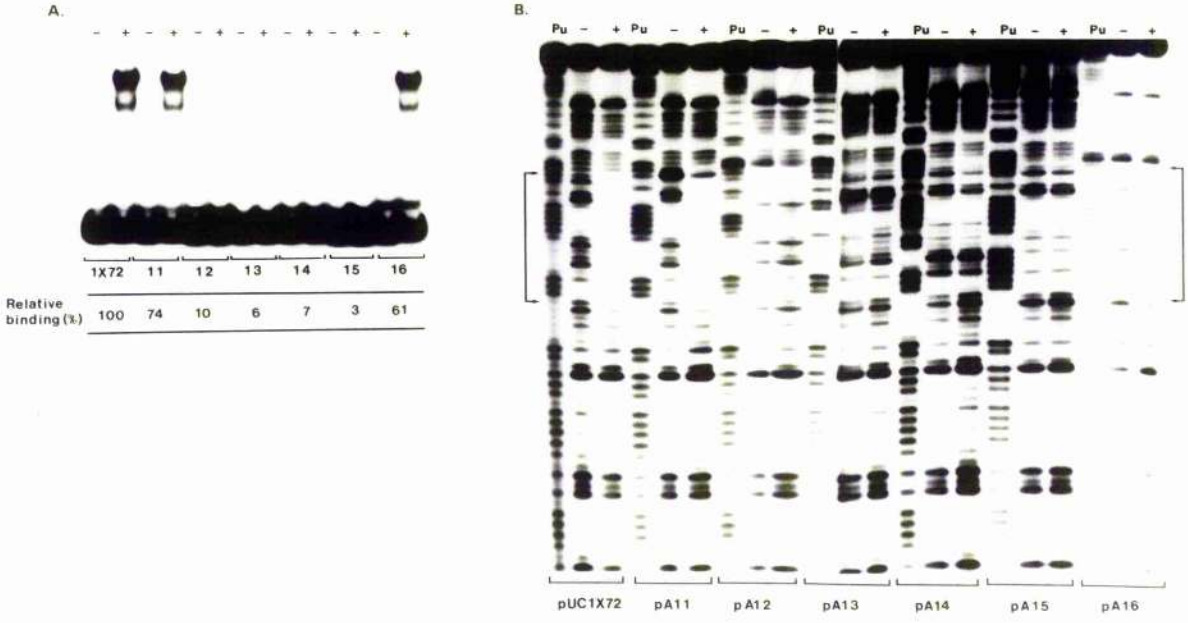
4.3 Mutational analysis of the EBP1 binding site in the SV40 enhancer

Although the critical bases for EBP1 binding in the SV40 enhancer were identified by analysis of double base pair mutations, it was not clear to what extent the base pairs between these critical residues contribute to EBP1 binding since bases between these two conserved blocks of guanine and cytosine residues (Figure 3.1C) appear to be mutable in the IRE (Figure 4.2). To address this question we have made use of the pA.. series (Figure 2M) of clustered triple point mutations in the SV40 enhancer (Zenke *et al.*, 1986) which have been used extensively to study enhancer function. These plasmids contain the SV40 enhancer region in which three neighbouring nucleotides have been mutated simultaneously *in vitro* by generating non-complementary transversions (A $\rightleftharpoons$ C, G $\rightleftharpoons$ T). The ability of EBP1 to bind to a set of pA mutants that span the binding site was determined by the gel electrophoresis DNA binding assay and by DNase footprinting. [^{32}P]-labelled restriction enzyme fragments containing the mutations were incubated with EBP1 and free DNA separated from DNA–protein complexes by electrophoresis in a native polyacrylamide gel. The extent of binding was quantitated by determining the [^{32}P]-radioactivity in the bound

Figure 4.3 Mutational analysis of the EBP1 binding site on the SV40 enhancer.

A. *Bam* HI to *Pvu* II fragments from pUC1X72 and pA11 to pA16, were 3'-[^{32}P] labelled at the *Bam* HI site, and incubated with 2 µl of affinity purified EBP1, in the presence of 0.5 µg of unlabelled poly[d(A-T)]: poly[d(G-C)], in the standard gel electrophoresis DNA binding assay. Binding reactions were fractionated in a 6% polyacrylamide gel and products visualised by autoradiography. The relative binding of EBP1 to the various templates (as determined by Cerenkov radiation counting of bound species) is indicated.

B. DNase I protection of mutated SV40 enhancer sequences by EBP1. The [^{32}P]-labelled fragments described in **A** were incubated with 0 (–) or 20 µl (+) of affinity purified EBP1 and treated with DNase I as described in "Materials and Methods". DNA was isolated and fractionated in 8% denaturing polyacrylamide gels and the cleavage products visualised by autoradiography. G+A (Pu) specific cleavage reactions of the labelled fragments were electrophoresed in parallel as markers. Sources of labelled fragments are indicated at the foot of the appropriate lanes.



fraction by liquid scintillation counting. Examination of Figure 4.3A indicates that mutations pA11 and 16, which lie on the boundary of the EBP1 recognition site, have only a marginal effect on EBP1 binding whereas mutations pA12, 13, 14 and 15, which lie within the recognition site, reduce binding by at least an order of magnitude. These results were confirmed by DNase I footprinting which also indicates that EBP1 binding to pA11 and 16 was unaltered by mutations on the boundary of the recognition site (Figure 4.3B).

4.4 DNA sequence requirement for EBP1 binding

The mutational analysis of EBP1 in the SV40 enhancer and the IRE is compiled in Table 4.1, where DNA sequences which bind EBP1 (A) are compared with sequences that do not (B). An invarient feature of all the DNA sequences that constitute high affinity binding sites for EBP1 is that they possess the sequence 5'-G-G separated by six base pairs from the complementary sequence 5'-C-C. Alteration of any one of these four bases reduces EBP1 binding considerably. The absence of the two C residues in related naturally occurring sequences from the GT-II motif of the SV40 enhancer (SV2) and two regions of the adenovirus type 2 E1a enhancer (enhA, enhB) explain why these sequences fail to bind EBP1 (Table 4.1B). One T residue located between these essential sequences is also conserved but is not absolutely required for binding since the H2TF1 sequence contains a C at this position (Figure 3.1C). However, the SVUP sequence represents a low affinity binding site for EBP1 and this may be explained by the absence of the conserved T residue. Single base pair changes between the conserved C's and G's can usually be tolerated, although this does not appear to be the case with the -62A mutant or the triple point mutations. The binding of EBP1 to 34 different DNA sequences that are related to the binding site for EBP1 on the SV40 enhancer were anaylsed. Nineteen different sequences bound EBP1 (Figure 3.1; Table 4.1) and the information obtained was used to determine the DNA sequence requirement for EBP1 binding. At each position in the sequence the frequency with which each base ocurrs has been determined and is presented graphically in Figure 4.4. As mentioned above, all sequences which bind EBP1 have G's at positions 1 and 2 and C's at positions 9 and 10. What is also apparent is the strong preference for purines in positions 1 to 6, and the equally strong preference for pyrimidines in positions 7 to 10, on the top strand.

Table 4.1 EBP1 binding sites

DNA sequences of wild type and mutant templates which bind EBP1 (A) or those which do not bind EBP1 (B). Nucleotide changes, with respect to the EBP1 recognition site present in the SV40 enhancer, are underlined.

A.

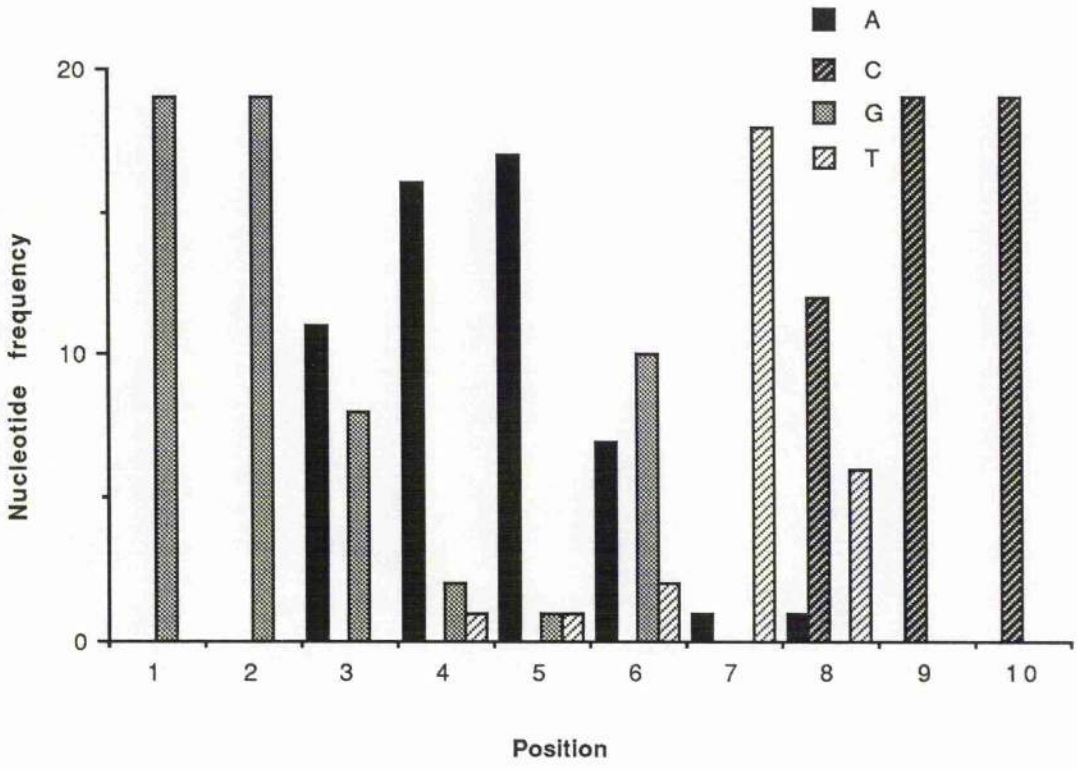
SV1	G T G G A A A G T C C C G
SYM BS	G G G G A A A T T C C C C A
SV1 M2	G T G G A A T G T C C C G
SV1 M4	C T G G A A A G T C C C G
SV pA11	T T G G A A A G T C C C C A
SV pA16	G T G G A A A G T C C C C C
IRE 61G	G T G G G G A A T T C C T C
IRE 65C	G C G G G A A A T T C C T C
IRE 60G	G T G G G A G A T T C C T C
IRE 57C	G T G G G A A A T C C C T C
IRE 54C	G T G G G A A A T T C C C C
IRE 61T	G T G G G T A A T T C C T C

B.

SV pA12	G G T T A A A G T C C C C
IRE 64A	G T A G G A A A T T C C T
IRE 63A	G T G A G A A A T T C C T
SV1 M1	G T C C A A A G T C C C G
IRE 62A	G T G G A A A A T T C C T
SV pA13	G T G G C C C G T C C C C
SV pA14	G T G G A A A T G A C C C
SV1 M3	G T G G A A A G T G G C C
IRE 56T	G T G G G A A A T T T C T
IRE 55T	G T G G G A A A T T C T T
SV pA15	G T G G A A A G T C A A A
SV2	G T G G A A T G T G T G
Enh A	G T G G T A A A A G T G
Enh B	G A G G A A G T G A A A

Figure 4.4 Frequency of base usage within EBP1 binding sites

Nineteen different EBP1 binding sites, including naturally occurring sequences and mutated derivatives, were analysed and the frequency with which each base occurs at each position determined. A = adenine, C = cytosine, G = guanine, and T = thymine nucleotides respectively.



Chapter 5. Contact point analysis of the EBP1 binding site

Although the SV40 enhancer is only ~100 bp long, it is an extremely complex transcriptional unit containing the binding sites for a large number of proteins (Wildeman *et al.*, 1986; Davidson *et al.*, 1986; for review see N.C. Jones *et al.*, 1988). Given the complexity of the situation in which a large number of proteins have the potential to interact with the same stretch of DNA, it is important to define precisely the recognition sequences for each of these proteins. To this end the points on the SV40 enhancer, IRE and HIV enhancer that are in close contact with EBP1 have been examined using nuclease protection and a variety of chemical probing techniques.

5.1 DNA protected by EBP1 from DNase I cleavage

In the presence of affinity purified EBP1, a total of nineteen phosphodiester bonds on both the top and bottom strands are protected by EBP1 from cleavage with DNase I. On the top strand (Figure 5.1A), the phosphodiester bond 3' to base 250 is resistant to DNase I cleavage in the presence of EBP1 and the protected region extends to the phosphodiester bond 3' to base 232. Partial protection of the bond 3' to base 231 is also apparent. The protected region on the bottom strand (Figure 5.1B) includes the bond 3' to base 233 and extends to the bond 3' of base 251 (BBB system of numbering employed; Tooze, 1982). However, interpretation of DNase I cleavage patterns should take into account the large size of DNase I itself. Consideration of X-ray crystallographic data on the structure of a DNase I–oligonucleotide complex (Suck *et. al.*, 1988) indicates that DNase I has an extended DNA binding surface with 6 bp on the 5'-side and 2 bp on the 3'-side of the cutting site being occupied by the enzyme. This would suggest that the area occupied by EBP1 on the SV40 enhancer is in fact significantly smaller than that demonstrated by DNase I "footprint" analysis. Thus EBP1 appears to occupy the DNA between bases 248 and 236 on the top strand, and 235 (or 238) and 247 on the bottom strand.

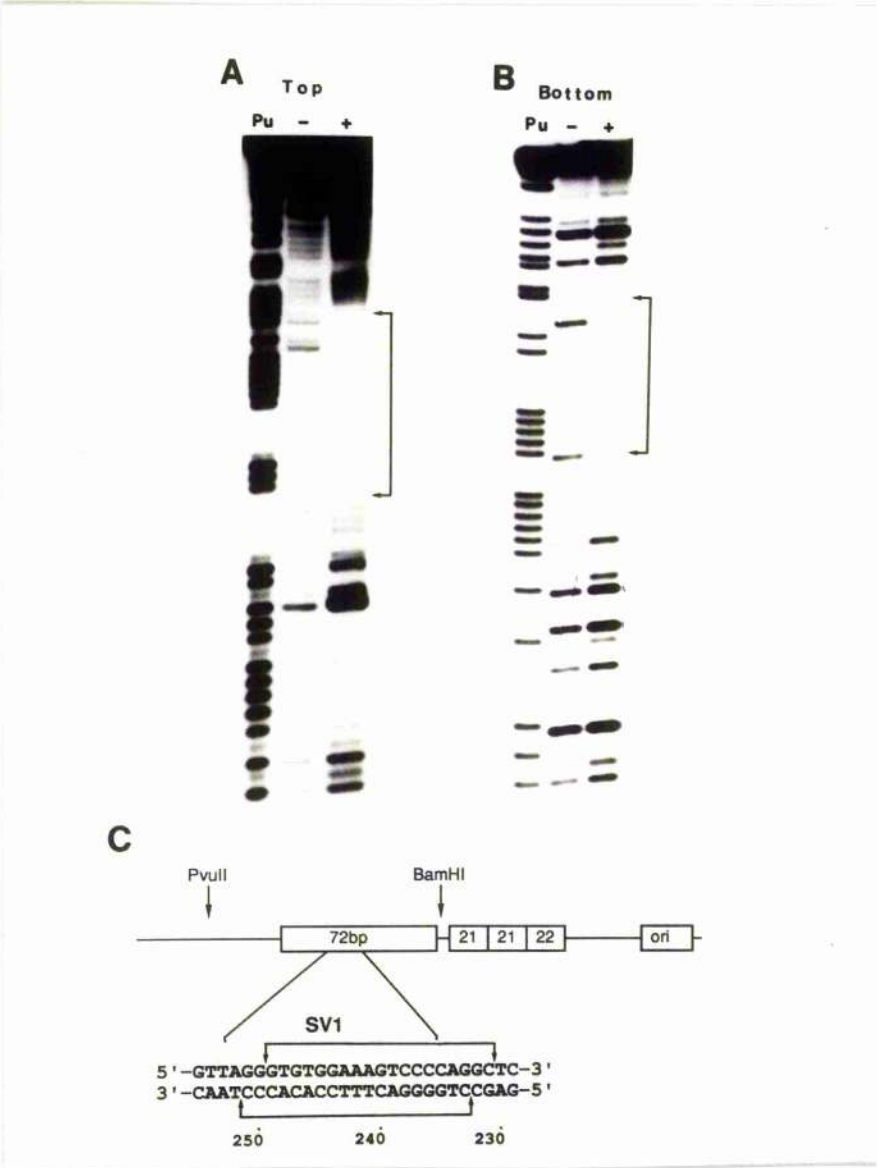
5.2 Methylation protection by EBP1 on the SV40 enhancer

The exposure of double stranded DNA to DMS results in the methylation of purines at the N-7 position of guanine and the N-3 position of adenine with the introduction of a net positive charge onto the purine ring (Singer, 1975; Ogata and Gilbert, 1978). The reaction is set up under mild,

Figure 5.1 DNase I protection of the SV40 enhancer by EBP1

A *Bam* HI to *Eco* RI fragment from pUC1X72, 3'-end labelled (**A**) or 5'-end labeled (**B**) at the *Bam* HI site, was incubated in the absence (-) or presence (+) of 10 µl affinity purified EBP1 and digested with DNase I as described in "Materials and Methods". DNA was isolated and fractionated by electrophoresis in 6% denaturing polyacrylamide gels and the cleavage products visualised by autoradiography. G+A (Pu) specific cleavage reactions of the labelled fragments were electrophoresed in parallel as markers. Phosphate bonds within the brackets are protected from DNase I cleavage.

C. Organisation of the SV40 transcriptional control region. The 21bp repeated elements of the early promoter, the origin of DNA replication (ori) and the enhancer, containing one copy of the 72 bp repeat, are indicated. Shown below is the DNA sequence of the enhancer region containing the EBP1 binding site. Phosphate bonds within the brackets are protected from DNase I cleavage. DNA sequences present in the double stranded SV1 oligonucleotide are also indicated.



non-denaturing conditions such that they do not interfere with normal protein-DNA binding reactions. The destabilising effect of modification, due to methylation, renders the DNA backbone susceptible to cleavage with alkali at the point of modification. The extent of methylation at each purine in a given DNA fragment, in the presence and absence of a DNA binding protein, can be determined by fractionation of the cleaved molecules in denaturing polyacrylamide gels followed by autoradiography. Differences in the level of methylation reflect the structure of the site of interaction on the DNA and identify the specific sites of purine-protein contact.

Binding of protein to the DNA can both decrease (protection) and increase (enhancement) the level of purine methylation in the contact region. It is thought that protection is most likely caused by steric hindrance. Less is known about the mechanism of enhancement but it may represent a close approach of the protein to DNA in such a way as to produce a hydrophobic pocket that increases the local concentration of the methylating reagent, DMS. Since the N-7 position of guanine lies in the major groove of DNA and the N-3 of adenine in the minor groove, a perturbation of the methylation reaction at guanine or adenine suggests a protein contact in the major or minor groove, respectively.

An end-labelled restriction fragment containing the cloned SV1 oligonucleotide, isolated from pSVI, was incubated with purified EBP1 and exposed to DMS. DNA was extracted and the backbone cleaved at modified G residues by treatment with piperidine. Analysis of the cleavage products by polyacrylamide gel electrophoresis (Figure 5.2A) revealed that guanines at positions 240, 244 and 245 on the top strand and positions 237 and 236 on the bottom strand are protected from modification by DMS. Although the sites present in pSVI and the complete SV40 enhancer bind EBP1 with equal affinity, we examined the region surrounding the binding site for changes in DMS sensitivity. Apparent in the fragment derived from the complete SV40 enhancer is a G residue at position 235 on the bottom strand that is hypersensitive to DMS methylation in the presence of EBP1 (Figure 5.2B). A summary of this data is presented in Figure 5.2C.

5.3 Methylated bases that interfere with EBP1 binding on the SV40 enhancer

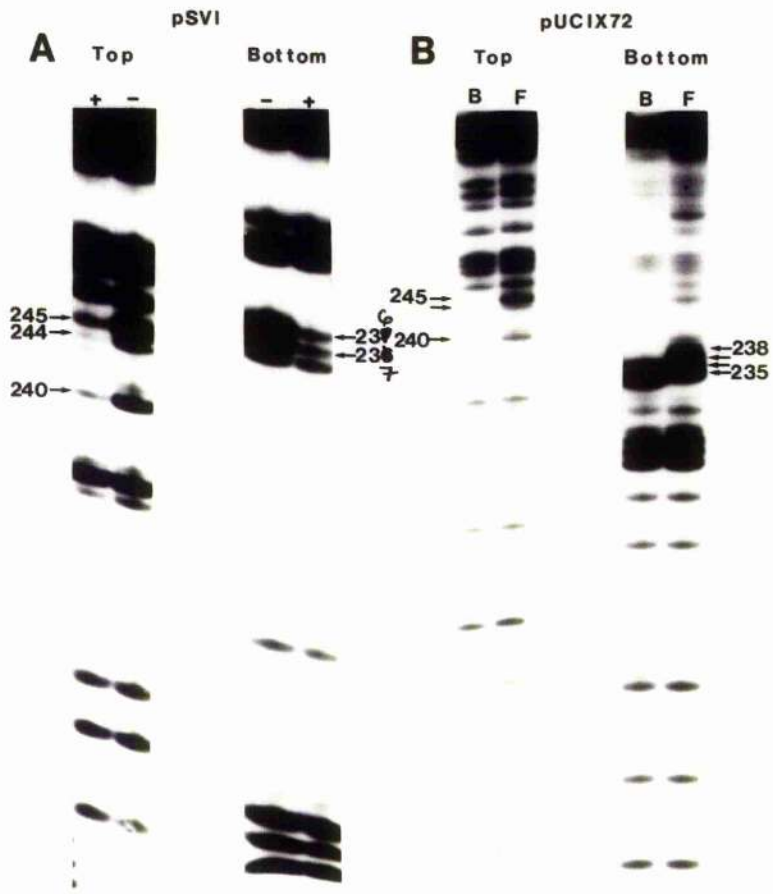
A previously methylated guanine or adenine residue in DNA can interfere with the binding of protein (Siebenlist and Gilbert, 1980). Alkylation of either base creates a positive charge on the

Figure 5.2 Methylation protection of the EBP1 binding site by EBP1

A. *Hind* III to *Kpn* I, or *Eco* RI to *Pst* I fragments from pSVI, 3'-end labelled at the *Hind* III (top) or *Eco* RI (bottom) sites, were incubated in the absence (-) or presence (+) of 10 µl affinity purified EBP1 and treated with DMS as described in "Materials and Methods". DNA was cleaved at modified G residues by treatment with piperidine, reaction products were fractionated by electrophoresis in 8% denaturing polyacrylamide gels and the cleavage products visualised by autoradiography. Positions of protected bases are indicated by arrows.

B. A *Bam* HI to *Eco* RI fragment from pUC1X72, 3'-end labelled (top) or 5'-end labelled (bottom), was incubated with 10 µl affinity pufiried EBP1, treated with DMS and free DNA (F) separated from EBP1- DNA complexes (B) in a native polyacrylamide gel. DNA was eluted from the gel, treated with piperidine and cleaved products fractionated by electrophoresis in 6% denaturing polyacrylamide gels. Positions of protected bases are indicated by arrows.

C. DNA sequence of the region of the SV40 enhancer containing the EBP1 binding site. Filled squares represent guanines protected from DMS methylation in the presence of EBP1; the open square represents a guanine hypersensitive to DMS methylation, and the hatched square a guanine protected from DMS methylation in pUC1X72, but not in pSV1.



C

5'-GTTAGGGTGTGGAAAGTCCCCAGGCTC-3'

3'-CAATCCCACACCTTTCAGGGGTCCGAG-5'

250 240 230

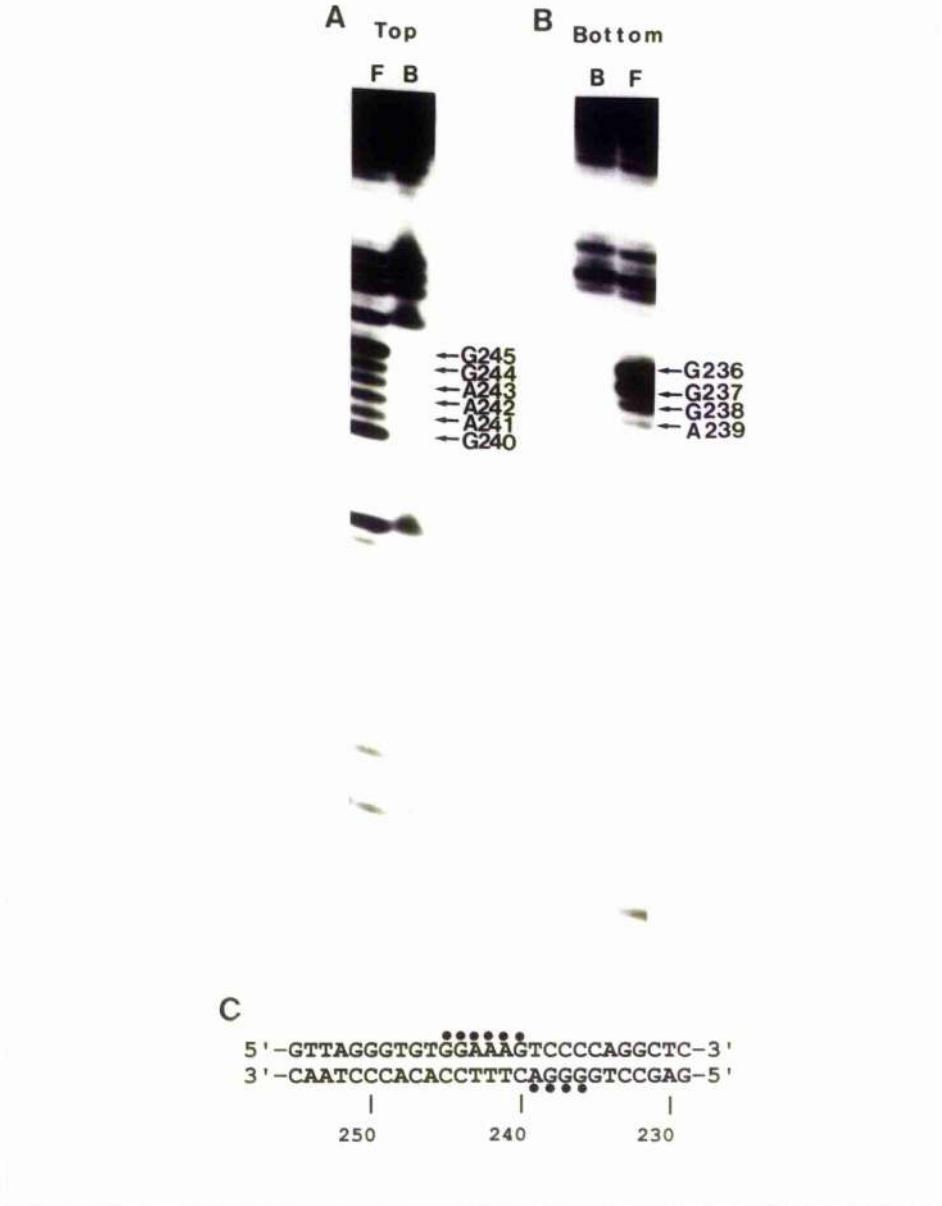
purine ring, rendering the modified residue base labile. Cleavage of the DNA at the point of modification and analysis of the products allows us to determine which N-7s of guanine and N-3s of adenine are critical for either the formation of the DNA-protein complex, or for its maintenance. Methylation interference experiments require that the DNA is modified such that each molecule in the population contains no more than one methylated base. The protein under study is then allowed to bind to the modified DNA and free DNA is separated from DNA-protein complexes on non-denaturing polyacrylamide gels. Modification at a site that makes an essential contact with the protein will interfere with binding; as a result that molecule will not be represented as DNA-protein complexes. DNA molecules with modifications that do not interfere with binding will be present as DNA-protein complexes. After cleavage of the DNA molecule at the point of modification, comparison of DNA present in the free and bound fractions by denaturing polyacrylamide gel electrophoresis identifies functional groups on the DNA that are essential for complex formation.

Partially methylated, [^{32}P]-labelled DNA representing each strand of the cloned SV1 oligonucleotide was incubated with purified EBP1 and free DNA separated from DNA-protein complexes by electrophoresis in a native polyacrylamide gel. Bound and free DNA was purified, treated with NaOH, to cleave the strands at modified purine bases, and the cleavage products resolved on a sequencing gel. On the top strand the G residues at positions 240, 244 and 245 that were protected from attack by DMS all interfere with binding (Figure 5.3A). Located between these interfering G residues, the A residues at positions 241, 242 and 243 also interfere with binding. On the bottom strand the two G residues at positions 236 and 237, which were protected from DMS modification, interfere with binding after modification Figure 5.3B). The two adjacent purine bases, G 238 and A 239, also interfere with binding of EBP1 when modified. Guanine 235, which is hypersensitive to DMS attack and is not required for binding, does not interfere with binding when modified (data not shown). Assessment of this data indicates that the arrangement of interfering purine bases is highly strand specific; a string of four purine bases on the bottom strand followed by six purine bases on the top strand all interfere with binding (Figure 5.3C). Dimethylsulphate introduces a methyl group onto the N-7 position of guanine, which projects into the major groove, and the N-3 position of adenine, which projects into the minor groove. The data presented above therefore suggest that EBPI makes contacts in both the major and minor grooves of the DNA

Figure 5.3 Methylated bases that interfere with EBP1 binding.

Hind III to *Kpn* I (**A**) or *Eco* RI to *Pst* I (**B**) fragments from pSV1, 3'-end labelled at the *Hind* III (top) or *Eco* RI (bottom) sites were treated with DMS and the methylated DNA incubated with 10 µl affinity purified EBP1. Free DNA (F) and EBP1–DNA complexes (B) were separated by electrophoresis in a native polyacrylamide gel. DNA was eluted from the gel and cleaved at modified purine bases by treatment with NaOH. Cleavage products were fractionated by electrophoresis in 8% denaturing polyacrylamide gels. Positions of interfering bases are indicated by arrows.

C. DNA sequence of the region of the SV40 enhancer containing the EBP1 binding site. Filled circles depict modified bases which interfere with EBP1 binding.



double helix.

5.4 Base specific contacts of EBP1 on the HIV enhancer

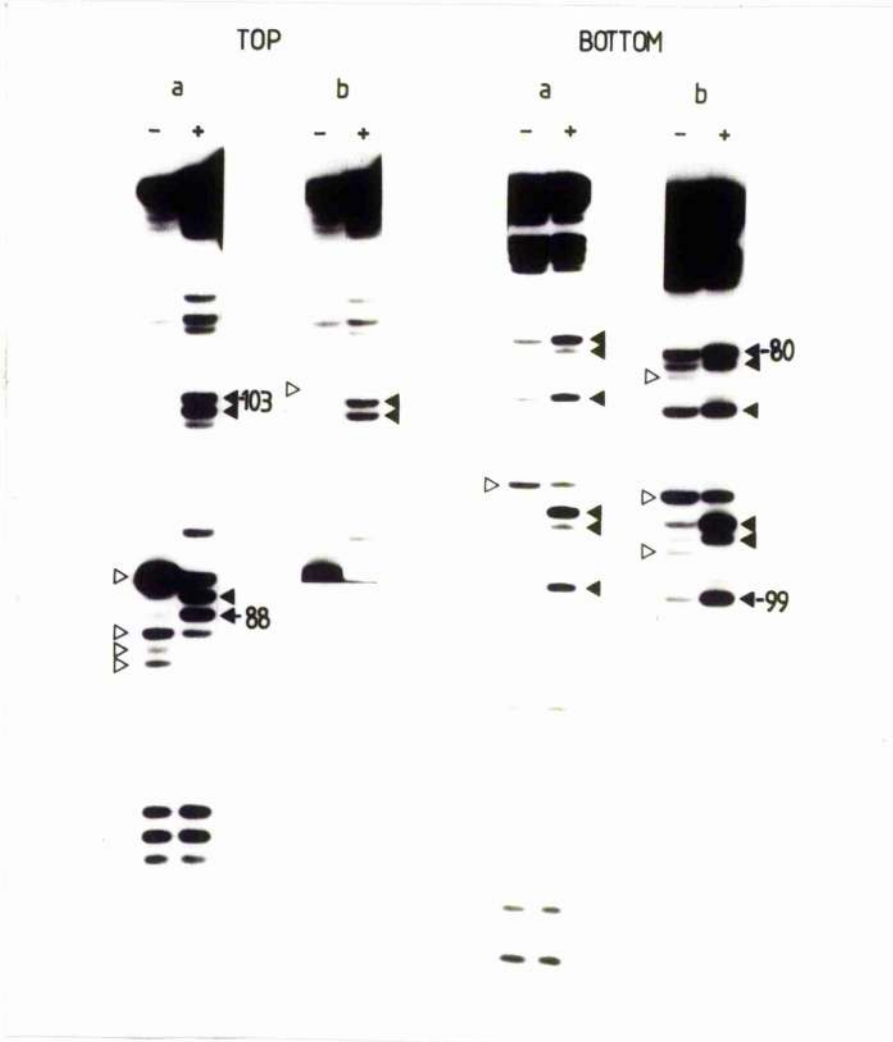
To determine which purine bases in the EBP1 recognition sites of the HIV enhancer are in close contact with EBP1, methylation protection and interference experiments were carried out. EBP1 was incubated with a restriction enzyme fragment containing the HIV enhancer, [^{32}-P]-labelled on either the top or bottom strand, and the resulting DNA-protein complexes treated with DMS. Guanine residues methylated at the N-7 position were identified by treating with piperidine to cleave the DNA at the point of DMS modification. The cleavage products were fractionated by electrophoresis in denaturing polyacylamide gels and visualised by autoradiography. X-ray film was allowed to expose for 15 hours, or 48 hours, for maximum resolution of protected/enhanced cleavages. Results indicate that guanines at positions -88, -89, -102 and -103 are protected by EBP1 from modification by DMS (Figure 5.4A). On the bottom strand, guanines at positions -80, -81, -85, -94, -95 and -99, are protected from DMS modification in the presence of EBP1 (Figure 5.4B). In addition, adenine residues at positions -86 and -104, guanine at position -87 and cytosine at position -85 on the top strand (Figure 5.4A) and adenines at positions -82 and -96 on the bottom strand (Figure 5.4B) are hypersensitive to DMS methylation in the presence of EBP1. A summary of this data is presented in Figure 5.4C.

A complementary series of experiments were carried out to determine which methylated purine bases interfered with the binding of EBP1. Specific binding of purified NF-kB to the HIV enhancer has been demonstrated by DNAse I protection and methylation interfence analyses (Kawakami *et al.*, 1988). In direct footprint analysis, NF-kB binding sites, which also correspond to EBP1 binding sites (Figure 3.1), were protected from DNase I cleavage in the presence of affinity purified NF-kB. Methylation interference studied revealed six guanine residues in the distal binding site which when methylated interfered with NF-kB binding. However, decreased signals indicative of methylation interference were not clearly observed, suggesting that interference was only partial. In addition, no methylation interference was observed in the proximal binding site in the presence of NF-kB. Kawakami and coworkers (1988) propose that the distal element can be occupied only by NF-kB, whereas the proximal element may be occupied by either a common cellular factor, for example

Figure 5.4 Methylation protection of bases in the HIV enhancer in the presence of EBP1

A *Hind* III to *Sac* I fragment, or *Eco* RI to *Pst* I fragment, from pHIVEn was 3'-end labelled at the *Eco* RI (Top; **A**) or *Hind* III (Bottom; **B**) site, was incubated in the absence (-) or presence of 20 µl of EBP1 (+) and treated with DMS as described in "Materials and Methods". DNA was cleaved at modified bases by treatment with piperidine and the reaction products fractionated by electrophoresis in 8 % denaturing polyacrylamide gels. Cleavage products were visualised by autoradiography after 15 (Top, b; Bottom, a) or 48 hours (Top. a; Bottom, b) exposure to X-ray film. Positions with altered sensitivity to DMS are indicated by filled triangles, in the case of protected bases, or open triangles, in the case of bases hypersensitive to DMS methylation.

C. DNA sequence of the HIV enhancer containing the two EBP1 binding sites. Positions of the bases with altered sensitivity to DMS are indicated as described above.



5'-TCTACAAGGGACTTTCCGCTGGGGACTTTCCAGG-3'
3'-AGATGTTCCCTGAAAGGCGACCCCTGAAAGGTCC-5'

H2TF1, or by NF-kB. In light of these findings, restriction fragments containing the proximal (right) or distal (left) EBP1 binding sites on the HIV enhancer, [^{32}-P]-labelled on either the top or bottom strands, were treated with DMS to generate, on average, less than one methylated base per labelled fragment. Partially methylated DNA fragments were incubated with EBP1 and DNA-protein complexes separated from free DNA in a non-denaturing polyacrylamide gel. DNA present in the bound and free fractions was eluted from the gel and treated with NaOH to cleave N-7 methylated guanines and N-3 methylated adenines at the point of modification. Comparing bound and free DNA on a denaturing polyacrylamide gel revealed which purine bases, when methylated, interfered with EBP1 binding. On the top strand of the distal site four consecutive purine bases, -100A, -101G, -102G and -103G, all interfere with EBP1 binding when methylated. A similar situation is evident on the bottom strand where six consecutive purines, -94G, -95G, -96A, -97A, -98A and -99G, all interfere with EBP1 binding when methylated (Figure 5.5A). An identical pattern of interfering bases is found for the proximal EBP1 binding site whereby four consecutive purine bases, -86A, -87G, -88G and -89G, on the top strand, all interfere with EBP1 binding when methylated and six consecutive purine bases on the bottom strand, at positions -80G, -82A, -83A, -84A and -85G, interfere with binding (Figure 5.5B). These data are summarised in Figure 5.5C and show that all purine bases in the 10 bp stretches from -80 to -89, and -94 to -103, interfere with EBP1 binding when methylated. The guanine nucleotides that were protected by EBP1 from methylation, all interfered with binding when methylated. The guanine residue at position -90, which demonstrated a very pronounced increase in reactivity towards DMS, in the presence of EBP1, did not interfere with EBP1 binding when methylated – this nucleotide lies between the two10 bp stretches of interfering purines.

(as confirmed by densitometric scanning of the autoradiographs)

5.5 Base specific contacts of EBP1 on the IRE

To establish which purine bases within the IRE are in close contact with EBP1, methylation protection and interference experiments were performed. EBP1 was incubated with a restriction enzyme fragment containing the IRE, [^{32}P]-labelled on either the top or bottom strand, and the resulting DNA–protein complexes exposed to DMS. Guanine residues methylated at the N-7 position were identified by treating with piperidine to cleave the DNA at the point of modification.

Figure 5.5 Methylated bases on the HIV enhancer that interfere with EBP1 binding

A.A *Hind* III to *Sac* I fragment, or *Eco* RI to *Pst* I fragment from pHIV-R, was 3'-end labelled at the *Eco* RI (Top) or *Hind* III (Bottom) site, treated with DMS and the methylated DNA incubated with 20 µl of affinity purified EBP1 as described in "Materials and Methods". Free DNA (F) and EBP1–DNA complexes (B) were separated by electrophoresis in a native polyacrylamide gel. DNA was eluted from the gel and cleaved at modified purine bases by treatment with NaOH. Cleavage products were fractionated by electrophoresis in an 8 % denaturing polyacrylamide gel. Nucleotide positions of the interfering bases are indicated by arrows. The DNA sequence of HIV-R ,containing the proximal EBP1 binding site, is shown with the positions of the interfering bases depicted by the stars.

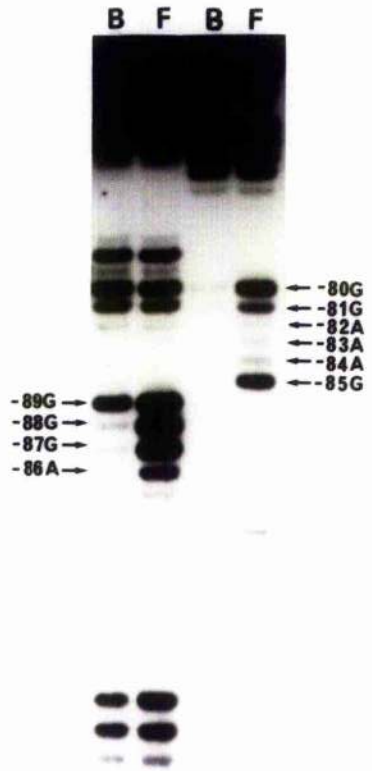
B. A *Hind* III to *Sac* I fragment, or *Eco* RI to *Pst* I fragment from pHIV-L, was 3'-end labelled at the *Eco* RI (Top) or *Hind* III (Bottom) site, treated with DMS, the methylated DNA incubated with EBP1 and processed as described in **A.** Nucleotide positions of the interfering bases are indicated by arrows. The DNA sequence of HIV-L, containing the distal EBP1 binding site, is shown with positions of interfering bases depicted by the stars.

Figure 5.5 Methylated bases on the HIV enhancer that interfere with EBP1 binding

A.A *Hind* III to *Sac* I fragment, or *Eco* RI to *Pst* I fragment from pHIV-R, was 3'-end labelled at the *Eco* RI (Top) or *Hind* III (Bottom) site, treated with DMS and the methylated DNA incubated with 20 μl of affinity purified EBP1 as described in "Materials and Methods". Free DNA (F) and EBP1–DNA complexes (B) were separated by electrophoresis in a native polyacrylamide gel. DNA was eluted from the gel and cleaved at modified purine bases by treatment with NaOH. Cleavage products were fractionated by electrophoresis in an 8 % denaturing polyacrylamide gel. Nucleotide positions of the interfering bases are indicated by arrows. The DNA sequence of HIV-R ,containing the proximal EBP1 binding site, is shown with the positions of the interfering bases depicted by the stars.

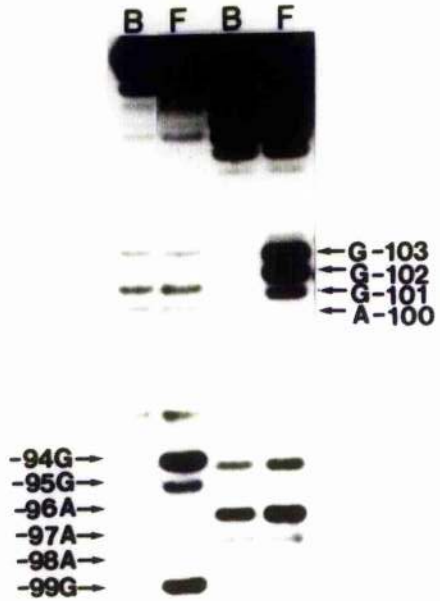
B. A *Hind* III to *Sac* I fragment, or *Eco* RI to *Pst* I fragment from pHIV-L, was 3'-end labelled at the *Eco* RI (Top) or *Hind* III (Bottom) site, treated with DMS, the methylated DNA incubated with EBP1 and processed as described in **A**. Nucleotide positions of the interfering bases are indicated by arrows. The DNA sequence of HIV-L, containing the distal EBP1 binding site, is shown with positions of interfering bases depicted by the stars.

A.



5′GATCTGGGGACTTTCCAGG 3′ HIV-R
3′ACCCCTGAAAGGTCCCTAG 5′

B.



5′GATCTAGGGACTTTCCGCG3′ HIV-L
3′ATCCCTGAAAGGCGCCTAG5′

On the top strand, G residues -63 and -64 are protected by EBP1 form modification by DMS. The G residue at -62 demonstrates the unusual property of being protected from methylation at low concentrations of EBP1 but at high concentrations of EBP1 appears to be hypermethylated (Figure 5.6A). On the bottom strand -55G and -56G are protected from methylation by EBP1 whereas the reactivity of DMS to the G residue at -53 is increased by the presence of EBP1 (Figure 5.6A). In a complementary series of experiments the ability of methylated purine bases to interfere with binding of EBP1 was examined. A restriction enzyme fragment containing the IRE, [^{32}P]-labelled on either the top or the bottom strand, was treated with DMS to generate less than one methylated base per fragment. Partially methylated DNA fragments were incubated with EBP1 and DNA–protein complexes resolved from free DNA in a native polyacrylamide gel. DNA present in the bound and free fractions was eluted from the gel and treated with NaOH to cleave N-7 methylated G residues and N-3 methylated A residues at the point of modification. Comparison of the bound and free DNA on a denaturing polyacrylamide gel reveals that on the top strand the six consecutive purine bases, -59, -60 and -61A, and -62, -63 and -64G, all interfere with EBP1 binding when methylated (Figure 5.6B). A similar situation is apparent on the bottom strand where the four consecutive purine bases -55 and -56G, and -57 and -58A, all interfere with binding when methylated (Figure 5.6B). These data are summarised in Figure 5.6C and indicate that all purine bases in the 10 base pairs from -55 to -64 interfere with EBP1 binding when methylated. As expected the G residues that were protected by EBP1 from methylation, interfered with binding when methylated, as did -62G which was hypermethylated at high concentrations of EBP1. Guanine residue -53, which demonstrated an increased reactivity towards DMS in the presence of EBP1, did not interfere with EBP1 binding when methylated.

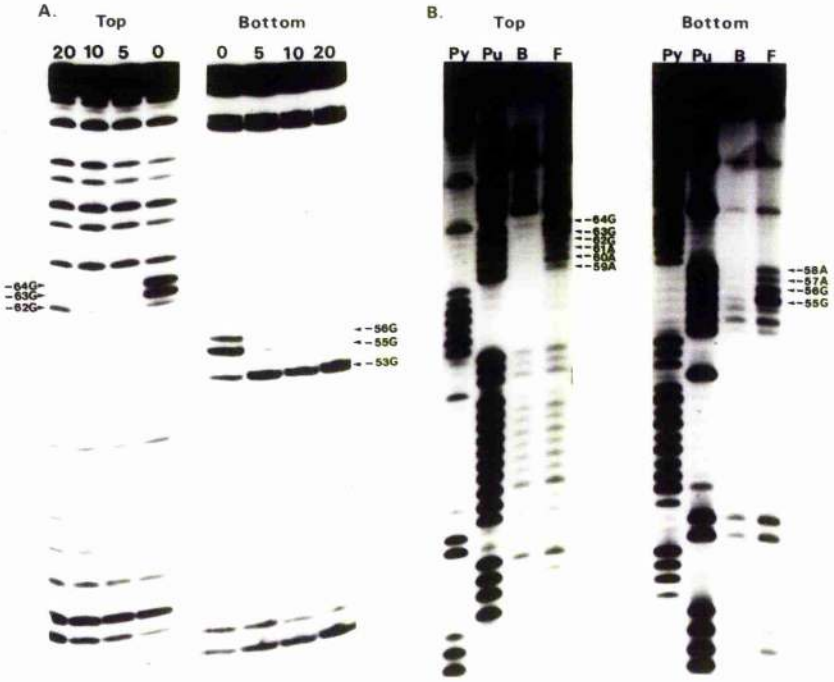
A summary of the bases in the SV40 enhancer, IRE and HIV enhancer that are in close contact with EBP1, as determined by DMS chemical probing, is presented in Figure 5.7. What is immediately apparent from the pattern of contacts is that the protein binds homologously to its recognition sites in the aforementioned enhancers. In each instance, guanine residues at positions 1 and 2 are protected from DMS methylation – these are separated by six bases from two protected guanines at positions 9 and 10 on the opposite strand. Assesement of the data indicates that the arrangement of purine bases, which when methylated interfere with EBP1 binding, is highly strand

Figure 5.6 Methylation protection of the EBP1 binding site and methylated bases that interfere with EBP1 binding on the IRE

A. A *Hind* III to *Bgl* II fragment from pSPIRE, 3'-[^{32}P] labelled (top) of 5'-[^{32}P] labelled (bottom) at the *Hind* III site, was incubated with 0, 5, 10 or 20 µl of affinity purified EBP1 and treated with DMS as described in "Materials and Methods". DNA was cleaved at modified guanine residues by treatment with piperidine and reaction products fractionated by electrophoresis in 8% denaturing polyacrylamide gels. Cleavage products were visualised by autoradiography. Positions of protected bases are indicated by arrows and the corresponding IRE nucleotide numbers.

B. The labelled fragments described in **A** were treated with DMS and the methylated DNA incubated with 20 µl of affinity purified EBP1. Free DNA (F) and EBP1–DNA complexes (B) were separated by electrophoresis in a native polyacrylamide gel. DNA was eluted from the gel and cleaved at modified purine bases by treatment with NaOH. Cleavage products were fractionated and visualised as described in **A**. C+T (Py) and G+A (Pu) specific cleavage reactions of the labelled fragments were electrophoresed in parallel as markers.

C. DNA sequence of the IRE containing the EBP1 binding site. Filled circles represent guanines protected from DMS methylation in the presence of EBP1 with open circles depicting guanines hypersensitive to DMS methylation. Open squares represent DMS- modified bases which interfere with EBP1 binding.


A.
Top
Bottom
20 10 5 0
0 5 10 20
-64G
-63G
-62G
-56G
-55G
-53G
B.
Top
Bottom
Py Pu B F
Py Pu B F
-64G
-58A
-57A
-56C
-55G
59A

C.


5'-GAGAATGAAAGTGGGAAATTCCTCTGAATAGAGAGAGGACC-3'
3'-CTCTTACTTTCACCCTTTAAGGAGACTTATCTCTCTCCTGG-5'
-70
-60
-50
-40

FIGURE 5.7 Summary of the EBP1/DNA DMS Methylation protection and interference analysis

The nucleotide sequences containing the EBP1 binding sites present in the SV40 enhancer, IRE and HIV (HIV-R and HIV-L) are shown. Filled circles represent bases which are protected from DMS methylation in the presence of EBP1 while open circles indicate bases with an increased reactivity towards DMS modification in the presence of EBP1. Purine bases which, when modified, interfere with EBP1 binding are depicted by the shaded boxes.

specific in that a string of four purine bases on one strand followed by six purine bases on the other strand all interfere with binding. Since DMS introduces a methyl group onto the N-7 position of guanine, which projects into the major groove of the DNA, and the N-3 position of Adenine, which projects into the minor groove, a straightforward interpretation of the data generated by DMS chemical probing studies, suggests that EBP1 makes sequence-specific contacts in both the major and minor grooves of the DNA double helix of the SV40 enhancer, IRE and HIV enhancer, in a similar manner.

5.6 Ethylated phosphates that interfere with EBP1 binding

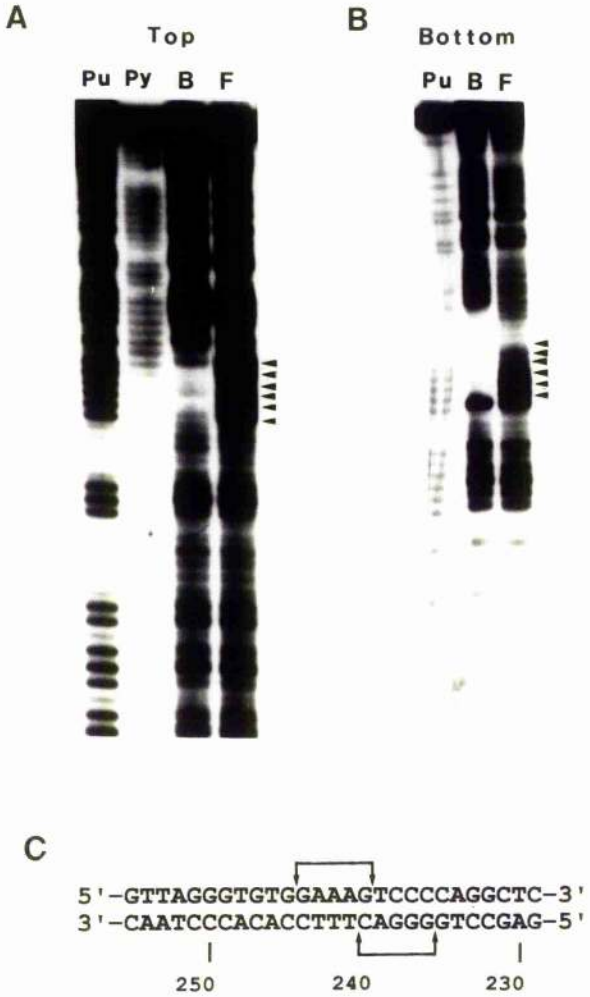
Modification of DNA with ethylnitrosourea creates a phosphotriester bond in the DNA backbone by the linkage of an ethyl group to one of the phosphate oxygens (Siebenlist and Gilbert, 1980). These phosphotriester bonds are more alkali labile that the phosphodiester bonds. An interference experiment can then be carried out using essentially the same experimental design as that described for methylation interference. Addition of the ethyl group removes the negative charge on the phosphate oxygen, thus blocking ionic interactions between protein and DNA. The bulky ethyl group may also sterically hinder the close approach of a binding protein. [^{32}P]-labelled, ethylated DNA was incubated with purified EBP1 and DNA-protein complexes were separated from free DNA by native polyacrylamide gel electrophoresis. DNA from each fraction was purified, treated with NaOH, to cleave the alkali sensitive phosphotriester bonds, and the cleavage products fractionated on denaturing polyacrylamide gels. Cleavage at a particular phosphotriester is indicated by the presence of a corresponding band on the gel. Absence of a band from the bound fraction and its presence in the free fraction indicates that ethylation of the DNA backbone at this position blocks binding.

On the top strand it is evident that modification of any one of six contiguous phosphate groups, stretching from the phosphate 3' to base 245 to the phosphate 3' to base 240, can interfere with binding (Figure 5.8A). A similar situation is also apparent on the bottom strand where six contiguous phosphates, stretching from the phosphate 5' to base 236 to the phosphate 5' to 241, consistently interfere with binding (Figure 5.8B). Ethylation of phosphates 5' to bases 242 and 243 also leads to a reduction in binding but the effect observed is less pronounced and less reproducible. The

Figure 5.8 Ethylated phosphotriesters that interfere with EBP1 binding

A *Bam* HI to *Eco* RI fragment from pUC1X72, 3'-end labelled (**A**, top) or 5'-end labelled at the *Bam* HI site(**B**, bottom), was treated with ethylnitrosourea as described in "Materials and Methods". Ethylated DNA was incubated with 10 µl affinity purified EBP1and free DNA (F) separated from EBP1–DNA complexes (B) in a native polyacrylamide gel. DNA was eluted and cleaved at phosphtriester bonds by treatment with NaOH. Reaction products were fractionated by electrophoresis in 6% denaturing polyacrylamide gels and cleavage products visualised by autoradiography. G+A (Pu) and C+T (Py) specific cleavage reactions of the labelled fragments were run in parallel as markers. Positions of modified phosphotriesters that interfere with binding are indicated by arrows.

C. DNA sequence of the region of the SV40 enhancer containing the EBP1 binding site. Modified phosphate bonds within the bracket interfere with EBP1 binding.



strand specificity of the interfering ethylated phosphate groups (summarised in Figure 5.8C) is reminiscent of that observed for methylated bases which interfere with binding. On the top strand contacted phosphates are all located on the 3' side of purine bases which, when methylated, interfere with EBP1 binding.

5.7 Orthophenanthroline/copper (OP/Cu+) chemical nuclease "footprinting" of the EBP1 binding site

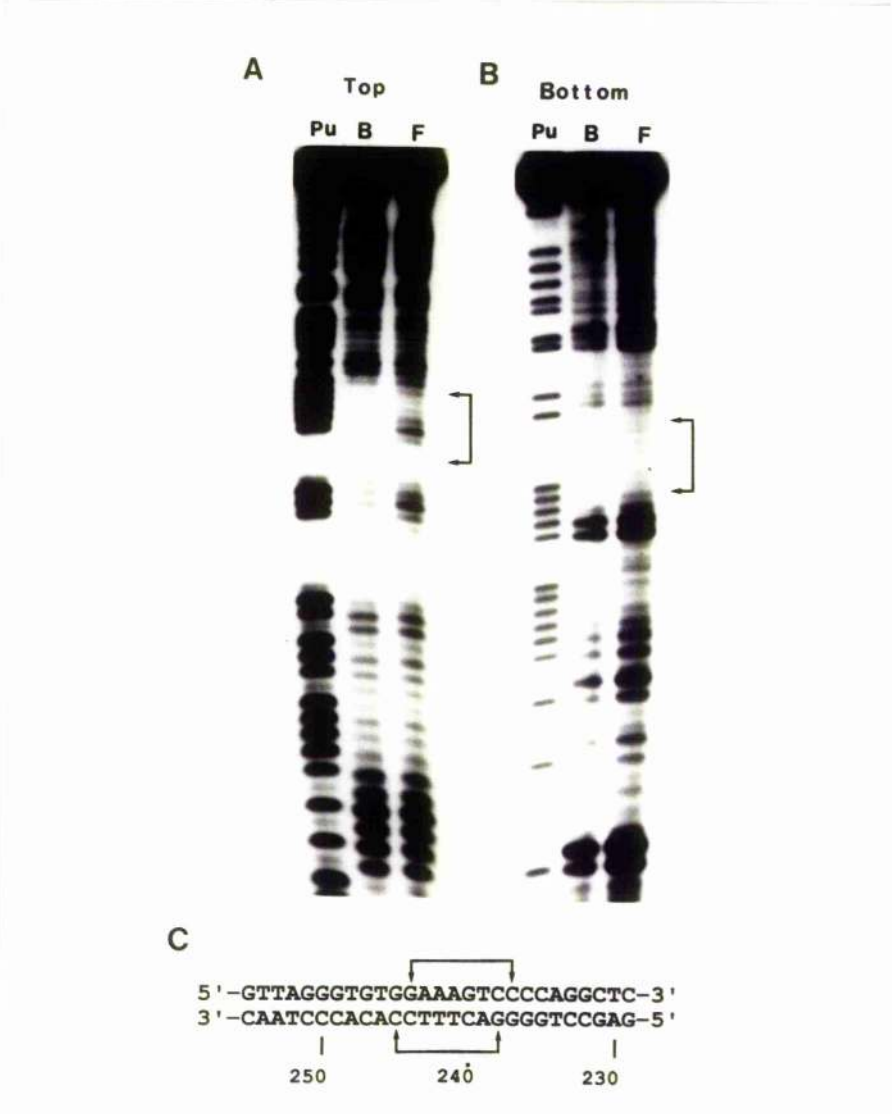
To examine further the disposition of EBP1 on the DNA we have determined the susceptibility of backbone deoxyribose residues to attack by hydroxyl radicals generated by OP/Cu+ (Spassky and Sigman, 1985). The 1,10-phenanthroline/ cuprous ion complex, together with mercaptopropionic acid as coreactant, nicks DNA in a reaction that requires the binding of the coordination complex to nucleic acid. It is thought that the OP/Cu+ ion binds in the minor groove of the DNA and generates a short-lived hydroxyl radical which can cause oxidative degradation of the deoxyribose ring. Since cleavage by OP/Cu+ produces predominantly 3'-phosphorylated and 5'-phosphorylated ends, reaction products can be accurately calibrated using Maxam and Gilbert chemical sequencing reactions. Protein bound to a site on the DNA may protect against hydroxyl radical damage by blocking binding of OP/Cu+ in the minor groove, or by directly shielding the DNA from the hydroxyl radical.

Taking advantage of the stability of DNA-protein complexes in native polyacrylamide gels we combined either a 5'- or 3'- [^{32}P]-labelled DNA fragment, containing an EBP1 binding site, with purified EBP1. Once equilibrium had been reached free DNA and DNA-protein complexes were separated by electrophoresis in a native polyacrylamide gel. Free DNA and DNA-protein complexes within the gel matrix were treated with OP/Cu+ *in situ* (Kuwabara and Sigman, 1987) and the cleaved DNA purified and analysed by denaturing polyacrylamide gel electrophoresis (Figure 5.9). It is immediately apparent that a small region of the DNA on each strand is protected from attack by the hydroxyl radical. On the top strand the region of protection encompasses eight deoxyribose residues and extends from the deoxyribose attached to base 244, to the deoxyribose attached to base 237 (Figure 5.9A). A similar situation is apparent on the bottom strand with eight deoxyribose residues protected from cleavage, extending from the deoxyribose attached to base 238, to the

Figure 5.9 1,10-orthophenanthroline/copper cleavage of the SV40 enhancer in the presence of EBP1

A *Bam* HI to *Eco* RI fragment from pUC1X72, 3'-end labelled (A, top) or 5'-end labelled (B, bottom) at the Bam HI site, was incubated with 10 µl affinity purified EBP1 and, after binding reactions had reached equilibrium, free DNA (F)was separated from EBP1–DNA complexes (B) by electrophoresis in a native polyacrylamide gel. Cleavage with OP/Cu+ was carried out within the acrylamide matrix and DNA eluted as described in "Materials and Methods". DNA was fractionated by electrophoresis in 6% denaturing polyacrylamide gels and cleavage products visualised by autoradiography. G+A (Pu) specific reactions of the labelled fragments were electrophoresed in parallel as markers. Deoxyribose residues within the bracketed region are protected from cleavage by OP/Cu+.

C. DNA sequence of the region of the SV40 enhancer containing the EBP1 binding site. Deoxyribose residues within the bracketed region are protected from OP/Cu+ cleavage in the presence of EBP1.



deoxyribose attached to base 244 (Figure 5.9B). A summary of the OP/Cu+ protection data is presented in Figure 5.9C. Hydroxyl radical footprinting, generated by an EDTA–Fe(II) complex, which does not interact with the DNA (Tullius and Dombroski, 1986),was carried out *in situ* and gave very similar results to that observed with OP/Cu+ (data not shown).

5.8 Summary of contact point data

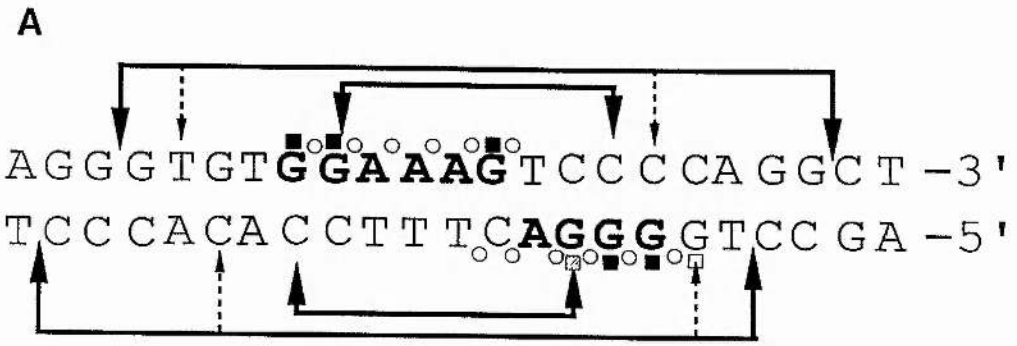
It has already been demonstrated using DMS chemical probing techniques that the pattern of contacts between EBP1 and its binding site on the SV40 enhancer is typical of the patterns observed on the site present in the IRE and the two sites present on the HIV enhancer (Figure 5.7). In addition, the pattern of protection by EBP1 on the IRE revealed by "footprinting" with hydroxyl radicals, generated by the OP/Cu+ complex (Figure 3.2B), is identical to that seen in similar studies of EBP1 on the SV40 enhancer (Figure 5.9). In view of this, results from the protection and interference data from the EBP1/SV40 enhancer interactions have been summarised in Figure 5.10. Figure 5.10B represents a double helical projection of the contact point data. To gain an insight into the DNA surfaces that are contacted by EBP1, we have used computer graphics to display the data generated in the contact point analysis (Figure 5.11). What is evident from these studies is that within a 10 bp region every purine base interferes with EBP1 binding when methylated and six phosphates on each strand interfere with binding when the attached oxygen groups are ethylated. The results suggest that EBP1 makes sequence-specific contacts with the DNA over one complete turn of the double helix and has binding determinants in both the major and minor grooves.

In summary, a 57,000 molecular weight polypeptide, EBP1, was identified which bound to the "core" region of the SV40 enhancer. Analysis of DNA sequences required for specific EBP1 binding distinguished this protein from a number of previously identified enhancer binding proteins. Multiple copies of the EBP1 binding site could enhance transcription *in vivo* when inserted into an expression vector. *In vivo* transcriptional activity of wild type and mutant SV40 enhancers (Zenke *et al.*, 1986) correlated with EBP1 binding. Mutations that abolish EBP1 binding also have a severe deleterious effect on transcription, indicating that this protein may be important for SV40 enhancer activity.

Figure 5.10 Summary of EBP1/DNA contact point analysis

A. Phosphodiester bonds within the large bracket are protected from cleavage with DNase I in the presence of EBP1. Deoxyribose residues within the smaller brackets are protected from cleavage with OP/Cu+. Bold letters indicate that these bases interfere with EBP1 binding when methylated. Filled squares indicate bases protected from methylation while the open square indicates that this base was hypermethylated in the presence of EBP1. The shaded square indicates that this base was protected in the plasmid pUC1X72 but not in pSV1. Open circles indicate the positions of phosphotriesters that interfere with EBP1 binding.

B. Helical projection of the EBP1 binding site. Modified bases (⊛) and phosphates (○) that interfere with binding are indicated, as are deoxyribose residues (★) that are protected from cleavage with OP/Cu+ in the presence of EBP1.



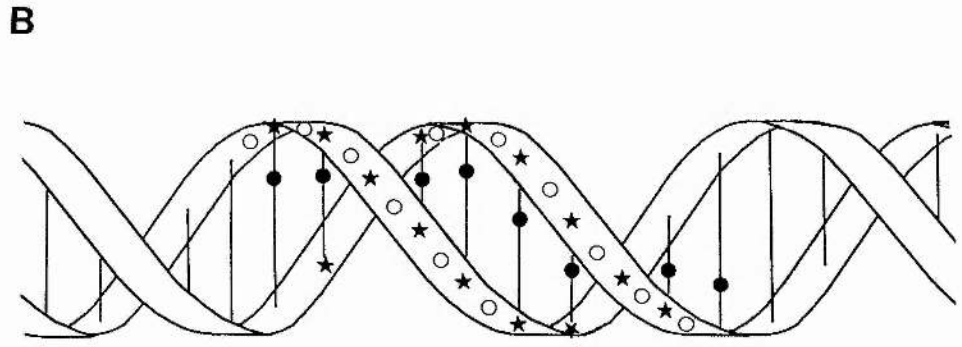
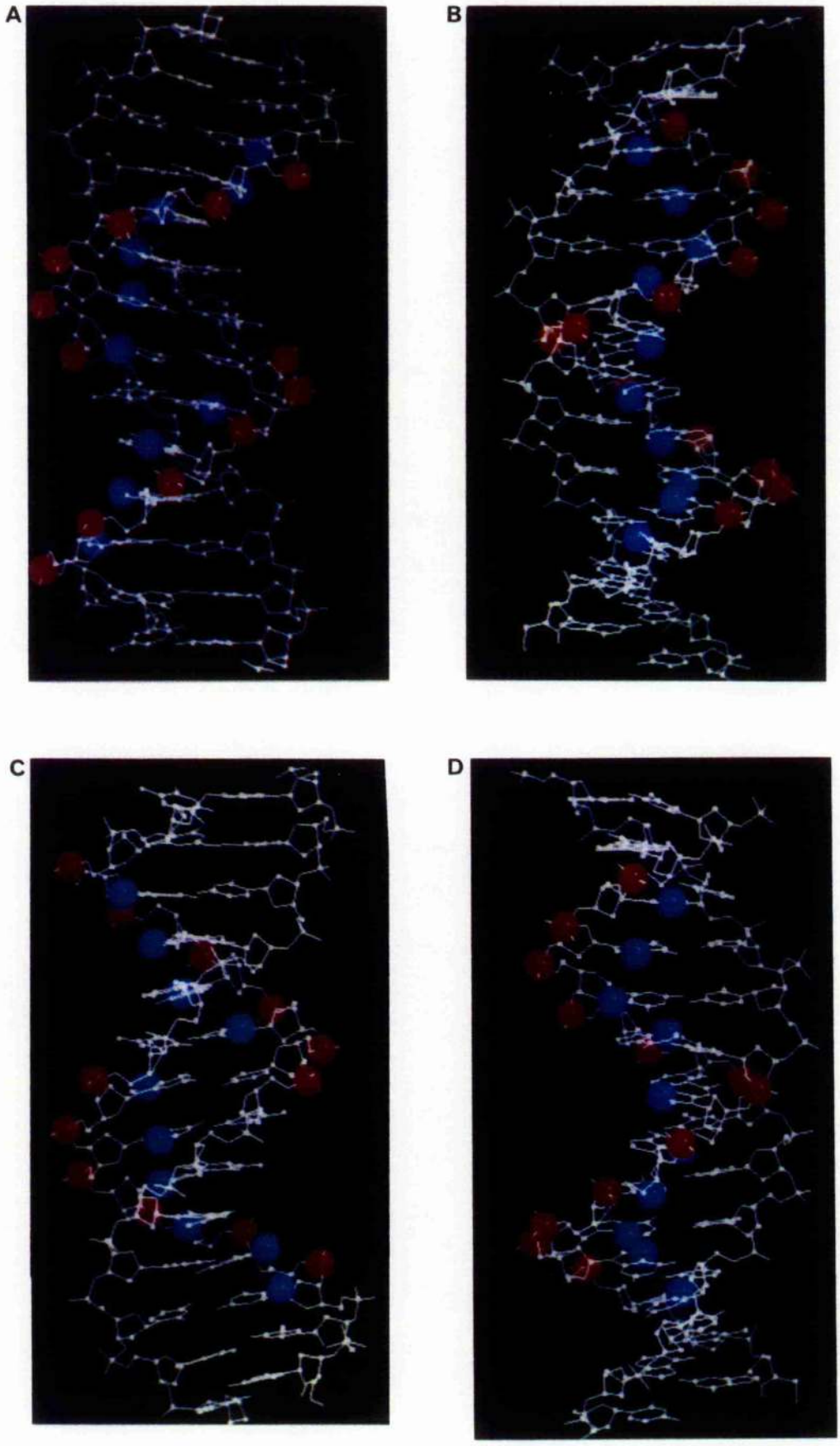


Figure 5.11 Computer graphic analysis of the contacts between EBP1 and its recognition site on the SV40 enhancer

A. Computer generated projection of the EBP1 binding site. The van der Waal's radii of nitrogen atoms that interfere with binding are indicated in blue while the modified phosphate groups that interfere with binding are indicated in red. The view in **A** is rotated through 90°, 180° and 270° in **B**, **C** and **D** respectively.


A
B
C
D

Methylation protection and alkylation interference studies identified purine bases and backbone phosphate groups which participate in the formation of a specific EBP1/DNA complex. In addition, "footprinting" with hydroxyl radicals revealed deoxyribose residues in the binding site which are protected from cleavage by EBP1. This information was displayed using computer graphics and indicated that EBP1 made specific base and backbone contacts over one complete turn of the DNA double helix, supporting a model in which the protein makes sequence-specific contacts in the major groove, although binding may also be influenced by interactions in the minor groove.

Competition and contact point analyses revealed that EBP1 bound similarly to sites present in several other viral and cellular enhancers including those present in the IRE and HIV enhancer. Mutational analysis of EBP1 binding sites identified base pairs important for specific EBP1/DNA complex formation. All high affinity binding sites contained the sequence 5'–GG$(N)_6$CC–3'. Although single base pair changes in the region between the conserved Gs and Cs can generally be tolerated, it is clear that the central six base pairs contribute to binding affinity. Mutations in the recognition site which could lead to gross structural changes in the DNA abolish EPB1 binding.

DISCUSSION

In this study, the identification, purification and characterisation of ~57,000 molecular weight HeLa cell nuclear protein, EBP1, and its interactions with DNA are described. A considerable body of evidence suggests that this protein, which binds to the enhancer "core" region of SV40 and to similar sequences present in a variety of viral and cellular enhancers, is involved in transcriptional activation.

1. The binding of EBP1 to the SV40 Enhancer and its Role as a Transcriptional Activator

EBP1 was detected in nuclear extracts from HeLa cells in gel electrophoresis DNA binding assays (Figure 1.3B) and the region of the enhancer bound by the protein was estimated, from DNase I protection data, to extend from position 230 to position 250 on the SV40 genome (Figure 1.5). This region includes the GT-I motif and spans the "core" sequence (5'-GTGG A/T A/T A/T G-3') identified by comparison of DNA sequences from a number of different enhancers (Laimins *et. al.*, 1982; Weiher *et. al.*, 1983). However, this sequence alone does not constitute the binding site for EBP1 since the SV2 double stranded oligonucleotide, which spans the GT-II motif and also contains the "core" sequence, fails to compete for EBP1 binding (Figures 1.3B and 1.5). While this sequence is clearly an important component in the recognition event, mutation of the G's at positions 244 and 245 abolishes binding (Figures 1.5; 2.5; 3.1; 4.1), additional sequences are required to constitute a functional binding site for EBP1. These additional sequences are identified by the mutations in the SV1.M3 oligonucleotide; C's at positions 237 and 238 are changed to G's with the result that the double stranded oligonucleotide can no longer be bound by EBP1 (Figures 3.1; 4.1).

The extension of the binding site past the GT-I motif and into the TC-II motif differentiates this factor from a number of previously identified proteins which bind to the SV40 enhancer (Barrett *et. al.*, 1987; Johnson *et. al.*, 1987; Xiao *et. al.*, 1987a). Purification of the transcription factors, AP-2 and AP-3 from HeLa cells, which bind close to this region have recently been described (Imagawa *et. al.*, 1987; Mitchell *et. al.*,1987). However,EBP1 can be distinguished from both these factors by virtue of the abilities of the different proteins to bind to mutated SV40 templates. The plasmid pA12

(Figure 2M) contains a mutated SV40 enhancer element in which the sequence 5'-TGG-3' (nucleotides 246-244) is altered to 5'-GTT-3' (Zenke *et al.*, 1986). AP-2 binds normally to this fragment (Mitchell *et al.*, 1987) whereas EBP1 binding is abolished (Figure 4.3), indicating that EBP1 and AP-2 have distinct recognition sites. EBP1 binding is also abolished by the mutation present in pA15 (Figure 4.3), in which 5'-CCC-3' (nucleotides 237-235) is altered to 5'-AAA-3' (Zenke *et al.*, 1986), whereas AP-3 binds normally to this template (Mitchell *et al.*, 1987), thus distinguishing EBP1 from AP-3.

The SV40 enhancer contains a number of DNA sequence motifs (Figure 1.1) that can be bound by a variety of *trans*- acting factors, and whose integrity is crucial for full enhancer activity (Zenke *et al.*, 1986; Herr and Clarke, 1986; Schirm *et al.*, 1987; Davidson *et al.*, 1986; Fromental *et al.*, 1988). Multiple tandem copies of the "core" enhancer sequence can function in a cell type-specific manner to activate transcription of a linked gene (Ondek *et al.*, 1987; Schirm *et al.*, 1987). To ascertain whether or not the DNA sequences in the "core" region of the SV40 enhancer, which bound EBP1, play a role in transcriptional activation, the double stranded SV1 oligonucleotide, present as 1, 2 or 3 copies, was introduced into an expression vector and the effects of these sequences on transient gene expression in HeLa cells determined. Results of these studies (Figure 1.6) showed quite clearly that multimerisation of this sequence led to an increase in transcriptional activation demonstrating that the sequences which constitute the EBP1 binding site on the SV40 enhancer can indeed function to activate transcription *in vivo*.

A role for EBP1 in transcriptional activation of the SV40 genome is also suggested by correlating the *in vivo* activity of wild type and mutated enhancers with EBP1 binding *in vitro*. As previously demonstrated, tandemly ligated copies of an oligonucleotide containing the EBP1 binding site can function to activate transcription when inserted into an appropriate expression vector (Ondek *et. al.*, 1987 ; this study Figure 1.6). Mutation of the two G's at positions 244 and 245 in the EBP1 binding site abolished the enhancer activity of the ligated oligonucleotides (Ondek *et. al.*, 1987) and also abolished binding of EBP1 (Figures 1.5; 2.5; 3.1; 4.1). Nomiyama *et al.* (1987) have examined the activity of the pA series of triple point mutants in the SV40 enhancer (Zenke *et al.*, 1986) in a variety of cell types and correlated the activity of the various enhancer motifs with the presence of *trans* -acting factors. One cell type-specific protein, GT-IC (probably equivalent to

AP-3), present in HeLa but not MCPII cells, binds to the GT-I motif in the SV40 enhancer (Xiao *et al.*, 1987a). An additional cell type-specific protein TC-IIA and a ubiquitous protein TL-IIB are reported to bind to the TC-II motif (Nomiyama *et al.*, 1987). TC-IIA and TC-IIB appear to have very similar recognition sites to that of EBP1. The mutation present in pA12 (described above) abolishes binding of all of these proteins and dramatically reduces transcription in all of the cell lines tested so far (Nomiyama *et al.*, 1987). Mutations in the GT-I motif (pA10 and pA11), which eliminate GT-IC binding but which have no effect on the binding of TC-IIA, TC-IIB or EBP1 (Figure 4.3), have no effect in MCPII cells where GT-IC is absent. In HeLa cells mutations pA10 and pA11 have only a minimal effect on enhancer activity whereas mutation pA15, which abolishes binding of TC-IIA, TC-IIB and EBP1 reduces transcription, *in vivo*, to 40-50% of wild type (Figure 1D). Since binding of GT-IC (AP-3) and EBP1 (TC-IIB) is eliminated by pA12, it suggests that in HeLa cells these proteins may compete for the same binding site. The minimal effect of mutations pA10 and pA11 in HeLa cells suggests that EBP1 (TC-IIB) is bound preferentially to the template as opposed to GT-IC. Elimination of EBP1 (TC-IIB) binding by mutant pA15 permits binding of GT-IC (AP-3) which also appears to stimulate transcription since a template which contains both the pA10 and pA15 mutations gives only 5% enhancer activity in HeLa cells (Nomiyama *et al.*, 1987). These data are summarised in Figure 1D, which shows the disposition on the SV40 enhancer of the transcription factors AP-2, AP-3 (Mitchell *et. al.*, 1987) and EBP1.

The binding site for EBP1 quite clearly overlaps the binding sites for both AP-2 and AP-3, as discussed above, and binding of EBP1 may exclude binding of these transcription factors or *vise versa* . This poses the question of which proteins are bound to the enhancer *in vivo* in cell types other than HeLa. Answering this question is important since it has been shown that the regions bound by AP-3 and EBP1 can act to mediate the transcriptional response to phorbol esters (Chiu *et. al.*, 1987). Thus the multimerised C element described by Ondek *et al.* (1987), which is equivalent to the SV1 oligonucleotide used in this study, can act as a phorbol ester responsive element (Chiu *et al.*, 1987). In addition, mutation of the two G's at positions 244 and 245 destroys the ability of the multimerised oligonucleotides to act as a phorbol ester inducible element (Chiu *et al.*, 1987) and, as we have shown here, abolishes the ability of EBP1 to recognize its binding site (Figures 1.5; 2.5; 3.1; 4.1). Correlation between the *in vivo* transcriptional activity of wild type and

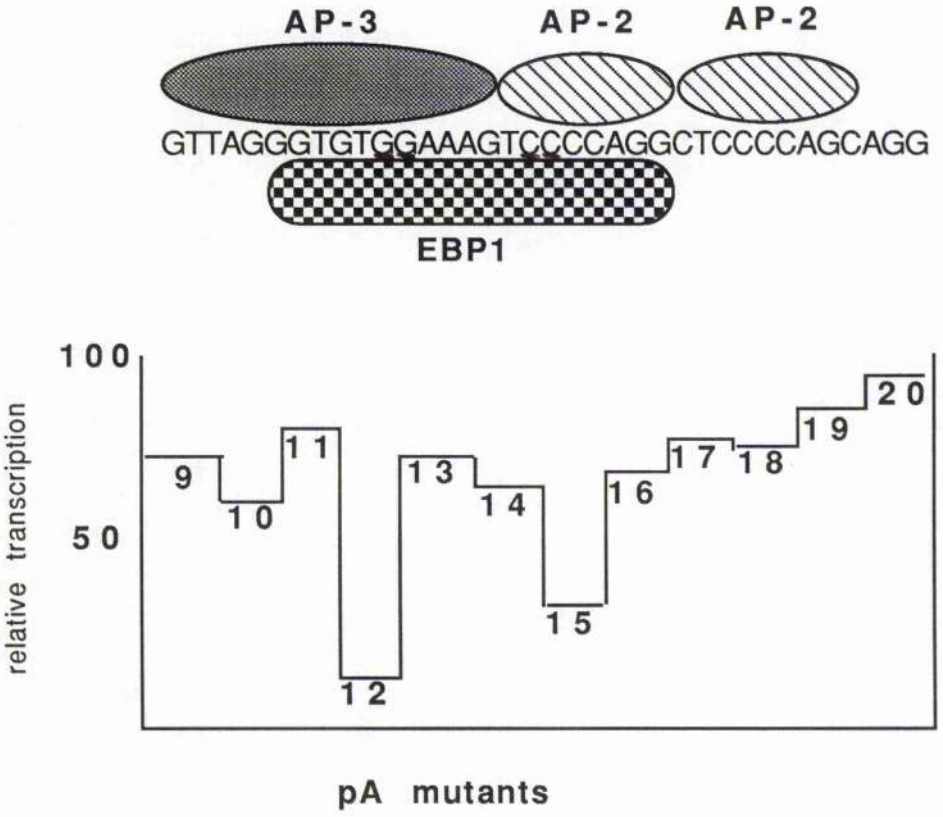


Figure 1.D Disposition of transcription factors AP-2, AP-3 and EBP1 on the SV40 enhancer

The DNase I "footprint" boundaries for AP-2 and AP-3 on the SV40 enhancer DNA sequence are from Mitchell *et al*. (1987) and EBP1 from this study. It should however be stressed that these boundaries probably represent an over estimate of the region occupied by the DNA binding proteins, since DNase I is a relatively large molecule and requires an extended binding surface for DNA cleavage (Suck *et al.*, 1988). The effect of mutations in this region on transcription *in vivo* are from Zenke *et al*., 1986. Altered bases in the mutated SV1 oligonucleotides SV1.M1 and SV1.M3, that abolish EBP1 binding, are underlined.

mutant enhancers with EBP1 binding *in vitro,* indicates that this protein may be important for SV40 enhancer activity since mutations that abolish EBP1 binding also have a severe deleterious effect on transcription *in vivo* (Zenke *et al.*, 1986).

The human ß-interferon gene regulatory element (IRE) (Goodbourn *et al.*, 1985) consists of two genetically separable positive regulatory domains, PRDI and PRDII, both of which are required for maximal induction by virus or poly(I).poly(C), and a negative control region, NRDI, which overlaps PRDII (Goodbourn and Maniatis, 1988). Interactions between the regulatory domains were investigated by analysing the effects of the IRE single base pair mutations on the expression of the ß-interferon gene (Goodbourn and Maniatis, 1988). Single point mutations in PRDII have the potential to affect binding of both negative and positive effectors such that the resulting phenotypes may be rather complex. PRDII was identified as being crucial for induction of ß-interferon expression since mutations at positions -64 to -55 inclusive are all defective in their ability to respond to agents which activate ß-interferon transcription. This effect has been ascribed to the inability of a positive transcription factor to interact with mutant PRDIIs. EBP1 binds to a region within PRDII which shows strong sequence similarity to the EBP1 site on the SV40 enhancer (Figures 3.1; 4.2). Examination of the IRE point mutants which alter the conserved guanine and cytosine blocks in the EBP1 binding site, that is -64A, -63A, -56T and -55T, reveals that these mutations reduce EBP1 binding to almost undetectable levels (Figure 4.2). In addition, the -64A and -56T mutations reduce induced transcription levels to virtually zero, whilst -63A and -55T mutations reduce transcription to approximately half that of wild type levels (Goodbourn and Maniatis, 1988). Mutant -61G, which increases EBP1 binding to higher than wild type levels, shows an almost two-fold increase in relative transcription level after induction, whereas mutation -62A, which abolishes EBP1 binding and becomes hypermethylated at high concentrations of EBP1, demonstrates a 50% reduction in induced transcription levels. It would therefore appear that mutations in PRDII of the IRE which increase or decrease EBP1 binding, show a corresponding increase or decrease in the relative transcriptional induction of ß-interferon mRNA lending support to the idea that EBP1 does indeed function as a transcriptional activator *in vivo.*

2. Nucleotide Sequence Requirements for Specific EBP1/DNA Complex Formation

To determine the DNA sequences required for specific EBP1 binding, a detailed mutational analysis of the EBP1 binding sites in the IRE and SV40 enhancer has identified the nucleotide requirements necessary for specific EBP1/DNA complex formation. More that thirty DNA sequences related to the binding site for EBP1, including those present in the mouse Ig kappa light chain gene enhancer, mouse class 1 H2 major histocompatibility gene enhancer, HIV enhancer, IRE, SV40 enhancer and mutated derivatives thereof, were analysed for their ability to bind EBP1. A common feature in all the DNA sequences which bind EBP1 is that they possess the 10 bp sequence 5'-GG$(N)_6$CC-3', within which every purine base interferes with EBP1 binding when methylated. Single base pair changes, as revealed in studies using mutant IRE templates (Figure 4.2), between the conserved Gs and Cs can generally be tolerated (Table 4.1), although it is clear that these bases contribute to binding affinity. This is exemplified at position -61 in the IRE where an A to T change reduces binding, but an A to G change increases binding.

Mutational analysis also reveals that alteration in the spacing between the conserved Gs at positions 1 and 2, and the conserved Cs at positions 9 and 10 (Figure 3.1; Table 4.1) abolish EBP1 binding. The SV1.M3 double stranded oligonucleotide alters this spacing to 7 bp, whilst the -64A IRE mutant alters it to 5 bp.; both of these are unable to bind EBP1 (Figures 4.1; 4.2). Furthermore, binding studies revealed that mutants pA11 and 16, which lie on the boundary of the SV40 EBP1 recognition site have only a slight effect on EBP1 binding (Figure 4.3) whereas mutations pA12, 13, 14 and 15, which lie within the recognition site, and also span the 10 bp region which, when methylated, interferes with EBP1 binding (Figure 5.3), reduce binding by at least an order of magnitude (Figure 3.1). That no changes are tolerated within the EBP1 recognition site in this instance may be due to the inability of the DNA to accommodate the gross structural changes which may arise as a result of the triple point mutations.

The importance of DNA structure in directing high affinity EBP1 binding is suggested by the effect of mutations at -62, in which a G to A alteration abolishes EBP1 binding. This was somewhat unexpected since an adenine is permitted at this position in the EBP1 binding site in the SV40 enhancer (Figures 1.5; 2.5). The -62A mutant creates a situation whereby four consecutive

adenine residues are present in the binding site. The unusual structural features of an oligo(dA).oligo(dT) tract and its biological implications have been described by Nelson *et al.*, (1987). Their studies revealed that these sequences have an unusually high propeller twist of the bases along their longitudinal axis with a consequent narrowing of the minor groove. This in turn causes the major groove side of each base to point towards the 3'-end of its strand, thereby facilitating the formation of three-centred bifurcated hydrogen bonds. The inability of the -62A DNA template to bind EBP1 may therefore be due to structural changes in the DNA in the region of the four consecutive adenine residues.

3. Contact Point Analysis of the EBP1 Binding Site

The area of the SV40 enhancer occupied by affinity purified EBP1 was estimated from the DNase I protection data (Figures 2.5; 5.1). Although 19 bp on each strand are protected from DNase I cleavage, the area occupied by EBP1 is probably much smaller since DNase I has an extended DNA binding surface with 6 bp on the 5' side and 2 bp on the 3' side of the cutting site being occupied by the enzyme (Suck et al., 1988). This would indicate that EBP1 occupies the DNA between bases 248 and 236 on the top strand and 235 (or 238) and 247 on the bottom strand.

Within the region occupied by EBP1 a total of 8 deoxyribose moieties on each strand are protected from hydroxyl radical attack (produced by OP-CU+) in the presence of EBP1. Comparison of the region of the DNA occupied by EBP1 (as predicted from DNase I cleavage patterns and the structure of DNase I) with that protected from cleavage with OP-Cu+ indicates that the two are not symmetrically disposed (Figure 5.9). Neither is the OP-Cu+ "footprint" symmetrically arranged with respect to the G residues 236 and 237, and , 244 and 245 which are protected from DMS methylation (Figure 5.2). Furthermore, EBP1 does not protect the sugar residues attached to G236, 237 and 245 from hydroxyl radical attack (Figure 5.9). Thus, even although EBP1 is interacting with bases in the major groove, the DNA backbone is still susceptible to attack by hydroxyl radicals generated by OP-Cu+, probably located in the minor groove.

On the top strand, the 5' boundary of the OP-Cu+ "footprint" coincides with the 5' boundary of the phosphate groups that interfere with binding when the attached oxygen atoms are ethylated

(Figure 5.8). However, at the 5' boundary of the binding site on the bottom strand, the situation is quite different. Modification of phosphate groups on the 3' side of bases 235 and 236 interferes with binding but the adjacent deoxyribose residues are not protected by EBP1 from hydroxyl radical attack. These data stress the asymmetry of the EBP1-DNA interaction and suggest that , particularly at the 5' boundaries of the binding site, EBP1 makes contacts with the phosphate oxygens without completely covering the DNA backbone. On each strand six phosphate groups reproducibly interfere with binding indicating that EBP1 makes backbone contacts on the SV40 enhancer over one complete turn of the DNA double helix.

The case for EBP1 making extensive major groove interactions is provided by the DMS protection data which indicate that on G residues 236, 237, 240, 244 and 245, the N-7 groups on the purine ring are protected from methylation (Figure 5.2). In B-DNA these N-7 groups project into the major groove and protection of these groups from DMS modification implies that at these positions EBP1 makes close contact with the DNA. Guanine 235 has an interesting effect on the pattern of DMS modification. This residue is not required for EBP1 binding in that the plasmid pSV1, which has a C at this position, is bound by EBP1 with a similar affinity as that of the complete SV40 enhancer (Figure 3.1). In the presence of EBP1 the DMS modification rate of G235 is greatly increased. However, when G235 is present, G238, which is not protected in pSV1, is protected from DMS modification by EBP1 (Figure 5.2). The increased rate of G235 methylation could be explained by an increase in N-7 nucleophilicity caused by bound EBP1. Alternatively, EBP1 could create a hydrophobic pocket close to the recognition site resulting in a local increase in DMS concentration. A high local concentration of DMS near residue 235 may be accompanied by a correspondingly reduced DMS concentration near residue 238. It is worth noting that G235 is predicted, from DNase I protection data, to lie precisely on the boundary of the region occupied by EBP1 (Figure 5.10).

All G residues which are protected from DMS modification also interfere with EBP1 binding when methylated. In addition, methylation of A residues 239, 241, 242 and 243 abolishes EBP1 binding. Modification of A residues by DMS results in the addition of a methyl group at the N-3 position. Since N-3 of adenine projects into the minor groove a straightforward interpretation of these results is that EBP1 makes base-specific contacts in the minor groove. It should however be

stressed that modification of the purine ring by DMS results in the addition of a positive charge to the ring which could have profound effects on specific DNA binding, quite distinct from the steric effects of introducing a bulky methyl group at the N-3 position. In particular, it has been shown that sequence-specific recognition of the operator sequence by the 434 bacteriophage repressor takes place by the formation of H-bonds between amino acid side chains and H-bond donors and acceptors arrayed along the edge of the bases that project into the major groove (Anderson *et al.*, 1985). One of the potential H-bond acceptors is the imidazole nitrogen at position 7 of the purine ring. The effect of introducing a positive charge into the purine ring would be to withdraw electrons from this nitrogen thus reducing its ability to act as a hydrogen bond acceptor. Consequently the effect of N-3 methylation on EBP1 binding may be as a result of the inability of EBP1 to make base-specific hydrogen bonds in the major groove and not of disruption of minor groove contacts. Indeed a number of observations suggest that the latter alternative is more likely. Examination of the binding of EBP1 to a variety of related sequences has shown that EBP1 binds with high affinity to sites that have guanine to adenine substitutions at positions 242 and 243(Table 4.1). Methylation of these substituted G residues also blocks EBP1 binding (Figure 5.3) suggesting sequence-specific contacts in the major groove.

To gain an insight into the DNA surfaces in the SV40 enhancer that are contacted by EBP1 computer graphics were employed to display data generated by the contact point analysis (Figure 5.11). What is apparent is that EBP1 makes contacts with each base pair over one complete turn of the DNA helix. Guanine residues 236 and 237, on the bottom strand, and 244 and 245 on the top strand, which are one turn of the helix apart, are crucial for EBP1 binding since their mutation reduces EBP1 binding to undetectable levels. Phosphate backbone contacts made by EBP1 also span one turn of the DNA helix which is unlike the pattern of contacts made by proteins which contain the well documented "helix-turn-helix" motif exemplified by the bacteriophage 434 repressor (Anderson et al., 1985). EBP1 binding is not reduced by exposure of the protein to chelating agents such as EDTA (10mM) and 1,10-orthophenanthroline (10mM) and the pattern of contacts is unlike that made by "zinc finger" proteins such as TFIIIA (Bogenhagen et al., 1980; Fairall et al., 1986). The observed contacts for EBP1 are also arranged differently from other well characterised eukaryotic DNA binding proteins such as SV40 large T antigen (Jones and Tjian,

1984), Drosophila heat shock transcription factor (Shuey and Parker, 1986) and human nuclear factor I (de Vries et al., 1987).

EBP1 binds to sequences in the PRDII region of the human ß-interferon gene regulatory element (IRE). This study demonstrates, using nuclease protection, chemical probing techniques and mutational analysis, that EBP1 binds to its sites in the IRE and SV40 enhancer in a very similar manner. DNase I protection studies show that EBP1 occupies a region of 19 bp on each strand of the IRE and that, within this region, a total of eight deoxyribose moieties on each strand of the DNA backbone are protected from hydroxyl radical attack (generated by OP/Cu+) in the presence of EBP1 (Figure 3.2).

EBP1 protects guanine residues -64, -63, -56 and -55 in the IRE from methylation by DMS (Figure 5.6) which suggests protein contacts in the major groove. All guanine residues which are protected from DMS modification also interfere with EBP1 binding when methylated. A total of 10 purine bases within the 10 bp from -55 to -64 of the IRE interfere with EBP1 binding when methylated. On the top strand the six consecutive purine bases -59, -60 and -61A, and -62, -63 and -64G all interfere with binding when methylated (Figure 5.6B). A similar situation is apparent on the bottom strand where the four consecutive purine bases -55 and -56G, and -57 and -58A also interfere with binding when methylated (Figure 5.6B).

Two EBP1 binding sites are present in the HIV enhancer and DNase I "footprinting" has revealed that both these sites are protected from cleavage in the presence of EBP1 (Figure 3.3). Guanine residues -88, -89, -102 and -103 on the top strand (Figure 5.4A), and -80, -81, -85, -94, -95 and -99 on the bottom strand (Figure 5.4B) are protected from DMS methylation in the presence of EBP1. These residues include the conserved guanines shown to be essential components of the EBP1 recognition site (Table 4.1; Figure 3.1). In addition, adenine residues at positions -82, -86, -90 and -104, guanine at position -86, and cytosine at position -85 show increased reactivity to DMS methylation in the presence of EBP1. This pattern of hypermethylated bases is somewhat unusual in that modified DNA due to adenine methylation is not generally cleavable by the action of piperidine. An explanation of this observation is that methylation probably occurs at a site other than N-3 of the purine ring, such as N-7. Alternatively perturbation of the DNA structure in this region may be responsible for the observed effects since it is known that DMS Methylation of cytosine residues occurs at the imidazole N-3 in the pyrimidine ring directly involved in H-bonding to its

complement on the DNA. Accessibility of this base to DMS methylation therefore implies that either the DNA in this region was single stranded (see Boroweic and Hurwitz, 1988), or that the DNA was quite dramatically distorted. The pronounced increase in reactivity towards DMS shown by guanine at position -90 could also stem from perturbation of the DNA in this region since this residue lies between the two EBP1 binding sites on the HIV enhancer.

It is apparent from the pattern of contacts resulting from DMS chemical probing studies that EBP1 binds similarly to different viral and cellular enhancers. In each case a similar pattern of protection is observed within the 10 bp recognition sequence is preserved in that guanines at positions 1 and 2 are protected from DMS methylation, as are the guanines at positions 9 and 10 on the opposite strand. In addition, the highly strand-specific pattern of purine bases which interfere with EBP1 binding is evident in each case. Taken together, the methylation protection and interference data indicates that EBP1 binds homologously to its sites in the SV40 enhancer, IRE and HIV enhancer and in each case makes sequence-specific contacts over one complete turn of the DNA double helix.

4. The "Identity" of EBP1

The binding site for EBP1 is similar to that described for the inducible factor NF-kB (Sen and Baltimore, 1986a,b) and the apparently ubiquitous factor H2TFI or KBFI (Baldwin and Sharp, 1987; 1988; Israel *et al.*, 1987; Yano *et al.*, 1987). These proteins may constitute a family of related, but distinguishable, proteins which recognize similar DNA sequences. Synthetic oligonucleotides containing the NF-kB and H2TFI binding sites compete for EBP1 binding with similar efficiency to the SVI oligonucleotide (Figure 3.1). However, EBP1 can be distinguished from NF-kB and H2TFI/KBFI by a number of criteria. NF-kB binds to a site in the Ig kappa light chain enhancer (Sen and Baltimore, 1986a) but, unlike EBP1 which is constitutively produced in HeLa cells, NF-kB binding activity is normally restricted to B cells and its activity is only apparent in other cell types after stimulation with agents such as bacterial lipopolysaccharide and phorbol esters. This induction of NF-kB binding activity is evident even when the stimulation is carried out in the presence of inhibitors of protein synthesis. It therefore appears that prior to induction NF-kB exists in the cell in an inactive form that is converted to an active form by an as yet undefined post-translational

modification (Sen and Baltimore, 1986b). Preliminary data has shown that NF-kB, induced in mouse B cells, binds to the SV1 oligonucleotide, but the DNA-protein complex formed has an increased electrophoretic mobility when compared to the EBP1/SV1 complex (data not shown).

The distinguishing features of H2TFI/KBFI are its high affinity for the sequence

-TGGGGATTCCCCA-
-ACCCCTAAGGGGT-

which is found in the enhancers of the mouse H-2 and B2-mircoglobulin genes, and its unique pattern of methylation interference (Baldwin and Sharp, 1988; Israel *et al.,* 1987). As mentioned previously, EBP1 binds the SV1 and H2TFI/KBFI sequences with equal affinities whereas H2TFI/KBFI binds the sequence indicated above with a 20-fold higher efficiency (Baldwin and Sharp, 1988). Methylation interference experiments with H2TFI/KBFI indicate that all of the G's in the sequence interfere with binding when methylated (Baldwin and Sharp, 1988; Israel *et al.,* 1987). This is in contrast to the pattern obtained on the SV40 enhancer and EBP1, where the outermost G's in the sequence do not interfere with binding when modified (Figure 5.6). An attractive explanation of these differences is that EBP1, NF-kB and H2TFI/KBFI represent the different modified forms of the same DNA binding protein.

Both the interleukin-2 receptor alpha (IL-2Rα) gene and the long terminal repeat (LTR) of HIV-1 can be activated by various T-cell mitogens, including PMA, PHA and the *tax* gene product of HTLV-1 (Nabel and Baltimore, 1987; Kaufman *et al.,* 1987). The HIV-I enhancer has been shown to interact with an inducible factor indistinguishable from NF-kB (Nabel and Baltimore, 1987). The binding of NF-kB to the *tax* responsive sequences in IL-2Rα suggest that NF-kB (or a protein with a similar DNA recognition sequence, such as EBP1) is involved in the *tax*- mediated *trans*- activation pathway. Bohnlein *et al* (1988) have investigated the *trans*- acting factor(s) and *cis*- acting sequences that regulate mitogen inducibility of the IL-2Rα promoter and the potential interplay of the same factors with the HIV-1 LTR. They were able to demonstrate the binding of an NF-kB-like inducible factor to the sites in the HIV-I enhancer which bind EBP1 (Figure 3.3). These sites bear marked homology to a region in the IL-2Rα promoter. In addition, they could show that the inducible, host cell encoded, HIVEN86A protein could associate specifically with these sites in the IL-2Rαpromoter and HIV-i enhancer. Isolation of the IL-2Rαpromoter binding site , or single copies of the individual HIV-1 sites, were sufficient to impart mitogen inducibility to an unresponsive heterologous promoter. This suggests that the normal action of an inducible nuclear DNA binding

protein(s) involved in the regulation of IL-2Rα gene expression can be subverted by the HIV-1 provirus to facilitate activation of retroviral gene transcription. This emphasises the important interplay of host cell factors, like NF-kB and probably EBP1, with this retrovirus in the regulation of latent versus lytic viral infection.

Kawakami and coworkers (1988) have demonstrated that purified NF-kB stimulates transcription from the HIV-1 promoter in HeLa cell extracts and that this stimulation is specific for templates containing NF-kB sites. However, their studies also showed that affinity purified NF-kB bound only to the distal site in the HIV-1 enhancer and suggest that this may reflect a steric problem due to the close proximity of the binding sites. Affinity purified EBP1, on the other hand, binds with high affinity to both these sites on the HIV-1 enhancer (Figure 3.3). Since the distal site can be occupied by NF-kB, the proximal site could, in principal, be occupied by a common cellular factor, such as EBP1 or H2TF1/KBF1, or by NF-kB itself. Binding and functional studies using the various purified factors will clearly be required to resolve the mechanisms involved in HIV-1 promoter activation.

Although we have no direct evidence for post-translational modification of EBP1, this phenomenon may account for the observation that two distinct DNA-protein complexes can be detected at early stages of purification (Figure 2.1). Both complexes have the same DNA-binding specificity and are cross-linked to polypeptides of identical molecular weight (Figure 2.4). After affinity purification however only the more slowly migrating complex is detected Figure 2.2). This may result from instability of the factor responsible for formation of the faster migrating DNA-protein complex, modification of the factor during extraction and purification or removal of associated factors during purification. An apparent "activation" of binding activity *in vitro,* by proteolytic cleavage, has also been demonstrated (Figure 2.6). An explanation for this observation could be that the increase in binding activity results merely in the release of active EBP1 that is otherwise trapped in aggregates. Both elastase and protease XIV remove the aggregates present at the top of the gel, but only protease XIV stimulates binding of EBP1 to the double stranded SV1 oligonucleotide indicating that the specificity of the proteolytic cleavage is important for the effects observed above. Whether proteolytic cleavage of EBP1 represents a post-translational modification of relevance *in vivo* remains to be determined.

The activation of DNA binding activity of what appears to be a cytoplasmic precursor of NF-kB

has recently been reported (Baeuerle and Baltimore, 1988). Dissociating agents were used to effect this activation which resulted in the appearance of as much cytosolic NF-kB in unstimulated pre-B cells as is found in the nucleus when the same cells are activated with phorbol ester. The activation of specific DNA binding could be due to covalent modification of NF-kB, a conformational change in the protein, or by the release of an inhibitor which prevents specific binding to DNA. They suggest that the activation of binding activity is attributable to the release of an inhibitor, present in the cytosol, which blocks translocation of NF-kB to the nucleus and subsequent binding to its cognate site on the DNA. An attractive interpretation *in vivo* of these effects might be that protein kinase C, by phosphorylating the inhibitor (or another factor that can inactivate the inhibitor) would have the same effect as the denaturing agents *in vitro*. The data of Baeuerle and Baltimore (1988) help to explain how NF-kB is induced in immature B cells, Hela cells and other cells (such as T cells) but does not explain how B cells constitutively express NF-kB in an active state – this requires further investigation.

EBP1, NF-kB and H2TF1/KBF1 may represent different modified forms of the same DNA binding protein which contains a conserved DNA binding domain. Isolation and expression of cDNAs corresponding to each of these proteins, as well as a quantitative evaluation of their binding to the different templates, will be required to establish the relationship between this family of proteins which appear to recognise the same DNA sequence. The recent cloning of the gene coding for the DNA binding domain of a protein which binds to the NF-kB and H2TFI recognition sequences lends support to this idea (Singh *et al.*, 1988).

5. Models for enhancer-mediated transcriptional activation

The ability of the individual sequence modules of enhancer elements to bind different transcription factors yet activate transcription in essentially the same way, suggest that these *trans*- acting factors exert their effects via a common mechanism. Of central importance in the scheme of enhancer action is the observation from the laboratories of Schaffner (Weber and Schaffner, 1985) and Maniatis (Treisman and Maniatis, 1985) that enhancers function by increasing the RNA polymerase density over a linked gene, i.e., they increase the rate of transcription.

In the first of these models, the polymerase entry site, or scanning model, the enhancer has a

very high affinity for RNA polymerase II. The polymerase binds to enhancer sequences and slides, in either direction, along the DNA, until it reaches promoter initiation signals where it begins to transcribe RNA (Moreau *et al.*, 1981). Additional support for this model came from the experiments of Brent and Ptashne (1984) in which the *lexA* repressor DNA binding site was inserted between the promoter and UAS of a yeast ß-galactosidase expression vector. This construct was introduced into yeast cells, together with a *lexA* expression vector, which resulted in reduction of ß-galactosidase activity. The same result was obtained when the *lexA* operator was replaced by a transcription termination signal. These results were interpreted as being due to RNA polymerase II binding at the UAS and sliding along the DNA to the promoter unless prevented from doing so by operator-bound *lexA* or a transcription terminator. Support for this model also comes from the observation that enhancers and tandem promoters in mammalian cells preferentially activate proximal promoters (reviewed by Muller *et al.*, 1988). Certain mechanistic problems, however could arise from this type of model. In particular, on activation from a 3' position, collisions would be expected between the sliding polymerase and polymerase actually involved in active transcription. Even so, *in vitro* experiments have demonstrated that a prokaryotic RNA polymerase can transcribe through a 5S RNA gene, displacing neither the specific 5S RNA transcription complex, nor the nucleosome cores (reviewed by Muller *et al.*, 1988). In this way it is feasible that two polymerases could pass each other without releasing from the DNA. (It should be stressed however, that this is still a highly controversial area since it is not known what actually controls the directionality of sliding. It is probably not the enhancer region since this can be moved 5' or 3' of the gene and maintained in the same orientation, resulting in an enhancer effect.)

The second model for enhancer action, and the more favourable of the two, is the looping model in which a remote enhancer and a promoter interact with one another via proteins bound to the DNA (Figure 2D). In prokaryotic systems, such DNA loops have been extensively investigated (reviewed by Ptashne, 1986c). There exists in prokaryotes examples of gene regulation by repressor binding sites on the same face of the DNA double helix. This effect is highly sensitive to the length of intervening DNA in that insertion of even one half turn of the helix can virtually abolish cooperativity (Dunn *et al.*, 1984). Hochschild and Ptashne (1986) examined cooperativity between two binding sites for lambda repressor and found that a variable number of full helical turns of the

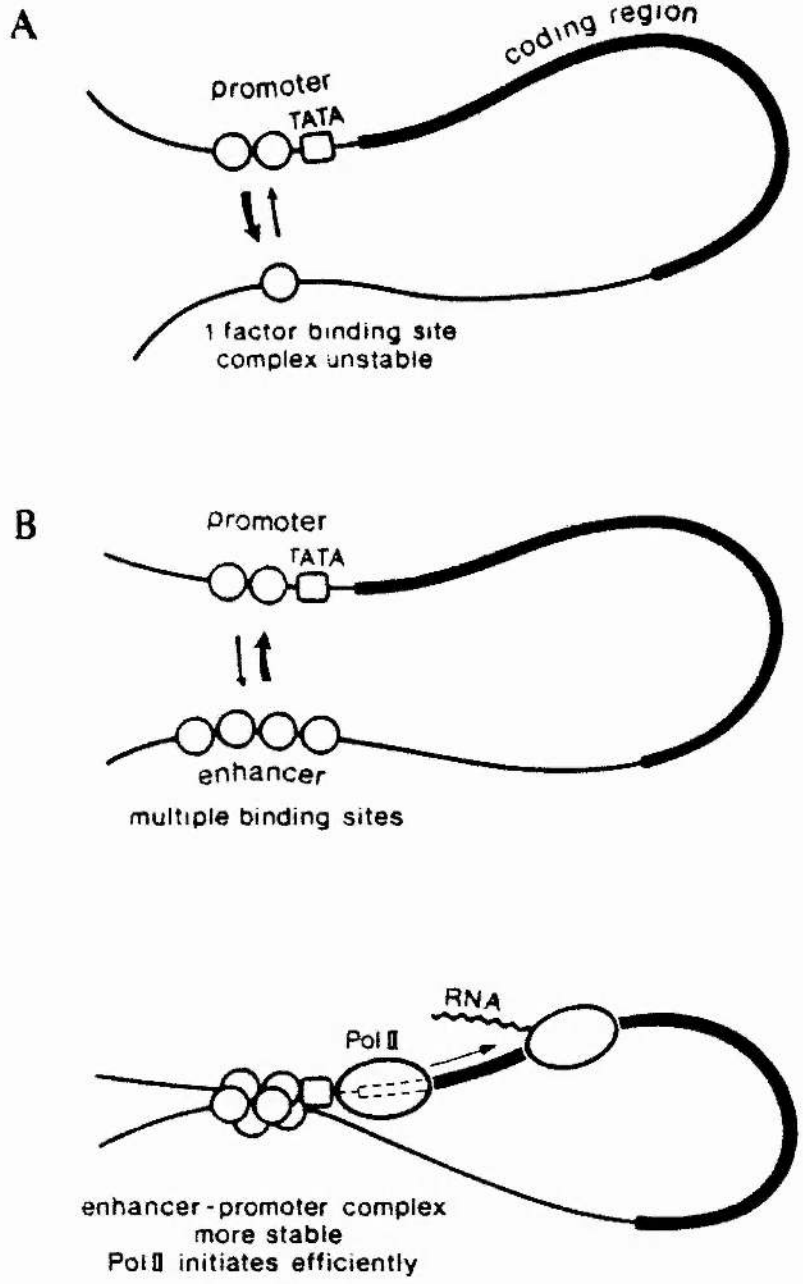


Figure 2D. Looping model of enhancer mechanism

A. A remote enhancer with a single binding site for a transcription factor cannot easily form a functional transcription complex, nor are there sufficient factors present in the promoter region to allow high transcriptional activity.

B. A larger number of factor binding sites are present in the enhancer which facilitates formation of a functional complex via protein–protein contacts (including the TATA box binding factor and RNA polymerase II) which leads to initiation of transcription. (Muller *et al.*, 1988).

DNA were required for cooperativity, but not if the binding sites were positioned on opposite sides of the helix. Indeed, the proposed intervening looped out structure resulting from protein–protein interactions has been visualised by electron microscopy (Griffith *et al.*, 1986). Whether this looping model for enhancer–promoter interaction actually exists remains to be determined but preliminary data on loop formation induced by the progesterone receptor lends some support to the idea (reviewed by Muller *et al.*, 1988).

Whether or not insertion of half helical turns into intervening DNA as a rule affects the activity of eukaryotic promoter–enhancer interactions is not clear, neither is the problem of long-range enhancer–promoter interactions since half helical turns often have little or no effect on eukaryotic transcriptional activity. This difference between prokaryotic and eukaryotic promoters could be attributable to the nature of the DNA binding proteins themselves, e.g., in eukaryotes it is feasible that two disinct domains on the transcription factor could be connected by long, flexible arms. Nevertheless, the looping model accommodates a method of activation over large distances and the orientation independence of enhancers, as does the preferential transcription of enhancer-proximal promoters versus distal promoters. There is no evidence to suggest that EBP1 acts via looping but the plethora of factors which have the ability to bind to the SV40 enhancer suggest that protein–protein interactions probably play a major role in transcription activation and one could easily envisage a situation whereby distantly bound proteins could interact with each other. In addition, it has been shown that EBP1 has a distinct, cleavable DNA binding domain (Figure). It is therefore possible that a component of the protein not involved in binding to DNA could be some sort of activation domain involved in direct binding to other proteins possibly via a negatively charged region (for review see Sigler, 1988). Transcription factor Sp1 has a region of the protein not associated with binding to DNA . This region has a very high glutamine content (Jones *et al.*, 1987) and it is feasible that these residues could be involved in protein–protein interactions. Similarly, the proline rich regions of the CTF/NF-I family of proteins (Santoro *et al.*, 1988) could provide the sites for interaction with other oppositely charged protein domains. Whether or not domain(s) of EBP1 are involved in such interactions will require detailed sequence information on the protein.

6. The interplay of cellular factors in transcriptional control and DNA replication

It is well established that eukaryotic cells can generate alternative messages from one gene using devices such as alternative promoters, terminators and splice sites. A factor has been described, CTF, which stimulated transcription from various viral and cellular promoters, and which can mediate, in part, some of the increased transcription caused by transforming growth factor-ß (reviewed by Short, 1988). During the purification of this factor, it became apparent that it was identical to the factor NF-I, shown previously to stimulate adenovirus DNA replication (Jones *et al.*, 1987). This NF-I, purified from HeLa cells, was able to bind to and stimulate DNA replication from the adenovirus origin of DNA replication and also bind to several promoters to activate transcription *in vitro.* However, preparations of the protein were inherently heterogeneous – repeated rounds of affinity chromatography failed to produce preparations with a single molecular mass. The reason for this heterogeneity has now been established by Santoro *et al.*, (1988) whereby they carried out a detailed molecular analysis of CTF/NF-I cDNA clones and demonstrated the presence of multiple mRNA species containing alternative coding regions, resulting from differential splicing. Expression of these cloned genes and analysis of there functions *in vivo* established that the individual gene products could bind to the GCCAAT recognition site common to both CTF and NF-I, and act as both promoter-selective transcriptional activators and as initiation factors for DNA replication. The functional differences between the three forms of CTF/NF-I isolated by Santoro and coworkers (1988) still remains somewhat obscure but it is conceivable that each form could have subtly different effects on transcription or DNA replication, perhaps mediated by their interaction, in different ways, with some other factor involved in these processes.

All three of the CTF/NF-I proteins have a C-terminal portion fairly similar to the C-terminal portion of the yeast RAP1 factor (Shore and Nasmyth, 1987) which binds to the silencer sequence in the *HMR* locus in yeast. This adds yet another degree of complexity to the CTF/NF-I story in that these silencer regions are the inverse of enhancers in that they repress transcription in both directions over long distances. RAP1 also influences the transmission of plasmids at mitosis, as well as binding to yeast telomeres. These findings lend support to the multifunctional nature of CTF/NF-I.

The association of transcription factors with origins of DNA replication may provide us with a mechanism for coordinate regulation of transcription and replication. It is possible that replication is

dependent on a template that requires activation by RNA polymerase II-mediated initiation of transcription, but this situation does not appear to be necessary for adenovirus, SV40 and polyoma DNA replication where RNA polymerase II activity is dispensable (reviewed by Santoro *et al.*, 1988). Alternatively, binding of the transcription factors to the template could directly alter the chromatin structure to facilitate DNA replication. A model favoured by Santoro and coworkers (1988) involves specific interaction between transcription factors bound to the DNA template and components of the replication machinery.

Does EBP1 have such a multifunctional role? There is no evidence to suggest such a role at present. However, it does appear that EBP1 belongs to a family of related DNA binding proteins, i.e., NF-kB and H2TF1/KBF1, all of which have similar molecular weights (Kawakami *et al.*, 1988; Singh *et al.*, 1988). To determine the relationship between these proteins will require a detailed molecular and functional analysis.

REFERENCES

Adams, R.L.P., J.T. Knowler and D.P. Leader. 1986.In, The biochemistry of the nucleic acids, 10th edition, Pub. by Chapman and Hall.

Adhya, S. and S. Garges. 1982. *Cell* **29:** 287–289.

Adhya, S. 1987. In, *Escherichia coli* and *Salmonella typhimurium:* Cellular and Molecular Biology, ed. by F.C. Neidhardt, ASM, Washighton, pp1503–1512.

Adhya, S. and A. Majumdar. 1987. In, RNA polymerase and regulation of transcription, eds., W.S. Reznikoff, R.R. Burgess, J.E. Dahlberg, C.A. Gross, M.T. Record and M.P. Wicken's, Elsevier Science Publishers, New York, pp129–135.

Aggarwal, A.K., D.W. Rodgers, M. Drottar, M. Ptashne and S.C. Harrison. 1988. *Science* **242:** 899–907.

Aiba, H. 1983. *Cell* **32:** 141–147.

Alberts, B. and G. Herrick. 1971. *Methods Enzymol.* **21:** 198–217.

Anderson, W.F., D.H. Ohlendorf, Y. Takeda and B.W. Matthews. 1981. *Nature* **290:** 754–758.

Anderson, J.A., M. Ptashne and S.C. Harrison. 1985. *Nature* **316:** 596–601.

Anderson, J.E., M. Ptashne and S.C. Harrison. 1987. *Nature* **326:** 846–852.

Angel, P., M. Imagawa, R. Chiu, B. Stein, R.J. Imbra, H.J. Rahmsdorf, C. Jonat, P. Herrlich and M. Karin. 1987. *Cell* **49:** 729–739.

Arcangioli, B. and B. Lescure. 1985.*EMBO J.* **4:**2627–2633.

Arndt-Jovin, D.J., T.M. Jovin, W. Bahr, A. Frischauf and M. Marquardt. 1975. *Eur. J. Biochem.* **54:** 411–418.

Ausubel, F.M. 1984. *Cell* **37:** 5–6.

Baeuerle, P.A. and D. Baltimore. 1988. *Cell* **53:** 211–217.

Baldwin, A.S., Jr. and P.A. Sharp. 1987. *Mol. and Cell. Biol.* **7:** 305–313.

Baldwin, A.S., Jr. and P.A. Sharp. 1988. *Proc. Natl. Acad. Sci. USA* **85:** 723–727.

Banerji, J., S. Rusconi and W. Schaffner. 1981. *Cell* **27:** 299–308.

Banerji, J., L. Olsen and W. Schaffner. 1983. *Cell* **33:** 729–740.

Barbier, B., M. Charlier and J.C. Maurizot. 1984. *Biochemistry* **23:** 2933–2939.

Barklay, M.D. and S. Bourgeois. 1978. In. The Operon, ed. by J.H. Miller and W.S. Reznikoff: CSH

Laboratory, Cold Spring Harbour, New York, pp 177–220.

Barrett, P., L. Clark and R.T. Hay. 1987. *Nucl. Acids Res.* **15:** 2719–2735.

Barrera-Saldana, H., K. Takahashi, M. Vigneron, A. Wildeman, I. Davidson and P. Chambon. 1985. *EMBO J.* **4:** 3839–3849.

Bass, S., P. Sugiono, D.N. Arvidson, R.P. Gunsalus and P. Youderian. 1987. *Genes & Dev.* **1:** 565–572.

Bass, S., V. Sorrells and P. Youderian. 1988. *Science* **242:** 240–245.

Bellard, M., F. Gannon and P. Chambon. 1977. *CSH Symp Quant. Biol.* **42:** 779.

Bender, A. and G.F. Sprague Jr. 1987. *Cell* ***50:*** 681–691.

Benezra, R., C.R. Cantor and R. Axel. 1986. *Cell* **44:** 697–704.

Benoist, C. and P. Chambon. 1981. *Nature* **290:** 304–315.

Benyon, J., M. Cannon, V. Buchanan-Wollaston and F. Cannon. 1983. *Cell* **34:** 665–671.

Berk, A.J. 1986. *Ann. Rev. Genet.* **20:** 45–79.

Berman, M.L. and A. Landy. 1979. *Proc. Natl. Acad. Sci. USA* **76:** 4303–4307.

Bird, A.P. 1986. *Nature* **321:** 209–213.

Birnboim, H.C. and Doly, J. 1979. *Nucl. Acids Res.* **7:** 1513–1523.

Bohmann, D., T.J. Bos, A. Admon, T. Nishimure, P.K. Vogt and R. Tjian. 1987. *Science* **238:** 1386–1392.

Bohnlein, E., J.W. Lowenthal, M. Siekevitz, D.W. Ballard, R.B. Franza and W.C. Greene. 1988. *Cell* **53:** 827–836.

Boroweic, J.A. and J.D. Gralla. 1986. *Biochemistry* **25:** 5051–5057.

Boroweic, J.A. and J. Hurwitz. 1988. *EMBO J.* **7:** 3149–3158.

Bradbury, E.M., N MacLean and H.R. Matthews. 1981. In, DNA, chromatin and chromosomes. Blackwell Scientific Publications, Oxford.

Bradford, M.M. 1976. *Anal. Biochem.* **72:** 248–254.

Brand, A.H., L. Breeden, J. Abraham, R. Sternglanz and K. Nasmyth. 1985. *Cell* **41:** 41–48.

Brand, A.H., G. Micklem and K. Nasmyth. 1987. *Cell* **51:** 709–719.

Breathnach, R. and P. Chambon. 1981. *Ann. Rev. Biochem.* **50:** 349–383.

Brent, R. 1985. *Cell* **42:** 3–4.

Brent, R. and M. Ptashne. 1984. *Nature* **312:** 612–615.

Brent, R. and M. Ptashne. 1985. *Cell* **43:** 729–736.

Briggs, M.R., J.T. Kadonaga, S.P. Bell and R. Tjian. 1986. *Science* **234:** 47–52.

Brown, R.S., C. Sander and P. Argos. 1985. *FEBS Lett.* **186:** 271–274.

Brown, M.E., J. Amin, P. Schiller, R. Voellmy and W.A. Scott. 1988. *J. Mol. Biol.* **203:** 107–117.

Buc, H. 1987. *Nucl Acids and Mol. Biol.* vol. **1:** ed. by F. Eckstein and D.M.J. Lilley, Springer-Verlag, Berlin, Heidelberg, pp186–195.

Bujard, H., M. Brunner, U. Deuschle, W. Kammerer and R. Knaus. 1986. 16th Steenbock Symp.,

Madison. Burgess, R.R., A.A. Travers, J.J. Dunn and E.K.F. Bautz. 1969. *Nature* **221:** 43–46.

Burton, Z.F., R.R. Burgess, J. Lin, D. Morse, S. Holder and C.A. Gross. 1981. *Nucl. Acids Res.* **9:** 2889–2903.

Burton, Z.F., L.G. Ortolan and J. Greenblat. 1986. *EMBO J.* **5:** 2923–2930.

Burton, Z.F., M. Killeen, M. Sopta, L.G. Ortolan and J. Greenblatt. 1988. *Mol. Cell. Biol.* **8:** 1602–1613.

Busby, S.J.W. 1986. In Regulation of gene expression, **39:** ed. by I.R. Booth and C.F. Higgins, Cambridge University Press, pp51–77.

Bushman, F.D., J.E. Anderson, S.C. Harrison and M. Ptashne. 1985. *Nature* **316:** 651–653.

Bushman, F.D. and M. Ptashne. 1986. *Proc. Natl. Acad. Sci. USA* **83:** 9353–9357.

Buttin, G. 1963. *J. Mol. Biol.* **7:**183–205.

Challberg, M.D. and T.J. Kelly. 1979. *Proc. Natl. Acad. Sci. USA* **76:** 655–659.

Chamberlin, M.J. 1982. *The Enzemes* **15:** 61–86.

Chandler, V.L., B.A. Maler and K.R. Yamomoto. 1983. *Cell* **33:** 489–499.

Chen, W. and K. Struhl. 1985. *EMBO J.* **4:** 3273–3280.

Chen, W., S. Tabor and K. Struhl. 1987. *Cell* **50:** 1047–1055.

Chiu, R., M. Imagawa, R.J. Imbra, J.R. Bockoven and M. Karin. 1987. *Nature* **329:** 648–651.

Chochran, M.D. and C. Weissman. 1982. *EMBO J.* ***3:*** 2453–2457.

Chodosh, L.A., R.W. Carthew and P.A. Sharp. 1986. *Mol. Cell. Biol.* **6:** 4723–4733.

Chodosh, L.A., J. Olesen, S. Hahn, A.S. Baldwin, L. Guarente and P.A. Sharp. 1988. *Cell* **53:** 25–35.

Chou, P.Y. and G.D. Fasman. 1978. *Ann. Rev. Biochem.* **47:** 251–276.

Coen, D.M., S.P. Weinheimer and S.L. McKnight. 1986. *Science* **234:** 53–59.

Concino, M.F., R.F. Lee, J.P. Merryweather and R. Weinmann. 1984. *Nucl. Acids Res.* **12:** 7423–7433.

Costlow, N. and J.T. Lis. 1984. *Mol. Cell. Biol.* **4:** 1853–1863.

Das, A.J., Urbanowski, H. Weissbach, J. Nestor and C. Yanofsky. 1983. *Proc. Natl. Acad. Sci. USA* **80:** 2879–2883

Davidson, I., C. Fromental, P. Augereau, A. Wildeman, M. Zenke and P. Chambon. 1986. *Nature* **323:** 544–548.

Davidson, I., J.H. Xiao, R. Rosales, A. Staub and P. Chambon. 1988. *Cell* **54:** 931–942.

Deb, S., A.L. DeLucia, C. P. Baur, A. Koff and P. Tegtmeyer. 1986. *Mol. Cell. Biol.* **6:** 1663–1670.

DeGrazia, H., S. Abhirman and R.M. Wartell. 1985. *Nucl. Acids Res.* **13:** 7483–7498.

DePamphilis, M.L. 1988. *Cell* **52:** 635–638.

Deuschle, U., W. Kammerer, R. Gentz and H. Bujard. 1986. *EMBO J.* **5:** 2987–2994.

deVilliers, J., W. Schaffner, C. Tyndall, S. Lupton and R. Kamen. 1984. *Nature* **312:** 242–246.

deVries, E., W. van Dreil, S.J.L. van den Heuvel and P.C. van der Vliet. 1987. *EMBO J.* **6:** 161–168.

Dickerson, R.E. 1983. *Sci. Am.* **244:** 86–102.

Dickson, R.C., J. Abelson, M.W. Barnes and W.S. Renzikoff. 1975. *Science* **187:** 27–35.

Doi, R.H. and F.F. Wang. 1986. *Microbiol. Rev.* **50:** 227–243.

Drew, H.R. and A.A. Travers. 1984. *Cell* **37:** 491–502.

Drew, H.R. and A.A. Travers. 1985a. *Mol. Biol.* **186:** 773–790.

Drew, H.R. and A.A. Travers. 1985b. *Nucl. Acids Res.* **13:** 4445–4467.

Drew, H.R., J.R. Weeks and A.A. Travers. 1986. *EMBO J.* **4:** 1025–1032.

Dreyfus, M., D. Kotlarz and S. Busby. 1985. *J. Mol. Biol.* **182:** 411–417.

Drouin, J. 1980. *J. Mol. Biol.* **140:** 15–34.

Dunn, T., S. Hahn, S. Ogden and R.F. Schleif. 1984. *Proc. Natl. Acad. Sci. USA* **83:** 3654–3658.

Dynan, W.S. and R. Tjian. 1983. *Cell* **32:** 669–680.

Dynan, W.S. and R. Tjian. 1985. *Nature* **316:** 774–778.

Elgy, J.M., N. Miyamoto, V. Moncollin and P. Chambon. 1984. *EMBO J.* **3:** 2363–2371.

Elliot, T. and E.P. Geiduschek. 1984. *Cell* **36:** 211–219.

Emerson, B., C.D. Lewis and G. Felsenfeld. 1985. *Cell* **41:** 21–30.

Everett, R.D., D. Baty and P. Chambon. 1983. *Nucl. Acids Res.* **11:** 2447–2464.

Fairall, L., D. Rhodes and A. Klug. 1986. *J. Mol. Biol.* **192:** 577–591.

Falkner, F.G., R. Mocikat and H.G. Zachau. 1986. *Nucl. Acids Res.* **14:** 8819–8827.

Felsenfeld, G. and J.D. McGhee. 1986. *Cell* **44:**375–377.

Fire, A., M. Samuels and P.A. Sharp. 1984. *J. Biol. Chem.* **259:** 2509–2516.

Fletcher, C., N. Heintz and R.G. Roeder. 1987. *Cell* **51:** 773–781.

Forsburg, S. and L. Guarente. 1987. *Mol. Cell. Biol.* **8:** 647–654.

Fried, M. and D.M. Crothers. 1981. *Nucl. Acids Res.* **9:** 6505–6525.

Fried, M.G. and D.M. Crothers. 1984. *J. Mol. Biol.* **172:** 263–282.

Fromental, C., K. Masamoto, H. Nomiyama and P. Chambon. 1988. *Cell* **54:** 943–953.

Fromm, M. and P. Berg. 1982. *J. Mol. Appl. Genet.* **1:** 457–481.

Galas, D.J. and A. Schmitz. 1978. *Nucl. Acids Res.* **5:** 3157–3170.

Garner, M.M. and A. Revzin. 1981. *Nucl. Acids Res.* **9:** 3047–3060.

Garoff, H. and W. Ansgore. 1981. *Anal. Biochem.* **115:** 450–457.

Gellert, M. 1981. *Ann. Rev. Biochem.* **50:** 879–910.

Gerster, T. P. Mathias, M. Thali, J. Jiricny and W. Schaffner. 1987. *EMBO J.* **6:** 1323–1330.

Ghosh, P.K. and P. Lebowitz. 1981. *J. Virol.* **40:** 224–240.

Gidoni, D., J.T. Kadonaga, H. Barrera-Saldana, K. Takahashi, P. Chambon and R. Tjian. 1985.*Science* **230:** 511–517.

Gilbert, W. 1976. In, RNA polymerase, ed. by R. Losick and M. Chamberlin, CSH Laboratory, New York, pp193–205.

Gill, G. and M. Ptashne. 1987. *Cell* **51:** 121–126.

Gill, G. and M. Ptashne. 1988. *Nature* **334:** 721–724.

Gillies, S.D., S.L. Morrison, V.T. Oi and S. Tonegawa. 1983. *Cell* **33:** 717–728.

Gilman, M.Z., R.N. Wilson and R.A. Weinberg. 1986. *Mol. Cell Biol.* **6:** 4305–4316.

Giniger, E., S.M. Varnum and M, Ptashne. 1985. *Cell* **40:** 767–774.

Golub, E.I. 1988. *Nucl. Acids Res.* **16:** 1641.

Goodbourn, S., K. Zinn and T. Maniatis. 1985. *Cell* **41:** 509–520.

Goodbourn, S., H. Burstein and T. Maniatis. 1986. *Cell* **45:** 601–610.

Gorman, C.M., L.F. Moffat and B.H. Howard. 1982. *Mol. Cell. Biol.* **2:**1044–1051.

Graham, F. and A. van der Eb. 1973. *Virology* **52:** 456–457.

Griffith, J., A. Hochschild and M. Ptashne. 1986. *Nature* **322:** 750–752.

Gross, D.S. and W.T. Garrard. 1987. *TIBS* **12:** 293–297.

Grosscheldl, R. and M.L. Birnsteil. 1980. *Proc. Natl. Acad. Sci. USA* **77:** 7102–7106.

Grossman, A.D., J.W. Erickson and C.A. Gross. 1984. *Cell* **38:** 383–390.

Gruss, P., R. Dhar and G. Khoury. 1981. *Proc. Natl. Acad. Sci. USA* **78:** 943–947.

Guarente, L. 1984. *Cell* **36:** 799–800.

Guarente, L and E. Hoar. 1984. *Proc. Natl. Acad. Sci. USA* **81:** 7860–7864.

Guarente, L., B. Lalonde, P. Gifford and E. Alani. 1984. *Cell* **36:** 503–511.

Guarente, L. 1988. *Cell* **52:** 303–305.

Hahn, S. and L. Guarente. 1988. *Science* **240:** 317–321.

Hall, M.N. and A.D. Johnson. 1987. *Science* **237:** 1007–1012.

Hamer, D.H. and G. Khoury. 1983. In, Enhancers and Eukaryotic Gene Expression, CSH Current Communications in Molecular Biology, ed., by Y. Gluzman and T. Shenk, pp1–15.

Hanahan, D. 1983. *J. Mol. Biol.* **166:** 557-580.

Harley, C.B. and R.P. Reynolds. 1987. *Nucl. Acids Res.* **15:** 2343–2361.

Harshman, K.D., W.S. Moye-Rowley and C.S. Parker. 1988. *Cell* **53:** 321–330.

Hawley, D.K. and W.R. McClure. 1982. *J. Mol Biol.* **157:**493–525.

Hawley, D.K. and W.R. McClure. 1983. *Nucl. Acids Res.* **11:** 2237–2255.

Hay, R.T. and M.L. DePamphilis. 1982. *Cell* **28:** 767–779.

Heine, J.W., R.W. Honess, E. Cassai and B. Roizman. 1974. *J. Virol.* **14:** 640–651.

Hentschel, C., J.C. Irminger, P. Bucher and M.L. Birnsteil. 1980. *Nature* **285:** 147–151.

Herr, W. and Y, Gluzman. 1985. *Nature* **313:** 711–714.

Herr, W. and J. Clarke. 1986. *Cell* **45:** 461–470.

Herskowitz, I. and Y. Oshima. 1981. In, Molecular Biology of yeast *Saccharomyces*. Strathern, ed. by

J.N., E.W. Jones and J.R. Broach, CSH Laboratory, New York, pp181–209.

Hill, D.E., I.A. Hope, J.P. Macke and K. Struhl. 1986. *Science* **234:** 451–457.

Hochschild, A., N. Irwin and M. Ptashne. 1983. *Cell* **32:** 319–325.

Hochschild, A. and M. Ptashne. 1986. *Cell* **44: 925–933.**

Hochschild, A., J. Douhan III and M. Ptashne. 1986. *Cell* **47:** 807–816.

Hochschild, A. and M. Ptashne. 1988. *Nature* **336:** 353–357.

Holland, M.J. and J.P. Holland. 1978. *Biochem. J.* **17:** 4900–4907.

Hope, I.A. and K. Struhl. 1985. *Cell* **43:** 177–188.

Hope, I.A. and K. Struhl. 1986. *Cell* **46:** 885–894.

Hope, I.A. and K. Struhl. 1987. *EMBO J.* **6:** 2781–2784.

Hope, I.A., S. Mahadevan and K. Struhl. 1988. *Nature* **333:** 635–640.

Hopper, J., J. Broach and L. Rowe. 1978. *Proc. Natl. Acad. Sci. USA* **75:** 2878–2882.

Hsieh, W.T., P.A. Whitson, K.S. Matthews and R.D. Wells. 1987. *J. Biol. Chem.* **262:** 14583–14591.

Hurstel, S., M. Granger-Schnarr, M. Daune and M. Schnarr. 1986. *EMBO J.* **5:** 793–798.

Hurstel, S,. M. Granger-Schnarr and M. Schnarr. 1988. *EMBO J.* **7:** 269–275.

Imagawa, M., R. Chiu and M. Karin. 1987. *Cell* **51:** 251–260.

Irani, M., L. Orosz and S. Adhya. 1983. *Cell* **32:** 783–788.

Israel, A., A. Kimura, M. Kieran, O. Yano, J. Kanellopoulos, O. Le Bail and P. Chambon. 1987. *Proc. Natl. Acad. Sci. USA* **84:** 2653–2657.

Jacob, F. and J. Monod. 1961. *J. Mol. Biol.* **3:** 318–356.

Jakobovits, E.B., S. Bratosin and Y. Aloni. 1980. *Nature* **285:** 263–265.

Joachimiak, A.J., R.L. Kelley, P.R. Gunsalus, C. Yanofsky and P.B. Sigler. 1983a. *Proc. Natl. Acad. Sci. USA* **80:** 668–672.

Joachimiak, A.J., R.W. Schevitz, R.L. Kelley, C. Yanofsky and P.B. Sigler. 1983b.*J. Biol. Chem.* **258:**12641–12643.

Johnson, A., B.J. Meyer and M. Ptashne. 1979.*Proc. Natl. Acad. Sci. USA* **76:** 5061–5065.

Johnson, A.D., A.R. Poteete, G. Lauer, R.T. Sauer, G.K. Ackers and M. Ptashne. 1981. *Nature* **294:** 217–233.

Johnson, A.D. and I Herskowitz. 1985. *Cell* **42:** 237–247.

Johnson, P.F., W. H. Landschulz, B.J. Graves and S.L. McKnight. 1987. *Genes & Dev.* **1:** 133–146.

Johnson, S.A., J.M. Salmeron Jr. and S.S. Dincher. 1987. *Cell* **50:** 143–146.

Jones, E.W. and G.R. Fink. 1982. In the molecular biology of the yeast *Saccharomyces* : metabolism and gene expresion, ed. by J.N. Strathearn, E.W. Jones and J.R. Broach,CSH Laboratory, New York, pp181–299.

Jones, K.A., J.T. Kadonaga, P.J. Rosenfeld, T.J. Kelly and R. Tjian. 1987. *Cell* **48:** 79–89.

Jones, N.C., P.W.J. Rigby and E.B. Ziff. 1988. *Genes & Dev.* **2:** 267–281.

Jones, R.H., S. Moreno, P. Nurse and N.C. Jones. 1988. *Cell* **53:** 659–667.

Jordan, S.R. and C.O. Pabo. 1988. *Science* **242:** 893–899.

Joyce, C.M. and N.D.F. Grindley. 1983. *Proc. Natl. Acad. Sci. USA* **80:** 1830–1834.

Kadonaga, J.T. and R. Tjian. 1986. *Proc. Natl. Acad. Sci. USA* **83:** 5889–5893.

Kadonaga, J.T., K. Jones and R. Tjian. 1986. *Trends Biochem.* **11:** 20–23.

Kadonaga, J.T., K.R. Carner, F.R. Masiarz and R.Tjian. 1987. *Cell* **51:** 1079–1090.

Kakidani, H. and M. Ptashne. 1988. *Cell* **52:** 161–167.

Kammerer, W., U. Deuschle, R. Gentz and H. Bujard. 1986. *EMBO J.* **5:** 2995–3000.

Karin, M., A. Haslinger, H. Holtgreve, R.I. Richards, P. Krauter, H.M. Westphal and M. Beato. 1984. *Nature* **308:** 513–519.

Kaufman, J.D., G. Valandra, G. Rodriguez, G. Bushar, C. Giri and M.D. Norcross. 1987. *Mol. Cell. Biol.* **7:** 3759–3766.

Kawakami, K., C. Scheidereit and R.G. Roeder. 1988. *Proc. Natl. Acad. Sci. USA* **85:** 4700–4704.

Keegan, L., G. Gill and M. Ptashne. 1986. *Science* **231:** 699–704.

Keleher, C.A., C. Goutte and A.D. Johnson. 1988. *Cell* **53:** 927–936.

Kelley, R.L. and C. Yanofsky. 1982. *Proc. Natl. Acad. Sci. USA* **79:** 3120–3124.

Keng, T. and L. Guarente. 1987. *Proc. Natl. Acad. Sci. USA* **84:** 9113–9117.

Kirchhausen, T., J.C. Wang and S.C. Harrison. 1985. *Cell* **41:** 933–943.

Klar, A.J.S., J.N. Strathearn and J.B. Nicks. 1984. . In, microbial development, ed. by R. Losick and L. Shapiro, CSH Laboratory, New York, pp151–195.

Klenow, H., K. Overgaarg-Hansen and S.A. Patkar. 1971. *Eur. J. Biochem.* **22:** 371–381.

Kmiec, E.G. and A. Worcel. 1985. *Cell* **41:** 945–953.

Koudelka, G.B., S.C. Harrison and M. Ptashne. 1987. *Nature* **326:** 886–891.

Koudelka, G.B., P. Harbury, S.C. Harrison and M. Ptashne. 1988. *Proc. Natl. Acad. Sci. USA* **85:** 4633–4637.

Kramer, H., M. Amouyal, A. Nordheim and B. Muller-Hill. 1988. *EMBO J.* **7:** 547–556.

Kronstad, J.W., J.A. Holly and V.L. MacKay. 1987. *Cell* **50:** 369–377.

Kutsu, S., K. Sei and J. Keener. 1986. In, Regulation of Gene Expression, ed. by I. Booth and C.J. Higgins, Cambridge University Press **39:** 139–154.

Kuwabara, M.D. and D.S. Sigman. 1987. *Biochemistry* **26:** 7234–7238.

Laimins, L.A., G. Khoury, C. Gorman, B. Howard and P. Gruss. 1982. *Proc. Natl. Acad. Sci USA* **79:** 6453–6457.

Landolfi, N.F., J.D. Carpa and P.W. Tucker. 1986. *Nature* **323:** 548–551.

Lee, W., A. Haslinger, M. Karin and R. Tjian. 1987a. *Nature* **325:** 368–372.

Lee, W., P. Mitchell and R. Tjian. 1987b. *Cell* **49:** 741–752.

Lenardo, M., J.W, Pierce and D. Baltimore. 1987. *Science* **236:** 1573–1577.

Levy, J.A., C. Cheng-Mayer, D. Dina and P.A. Luciw. 1985. *Science* **232:** 998–1001.

Lewin, B. 1985. Genes II, Peb. by John Wiley and Sons.

Lewis, M., A. Jeffrey, J. Wang, R. Ladner, M. Ptashne and C.O. Pabo. 1983. *CSH Symp. Quant. Biol.* **47:** 435–440.

Little, J.W. and D.W. Mount. 1982. *Cell* **29:** 11–22.

Little, J.W. and S.A. Hill. 1985. *Proc. Natl. Acad. Sci. USA* **82:** 2301–2305.

Losick, R. and J. Pero. 1981. *Cell* **25:** 582–584.

Losick, R. and P. Youngman. 1984. In. Microbial Development, ed. by R. Losick and L. Shapiro, CSH Laboratory, New York, pp63–88.

Ma, J. and M. Ptashne. 1987a. *Cell* **48:** 847–853.

Ma, J. and M. Ptashne. 1987b. *Cell* **50:** 137–142.

Majumdar, A. and S. Adhya. 1984. *Proc. Natl. Acad. Sci. USA* **81:** 6100–6104.

Majors, J. 1975. *Proc. Natl. Acad. Sci. USA* **72:** 4394–4398.

Maki, Y., T.J. Bos, C. Davis, M. Srarbuck and P.K. Vogt. 1987. *Proc. Natl. Acad. Sci. USA* **84:** 2848–2852.

Maniatis, T., E.F. Fritsch and J. Sambrook. 1982. Molecular cloning: a laboratory manual. CSH Laboratory, New York.

Manly, S.P., G.N. Bennett and K.S. Matthews. 1983. *Proc. Natl. Acad. Sci. USA* **80:** 6219–6223.

Mathis, D.J. and P. Chambon. 1981. *Nature* **290:** 310–315.

Maxam, A.M. and W. Gilbert. 1977. *Proc. Natl. Acad. Sci. USA* **74:** 560–564.

Maxam, A.M. and W. Gilbert. 1980. *Methods in Enzymology* **65:** 499–560.

May, E., O. Francis, M. Ernoult-Lange, M. Zenke and P. Chambon. 1987. *Nucl. Acids Res.* **15:** 2445–2461.

McClure, W.R. 1980. *Proc. Natl. Acad. Sci. USA* **72:** 5634–5638.

McClure, W.R. 1985. *Ann. Rev. Biochem.* **54:** 171–204.

McKay, D.B. and T.A. Steitz. 1982. *Nature* **290:** 744.

McKay, D.B., I.T. Weber and T.A. Steitz. 1982. *J. Biol. Chem.* **257:** 9518–9524.

McKnight, S.L. and R.C. Kingsbury. 1982. *Science* **217:** 316–324.

McKnight, S. and R. Tjian. 1986. *Cell* **46:** 795–805.

Meisenberger, O., H. Henmann and I. Pilz. 1980. *FEBS Lett.* **122:** 117–120.

Messing, J. 1979. *Recomb. DNA Tech.* Bull. **2:** 43–48.

Meyer, B., R. Maurer and M. Ptashne. 1980. *J. Mol. Biol.* **139:** 163–194.

MIller, J.H. and Reznikoff. 1980, eds., The Operon, CHS Laboratory, New York.

Miller, J., A.D. McLachlan and A. Klug. 1985. *EMBO J.* **4:** 1609–1614.

Mitchell, P., C. Wang and R. Tjian. 1987. *Cell* **50:** 847–861.

Moncollin, V., N.G. Miyamoto, M.X. Zheng and J.M. Elgy. 1986. *EMBO J.* **5:** 2577–2584.

Monod, J. 1947. *Growth* **11:** 223–289.

Moreau, P., R. Hen, B. Wasylyk, R. Everett, M.P. Gaub and P. Chambon. 1981. *Nucl. Acids Res.* **9:** 6047–6068.

Morse, R.H. and R.T. Simpson. 1988. *Cell* **54:** 285–287.

Muller, M.M., T. Gerster and W. Schaffner. 1988. *Eur. J. Biochem.* **176:** 485–495.

Myers, R. M., D.C. Rio, A.K. Robbins and R. Tjian. 1981. *Cell* **25:** 373–384.

Myers, R.M., K. Tilly and T. Maniatis. 1986. *Science* **232:** 613–618.

Nabel, G. and D. Baltimore. 1987. *Nature* **326:** 711–713.

Nakanishi, S., S. Adhya, M. Gottesman and I. Pastan. 1973. *Proc. Natl. Acad. Sci. USA* **70:** 334–338.

Nelson, H.C.M., J.T. Finch, B.F. Luisi and A. Klug. 1987. *Nature* **330:** 221–226.

Neuberger, M.S. 1983. *EMBO J.* **2:** 1373–1378.

Nomiyama, H., C. Fromental, J.H. Xiao and P. Chambon. 1987. *Proc. Natl. Acad. Sci. USA* **84:** 7881–7885.

Oakley, B.R., D.R. Kirsch and N.R. Morris. 1980. *Anal. Biochem.* **105:** 361–363.

Ogata, R.T. and W. Gilbert. 1977. *Proc. Natl. Acad. Sci. USA* **74:** 4973–4976.

Ogata, R.T. and W. Gilbert. 1978. *Proc. Natl. Acad. Sci. USA* **75:** 5851–5854.

Ogden, J., C. Stanway, S. Kim, J. Mellor, A.J. Kingsman and S.M. Kingsman. 1986. *Mol. and Cell.Biol.* **6:** 4335–4343.

Ohlendorf, D.H., W.F. Anderson, R.G. Fisher, Y. Takeda and B.W. Matthews, 1982. *Nature* **298:** 718–723.

Ohlsson, H. and T. Edlund. 1986. *Cell* **45:** 35–44.

Olesen, J., S. Hahn and L. Guarente. 1987. *Cell* **51:** 953–961.

Ondek, B., A. Shepard and W. Herr. 1987. *EMBO J.* **6:** 1017–1025.

Oshima, Y. 1982. In, Molecular biology of the yeast *Saccharomyces cerevisiae:* metabolism and gene expression, ed. by J.N. Strathearn, E.W. Jones and J.R. Broach, CSH Laboratory, New York pp159–180.

Otwinowski, Z., R.W. Schevitz, R-G. Zhang, C.L. Lawson, A. Joachimiak, R.Q. Marmorstein, B.F. Luisi and P.B. Sigler. 1988. *Nature* **335:** 321–329.

Pabo, C.O. and M. Lewis. 1982. *Nature* **298:** 443–447.

Pabo, C.O. and R.T. Sauer. 1984. *Ann. Rev. Biochem.* **53:** 293–321.

Parker, C.S. and J. Topal. 1984. *Cell* **36:** 357–369.

Paulmier, N., M. Yaniv, B. von Wilcken-Bergmann and B. Muller-Hill. *E. coli. EMBO J.* **6:** 3539–3542.

Perez-Mutul, J., M. Macchi and B. Wasylyk. 1988. *Nucl. Acids Res.* 6085–6096.

Pfeifer, K., B. Arcangioli and L. Guarente. 1987a. *Cell* **49:** 9–18.

Pfeifer, K., T. Prezant and L. Guarente. 1987b. *Cell* **49:** 19–27.

Pinkham, J. and L. Guarente. 1985. *Mol .Cell .Biol* . **12:** 3410–3416.

Pruijn, G.J.M., W. vad Driel and P.C. van der Vliet. 1986. *Nature* **322:** 656–659.

Ptashne, M, A. Jeffrey, A.D. Johnson, R. Maurer, B.J. Meyer, C.O.I Pabo, T.M. Roberts and R.T. Sauer. 1980. *Cell* **19:** 1–11.

Ptashne, M. 1986a. In, A genetic switch,Blackwell Scientific Publications and Cell Press.

Ptashne, M. 1986b. *Nature* **322:** 697–701.

Ptashne, M. 1986c. *Nature* **322:** 697–701.

Reinberg, D., M. Horikoshi and R.G. Roeder. 1987. *J. Biol. Chem.* **262:** 3322–3330.

Reinberg, D. and R.G. Roeder. 1987. *J. Biol. Chem.* **262:** 3310–3321.

Rhodes, D. and A. Klug. 1986. *Cell* **46:** 123–132.

Ricchetti, M., W. Metzger and H. Heumann. 1986. *Proc. Natl. Acad. Sci. USA* **85:** 4610–4614.

Richet, E., P. Abcarian and H,A. Nash. 1986. *Cell* **46:** 1011–1021.

Roberts, J. and R. Devoret. 1983. In Lambda II, R.W. Hendrix, ed. by J.W. Roberts, F.W. Stahl and R. Weinberg, CSH Laboratory, New York, pp 123–145.

Rosenberg, M. and D. Court. 1979. *Ann. Rev. Genetics* **13:** 319–353.

Rosenfeld, P.J. and T.J. Kelly. 1986. *J. Biol. Chem.* **261:** 1398–1408.

Ruben, S., H. Poteat, T-H. Tan, K. Kawakami, R. Roeder, W. Haseltine and C.A. Rosen. 1988. *Science* **241:** 89–92.

Ruden, D.M., J. Ma and M. Ptashne. 1988. *Proc. Natl. Acad. Sci. USA* **85:** 4262–4266.

Sanger, F., S. Nicklen and A.R. Coulson. 1977. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5476.

Santoro, C., N. Mermod, P.C. Andrews and R. Tjian. 1988. *Nature* **334:** 218–224.

Saragosti, S., G. Moyne and M. Yaniv. 1980. *Cell* **20:** 65–72.

Sauer, R.T., C.O. Pabo, B.J. Meyer, M. Ptashne and K.C. Backman. 1979. *Nature* **279:** 396–400.

Sauer, R.T., R.P. Yocum, R.F. Doolittle, M. Lewis and C.O.Pabo. 1982. *Nature* **298:** 447–451.

Sauer, R.T., D.L. Smith and A.D. Johnson. 1988. *Genes & Dev.* **2:** 807–816.

Sawadogo, M. and R.G. Roeder. 1984. *J. Biol. Chem.* **259:** 5321–5326.

Sawadogo, M. and R.G. Roeder. 1985. *Cell* **43:** 165–175.

Schaffner, W. 1986. In, Eukaryotic Transcription – The role of *cis*- and *trans*- acting elements in initiation, CSH Current Communications in Molecular Biology, pp1–18.

Schevitz, R.W., Z. Otwinowski, A. Joachimiak, C.L. Lawson and P.B. Sigler. 1985. *Nature* **317:** 782–786.

Schiller, P., J. Amin, J. Aranthan, M.E. Brown, W.A. Scott and R. Voellmy. 1988. *J. Mol. Biol.* **203:** 97–105.

Schirm, S., J. Jiricny and W. Schaffner. 1987. *Genes & Dev.* **1:** 65–74.

Seed, B. 1982. *Nucl. Acids Res.* **10:** 1799–1810.

Sen, R. and D. Baltimore. 1986a. *Cell.* **46:** 705–716.

Sen, R. and D. Baltimore. 1986b. *Cell.* **47:** 921–928.

Sentenac, A. 1985. *CRC Crit. Rev. Biochem.* **18:** 31–90.

Serfling, E., A. Lubbe, K. Dorsch-Hasler and W. Schaffner. 1985. *EMBO J.* **4:** 3851–3859.

Singer, B. 1975. In, Progression in Nucleic Acid Research and Molecular Biology, ed. by W. Cohn, Academic Press, Vol. **15:** 219–284.

Shaw, W. 1975. *Methods Enzymol.* **53:** 737–754.

Shore, D. and K. Nasmyth. 1987. *Cell* **51:** 721–732.

Shore, D., D.J. Stillman, A.H. Brand and K.A. Nasmyth. 1987. *EMBO J.* **6:** 461–467.

Short, N.J. 1988. *Nature* **334:** 192–193.

Shuey, D.J. and C.S. Parker. 1986. *J. Biol. Chem.* **17:** 7934–7940.

Siebenlist, U. and W. Gilbert. 1980. *Proc. Natl. Acad. Sci. USA* **77:** 122–126.

Siebenlist, U., R.B. Simpson and W. Gilbert. 1980. *Cell* **20:** 269–281.

Sigman, D., A. Spassky, S. Rimsky and H. Buc. 1985. *Biopolymers* **24:** 183–197.

Singh, H., R. Sen, D. Baltimore and P. Sharp. 1986. *Nature* **319:** 154–157.

Singh, H., J.H. LeBowitz, A.S. Baldwin, Jr. and P.A. Sharp. 1988. *Cell* **52:** 415–423.

Somasekhar, M.B. and J.E. Mertz. 1985. *J. Virol.* **56:** 1002–1013.

Sorger, P.K. and H.R.B. Pelham. 1987. *EMBO J.* **6:** 3035–3041.

Sorger, P.K., M.J. Lewis and H.R.B. Pelham. 1987. *Nature* **329:** 81–84.

Spassky, A. and D.S. Sigman. 1985a. *Biochemistry* **24:** 8050–8056.

Spassky, A., K. Kirkegaard and H. Buc. 1985b. *Biochemistry* **24:** 2723–2731.

St. John, T. and R Davis. 1981. *J. Mol. Biol.* **152:** 285–315.

Staudt, L.M., H. Singh, R. Sen, T. Wirth, P.A. Sharp and D. Baltimore. 1986. *Nature* **323:** 640–643.

Stragier, P., C. Parsot and C. Bouvier. 1985. *FEBS Lett.* **187:** 11–15.

Straney, D.C. and D.M. Crothers. 1985.*Cell* **43:** 449–459.

Straney, S.B. and D.M. Crothers. 1987. *Cell* **51:** 699–707.

Strauss, H.S., R.R. Burgess and T.M. Record. 1980. *Biochemistry* ***19:*** 3504–3515.

Strauss, F. and A. Varshavsky. 1984. *Cell* **37:** 889–901.

Struhl, K. 1984. *Proc. Natl. Acad. Sci. USA* **81:** 7865–7869.

Struhl, K. 1986. *Mol. Cell. Biol.* **6:** 3847–3853.

Struhl, K. 1987a. *Cell* **49:** 295–297.

Struhl, K. 1987b. *Cell* **50:** 841–846.

Struhl, K. 1988. *Nature* **332:** 649–650.

Sturm, R., B.R. Bauruker, J.R. Franza and W. Herr. 1987. *Genes & Dev.* **1:** 1147–1160.

Suck, D., A. Lahm and C. Oefner. 1988. *Nature* **332:** 464–468.

Swimmer. C. and T. Shenk. 1984. *Proc. Natl. Acad. Sci. USA* **81:** 6652–6656.

Takahashi, K., M. Vigneron, H. Matthes, A. Wildeman and P. Chambon. 1986. *Nature* **319:** 121–126.

Thompson, J.F., L.M. deVargas, S.E. Skinner and A. Landy. 1987. *J. Mol. Biol.* **195:** 481–493.

Tooze, J. 1982. Molecular biology of tumour viruses second edition, part 2. DNA tumour viruses. CSH Laboratory, New York.

Travers, A.A., A.I. Lamond, H.A.F. Mace and M.L. Berman. 1983. *Cell* **35:** 265–273.

Travers, A.A. 1986b. *Biochem. Soc. Trans.* **14:** 199–200.

Travers, A.A. 1987a. *CRC Cri. Rev. Biochem.* **22:** 181–219.

Travers, A.A. 1987b. *Trends Biochem. Sci.* **12:** 108–112.

Travers, A.A. and A. Klug. 1987. *Nature* **327:** 280–281.

Treisman, R. 1985. *Cell* **42:** 889–902.

Treisman, R. and T. Maniatis. 1985. *Nature* **315:** 72–75.

Treisman, R. 1987. *EMBO J.* **6:** 2711–2717.

Tsapakos, M.J., P.V. Haydock, M. Hermodson and R.L. Somerville. 1985. *J. Biol. Chem.* **260:** 16383–16394.

Tullius, T.D. and B. Dombroski. 1986. *Proc. Natl. Acad Sci. USA* **83:** 5469–5473.

Varshavsky, A.J., O.H. Sundin and M.J. Bohn. 1979. *Cell* **16:** 453–466.

Veldman, G.M., S. Lupton and R. Kamen. 1985. *Mol. Cell. Biol.* **5:** 649–658.

Vogt, P.K., T.J. Bos and R.F. Doolittle. 1987. *Proc. Natl. Acad. Sci. USA* **84:** 3316–3319.

von Hippel, P.H., G.D. Bear, W.D. Morgan and J.A. McSwiggen. 1984. *Ann. Rev. Biochem.* **53:** 389–446.

von Wilcken-Bergmann, B., M. Koenen, H-W. Griesser and B. Muller-Hill. 1982. *EMBO J.* **2:** 1271–1274.

Walter, G., W. Zillig, P. Palm and E. Fuchs. 1967. *Eur. J. Biochem.* **3:**194–201.

Wasylyk, W., C, Wasylyk, H. Matthes, M. Wintzerith and P. Chambon. 1983. *EMBO J.* **2:** 1605–1611.

Wasylyk, B. 1988. *CRC Crit. Rev. in Biochem.* **23:** 77–120.

Weber, I.T., D.B. McKay and T.A. Steitz. 1982. *Nucl. Acids Res.* **10:** 5085–5102.

Weber, I.T. and T.A. Steitz. 1984. *Proc. Natl. Acad. Sci. USA* **81:** 3973–3979.

Weber, F., J. de Villiers and W. Schaffner. 1984. *Cell* **36:** 983–992.

Weber, F. and W. Schaffner. 1985. *Nature* **315:** 75–77.

Webster, N., J.R. Jin, S. Green, M. Hollis and P. Chambon. 1988. *Cell* **52:** 169–178.

Weiderrecht, G., D.J. Shuey, W.A. Kibbe and C.S. Parker. 1987. *Cell* **48:** 507–515.

Weiher, H., M. Konig and P. Gruss. 1983. *Science* **219:** 626–631.

Weintraub, H. and M. Groudine. 1976. *Science* **193:** 848–856.

Weintraub, H. 1985. *Cell* **42:** 705–711.

Wharton, R.P. and M. Ptashne. 1985. *Nature* **316:** 601–605.

Whitson, P.A., J.S. Olson and K.S. Matthews. 1986. *Biochemistry* **25:** 3852–3858.

Whitson, P.A., W.T. Hsieh, R.D. Wells and K.S. Matthews. 1987. **262:** 14592–14599.

Wildeman, A.G., M. Zenke, C. Schatz, M. Wintzerith, T Grundstrom, H. Matthes, K. Takahashi and P. Chambon. 1986. *Mol. Cell Biol.* **6:** 2098–2105.

Wing, R., H. Drew, T. Takano, C. Broka, S. Takana, H. Hakura and R. Dickerson. 1980. *Nature* **287:**

755–758.

Winkler, M.E. and C. Yanofsky. 1981. *Biochemistry* **20:** 3738–3744.

Winter, R.B. and P.H. von Hippel. 1981. *Biochemistry* **20:** 6948–6960.

Wirth, T., L. Staudt and D. Baltimore. 1987. *Nature* **329:** 174–178.

Wolberger, C., Y. Dong, M. Ptashne and S.C. Harrison. 1988. *Nature* **356:** 789–795.

Wu, C. 1980. *Nature* **286:** 854–860.

Wu, C., B. Hunt and R. Morimoto. 1985. *Mol. Cell. Biol.* **5:** 330–341.

Wu, F., J. Garcia, R. Mitsuyasu and R. Gaynor. 1988. *J. Virol.* **62:** 218–225.

Xiao, J.H., I. Davidson, M. Macchi, R. Rosales, M. Vigneron, A. Staub and P. Chambon. 1987a. *Genes & Dev.* **1:** 794–807.

Xiao, J.H., I. Davidson, D. Ferrandon, R. Rosales, M. Vigneron, M Macchi, F. Ruffenach and P. Chambon. 1987b. *EMBO J.* **6:** 3005–3013.

Yaniv, M. 1986. *CRC Crit. Rev. Biochem.*, **21:** 1–26.

Yano, O., J. Kanellopoulos, M. Kieran, O. Le Bail, A. Israel and P. Kourilsky. 1987. *EMBO J.* **6:** 3317–3324.

Zenke, M., T. Grundstrom, H. Matthes, M. Wintzerith, C. Schatz, A. Wildeman and P. Chambon. 1986. *EMBO J.* **5:** 387–397.

Zhang, R.G., A. Joachimiak, C.L. Lawson, R.W. Schevitz, Z. Otwinowski and P.B. Sigler.1987. *Nature* **327:** 591–597.

Zinn, K. and T. Maniatis. 1986. *Cell* **45:** 611–618.

Zurawski, G., R.P. Gunsalus, K.D. Brown and C. Yanofsky. 1981. *J. Mol. Biol.* **145:** 47–73.

Figure 2.3 The binding of EBP1 to bromodeoxyuridine-substituted DNA

A. A bromodeoxyuridine-substituted, [^{32}P]-labelled, double stranded DNA fragment containing the SV1 sequence was incubated in the presence (+), or absence(-) of partially purified EBP1 under conditions employed in the standard gel electrophoresis DNA binding assay, DNA–protein complexes resolved on 6% polyacrylamide gels and visualised by autoradiography. Arrows depict the positions of the DNA–protein complexes (B) or free DNA (F).

B. Competition analysis of EBP1 binding. Reactions were prepared as in **A** but with the inclusion of 10 (lane 1), 25 (lane 2), 50 (lane 3) or 100 ng (lane 4) of unlabelled, double stranded SV1 oligonucleotide as competitor, in the binding mixture. Complexes were resolved and visualised as described in **A**.

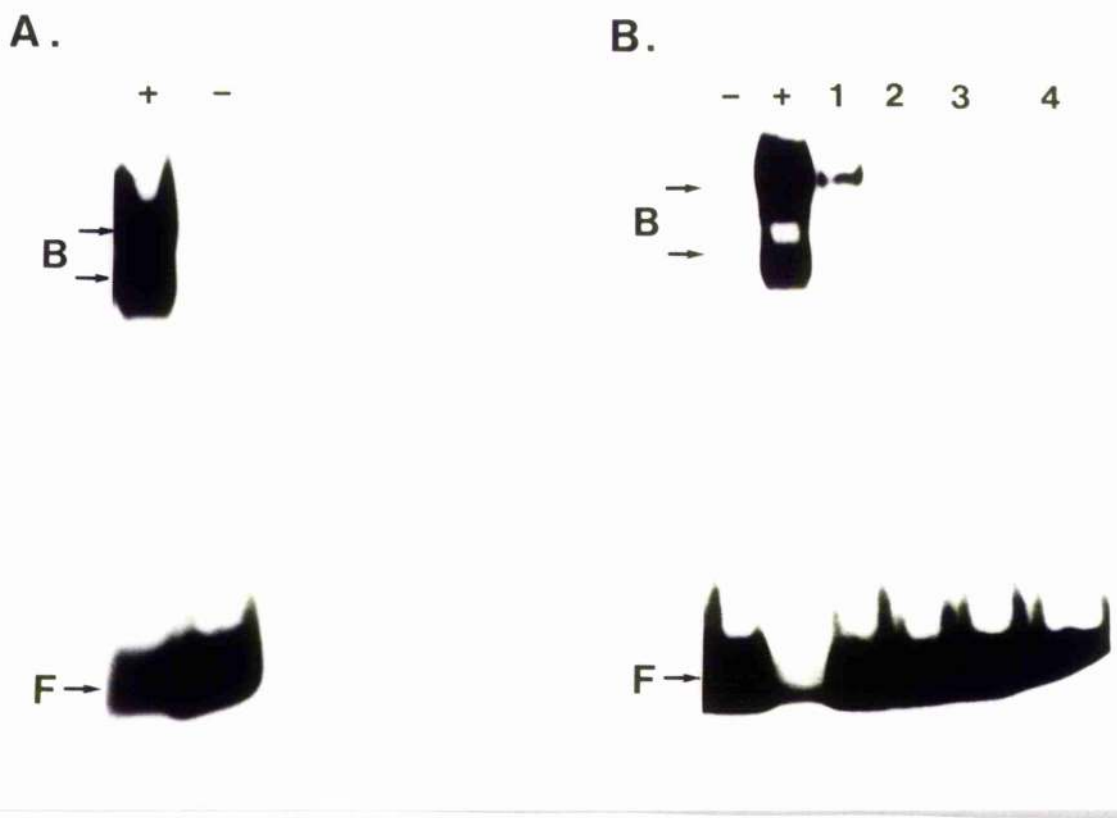


Figure2.4 Bromodeoxyuracil cross-linking of EBP1 to its binding site on the SV40 genome

A bromodeoxyuridine substituted, [^{32}P]-labelled double stranded DNA fragment containing the EBP1 binding site was prepared as described in "Materials and Methods". Probe DNA was incubated with affinity purified EBP1 under conditions employed in the gel electrophoresis DNA binding assay. Once equilibrium had been reached, samples were irradiated with UV light followed by treatment with DNase I and micrococcal nuclease as indicated. Reaction products were fractionated on 10% SDS polyacrylamide gels and visualised by autoradiography. The positions of the 66,000 and 45,000 molecular weight protein standards are indicated by the arrows. UV irradiation, in the absence of competitor was for 0, 5, 10 and 20 minutes. In the competition analysis EBP1 was allowed to bind in the presence of 0.1 µg SV1, 0.5 µg SV1, 0.5 µg NFI or 0.5 µg NFIII double stranded oligonucleotides prior to UV irradiation for 10 minutes. Reaction productes which were UV irradiated for 10 minutes without subsequent DNase I or mirococcal nuclease digestion are also shown (-MNase,DNase).

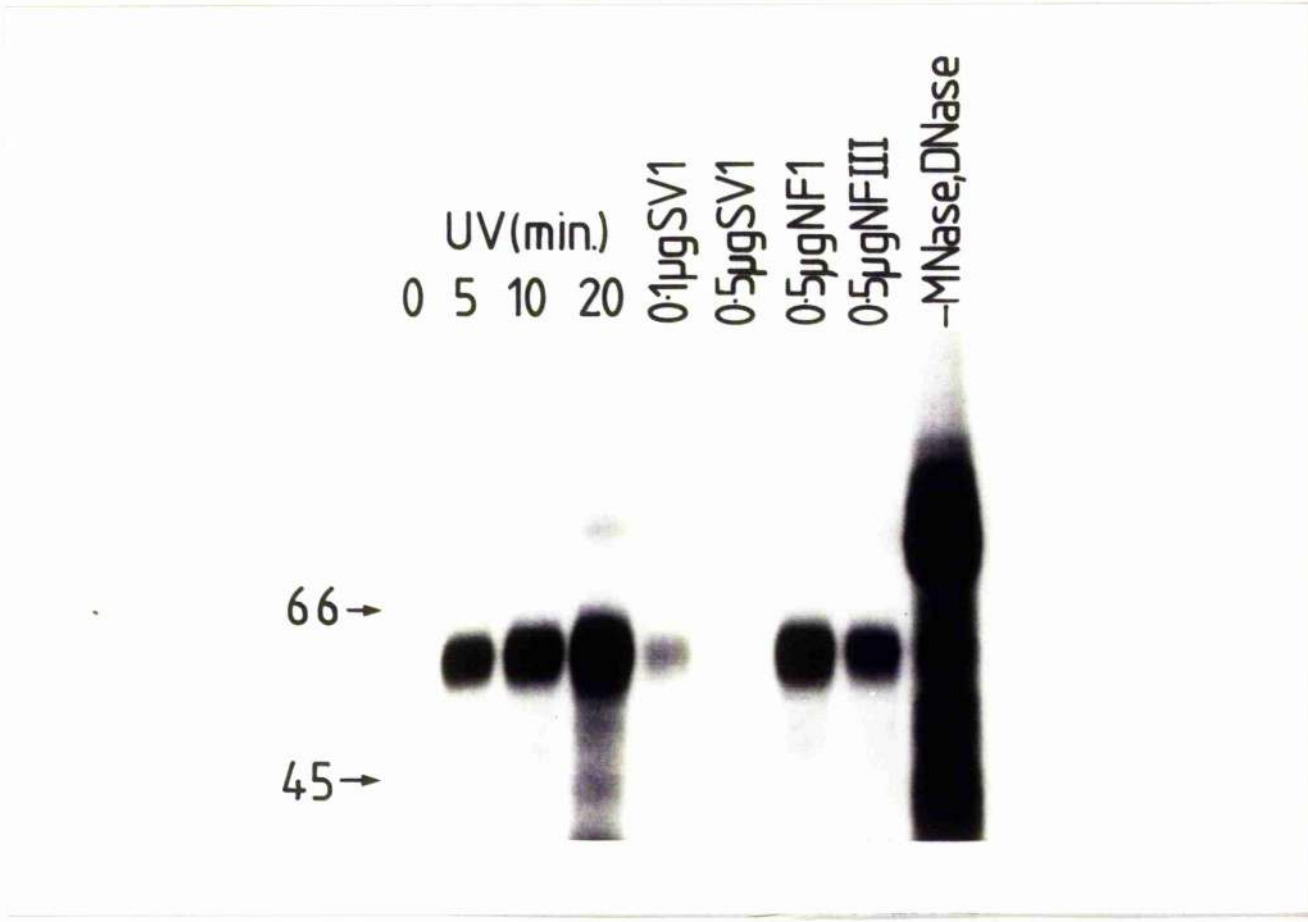


Figure 2.5. DNase I protection of the SV40 enhancer by affinity-purified EBP1

The *Bam* HI to *Pvu* II fragment from pUC1X72, 3'-[^{32}P] labIled at the *Bam* HI site was incubated with 0, 5, 10 or 20 µl of affinity purified EBP1 as described in "Materials and Methods". Competition assays were carried out in which 10 µl of EBP1 was incubated with labelled probe in the presence of 500 ng unlabelled SV1, SV1.M1 or SV1.M2 double stranded oligonucleotide. After treatment with DNase I, reaction products were electrophoresed in a 6% denaturing polyacrylamide gel. C+T and A+G specific cleavage reactions of the labelled fragment were electrophoresed in parallel as markers.

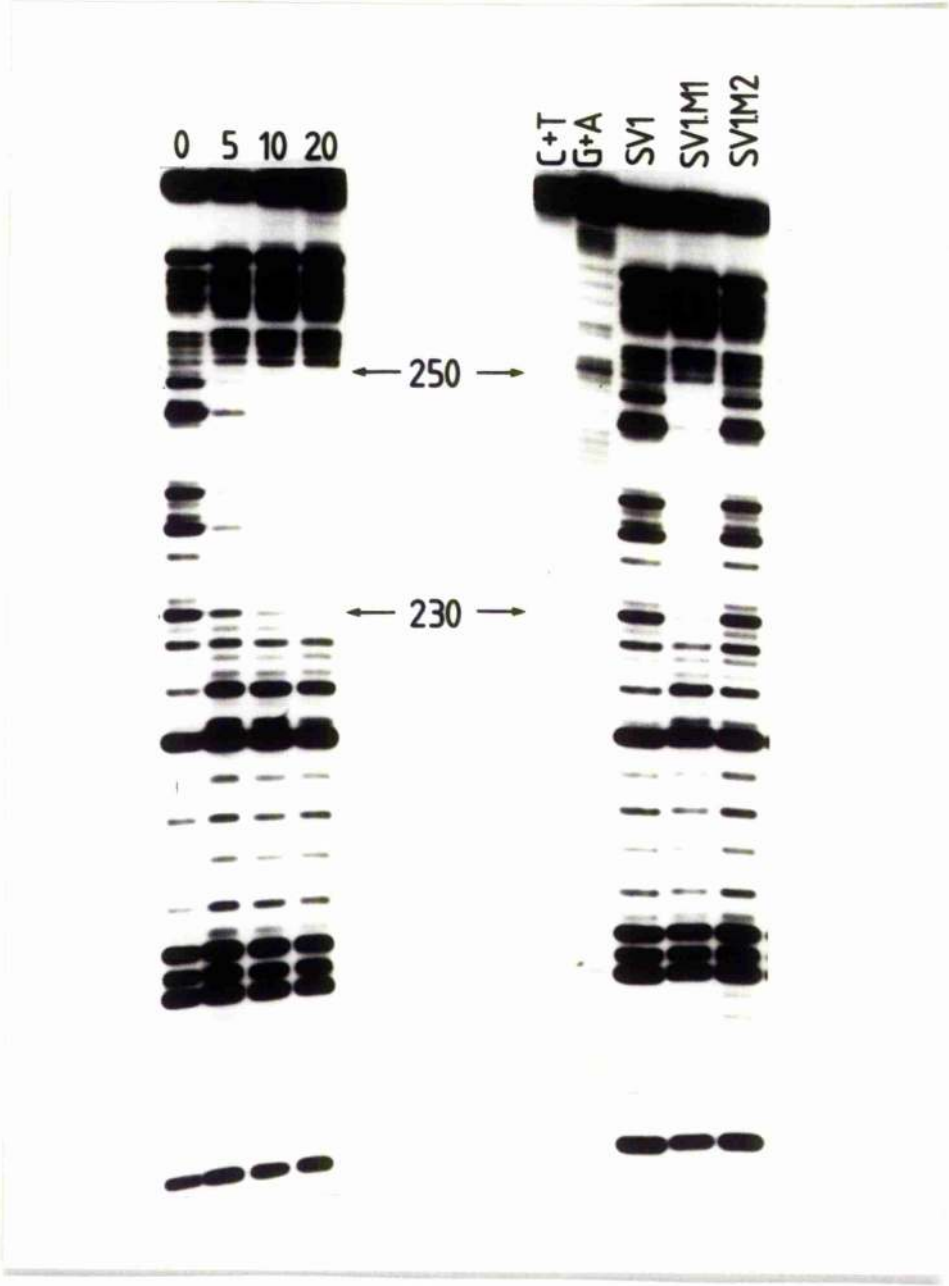


Figure 2.6 Proteolytic cleavage of EBP1

Reactions contained 0.5 ng [^{32}P]-labelled, double stranded SV1 oligonucleotide, 1.0 μg of unlabelled poly [d(A-T)] poly [d(G-C)] and 2 μl of affinity purified EBP1; (–) indicates that EBP1 was omitted from these reactions. In addition, reactions contained 0 (lanes 0), 5 (lanes 1), 10 (lanes 2), 20 (Elastase and Type XIV lanes 3) or 5000 μg (Prot.K lane 3) of proteinase K, elastase or protease XIV. Proteolytic enzymes were added after the binding reactions had reached equilibrium and digestion was allowed to proceed for 15 minutes at 20°C. Reaction products were fractionated on native 6% polyacrylamide gels.

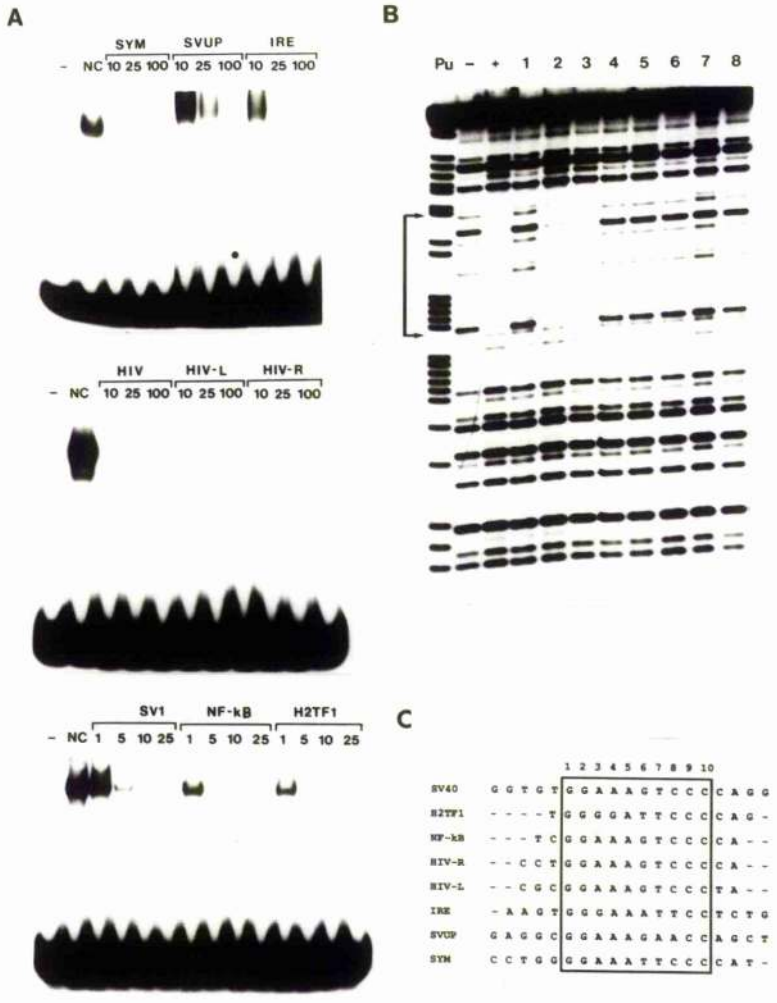

A
SYM
SVUP
IRE
- NC 10 25 100 10 25 100 10 25 100
HIV
HIV-L
HIV-R
- NC 10 25 100 10 25 100 10 25 100
SV1
NF-kB
H2TF1
- NC 1 5 10 25 1 5 10 25 1 5 10 25
B
Pu - + 1 2 3 4 5 6 7 8
C
1 2 3 4 5 6 7 8 9 10
SV40 G G T G T G G A A A G T C C C C A G G
H2TF1 - - - - T G G G G A T T C C C C A G -
NF-kB - - - T C G G A A A G T C C C C A - -
HIV-R - - C C T G G A A A G T C C C C A - -
HIV-L - - C G C G G A A A G T C C C C T A - -
IRE - A A G T G G G A A A T T C C T C T G
SVUP G A G G C G G A A A G A A C C A G C T
SYM C C T G G G G A A A T T C C C C A T -

Figure 3.2 DNase I protection and OP/Cu+ cleavage of the IRE in the presence of EBP1

A. A *Hind* III to *Bgl* II fragment from pSPIRE, was 3'-[^{32}P]-labelled (top) or 5'-[^{32}P] labelled (bottom) at the *Hind* III site, incubated with 0, 5, 10 or 20 µl of affinity purified EBP1 and digested with DNase I as described in "Materials and Methods". DNA was isolated, fractionated by electrophoresis in 8% denaturing polyacrylamide gels and the cleavage products visualised by autoradiography. C+T (Py) and G+A (Pu) specific cleavage reactions of the labelled fragment were electrophoresed in parallel as markers.

B. The aforementioned labelled fragments were incubated with 20 µl of affinity purified EBP1 and, after binding reactions had reached equilibrium, free DNA (F) was separated from EBP1–DNA complexes (B) by electrophoresis in a native polyacrylamide gel. Cleavage with OP/Cu+ was carried out within the acrylamide matrix and DNA eluted as described in "Materials and Methods". DNA was fractionated, and products visualised as described in **A**. Deoxyribose residues within the bracketed region are protected from cleavage by OP/Cu+.

C. DNA sequence of the IRE containing the EBP1 binding site. Phosphate bonds within the large brackets are protected from DNase I cleavage broken lines depict potential protected bonds). Deoxyribose residues within the small bracketed regions are protected from OP/Cu+ cleavage in the presence of EBP1.

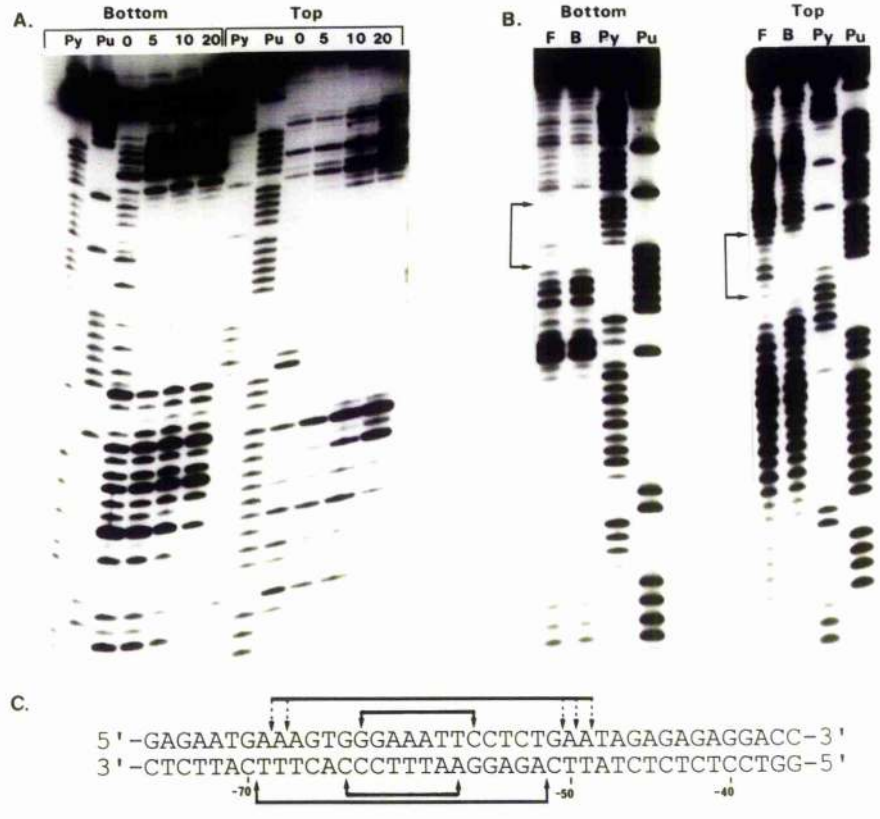
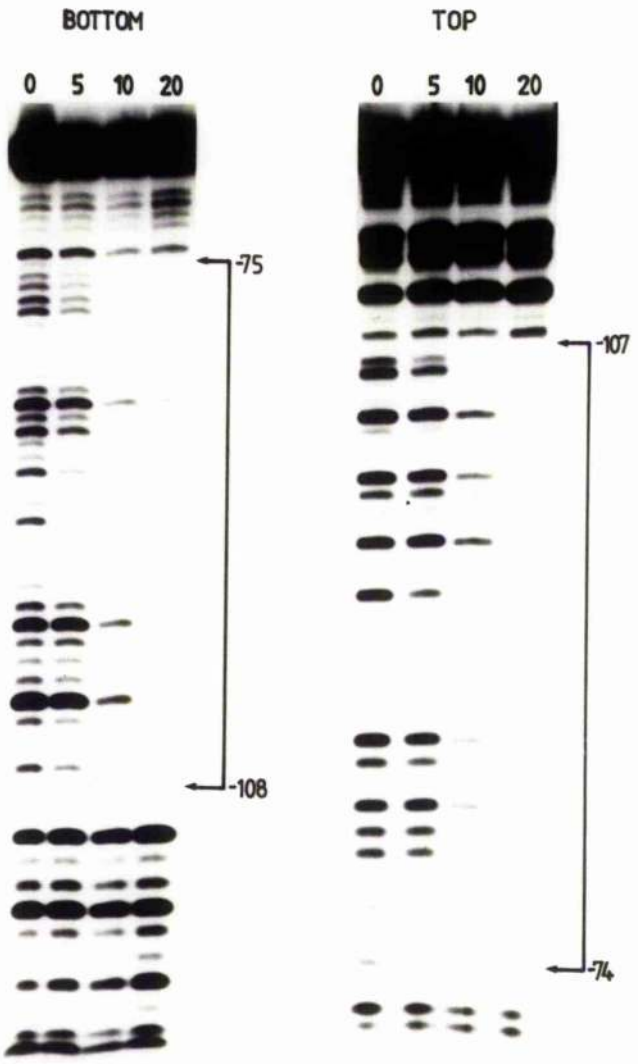


Figure 3.3 DNase I protection of the HIV enhancer in the presence of EBP1

A *Hind* III to *Sac* I fragment, or *Eco* RI to *Pst* I fragment, from pHIVEn was 3'-end labelled at the *Hind* III (Bottom; **A**) or *Eco* RI (Top; **B**) site, incubated in the absence (0) or presence of 5, 10 or 20 µl of affinity purified EBP1 and digested with DNase I as described in "Materials and Methods". DNA was isolated and fractionated by electrophoresis in 8 % denaturing polyacrylamide gels and the cleavage products visualised by autoradiography.

C. DNA sequence of the HIV enhancer containing the EBP1 binding sites (as depicted by the horizontal bars). Phosphate bonds within the bracketed regions are protected from DNase I cleavage in the presence of EBP1.



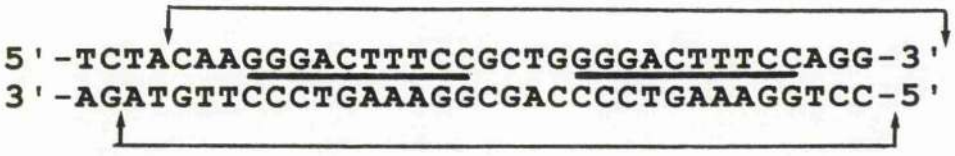


Figure 4.1 DNA binding specificity of affinity purified EBP

Competition analysis of EBP1 binding. Reactions contained 0.5 ng [^{32}P]-labelled double stranded SV1 oligonucleotide, 1.0 μg of unlabeled poly [d(A-T)]: poly [d(C-G)] and 1.0 μl of affinity purified EBP1 (lane (-) contained no EBP1). In addition, reactions contained 25 (lanes 1), 50 (lanes 2) or 100 ng (lanes 3) of unlabelled double stranded SV1, SV1.M1 or SV1.M3 oligonucleotides. Reaction products were analysed on 6% polyacrylamide gels. The nucleotide sequence of the SV1 oligonucleotide is shown – the corresponding changes incorporated into SV1.M1 and SV1.M3 are indicated.

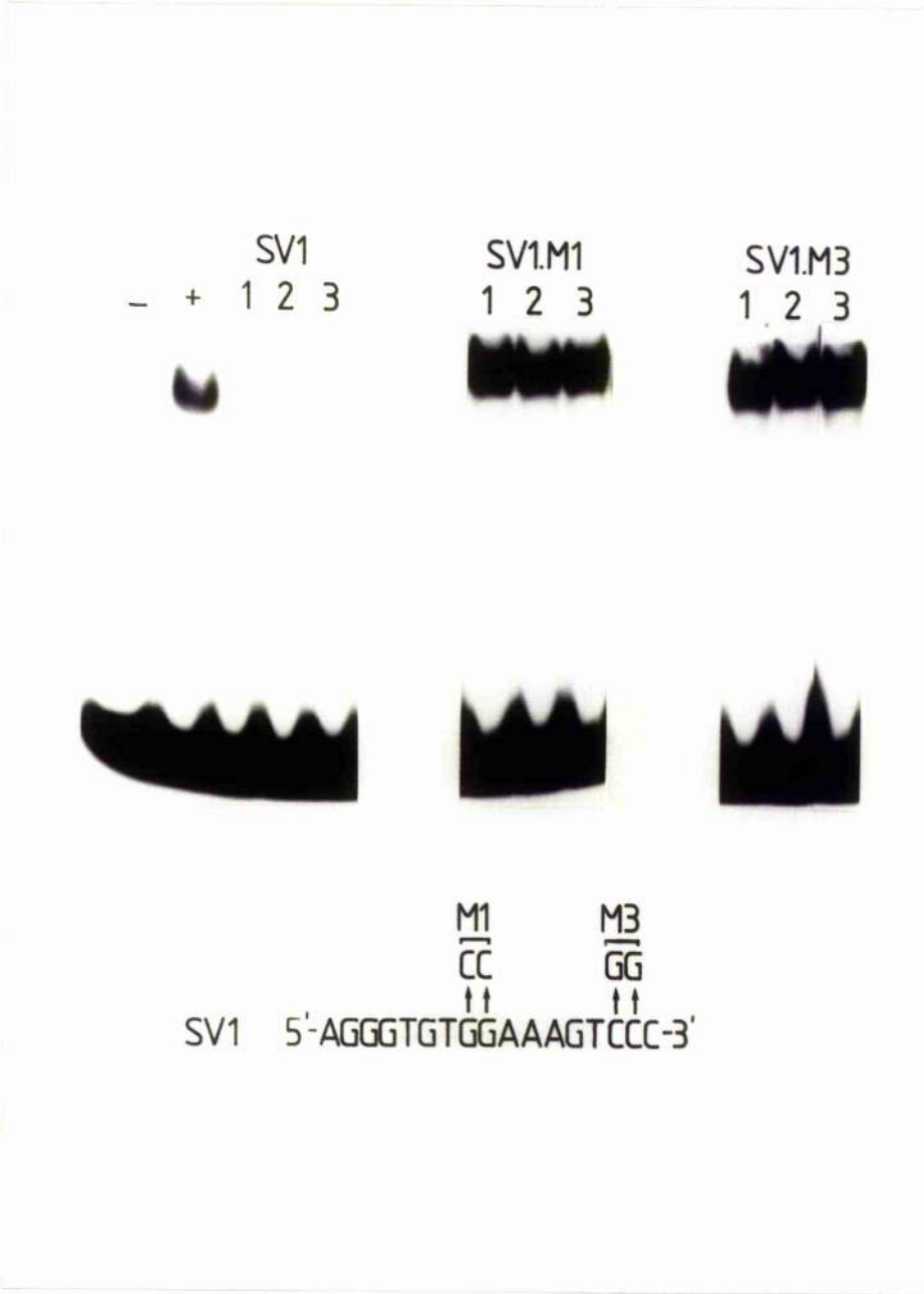


Figure 4.2 Mutational analysis of the EBP1 binding site present in the human ß-interferon gene regulatory element (IRE)

A. *Bgl* II to *Nco* I fragments were isolated from the pIF10 series of plasmids, which contain single point mutations throughout the IRE, and [^{32}P]-labelled at the *Bgl* II site. Labelled fragments (0.5 ng) were incubated with 0.5 µg of unlabelled poly[d(A-T)]: poly [d(G-C)] and 2 µl of affinity purified EBP1 in the standard gel electrophoresis DNA binding assay. Binding reactions were fractionated on 6% polyacrylamide gels and products visualised by autoradiography. The individual nucleotide changes incorporated into the mutant templates are indicated above the corresponding lanes. No EBP1 was present in the reaction represented by the dash.

B. Nucleotide changes incorporated in the pIF10 mutants used in this study. The top line represents the wild type (wt) IRE sequence together with the relevant nucleotide numbers. Arrows represent the bases which have been altered with the bottom line depicting the corresponding changes in these altered bases.

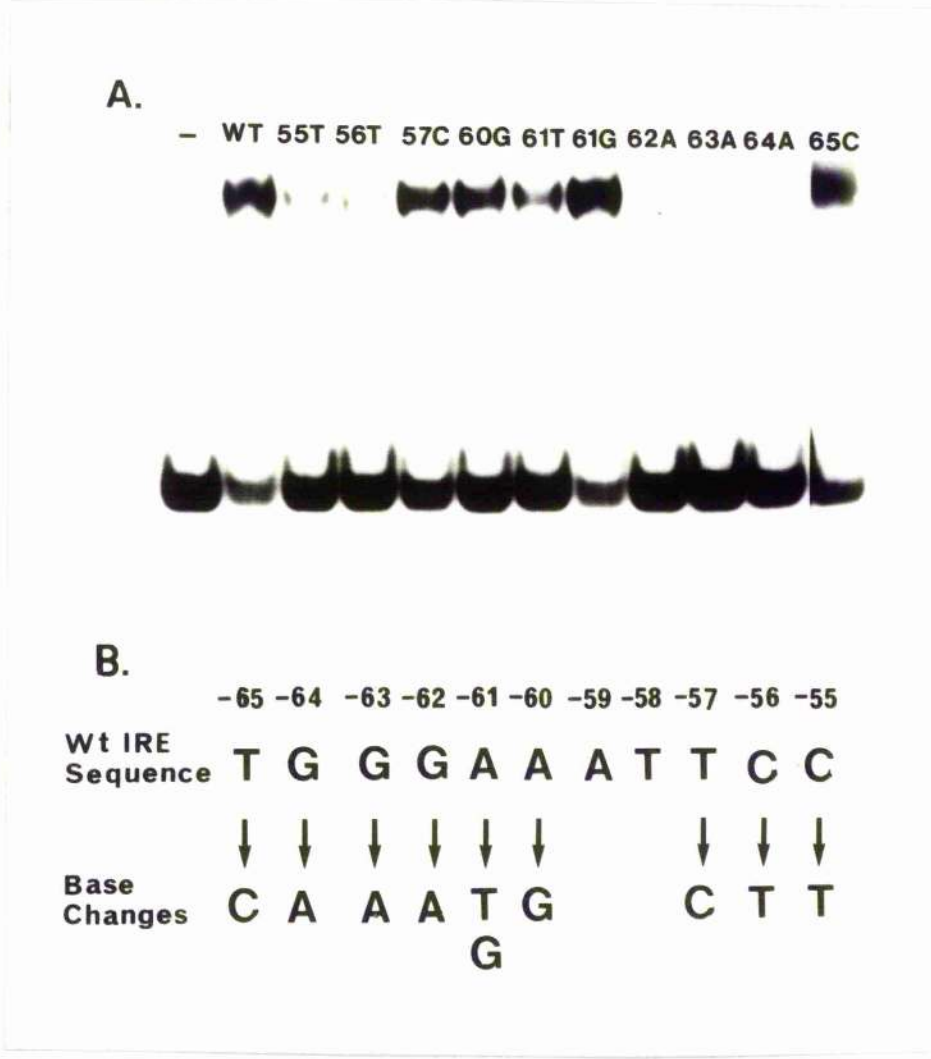


Figure 4.3 Mutational analysis of the EBP1 binding site on the SV40 enhancer.

A. *Bam* HI to *Pvu* II fragments from pUC1X72 and pA11 to pA16, were 3'-[^{32}P] labelled at the *Bam* HI site, and incubated with 2 µl of affinity purified EBP1, in the presence of 0.5 µg of unlabelled poly[d(A-T)]: poly[d(G-C)], in the standard gel electrophoresis DNA binding assay. Binding reactions were fractionated in a 6% polyacrylamide gel and products visualised by autoradiography. The relative binding of EBP1 to the various templates (as determined by Cerenkov radiation counting of bound species) is indicated.

B. DNase I protection of mutated SV40 enhancer sequences by EBP1. The [^{32}P]-labelled fragments described in **A** were incubated with 0 (–) or 20 µl (+) of affinity purified EBP1 and treated with DNase I as described in "Materials and Methods". DNA was isolated and fractionated in 8% denaturing polyacrylamide gels and the cleavage products visualised by autoradiography. G+A (Pu) specific cleavage reactions of the labelled fragments were electrophoresed in parallel as markers. Sources of labelled fragments are indicated at the foot of the appropriate lanes.

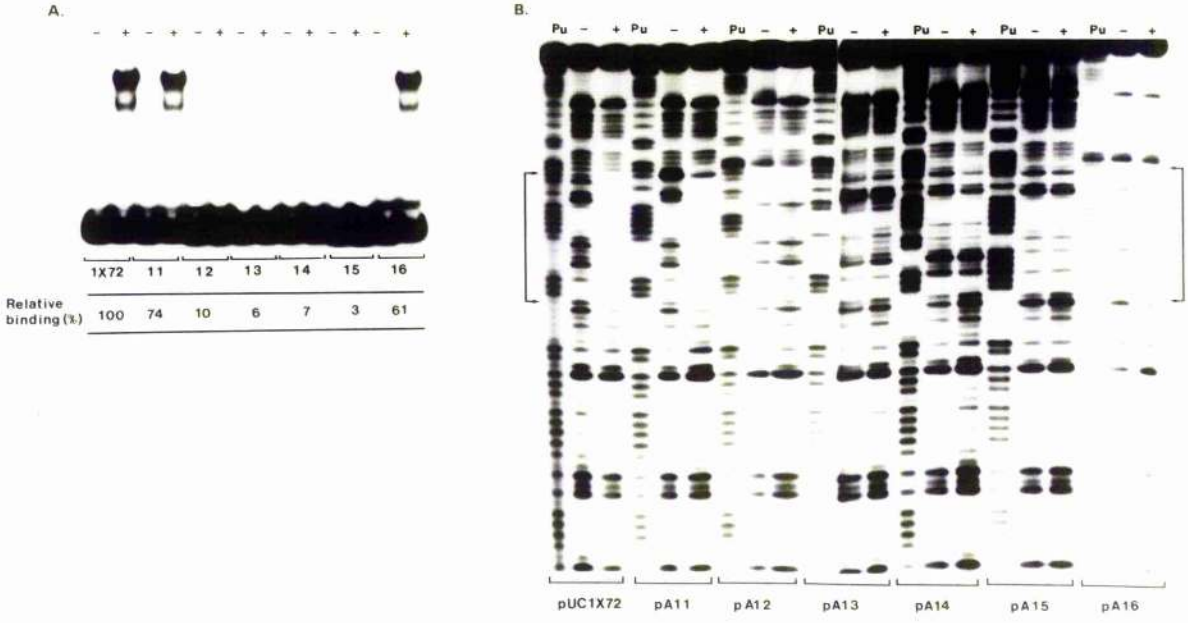


Figure 4.4 Frequency of base usage within EBP1 binding sites

Nineteen different EBP1 binding sites, including naturally occurring sequences and mutated derivatives, were analysed and the frequency with which each base occurs at each position determined. A = adenine, C = cytosine, G = guanine, and T = thymine nucleotides respectively.

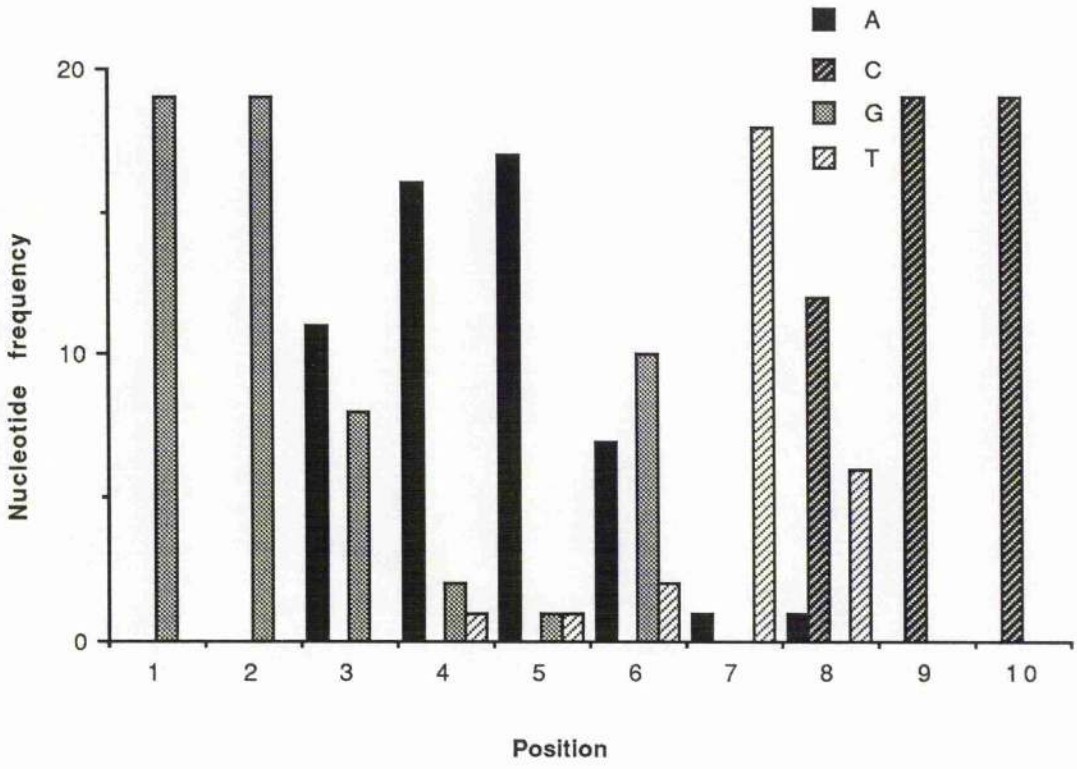
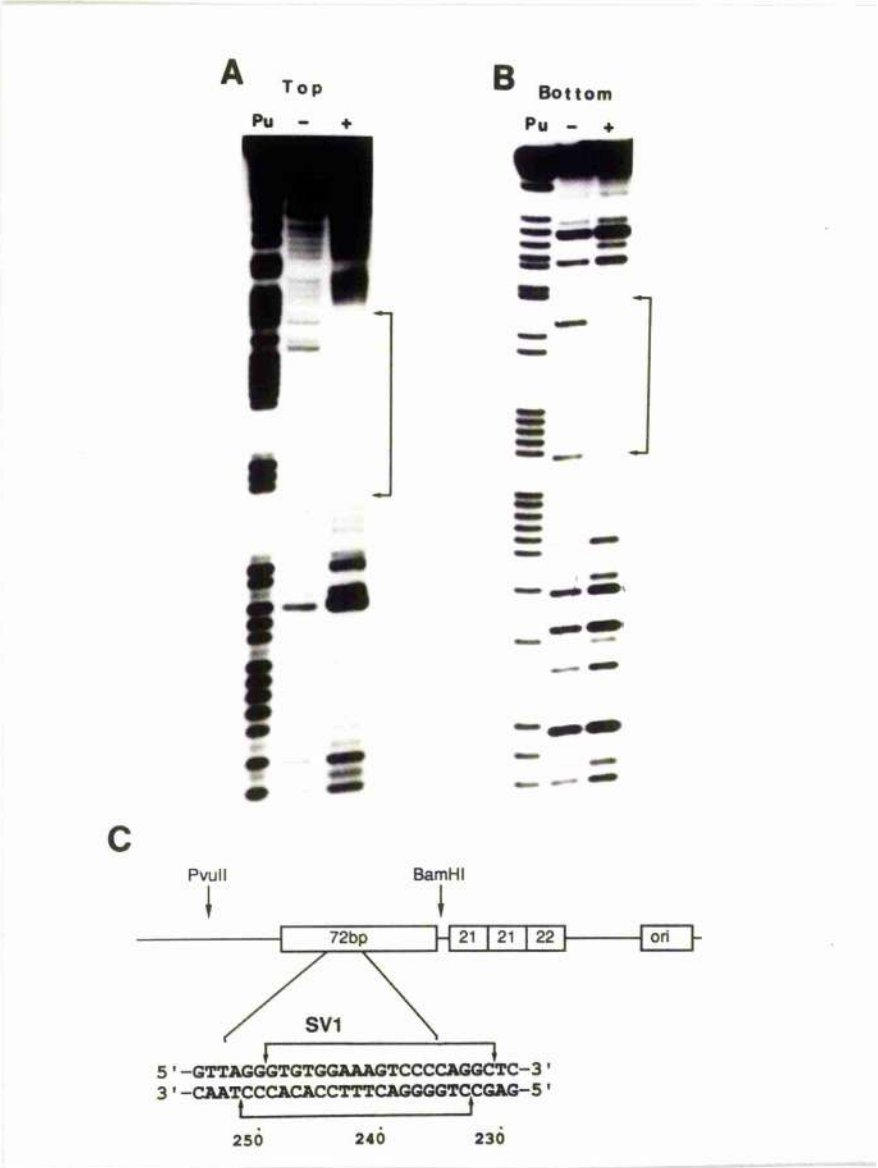
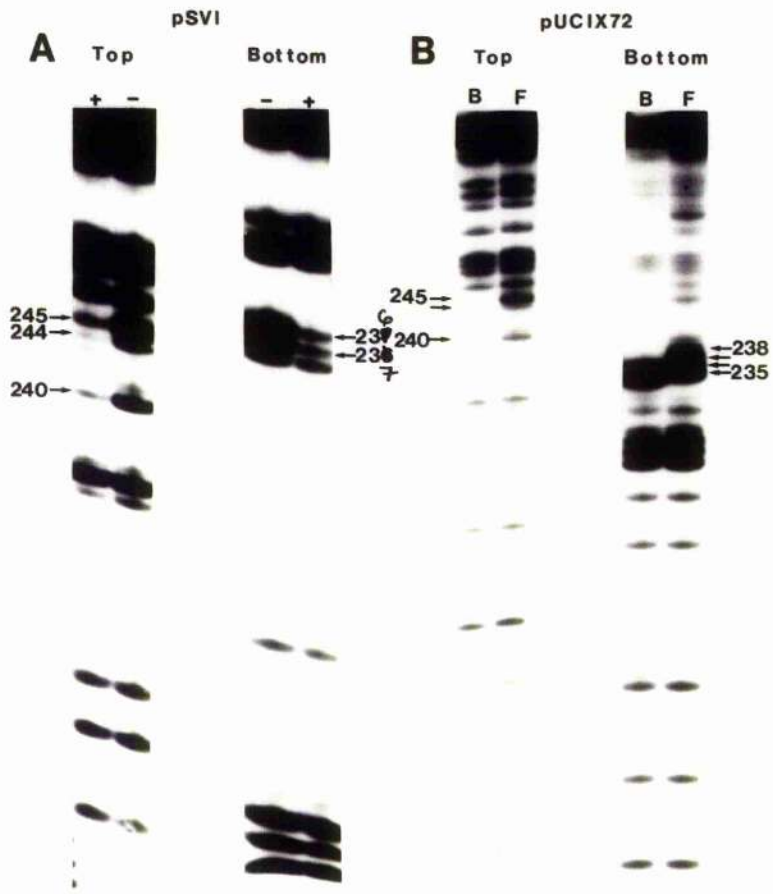


Figure 5.1 DNase I protection of the SV40 enhancer by EBP1

A *Bam* HI to *Eco* RI fragment from pUC1X72, 3'-end labelled (**A**) or 5'-end labeled (**B**) at the *Bam* HI site, was incubated in the absence (-) or presence (+) of 10 µl affinity purified EBP1 and digested with DNase I as described in "Materials and Methods". DNA was isolated and fractionated by electrophoresis in 6% denaturing polyacrylamide gels and the cleavage products visualised by autoradiography. G+A (Pu) specific cleavage reactions of the labelled fragments were electrophoresed in parallel as markers. Phosphate bonds within the brackets are protected from DNase I cleavage.

C. Organisation of the SV40 transcriptional control region. The 21bp repeated elements of the early promoter, the origin of DNA replication (ori) and the enhancer, containing one copy of the 72 bp repeat, are indicated. Shown below is the DNA sequence of the enhancer region containing the EBP1 binding site. Phosphate bonds within the brackets are protected from DNase I cleavage. DNA sequences present in the double stranded SV1 oligonucleotide are also indicated.





C

5'-GTTAGGGTGTGGAAAGTCCCCAGGCTC-3'

3'-CAATCCCACACCTTTCAGGGGTCCGAG-5'

250 240 230

Figure 5.3 Methylated bases that interfere with EBP1 binding.

Hind III to *Kpn* I (**A**) or *Eco* RI to *Pst* I (**B**) fragments from pSV1, 3'-end labelled at the *Hind* III (top) or *Eco* RI (bottom) sites were treated with DMS and the methylated DNA incubated with 10 µl affinity purified EBP1. Free DNA (F) and EBP1–DNA complexes (B) were separated by electrophoresis in a native polyacrylamide gel. DNA was eluted from the gel and cleaved at modified purine bases by treatment with NaOH. Cleavage products were fractionated by electrophoresis in 8% denaturing polyacrylamide gels. Positions of interfering bases are indicated by arrows.

C. DNA sequence of the region of the SV40 enhancer containing the EBP1 binding site. Filled circles depict modified bases which interfere with EBP1 binding.

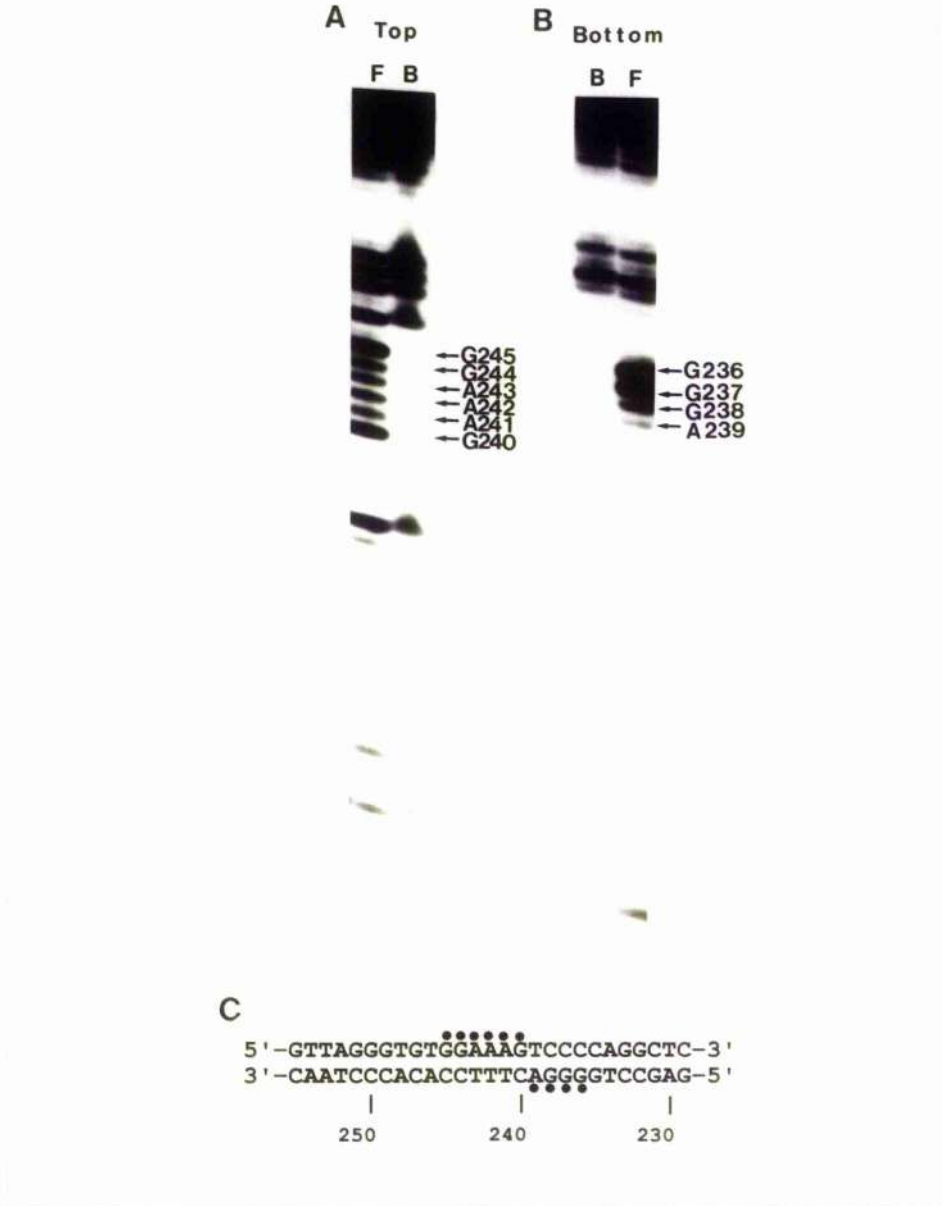
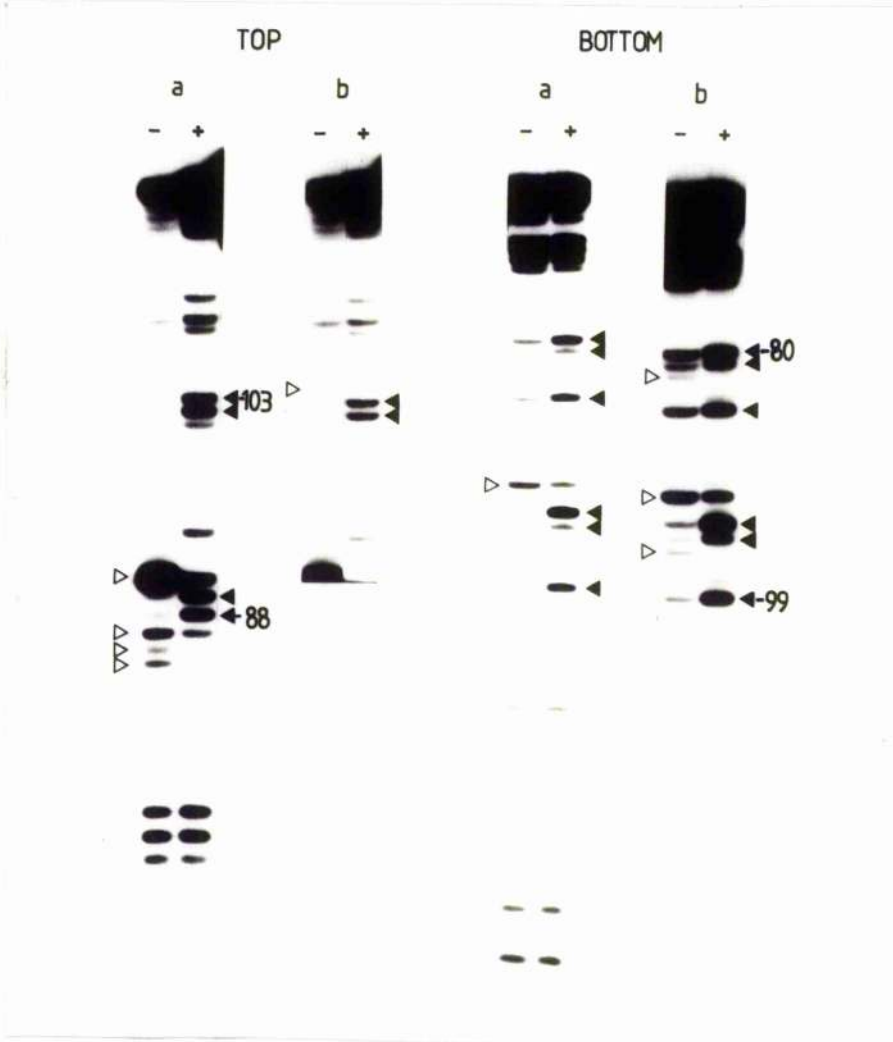


Figure 5.4 Methylation protection of bases in the HIV enhancer in the presence of EBP1

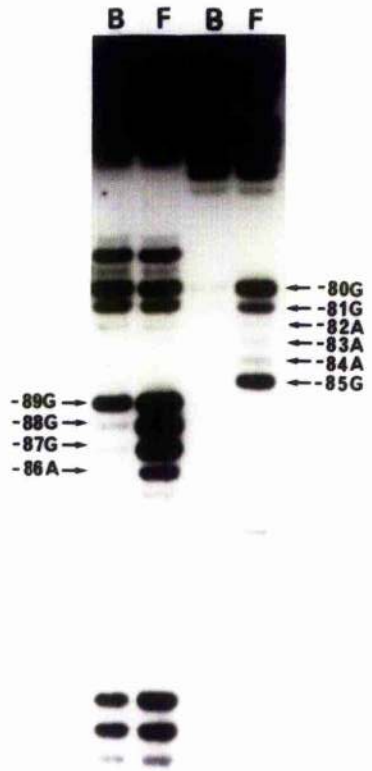
A *Hind* III to *Sac* I fragment, or *Eco* RI to *Pst* I fragment, from pHIVEn was 3'-end labelled at the *Eco* RI (Top; **A**) or *Hind* III (Bottom; **B**) site, was incubated in the absence (-) or presence of 20 µl of EBP1 (+) and treated with DMS as described in "Materials and Methods". DNA was cleaved at modified bases by treatment with piperidine and the reaction products fractionated by electrophoresis in 8 % denaturing polyacrylamide gels. Cleavage products were visualised by autoradiography after 15 (Top, b; Bottom, a) or 48 hours (Top. a; Bottom, b) exposure to X-ray film. Positions with altered sensitivity to DMS are indicated by filled triangles, in the case of protected bases, or open triangles, in the case of bases hypersensitive to DMS methylation.

C. DNA sequence of the HIV enhancer containing the two EBP1 binding sites. Positions of the bases with altered sensitivity to DMS are indicated as described above.



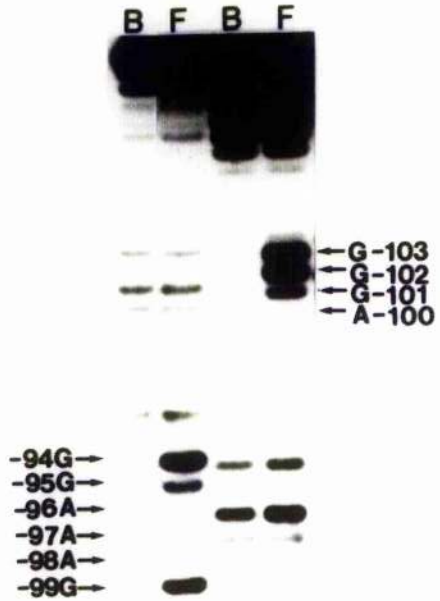
5'-TCTACAAGGGACTTTCCGCTGGGGACTTTCCAGG-3'
3'-AGATGTTCCCTGAAAGGCGACCCCTGAAAGGTCC-5'

A.

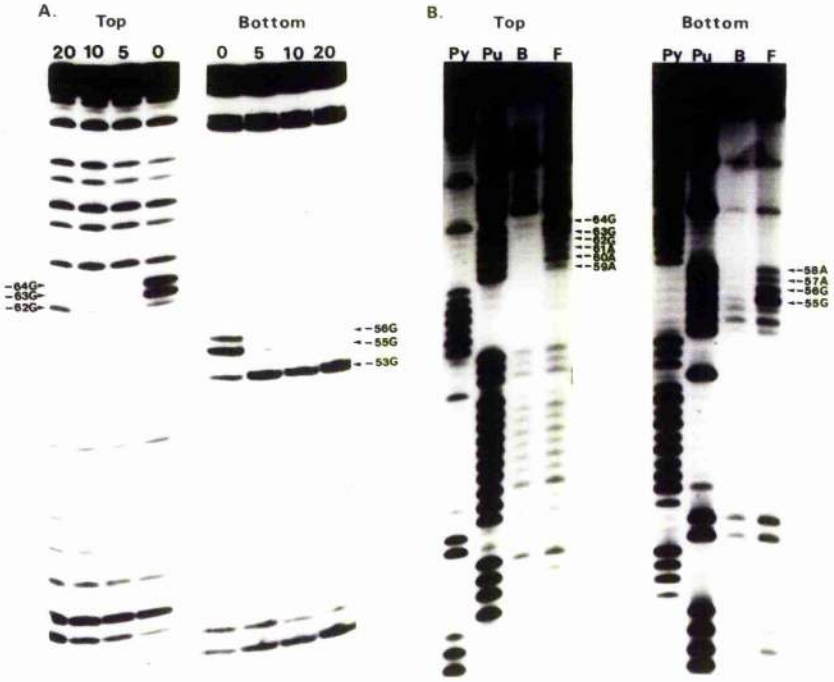


5′GATCTGGGGACTTTCCAGG 3′ HIV-R
3′ACCCCTGAAAGGTCCCTAG 5′

B.



5′GATCTAGGGACTTTCCGCG3′ HIV-L
3′ATCCCTGAAAGGCGCCTAG5′


A.
Top
Bottom
20 10 5 0
0 5 10 20
-64G
-63G
-62G
-56G
-55G
-53G
B.
Top
Bottom
Py Pu B F
Py Pu B F
-64G
-58A
-57A
-56C
-55G
59A

C.


5'-GAGAATGAAAGTGGGAAATTCCTCTGAATAGAGAGAGGACC-3'
3'-CTCTTACTTTCACCCTTTAAGGAGACTTATCTCTCTCCTGG-5'
-70
-60
-50
-40

FIGURE 5.7 Summary of the EBP1/DNA DMS Methylation protection and interference analysis

The nucleotide sequences containing the EBP1 binding sites present in the SV40 enhancer, IRE and HIV (HIV-R and HIV-L) are shown. Filled circles represent bases which are protected from DMS methylation in the presence of EBP1 while open circles indicate bases with an increased reactivity towards DMS modification in the presence of EBP1. Purine bases which, when modified, interfere with EBP1 binding are depicted by the shaded boxes.

Figure 5.8 Ethylated phosphotriesters that interfere with EBP1 binding

A *Bam* HI to *Eco* RI fragment from pUC1X72, 3'-end labelled (**A**, top) or 5'-end labelled at the *Bam* HI site(**B**, bottom), was treated with ethylnitrosourea as described in "Materials and Methods". Ethylated DNA was incubated with 10 µl affinity purified EBP1and free DNA (F) separated from EBP1–DNA complexes (B) in a native polyacrylamide gel. DNA was eluted and cleaved at phosphtriester bonds by treatment with NaOH. Reaction products were fractionated by electrophoresis in 6% denaturing polyacrylamide gels and cleavage products visualised by autoradiography. G+A (Pu) and C+T (Py) specific cleavage reactions of the labelled fragments were run in parallel as markers. Positions of modified phosphotriesters that interfere with binding are indicated by arrows.

C. DNA sequence of the region of the SV40 enhancer containing the EBP1 binding site. Modified phosphate bonds within the bracket interfere with EBP1 binding.

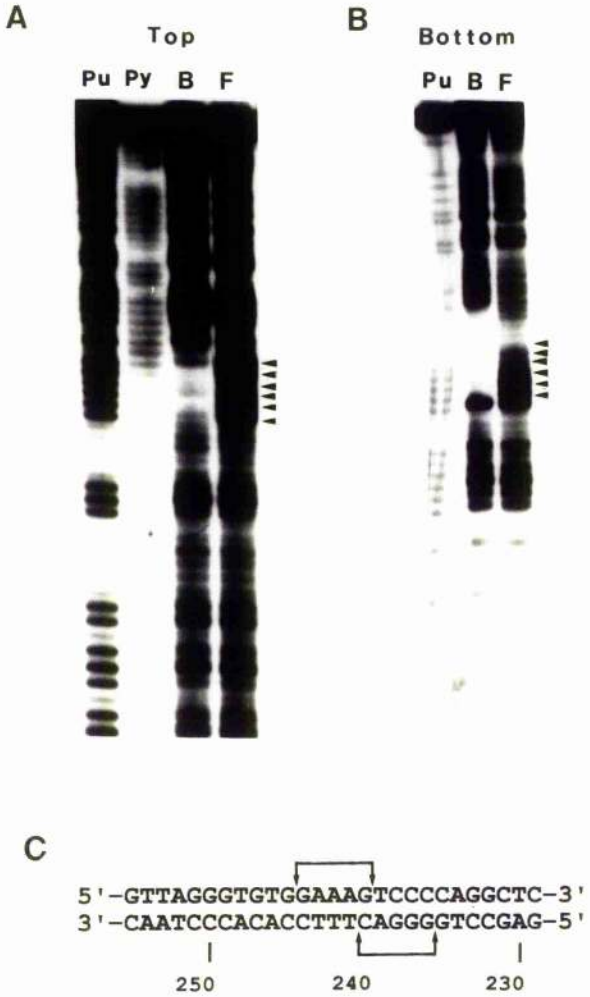


Figure 5.9 1,10-orthophenanthroline/copper cleavage of the SV40 enhancer in the presence of EBP1

A *Bam* HI to *Eco* RI fragment from pUC1X72, 3'-end labelled (A, top) or 5'-end labelled (B, bottom) at the Bam HI site, was incubated with 10 µl affinity purified EBP1 and, after binding reactions had reached equilibrium, free DNA (F)was separated from EBP1–DNA complexes (B) by electrophoresis in a native polyacrylamide gel. Cleavage with OP/Cu+ was carried out within the acrylamide matrix and DNA eluted as described in "Materials and Methods". DNA was fractionated by electrophoresis in 6% denaturing polyacrylamide gels and cleavage products visualised by autoradiography. G+A (Pu) specific reactions of the labelled fragments were electrophoresed in parallel as markers. Deoxyribose residues within the bracketed region are protected from cleavage by OP/Cu+.

C. DNA sequence of the region of the SV40 enhancer containing the EBP1 binding site. Deoxyribose residues within the bracketed region are protected from OP/Cu+ cleavage in the presence of EBP1.

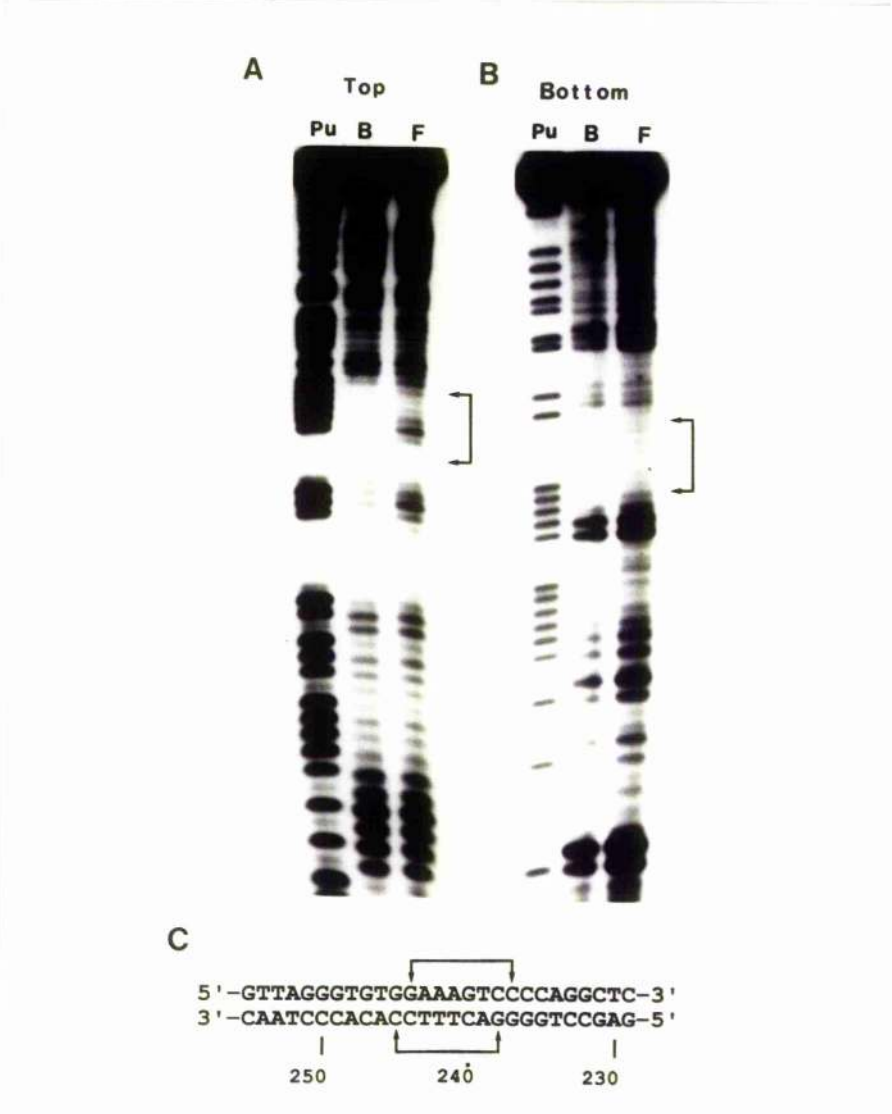
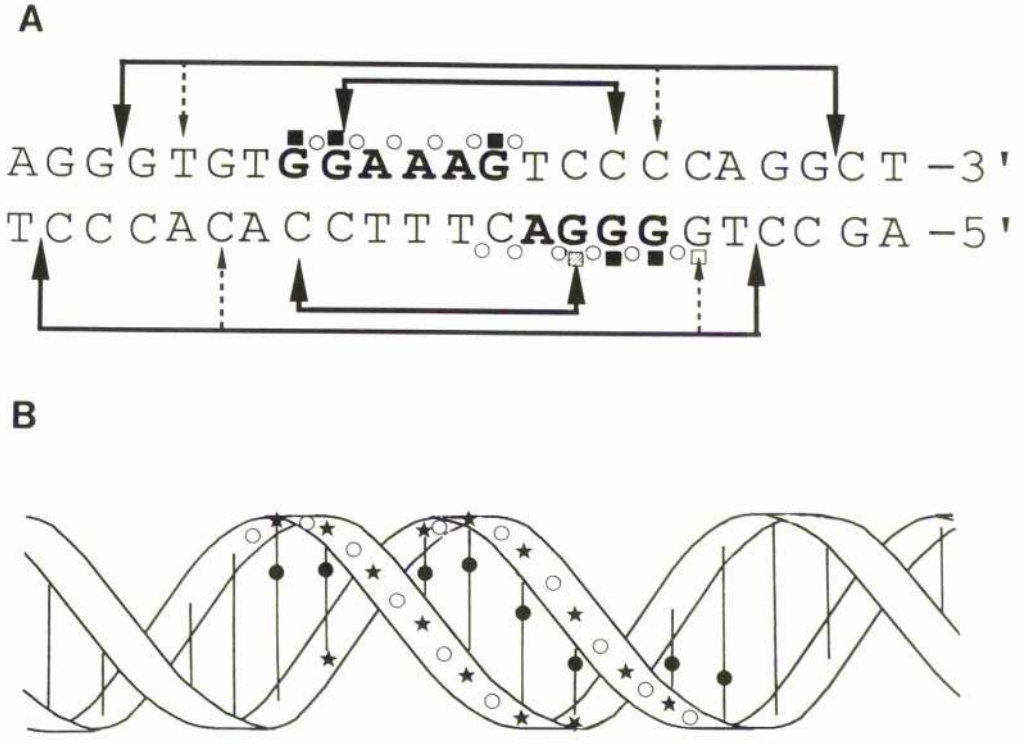


Figure 5.10 Summary of EBP1/DNA contact point analysis

A. Phosphodiester bonds within the large bracket are protected from cleavage with DNase I in the presence of EBP1. Deoxyribose residues within the smaller brackets are protected from cleavage with OP/Cu+. Bold letters indicate that these bases interfere with EBP1 binding when methylated. Filled squares indicate bases protected from methylation while the open square indicates that this base was hypermethylated in the presence of EBP1. The shaded square indicates that this base was protected in the plasmid pUC1X72 but not in pSV1. Open circles indicate the positions of phosphotriesters that interfere with EBP1 binding.

B. Helical projection of the EBP1 binding site. Modified bases (●) and phosphates (○) that interfere with binding are indicated, as are deoxyribose residues (★) that are protected from cleavage with OP/Cu+ in the presence of EBP1.



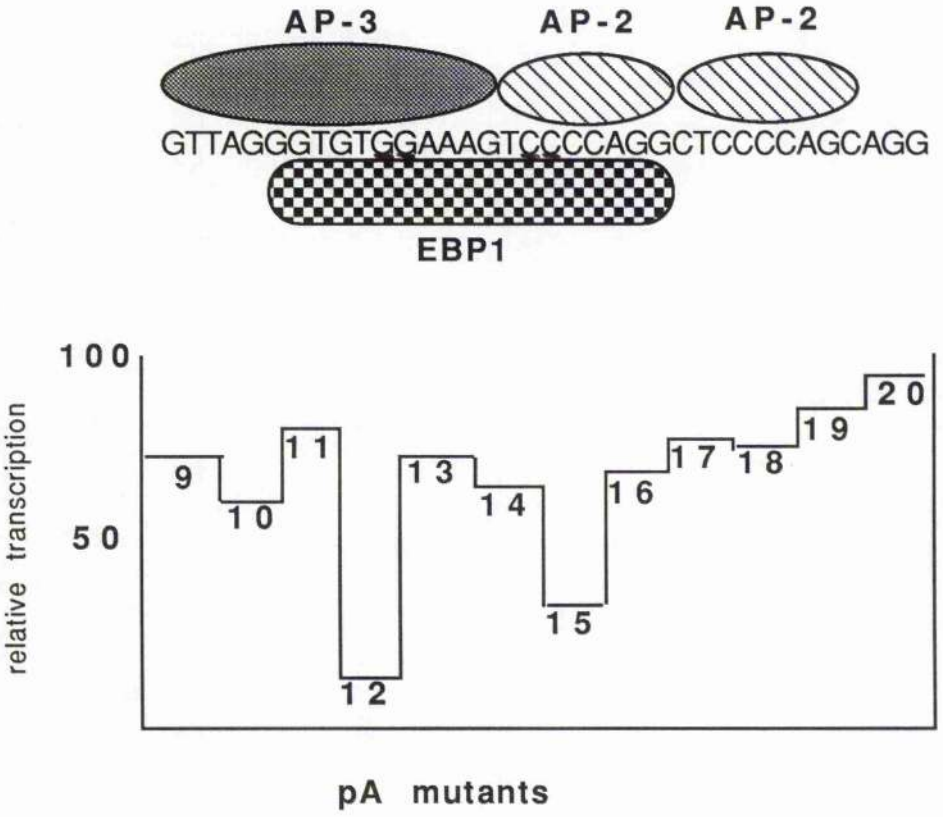


Figure 1.D Disposition of transcription factors AP-2, AP-3 and EBP1 on the SV40 enhancer

The DNase I "footprint" boundaries for AP-2 and AP-3 on the SV40 enhancer DNA sequence are from Mitchell *et al*. (1987) and EBP1 from this study. It should however be stressed that these boundaries probably represent an over estimate of the region occupied by the DNA binding proteins, since DNase I is a relatively large molecule and requires an extended binding surface for DNA cleavage (Suck *et al.*, 1988). The effect of mutations in this region on transcription *in vivo* are from Zenke *et al*., 1986. Altered bases in the mutated SV1 oligonucleotides SV1.M1 and SV1.M3, that abolish EBP1 binding, are underlined.