THE REGULATION OF HERPES SIMPLEX VIRUS IMMEDIATE EARLY GENE EXPRESSION

by

Michael Alexander Dalrymple

A Thesis presented for the degree of
Doctor of Philosophy

in

the Faculty of Science,
University of Glasgow

Institute of Virology,
Church Street,
Glasgow G11 5JR.

February 1986

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

Table of Contents

Acknowledgements

I thank Professor J.H. Subak-Sharpe for provision of the facilities in the Institute of Virology and for his overall supervision of this work. I am further indebted to him for employing me on a temporary basis towards the end of the production of this thesis. Without this financial support the end point would have been much longer in arriving.

I am grateful to my supervisor, Dr. Chris Preston, for his constant encouragement and advice, a willingness to listen to my problems and (not least) for introducing me to the game of cricket. His criticism of the manuscript was invaluable.

The sequencing project was only possible thanks to the technical advice and support of Dr. Duncan McGeoch. In this respect, I am also indebted to: Dr. Andrew Davison for advice and communication of results prior to publication, Dr. Phillip Taylor for help with the computing side and Dr. David Bzik. The friendship of all in labs 204 and 200 was much appreciated.

There have been many valuable discussions held with members of the scientific staff and fellow students, but I would particularly like to thank Dr. Fraser Rixon, Dr. Richard Elliot and Dr. Nigel Stow.

Lastly and most of all, thanks to my wife Karen, and daughters Michelle and Nicola, for their patience and endurance. Also to my parents and in-laws for believing I could do it, and to the support of the whole family especially Biff and George.

The author was the recipient of a Training Fellowship from the Medical Research Council. Except where otherwise specified, all the results reported herein were obtained by the author's own efforts.

Summary

The experiments presented in this thesis deal with two aspects of the transcriptional control of herpes simplex virus type 1 (HSV-1) immediate early (IE) gene expression.

Stimulation of the IE gene set is mediated by a protein component of the virion, Vmw65, whose gene has been mapped on the viral genome (Campbell et al, 1984). A specific DNA fragment, BamHI F, containing the entire coding and flanking sequences for Vmw65 has been sequenced using a bacteriophage M13 shot-gun cloning strategy and the dideoxy sequencing technology. The fragment is 8,055 base pairs in size. The mRNA for Vmw65 has been positioned precisely, by nuclease S1 mapping. The predicted open reading frame for Vmw65 consists of 490 codons and translates to a protein of molecular weight 54,342. The protein would appear to have no outstanding physical characteristics, such as regions of extreme hydrophobicity or hydrophilicity, but has significantly more acidic residues than the average protein, especially towards the carboxy terminus. The 5' and 3' flanking sequences of the gene exhibit recognisable signals known to be involved in the regulation of transcription and termination of eukaryotic genes. A homologue to Vmw65 has been identified in the genome of varicella-zoster virus (VZV; A.J. Davison, personal communication). The proteins from the two viruses are highly conserved at the level of amino acid sequence and approximately colinear, although the VZV equivalent is 80 amino acids shorter than Vmw65 at the carboxy terminus.

Three other genes can be detected in the sequence of BamHI F, consistent with the mRNA mapping of Hall et al (1982). The functions of the products of these genes are unknown. All three genes have homology to genes on the VZV genome at the level of amino acid sequence, but the degree of conservation is variable. The region of VZV which is analogous to the BamHI F fragment shows an identical arrangement of genes to HSV-1.

The second aspect of IE gene control which has been investigated is the reported autoregulation of IE gene 3 expression by its product Vmw175. This is based on experiments which showed Vmw175 to be present in lower abundance at early and late times, when compared to IE times, and the observation that a temperature sensitive mutant in Vmw175 (tsK) overproduces IE gene products. The investigation involved the use of plasmid constructs in which the HSV-1 thymidine kinase (TK) was placed under the control of IE promoters. These plasmids were analysed for expression of TK activity after introduction into tissue culture cells, together with the cloned Vmw175 gene from wild type or tsK virus, by calcium-phosphate transfection. The results presented confirm that autoregulation does occur. Polypeptide Vmw175 is able to stimulate expression from the promoters of IE genes 2 and 4/5, but not from its own promoter. This finding suggests that autoregulation may occur indirectly via competitive exclusion of the IE gene 3 promoter from the transcription machinery at post-immediate early times. The reason why Vmw175 cannot activate its own promoter is unclear, however some evidence is presented implicating the enhancer region of IE gene 3. The observation that Vmw175 activates other IE promoters, presumably in an analogous manner to the activation of early and late gene expression, may provide an insight into the role of these IE gene products in the viral lytic cycle.

ABBREVIATIONS

Most of the frequently used abbreviations have been defined in the relevant area of the text, however those which have not are listed below. Abbreviations for buffers and media are given in the Materials and Methods section.

A	adenine
Ad	adenovirus
ATP	adenosine-5'-triphosphate
b, bp	base(s), base pairs(s)
BHK	baby hamster kidney
bis-acrylamide	N,N'-methylene bisacrylamide
Bq	Becquerel(s)
BSA	bovine serum albumin
C	cytosine
Ci	Curie(s)
cpe	cytopathic effect
cpm	counts per minute
dATP	2'-deoxyadenosine-5'-triphosphate
dCTP	2'-deoxycytidine-5'-triphosphate
dGTP	2'-deoxyguanosine-5'-triphosphate
dTTP	2'-deoxythymidine-5'-triphosphate
dNTP	any 2'-deoxyribonucleoside-5'-triphosphate
ddATP	2',3'-dideoxyadenosine-5'-triphosphate
ddCTP	2',3'-dideoxycytidine-5'-triphosphate
ddGTP	2',3'-dideoxyguanosine-5'-triphosphate
ddTTP	2',3'-dideoxythymidine-5'-triphosphate
ddNTP	any 2',3'-dideoxyribonucleoside-5'-triphosphate
DMSO	dimethyl sulphoxide
DNA	deoxyribose nucleic acid
DNAase	deoxyribonuclease
DTT	dithiothreitol
EDTA	sodium ethylenediamine tetra-acetic acid
EtBr	ethidium bromide
G	guanine

g	gram(s)
%G+C	moles per cent guanosine + cytosine moieties
HeBS	hepes buffered saline
hepes	4-(2-hydroxyethyl)-1-piperazine ethane sulphonic acid
hr(s)	hour(s)
IPTG	isopropyl-D-thiogalactosidase
kb, kbp	kilobase(s), kilobase pair(s)
l	litre
M	molar
min(s)	minute(s)
ml	millilitre(s)
mm	millimetre(s)
mM	millimolar
m.o.i.	multiplicity of infection
mol	mole(s)
mRNA	messenger ribonucleic acid
N	unspecified nucleotide (A, G, C or T)
NaOAc	sodium acetate
NPT	non-permissive temperature
OD	optical density
^{32}P	radiolabelled phosphate
PAA	phosphonoacetic acid
PEG6000	polyethylene glycol 6000
p.f.u.	plaque forming units
PI	post infection
PIPES	piperazine-N,N'-bis(2-ethane sulphonic acid)
poly A^+	polyadenylated
R	purine moiety
RNA	ribonucleic acid
RNAase	ribonuclease
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
s	second(s)
SV40	simian virus 40
T	thymine
TdR	2'-deoxythymidine

TEMED	N,N,N',N'-tetramethylethylene diamine
TK	thymidine kinase
tRNA	transfer ribonucleic acid
ts	temperature-sensitive
UV	ultra-violet
V	volt(s)
v/v	volume/volume
Vmw	molecular weight of viral polypeptide in kilodaltons
W	watts
w/v	weight/volume
Xgal	5-bromo, 4-chloro, 3-indolyl, β-D-galactopyranoside
Y	pyrimidine moiety
µCi	microcurie
µg	microgram
µl	microlitre

INTRODUCTION

This introduction is designed to cover areas of knowledge pertinent to the study of gene expression in HSV-1. It is important to examine HSV-1 transcription in the light of what is known about eukaryotic transcription in general, however, the reader will quickly realise that the reverse is also true and the study of the control of viral gene expression has been invaluable in constructing rules for transcriptional control of all genes. Constraints on time and space have forced me to treat the fascinating biology and molecular biology of herpesviruses rather abruptly.

1. Eukaryotic Transcription

1.1 The Structure of Messenger RNA

Production of mRNA in eukaryotic cells is often not simply a matter of transcribing DNA into RNA. Many transcription units are much longer than the final mRNA molecule (Gilbert, 1978), non-coding intervening sequences (introns) being removed, selectively, by a process called "splicing". Furthermore, mRNA molecules are "capped" by the addition of a methylated guanylate residue at the 5' end of the primary transcript (Banerjee, 1980) and 3' polyadenylated, probably via a mechanism similar to splicing (Moore & Sharp, 1984). A diagrammatic representation of these processes is presented in Figure 1. Available evidence suggests that capping and transcription initiation are tightly linked, or coincident, and that the cap has an essential role in translation (reviewed in Banerjee, 1980) and possibly in splicing (Konarska *et al*, 1984). The recent development of *in vitro* cell extracts, which support the accurate splicing of exogenous mRNA precursor substrates, has allowed the elucidation of several steps in the splicing pathway (reviewed by Padgett *et al*, 1985). Such systems should result in a thorough understanding of this complex process, an extreme example of which has recently been revealed by the determination of the structure of the human

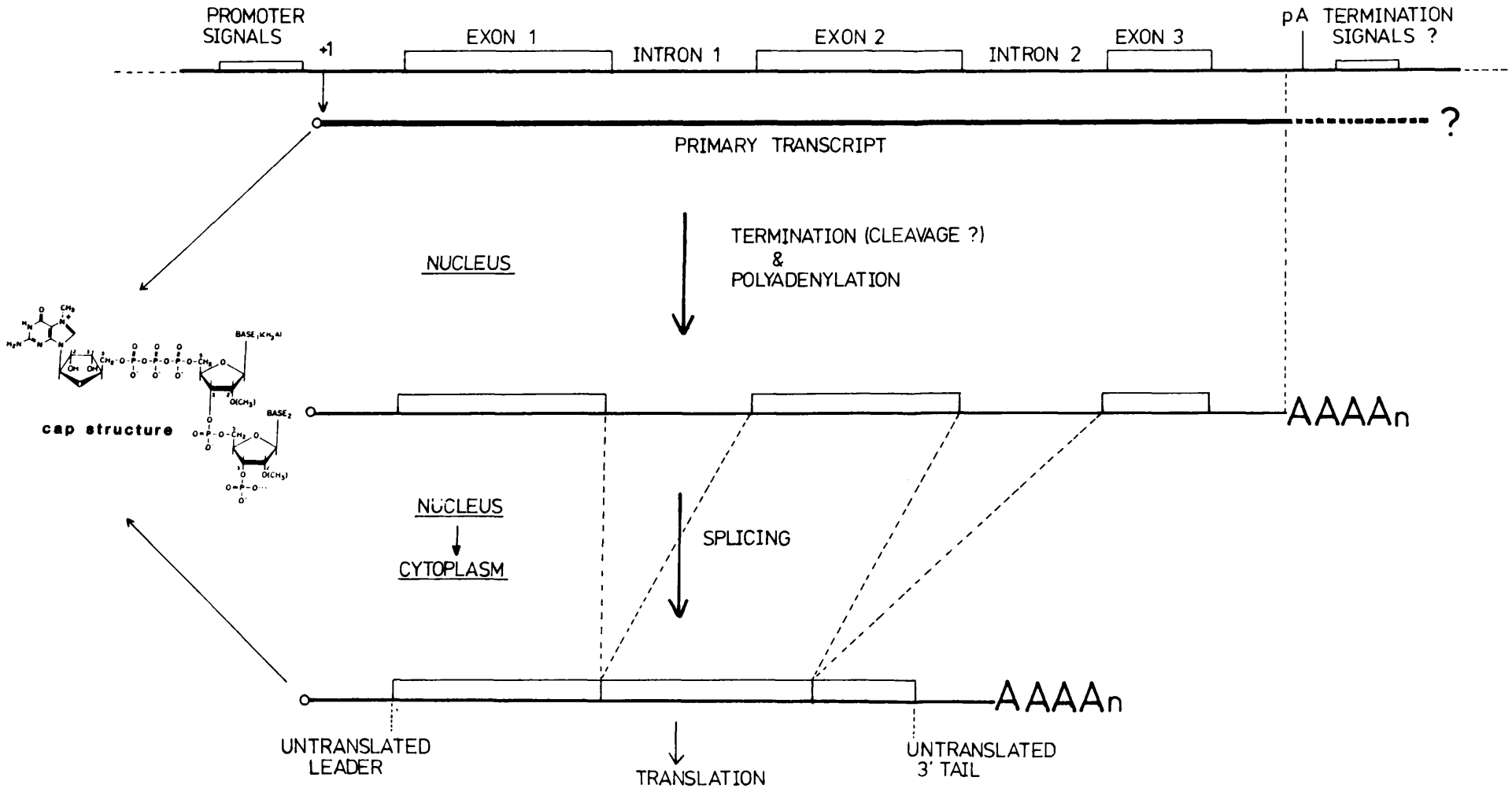

PROMOTER SIGNALS
+1
EXON 1
INTRON 1
EXON 2
INTRON 2
EXON 3
pA TERMINATION SIGNALS ?
PRIMARY TRANSCRIPT
?
NUCLEUS
TERMINATION (CLEAVAGE ?)
&
POLYADENYLATION
cap structure
AAAAn
NUCLEUS
CYTOPLASM
SPLICING
AAAAn
UNTRANSLATED LEADER
TRANSLATION
UNTRANSLATED 3' TAIL

Figure 1: The Structure of Eukaryotic Messenger RNA

This is a diagrammatic representation of the steps involved in post-transcriptional maturation of a typical mRNA. The top line represents the gene and shows the position of 5' (promoter) and 3' (polyadenylation [pA] and termination) flanking signal sequences. Coding sequences (EXONS) are shown as large open boxes, non-coding intervening sequences (INTRONS) are shown as straight lines. The structure of the "cap" is drawn (after Banerjee, 1980) and is represented on the mRNA as an open circle. Polyadenylation is denoted by $AAAA_n$. Splicing probably occurs during, or just before, transport from the nucleus to the cytoplasm.

gene for blood clotting factor VIII. This gene spans 186,000 bp of sequence and consists of 26 exons (ranging in size from 69 to 3,106 bp) with introns as large as 32.6 kb (Gitschier et al, 1984).

1.2 The Transcription Machinery

Eukaryotic cells contain three different RNA polymerase complexes called RNA polymerase I, II, and III. These activities were originally seperated by their elution order from a DEAE-Sephadex column, but they can be more rapidly discriminated by a differential sensitivity to α-amanitin, a bicyclic octapeptide toxin isolated from the mushroom Amanita phalloides. RNA polymerase (pol) I, refractory to α-amanitin, synthesises ribosomal RNA (rRNA) precursors and is concentrated in the nucleolus. RNA pol II is responsible for the synthesis of mRNA, 50% inhibition of its activity can occur at toxin concentrations as low as 0.01 µg/ml. RNA pol III transcribes 5S and tRNA genes and is also sensitive to inhibition by the toxin, but higher levels (5-10 µg/ml) are required. Each of these enzymes is a multi subunit complex, the subunits being classified as large (greater than 100,000 molecular weight) or small (less than 50,000). Only RNA pol III has a subunit of intermediate size. While subunit compositions vary between organisms there are usually 9-14 protein species as revealed by SDS-polyacrylamide gel electrophoresis (Paule, 1981). Studies using monospecific antibodies have revealed shared subunits between the three polymerase classes but little is known of the function of these subunits. Purified RNA polymerases prefer a single stranded DNA template and do not specifically initiate transcription at sites used in vivo (for reviews see Lewis & Burgess, 1982; Sentenac, 1985). The apparent non-specific transcription by the purified enzymes has led to the search for crude cell extracts which mimic the in vivo sequence requirements of selected promoters. These may then be fractionated and semi-purified components added back to purified polymerase in an attempt to reconstitute bone fide promoter recognition.

1.2.1 RNA Pol I

Pol I transcription poses a particular challenge because of its localisation to the nucleolus, a potential control point (Maul & Hamilton, 1967; Roeder & Rutter, 1970), and the possibility that organisation of rDNA in the form of tandem repeats, separated by non-transcribed spacer sequences (Long & Dawid, 1980), plays a role in transcriptional regulation. Despite these reservations, nearly all the present information on cis-acting sequences responsible for promoter selection (discussed in section 1.3.1) has come from the development of in vitro extracts. In vitro systems which support the transcription of rRNA genes have proved the most difficult to obtain. Nevertheless several groups have succeeded in identifying protein factors which play a role in the faithful transcription of these genes. Mishimi et al (1982) were able to distinguish four fractions, from mouse and human cell extracts, which were required for pol I transcription. Fraction C was mainly the polymerase itself, fraction D contained proteins which mediated accurate initiation, B suppressed random initiation at nicks and ends, and finally fraction A enhanced transcription in a reaction containing C and D. It appeared that some factors which determined species specificity resided in fraction D as mouse and human templates absolutely required the homologous fraction. This is similar to work by Miesfeld and Arnheim (1984) in which extracts from HeLa and mouse L cells were fractionated. The authors noted that RNA pol I is unlike II and III in that it appears to have a narrow species range.

1.2.2 RNA Pol III

Pol III systems which operate in vitro to transcribe tRNA and 5S RNA gene templates have been available for a relatively long time. Parker and Roeder (1977) were the first to demonstrate that isolated Xenopus laevis chromatin, but not a naked 5S DNA template, was transcribed efficiently by a purified RNA pol III preparation and thus at least one chromatin bound transcription factor was necessary. Cell

extracts were later produced which could faithfully transcribe cloned DNA templates (Weil et al, 1979b). A protein, of molecular weight 38,600, was subsequently isolated which facilitated 5S DNA transcription in an otherwise non-permissive Xenopus oocyte extract (Engelke et al, 1980). This protein, named TFIIIA, specifically bound to an intragenic region in the gene for 5S RNA and a further series of experiments showed that this binding site lies in a region previously shown to be required for expression of 5S RNA genes in vivo (Bogenhagen et al, 1980; Sakonju et al, 1981; Sakonju & Brown, 1982).

TFIIIA is now the best characterised eukaryotic transcription factor in any class. The factor has been shown to bind stoichiometrically to 5S RNA, which immediately suggests a model for autoregulation of 5S RNA transcription whereby product and promoter are in direct competition for a limiting transcription component (Pelham & Brown, 1980). The protein is zinc dependent (Hanas et al, 1983b) and binds cooperatively to the 5S gene (Hanas et al, 1983a) where it is able to locally unwind DNA (Hanas et al, 1984). Recently a cDNA clone of TFIIIA has been isolated and sequenced (Ginsberg et al, 1984). It has been postulated that the protein contains 7-11 repeated zinc-binding domains which are visible in the amino acid sequence as nine tandem repeats of approximately 30 residues (Miller et al, 1985), suggesting a model for TFIIIA binding much like a caterpillar in that each segment of the elongated protein molecule can independently interact with the DNA sequence and transcription through the region displaces only a small proportion of the binding potential at any moment.

Despite this wealth of knowledge it is clear that TFIIIA is not the sole transcriptional requirement and that at least two other factors play a role in the synthesis of 5S RNA (Lassar et al, 1983; Bieker et al, 1985). These additional factors are required to form a "stable complex", a preinitiation complex of promoter specific proteins and DNA, which is relatively resistant to disruption and is postulated to allow multiple rounds of transcription from

selected promoters and hence maintain expression of specific gene sets (Bogenhagen et al, 1982; Brown, 1984). Furthermore, TFIIIA is not required for the synthesis of tRNA precursors. Several groups have fractionated and reconstituted crude extracts permissive for expression of tRNA precursors (Segall et al, 1980; Ruet et al, 1984; Burke & Söll, 1985). These experiments suggest that there are both gene specific and general factors involved in transcription by RNA pol III.

1.2.3 RNA Pol II

The first cell extracts that supported faithful RNA pol II transcription in vitro, from cloned DNA templates, were prepared by Weil and co-workers (1979a). In this system, a crude cytoplasmic extract was supplemented with purified RNA pol II and it was possible to detect RNA initiation at the well characterised in vivo start site of the adenovirus 2 (Ad2) major late promoter. A different, but related, system was described by Manley et al (1980) in which whole cell extracts of HeLa cells contained enough endogenous RNA pol II and associated factors to faithfully initiate transcription at several Ad2 promoters. These two systems have been extensively used to study the cis-acting sequences responsible for promoting mRNA synthesis and these sequences will be discussed at some length later. In general, however, in vitro extracts have failed to mimic several effects observed when DNA is re-introduced into intact cells, and therefore a brief summary of these results is in order.

Initially it appeared, for a variety of promoters, that only the TATA box (q.v) was required for maximal transcription in vitro (Hu & Manley, 1981; Wasylyk et al, 1980a; Corden et al, 1980; Tsai et al, 1981; Grosveld et al; 1981). Indeed a single base change (T→G) resulted in a drastic decrease in transcription efficiency (Wasylyk et al, 1980a). The first report of upstream sequences modulating transcription in vitro was provided by Tsuda and Suzuki (1981). Experiments were performed on the silkworm fibroin gene either in homologous silk gland extracts or in

heterologous HeLa cell extracts. The authors found that only in the homologous system did sequences upstream of the TATA box modulate the efficiency of mRNA synthesis. Furthermore, in silk gland extracts the fibroin gene was able to compete much better for the available transcription machinery than either the mouse β-globin or the Ad2 major late promoters.

In general, obtaining a response from sequences upstream of the TATA box has proved more difficult *in vitro* than expected and only a few published reports exist, for example it has been shown that the histone H2A far upstream sequences are important for transcription of this gene in HeLa cell extracts (Grosschedl & Birnstiel, 1982). One approach to the study of effects mediated by upstream sequences has been to perform competition experiments between two promoters. In this way, Gregory and Butterworth (1983) compared a trout protamine promoter with that of the mouse β-globin gene and observed inhibitory effects of upstream globin promoter sequences on transcription from the protamine promoter. In a similar way Mishoe *et al* (1984) have shown *in vitro* effects mediated by the SV40 21 bp repeats. Recently, investigators have reported that *trans*-acting factors are responsible for the activity of the SV40 enhancer (q.v.) *in vitro* (Sassone-corsi *et al*, 1984; Sergeant *et al*, 1984; Wildeman *et al*, 1984; Sassone-corsi *et al*, 1985). It is clear that the factor or factors which interact with the SV40 enhancer are not the same as those which interact with promoter regions, that the interaction spans a relatively large sequence area and that other enhancers are able to compete for these factors.

Fractionation of extracts of human cells by chromatography on ion exchange columns has resolved several components that must be added to purified pol II and template in order to reconstitute accurate initiation (Matsui *et al*, 1980; Samuels *et al*, 1982). Like the RNA pol III systems, the requirements seem to reside in only 3 or 4 fractions, one of which supresses random initiation. Further characterisation of these systems resulted in this factor being purified to homogeneity (Slattery *et al*, 1983) and

identified as the well known enzyme poly(ADP-ribose)polymerase. The significance of this observation is unclear but may simply be a consequence of the preferential binding to "nicks" and "ends" of DNA templates by this enzyme. Another fraction, named TFIIE, was associated with a DNA-dependent ATPase activity, a significant observation since hydrolysis of ATP or dATP is a pre-requisite for initiation in vitro (Bunick et al, 1982; Sawadogo & Roeder, 1984). Other workers have independently purified a stimulatory factor, called SII, which interacts weakly with RNA pol II. This is a 40,000 MW phosphoprotein which is essential for transcription, as antibodies to SII inhibited transcription when added to isolated nuclei. A stoichiometric amount of SII was shown to be bound to the polymerase during chain elongation (Ueno et al, 1979; Sekimizu et al, 1982; Horikoshi et al, 1984) suggesting a role in chain extension rather than initiation.

At least two promoter-specific transcription factors have been isolated and purified. Sp1 is a component of HeLa cells which was initially identified as being a specific activator of the SV40 early and late promoters in vitro (Dynan & Tjian, 1983a). A second fraction, named Sp2, was a more general transcription activator. The authors concluded that Sp1 was involved in promoter selection by RNA pol II. Sp1 has been shown to bind to the 21 bp repeated region of the SV40 early promoter by use of DNAase I footprinting (Dynan & Tjian, 1983b) and to specifically bind to one strand of the DNA duplex in the sequence GGGCGG, for both SV40 and a related monkey promoter (Gidoni et al, 1984). Recently, Sp1 was shown to bind to the HSV-1 promoters for IE gene 3 (Jones & Tjian, 1985) and thymidine kinase (TK; Jones et al, 1985) and to be required for their transcription in vitro. In the case of IE gene 3, only the Sp1 site nearest the mRNA 5' terminus was absolutely required in vitro, but the presence of upstream sites enhanced the reaction by a factor of 3-5 fold.

A factor (HSTF, for heat-shock transcription factor), which acts in an apparently analogous way to Sp1, has been

isolated from Drosophila melanogaster cells (Parker & Topol, 1984b). This factor is involved in selective activation of heat shock genes and binds specifically to a 55 bp region upstream from the TATA box, in the absence of RNA pol II. Another factor, named the B factor, binds specifically to the TATA box region of several promoters (Parker & Topol, 1984a). Both HSTF and B can bind to the hsp70 gene at the same time, however, once cells have been heat shocked, binding of the B factor is reduced. It may be that HSTF is a B analogue or that residual B protein is present in sufficient quantity to allow transcription from promoters marked by the presence of HSTF. Davison et al (1983) have also described a TATA box binding activity from HeLa cells. It is not known whether this is equivalent to the B factor in Drosophila.

In conclusion, several protein factors have been isolated, and in some cases purified, which act to limit RNA pol I, II and III transcription initiation to bone fide promoter sequences. The working hypothesis would appear to be that promoter sequences will fall into a limited number of groups (for example genes for house keeping enzymes or heat shock genes) which will be distinguished by a specific set of factors and thus co-ordinately regulated. As already described for genes transcribed by RNA pol III, one possible mechanism by which such selective activation could be initiated and maintained is via the formation of stable complexes. Such complexes have been reported in RNA pol II systems (Fire et al, 1982; Davison et al, 1983; Ackerman et al, 1983; Cullota et al, 1985).

1.3 Cis-acting Sequences

Most of the results and conclusions discussed in this section have come from an experimental strategy known as "reverse" or "surrogate" genetics. While individual experimental details differ, the general approach has two stages. The first is to mutagenise DNA sequences in vitro, by means of standard recombinant DNA techniques. Deletion mutants are the easiest to construct and can quickly define

regions of interest. Several groups have taken such analysis to its logical conclusion and studied point mutations, but this requires a correspondingly larger input of time and materials. The second stage is to re-introduce these mutated sequences into an environment in which they have the potential to be transcribed. More often than not this is by Calcium Phosphate-mediated transfection into tissue culture cells (Graham *et al*, 1973b). Alternatively, microinjection into tissue culture cells and *Xenopus* oocytes or analysis of transcription in cell-free extracts have been used.

1.3.1 *RNA Pol I*

Information about *cis*-acting transcription regulatory sequences for this gene class has almost totally been derived from the use of *in vitro* transcription systems. Grummt (1982) was the first to report a promoter analysis by investigation of transcription from deleted and non-deleted mouse rDNA templates *in vitro*. Her results showed that sequences between -33 and -13, with respect to the RNA start site, were absolutely required for efficient initiation and it was suggested that these sequences were functionally analogous to the TATA box (q.v.) of RNA pol II promoters. Competition experiments between wild-type and mutant templates suggested that sequences upstream may also exert an influence. These findings have essentially been repeated by other workers (Kohorn & Rae, 1982; Learned *et al*, 1983). The -33 to -13 region contains sequences conserved among mouse, rat and human rDNA genes: ATCTT (-38 to -33) and TATTG (-20 to -16). A G→A base change at position -16 reduced transcription by 95% suggesting that this position is of importance (Skinner *et al*, 1984). It was also noted that extracts made from mouse, human and protozoan cells were unable to transcribe heterologous rDNA templates, showing that rRNA synthesis *in vitro* required species-specific transcription factors (Grummt *et al*, 1982).

Moss (1982) has attempted to map, *in vivo*, the promoter of a *Xenopus laevis* rDNA gene after microinjection into *X. borealis* oocytes. It was found that only sequences

between -145 and +16 were required for accurate initiation and that promoter elements lay between -13 to -30 and -36 to -61, correlating well with the results from *in vitro* systems. Smale and Tjian (1985) have recently attempted to apply an approach classically used for genes transcribed by RNA pol II to identify promoter elements for RNA pol I transcription. The system consists of human rDNA 5' flanking sequences fused to the coding sequences of the gene for HSV-1 TK and expression assayed after transfection into tissue culture cells. Several interesting observations were made: (a) transcription of the TK gene was resistant to α-amanitin, although some RNA pol II transcription could be detected, (b) previously reported *in vitro* species-specificity of RNA pol I was similarly observed *in vivo*, (c) the RNA pol I promoter consisted of at least two elements - a "core" of about 40 bp overlapping the rRNA initiation site and a second, stimulatory, element between -234 and -131, and (d) the SV40 enhancer did not stimulate RNA pol I transcription from the rDNA promoter. This latter finding appears to be incompatible with models of enhancer action (q.v.) which involve only the formation of an open chromatin structure.

1.3.2 *RNA Pol III*

The first successful application of deletion analysis to the determination of promoter requirements was performed by Brown and his colleagues on the somatic 5S RNA gene of *Xenopus laevis.* Two series of deletions were constructed from either the 5' or 3' side of the gene and their activities in an *in vitro* transcription system determined (Sakonju *et al*, 1980; Bogenhagen *et al*, 1980). In this way the authors were able to show that intragenic sequences, between +50 and +83, directed RNA pol III to initiate transcription. This work correlates well with the discovery and characterisation of a transcription factor, TFIIIA, which binds to the essential region and helps in the formation of a stable transcription complex (see section 2.2).

A similar resection approach has revealed that tRNA genes also contain intragenic promoters, with boundaries between +8 and +62 (Hofstetter et al, 1981; Sharp et al, 1981; Galli et al, 1981; Ciliberto et al, 1982a). However, unlike the 5S gene promoter, that of a typical tRNA gene is discontinuous being split into two blocks (A and B) with approximate coordinates +8 to +19 and +52 to +62 respectively. Two lines of evidence support this conclusion: first, chaemeric tRNA genes, in which the 5' and 3' portions originate from different genes, are efficiently transcribed; and second, transcription can occur after replacement of the central region (between blocks A and B) with unrelated DNA sequences. It has been shown, however, that blocks A and B require a critical seperation distance which is between 30 and 40 bp (Ciliberto et al, 1982b). Efficient transcription of certain insect tRNA genes, in vitro, requires 5' flanking sequences (Dingerman et al, 1982; Sprague et al, 1980; Larson et al, 1983) and if extended to the in vivo situation would seem to be a major difference between 5S and tRNA genes.

1.3.3 RNA Pol II

One of the earliest approaches to identifying sequences involved in the control of transcription by RNA pol II was to look for conserved sequence patterns in the predicted regulatory regions of various genes. This approach had been successful in prokaryotes (Rosenberg & Court, 1979; Siebenlist et al, 1980). Initial searches revealed an AT-rich region of homology (called the "TATA" Box), first identified by Goldberg and Hogness, centred around 25-30 bp upstream of the mRNA cap site (Proudfoot, 1979; Gannon et al, 1979). A second region of homology was identified in the area around -70 to -80 of several cellular and viral protein coding genes and was named the "CAAT" box (Benoist et al, 1980; Corden et al, 1980). The nucleotide sequence around the cap site of sea urchin histone genes is also conserved (Busslinger et al, 1980) and it has been suggested that transcription initiates at a purine (R) flanked by a

pyrimidine (Y), often CAY (Corden et al, 1980). It should be noted, however, that the conservation of these sequences is far from absolute, despite their widespread occurrence, and the only true test of function is to introduce a mutation and assay for expression. Experiments of this type will now be reviewed.

1.3.3.1 The TATA box

As previously stated, this region was identified as a putative regulatory element due to its relatively high degree of conservation between many different protein coding genes. Its functional significance was first shown unequivocally in vitro for the chicken conalbumin gene. Upon introduction of a single T→G change in the third position of the sequence TATAAAA, a 20-fold drop in the level of mRNA synthesis was observed, and a T→A change in the same position resulted in a 40-fold drop in transcription (Wasylyk et al, 1980a). Mathis and Chambon (1981), working on the SV40 early promoter, reported that the TATA box was essential and functioned in vitro to position initiation within a narrow region, consistent with its conserved position with respect to the cap site. The sequence was shown to perform a similar function in vivo (Benoist & Chambon, 1981), although deletion of the region resulted in production of novel multiple initiation sites not usually seen with wild-type templates, a similar finding to that reported in the case of the sea urchin histone 2A gene (Grosschedl & Birnstiel, 1980). It is now believed that the discrepancy between in vitro and in vivo experiments reflected an inadequacy in the fidelity of the in vitro systems, as has been discussed. The general conclusions, that altered TATA boxes result in lowered transcriptional efficiency and that deletion of the same sequences leads to production of heterologous 5' termini, are supported by many independent experiments (Gluzman et al, 1980; Ghosh et al ,1981; Grosveld et al, 1981; Lebowitz & Ghosh, 1982; Grosveld et al, 1982). Furthermore, deletions 3' to the TATA box of the rabbit β-globin gene resulted in initiation

occuring at new sites but still approximately 30 bp downstream (Grosveld et al, 1981), consistent with the hypothesis that a factor or factors bind to the TATA region and direct the transcription complex to an initiation site at a fixed distance downstream.

It is now clear that the precise sequence which constitutes a "TATA" box is relatively variable. Several genes lack "obvious" homology to the classical consensus (G-GTATA(A/T)A(A/T)-G; Breathnach & Chambon, 1981), for example the SV40 late-early promoter (Ghosh & Lebowitz, 1981), the SV40 late promoter (Haegeman & Fiers, 1978) and the Ad2 EIIA and IVA2 promoters (Baker & Ziff, 1981). However, when sequences in the equivalent region are subjected to mutagenesis, it is apparent that they exhibit some of the properties associated with more conventional TATA regions. For example, in the SV40 late promoter lies a sequence GGTACCTAACC, 21-31 bp upstream of the major late start site. C→T substitutions, which increased the homology of this sequence to the TATA consensus, showed increased levels of mRNA initiated at the major late position (Brady et al, 1982). Similarly, an element has been found in the EIIA region of Ad2 which appears to function by positioning the mRNA cap sites, albeit rather less tightly than a classic TATA box (Zajchowski et al, 1985; Goding et al, 1985). Both of these results are consistent with the concept of a spectrum of "TATA" box elements which exhibit differing degrees of control over the site of initiation. The kinds of gene sequenced to date have been those which have been relatively easy to clone, for example viral genes and those for inducible proteins whose mRNAs and products accumulate to high levels. There must always be some doubt as to whether these genes are "typical" or represent a specific sub-class with peculiar properties. Recently the isolation and sequence analysis of some "housekeeping" genes has been reported. Reynolds et al (1984) have isolated the hamster gene for 3-hydroxy, 3-methylglutaryl coenzyme A reductase (HMGCoA reductase), a trans-membrane glycoprotein involved in cholesterol biosynthesis. This gene is transcribed from

multiple initiation sites and apparently lacks a TATA homologue. The gene for hypoxanthine phosphoribosyl transferase (HPRT; Melton et al, 1984) similarly lacks an obvious TATA box, but unlike HMGCoA reductase has only a single predominant 5' mRNA terminus. Finally the gene for human adenosine deaminase (ADA) has been isolated (Valerio et al, 1985). ADA has promoter activity within 135 bp of the cap site but the only possible TATA homologue is TTAA at -27. All three promoters of these housekeeping genes exhibit a very high %G+C (82% in the case of ADA) and therefore have GC-rich motifs which have been shown, in other systems (see below), to be involved in promoter function. These observations raise the interesting possibility that there exist sets of cellular genes differentiated by specific kinds of promoter.

1.3.3.2 Other Promoter Elements

It is in the search for so-called promoter elements that the concept of constitutive versus induced expression, taken from prokaryotic expression studies, has been invoked. Simply stated, it has long been held that all genes have sequence elements (the promoter) which will allow a certain basal level of expression, superimposed upon this feature would be found additional sequences which would confer the ability to be regulated up and down and these elements would be easily separable from the basal controls. As will become clear, this delineation is increasingly difficult to define as it has been shown that at least some proteins (trans-activators) probably induce transcription to high levels by interacting with proteins which bind to basal promoter elements (Feldman et al, 1982; Green et al, 1983) and not directly with special sequences. This is especially true in the viral situation, e.g. adenovirus (q.v.) and herpesvirus (q.v.), and therefore analysis of so-called constitutive expression of many viral genes, in the absence of viral stimulatory factors, is artificial.

I. GC-rich motifs

Elements of the form CCGCCC or GGGCGG have been shown to be important for promoter function in several systems. Their relevance was first identified in the case of the HSV-1 TK gene promoter and in the 21 bp repeated sequences of SV40, and it is in these two systems that the most intensive mutagenic studies have been performed. The SV40 promoters are discussed in detail in section 1.4.2.1.

The HSV-1 TK promoter has been dissected by a method known as linker-scanning (LS) mutagenesis (McKnight & Kingsbury, 1982). Two series of sequential deletions are made from opposite ends of the DNA region of interest. These deletion mutants are designed such that they are closely spaced and terminate in an oligonucleotide linker containing a unique restriction site. By combining suitable members of the two series it is possible to produce a third series of constructs in which the number of nucleotides in the linker replace, exactly, the same number of wild-type nucleotides. The sequence in this region therefore differs from the wild-type gene only by the linker substitution. This process, which essentially introduces clusters of point mutations, resulted in the identification of three areas important for transcription in Xenopus oocytes - an in vivo system generally regarded as detecting constitutive promoter activity. One of these regions corresponded to the TATA box, which has been discussed previously, the other two were located between -47/-61 (the first distal signal) and between -80/-105 (the second distal signal) (McKnight et al, 1981; McKnight & Kingsbury, 1982; McKnight & Gavis, 1980). Double LS mutants were constructed to determine whether the signals acted synergistically or independently (McKnight et al, 1981). These constructs showed that the mutation of one distal signal substantially reduced the contribution of the other, suggesting that the two signals act in a single step, possibly binding a dimeric transcription factor. Results from experiments designed to examine how this interaction operates suggest that the two sequence elements bind a transcription factor or factors in a co-operative manner and

that there is no direct intra-strand interaction, for example by formation of a stem-loop structure (McKnight et al, 1984). It has been shown that the transcription factor Spl (q.v.; Dynan & Tjian, 1985; Jones et al, 1985) binds to the first and second distal signals in the TK promoter. LS mutants in these signals reduce, or abolish, the ability of Spl to bind and so it seems likely that this factor is involved in the transcriptional control of the TK gene.

Many other promoters contain such GC-rich motifs: cellular genes such as human metallothionein-I_A (Richards et al, 1984), hamster HMGCoA reductase (Reynolds et al, 1984), mouse HPRT (Melton et al, 1984) and many viral genes, but more detailed mutagenic analyses are required to implicate the sequence in the expression of these genes.

II. The CAAT Box

This structure was identified as being conserved between several different promoters, in the region -70 to -80, and has canonical sequence 5'-GG(C/T)CAATCT-3' (Benoist et al, 1980; Corden et al, 1980; Efstradiatis, 1980). Functional analysis has been performed by deletion of the sequence and assay for the production of mRNA in a variety of systems. In most cases deletion of the CAAT box results in a significant, but not drastic, reduction in mRNA levels, for example in the following cases: rabbit β-globin (Dierks et al, 1981; 1983; Grosveld et al, 1981; 1982), the Ad2 major late promoter (Hen et al, 1982) and the human α-globin promoter (Mellon et al, 1981). However, deletion of the structure from the sea urchin histone 2A promoter actually results in an increase in transcription (Grosschedl & Birnstiel, 1982). Initial results suggested that the CAAT homology present in the HSV-1 TK promoter was not required for maximal expression (McKnight, 1982), but recent studies using in vitro transcription systems suggest that the CAAT box is the binding site for a factor important in promoter function (Jones et al, 1985). In general, it is true to say that the CAAT box has been less well studied than the TATA box and there is yet no clearly defined role for it.

III. Enhancers

The upstream sequences known as enhancers have been defined by the properties which initially set them apart from other promoter elements. These sequences have several special features:

(a) they stimulate RNA pol II transcription, by many fold, from both homologous and heterologous promoters.

(b) this effect can be seen even when the sequences are placed at a considerable distance from the activated promoter.

(c) the stimulation can occur in an orientation independent manner.

(d) cell-type specificity.

(e) there is some evidence that the increased transcription can be seen even when the enhancer is placed 3' to the promoter.

The first enhancer to be described lay within the 72 bp repeated elements located between the early and late promoters of SV40. Deletion of one of the two repeats was not lethal and viable viral progeny were still produced, but deletion of both elements abolished transcription from the early promoter and the subsequent production of virus (Benoist & Chambon, 1981; Gruss et al, 1981). It was quickly shown that this element could also positively influence transcription from other promoters (Banerji et al, 1981; Moreau et al, 1981) including those for rabbit β-globin, chicken conalbumin and the Ad2 major late transcription unit. The effect of stimulating transcription from heterologous promoters was observed even if the orientation of the 72 bp repeats was inverted with respect to the cap site. Furthermore, it was not absolutely dependent upon the position of the repeat (Moreau et al, 1981; Fromm & Berg, 1983; Wasylyk et al, 1983a). Similar experiments revealed that proximal promoters were activated in preference to more distal ones (Wasylyk et al, 1983a; de Villiers et al, 1982). This effect was even seen on plasmid vector sequences which had similarity to eukaryotic promoter elements. An analysis of the 72 bp repeat by point mutagenesis revealed only one

critical nucleotide, which is underlined, in the sequence GGTGTGGAAAG (Weiher et al, 1983), although G→A transitions for the first three G residues resulted in a decrease in activity. Using this short sequence, weak homologies were found to exist in the control region of Moloney murine sarcoma virus (MSV), the human papovavirus BKV and polyoma virus (Weiher et al, 1983). Functional analysis of the MSV control region had previously shown that the relevant sequence could replace the SV40 72 bp repeat and yield viable recombinant virus (Levinson et al, 1982). Close examination of several other RNA and DNA viruses has demonstrated the presence of enhancer-like elements, e.g. in Rous sarcoma virus (RSV; Luciw et al, 1983), polyoma virus (De Villiers & Schaffner, 1981; Tyndall et al, 1981), BK virus (Rosenthal et al, 1983), the adenovirus EIA gene (Hearing & Shenk, 1983; Weeks & Jones, 1983), bovine papilloma virus (Campo et al, 1983), human cytomegalovirus (HCMV; Boshart et al, 1985), MSV (Jolly et al, 1983), HSV-1 (Lang et al, 1984; Preston & Tanahill, 1984) and herpesvirus saimiri (Schirm et al, 1985).

A variety of experiments have shown that several enhancers exhibit a cell or tissue preference, suggesting that this might constitute one determinant in the host range of the above viruses. SV40/MSV recombinants showed less activity in monkey kidney cells than mouse fibroblasts, the natural host for MSV (Laimins et al, 1982). A similar conclusion was drawn from experiments with SV40/Harvey murine sarcoma virus recombinants (Kriegler & Botchan, 1983) and SV40/murine leukaemia virus recombinants (Celander & Haseltine, 1984). De Villiers and co-workers (1982) have shown SV40 and polyoma enhancers to have a cell type preference, as have Berg et al (1983). These observations immediately suggested that viruses were utilising cellular transcription networks and that it should be possible to identify cellular enhancers. One strategy was to use SV40 or BK virus enhancers as probes to screen genomic libraries (Conrad & Botchan, 1982; Rosenthal et al, 1983). Such work identified homologous sequences with the requisite enhancer

properties, although apparently less effective than their viral counterparts. Another approach to locating cellular enhancers used direct selection. Fried et al (1983) located sequences in transformed cell lines near the integration site of an enhancer-defective polyoma virus. Schaffner's group have cloned random fragments of cellular DNA into an SV40 vector which lacks an enhancer - the so-called "enhancer trap". Interestingly, they discovered that by a simple duplication of the residual promoter region enhancer function was restored (Weber et al, 1983).

Finally, enhancer elements have been defined in well characterised cellular genes. The first such enhancer was found to reside in the intron separating the variable (V) region, joint (J) segment, and the constant (C) region of the mouse immunoglobulin heavy chain genes (Gillies et al, 1983; Neuberger, 1983; Banerji et al, 1983). Similarly, enhancers have been found in the kappa light chain gene (Queen & Baltimore, 1983). These enhancer elements share the property of being extremely tissue-specific and are only active in B cell lineages, not in non-B cells. Competition experiments have suggested the existence of one or more tissue-specific protein factors (Mercola et al, 1985) and these have been indirectly visualised, binding to the DNA sequences, by dimethyl sulphate protection studies and genomic sequencing (Ephrussi et al, 1985).

There has been much speculation as to the mechanism by which enhancers operate. One early hypothesis, which attempted to explain the bidirectional and long distance nature of the effects, was that they acted as chromatin organisers and there is evidence in support of this model (Jongstra et al, 1984). However, it is interesting to note that enhancers are unable to stimulate expression of genes transcribed by RNA pol I (Smale & Tjian, 1985) suggesting that it may not simply be due to open chromatin structure. An alternative model is that enhancers are recognised by factors which facilitate the formation of stable transcription complexes. Enhancer-like sequences in the Xenopus U2 small nuclear RNA gene have been shown to operate

by such a mechanism (Mattaj et al, 1985), and two independent groups have demonstrated that the concentration of RNA pol II is increased on genes driven by enhancers (Weber & Schaffner, 1985; Treisman & Maniatis, 1985). It seems likely that the mechanism of enhancer action and the basis for the cell specificity is by their differential affinities for endogenous cell factors.

1.4 Inducible Promoter Elements

1.4.1 Cellular Genes

The "classical" enhancers, such as those of SV40, polyoma virus and the immunoglobulin genes, appear to act on an all or nothing basis. It is clearly necessary for some genes to be regulated over a wider spectrum and respond to various stimuli. Many such genes have been described and studied, and a small selection of the most detailed analyses are presented below.

The phenomenon of heat induction of a small number of genes, presumably coordinately regulated, has been described in virtually every organism studied and has been termed the heat shock response (Schlesinger et al, 1982; Ashburner & Bonner, 1979). In each case the new proteins are synthesised during an otherwise global inhibition of protein synthesis. Except in one case (Bienz & Gurdon, 1982), this regulation has been shown to occur at the level of transcription (Bienz, 1985). Several heat shock genes have been cloned and sequenced. The most extensively studied has been the hsp (heat shock protein) 70 gene of Drosophila melanogaster. Re-introduction of an hsp70-lac Z hybrid gene into flies suggested that the 5' flanking sequences to -194 were required for co-ordinate induction (Lis et al, 1983). It was observed that hsp70 was still induced by heat even when placed in heterologous environments such as mouse, monkey or frog cells (Corces et al, 1981; Pelham, 1982; Mirault et al, 1982) suggesting that the induction mechanism has been highly conserved. This fact has been used to systematically mutate and assay the hsp70 promoter in monkey COS cells and Xenopus oocytes, revealing that only sequences from -47 to

-66 are absolutely required for induction (Pelham, 1982; Mirault et al, 1982; Bienz & Pelham, 1982). This region contains an 11 bp element, with dyad symmetry, which is conserved in nearly all heat shock genes and has a consensus sequence C-GAA-TTC-G, called the heat shock element (HSE). This element is sufficient to confer inducibility on a completely non-responsive gene (Pelham & Bienz, 1982). Subsequent studies identified a transcription factor (HSTF) which specifically binds to the HSE in vitro (Parker & Topol, 1984b). Interestingly, this factor could be isolated from both heat shocked and normal cells, suggesting that the protein must somehow be activated by heat shock. In a totally different approach Wu (1984) demonstrated that two factors bound to hsp promoters in isolated nuclei. One, presumably HSTF, bound to heat shock promoters during heat shock but not detectably before, while a second factor bound tightly to the TATA box region under both conditions. Several of the known heat shock genes have multiple HSEs which appear to be functional (Bienz, 1985) and it has been argued that multicopy expression systems such as COS cells, and oocyte injection, are artificial and that, in fact, far upstream sequences may be required for optimal expression in a situation where only one or a few gene copies are present. Dudler and Travers (1984) sought to overcome this objection by introducing only a small number of copies of promoter deleted genes into flies, using P element transformation (Spradling & Rubin, 1982). Their results showed that 97 bp of 5' flanking sequences were required for induction of the genes to wild-type levels. This conflicts with previous results which suggested that sequences to -66 would suffice, but is in agreement with the biochemical studies on the binding of HSTF. A sequence with an 80% match to the -66/-47 HSE lies between -72 and -85, suggesting that full induction of hsp 70 in situ may require binding of two molecules of HSTF, perhaps co-operatively.

The metallothionein (MT) genes comprise a small multigene family which responds, at the transcriptional level, to heavy metal ions such as those of zinc and cadmium

(Karin et al, 1980; Durnam & Palmiter, 1981), to glucocorticoid hormones (Karin & Herschman, 1979; Hager & Palmiter, 1981) and to interferon (Friedman et al, 1984). The function of the MT proteins is less clear but may involve storage of zinc, heavy metal detoxification or free radical scavenging (reviewed by Karin, 1985). Brinster et al (1982) were able to show that the 5' flanking sequences responsible for heavy metal induction of the mouse MT-I gene lay in the first 90 bp upstream of the cap site. A more detailed analysis revealed that a proximal metal regulatory element (MRE) lay between -59 and -46, and on the basis of homology with human MT-I and MT-II genes it was postulated that the MRE sequence was CCTTTGCGCCCG (Searle et al, 1984; Stuart et al, 1984; Karin et al, 1984a, b; Richards et al, 1984). However, multiple copies of the MRE (named a-e in order of increasing distance from the cap site) can be discerned (Carter et al, 1984; Karin et al, 1984b) and it was not clear whether they all contribute to induction in situ. Searle et al (1985), making use of the unique BamHI sites in McKnight's TK LS mutants (q.v.), have studied the effect of inserting MRE sequences at many different points within the TK promoter and in combinations. The authors found that two MRE sequences were required to obtain significant induction and that inducibility could be increased by addition of more copies. Furthermore, the MRE was most effective in constructions retaining one or more of the TK distal elements and was relatively sensitive to distance effects, suggesting that it is not enhancer-like.

Human α- and β-interferon (IFN) genes are expressed by many cell types as a response to viral infection (reviewed in Lengyel, 1982). IFN- genes are also induced by synthetic double stranded RNA, e.g. poly(rI).poly(rC). The increase in expression of IFN is due, at least in part, to an increase in the rate of transcription (Raj & Pitha, 1983; Nir et al, 1984; Hiscott et al, 1984a), and induction must be mediated by pre-existing factors as the response does not require protein synthesis (Cavalieri et al, 1977).

In the case of IFN-α Weissman's group have clearly shown that the 5' flanking region mediates viral induction and that this induction is transcriptional (Weidle & Weissman, 1983). They also demonstrated that not more than 117 bp of the flanking region was required for inducible expression and noted that it contained a purine-rich sequence which is highly conserved amongst the known IFN-α genes. Recently, this group reported that a 46 bp promoter segment from IFN-α rendered an unrelated β-globin promoter inducible by Newcastle Disease virus. This segment lies between -109 and -64 (Ryals et al, 1985).

While the IFN-α and IFN-β genes are clearly related by virtue of their homology, both in coding and 5' flanking sequences, there is evidence to suggest that they are not co-ordinately regulated. After induction of fibroblasts, only IFN-β mRNA is produced, whereas leukocytes express predominantly IFN-α mRNA (Hiscott et al, 1984). Experiments designed to examine the sequence requirements for IFN-β induction give contradictory results. Maroteaux et al (1983b), using an SV40-based vector in a transient expression system, concluded that only the TATA box and cap site are absolutely required. In these studies, however, cycloheximide was used together with poly(rI).poly(rC) for optimal induction and the relevant controls were not reported which is puzzling since the same workers have also shown that cycloheximide alone can stimulate expression of IFN-β (Maroteaux et al, 1983a), and thus these workers cannot distinguish the two effects. From experiments in which IFN-β cDNA clones were stably introduced into mouse Ltk$^-$ cells it was concluded that only the coding sequences were absolutely required for induction (Pitha et al, 1982). Tavernier et al (1983) monitored IFN activity in monkey cells transfected with SV40/IFN-β plasmids. By deletion analysis of the 5' flanking sequences they were able to show that the region between -186 and -144 was required for induction and noted homology in this region to sequences involved in steroid hormone responses, however IFN-β mRNA levels were not measured. Zinn and co-workers (1983) studied

the regulation of IFN-β genes introduced into mouse cells on a bovine papilloma virus vector. Cell lines were derived which contained many copies of the hybrid constructs, but which were maintained extra-chromosomally. In this system, sequences to -107 were shown to be responsible for maximal constitutive expression and also induced expression. A sequence GAGA (-77 to -74) was part of this constitutive promoter but its deletion had a relatively greater effect on induced expression than other mutations and the authors argued that it formed part of an inducer element. Evidence was also provided for the existence of a second element between -210 and -107 which acted to depress constitutive expression and alter the kinetics of induction, suggesting that it could be the site of action of a cellular repressor.

Clearly the use of different systems to determine the induction sequences might explain some of the discrepancies. The suggestion of direct cellular inhibition of transcription from the IFN promoter is consistent with the observed stimulation of expression by cycloheximide. Further experiments revealed that the region from -37 to -77, called the interferon regulatory element (IRE), acted like an inducible enhancer in that it could act regardless of orientation and in a relatively distance independent manner (Goodbourn et al, 1985). A minimal regulator of 14 bp lay within this region (5'-AA(A/T)A--GAAAGGR-3') between -77 to -64 and is repeated five times within the 5' flanking sequences of both α- and β-interferon genes, which might explain the inability to completely abolish induction.

The above examples serve to show that one general mechanism for the co-ordinate induction of cellular genes is by the interaction of common regulatory molecules with conserved cis-acting sequences elements. These elements tend to be present in multiple copies which are often degenerate and probably have different affinities for their putative regulators. This duplication may be a simple evolutionary mechanism by which single copy genes have increased their ability to compete for regulatory molecules. To date there are few examples where the regulatory molecules have been

identified, but two such systems are the co-ordinate induction of enzymes involved in certain yeast metabolic pathways and mammalian steroid hormone responsive genes.

In the yeast Saccharomyces cerevisiae a number of genes have been described which appear to be co-ordinately regulated by means of cis-acting, so called, upstream activation sites (UAS). For example, transcription of at least three genetic loci involved in galactose utilisation, GAL1, 7 and 10, fail to be induced in strains harbouring recessive mutations at the GAL4 locus, which suggests that the GAL4 gene product is the trans-inducer of galactose induction (Johnston & Hopper, 1982; Laughon & Gesteland, 1982). The GAL10 gene has sequences more than 100 bp upstream of the TATA box which mediate induction of transcription by galactose (Guarente et al, 1982). This element also exhibits properties one might expect from an inducible enhancer in being orientation independent (Struhl, 1984) and is able to confer galactose induction on the usually non-responsive CYC1 gene, an effect still dependent on a functional GAL4 gene product.

A similar network appears to exist for heme induction of certain yeast genes. The gene CYC1 has two UAS sequences, at about -275 and -225, that respond to heme, and mutations at the HAP1 locus reduce heme activation, suggesting that the HAP1 gene product mediates the induction process. Again these UAS regions are able to mediate induction even when placed upstream of the heterologous LEU2 promoter (Guarente et al, 1984). It will clearly be of interest to further characterise these two regulatory genes and to determine their mode of action.

Steroid hormones bind to specific receptor proteins which are then translocated from the cytoplasm to the nucleus. These "activated" receptors are able to bind DNA (Mulvihill et al, 1982; Payvar et al, 1981). One well studied example is the hormonal regulation of expression from the retrovirus mouse mammary tumour virus (MMTV). Stably integrated provirus can be stimulated to produce mRNA

by the hormone glucocorticoid (Ucker et al, 1981; Ringold et al, 1977), and this transcription initiates from a unique site in the long terminal repeat (LTR; Hynes et al, 1981). Purified glucocorticoid receptor selectively binds to MMTV DNA in vitro (Payvar et al, 1981; 1983; Scheidereit et al, 1983) and it has been shown that this is due to the presence of a "glucocorticoid response element" (GRE) in the LTR which activates adjacent promoter sequences (Geisse et al, 1982; Chandler et al, 1983). Similar work has been done on two other glucocorticoid responsive genes - the human metallothionein-II gene (Karin et al, 1984b) and the chicken lysozyme gene (Renkawitz et al, 1984). This has led to the proposal of a consensus glucorticoid-dependent enhancer of sequence: 5'-(C/T)GGT-(A/T)CA(A/C)(A/T)-TGT(C/T)CT-3'. Homologues to this sequence are also present in two other glucocorticoid responsive genes - the rat growth hormone and human proopiomelanocortin genes (Cochet et al, 1982).

The last example serves to describe the most likely model for specific activation of genes: direct binding of regulatory molecules to specific DNA sequences in the 5' flanking region of the target gene. These activator molecules must then somehow make available the promoter for RNA pol II transcription or mark it as "activated". The following description of two viral systems suggest that this is not the only stimulatory mechanism.

1.4.2 SV40 and Large T Antigen

The simian virus 40 is a member of the papovavirus family. It has a small double stranded genome, which has been completely sequenced, and consists of 5243 bp (Fiers et al, 1978). Other features which have made it an attractive virus to study are: it is readily propagated in tissue culture, producing large yields of virus (Tooze, 1980), in certain cell types the virus has the capacity to initiate transformation in vitro (Shein & Elders, 1962), and it is oncogenic in hamsters (Eddy et al, 1962).

During a lytic cycle the genome is expressed in two temporal phases, early (E) and late (L), punctuated by the

onset of viral DNA replication (see Figure 2; Tooze, 1980). Various species of mRNA are formed by differential splicing (Berk & Sharp, 1978a; Aloni et al, 1977). The E temporal class consists of two mRNA species which encode two proteins called large T (MW 90,000) and small t (17,000) antigens (Thimmappaya & Weisman, 1977; Prives et al, 1977). During a lytic infection, large T is responsible for the induction of a number of cellular enzymes (Tjian et al, 1978), it negatively regulates its own synthesis (Tegtmeyer et al, 1975; Khoury & May, 1977), it is involved in initiating viral DNA replication (Tegtmeyer, 1972) and is required for maximal expression of late mRNA (Khoury & May, 1977; Keller & Alwine, 1984; Brady et al, 1984a). The role of small t is less well understood but some evidence suggests that it may be required for transformation of resting cells (Sleigh et al, 1978). Two abundant L mRNAs, 16S and 19S, encode three major proteins: VP1, the major capsid protein, and VP2 and VP3, minor structural components. A fourth, small, open reading frame has been noted in the leader region of the late 16S mRNA (Fiers et al, 1978) and termed the agnogene (Dhar et al, 1977). Jay et al (1981) have detected a 7,900 MW protein species, which is probably the product of this gene, accumulating late in infection. The protein has a short half-life and binds to single and double stranded DNA. It has been suggested (Hay et al, 1982; Hay & Aloni, 1985) that the agnoprotein directs premature termination of late transcription, late in the infectious cycle, by a mechanism akin to attenuation in bacteria (Yanofsky, 1981).

Despite these interesting observations, it is clear that control of both E and L gene expression occurs mainly at the level of transcription initiation and is mediated by sequences in a relatively small region of the genome around the origin of DNA replication (see Figure 2; Tooze, 1980). This, approximately 400 bp, DNA sequence contains the promoters for both E (Hansen et al, 1981; Benoist & Chambon, 1980) and L (Brady et al, 1982) transcripts, both of which have been subjected to intensive in vitro mutagenesis in order to distinguish functional elements.

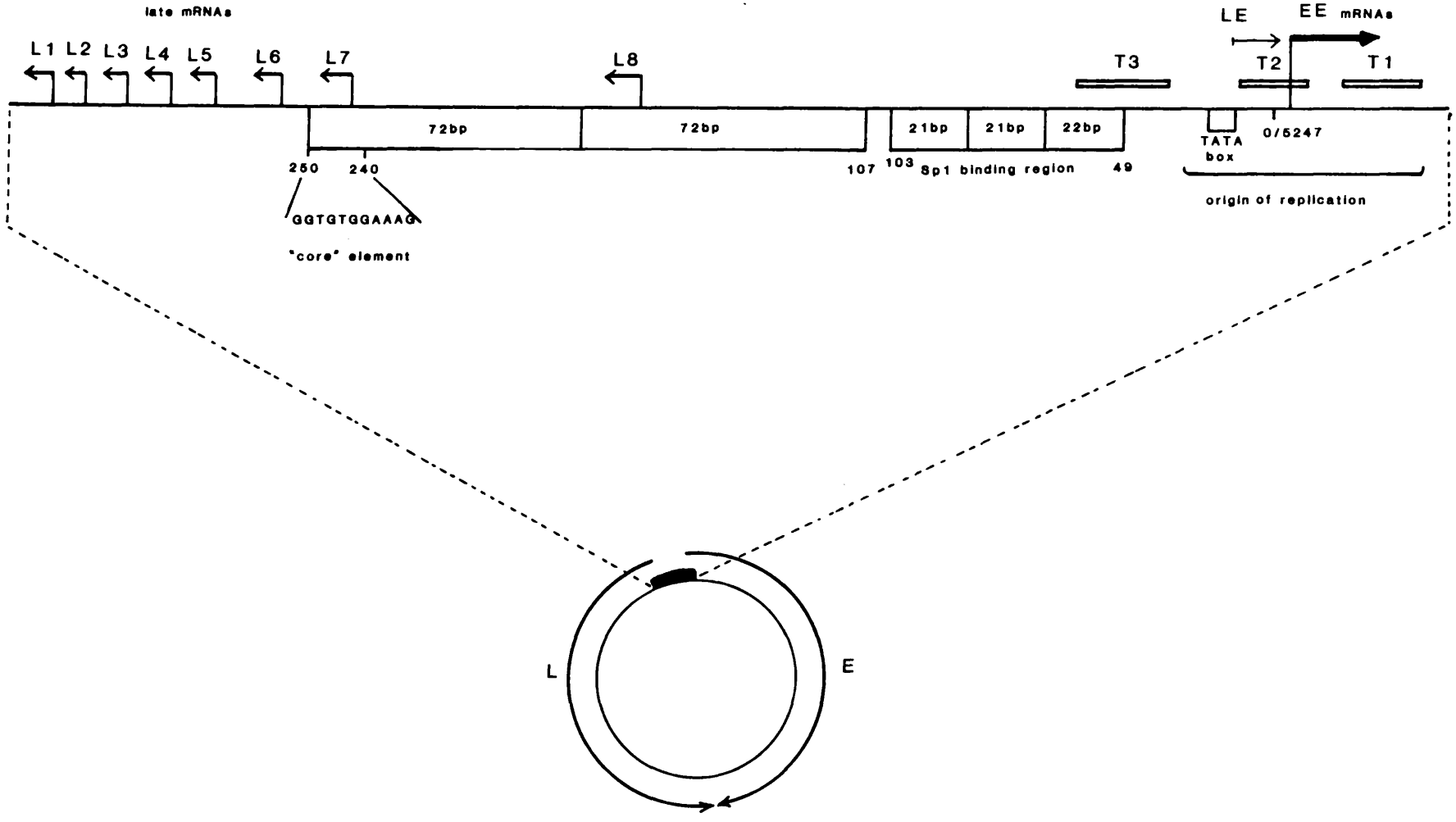

late mRNAs
L1 L2 L3 L4 L5 L6 L7 L8
LE EE mRNAs
T3 T2 T1
72bp 72bp 21bp 21bp 22bp
250 240 107 103 Sp1 binding region 49
TATA box
0/5247
GGTGTGGAAAG
"core" element
origin of replication
L E

Figure 2: The Control Region of SV40

A cartoon of the SV40 control region is drawn, noting the salient features discussed in section 1.4.2. L1-8 represent the known late mRNA initiation points. T1-3 represent large T-antigen binding sites. E and L indicate Early and Late transcription units, not specific transcripts, which are discussed in the text.

Early transcription can be detected both at early and late times. Before the onset of DNA replication, Early-early (EE) mRNA is initiated in the "core" sequences required for DNA replication origin function (Reddy et al, 1979). These EE mRNA 5' termini lie 21-25 bp downstream of a sequence TATTTAT, a reasonable candidate for a TATA box, which Ghosh and co-workers (1981) showed to be responsible for the positioning of these cap sites. Benoist & Chambon (1980, 1981) also concluded that this TATA box positioned the EE mRNA 5' ends in vivo. Two further early promoter elements have been defined: a 72 bp repeat region was shown to have a crucial role (Benoist & Chambon, 1980; 1981; Gruss et al, 1981) and also the 21 bp repeated sequences are important (Fromm & Berg, 1982; Everett et al, 1983). Deletion and point mutations within the 21 bp repeats indicated that the CCGCCC motifs, present as multiple copies, constituted an important component of the early promoter, but are only able to act in the presence of the 72 bp repeat (Benoist & Chambon, 1981; Gruss et al, 1981). As described above, the 72 bp repeats contain an enhancer function, capable of cis-activating transcription from heterologous promoters (see section 3.3.2). After DNA replication is initiated, there is a switch from mRNA intiation at the EE sites, which lie in the origin, and residual late early (LE) transcription initiates from sites further upstream (Ghosh & Leibowitz, 1981; Hansen et al, 1981), this change being mediated by large T antigen. It is now clear that the 21 bp repeats contain overlapping elements involved in promoting EE and LE transcription (Baty et al, 1984; Hansen et al, 1981), GC elements III to VI are involved in the function of both promoters while repeats I and II are important only for EE transcription. When EE promoter function is impaired, the LE promoter increases in activity suggesting that the two promoters are in competition (Wasylyk et al, 1983b).

Parallel studies on the interaction of large T antigen with the early promoter were pursued in several laboratories. Purified T antigen binds specifically to three closely spaced sites around the origin of replication in a

sequential manner (Tjian, 1978; Shalloway et al, 1980). This binding protects 30-35 bp of DNA from digestion by DNAase at each of the three sites, which also span the EE mRNA start sites and TATA box. Site I has the greatest affinity and site III the lowest. Binding to sites II and III is reduced by mutations which eliminate binding to site I (Myers et al, 1981; McKay & DiMaio, 1981; Tenen et al, 1982) suggesting that binding occurs co-operatively. Jones and Tjian (1984), using methylation and ethylation interference as a probe, concluded that clusters of G residues were important in the binding sites and that interaction was localised to only one face of the DNA helix. In vitro and in vivo transcription studies suggested that the binding of T antigen to site I is directly responsible for repression of EE mRNA synthesis (DiMaio & Nathans, 1982; Hansen et al, 1981; Rio et al, 1980). Further studies by Rio and Tjian (1983), in which site II was completely replaced by an equivalent sized piece of bacterial DNA, indicate that site II also plays a role in the repression of early transcription. Deletion of site III appeared to have no adverse effect on viral growth and the viruses did not over produce T antigen (DiMaio & Nathans, 1982) suggesting that this site had no role in repression of EE transcription. However, single base substitutions in the promoter proximal region of site III showed altered transcriptional properties both in vivo and in vitro (Das & Salzman,1985). Host cell proteins have also been shown to be required for the transcription of SV40 promoters. The factor Sp1 is necessary for efficient in vitro transcription from the early promoter (Dynan & Tjian, 1983a, b), discussed in section 2.3.

Sequences involved in the promotion of late (L) mRNA transcription have been analysed in considerably less detail. The L mRNA initiation sites are not preceded by a recognisable TATA homology although the sequence GGTACCTAACC (located 21-31 bp upstream from the major late start site) may function as such in vitro (Brady et al, 1982). Hartzell and co-workers (1984a) have implicated the 21 bp repeats in late promoter function and this finding correlates with the

observation that these elements are able to function bidirectionally (Everett et al, 1983). Further studies (Hartzell et al, 1984b), using a transient expression system, identified two sequence regions which affected maximal late expression. One sequence element lies in the minimal origin of replication and probably reflects the requirement for template amplification, the second falls within the origin proximal portion of the 72 bp repeats - a region which contains L mRNA initiation sites. In vitro analysis of a series of deletion mutants suggested that an important promoter element lay in the origin distal 21 bp repeat (Brady et al, 1984b). At present the situation is confused and a more subtle approach may be needed to differentiate the EE, LE and L overlapping promoter elements.

It has been proposed that the increase in L mRNA after the onset of DNA replication is also dependent on the activity of T antigen (Cowan et al, 1973; Khoury & May, 1977). Studies using a temperature sensitive mutant of large T, tsA, suggested that the ability of T antigen to positively influence transcription from the L promoter was dependent on cell type and, by implication, on the interaction of T antigen with host proteins (Alwine & Khoury, 1980). In order to disengage transcription from the obvious increase in template copy number, also mediated by large T, two approaches have been taken. Keller and Alwine (1984) used recombinant plasmids carrying the bacterial chloramphenicol acetyl transferase (CAT) gene under SV40 late control. By deleting origin sequences and co-transfection with plasmids which express T antigen, they were able to show that expression from non-replicating L-CAT constructs was significantly increased in the presence of T antigen. Origin competent plasmids showed even higher levels of L promoter activity. Brady and colleagues (1984a) obtained qualitatively similar results using either intact and origin deleted plasmids or intact SV40 DNA and DNA synthesis inhibitors, detecting expression of the late transcription unit by immunoblotting of VP1. Both groups

used mutant T antigen systems to conclude that the trans-activating function of the protein required an intact DNA binding domain. In order to reconcile the apparent paradox of a bidirectional enhancer (Gruss et al, 1981; Moreau et al, 1981) not stimulating transcription from the late promoter early in infection, it has been suggested that the mechanism of T antigen trans-activation is by competition with a cellular repressor which acts at the early time (Brady et al, 1984a). Lastly, Keller and Alwine (1985) have detected two sequence elements which mediate late gene expression in the presence of T antigen. One element contributed about 35% of maximal transcription and was active only when the origin sequences were intact. This element lies within the origin distal 72 bp repeat and may correspond to the TATA box-like element of Brady et al (1982). The second element lies in a 33 bp segment around the junction of the two 72 bp repeats. This sequence functioned in the absence of the origin and T antigen binding sites and may be responsible for trans-activation.

In summary, the present model for control of SV40 gene expression must accomodate three distinct properties of the controlling region:

(a) the relatively tight control of EE mRNA initiation by conventional TATA box and CCGCCC promoter elements, selectively activated by the upstream enhancer.

(b) the binding of T antigen with concomitant down-regulation of E mRNA expression (including a shift in the promoter utilisation), activation of DNA replication and increase in L transcription. It seems likely that the effect on transcription, at least in part, may simply be by steric hindrance of RNA pol II upon T antigen binding.

(c) the role of T antigen, distinguishable from its role in increasing template copy number, in trans-activating L promoter elements.

1.4.3 Adenovirus and the EIA gene product

The EIA coding region of adenoviruses 2 and 5 (Ad2 and Ad5; see Figure 3) has been intensively studied due to its role in both adenovirus gene expression and transformation (Berk et al, 1979; Jones & Shenk, 1979a; Sambrook et al, 1974; Johansson et al, 1978). Two main transcripts, the 12S and 13S mRNAs, are produced from the EIA region and these differ internally only by the size of a small intron encoding about 40 in phase amino acids (Perricaudet et al, 1979; Berk & Sharp, 1978b; Bos & van der Eb, 1985). Mutant viruses containing deletions in the EIA region fail to accumulate early mRNAs (Berk et al, 1979; Jones & Shenk, 1979b). Nevins (1981), by means of pulse-chase experiments, showed that this regulation occurred during transcription initiation. The phenomenon of promoter stimulation by the products of certain viral immediate-early genes has been termed trans-activation (see also sections on HSV-1 and SV40 transcription) and this term will be used here to describe stimulation of adenovirus early and late transcription units by EIA.

Several groups have shown that the product of the 13S transcript alone is capable of trans-activating adenovirus promoters (Carlock & Jones, 1981; Riccardi et al, 1981; Montell et al, 1981; Weeks & Jones, 1983). This transcript encodes a number of phosphoproteins, as determined by immunoprecipitation and SDS gel electrophoresis (Yee et al, 1983). The protein product of the 13S mRNA, purified from an over-expressing E.coli strain, is able to activate the EIII promoter upon co-injection into Xenopus oocytes (Ferguson et al, 1984), suggesting either that the post-translational processing is unnecessary for function or that the oocyte is able to make the required modifications. It has also been shown that similarly purified EIA is active in stimulating transcription from EIIA and the major late promoter (MLP) upon injection into mammalian cells (Krippl et al, 1984). Winberg & Shenk (1984) reported that the 12S product was able to activate various genes in the absence of 13S mRNA, although less effectively, and it has also been reported

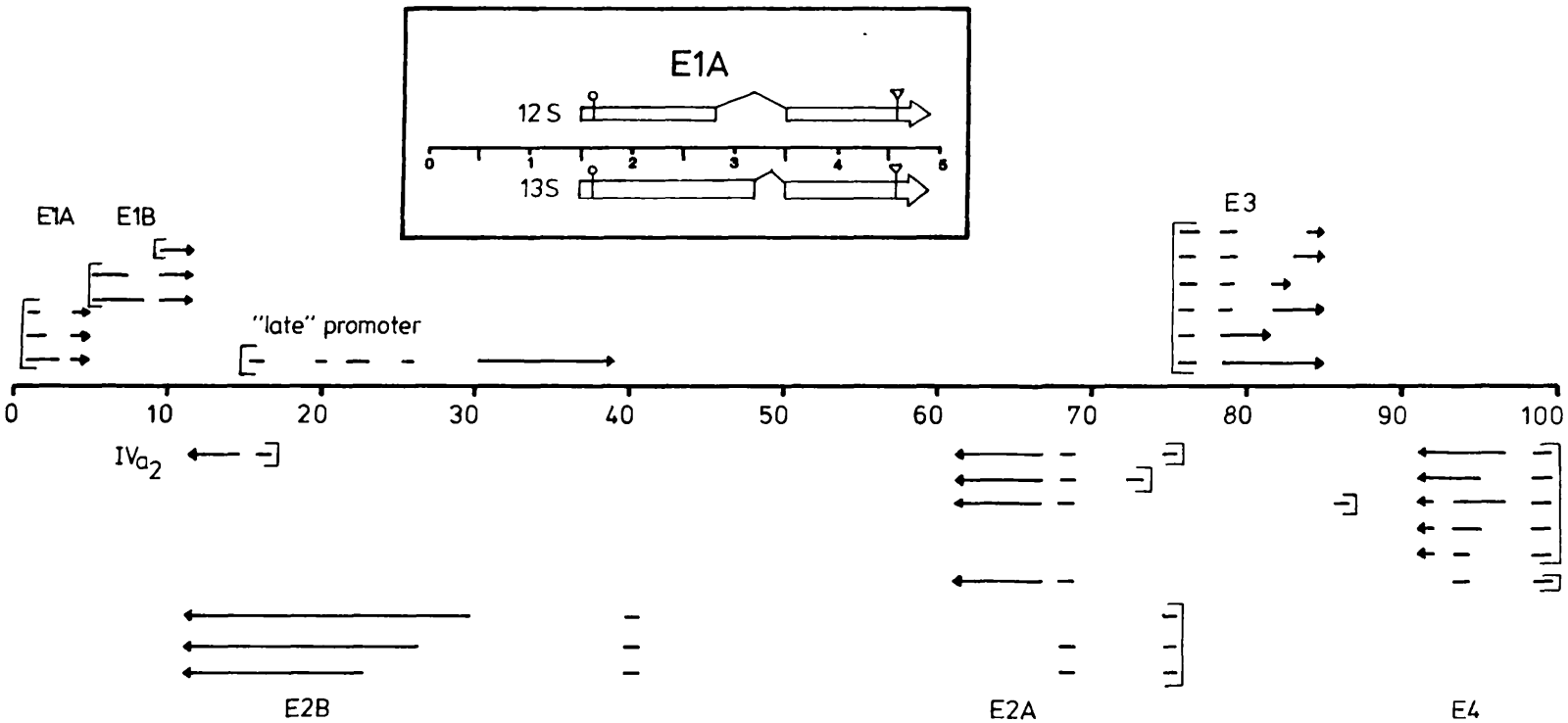

E1A
12 S
0
1
2
3
4
5
13S
E1A
E1B
"late" promoter
E3
0
10
20
30
40
50
60
70
80
90
100
IVa2
E2B
E2A
E4

Figure 3: Adenovirus type 2 Early trancription

The main Ad2 early transcription units are indicated. The inset shows a detail of the two major E1A transcripts, discussed in section 1.4.3. Promoters are indicated by the presence of square brackets and arrow heads show the direction of transcription.

that both 12S and 13S products stimulated expression from the EIIA and EIII transcription units and that this stimulation was at the transcriptional level (Leff et al, 1984). Rossini (1983) has shown that the EIA region activates transcription from the EIIA early (map coordinate 75; EIIAE) but inhibits expression from the EIIA late promoter (map coordinate 72; EIIAL). Both promoters transcribe mRNAs encoding the adenovirus DNA binding protein but at different times. Recently, evidence has been presented to support the view that the product of the 13S mRNA is responsible for EIIAE activation and that the product of the 12S mRNA is involved in EIIAL inhibition (Guilfoyle et al, 1985).

The observation that EIA products were able to activate cellular genes, both in situ as shown for the hsp70 gene (Kao & Nevins, 1983) and the β-tubulin gene (Stein & Ziff, 1984), and when introduced as plasmid during co-transfection experiments (Green et al, 1983; Svensson & Akusjarvi, 1984; Gaynor et al, 1984) precipitated the search for EIA-specific activator sequences. Bos and ten Wolde-Kraamwinkel (1983) showed that the EIB promoter of Ad12 is activated by the EIA 13S product and that sequences within 135 bp of the EIB mRNA cap site were important. Imperiale & Nevins (1984) failed to find specific sequences required for activation by EIA, although trans-activation was correlated with the presence of a functional promoter region which could operate in either orientation and at a considerable distance from the cap site. Similarly, Weeks & Jones (1985) found that sequences from -105 to -82 in the EIII gene upstream region were required for both EIA induced and uninduced expression. Placing these sequences at -79 with respect to an insensitive TK gene conferred inducibility by EIA. A linker scanning analysis of the EIIA region revealed two sequences important for promoter function, at -29 to -21 and -82 to -66 (Murthy et al, 1985), the first region corresponding to the TATA box. These authors were unable to differentiate specific sequences required for activation by EIA. Trans-activation of heterologous promoters, in short-term

transfection experiments, did not appear to be dependent on specific activator sequences (Green et al, 1983; Everett & Dunlop, 1984). Such results have led to the hypothesis that EIA acts in a general way to catalyse the formation of stable complexes (Gaynor & Berk, 1983). Evidence in support of this suggestion comes from the observation that EIA can repress activity from certain enhancers, e.g. those of SV40 and polyoma virus (Borelli et al, 1984; Velcich & Ziff, 1985) perhaps indicating that stimulation by EIA, in trans, and by enhancers, in cis, are mutually exclusive regulatory pathways which compete for the same transcription machinery. Interestingly it has been shown that the EIA transcription unit is driven by upstream promoter sequences with enhancer-like properties (Hearing & Shenk, 1983). The EIA enhancer region lies between -141 and -305 in Ad5 and can operate both 5' and 3' to the EIA coding sequences, in either orientation. Chambon's group, working on Ad2, located the enhancer to within 340 bp of the cap site and showed that stimulation in cis decreased with distance from the promoter but was unidirectional (Hen et al, 1983; Sassone-corsi et al, 1983). Their results also suggested that the EIA product could replace the cis-acting element in achieving maximal expression from the EIA promoter.

To complicate further the story of adenovirus gene control, two groups have now reported the ability of EIV products to stimulate gene expression. Goding et al (1985) suggest that EIV trans-activation of EIIAE transcription requires sequences between -48 and -19, acting independently of the EIA products. A mutational analysis of the EIV region in virus showed that the product exerts pleiotropic effects on DNA replication, translation and mRNA synthesis (Halbert et al, 1985). The stimulatory function of EIV product may account for the sequential induction of adenovirus early gene transcription (see review by Flint, 1982). Two studies have shown that the EIIA gene product (a 72K DNA binding) protein can supress transcription from the EIV promoter (Nevins & Winkler, 1980; Handa et al, 1983) suggesting a possible homeostatic mechanism whereby the levels of EIIA and EIV products assume a steady state.

2. The Herpesviruses

2.1 Basis for classification

Five members of the family herpetoviridae have man as a natural host (Roizman et al, 1981). To date at least another 80 distinct herpesviruses have been described and isolated from a wide range of vertebrate and invertebrate animal species. In all cases the initial classification is based mainly on virus morphology and therefore all members of the herpesvirus family share a consistent structure (see Figure 4) which has four main features:

(1) an electron dense core, incorporating the double-stranded DNA genome (Epstein, 1962; Ben-Porat & Kaplan, 1962).

(2) an icosahedral capsid, containing 12 pentameric and 150 hexameric capsomeres (Wildy et al, 1960).

(3) an envelope, derived from the nuclear membrane and which is punctuated with virally encoded glycoprotein "spikes" (Wildy et al, 1960; Spear & Roizman, 1972).

(4) a structure referred to as the tegument (Roizman & Furlong, 1974) which is largely undefined and consists of a layer of proteinacious material lying between the capsid and envelope.

In these respects, the virus particles of the various herpesviruses cannot be differentiated when observed under the electron microscope. However, on the basis of biological properties and the size, base composition and arrangement of their genomes few virus families exhibit as much variation as the herpetoviridae (Roizman, 1982). Using this cornucopia of biochemical and biological characteristics, a tripartite classification of the herpesviruses has been proposed (Roizman et al, 1981). A detailed rational for the scheme is presented in Table 1, but the basic feature is a division into alpha-, beta- and gammaherpesvirinae based on host-range, duration of reproductive cycle, cytopathology and location of the viral genome during latent infection.

Table 1: The Subdivision of Herpesviruses According to Biological Properties.

Subfamily	Examples
Alphaherpesvirinae	
HOST RANGE: Varies from wide to very narrow both *in vitro* and *in vivo*.	HSV-1 HSV-2
DURATION OF REPRODUCTIVE CYCLE: Short (<24hr)	EHV-1
CYTOPATHOLOGY: Rapid spread of infection in cell culture resulting in mass destruction of susceptible cells.	PRV BMV
LATENT INFECTION: Frequently, but not exclusively, in ganglia.	
Betaherpesvirinae	
HOST RANGE: Narrow, frequently restricted to the species or genus to which host belongs. *In vitro* replicates best in fibroblasts.	HCMV EHV-2
DURATION OF REPRODUCTIVE CYCLE: Relatively long (>24hr)	
CYTOPATHOLOGY: Slowly progressing lytic foci in cell culture. Infected cells become enlarged (cytomegalia) both *in vitro* and *in vivo*. Carrier cultures easily established.	
LATENT INFECTIONS: Possibly in secretory glands, lymphoreticular cells, kidneys and other tissues.	
Gammaherpesvirinae	
HOST RANGE: *In vivo* usually limited to the same family or order as the host it naturally infects. *In vitro* all members of this subfamily replicate in lymphoblastoid cells and some also cause lytic infections in certain types of epithelioid and fibroblastoid cells. Viruses in this group are specific for either B or T lymphocytes.	EBV MDV HVS
DURATION OF REPRODUCTIVE CYCLE: Variable.	
CYTOPATHOLOGY: Variable.	
LATENT INFECTION: Latent virus is frequently demonstrated in lymphoid tissue.	

Abbreviations:
HSV-1, -2: herpes simplex virus type 1, or type 2
EHV-1, -2: equine herpes virus type 1, or type 2
PRV : pseudorabies virus
BMV : bovine mamillitis virus
HCMV : human cytomegalovirus
EBV : Epstein Barr virus
MDV : Marek's disease virus
HVS : herpesvirus saimiri

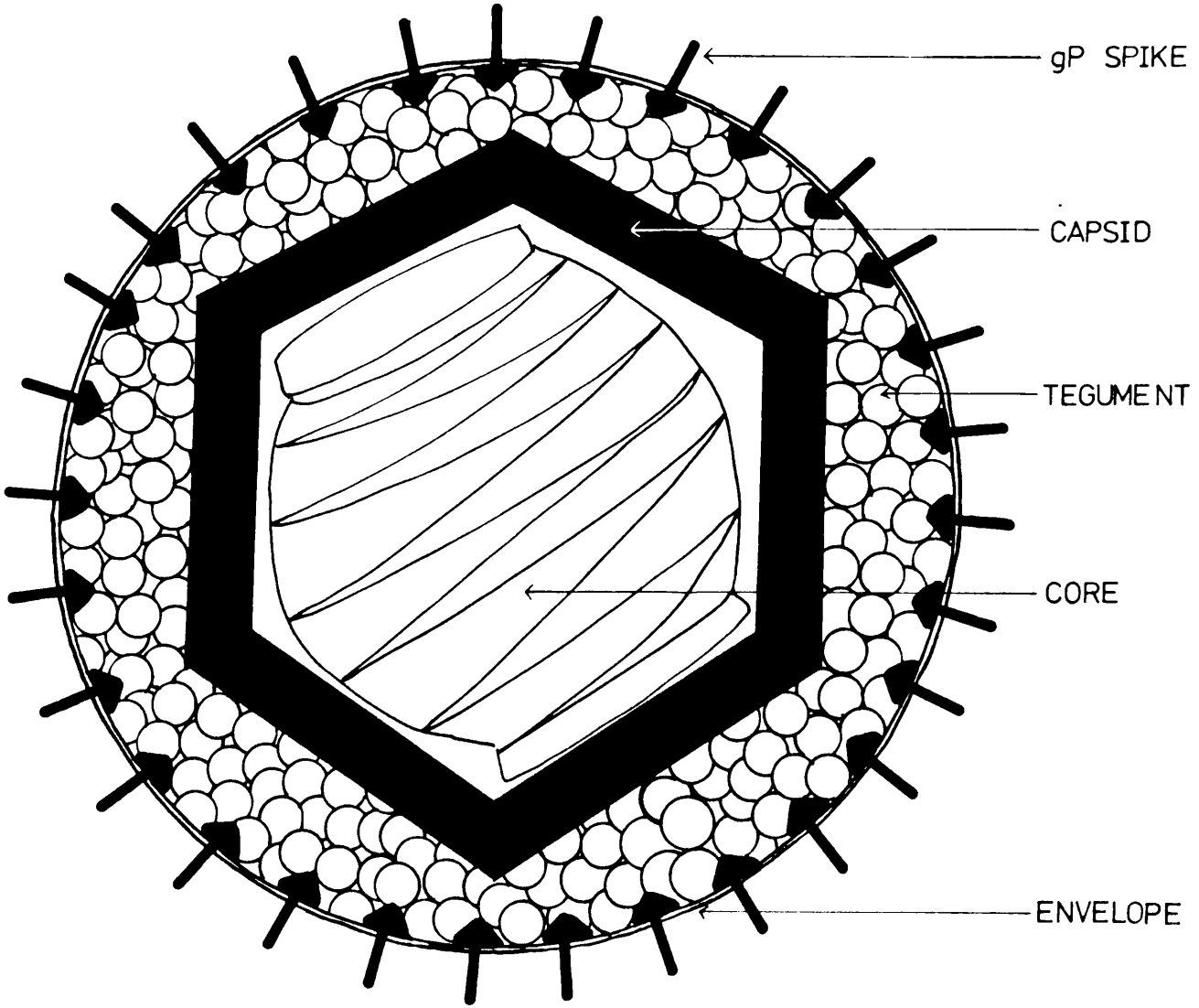

gP SPIKE
CAPSID
TEGUMENT
CORE
ENVELOPE

Figure 4: A Diagrammatic Representation of a Herpes Virus Particle.

The major features of a herpesvirus particle are shown. The core probably consists of DNA wound around a protein spindle. gP represents glycoprotein.

2.2 Human disease and herpesviruses

Herpesviruses which infect man are found in all three sub-families. The alphaherpesvirinae is represented by HSV-1 (the sub-family prototype) and HSV-2, and varicella-zoster virus (VZV). HSV-1 and -2 most commonly infect skin around the genitals, lips and oral cavities (Nahmias & Dowdle, 1968; Nahmias & Roizman, 1973). Less frequently several internal organs are involved resulting in severe or fatal disease. These two viruses are very closely related but can be differentiated antigenically (Plummer et al, 1981). HSV-1 and -2 display an important property of the herpes family in their ability to establish a so-called latent infection, during which the virus does not replicate and is believed to lie dormant in nerve ganglia (reviewed by Hill, 1985). Such a state, in some individuals, is punctuated by a series of recurrent disease episodes (reactivation) at the initial infectious site, or elsewhere, the causes of which are unknown.

VZV is the causative agent of two distinct clinical conditions: varicella (chickenpox) and herpes zoster (shingles). Chickenpox is a relatively benign infection of childhood whose manifestations are small, fluid-filled, skin eruptions. Zoster takes its name from the greek zona, meaning girdle, describing a creeping skin condition that encircles the torso, tends to occur later in life, and is almost certainly the result of a recurrent, previously latent, varicella infection, as the viruses isolated from the two conditions appear to be identical (Weller et al, 1958; Richards et al, 1979).

Human cytomegalovirus (HCMV; a member of the betaherpesvirinae) has been established as the most common infectious cause of intrauterine foetal brain damage and is a major cause of infection after organ transplant (Alford & Britt, 1985). Epstein Barr virus (EBV; the gammaherpesvirinae) is one of the most common and widely disseminated human viruses. It is the causative agent of infectious mononucleosis, a rarely fatal disease of the affluent West, persisting in the lymphoreticular system and

excreted intermittently in the oropharynx (Neiderman et al, 1976). Of more importance is the strong association of EBV with two important killer diseases in the third world and Far-East: Burkitt's lymphoma (De The et al, 1978) and nasopharyngeal carcinoma (Raab-Traub et al, 1983).

2.3 HSV-1

The virus studied in the work presented here was HSV-1 and I would briefly like to discuss some of the salient features of the virus before concentrating on HSV-1 gene expression and its control. HSV-1 has a linear, double-stranded, genome of approximately 100 million molecular weight, corresponding to about 140-150 kbp (Becker et al, 1968; Kieff et al, 1971), which is transcribed and replicated in the cell nucleus. It is now well established that the genome has a remarkable structure in which a large sequence region from each end is present internally as an inverted repeat and a shorter sequence is directly repeated, at each end and between the internal inverted repeats (see Figure 5; Sheldrick & Berthelot, 1974; Wadsworth et al, 1975; Wilkie, 1976; Delius & Clements, 1976). Thus the structure is described as two covalently bound segments: the long segment (82% of the genome) has a unique region (U_L) bounded by 9 kbp inverted repeats, and a short segment (18% of the genome) consisting of unigue sequences (U_S) bounded by 6 kbp inverted repeats. The direct terminal redundancy is known as the a sequence (Wadsworth et al, 1975) and varies in length, from about 280 to 550 bp, depending on the virus strain (Davison & Wilkie, 1981; Mocarski & Roizman, 1982). An interesting feature of this genomic organisation is the ability of the short and long segments to invert with respect to each other, resulting in a viral population within which the four possible genomic isomers are equally represented (Sheldrick & Berthelot, 1974; Hayward et al, 1975; Delius & Clements, 1976). This inversion process requires the a sequences (Mocarski et al, 1980; Mocarski & Roizman, 1981; Smiley et al, 1981; Davison & Wilkie, 1983), although the ability to invert is not a prerequisite for viral viability (Poffenberger et al, 1983).

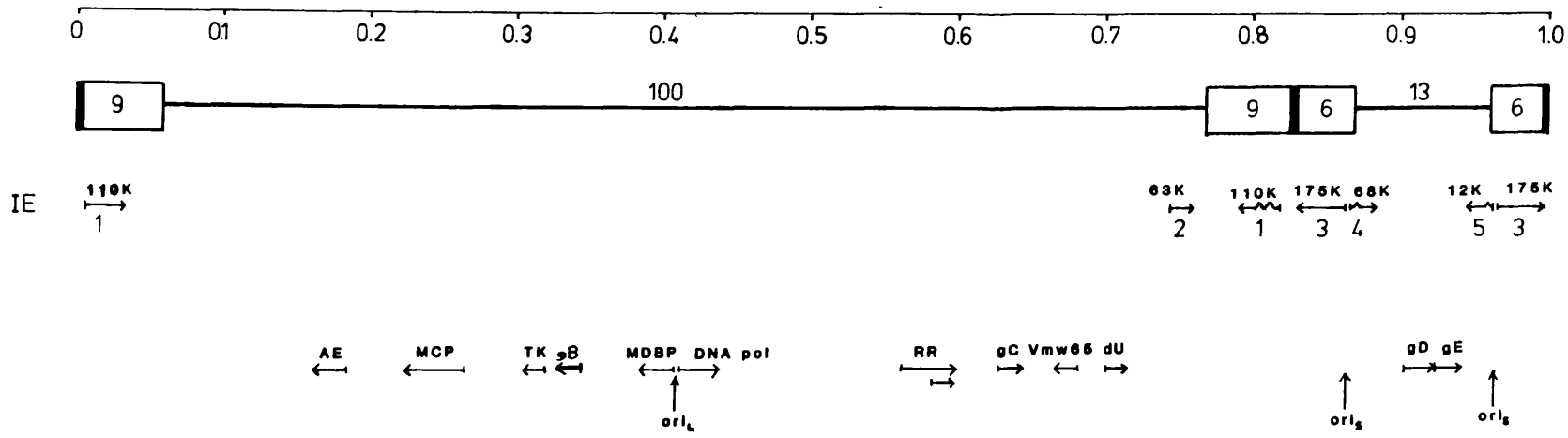

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
9
100
9
6
13
6
IE
110K
1
63K
2
110K
1
175K
3
68K
4
12K
5
175K
3
AE
MCP
TK
gB
MDBP
DNA pol
RR
gC
Vmw65
dU
gD
gE
ori_L
ori_S
ori_S

Figure 5: The Structure of the HSV-1 genome

The linear, double-stranded, DNA genome of HSV-1 is shown in prototype orientation. Open boxes represent inverted repeats, closed boxes represent the "a" sequence. The numbers above the genome detail the approximate length in kbp of U_L and U_S. The numbers in the boxes show the same statistic for each repeat. Below the genome are mapped the IE genes and the sizes of their protein products, and below this are indicated the approximate positions of the best characterised HSV-1 genes, which are as follows:

AE	alkaline exonuclease
MCP	major capsid protein
TK	thymidine kinase
MDBP	major DNA binding protein
DNA pol	DNA polymerase
RR	ribonucleotide reductase
gC	glycoprotein C
Vmw65	virion transcription activator
dU	dUTPase
gD	glycoprotein D
gE	glycoprotein E

The positions of origins of replication are shown. Finally, the relative orientation of the long and short unique region are detailed for each of the four genomic isomers.

Note that the mRNA for Vmw110 contains two introns.

The viral DNA is probably replicated by a rolling circle mechanism which generates long concatamers of newly replicated, unit length, genomes (Jacob et al, 1979). The information necessary for this replication is probably substantially virally encoded in the sense that both cis-acting control signals, and trans-acting factors, involved in DNA replication have been mapped to the genome. Each competent viral DNA molecule has three origins of replication. An origin is present in each of the two copies of the short repeat (Stow, 1982; Stow & McMonagle, 1983) and there is third origin in U_L (Spaete & Frenkel, 1982; Weller et al, 1985). The a sequences contain signals which mediate cleavage and packaging of newly replicated viral genomes (Stow et al, 1983; Stow & McMonagle, 1983; Varmuza & Smiley, 1985).The genome encodes its own DNA polymerase (Keir & Gold, 1963; Keir et al, 1966; Chartrand et al, 1979; 1980) and several enzymes involved in DNA metabolism, including: a thymidine kinase (Kit & Dubbs, 1963; Klemperer et al, 1967; Jamieson & Subak-Sharpe, 1974; McKnight, 1980), a ribonucleotide reductase (Cohen, 1972; Ponce de Leon et al, 1977; Dutia, 1983), an exonuclease (Keir & Gold, 1963; Preston & Cordingley, 1982; Banks et al, 1983) and a dUTPase (Wohlrab & Francke, 1980; Preston & Fisher, 1984). These features are shown in figure 5.

2.4 Transcription

The transcription pattern of HSV-1 has been investigated from two stand points: the study of virus gene organisation by detailed mRNA mapping, and as a model system for the study of eukaryotic gene control. It is with the latter that most of the work to be described is concerned and, therefore, the control of HSV-1 gene expression will be treated in most detail.

Transcript mapping has been performed, most notably by Wagner's group (reviewed by Wagner, 1984; 1985), as a prelude to identification of gene products and their function. This work, coupled with large and small scale DNA sequencing, has greatly increased our knowledge of the

virus. An almost complete transcript map of the genome now exists (Wagner, 1984; 1985; McGeoch et al, 1985; Rixon & McGeoch, 1984). The powerful ability to produce monospecific antibodies from synthetic oligopeptides, predicted from sequence data to occur in natural proteins, will allow the correlation of gene and product (for example see Frame et al, 1985). A detailed account of the HSV-1 mRNA organisation is outwith the scope of this introduction, however, certain generalisations about the features of HSV-1 trancripts can be made and are as follows:

(a) each transcript is controlled by its own promoter, located in the 5' flanking sequences.

(b) only a minority of the known transcripts are spliced (Costa et al, 1985).

(c) a relatively high density of transcript "packing" is observed often with a complex arrangement of overlapping mRNAs which are either 3' co-terminal and encode different polypeptides (Rixon & McGeoch, 1984), or 5' co-terminal and thus appear to be redundant (Anderson et al, 1981; Hall et al, 1982).

(d) most of the 5' flanking regions contain sequences similar to those implicated in eukaryotic promoter function, e.g. TATA and CAAT boxes (q.v.); similarly 3' flanking regions exhibit recognisable polyadenylation signals (AATAAA; Proudfoot & Brownlee, 1976; Fitzgerald & Shenk, 1981) and a possible candidate for a termination signal (YGTGTTYY; Taya et al, 1982; McLauchlan et al, 1985).

The HSV-1 genome is transcribed in at least three temporal phases, named immediate early (IE), early (E) and late (L) (Clements et al, 1977) or α, β, γ (Honess & Roizman, 1974; Jones & Roizman, 1979). Historically, the three mRNA classes have been differentiated on the basis of transcript abundance in the presence and absence of two kinds of metabolic inhibitor - inhibitors of protein synthesis (e.g. cycloheximide) and inhibitors of DNA synthesis (e.g. phosphonoacetic acid or ara C) (Swanstrom et al, 1975) - coupled to analysis of transcripts using hybridisation techniques. Thus, IE transcription does not

require de novo protein synthesis; E transcription requires the presence of IE products and is most abundant prior to the onset of DNA replication; L transcripts only accumulate to high levels after DNA synthesis is initiated. Present methods, which allow the isolation of individual genes, and the ability to study their expression with very sensitive probes, have tended to cloud the neat tripartite division, but in general it is still valid.

It is clear that transcription of the viral genome is mediated by host cell RNA pol II throughout infection, as all detectable viral transcripts are sensitive to α-amanitin (Constanzo et al, 1977; Ben-Zeev & Becker, 1977). Pertinent to this observation is the fact that naked HSV-1 DNA is infectious (Graham et al, 1973b; Sheldrick et al, 1973), although this does not exclude the possibility that some viral transcription requires virally encoded modifications to the host polymerase, indeed there is evidence to suggest that RNA pol II from infected cells has different biochemical properties to that from uninfected cells (Ben-Zeev et al ,1976).

2.4.1 IE mRNA Transcripts

The most distinct transcript class and the best studied comprises the IE mRNAs. Transcription of this small group of messages does not require de novo protein synthesis and infection in the presence of cycloheximide results in their over-production (Kozak & Roizman, 1974; Swanstrom et al, 1975; Clements et al, 1977; Jones & Roizman, 1979; Harris-Hamilton & Bachenheimer, 1985). Transcripts in the IE class clearly represent only a very small proportion of the genome (Clements et al, 1977; Watson & Clements, 1978; Jones & Roizman, 1979; Easton & Clements, 1980). There are only five distinct species, named IE mRNAs 1-5, whose sizes in HSV-1 are 4.7, 3.0 and three of about 2.0 kbp (Watson et al, 1979; Anderson et al, 1980). The map locations and size of the translation products are summarised in Figure 5.

2.4.2 Sequences Controlling IE Transcription

Post et al (1981) were the first to apply promoter fusion techniques to the study of IE promoters. These authors fused IE gene 3 sequences, extending 5' from a region in the 5' untranslated mRNA leader, to the coding sequences of the HSV-1 TK gene. This plasmid construct was then used either to produce recombinant virus in which TK was flanked by IE sequences (IE-TK), or to biochemically transform LTK^- cells to the TK^+ phenotype. It was discovered that the IE-TK containing virus effectively had its TK gene under IE control, that is, TK activity was detectable at IE times (1-3 hours post-infection). This suggested that the sequence elements which differentiated IE promoters from E and L promoters lay in the 5' flanking sequences. The most interesting observation came when biochemically transformed cells were superinfected with TK^- virus. It was shown that this resulted in a powerful stimulation of the integrated HSV-1 TK gene, which had previously been shown to have acquired IE characteristics of expression. Post and co-workers postulated that the virus inoculum contained a trans-acting factor (the so-called virion component) which effected a positive stimulation of IE promoters.

The transformation experiments were extended by contructing several deletions in the IE gene 5' flanking sequences and measuring transformation efficiency. These results suggested that the 5' flanking sequences of IE gene 3 contained seperate promoter and regulator elements. The sequences which conferred inducibility on the IE3 gene were mobile in that they could be placed upstream of a heterologous promoter (that of the natural TK gene) and remain functional. Furthermore, these cis-acting responder elements were shown to fall within the region -110 and -331 relative to the cap site (Mackem & Roizman, 1982a). Detailed deletion analysis of these upstream sequences have been performed and activity determined by transformation or short-term transfection assay (Cordingley et al, 1983; Herz & Roizman, 1983; Lang et al, 1984; Kristie & Roizman, 1984). The conclusions from these studies are generally in accord

and are that the major regulator function resides between -174 and -331 and operates through "basal" promoter elements which lie between +1 and -108. The regulator element stimulates expression from the IE3 promoter even in the absence of the virion component. This stimulatory ability has many of the hall-marks of classical enhancer function, that is, it does not require its own promoter, it can operate in an orientation independent manner, at a considerable distance from the promoter and it shows some cell type specificity (Lang et al, 1984; Preston & Tannahill, 1984).

Fusion of other IE gene promoters to TK, namely IE1, IE2 and IE4/5, has revealed that they too respond to the virion component but, at least in the cases of IE2 and IE4/5, do not have enhancer-like sequences (Mackem & Roizman, 1982b; 1982c; Preston et al, 1984; see section 5.6). Therefore, it would appear that even within a relatively well defined gene class, the IE genes, it is possible to make further divisions based on promoter strength in the absence of the trans-acting virion component.

Mackem and Roizman (1982a, b) noted that upstream of the various IE genes there were several conserved blocks of nucleotides, in particular, GC-rich rgions with the potential to form hairpin loops and an AT-rich motif for which they have published two consensus sequences: TAATGARAT and GYATGNTAATGARATTCYTTGNGG. A thorough analysis of sequences conserved between HSV-1 IE promoters, and between HSV-1 and -2 IE promoters, confirmed the conservation of these features (Whitton & Clements, 1984; Whitton et al, 1983) and it was suggested that these elements were responsible for the coordinate induction of HSV IE genes. Functional analysis of the IE4/5 promoter sequences has provided direct evidence for the role of TAATGARAT in IE gene response to the virion component, however, these experiments suggested that sequences flanking the element modulated the effeciency of the response (Preston et al, 1984). Gaffney et al (1985) have demonstrated, by using a

synthetic oligonucleotide containing the TAATGARAT element, that this sequence is able to confer responsiveness to the HSV virion component, however, this is not to say that other sequences do not play a role. Further evidence in support of sequences other than the TAATGARAT playing a role in stimulated and unstimulated expression from IE3 comes from experiments by Bzik and Preston (1986). These authors made a series of small deletions into the IE3 regulator element, from both sides. Their results suggest that three distinct sequence elements in this region play a role in the control of transcription and that their effects are cumulative. Two of these elements are consistent with the TAATGARAT consensus, but, a third motif consists of $^{-298}$/CGGAACGGAAGCGGAAACCGCCG/$^{-275}$ and is able to activate an otherwise non-functional TAATGARAT element. Similar sequences have been shown to be functionally important in adenovirus EIA and polyoma virus enhancers (Hearing & Shenk, 1983; Herbommel et al, 1984; Veldman et al, 1985). Recently there have been reports of the identification of enhancer elements in two other herpesviruses: herpes virus saimiri (Schirm et al, 1985) and HCMV (Boshart et al, 1985), and that HCMV IE promoters are trans-activated by a component of the virion (Stinski & Roehr, 1985; Spaete & Mocarski, 1985).

2.4.3 The Virion Component

Studies by Batterson and Roizman (1983) confirmed that the virion component was trans-acting and probably consisted of one or more peptides located outside the nucleocapsid. These workers superinfected cells, biochemically transformed with the HSV-1 TK gene under IE transcriptional control, with different virus preparations. Stimulation of TK activity was seen using UV light-irradiated virus preparation and, at NPT, using a mutant virus, tsB7, which fails to release viral DNA and whose nucleocapsids accumulate at the nuclear pores (Batterson et al, 1983). Furthermore, it was shown that the related herpesviruses PRV, HCMV and BMV were unable to stimulate expression from the HSV-1 IE promoter, unlike HSV-2. Campbell, Palfreyman

and Preston (1984) applied a novel strategy to identify the herpes gene or genes responsible for the specific stimulatory response. Cloned restriction fragments of the HSV-1 genome were transfected into cells, together with a plasmid construction in which the IE3 promoter regulated expression of the HSV-1 thymidine kinase (TK) gene. It was possible to identify those regions of the genome which stimulated levels of TK activity. Initial experiments located the IE specific trans-activator to the EcoRI I region (map coordinates 0.635-0.721). By a process of sub-cloning and insertion mutagenesis the authors were able to exclude all mRNAs known to map in EcoRI I except one (Hall et al, 1982). Hybrid arrested in vitro translation and immunoprecipitation studies identified the gene responsible as that for Vmw65, a major virion tegument phosphoprotein (Marsden et al, 1978; Lemaster & Roizman, 1980; McLean et al, 1982). A considerable part of the work presented here is concerned with the determination of the DNA sequence of the gene for Vmw65 (Dalrymple et al, 1985).

2.4.4 The Products of HSV-1 IE Genes

The five IE mRNA species: IE1, 2, 3, 4 and 5, translate to the following respective products: Vmw110, 63, 175, 68 and 12 (Pereira et al, 1977; Preston et al, 1978). The map positions of these (see Figure 5), and many other viral genes, have been identified by exploiting the detailed differences in electrophoretic mobility of the homologous polypeptides of HSV-1 and HSV-2, and the construction of intertypic recombinants of verifiable genome structure (Marsden et al, 1978; Morse et al, 1978; Halliburton, 1980). Due to their position in the major repeated regions, IE3 and IE1 are present as two copies in the genome and the promoters of IE4 and IE5 are exactly equivalent.

Of the IE polypeptides, Vmw175 is the best characterised. The protein has an apparent molecular weight of 175,000 under SDS gel electrophoresis, hence its name, and is heavily phosphorylated, although the molecular weight predicted from the DNA sequence is much lower (132,835; McGeogh et al, in press). There are three phosphorylated

forms, named a, b, and c in descending order of electrophoretic mobility (Pereira et al, 1977), which appear to cycle their phosphate moeities (Fenwick & Roizman, 1977; Wilcox et al, 1980). A mutant, tsK, which probably has only a single base pair change (Davison et al, 1984), in a total coding region of 3,894 bp (McGeoch et al, in press), is defective in the synthesis of the c form at NPT (Preston, 1979b). TsK, and a similar mutant (tsB2), accumulate the cytoplasmic form Vmw175a and transport of the b form to the nucleus is reduced (Courtney & Powell, 1975; Preston, 1979b; Cabral et al, 1980). The phosphorylation of the protein is probably of two kinds: poly(ADP-ribosyl)ation (Preston & Notarianni, 1983) and direct formation of phosphoesters with serine and threonine residues (Metzler & Wilcox, 1985).

The role of Vmw175 in the transcriptional control of E and L gene expression, can be introduced at this point. It is clear that early and late genes have an absolute requirement for functional Vmw175 in order that they be efficiently transcribed (Watson & Clements, 1978; 1980; Preston, 1979a; Dixon & Schaffer, 1980; Leung et al, 1980; Persson et al, 1985) and, in the absence of functional Vmw175, IE transcripts and polypeptides are over produced. The mutant tsK is unable to synthesise E and L mRNAs at NPT. Temperature shift experiments have shown conclusively that the ts lesion is reversible, throughout the infectious cycle (Watson & Clements, 1978; 1980; Preston, 1979a,b). Other mutants, tsB2lu and tsLB2 (Dixon & Schaffer, 1980), have been shown to exhibit a similar phenotype to tsK. It must be noted, however, that these mutants may mask more subtle, pleiotropic, effects of Vmw175. Experiments in which the arginine analog canavanine has been used, to disrupt IE protein function, suggest that subsets of E and L genes are regulated differently (Honess & Roizman, 1974; Pereira et al, 1977; Harris-Hamilton & Bachenheimer, 1985). This view is supported by reports of ts mutants in Vmw175 which are partially or completely permissive for E gene expression (Marsden et al, 1976; Preston, 1981; DeLuca et al, 1984) and it is clearly possible to delete small regions of the coding

sequences without affecting viral replication (Schröder et al, 1985).

The functions of the other IE gene products are less well understood. Only Vmw12 is not phosphorylated (Pereira et al, 1977; Marsden et al, 1978; Fenwick & Walker, 1979; Marsden et al, 1982). All four phosphorylated IE products assume nuclear locations and can be found tightly bound to isolated chromatin (Hay & Hay, 1980; 1981). Deletion of one copy of the gene for Vmw110, which is diploid, appears to have little phenotypic effect (Davison et al, 1981), however, there is evidence that this gene product plays a role in transcriptional activation of E and L genes (Everett, 1984; O'Hare & Hayward, 1985a; Quinlan & Knipe, 1985). It must be noted that the latter work is based on plasmid transfection assays and has not been shown to occur in the virus situation. Deletion of a large part of the coding sequences of the gene for Vmw68 did not reduce the yield of viral progeny (Post & Roizman, 1981), although a residual fragment of the protein was translated and may be essential in certain cell types (Sears et al, 1985). Recently, Schaffer's group have isolated mutant viruses with ts lesions in IE gene 2 (Sacks et al, 1985). These authors found that E proteins were synthesised at near wild-type levels at NPT, but certain IE products were over-produced and L protein synthesis was drastically reduced. They concluded that Vmw63 has an essential role in virus replication and that the protein is required after the onset of early gene expression.

2.4.5 Autoregulation of IE transcription

The experiments of Clements and co-workers (1977) suggested that IE transcripts decreased in abundance at E and L times. This was interpreted to mean that a viral protein, probably an E gene product, "switched-off" IE transcription (Honess & Roizman, 1974). It has been shown that an upshift of tsK, or similar mutants, to NPT after the onset of early gene expression results in the resumption of IE protein synthesis and the termination of E and L protein

synthesis (Preston, 1979a; Dixon & Schaffer, 1980). This result was interpreted to mean that Vmw175 has a direct role in suppression of IE gene expression and, on the basis of metabolic inhibitor experiments, it was concluded that "autoregulation" by Vmw175 occurs at the level of transcription (Preston, 1979a, b; Dixon & Schaffer, 1980; Watson & Clements, 1980). A similar conclusion has been reached in the case of PRV, whose single IE gene product appears to autoregulate (Ihara et al, 1983). However, little work has been done on the actual levels of IE mRNA, as detected by sensitive hybridisation techniques. Using nuclease S1 mapping and Northern blot analysis, it has been shown that significant levels of IE mRNAs are present, in the cytoplasm, at E and L times (F. Rixon, personal communication) and that Vmw175 is able to stimulate expression from the promoters of IE2 and IE4/5 (see section 5.6). The presence of IE mRNA at these times is inconsistent with the finding that IE proteins are not detectable (Honess & Roizman, 1974; 1975). Preston (1979b) has noted that IE mRNA becomes unstable at a certain stage in the infectious cycle which might suggest that there is some translational control of IE gene expression. Thus, the evidence in favour of a transcriptional down-regulation of all IE genes at E and L times must be re-examined in the light of modern techniques.

2.4.6 Transcription of Early and Late Genes

The methods used to study expression of E genes, in isolation from the viral chromosome, have produced a rather paradoxical situation. As has been discussed, it is quite clear that functional Vmw175 is required for the efficient expression of mRNA from E gene promoters when that promoter is resident in the virus. When the isolated TK gene (with its natural promoter) is transfected into cells and biochemically transformed cells selected however, it is relatively easy to obtain cell lines expressing this HSV gene (Wigler et al, 1977; Maitland & McDougall, 1977; Minson et al, 1978). This observation suggests that the TK gene can

be transcribed without trans-activation. Furthermore, TK genes integrated into the chromosomes of transformed cells are still able to respond to superinfecting virus in the way expected for a virally resident gene (Kit & Dubbs, 1977; Smiley et al, 1983). Similarly, it has been shown that a plasmid containing the HSV TK gene can be efficiently transcribed and translated upon injection into Xenopus oocytes (McKnight & Gavis, 1980; Cordingley & Preston, 1981). These facts led to the conclusion that the TK gene, and E and L genes in general, have promoter sequences which are able to be transcribed at a basal level, upon which is superimposed the viral stimulatory mechanism. Indeed, McKnight and co-workers have used the Xenopus oocyte assay to map, in considerable detail, the sequence elements resposible for the transcriptional activity of the TK promoter (see section 3.3.2.). However, given that E and L genes are activated by one or more IE gene products, the question was asked, which sequences confer the ability to be induced ?

Several groups have claimed to be able to delineate specific sequences which mediate trans-activation (Zipser et al, 1981; Smiley et al, 1983; El-Kareh et al, 1985), but these conclusions must be tempered by the finding that several non-viral promoters can be similarly stimulated (Green et al, 1983; Everett, 1984; see section 4.2). Everett (1983; 1984), using a short-term transfection assay and analysis of mRNA levels, has studied the glycoprotein D promoter in considerable detail. It was shown that trans-activation of gD by superinfecting virus required GA-rich sequences between -40 and -83, but this same sequence requirement was also found for cis-activation by the SV40 enhancer, suggesting that this region represents a basal promoter and not a virus E gene responder sequence. Recently, the TK linker scanning mutants of McKnight have been re-evaluated in a transient expression system and the effects of the various promoter mutations on trans-activation studied (Eisenberg et al, 1985). These workers conclude that the same elements which mediate

expression in the constitutive environment of Xenopus oocytes are important for the property of trans-activation by superinfecting HSV-1 in mouse cells, although there may be different emphasis placed on the importance of the various promoter elements by oocytes and tissue culture cells. Both these studies are consistent with viral IE gene products having the ability to generally activate any available promoter, probably by binding to, sequestering, or activating cellular factors which are important for gene expression in general.

There are no published accounts of the mutational dissection of L promoters, however some preliminary results were presented at the 10th International Herpesvirus Workshop, Ann Arbor, Michigan, 1985. Johnson and Everett constructed a plasmid which contained the promoter for US11 (a known late gene) with or without a viral origin of replication. By short-term transfection of tissue culture cells followed by superinfection with virus to provide replicative functions, they showed that transcription from the US11 promoter was drastically reduced, but not abolished, either in the presence of PAA (which inhibits DNA synthesis) or in the absence of a viral origin of replication, suggesting that template copy number may be an important level of control of L genes. Introduction of gC promoter deletions into viral genomes defined the region from +126 to -140 as important for the transcriptional activation of this gene (F. L. Homa, J. Glorioso, M. Levine). Knipe's group have published evidence for the major DNA binding protein (MDBP) playing a role in the control of some E and L genes, based on the finding that a ts lesion in this protein causes some IE, E and L mRNAs to be overproduced at NPT (Godowski & Knipe, 1983). Recently they have suggested that L gene expression is actively suppressed by the MDBP prior to DNA replication, but it is not established whether this is a specific effect (Godowski & Knipe, 1985).

2.4.7 Summary

There would appear to be a very sharp distinction between IE and E/L promoters in terms of transcriptional sensitivity to protein synthesis inhibitors such as cycloheximide. This has allowed the detailed characterisation of the IE genes and some aspects of their control. More work is being done on the role of IE products in subsequent lytic development and it is probable that Vmw175 will be joined by Vmw110 and Vmw63 in being assigned roles in viral transcription. The possibility of E and L gene products acting to regulate IE or L transcription has not been studied in detail, although the concept of the MDBP acting to repress L transcription is an intriguing one.

MATERIALS & METHODS

3. Materials

3.1 Chemicals

Most chemicals and other materials were obtained from either the Sigma Chemical Company Ltd. or BDH Chemicals U.K.. Acrylamide and boric acid were purchased from Koch Light Laboratories; Ammonium persulphate and TEMED from Biorad Laboratories; Tris[hydroxymethyl]-aminoethane (Tris) from Boehringer Mannheim GmbH and analytical grade ethanol from James Burroughs U.K. Ltd.. The sequencing primer (5'-GTAAAACGACGGCCAGT-3') was obtained from New England Biolabs. Deoxynucleoside and dideoxynucleoside triphosphates were bought from Pharmacia P-L Biochemicals. Wacker silane (methacryloxypropyl-tri[methoxyethoxy]silane) was a gift from Wacker-Chemie GmbH.

3.2 Radiochemicals

All radiochemicals were purchased from Amersham International plc at the following specific activities:

5'-{α-^{32}P} dATP; 3000 Ci (110 TBq)/mmol
5'-{γ-^{32}P} dATP; 5000 Ci (185 TBq)/mmol
(Methyl-^{3}H) thymidine; 40-60 Ci (1.5-2.2 Tbq)/mmol

3.3 Enzymes

Restriction enzymes were mainly purchased from Bethesda Research Laboratories, New England Biolabs or Nbl Enzymes Ltd.. XcyI was the product of Pharmacia P-L. S1 nuclease,T4 polynucleotide kinase and Klenow polymerase were bought from Boehringer Mannheim GmbH. Most of the sequencing was performed with a gift of Klenow polymerase from Dr. A. Davison. T4 DNA ligase was purchased from Bethesda Research Laboratories.

3.4 Viruses

Herpes simplex virus type 1 strain 17 TK- (P.G.Saunders, 1981) and a tsK TK- double mutant (C.M.Preston, unpublished results) were used.

3.5 Cells

BHK21 clone 13 (BHK C13; MacPherson and Stoker, 1962), a continuous cell line, derived from baby hamster kidney, and used throughout this work.

3.6 Bacterial Strains

Two strains of Escherichia coli K12 were employed. DH1 (recA1, nalA, r_R^-, m_R^-, endoI$^-$, B$^-$, relA1?; Hanahan, 1983) was used as host in all recombinant plasmid experiments. The host strain used for maintenance of M13 bacteriophage and recombinants was JM101 (Δ(lac;pro), supE, thi$^-$, F'traD36, proAB, lacI^q, Z Δ M15; Messing et al, 1981) which harbours an F factor essential for the tramsmittance of the male specific phage.

3.7 Bacteriophage

The single stranded DNA bacteriophage M13mp8 was used as the vector for production of cloned sequencing templates (see Figure 6; Messing and Vieira, 1982). Double stranded replicative form (RF) was a gift from Dr. D. Bzik. Initial cloning was performed using SmaI cleaved and phosphatase treated RF, a gift from Dr. D. M^{c}Geoch.

3.8 Plasmids

The plasmid vector employed was pAT153 (Twigg and Sherratt, 1980), a derivative of pBR322. The plasmid carries genes which confer resistance to the antibiotics ampicillin and tetracycline. It contains several unique restriction endonuclease sites. The main recombinant plasmid used in this study was pGX158, a clone of the BamHI F fragment of the HSV-1 genome (Campbell et al, 1984). A variety of IE promoter containing plasmids, whose construction has been described previously, were used and are described in the text.

3.9 Bacterial Culture Media

E.coli K12 DH1 was grown in Luria Broth (L broth) which is: 177mM NaCl, 10g/l Bactopeptone, 5g/l yeast extract, pH7.5 prior to sterilisation. Solid media for the growth of

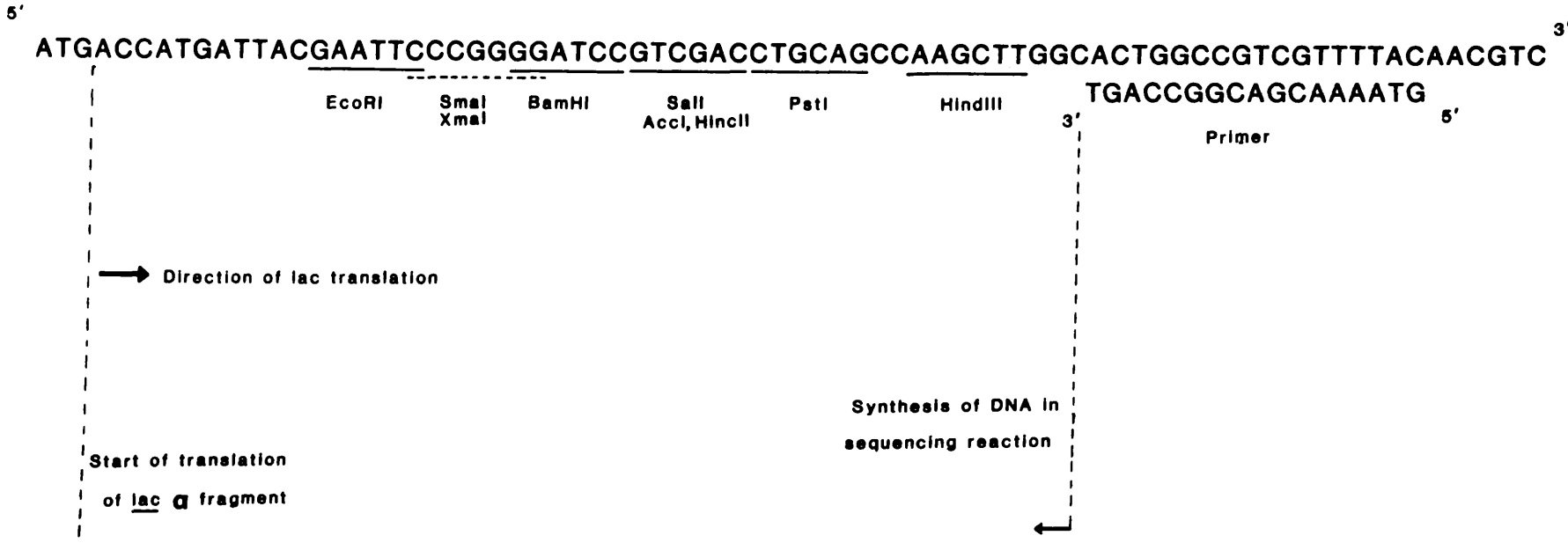

5'
ATGACCATGATTACGAATTCCCGGGGATCCGTCGACCTGCAGCCAAGCTTGGCACTGGCCGTCGTTTTACAACGTC
3'
EcoRI
SmaI
XmaI
BamHI
SalI
AccI, HincII
PstI
HindIII
TGACCGGCAGCAAAATG
3'
5'
Primer
Direction of lac translation
Synthesis of DNA in
sequencing reaction
Start of translation
of lac α fragment

Figure 6: The Multi-purpose Cloning Site of M13mp8

The upper sequence is that of M13mp8 from the ATG start codon of the β-galactosidase α-fragment, through the multi-purpose cloning region. The horizontal bars indicate the recognition sites for the enzymes shown. The primer sequence is shown aligned against the complementary region of M13mp8. The directions of translation of the galactosidase enzyme, and of the synthesis of DNA in a sequencing reaction are also indicated.

colonies was L broth agar, as above plus 1.5% (w/v) agar. When appropriate these media were supplemented with ampicillin (50µg/ml) or tetracycline (5µg/ml). Host strain JM101 was routinely cultured in 2YT broth composed of 16g/l Bactopeptone, 10g/l yeast extract, 5g/l NaCl. Colonies of JM101 were maintained on minimal essential agar (M9 salts, 0.2% glucose, 5mM $MgCl_2$, 2mM thiamine, 1.5% (w/v) agar).

3.10 Tissue Culture Solutions

Trypsin solution (0.25% in Tris/saline) and versene (6mM EDTA in PBS, 0.002% (w/v) phenol red) were used to harvest BHK C13 cells from support vessels. The cells were cultured in Glasgow Modified Eagle's Medium (Busby et al, 1964) to which was added 10% Tryptose phosphate, 100 units/ml Penicillin, 100µg/ml Streptomycin, 0.002% (w/v) phenol red and either 5% (ETC5) or 10% (ETC10) bovine calf serum.

3.11 Frequently Used Buffers and Solutions

TE : 10mM Tris.HCl (pH8.0), 1mM EDTA.

TBE : 90mM Tris, 90mM Boric acid, 1mM EDTA, (pH 8.3).

E : 36mM Tris, 30mM NaH_2PO_4, 1mM EDTA, (pH 6.8).

PBS : 170mM NaCl, 3.4mM KCl, 2mM KH_2PO_4, (pH 7.2).

Hx10: 1M NaCl, 100mM Tris.HCl (pH 7.4), 100mM $MgCl_2$.

Mx10: 500mM NaCl, 100mM Tris.HCl (pH 7.4), 100mM $MgCl_2$, 10mM DTT.

Lx10: 100mM Tris.HCl (pH 7.4), 100mM $MgCl_2$, 10mM DTT.

(High, medium and low salt restriction buffers)

TGM : 150ml 40% acrylamide/ N,N'-methylene bis-acrylamide (20:1), 50ml 10xTBE, 540g urea, made up to 1L with distilled water.

BGM : 150ml 40% acrylamide/ N,N'-methylene bis-acrylamide (20:1), 62.5ml 40xTBE, 540g urea, 50g sucrose, 10ml 1% bromophenol blue, made up to 1L with distilled water.

TNE : 300mM NaCl, 10mM Tris.HCl (pH8.0), 1mM EDTA.

TM : 100mM Tris.HCl (pH8.0), 100mM $MgCl_2$.

4. Methods

4.1 Growth and Maintenance of Bacterial Cells

Host E.coli strains were grown in L broth or 2YT at 37°C with shaking. Stock cultures were maintained by growing up a single colony in 5ml L broth, overnight, at 37°C with shaking. The cells were harvested by centrifugation in an MSE benchtop microfuge at 3000 rpm for 2 minutes. The excess medium was decanted off and 1ml of fresh L broth and 2ml of 80% sterile glycerol was added to the resultant pellet. Stocks were maintained at -20°C for an indefinite period.

4.2 Tissue Culture and Preparation of Virus Stocks

Cells were passaged as monolayers in 80oz. roller bottles containing 200ml of ETC10 and incubated in an atmosphere of 95% air, 5% CO_2 at 37°C.

For the preparation of viral stocks, five 80oz. bottles, containing monolayers about 90% confluent, were seeded with virus at an m.o.i. of 1:300, i.e. approximately 10^6 p.f.u. per bottle. The medium was decanted, the virus added in a volume of 40ml ETC10 and the cells incubated at 31°C. After 1-2 days CPE was obvious, and the roller bottles were harvested at 4-7 days when cells were rounded and falling from the surface. The cells were harvested by shaking into the medium and the contents removed to 250ml glass MSE centrifuge bottles which were centrifuged at 2000rpm for 10 minutes. The pellet was resuspended in 5ml of supernatant, transferred to a universal bottle and stored at -70°C. This was termed the cell associated (CA) portion. The remaining supernatant was centrifuged at 12000 rpm for 2hrs, to pellet the virus, the pellet was resuspended in 5ml of supernatant and stored at -70°C. This was termed the cell released (CR) portion. Both portions were sterility checked then treated as follows:

CA: Thawed, sonicated and frozen. This procedure was repeated then samples were centrifuged at 2000 rpm for 10 minutes. The supernatant was stored at -70°C. The pellet was resuspended in 5ml fresh ETC10 and the freeze/thawing

process repeated. The final supernatant was added to the first to give 10ml total of CA stock which was aliquoted, and titrated by established procedures (Brown et al, 1973)

CR: this was sonicated to give 5ml of CR stock, aliquoted and titrated.

4.3 Plasmid Preparation by Hard Lysis

A bacterial colony or loop from a glycerol stock was used to inoculate 5ml of L broth (plus appropriate antibiotic) in a 20ml bottle and incubated overnight. This "overnight" was then diluted into 200ml of L broth (plus antibiotic) and the culture shaken vigorously for 5-6 hrs, at which point chloramphenicol was added, to a final concentration of 50μg/ml, and the culture shaken overnight.

The bacteria were harvested by centrifugation for 5 minutes, at 4°C, in a Sorvall GS3 rotor. The supernatant was carefully removed and the bacterial pellet resuspended in 10ml of ice-cold 50mM Tris.HCl (pH 8.0)/ 25% sucrose solution. To this was added 5ml of fresh 5mg/ml lysozyme in 50mM Tris.HCl (pH 8.0) and the suspension incubated on ice for 30 minutes. At this point was added 3ml of 200mM EDTA and, after a further 5 minute incubation period, 3ml of 5M NaCl followed by 2ml of 10% SDS. This mixture was mixed carefully until white strands of an SDS/protein/DNA complex were visible. The mixture was then left to sit on ice for at least 2 hrs, up to a maximum of 16 hrs, and centrifuged in a Sorvall SS34 rotor at 21,000 rpm for 1 hr to remove cell debris and chromosomal DNA. The supernatant was then decanted carefully and treated in one of two ways:

(a) In order to thoroughly purify the plasmid DNA from contaminating chromosomal DNA, for the purposes of sequencing, the supernatant was subjected to ultra-centrifugation. To the supernatant was added 1 g/ml of $CsCl_2$ and EtBr to a final concentration of 500 μg/ml. This solution was poured into heat sealable polypropylene centrifuge tubes and centrifuged at 45,000 rpm, in a Beckman Ti50 rotor, for 48 hrs. At this point an equilibrium gradient had formed and the supercoiled plasmid band was

easily visible under long wave ultraviolet irradiation. The band was extracted by syringe and EtBr was removed by multiple extraction with distilled water saturated butan-2-ol followed by dialysis against two changes of TE buffer. Finally, the DNA was made 0.3M for sodium acetate and ethanol precipitated.

(b) For routine transfection experiments a different purification procedure was used. The initial supernatant was ethanol precipitated at -20°C for 30 minutes, centrifuged at 2000 rpm in an MSE coolspin, and the pellet air dried. This pellet was dissolved in 5ml of TE, extracted twice with equal volumes of phenol/chloroform and then once with chloroform alone. A second ethanol precipitation was performed. This time the air dried pellet was redissolved in 5ml TE, DNAase-free RNAase was added to a final concentration of 50µg/ml and incubated at 37°C for 1-2 hrs. Then a few crystals of proteinase K were added and the incubation continued at room temperature overnight. The phenol/chloroform extraction procedure was repeated, the aqueous phase made 0.3M for NaOAc and isopropanol precipitated by addition of a half volume of isopropanol and incubation at room temperature for longer than 30 minutes. After centrifugation, the pellet was washed with 100% ethanol, air dried, and redissolved in 1-2ml TE.

4.4 Determination of DNA Concentration

Two or three different volumes of the final plasmid preparation, estimated from experience to cover the range 0.4-1.0 µg of DNA, was digested with an appropriate restriction enzyme such that each molecule was linearised. The restriction digest was loaded onto a 1.0% agarose gel and electrophoresis performed overnight, in the presence of EtBr, alongside a standard of known concentration. The DNA was visualised under long wave ultraviolet trans-illumination and a negative prepared using polariod 667 film. The negative was then used to measure and compare concentrations by densitometry as determined by a Joyce-Loebel densitometer.

4.5 Gel Electrophoresis

4.5.1 Electrophoresis of DNA in Non-Denaturing Agarose Gels

DNA samples were run on horizontal slab gels of dimension 260mm x 160mm, comprising 200ml of 0.5-1.5% (w/v) agarose in 1xE buffer and 0.5µg/ml EtBr, eletrophoresed at 2 to 12 V/cm at room temperature. 0.2 volumes of loading dye (50% (v/v) Ficoll, 0.02% (w/v) Bromophenol Blue) was added to the samples prior to loading.

4.5.2 Non-Denaturing Polyacrylamide Gel Electrophoresis of DNA

These gels were prepared according to the method of Maniatis et al (1975). Vertical gel sandwiches, of dimensions 260mm x 160mm x 1mm, were sealed with suitable adhesive tape. Acrylamide stock (29% acrylamide, 1% N,N'-methylene bis-acrylamide stored at 4°C) was diluted to the required concentration (4-10%) and made 0.55x for TBE in a final volume of 80ml. To this was added 0.52ml of 10% ammonium persulphate (APS) and 80µl TEMED to initiate polymerisation. The mixture was poured quickly into the gel mould and allowed to stand for at least 45 minutes, during which time the acrylamide had fully polymerised, prior to electrophoresis. Samples were applied to the gel in 0.55xTBE, 10% (v/v) glycerol and 0.02% xylene cyanol/bromophenol blue. Electrophoresis was performed at 50-300 V/gel. After electrophoresis, the sandwich was dismantled and the gel either stained in 0.5µg/ml EtBr for 30 minutes, for visualisation under ultra violet light, or covered in plastic and exposed to X-ray film in order to detect radiolabelled DNA.

4.5.3 Denaturing Polyacrylamide/Urea Gel Electrophoresis

The products of S1 nuclease mapping reactions were run on ultra-thin polyacrylamide gels containing urea (Maxam & Gilbert, 1980). Gel sandwiches were 230mm x 450mm x 0.35mm. The final gel concentration of 8% was made 0.55x for TBE from a stock solution containing 29% acrylamide, 1% N,N'-methylene bis-acrylamide, 7M urea, and made up to a

final volume of 50ml with 10M urea. To this was added 50µl 25% (w/v) APS and 50µl TEMED to catalyse polymerisation. The gel was cast using a 50ml syringe without a needle and left, clamped on three sides, for at least 45 minutes. The gel was "pre-run" for 20 minutes at 40W prior to loading. Samples were loaded in a volume of 3-5µl containing 90% formamide, 0.02% (w/v) xylene cyanol, 0.02% bromophenol blue (formyl dyes), after heating at 100°C for 3 minutes and quenching on ice. Gels were electrophoresed at 40W until the bromophenol blue dye front was at the bottom of the gel. Radiolabelled species were detected by autoradiography.

4.5.4 Gradient Polyacrylamide Gel Electrophoresis for Sequencing.

By constructing a gradient of potential difference down the gel, small DNA fragments preferentially slow down as electrophoresis proceeds. This means that the spacing between bands becomes more even over the whole gel and allows more sequence to be read from each gel. The gradient of potential difference is constructed by using different ionic strength buffers, a low concentration of TBE is used for the top of the gel and a high concentration for the bottom (Biggin et al, 1983).

Two gel plates (400mm x 430mm x 0.35mm) were carefully cleaned, one notched and the other plain. The inside surface of the notched plate was wiped all over with about 1.5ml of "Repelcote" (a 2% solution of dimethylchlorosilane in 1,1,1-trichloroethane) to aid plate seperation after electrophoresis. The inside surface of the plain plate was similarly treated with about 1ml of a solution of Wackersilicone (100ml ethanol, 3ml 10% acetic acid, 500µl Wackersilicone). This compound allows a firm bond to occur between the glass surface and the polyacrylamide gel and so the gel can be dried to the plate by heating in an oven without gel distortion. The two plates were taped together as before.

Two standard solutions were made (see section 3.11), Top Gel Mix (TGM) which contains 0.5xTBE and Bottom Gel Mix

(BGM) containing 2.5xTBE. The latter also contained sucrose to act as ballast and bromophenol blue as an indicator to assess the smoothness of the gradient. Polymerisation was initiated in 80ml of TGM by the addition of 80µl each of 25% APS and TEMED, and also in 15ml of BGM by the addition of 15µl each of 25% APS and TEMED. Into a 50ml syringe was taken 10ml of this TGM followed by 14ml of BGM. The heavier BGM phase could be clearly seen due to the presence of the dye. A few bubbles of air were drawn into the vessel to allow some mixing of the two phases and then the mixture was carefully injected into the gel mould and allowed to settle. The remaining TGM was then taken up and injected into the mould, moving the flow from side to side in order not to skew the BGM to one side. Combs (with teeth sizes 2.5mm wide, 5mm deep and 1.5mm apart) were placed into the gel mould and the gel left to stand for at least 45 minutes.

The gels were run at 40W with 0.5xTBE in the top buffer reservoir and 1xTBE in the bottom reservoir.

4.6 Techniques Used in DNA Cloning

4.6.1 Restriction Enzyme Digests

DNA and restriction enzyme were incubated, in a suitable salt solution containing 100µg/ml BSA, for a pre-determined time period at 37°C (30°C for PstI). The amount of enzyme used and time of incubation varied according to the amount of DNA to be cleaved. One unit of restriction enzyme activity is defined as that amount required to digest 1µg of a standard DNA (usually bacteriophage λ) in 1 hour at the optimum temperature. Usually a three-fold excess was used, e.g. 1µg of DNA cleaved with one unit for 3 hours, or with 3 units for 1 hour. For the commonly used enzymes, incubation conditions were as described by Maniatis et al (1982) and consisted essentially of high (H), medium (M) or low (L) salt conditions (see "commonly used buffers"). Notable exceptions were the conditions used for SmaI (20mM KCl, 10mM Tris.HCl (pH8.0), 10mM $MgCl_2$, 1mM DTT), SalI (150mM NaCl, 10mM Tris.HCl (pH7.4), 10mM $MgCl_2$, 1mM DTT) and XcyI (10mM Tris.HCl (pH8.0), 10mM $MgCl_2$, 10mM 2-mercaptoethanol).

4.6.2 Recovery of DNA from Gel Slices

DNA bands were visualised under long wave ultra violet illumination after staining with EtBr. The required DNA fragments were cut from the gel and the DNA electroeluted from the agarose slice in 40mM Tris.HCl (pH8.5), 5mM sodium acetate, 1mM EDTA at 30mA for 1-2 hours. Electroelution was performed in special wells which allowed recovery of the DNA in a small volume (200µl). This solution was purified by extraction with phenol/chloroform and chloroform, adjusted to 0.3M sodium acetate and the DNA precipitated with ethanol at -20°C for greater than 1 hour or in a dry ice/ethanol bath for 10 minutes. If necessary, residual amounts of EtBr were removed by a butan-2-ol extraction prior to phenol/chloroform treatment. After precipitation, the DNA was pelleted by centrifugation at 13000g for 5 minutes, the pellet washed with 80% ethanol and dried under vacuum.

4.6.3 DNA-DNA Ligation

Ligations were performed in 20mM Tris.HCl (pH7.6), 10mM $MgCl_2$, 10mM DTT, 0.6mM ATP. Insertion of fragments into plasmids, or M13mp8 RF, was achieved by incubating an, approximately, ten-fold molar excess of "insert" with 20ng of vector in a volume of 10µl containing 1 unit of T4 DNA ligase overnight at 15°C. Vector DNA was treated with either bacterial alkaline phosphatase (BAP) or calf intestinal phosphatase (CIP) in order to remove 5' terminal phosphate groups and reduce re-circularisation. The BAP reaction was performed in a volume of 50µl containing 50mM Tris.HCl (pH8.5) at 65°C for 1 hour. The CIP enzyme is active under restriction digest conditions and was simply added to the digest along with the other ingredients. The phosphatase treated DNA was then phenol/chloroform extracted, made 0.3M for sodium acetate and ethanol precipitated. After centrifugation the pellet was washed and dried under vacuum before redissolving in 1xTE.

4.6.4 Transformation and Transfection of E.coli

I. Induction of Competence.

10µl of the stock glycerol culture of DH1, or a single colony of JM101, was used to innoculate 5ml of L broth or 2YT and incubated overnight at 37°C with shaking. The following day 100mls of fresh liquid medium was innoculated with 1ml of the overnight and shaken at 37°C, until the OD_{630} reached 0.3. The culture was chilled on ice for 5 minutes before harvesting in a Sorvall SS34 rotor (10000rpm, 4°C, 1 minute). The pellets were resuspended in a total of 40ml of ice cold 100mM $CaCl_2$ and incubated on ice for 1hr. The cells were then re-harvested and finally resuspended in a total of 1ml of ice cold 100mM $CaCl_2$. Competent cells were used within 24hr.

II. Transformation by Plasmid

10µl of plasmid DNA, usually a ligation mix, was added to 100µl of competent DH1. After incubation on ice for 45 minutes the transformation mixture was incubated at 42°C for 3 minutes and transferred to 2ml of fresh L broth. Incubation was continued for a further 2 hours, with shaking, at 37°C. The transformation mixture could then be plated out on L broth or selective agar plates containing antibiotic.

III. Transfection by M13 RF

The mp system of M13 vectors (Messing et al, 1977; Messing & Vieira, 1982) allows positive identification of recombinants by the production of different coloured plaques. The colour change is based on the bacterial enzyme β-galactosidase. Enzymatic activity is contributed from two distinct parts of the protein which can be artificially separated but still complement each other in trans. Thus, an M13 phage was engineered to contain the coding sequences for the first 145 amino acids of the β-galactosidase molecule, which is enzymatically inactive but can complement the lacZ ΔM15 deletion (see Figure 6). Full enzyme activity could be obtained upon growth in an E.coli strain (JM101; Messing et al, 1981) expressing this deleted gene and transcriptionally induced by the allolactose analogue IPTG.

The substrate Xgal is hydrolysed by β-galactosidase to bromochloroindole which imparts a blue colour to the infected plaque on a bacterial lawn. The vector M13mp8 has several unique restriction enzyme sites engineered into the α-peptide coding sequences such that if an insertion occurs the enzyme is inactivated and the plaque remains white.

10µl of M13 RF, usually a ligation mix, was added to 100µl of competent JM101 and incubated on ice for 45 minutes. After heat shock at 42°C for 3 minutes, these cells were added to a tube containing 3ml of molten top agar (maintained at 45°C), 25µl each of IPTG (25mg/ml in water) and Xgal (25mg/ml in dimethylformamide) and 100µl of mid-log phase JM101 in 2YT broth (a few ml of which were kept on ice from the initial 0.3 A_{630} culture). The final mixture was shaken quickly before being poured onto L broth agar plates, pre-warmed to room temperature. The top agar was allowed to set before incubation of the plates at 37°C overnight.

4.6.5 Small Scale Preparation of Plasmid DNA by Miniprep Lysis

Bacterial colonies were picked into 2ml L broth plus antibiotic (where necessary) and grown at 37°C with shaking for 5-6 hours. Chloramphenicol was added to 25µg/ml and incubation was continued overnight. A 1.5ml aliquot was placed in a 1.5ml reaction vial and centrifuged for 1 minute at 3000g (using an MSE benchtop microfuge). The resultant pellet was resuspended in 100µl STET (8% (w/v) sucrose, 5% (v/v) triton X100, 50mM EDTA, 50mM Tris HCl pH 8.0) containing 1mg/ml lysozyme. The suspension was then placed in a boiling water bath for 50 seconds before centrifugation at 13000g for 10 mins. The large fluffy pellet was removed with an automatic pipette and the nucleic acid precipitated by adding 300µl of 0.3M sodium acetate, an equal volume of isopropanol and incubation at -20°C for 30 minutes. After centrifugation for 5 minutes at 13000g, the supernatant was carefully decanted. The pellet was washed with absolute alcohol and then dried under vacuum. Finally, the pellet was resuspended in 50µl TE buffer. This level of purity was

sufficient to allow restriction endonuclease screening of large numbers of clones.

4.7 Gene Expression Experiments

4.7.1 Transfection of Plasmid DNA into BHK C13 Cells

About $2x10^5$ cells were seeded onto 30mm plates and grown overnight, in ETC10, at which point they were about 70-80% confluent.

A total of 3µg of plasmid DNA, composed of test plasmids and pBR322 carrier, was made up to 20µl with distilled water. To this was added 50µl of 2x HeBS solution and the tube vortexed. 44µl of 286mM $CaCl_2$ (0.5ml of 2M $CaCl_2$ + 3.0ml of distilled water) was added and the tube quickly vortexed for 5-10 seconds. The mixture was then left for 30mins at room temperature. During this time a calcium phosphate precipitate formed. The medium was removed from the tissue culture plates and 95ul of the precipitate was gently layered onto the cells. The plates were then incubated at 37°C for 45 minutes, rocking occasionally. During this time a special plating medium was prepared in the following proportions: 20ml ETC10, 2ml 2xHeBS, 17.75ml of used ETC10, 0.25ml 2M CaC12. 2ml of this medium was then added to each plate and the cells incubated normally at 37°C for 3-4hrs. The cells were then "boosted" by removing the plating medium and adding 0.3ml 25% DMSO/1xHeBS for 4 minutes (Stow & Wilkie, 1976). The DMSO was removed and the cells washed three times with ETC10. Finally 3ml of ETC10 was added and the cells incubated overnight at 37°C.

4.7.2 Preparation of Cell Extracts

The medium was removed from 30mm plates of transfected cells and 2ml of ice cold PBS added. The PBS was removed by Pasteur pipette and 0.5ml of fresh, ice cold, PBS added. The cells were carefully scraped into the PBS and decanted into a 1.5ml reaction tube. The tubes were then centrifuged at low speed, on a bench microfuge, for 2 minutes and the PBS supernatant removed. To the pelleted cells was added 100µl

of ice cold lysis buffer [10mM Tris.HCl(pH7.5), 10mM NaCl, 2mM $MgCl_2$, 0.5% v/v NP40]. The pellets were then disrupted by vortexing for 10 seconds and the tubes placed on ice. After 3 minutes the tubes were vortexed again and placed on ice for a further 5 minutes. Finally the tubes were vortexed and centrifuged at low speed for 2 minutes. The supernatant was removed to a fresh tube and either placed on ice, for immediate use, or quickly frozen in dry ice and stored at -70°C.

4.7.3 Thymidine Kinase Assays

Samples of cell extract were incubated for a fixed time period in an assay mix consisting of 100mM sodium phosphate buffer (pH6.0), 50µM dTTP, 5mM ATP, 100µCi {^{3}H}-thymidine/ml, 10mM $MgCl_2$. Unless otherwise stated 3µl of extract was incubated at 30°C for 1.5 hours. 10 µl of cold 2mM TdR was then added to quench the reaction and the tubes heated at 100°C in a heating block for 3 minutes. The tubes were incubated on ice for at least 5 minutes and centrifuged at high speed for 2 minutes in the bench microfuge. 50 µl of supernatant was spotted onto DE81 discs. The discs were washed 3 times in 4mM ammonium formate/ 10mMTdR for 10 minutes then briefly washed in absolute alcohol. They were then dried thoroughly under a heat lamp, placed in 5ml of scintillation fluid (0.1% PPO in toluene) in scintillation vials and counted in a scintillation counter for 1 minute.

4.7.4 Determination of Protein Concentrations

The method of Lowry et al (1951) was used. A standard curve was contructed using BSA.

4.8 DNA Sequencing

Principle

The dideoxy/chain termination method was used throughout, combined with a "shot-gun" cloning strategy. This strategy involved the cloning of several hundred small (300-600 bp), randomly generated, fragments. These clones

were then sequenced, as will be described, and collated by means of a computer and the Staden database software (Staden, 1979; 1980). The object is to sequence enough overlapping clones to provide one contiguous sequence which spans the region of interest.

The chain termination method requires a single stranded DNA template. This is obtained using the M13 cloning vectors of Messing and co-workers (Messing et al, 1977; Messing & Vieira, 1982). This filamentous bacteriophage replicates as a double stranded circle (the replicative form, RF) which may be manipulated and handled like a plasmid. Infectious phage particles, however, contain a genome of single stranded DNA and are continually secreted from the host cell, accumulating to high titre in the culture medium.

A short synthetic oligonucleotide primer, which is complementary to the insert site, is annealed to the single stranded template. The annealed template is then a substrate for the Klenow fragment of E.coli DNA polymerase when incubated with the four nucleoside triphoshates. Four separate reactions are set up, each containing an empirically determined concentration of one of four dideoxynucleoside triphosphates (Sanger et al, 1977; Sanger et al, 1980) together with the nucleoside triphosphates, one of which is radiolabelled (α-{^{32}P} dATP). When incorporated into growing DNA chains the dideoxy form results in abrupt termination of synthesis, preventing further elongation. The ratio of dideoxy to deoxy form is set such that only a small proportion of chains terminate at any moment in time and termination is essentially at random. Therefore, the four reactions provide sets of radiolabelled chains which have been randomly terminated, at A, C, G or T residues, but extending from a fixed position (the primer). The radiolabelled reactions are run side by side on ultra-thin denaturing polyacrylamide gels which can separate chains differing by only one nucleotide. The sequence may then be read directly from an autoradiograph (see Figures 10 and 11).

4.8.1 Cloning

A bank of M13 clones was constructed by sonication of self-ligated BamHI F fragment and isolation of sub-fragments in the size range 300-600 bp. These were then randomly inserted into the unique SmaI site of the vector M13mp8.

200µg of supercoiled pGX158 (see Figure 7) was digested with BamHI and fractionated on a 0.5% agarose gel. The region of the gel containing the BamHI F fragment was cut out and the DNA isolated by electroelution. 20µg of purified fragment was self-ligated until greater-than 95% of the DNA was in multimeric forms, thereby decreasing the proportion of "ends" which are more likely to be present as small molecules after sonication . This DNA was then sonicated (Deininger, 1983) and end-repaired by incubation with 10 units of T4 DNA polymerase, in a reaction 10µM for all four dNTPs and 1xTM, for 4 hours at 30°C. Repaired DNA was size fractionated on a 1% agarose gel, using HinfI or AluI cut pBR322 as markers. Gel slices containing fragments in the range 300-600 bp were isolated and the DNA extracted by electroelution.

The end-repaired, sonicated, BamHI F DNA was ligated with 20ng of SmaI cleaved and phosphatase treated M13mp8 RF under standard conditions. In order to optimise the ratio of insert to vector an initial experiment involved titrating the insert DNA for insertion efficiency. The ligation mix was transfected into competent JM101 and plated on L broth agar.

4.8.2 Preparation of Single Stranded Template

A single colony of JM101 was picked from a minimal agar plate into 5ml of 2YT broth and shaken, overnight, at 37°C. 100 ml of 2YT broth was inoculated with 0.5ml of the overnight, mixed, and 1.5ml aliquots were placed in 20ml universal bottles. White plaques were picked into the bottles which were shaken vigorously for 5-6 hours at 37°C. The phage cultures were decanted, carefully, into 1.5ml reaction vials and the bacteria pelleted by centrifugation at 13000 rpm for 2 minutes in a bench-top microcentrifuge.

Approximately 0.8-0.9ml of supernatant was removed to a second tube containing 200µl of 2M NaCl/20% PEG6000, mixed thoroughly, and left to stand at 4°C for at least 30 minutes in order to precipitate the bacteriophage. The phage were pelleted at 13000 rpm for 5 minutes, the supernatant removed, carefully with a Pasteur pipette without touching the pellet, and the tubes re-centrifuged at 13000 rpm for 1 minute. The remaining supernatant was very carefully removed using an automatic pipette. The phage pellet was resuspended in 200µl of 1xTNE, extracted twice with TE equilibrated phenol and ethanol precipated at -20°C overnight. The phage DNA was pelleted, washed with 80% ethanol and then dried under vacuum. Lastly, the pellet was redissolved in 50µl of 1xTE.

4.8.3 DNA Sequencing Reactions

The amount of single stranded template produced from a 1.5ml phage culture was usually far in excess of the amount required for a successful set of sequence reactions. About 3µl of template DNA was added to 1µl of 10xTM buffer, 1µl of oligonucleotide primer (0.2pmol/µl) and 5µl of distilled water in a 1.5ml reaction vial. This mix was incubated at 37°C for 30 minutes to allow annealing. Usually 16-18 clones were sequenced at once.

1.5ml reaction vials, with their caps removed, were set up for each of the T, C, G and A reactions. 2µl of the annealed template was added to each of these tubes. On ice, a "Klenow mix" had been prepared containing 2µl of {α-^{32}P} dATP, 16µl 12µM dATP and about 20 units of Klenow polymerase, made up to 26µl with distilled water. To each of four tubes, labelled and containing 32µl of T, C, G, or A "sequence mix" (see Table 2), was added 6µl of Klenow mix. 2µl of this final "reaction mix" was added to each 2µl sample of annealed template according to the reaction (T, C, G or A) required. The extension reactions were allowed to stand for 15 minutes at room temperature before the addition of 2µl of "chase mix" (containing 0.25mM dTTP, dCTP, dGTP and dATP) and a further 30 minute incubation at room

Table 2: Preparation of Sequencing Nucleotide Mixes

(a) Deoxynucleotide Triphosphate Mixes

	dA-0	dT-0	dC-0	dG-0
5mM dTTP	20	1	20	20
5mM dCTP	20	20	1	20
5mM dGTP	20	20	20	1
10xTE	50	50	50	50
dH_2O	540	370	370	370

(b) Sequencing Mixes

	dN-0	ddNTP	dH_2O
T mix	200	200µl 0.3mM ddTTP	0
C mix	200	56µl 70µM ddCTP	144
G mix	200	100µl 0.1mM ddGTP	100
A mix	200	100µl 70µM ddATP	100

Numbers are in µl.

temperature. 3µl of Formyl dyes was added, the reactions heated for 2-3 minutes at 100°C in a heat block, quenched on ice and then loaded on a sequencing gel.

4.9 Mapping of 5' and 3' Termini of mRNAs

The termini of mRNAs can be positioned accurately on a DNA sequence using the nuclease digestion procedure developed by Berk and Sharp (1978a). The method has been modified by the use of 5' and 3' terminally end-labelled DNA probes in order to locate the 5' and 3' termini of the mRNA respectively (Weaver & Weissman, 1979).

4.9.1 5' Terminal Labelling of DNA Restriction Fragments

1-5µg of DNA (free from contaminating RNA) was digested with an appropriate restriction enzyme and treated with calf intestinal phosphatase to remove the 5' terminal phosphate groups. The reaction mix was ethanol precipitated and the DNA pellet dried *in vaccuo*. The pellet was taken up in 20µl of 50mM Tris.HCl (pH7.6), 10mM $MgCl_2$, 5mM DTT, 0.5mM spermidine, 50µCi of γ-{^{32}P} ATP and 2 units of T4 polynucleotide kinase. The reaction was performed at 37°C for 2-3 hours. The end-labelled DNA was then fractionated on a non-denaturing polyacrylamide gel, the required fragment cut from the gel and isolated by electroelution.

4.9.2 3' Terminal Labelling of DNA Restriction Fragments

DNA restriction fragments produced with an enzyme leaving recessed 3' termini may be end-labelled using T4 DNA polymerase and one or more α-{^{32}P} dNTPs. The XcyI restricted DNA was recovered by ethanol precipitation, as before. The dried pellet was resuspended in 20µl containing 1xTM, 25µM dCTP, 25µM dATP, 25µM dTTP, 10µCi α-{^{32}P} dGTP and 2 units of T4 DNA polymerase. The reaction was for 1-2 hours at 31°C and the required fragment was isolated as before.

4.9.3 Mapping Termini of mRNA using Nuclease S1

Approximately 0.5-1.0µg of end-labelled DNA was co-precipated with 10µg of mock infected, or HSV-1 infected,

cell RNA (BSC-1 cells harvested at 6 hours PI, a gift from C.M. Preston). The RNA/DNA pellet was dried and resuspended in 20µl of 90% (v/v) formamide (deonised with Amberlite monobed resin, MB-1), 0.4M NaCl, 40mM PIPES (pH6.8), 1mM EDTA. The mixture was heated at 90°C for 3 minutes, in order to denature the nucleic acids, then rapidly transfered to a water bath set at precisely 58°C and incubated for about 16 hours. The incubations were quenched on ice and diluted ten fold with a nuclease S1 mixture containing 0.25M NaCl, 30mM NaOAc (pH4.5), 1mM $ZnSO_4$ and 4000 units of enzyme. The S1 reaction was performed at 37°C for 1 hour. The digests were extracted with phenol/chloroform and chloroform, made 0.3M for NaOAc and ethanol precipitated. The dried pellet was resuspended in formyl dyes and analysed by denaturing polyacrylamide gel electrophoresis.

4.10 Computing

The computer and associated software are important tools in the accumulation, handling and analysis of sequence data. The Institute of Virology has a Digital Research PDP 11/44 machine, running under the RSX-11M operating system. This was used for the day to day input and processing of raw sequence data, editing, and the bulk of the subsequent analysis. Searches of the large protein and nucleic acid databases were performed on remote machines, namely the Digital Research VAX/VMS (version V4.1) system in Edinburgh and the Phoenix 3 system in Cambridge, based on an IBM 3081.

The programs utilised can essentially be divided into three groups: (a) those used during the sequencing period to input, accumulate and process the gel readings, (b) those used in the subsequent interpretation and analysis of the completed sequence, and (c) general utilities which allowed the handling of the sequence in manageable pieces. All these programs have been either designed or implemented by Dr. P. Taylor, whose help is gratefully acknowledged. These groups are now described in more detail.

4.10.1 The Staden Database (DB) Package.

Staden has published detailed descriptions of a series of programs which can be used to inter relate the individual gel readings from a large shot-gun sequencing project (Staden, 1979, 1980). These include the following:

BATIN - gel readings are input via the keyboard and allocated a filename (e.g. G5.100 relates to gel number 5, clone number 100), called the archive. The operator defines a "file of file" name into which records of the individual archive filenames are placed for use by subsequent programs.

DBUTIL - this comprises the core of the sequencing project. Three large files are made within which are kept information relating each individual archive file to overlapping files on either side. This is termed the "file of relationships". In this way only the relative coordinates need be retained and, together with the archive files, a complete map of overlapping fragments can be constructed. In order to simplify the description of the database, a new word was coined- the "contig". A contig was defined as a set of gel readings that are related to one another by overlap of their sequences. Initially the number of contigs will be close to the number of gel readings, especially in a very large project, however, as the gaps between gel readings are filled by new overlapping sequences, the number of contigs will drop. The aim is to achieve one contig, that is, a single contiguous sequence within which all the input gel readings overlap. DBUTIL incorporates several features which allow difficult (poor) gel readings to be input "by hand", the database and/or individual gel readings to be edited, contigs to be joined and the state of the database to be monitored. The program is also able to calculate a consensus sequence from the assembled gel readings.

DBAUTO - input consists of a "file of files" containing several individual gel readings. These are compared with the existing database, held in DBUTIL, for any matches. If the number of consecutive matching characters is below a pre-set minimum, the gel reading is assumed to be novel sequence and

inserted into the database as a new gel. Valid matches which fall within the limits of certain user defineable parameters (e.g. one can set limits to the tolerated percentage mismatch and allow the program to insert a certain number of "padding" characters into both the new gel or the existing data) are automatically aligned and inserted into the existing database which is then updated. Gel readings which match but fail to satisfy the pre-set parameters are placed into a "failed" file. The results of all these deliberations are written to a file (JOB.DAT) which can be printed out on a line printer and read at leisure.

DBX - this performs functions analogous to those of DBAUTO but does not automatically interact with the database. Thus, it simply reports the results of attempts to match gel readings with a database consensus sequence. In this way it is possible to pre-screen gel readings for matches to vector sequences, for the occurrence of certain restriction enzyme sites (important when the sonicated DNA was initially self-ligated otherwise the program will attempt to join the ends of the fragment being sequenced!) or for "problem" gels and eliminate them before proceeding with DBAUTO.

DPAD - this is not a Staden program, it was written by Dr. Phillip Taylor, but it is an important part of the data acquisition apparatus. This allows gels to be read on a digitising tablet from which the data is input directly into the computer. The data is automatically put into a form which can be read by the DB system. The output is similar to BATIN, i.e. a file of files and a collection of user named archives.

4.10.2 Analysis of the Completed Sequence

AASCAN - This program, based on that of Bishop and Thompson (1984), allowed a rapid search of all predicted ORFs in the genome of EBV (Baer et al, 1984) for homology to ORFs in BamHI F. It is based on the occurrence of unique heptapeptide "words" within a given protein sequence. Thus, the sequence(s) to be searched are held as a so-called

"dictionary" of unique heptapeptide "words". The probe sequence is first broken down to an identical form and then the constituent parts are quickly matched against the dictionary. Perfect matches are noted, together with their positions relative to each other, and a score calculated based on the difference between the actual number of heptapeptides found and the number expected to occur at random.

BASES - counts the numbers of each base in a sequence and outputs a table of values.

CINTHOM - a matrix comparison program (Pustell & Kafatos, 1982). Two amino acid or nucleic acid sequences may be compared for homology using this program. Homology is displayed in two-dimensions, one sequence being aligned on the X-axis and the other on the Y-axis. The output uses letters to indicate varying percentage matches, capital "A" being the highest (99-100%) and small "z" the lowest, relative to the best possible score for a perfect match. Rows of letters forming a 45^{o} diagonal indicate a region of homology, lateral displacements of the diagonal correspond to deletions or insertions, and parallel lines suggest duplications. The matrix may be "compressed", with very little loss of information, in order to reduce the physical size of the printed output. The program filters "noise" by exponentially weighting matches over long segments. Matches are assessed over a long span, centred on any two bases or residues at a time, and the contributions of progressively more distant matches are given less weight. The alignment is displayed by placing one letter per base at the coordinate corresponding to the highest homology score for that base. The minimum match to be displayed is set by the operator providing another level of noise filtering. An optional modification to the standard algorithm has been developed by Dr. Phillip Taylor for the comparison of protein sequences. This treats non-identical, but biochemically similar, amino acids as an evolutionarily conserved change and scores such matches higher than completely dissimilar residues. The criteria for relatedness were based on the replacement

mutation matrix of Schwartz and Dayhoff (1979). All CINTHOM plots shown have utilised this feature.

CHART - converts simple numerical data into a histogram which can be printed on an Epson FX-80 dot matrix printer (Epson).

CLEAN - removes low scores from a CINTHOM plot.

DIAG - takes the large scale output from CINTHOM and converts it to a simple dot plot to be output on the Epson.

DSPLAY - (Blumenthal et al, 1982) searches for open reading frames (ORFs) of user defined length and displays them on a VDU terminal or outputs the display to a file for printing. The ORFs need not begin with an ATG.

FRMSCN - (Staden & McLachlan, 1982) this program identifies potential protein coding regions in long DNA sequences. The basic assumption is that all genes within a genome have similar codon preferences which are strong enough to be used as a means of discriminating between coding and non-coding regions (Grantham et al, 1980). In the case of HSV-1, which has a high %G+C, the codon usage is sufficiently skewed to be diagnostic (McGeoch, 1984). The program takes a standard gene (initially the coding sequences of the HSV-1 TK gene were used, but Figure 13 shows an analysis done with the total predicted coding capacity of BamHI F) and compiles a table of its codon frequencies. The sequence of interest is then scanned using a moving window of user defined length and the standard frequencies are used to analyse each of the three reading frames. The probability of the sequence being coding is plotted on a logarithmic scale. Continuous ORFs of high coding probability are drawn with an unbroken line at the 50% mark. Stop codons are also marked. An alternative, non-graphical, output lists the probabilities and the mid-point coordinate of the window used and thus it is possible to quickly identify regions where a dramatic shift occurs in the predicted coding frame. This could be due to an error in the primary sequence data, the sequence in this region can then be re-examined and either confirmed or modified.

HOMOL - aligns amino acid or nucleic acid sequences. One optimal alignment of two sequences is found by balancing the length and frequency of gaps, introduced into one or both sequences, against the number of allowed mismatches (Taylor, 1984). The relative weights assigned to a mis-match or gap can be altered. The alignment produced is the best possible under the given conditions but is probably only one of many equivalent alignments.

LINK - telephone link to the campus network. The link can be used to communicate with other machines almost anywhere in the U.K. if the requisite "addresses" and passwords are known. Files can be transmitted to other machines or received from them. The Edinburgh system was accessed by calling ERCVAX. The Cambridge system was accessed via the JANET (Joint Academic Network).

MWCALC - calculates the molecular weight and amino acid composition of a protein sequence.

PROFIL - produces a high resolution graphical display of the hydropathicity (Kyte & Doolittle, 1982) or hydrophilicilty (Hopp & Woods, 1981) of the amino acid side chains along a protein backbone. The display can be output to the Epson.

PTRANS - (Taylor, 1986) translates a DNA or RNA sequence into amino acids and outputs a listing with the translation above the coding strand of the nucleic acid sequence. Optionally, tables of codon usage and amino acid composition can be produced.

SEARCH - searches a file for specific sequences. Most often used to look for restriction enzyme cleavage sites. An up to date file of the known restriction sites is maintained for use by this program and subsets can be designed to look for enzymes with 4, 5 or 6 base pair recognition sites.

SEQLIST - lists a DNA or protein sequence and numbers the residues.

4.10.3 Utilities

CHOP - cuts a large sequence into smaller pieces.

TURN - produces the reverse/complement of a DNA sequence.

Searches of the large protein and nucleic acid databases (for a review see Kneale & Bishop, 1985) were not possible on the departmental computer. The largest databases are compiled by the European Molecular Biology Laboratory (EMBL; nucleic acids; Hamm & Cameron, 1986), release 6 of which holds 4,835 entries comprising 4,567,592 bases, and the National Biomedical Research Foundation (NBRF; proteins; Chen & Barker, 1985; George et al, 1986) which holds 3,182 entries comprising 694,014 residues (release 5.0). Rapid searches could be made on the Cambridge system using PEPSCAN, a program analogous to AASCAN. More exhaustive searches were performed on the ERCVAX. On the Edinburgh machine are maintained the programs of the University of Wisconsin Genetics Computer Group (UWGCG; Devereux et al, 1984). This includes online access to the NBRF database and the program WORDSEARCH which uses the search algorithm of Wilbur and Lipman (1983). Sequences identified out by these methods were further analysed using CINTHOM and HOMOL to determine their significance.

RESULTS & DISCUSSION

5. RESULTS

The results presented here seek to illuminate two different aspects of the control of expression of HSV-1 IE promoters, using two quite disparate techniques.

As discussed in the introduction, HSV-1 IE promoters are stimulated by a virion protein whose gene has been mapped to the BamHI F fragment of the genome. It is believed that autoregulation occurs subsequently, at least in the case of IE3. The first section deals with the determination of the DNA sequence of the BamHI F fragment and computer aided analysis of its genetic content, with particular emphasis on the gene for Vmw65. The second section seeks to address the problem of autoregulation of the HSV-1 IE3 gene by analysis of promoter function in short-term transfection assays.

5.1 Positive Control

5.1.1 The BamHI F Sequence Data

The genomic location of BamHI F is shown in Figure 7. Sequence data was obtained using a "shotgun" cloning approach and the dideoxy/chain termination technology. The principle is that by sequencing sufficient numbers of small clones, obtained at random from the DNA fragment of interest, overlaps will occur which eventually span the entire region and provide one, so-called contiguous, sequence. The mechanics of the cloning and sequencing reactions, together with a description of the collation of data by computer, have been described in the methods section.

In the BamHI F database there were 79,980 characters which produced a final contiguous length sequence of 8055bp. For each contig character, an average of 9.93 characters were read from gels. There were 462 independent gel readings present in the database and 96.4% of the sequence was obtained on both strands. A detailed listing of gel lines and overlaps, plus a key to their meaning, is presented in Figures 8 and 9.

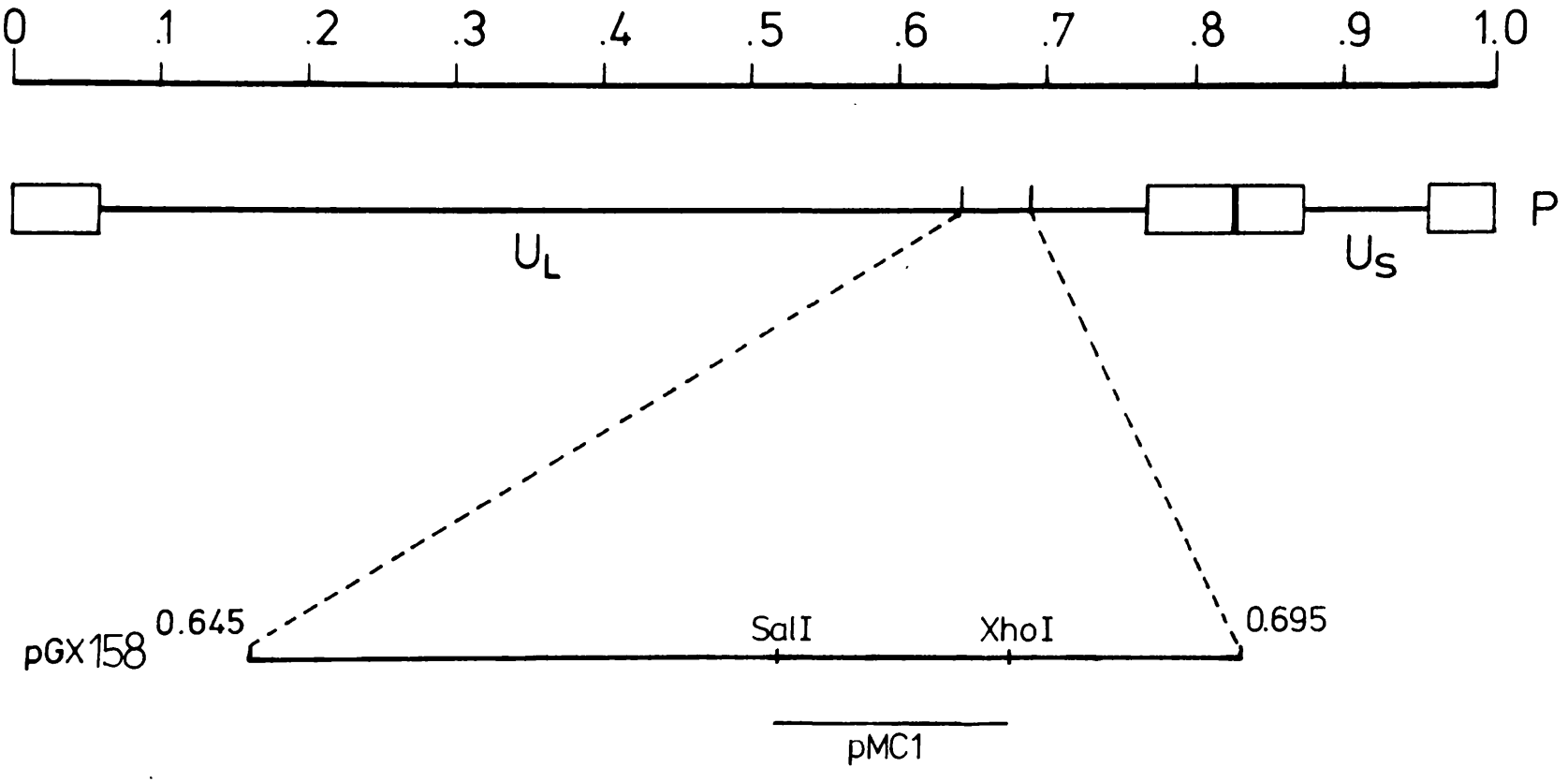

0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
P
U_L
U_S
pGX158
0.645
SalI
XhoI
0.695
pMC1

Figure 7: Location of the BamHI F fragment in the HSV-1 genome.

A diagrammatic representation of the HSV-1 prototype genome. BamHIF is located between 0.645 and 0.695 map units and is cloned in the plasmid pGX158. The region covered by the smallest clone with stimulatory activity (pMC1; Campbell et al, 1984) is marked as this plasmid was used to produce the end-labelled S1 probes (see Figure 12).

Figure 8: List of Gel lines in the BamHI F database.

The "contig line" section shows the number of the final contig (595, an arbitrary number) followed by the length of the contig (8055 bp) and the database numbers of the gels at the extreme left (240) and extreme right (401) hand ends. The "gel lines" section has six columns. The first column shows the experimental name assigned to the clone, the next five columns are headed A-E and are:

A - gel database number, assigned by the program.
B - position of the left-most base of the gel in the contig.
C - length of the gel reading. Negative indicate that the gel reading overlaps on the opposite strand.
D - number of the gel overlapping to the left.
E - number of the gel overlapping to the right.

CONTIG LINES

	595	8055	0	240	401
GEL LINES	A	B	C	D	E
ENDBIT.6	240	1	40	0	337
ENDBIT.10	337	1	-40	240	400
ENDBIT.13	400	1	142	337	399
ENDBIT.12	399	1	-209	400	233
ENDBIT.4	233	1	240	399	334
ENDBIT.7	334	1	247	233	317
G69.81	317	13	231	334	228
G51.617	228	32	151	317	172
G42.459	172	32	229	228	198
G48.608	198	37	216	172	298
G71.815	298	57	197	198	454
1177.G87	454	79	216	298	435
1155.G59	435	92	-242	454	162
G44.467	162	103	136	435	170
G45.485	170	105	-206	162	343
G82.926	343	108	-206	170	163
G44.470	163	114	-176	343	494
914EXT.RR	494	125	115	163	313
G69.82	313	149	-258	494	159
G42.452	159	154	-241	313	364
1012.G77	364	163	134	159	257
REP.452	257	181	-214	364	43
G11.148	43	207	87	257	345
G82.930	345	227	194	43	203
GELA.564	203	253	126	345	174
G42.449	174	268	-172	203	333
G81.947	333	295	266	174	119
G34.339	119	303	90	333	320
GX.786	320	307	179	119	289
G70.811	289	318	158	320	301
G73.833	301	380	211	289	495
870EX.PRR	495	418	97	301	485
1213.G90	485	435	-185	495	260
G67.870	260	515	-74	485	83
G18.240	83	570	172	260	367
1018.G77	367	603	144	83	244
G56.700	244	620	-291	367	438
1080.G53	438	681	-238	244	149
L.85	149	686	-216	438	492
1018EX.RR	492	719	86	149	493
240EXT.PRR	493	722	65	492	299
G71.827	299	768	190	493	56
G13.85	56	784	-168	299	388
1064.G79	388	785	-193	56	16
G3.27	16	793	182	388	184
G45.483	184	825	-241	16	404
1107.G85	404	827	130	184	21
G4.45	21	833	-152	404	51
G10R.133	51	834	-154	21	215
G51.594	215	889	237	51	73
G16.204	73	914	209	215	390
1058.G76	390	925	200	73	58
G15.186	58	937	-168	390	38
G11.142	38	963	213	58	407
1112.G85	407	974	-175	38	437
1067.G79	437	1006	230	407	264
G67.869	264	1045	-136	437	455
1178.G87	455	1061	-231	264	462
1191.G87	462	1087	-194	455	316
L.455	316	1100	-165	462	344
G82.929	344	1112	-257	316	321
GX.797	321	1116	186	344	449
1143.G59	449	1127	201	321	253
G55.682	253	1129	-218	449	467
1179.G87	467	1136	-240	253	3
G2R.11	3	1156	-240	467	124
G33.270	124	1166	-242	3	328
GX.796	328	1172	-217	124	284
GELH.682	284	1205	-142	328	125
G33.434	125	1205	-222	284	160
G42.455	160	1231	-114	125	2
G2R.10	2	1232	-165	160	456
1180.G87	456	1248	-215	2	158
R.270	158	1269	-139	456	255
REP.455	255	1278	-67	158	329
G81.958	329	1293	181	255	420
1148.G59	420	1377	-222	329	216
G51.595	216	1381	-183	420	265
G67.875	265	1395	-217	216	469
1184.G87	469	1399	-209	265	179
G42.462	179	1445	-249	469	414
1030.G58	414	1456	164	179	57
G13.86	57	1468	117	414	128
G34.333	128	1473	-103	57	432
G66.638	432	1481	189	128	102
G19.259	102	1498	235	432	353
G82.943	353	1517	180	102	150
L.86	150	1518	174	353	105
G21.295	105	1526	-76	150	130
G33.362	130	1532	191	105	92
G21.283	92	1569	227	130	406
1119.G85	406	1571	-151	92	193
G46.500	193	1577	143	406	154
R.274	154	1580	119	193	146
g20.274	146	1589	175	154	391
1040.G79	391	1599	269	146	88
G17.229	88	1654	166	391	272
G68.B6	272	1657	239	88	15
G3.25	15	1662	-84	272	109
G20.267	109	1698	183	15	194
G46.502	194	1721	175	109	305
G6.57	305	1721	239	194	55
G13.5	55	1730	92	305	491
274EXT.RR	491	1737	103	55	87
G17.220	87	1755	-200	491	306
G6.56	306	1772	-275	87	4
G2R.12	4	1802	-128	306	497
G5EXT.PRR	497	1807	41	4	143
G41.440	143	1842	-178	497	310
G75.906	310	1844	-213	143	461
1190.G87	461	1904	-132	310	483
1200.G88	483	1904	243	461	314
G11EXT.139	314	1913	-150	483	484
1201.G88	484	1927	-222	314	54
G13.4	54	1930	-108	484	278
A12.G70	278	1959	104	54	273
A2.G70	273	1959	105	278	275
A5.G70	275	1959	105	273	423
1047.G55	423	1990	195	275	37
G11.139	37	2039	-206	423	288
G70.806	288	2040	-209	37	12
G3.21	12	2083	-174	288	180
G42.451	180	2125	-295	12	290
G70.813	290	2146	189	180	349
G82.916	349	2158	-248	290	141
G41.444	141	2166	127	349	416
1032.G58	416	2213	-162	141	68
G16.202	68	2220	-187	416	370
1026.G78	370	2240	258	68	332
G82.935	332	2246	186	370	204
GELA.565	204	2264	-90	332	397
1061.G76	397	2280	263	204	368
1019.G77	368	2342	-103	397	464
1197.G88	464	2461	205	368	286
G70.802	286	2462	201	464	319
GX.784	319	2465	113	286	376
1024.G78	376	2520	278	319	458
1185.G87	458	2586	174	376	205
GELA.568	205	2617	189	458	157
R.266	157	2696	-63	205	354
G82.939	354	2696	-137	157	22
G4.46	22	2710	92	354	318
G80.768	318	2714	108	22	419
1147.G59	419	2722	198	318	450
1145.G59	450	2727	224	419	238
G56.702	238	2743	-233	450	342
G82.921	342	2758	-145	238	281
A7.G70	281	2769	117	342	308
G69.78	308	2797	-151	281	348
G83.994	348	2807	173	308	460
1189.G87	460	2861	-171	348	417
1034.G58	417	2861	214	460	190
G46.501	190	2892	-197	417	263
G67.868	263	2921	218	190	84
G18.243	84	2938	159	263	327

GX.793	327	2940	183	84	445
1075.G61	445	2964	-276	327	398
1060.G76	398	3008	-224	445	302
G75.900	302	3034	-178	398	191
G46.511	191	3034	-184	302	389
1057.G76	389	3040	160	191	276
A8.G70	276	3068	-120	389	122
G34.344	122	3075	63	276	161
G44.466	161	3108	112	122	112
G20.278	112	3119	-78	161	189
G46.498	189	3162	-194	112	429
1162.G86	429	3165	-181	189	411
1025.G58	411	3165	-187	429	267
G68.B3	267	3174	-185	411	71
G16.212	71	3179	-203	267	91
G18.248	91	3179	214	71	280
A14.G70	280	3189	-104	91	415
1031.G58	415	3203	-166	280	147
G20.277	147	3226	-130	415	165
G44.472	165	3229	75	147	195
G47.601	195	3243	166	165	307
G75.905	307	3246	190	195	208
GELA.569	208	3258	-190	307	357
G82.920	357	3259	-247	208	219
G51.618	219	3271	-128	357	426
1167.G86	426	3286	178	219	470
1188.G87	470	3289	-234	426	177
G44.474	177	3295	159	470	430
G66.635	430	3316	-222	177	434
1146.G59	434	3324	193	430	173
G42.463	173	3335	-121	434	422
1149.G59	422	3346	253	173	234
G57.730	234	3351	-226	422	156
R.269	156	3367	-156	234	431
G66.637	431	3391	258	156	405
1114.G85	405	3409	252	431	140
G41.443	140	3420	-170	405	377
1025.G78	377	3441	-234	140	472
1096.G74	472	3455	231	377	229
G57.727	229	3481	156	472	359
G83.983	359	3501	-186	229	338
G81.948	338	3519	-236	359	82
G17.233	82	3522	-150	338	42
G11.147	42	3587	-104	82	274
A3.G70	274	3587	-104	42	182
G44.476	182	3609	-178	274	138
G40.421	138	3613	-140	182	79
G17.221	79	3638	-227	138	95
G19.250	95	3645	-165	79	403
1106.G85	403	3647	190	95	155
R.271	155	3657	96	403	283
H2PT1.G70	283	3659	136	155	424
1165.G86	424	3688	203	283	86
G21.286	86	3725	180	424	365

1013.G77	365	3734	105	86	121
G34.342	121	3747	-140	365	100
G19.252	100	3762	-151	121	211
GELB.575	211	3782	-230	100	433
1022.G58	433	3784	179	211	277
A11.G70	277	3843	-161	433	410
1024.G58	410	3857	-195	277	8
G5.97	8	3881	-252	410	279
A13.G70	279	3926	-76	8	28
G10R.135	28	3942	-205	279	218
G51.599	218	3996	189	28	227
G51.616	227	3997	130	218	440
G57.616	440	3997	266	227	199
G48.609	199	4005	140	440	196
G47.603	196	4012	253	199	294
G70.808	294	4018	-245	196	446
1083.G53	446	4061	243	294	101
G19.254	101	4142	-208	446	207
G47.606	207	4160	-195	101	363
1009.G77	363	4164	-141	207	49
G12.166	49	4190	161	363	292
G70.805	292	4215	-191	49	164
G44.471	164	4238	-120	292	45
G12.154	45	4280	-139	164	166
G44.477	166	4293	-173	45	39
G11.144	39	4316	187	166	271
G68.M7	271	4329	145	39	168
G44.480	168	4341	-159	271	258
G67.873	258	4362	176	168	183
G44.478	183	4396	150	258	309
G68.51	309	4444	-120	183	361
TGG	361	4457	141	309	188
G46.497	188	4471	-83	361	452
1027.G58	452	4486	-177	188	60
G15.191	60	4493	195	452	30
G10R.132	30	4533	-168	60	285
G70.799	285	4544	186	30	270
G68.M5	270	4558	-181	285	13
G3.20	13	4571	176	270	256
RG3.20	256	4571	237	13	206
GELB.579	206	4581	-154	256	402
1104.G85	402	4581	-207	206	340
G81.953	340	4588	-256	402	362
G73.834	362	4603	259	340	478
1210.G88	478	4636	-259	362	65
G12.162	65	4645	-181	478	237
G54.662	237	4681	-302	65	167
G44.479	167	4690	-105	237	243
G56.704	243	4717	223	167	386
1041.G79	386	4727	223	243	103
G21.290	103	4742	160	386	63
G13.2	63	4749	-162	103	413
1029.G58	413	4749	-195	63	44
G11.149	44	4750	-109	413	25

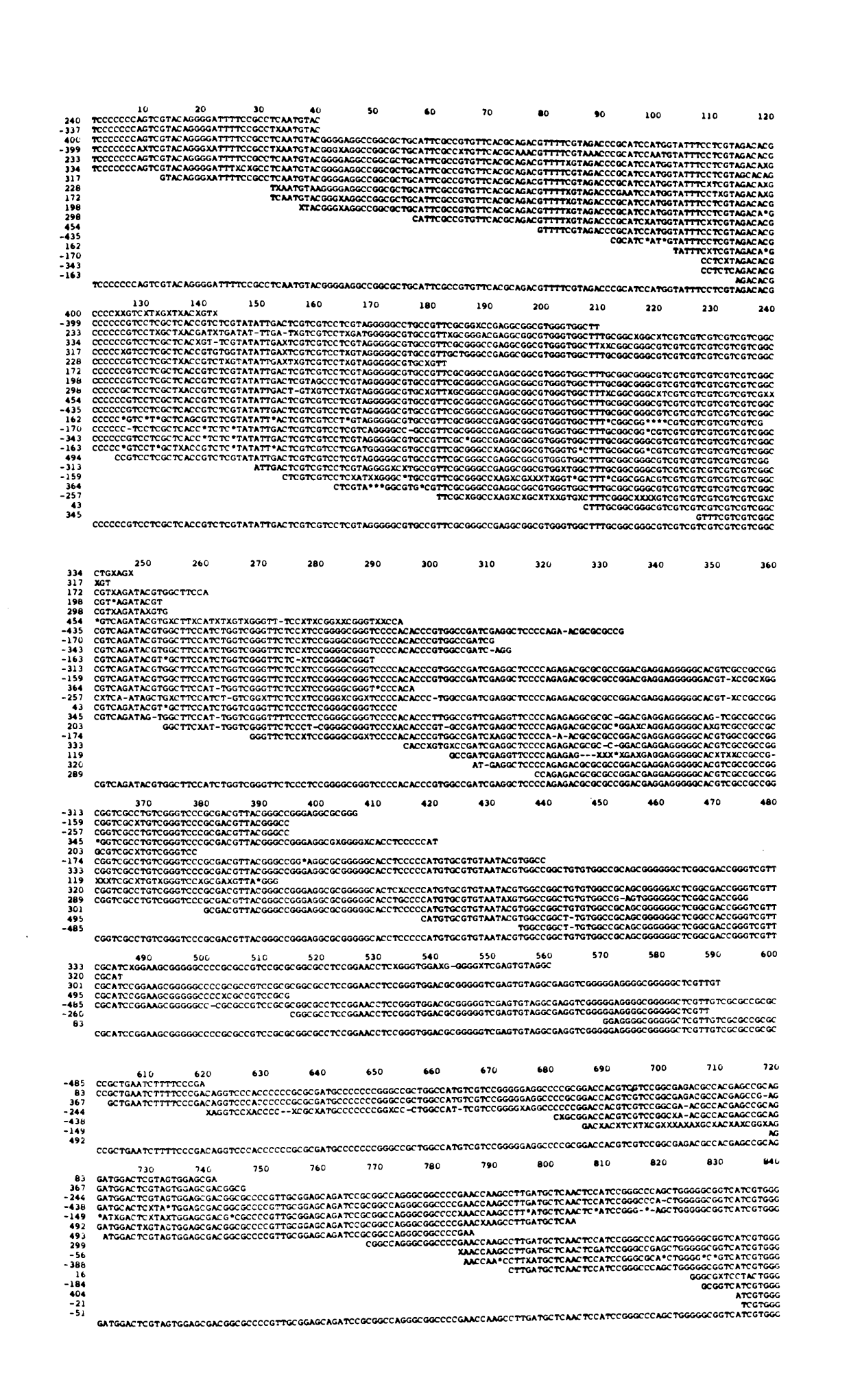
G4.41	25	4775	-115	44	129
G33.338	129	4775	-193	25	235
G57.731	235	4778	-281	129	352
G82.941	352	4788	-124	235	118
G34.335	118	4812	71	352	61
G10R.128	61	4819	-222	118	67
G12.163	67	4823	-139	61	76
G16.213	76	4852	-140	67	1
G2R.7	1	4853	-206	76	117
G34.332	117	4897	-76	1	384
1015.G77	384	4898	-220	117	212
GELB.578	212	4915	96	384	480
1218.G90	480	4920	-202	212	436
1073.G60	436	4928	-268	480	32
G10R.127	32	4934	-172	436	408
1113.G85	408	4993	190	32	382
1002.G77	382	4997	186	408	5
G3.19	5	5004	188	382	266
G67.877	266	5022	-182	5	465
1198.G88	465	5028	244	266	341
G82.932	341	5059	-184	465	296
G71.828	296	5081	124	341	297
G71.829	297	5109	-183	296	33
G10R.126	33	5117	218	297	231
G54.635	231	5121	142	33	245
G54.665	245	5121	143	231	287
G70.804	287	5128	-207	245	230
G54.664	230	5176	224	287	169
G45.481	169	5223	-156	230	418
1034.G58	418	5304	167	169	303
G75.909	303	5309	225	418	351
G82.940	351	5317	147	303	151
L.98	151	5340	-161	351	75
G16.210	75	5365	-212	151	350
G80.772	350	5423	170	75	9
G5.98	9	5439	-119	350	479
1219.G90	479	5442	193	9	409
1118.G85	409	5443	274	479	148
L.23	148	5467	-168	409	23
G3.23	23	5510	-196	148	226
G51.596	226	5532	-132	23	142
G40.433	142	5547	-126	226	481
1209.G88	481	5554	169	142	383
1006.G77	383	5559	-133	481	269
G68.B5	269	5590	178	383	197
G47.607	197	5596	-79	269	412
1028.G58	412	5606	172	197	489
103OEXT.RR	489	5638	59	412	490
631EXT.PRR	490	5638	62	489	451
1150.G59	451	5661	245	490	379
1030.G78	379	5680	-129	451	443
G66.631	443	5680	-153	379	291
G70.814	291	5687	-231	443	152
L.165	152	5689	-160	291	97

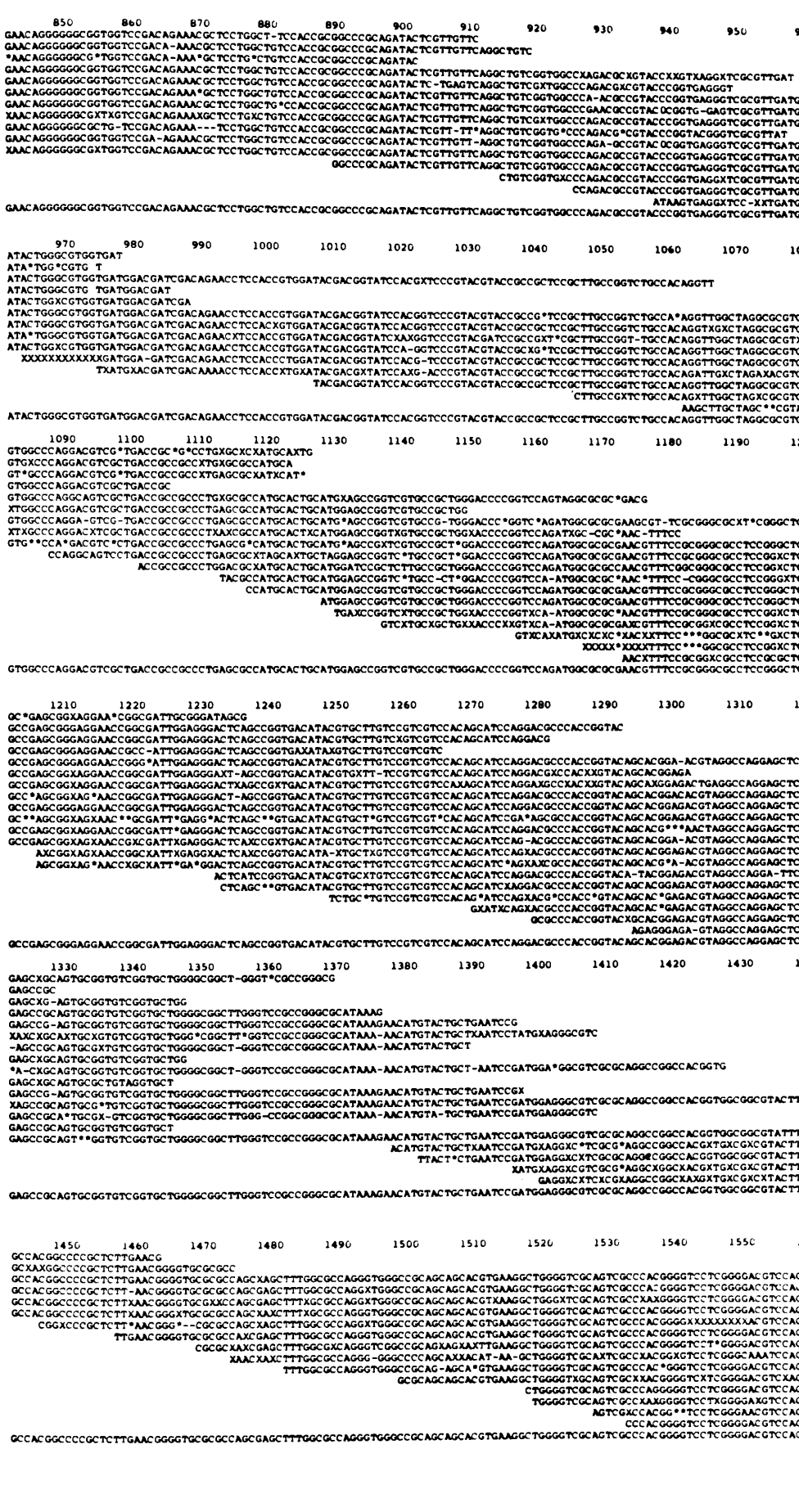
G19.260	97	5737	165	152
G12.165	48	5762	-132	97
G12.164	47	5787	165	48
G45.489	185	5793	-150	47
G80.769	326	5825	-146	185
G80.767	322	5838	-90	326
1176.G87	453	5860	191	322
G33.439	131	5874	104	453
G41.439	139	5874	145	131
97EXT.PRR	488	5880	108	139
G20.280	114	5903	180	488
G83.993	347	5939	-155	114
G68.84	268	5963	-159	347
L.280	315	5971	226	268
1182.G87	468	6021	-225	315
1195.G88	463	6023	-238	468
H2PT2.G70	282	6038	99	463
G49.541	200	6040	146	282
G67.867	259	6041	178	200
GX.792	324	6061	-172	259
G17.218	77	6068	-271	324
G10R.134	29	6072	186	77
G44.475	181	6084	-159	29
G45.487	187	6137	-182	181
G10R.124	34	6162	166	187
G42.450	175	6211	-85	34
G16.206	74	6218	191	175
1084.G53	439	6223	-223	74
G16.214	72	6225	78	439
1186.G87	459	6238	208	72
1017.G77	366	6239	126	459
G21.288	132	6252	-186	366
G82.925	358	6294	239	132
G68.49	373	6331	266	358
1151.G59	421	6340	208	373
G4.42	19	6342	-102	421
G49.542	201	6362	-101	19
1159.G86	428	6390	195	201
1023.G78	369	6398	-154	428
G68.48	372	6416	-240	369
G33.436	126	6420	-137	372
1205.G88	476	6423	251	126
1022.G78	375	6425	-233	476
G68.63	374	6429	162	375
G41.436	144	6455	-102	374
G34.340	120	6476	65	144
G45.490	186	6487	203	120
G82.942	356	6490	136	186
G15.190	62	6518	-180	356
1141.G59	447	6518	-200	62
G82.932	331	6530	-210	447
L.190	153	6532	-123	331
1045.G79	387	6533	-133	153
G4.35	17	6567	198	387
48	1029.G78	378	6603	282	17	104
---	---	---	---	---	---	---
47	G21.292	104	6618	-190	378	26
185	G10R.137	26	6636	87	104	355
326	G80.775	355	6642	158	26	381
322	1003.G77	381	6651	214	355	323
453	GX.788	323	6700	-228	381	171
131	G45.488	171	6721	-79	323	496
139	137HG10.EX	496	6723	140	171	441
488	G66.634	441	6739	199	496	176
114	G44.473	176	6765	-155	441	346
347	G83.992	346	6783	-151	176	217
268	G51.598	217	6785	-131	346	293
315	G70.812	293	6802	-232	217	202
468	GELA.559	202	6814	-70	293	123
463	G34.345	123	6837	78	202	6
282	G5.95	6	6847	-207	123	475
200	1199.G88	475	6890	194	6	394
259	1065.G79	394	6941	177	475	46
324	G12.156	46	7003	111	394	487
77	59EXT.RR	487	7053	80	46	360
29	G83.991	360	7053	-135	487	220
181	G51.621	220	7053	-160	360	486
187	145EXT.PRR	486	7101	88	220	59
34	G15.187	59	7124	-141	486	145
175	G20.268	145	7165	-152	59	371
74	G68.47	371	7173	-293	145	78
439	G17.219	78	7183	286	371	393
72	1044.G79	393	7190	191	78	7
459	G5.96	7	7191	216	393	241
366	G56.698	241	7192	204	7	252
132	G50.531	252	7198	152	241	64
358	G12.159	64	7200	127	252	93
373	G21.285	93	7212	238	64	213
421	G47.600	213	7241	178	93	425
19	1166.G86	425	7244	-197	213	66
201	G11.140	66	7253	262	425	24
428	G3.32	24	7261	146	66	127
369	G34.330	127	7269	217	24	312
372	G75.902	312	7274	206	127	242
126	G56.703	242	7308	-295	312	52
476	G10R.129	52	7319	-169	242	466
375	1187.G87	466	7334	223	52	90
374	G18.245	90	7357	-143	466	444
144	G66.633	444	7362	-176	90	137
120	G40.419	137	7380	-126	444	236
186	G57.736	236	7402	-197	137	178
356	G42.461	178	7402	-289	236	10
62	G5.99	10	7409	187	178	395
447	1066.G79	395	7423	-235	10	210
331	GELA.574	210	7426	-187	395	448
153	1142.G59	448	7453	-135	210	457
387	1181.G87	457	7480	-221	448	99
17	G19.251	99	7489	-165	457	107
378	G21.294	107	7502	111	99	133

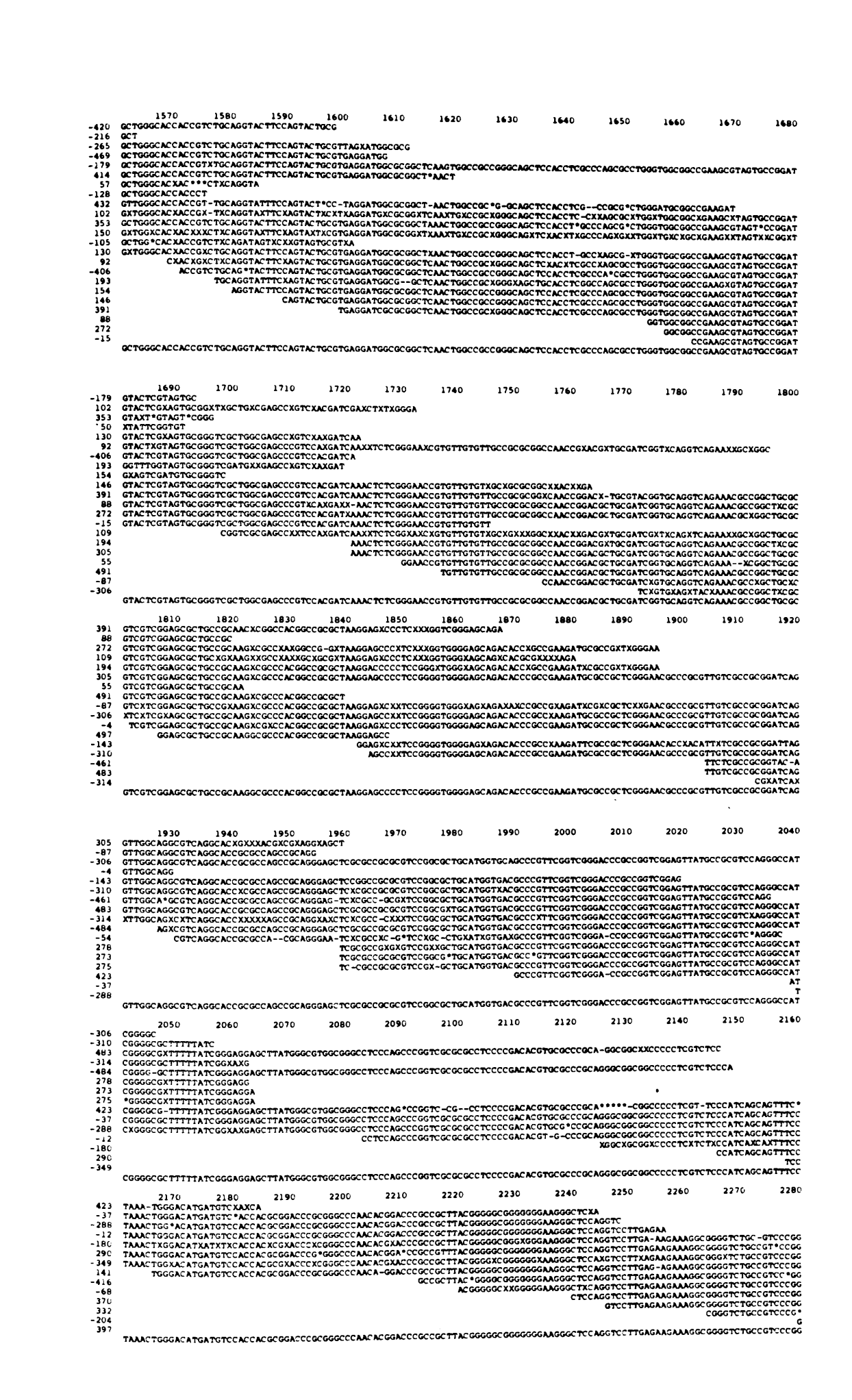
G40.418	133	7537	-62	107	339
G81.951	339	7539	-246	133	70
G16.211	70	7555	189	339	311
G75.904	311	7581	-172	70	209
GELA.571	209	7600	174	311	89
G18.242	89	7611	165	209	94
G17.224	94	7653	105	89	380
1933.G78	380	7671	-298	94	477
1207.G88	477	7678	161	380	482
1206.G88	482	7705	242	477	50
G12.167	50	7763	147	482	471
1095.G74	471	7764	125	50	392
1043.G79	392	7781	203	471	261
G67.871	261	7794	-117	392	396
1055.G76	396	7811	-229	261	442
G66.629	442	7811	-238	396	221
G51.627	221	7840	-209	442	254
G5.101	254	7851	-180	221	295
G71.821	295	7857	-160	254	31
G10R.131	31	7901	-105	295	336
ENDBIT.9	336	7915	-141	31	239
ENDBIT.5	239	7936	120	336	232
ENDBIT.3	232	7987	69	239	335
ENDBIT.8	335	8022	34	232	401
ENDBIT.14	401	8023	33	335	0

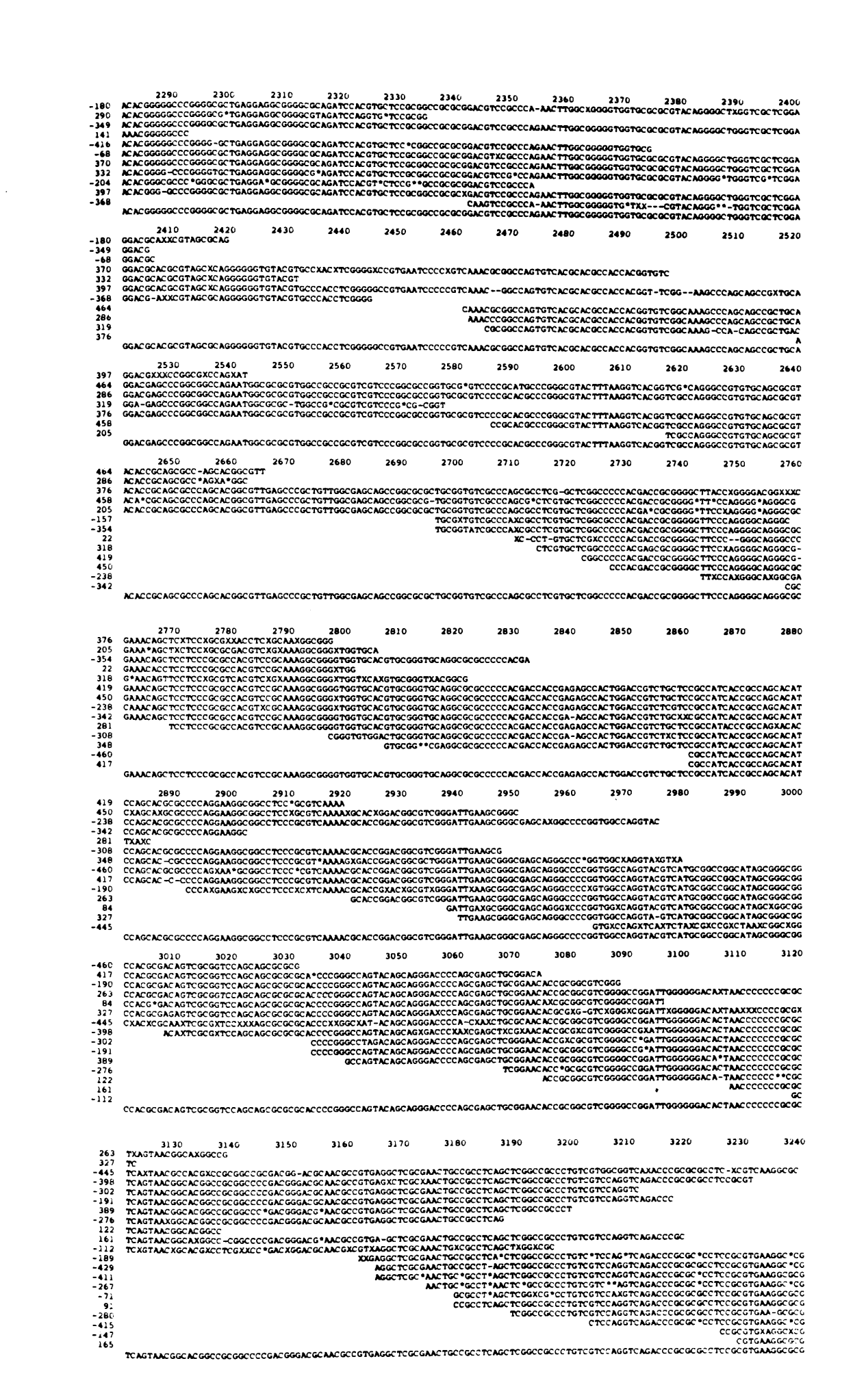
Figure 9: The BamHI F database.

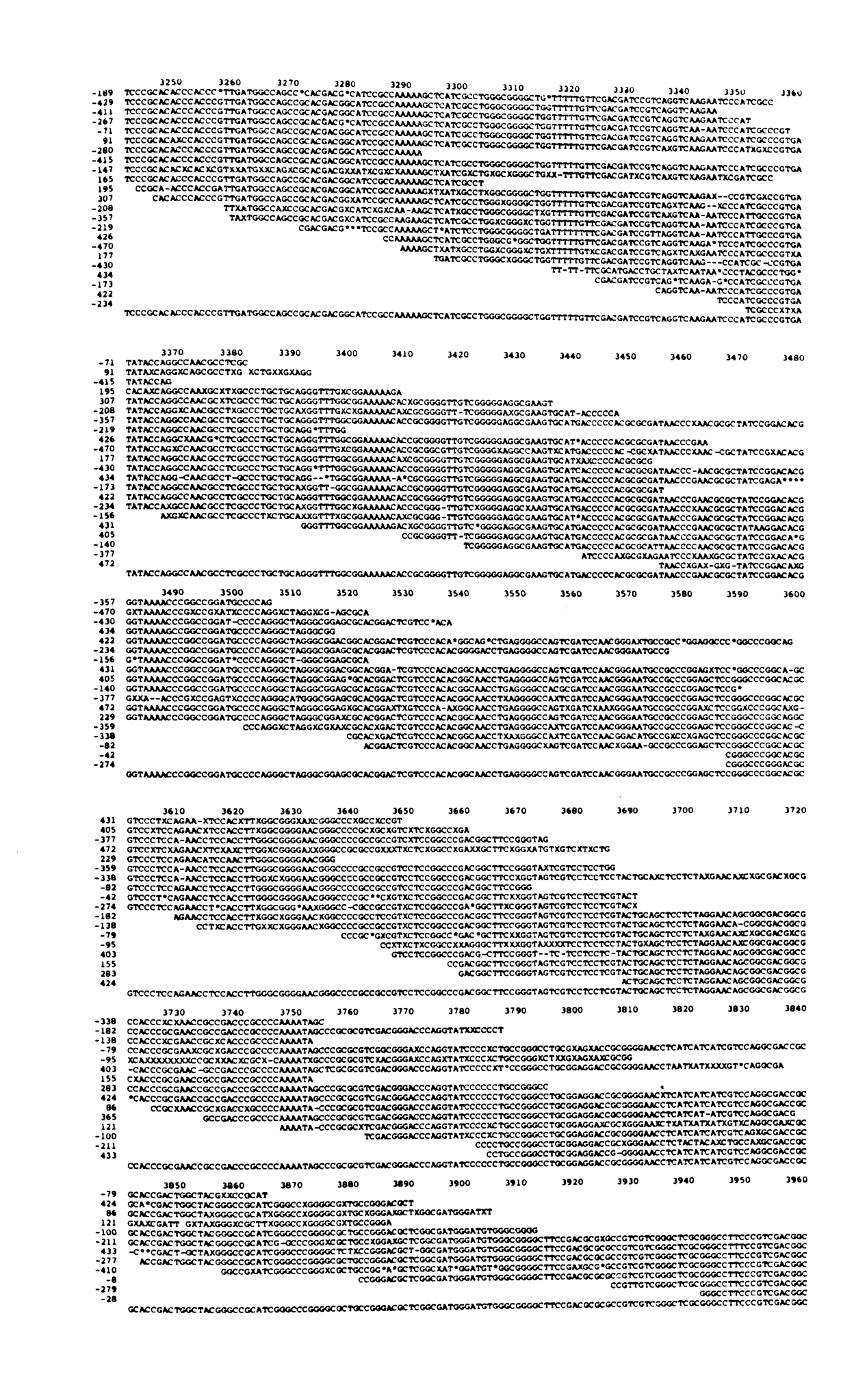
All gel readings are shown with their corresponding overlaps. The figures down the right side are gel numbers, which can be cross-referenced to figure 8, column A. Computer inserted "padding characters" are * and those inserted by the operator are X.

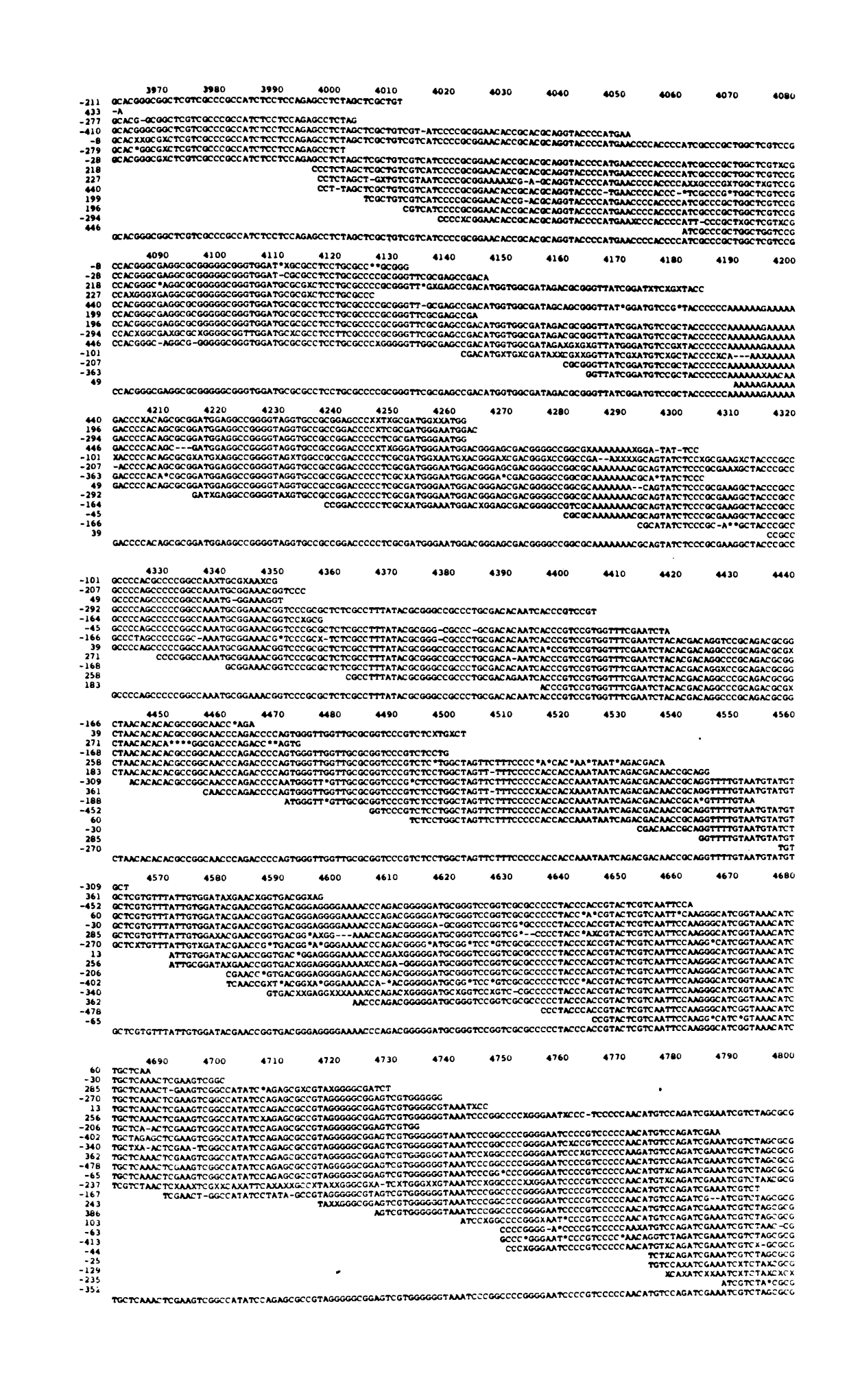


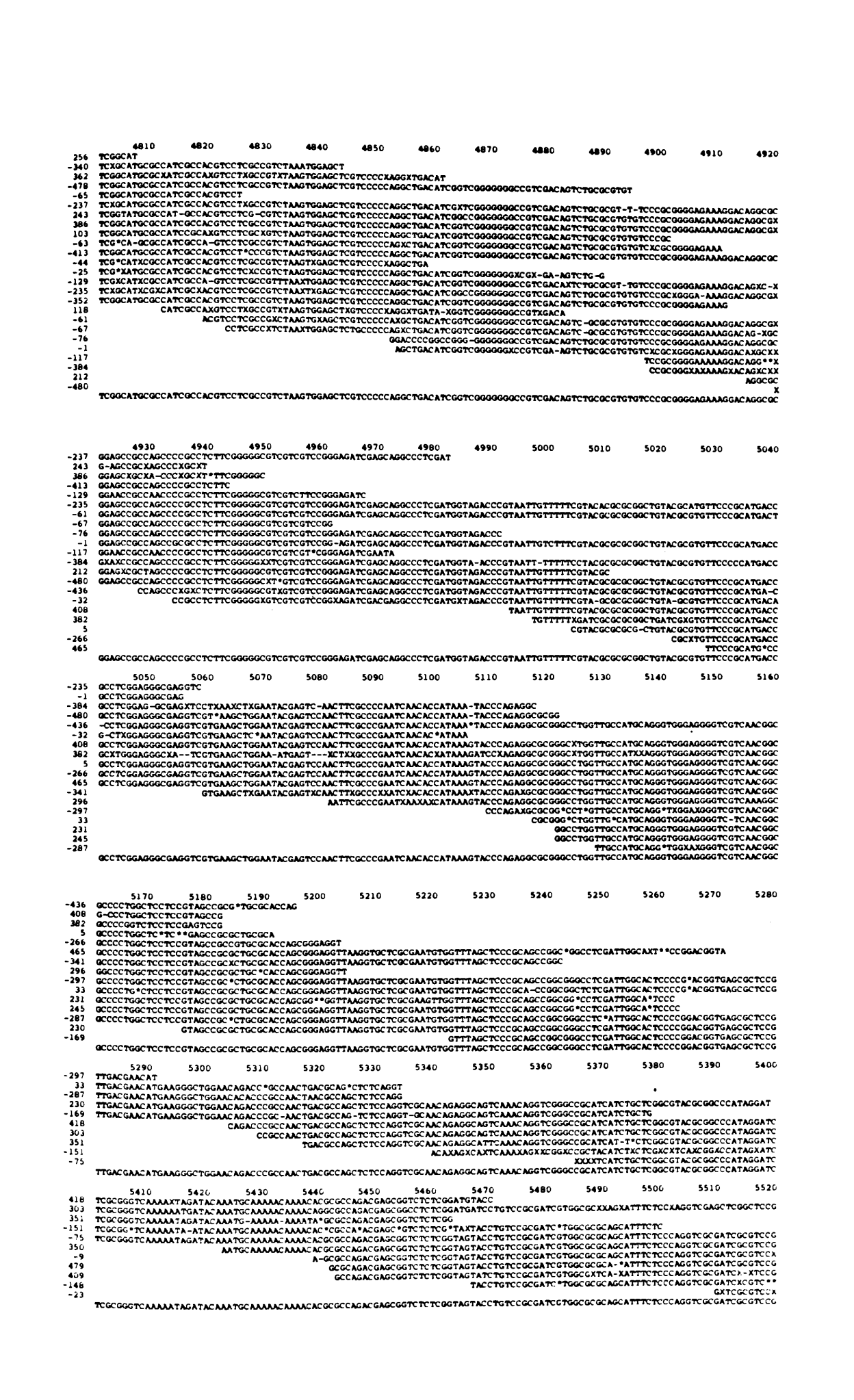


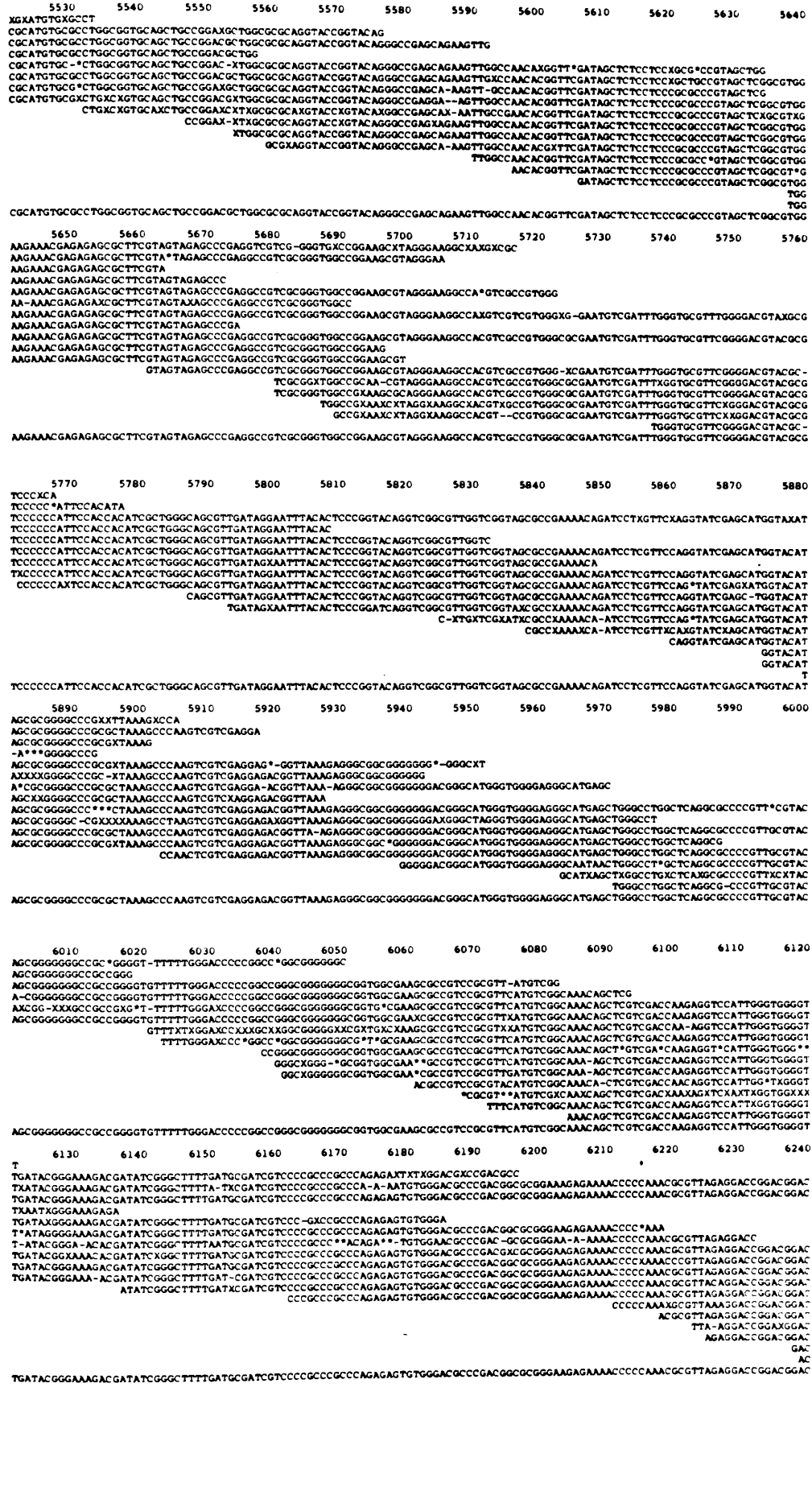


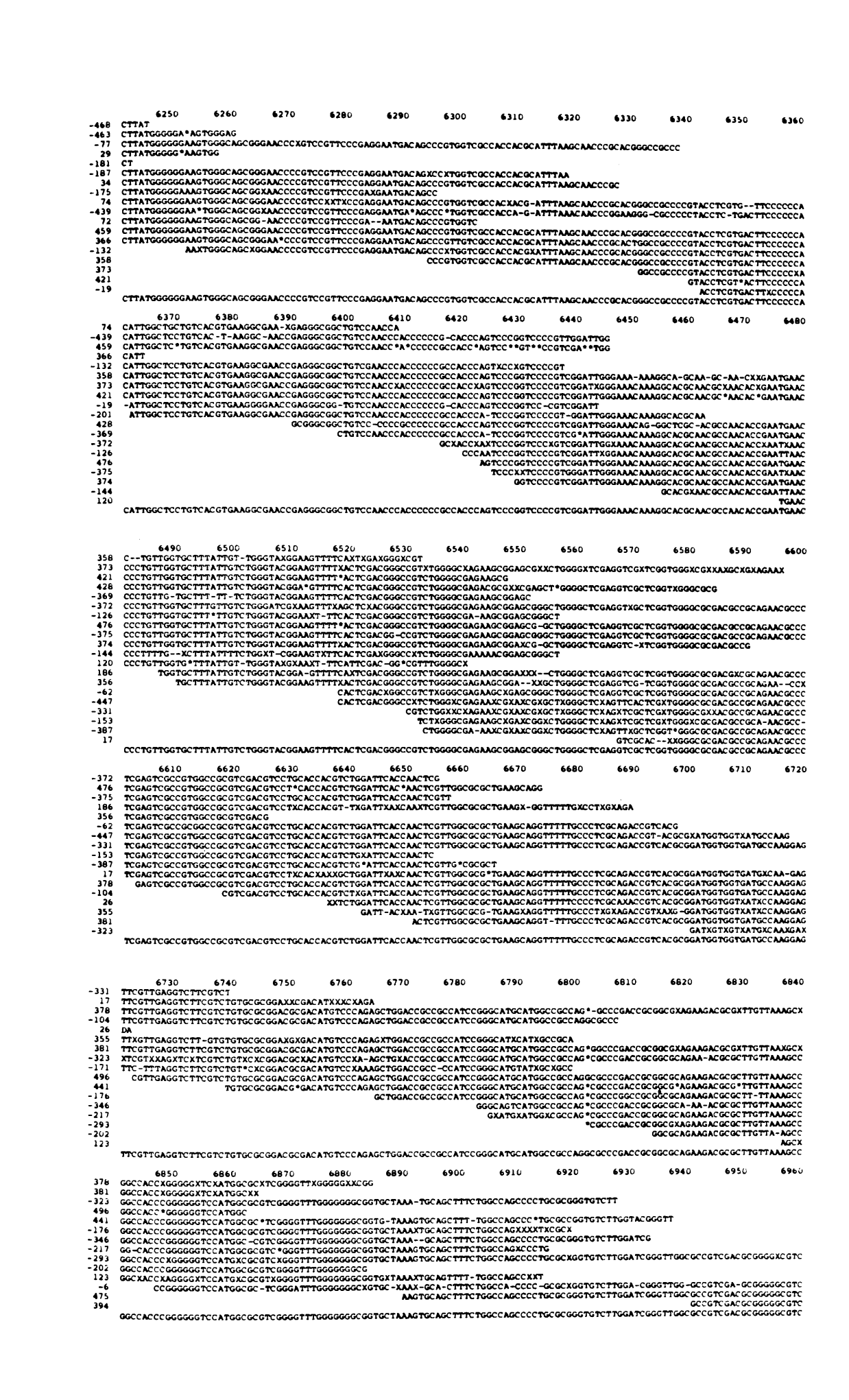


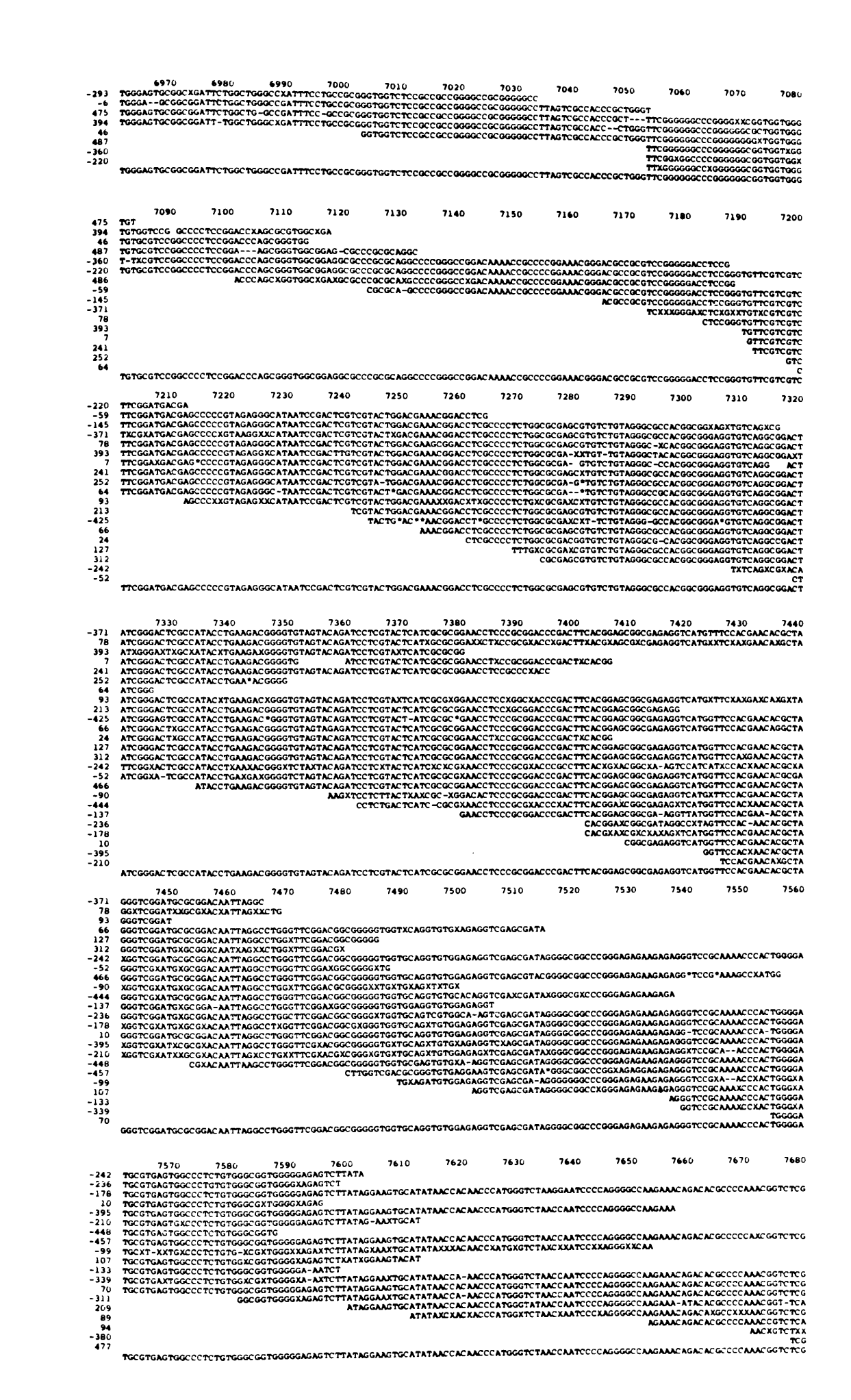


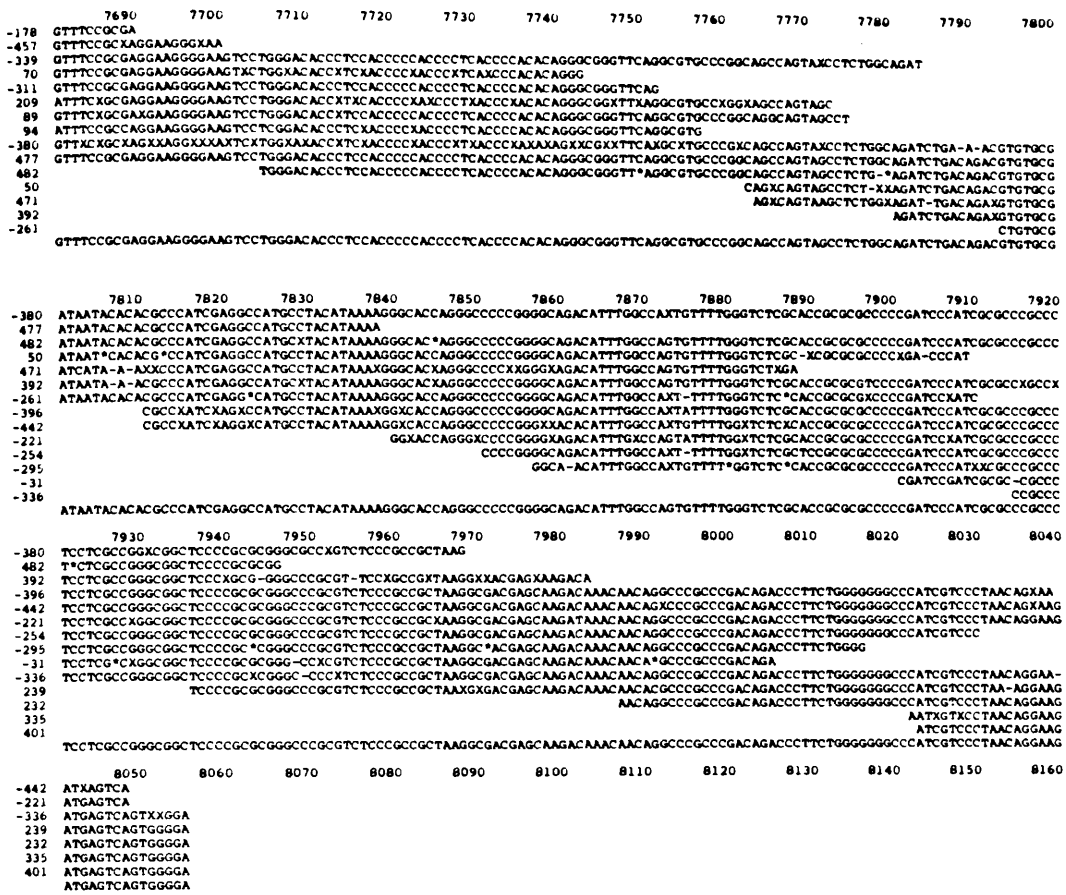












Most of the data were obtained using the gradient gel system. However, there were several regions of ambiguity as a result of so-called gel compressions. These are caused by formation of local areas of secondary structure, which are stable enough to resist denaturation under the usual gel conditions, in the newly synthesised strand. Such molecules have aberrant mobilities and give false gel readings which were recognised either as obvious peculiarities on the autoradiograph or by the detection of frameshifts in a clear ORF, by use of the FRMSCN program. The solution to the problem is to design gel conditions which destabilise the secondary structure. The method used throughout this study was to run "hot-gels", that is, standard sequence reactions run on single concentration gels which were artificially heated by means of a water jacket. Typical examples of sequencing autoradiographs, which include examples of compressions and their resolution, are shown in Figures 10 and 11.

5.1.2 Interpretation of the Sequence Data

The location of protein coding sequences in BamHI F was based on several lines of evidence (the principles by which named computer programs work are discussed in section 4.9):

(a) the identification of ORFs with the DSPLAY program.

(b) the existance of a typical HSV-1 protein codon usage within ORFs, as revealed by the FRMSCN program.

(c) the published mRNA map (Hall et al, 1982; see Figure 12a). In the case of the mRNA for Vmw65 detailed S1 nuclease mapping was performed.

(d) homology to ORFs in the analogous region of the VZV genome (Dr.A.Davison, personal communication).

Figure 12b shows all open reading frames in the sequence of BamHI F in the prototype orientation. Those open reading frames which coincide with regions which exhibit a codon usage consistent with a standard HSV-1 protein coding usage are shown by heavy lines. Fortunately the relatively high %G+C content of HSV-1 DNA, which results to a great

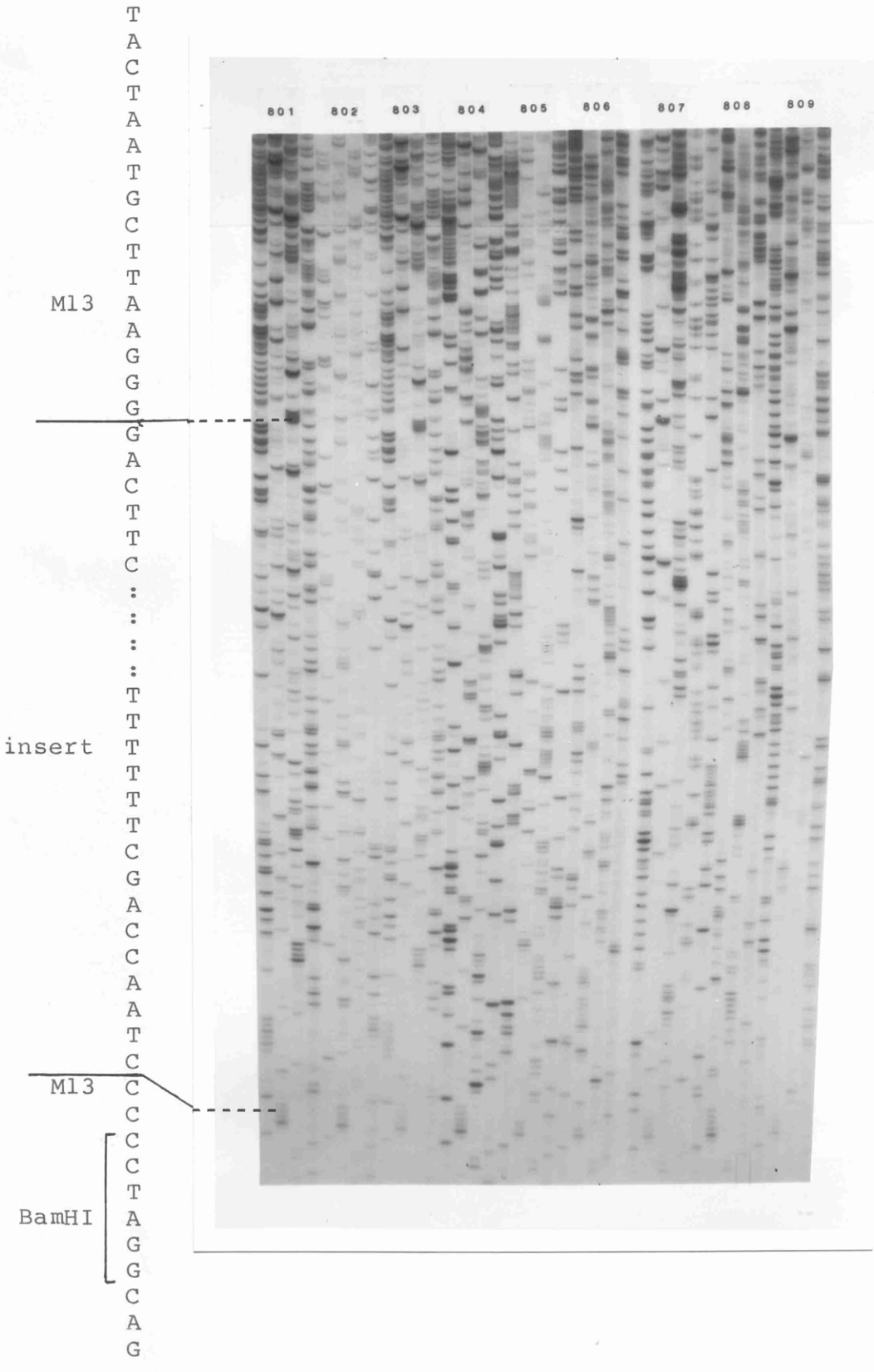

801
802
803
804
805
806
807
808
809
M13
insert
M13
BamHI
T
A
C
T
A
A
T
G
C
T
T
A
A
G
G
G
G
G
A
C
T
T
C
T
T
T
C
G
A
C
C
A
A
T
C
C
C
C
C
T
A
G
G
C
A
G

Figure 10: Examples of Sequence Autoradiographs I.

This figure shows half of the results obtained from a typical wide gradient sequencing gel. The clone numbers are marked along the top. Down the side of the figure are detailed some features of the sequences, the order of reactions being TCGA. In most clones the "top " and "bottom" of the insert can be seen (marked on figure), this being discerned from the characteristic M13 sequences shown. In particular, the pattern of T residues is immediately recognisable as M13 sequence. The size of the inserts from this cloning experiment vary considerably and some are very small. Small inserts are preferentially cloned into M13 phage and if possible should be avoided. Other features of the sequence to be noted are:

(a) The first C or G band in a run of C's or G's is often feint and immediately followed by a very intense band which may mask the first.

(b) The first A or T band in a run of A's or T's is almost always very intense compared to subsequent bands.

These features can often be used diagnostically to confirm the presence of a particular base during a gel reading.

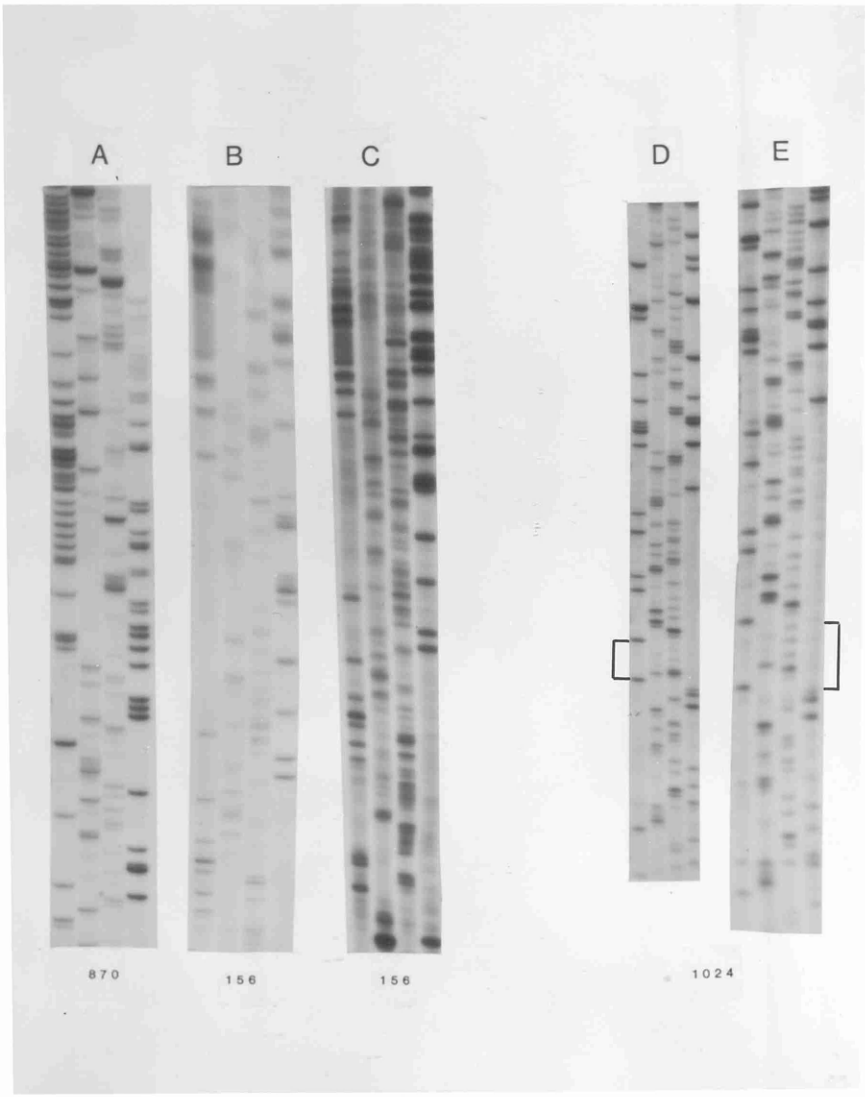

A
B
C
D
E
870
156
156
1024

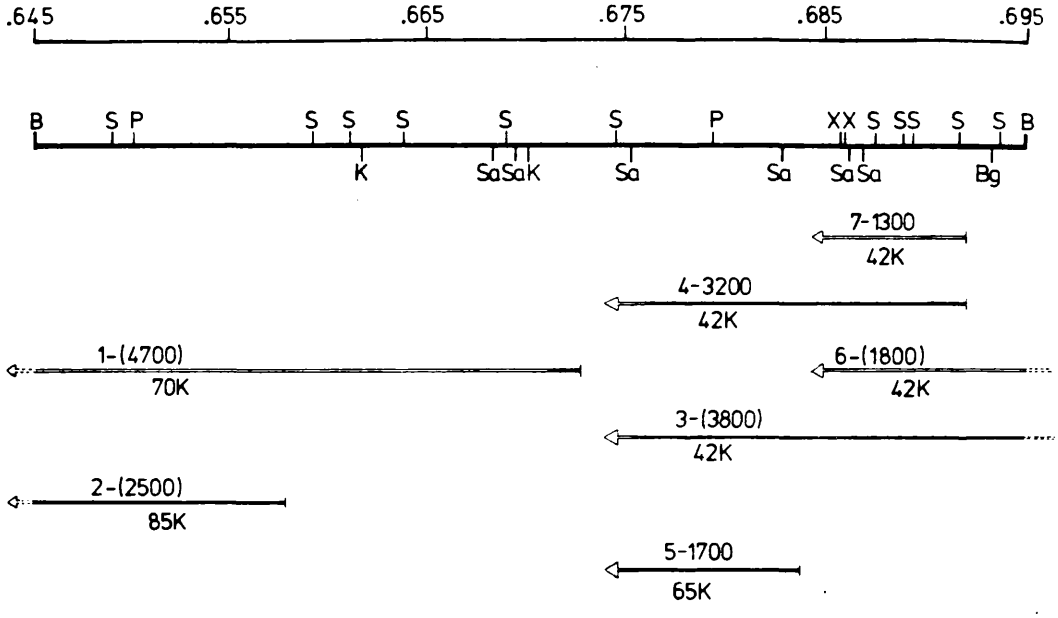
Figure 11: Examples of Sequence Autoradiographs II.

Two examples of sequence artifacts are shown.

Panel A shows a typical sequence, much like those presented in Figure 10, however, panel B shows a similar region of a clone which is particularly high in G and C residues. It was noted during the project that long clones reading into this region gave "fuzzy" bands after about 100-150 bases, presumably due to the formation of some secondary structure in the newly synthesised strand which was not resolved on a gradient sequence gel. Panel C shows a sequence reaction with the same clone (156) but run on a "hot gel" at 87°C. Although not perfect, the bands are much sharper and more intense. Such gels, coupled with shorter clones in this region, allowed the unambigous resolution of the sequence.

Panels D and E show the resolution of a typical compression due to the aberrant migration of newly synthesised DNA able to form a stable secodary structure. D. shows the result of running the sequence reaction on a normal gel, the boxed region defines the compression which might be read as: T, G, C, G, G, C, G, C, G, T. The region from reactions run on a hot gel (panel E) reads: T, G, G, C, G, C, G, C, G, T. Compressions are often obvious, as in this case, but may be indicated only by the absence of a band.

(a)


.645
.655
.665
.675
.685
.695
B S P S S S S S P XX S SS S S B
K Sa Sa K Sa Sa Sa Sa Bg
7-1300
42K
4-3200
42K
1-(4700)
70K
6-(1800)
42K
3-(3800)
42K
2-(2500)
85K
5-1700
65K

(b)

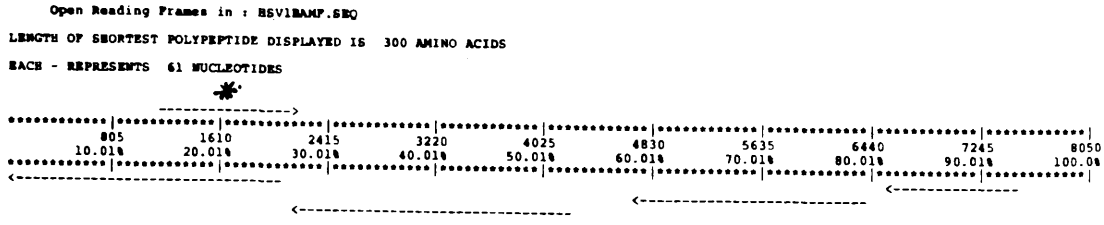

Open Reading Frames in : BSV1BAMF.SEQ
LENGTH OF SHORTEST POLYPEPTIDE DISPLAYED IS 300 AMINO ACIDS
EACH - REPRESENTS 61 NUCLEOTIDES
805 1610 2415 3220 4025 4830 5635 6440 7245 8050
10.01% 20.01% 30.01% 40.01% 50.01% 60.01% 70.01% 80.01% 90.01% 100.0%

Figure 12:

(a) The mRNA Map of Hall et al (1982)

The positions of the seven mRNAs mapped to BamHI F are shown aligned against a partial restriction enzyme map of BamHI F taken from the sequence data. The number assigned to each mRNA is the same as that used by Hall and co-workers. The approximate, experimentally determined, length of each mRNA is given in nucleotides. Brackets indicate that the whole mRNA does not map in BamHI F. Abbreviations used for restriction enzyme cleavage sites are: B, BamHI; Sa, SalI; P, PvuII; K, KpnI; S, SmaI; X, XhoI; Bg, BglII.

(b) Output from the program DSPLAY

This program records the positions of ORFs beside a rough map of the DNA sequence of interest.

This figure is available as a loose sheet at the back of the thesis.

There is an additional ORF, denoted by *, for which no mRNA was mapped by Hall et al (1982; see above). Evidence, discussed in the text, would predict that no such gene is transcribed during the HSV-1 lytic cycle.

extent from changes in the third base position, gives coding regions a particularly biased codon usage. This means that such an analysis can, in many cases, positively identify protein coding sequences and obvious frameshift errors in the primary data. A typical graphical output from FRMSCN is shown in Figure 13.

These two pieces of evidence were combined with the mRNA mapping data of Hall and co-workers (1982) to predict the existence of four protein coding sequences in the sequence of BamHI F, consistent with the same authors' *in vitro* translation data. A complete sequence listing and translation of all four hypothetical ORFs is presented in Figure 14. A provisional nomenclature of the form BmFX has been decided upon to indicate that the ORF lies in the BamHI F fragment and has approximate molecular weight X,000. It is likely that this nomenclature will be changed, when the complete sequence of the long unique region is known, to tie in with the system used to name hypothetical ORFs in the short unique region (McGeoch *et al*, 1985). The ORF which corresponds to the gene for Vmw65 has been labelled as such. The prime notation given to BmF72', the leftmost ORF, is meant to indicate that the ORF is incomplete and continues into the adjacent restriction fragment.

5.2 The Gene for Vmw65

A subclone of that region of BamHI F covering the gene for Vmw65 had been constructed previously and named pMC1 (Campbell *et al*, 1984). It consists of a SalI (position 4099) to XhoI (position 1490) fragment cloned into the vector pUC9. An S1 mapping experiment was performed in order to map the 5' terminus of the mRNA for Vmw65. Ten micrograms of infected or mock-infected cell RNA (isolated 6 hours after infection) was hybridised to a fragment, uniquely 5' end-labelled at an MluI site (position 1833) or SalI site (position 1958), prepared from pMC1. A similar experiment was performed to locate the 3' terminus of the mRNA using a

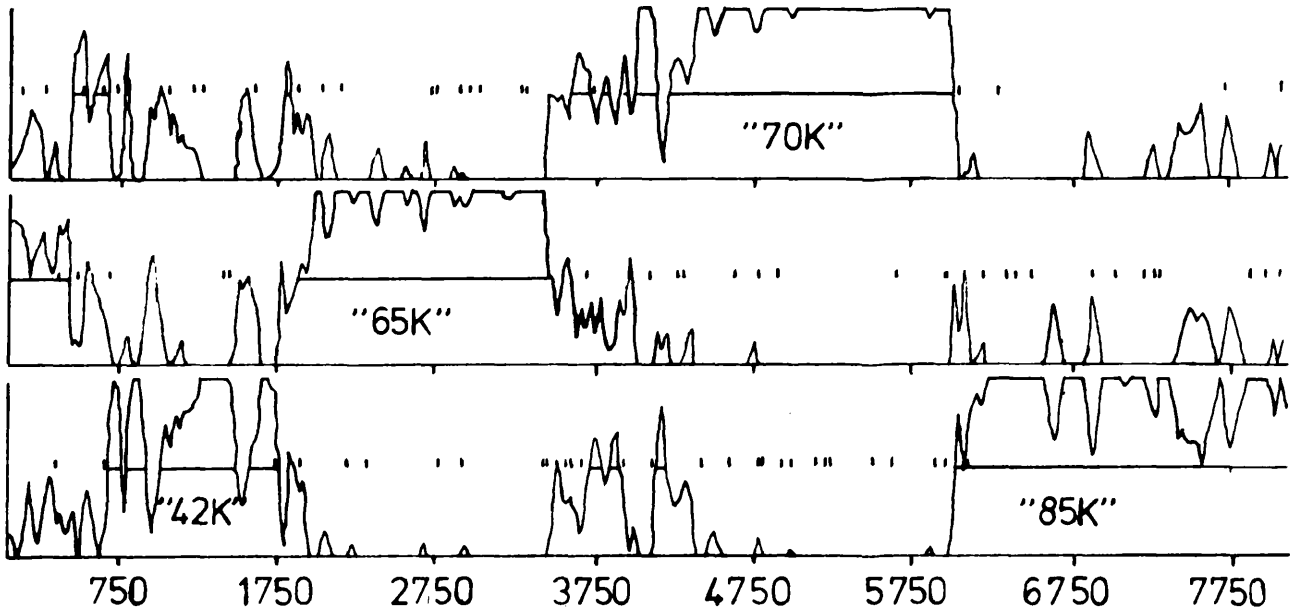

"70K"
"65K"
"42K"
"85K"
750
1750
2750
3750
4750
5750
6750
7750

Figure 13: Graphical Output from the Program FRMSCN

This program (described in section 4.9.2) clearly reveals the positions of the four genes in BamHI F, which are labelled according to the molecular weights of the *in vitro* translated products (Hall *et al*, 1982). The x-axis is the sequence of BamHI F and the y-axis is a logarithmic scale of the probability of the region being coding. Note that the three graphs represent the analysis in each of the three possible reading frames.

An equivalent analysis, for the opposite strand, is presented below.

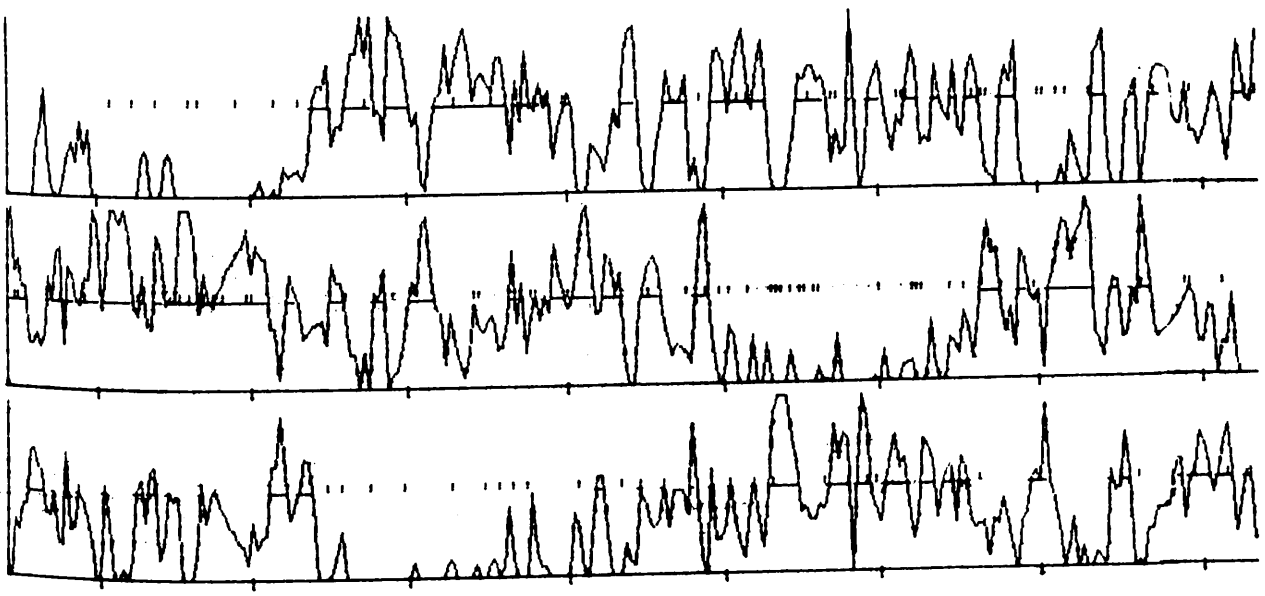
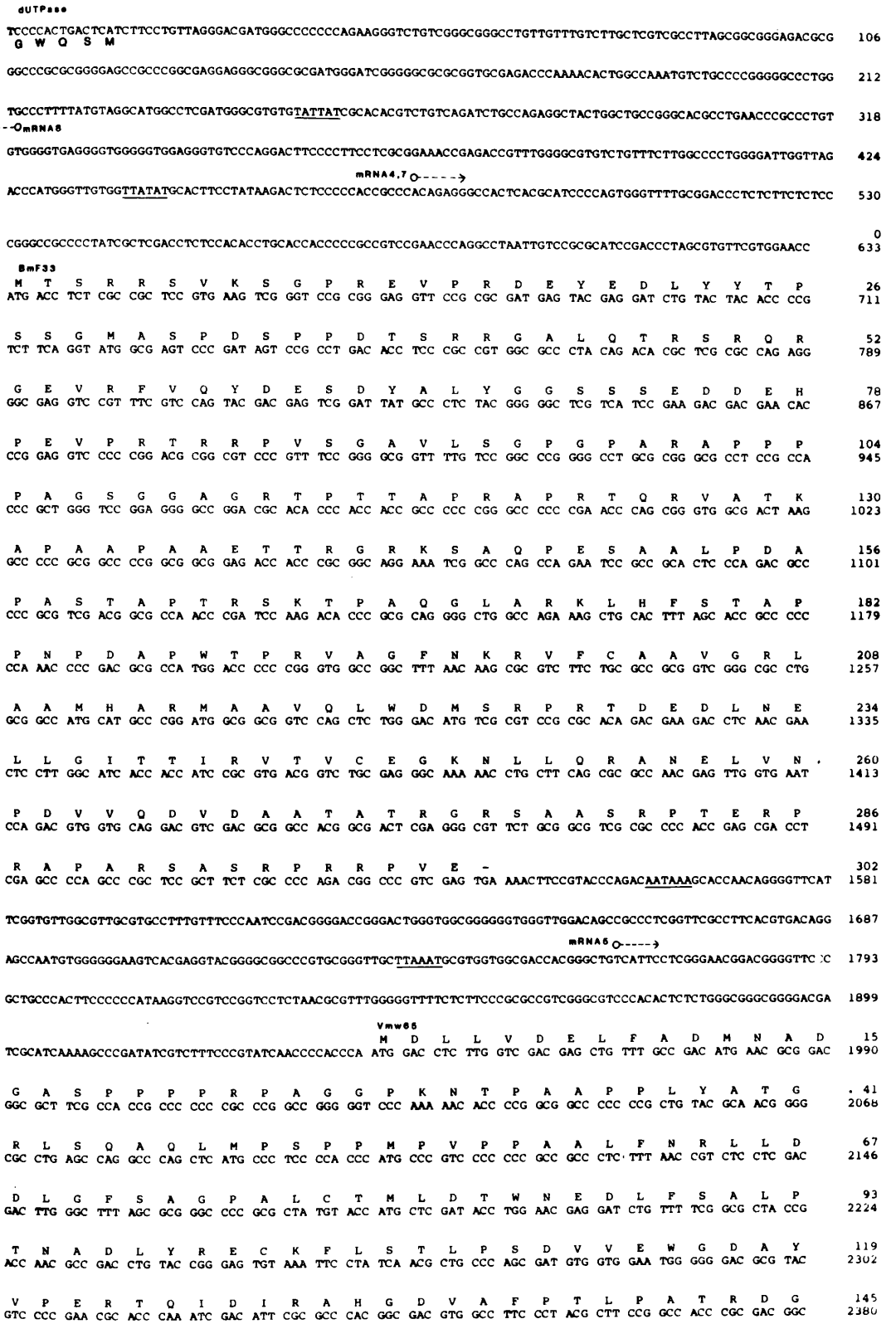


Figure 14: The sequence of BamHI F

Presented here is 8055 base pairs of sequence. All predicted ORFs are shown with their translation above the DNA sequence, except the N-terminus of the dUTPase which is shown below to indicate that it is translated from the opposite strand. The initiation sites of the transcripts mapped by Hall et al (1982) are shown as O-----, except where the degree of precision merits a less specific marker in which case an open box of variable length is shown. The terminus of the mRNA for Vmw65 is marked as -----X. The precise 5' and 3' ends of the mRNA for Vmw65 were located by S1 mapping (see Figure 15). Sequences for potential TATA boxes are underlined, as are probable polyadenylation sites.


dUTPase
mRNA8
mRNA4,7
BmF33
Vmw65
mRNA6

```
 L  G  L  Y  Y  E  A  L  S  R  F  F  H  A  E  L  R  A  R  E  E  S  Y  R  T  V      171
CTC GGG CTC TAC TAC GAA GCG CTC TCT CGT TTC TTC CAC GCC GAG CTA CGG GCG CGG GAG GAG AGC TAT CGA ACC GTG    2458

 L  A  N  F  C  S  A  L  Y  R  Y  L  R  A  S  V  R  Q  L  H  R  Q  A  H  M  R      197
TTG GCC AAC TTC TGC TCG GCC CTG TAC CGG TAC CTG CGC GCC AGC GTC CGG CAG CTG CAC CGC CAG GCG CAC ATG CGC    2536

 G  R  D  R  D  L  G  E  M  L  R  A  T  I  A  D  R  Y  Y  R  E  T  A  R  L  A      223
GGA CGC GAT CGC GAC CTG GGA GAA ATG CTG CGC GCC ACG ATC GCG GAC AGG TAC TAC CGA GAG ACC GCT CGT CTG GCG    2614

 R  V  L  F  L  H  L  Y  L  F  L  T  R  E  I  L  W  A  A  Y  A  E  Q  M  M  R      249
CGT GTT TTG TTT TTG CAT TTG TAT CTA TTT TTG ACC CGC GAG ATC CTA TGG GCC GCG TAC GCC GAG CAG ATG ATG CGG    2692

 P  D  L  F  D  C  L  C  C  D  L  E  S  W  R  Q  L  A  G  L  F  Q  P  F  M  F      275
CCC GAC CTG TTT GAC TGC CTC TGT TGC GAC CTG GAG AGC TGG CGT CAG TTG GCG GGT CTG TTC CAG CCC TTC ATG TTC    2770

 V  N  G  A  L  T  V  R  G  V  P  I  E  A  R  R  L  R  E  L  N  H  I  R  E  H      301
GTC AAC GGA GCG CTC ACC GTC CGG GGA GTG CCA ATC GAG GCC CGC CGG CTG CGG GAG CTA AAC CAC ATT CGC GAG CAC    2848

 L  N  L  P  L  V  R  S  A  A  T  E  E  P  G  A  P  L  T  T  P  P  T  L  H  G      327
CTT AAC CTC CCG CTG GTG CGC AGC GCG GCT ACG GAG GAG CCA GGG GCG CCG TTG ACG ACC CCT CCC ACC CTG CAT GGC    2926

 N  Q  A  R  A  S  G  Y  F  M  V  L  I  R  A  K  L  D  S  Y  S  S  F  T  T  S      353
AAC CAG GCC CGC GCC TCT GGG TAC TTT ATG GTG TTG ATT CGG GCG AAG TTG GAC TCG TAT TCC AGC TTC ACG ACC TCG    3004

 P  S  E  A  V  M  R  E  H  A  Y  S  R  A  R  T  K  N  N  Y  G  S  T  I  E  G      379
CCC TCC GAG GCG GTC ATG CGG GAA CAC GCG TAC AGC CGC GCG CGT ACG AAA AAC AAT TAC GGG TCT ACC ATC GAG GGC    3082

 L  L  D  L  P  D  D  D  A  P  E  E  A  G  L  A  A  P  R  L  S  F  L  P  A  G      405
CTG CTC GAT CTC CCG GAC GAC GAC GCC CCC GAA GAG GCG GGG CTG GCG GCT CCG CGC CTG TCC TTT CTC CCC GCG GGA    3160

 H  T  R  R  L  S  T  A  P  P  T  D  V  S  L  G  D  E  L  H  L  D  G  E  D  V      431
CAC ACG CGC AGA CTG TCG ACG GCC CCC CCG ACC GAT GTC AGC CTG GGG GAC GAG CTC CAC TTA GAC GGC GAG GAC GTG    3238

 A  M  A  H  A  D  A  L  D  D  F  D  L  D  M  L  G  D  G  D  S  P  G  P  G  F      457
GCG ATG GCG CAT GCC GAC GCG CTA GAC GAT TTC GAT CTG GAC ATG TTG GGG GAC GGG GAT TCC CCG GGG CCG GGA TTT    3316

 T  P  H  D  S  A  P  Y  G  A  L  D  M  A  D  F  E  F  E  Q  M  F  T  D  A  L      483
ACC CCC CAC GAC TCC GCC CCC TAC GGC GCT CTG GAT ATG GCC GAC TTC GAG TTT GAG CAG ATG TTT ACC GAT GCC CTT    3394

 G  I  D  E  Y  G  G  -                                                          491
GGA ATT GAC GAG TAC GGT GGG TAG GGGGCGCGACCGGACCCGCATCCCCCGTCTGGGTTTTCCCCTCCCGTCACCGGTTCGTATCCACAATAAACACG 3492

                -----X
AGCACATACATTACAAAACCTGCGGTTGTCGTCTGATTATTTGGTGGTGGGGGAAAGAACTAGCCAGGAGACGGGACCGCGCAACCAACCCACTGGGGTCTGGGTT 3598

GCCGGCGTGTGTGTTAGCCGCGTCTGCGGGCCTGTCGTGTAGATTCGAAACCACGGACGGGTGATTGTGTCGCAGGGCGGCCCGCGTATAAAGGCGAGAGCGCGGG 3704

      mRNA1
ACCGTTTCCGCATTTGGCCGGGGGCTGGGGCGGCGGGTAGCCTTCGCGGGAGATACTGCGTTTTTTTGCGCCGGCCCCGTCGCTCCCGTCCATTCCCATCGCGAGG 3810

                                                                          BmF73
                                                                          M  S      2
GGGTCCGGCGGCACCTACCCCGGCCTCCATCCGCGCTGTGGGGTCTTTTTCTTTTTTGGGGGGTAGCGGACATCCGATAACCCGCGTCTATCGCCACC ATG TCG 3914

 A  R  E  P  A  G  R  R  R  R  A  S  T  R  P  R  A  S  P  V  A  D  E  P  A  G       28
GCT CGC GAA CCC GCG GGG CGC AGG AGG CGC GCA TCC ACC CGC CCC CGC GCC TCG CCC GTG GCG GAC GAG CCA GCG GGC    3992

 D  G  V  G  F  M  G  Y  L  R  A  V  F  R  G  D  D  D  S  E  L  E  A  L  E  E       54
GAT GGG GTG GGG TTC ATG GGG TAC CTG CGT GCG GTG TTC CGC GGG GAT GAC GAC AGC GAG CTA GAG GCT CTG GAG GAG    4070

 M  A  G  D  E  P  P  V  R  R  R  R  E  G  P  R  A  R  R  R  R  A  S  E  A  P       80
ATG GCG GGC GAC GAG CCG CCC GTG CGC CGT CGA CGG GAA GGC CCG CGA GCC CGA CGA CGG CGC GCG TCG GAA GCC CCG    4148

 P  T  S  H  R  R  A  S  R  Q  R  P  G  P  D  A  A  R  S  Q  S  V  R  G  R  L      106
CCC ACA TCC CAT CGC CGA GCG TCC CGG CAG CGC CCC GGG CCC GAT GCG GCC CGT AGC CAG TCG GTG CGC GGT CGC CTG    4226

 D  D  D  D  E  V  P  R  G  P  P  Q  A  R  Q  G  G  Y  L  G  P  V  D  A  R  A      132
GAC GAT GAT GAT GAG GTT CCC CGC GGT CCT CCG CAG GCC CGG CAG GGG GGA TAC CTG GGT CCC GTC GAC GCG CGG GCT    4304

 I  L  G  R  V  G  G  S  R  V  A  P  S  P  L  F  L  E  E  L  Q  Y  E  E  D  D      158
ATT TTG GGG CGG GTC GGC GGT TCG CGG GTG GCG CCG TCG CCG CTG TTC CTA GAG GAG CTG CAG TAC GAG GAG GAC GAC    4382

 Y  P  E  A  V  G  P  E  D  G  G  G  A  R  S  P  P  K  V  E  V  L  E  G  R  V      184
TAC CCG GAA GCC GTC GGG CCG GAG GAC GGC GGC GGG GCC CGT TCC CCG CCC AAG GTG GAG GTT CTG GAG GGA CGC GTG    4460

 P  G  P  E  L  R  A  A  F  P  L  D  R  L  A  P  Q  V  A  V  W  D  E  S  V  R      210
CCG GGC CCG GAG CTC CGG GCG GCA TTC CCG TTG GAT CGA CTG GCC CCT CAG GTT GCC GTG TGG GAC GAG TCC GTG CGC    4538

 S  A  L  A  L  G  H  P  A  G  F  Y  P  C  P  D  S  A  F  G  L  S  R  V  G  V      236
TCC GCC CTA GCC CTG GGG CAT CCG GCC GGG TTT TAC CCG TGT CCG GAT AGC GCG TTC GGG TTA TCG CGC GTG GGG GTC    4616
```

```
 M  H  F  A  S  P  D  N  P  A  V  F  F  R  Q  T  L  Q  Q  G  E  A  L  A  W  Y      262
ATG CAC TTC GCC TCC CCC GAC AAC CCC GCG GTG TTT TTC CGC CAA ACC CTG CAG CAG GGC GAG GCG TTG GCC TGG TAT     4694

 I  T  G  D  G  I  L  D  L  T  D  R  R  T  K  T  S  P  A  Q  A  M  S  F  L  A      288
ATC ACG GGC GAT GGG ATT CTT GAC CTG ACG GAT CGT CGA ACA AAA ACC AGC CCC GCC CAG GCG ATG AGC TTT TTG GCG     4772

 D  A  V  V  R  L  A  I  N  G  W  V  C  G  T  R  L  H  A  E  A  R  G  S  D  L      314
GAT GCC GTC GTG CGG CTG GCC ATC AAC GGG TGG GTG TGC GGG ACG CGC CTT CAC GCG GAG GCG CGC GGG TCT GAC CTG     4850

 D  D  R  A  A  E  L  R  R  Q  F  A  S  L  T  A  L  R  P  V  G  A  A  A  V  P      340
GAC GAC AGG GCG GCC GAG CTG AGG CGG CAG TTC GCG AGC CTC ACG GCG TTG CGT CCC GTC GGG GCC GCG GCC GTG CCG     4928

 L  L  S  A  G  G  L  V  S  P  Q  S  G  P  D  A  A  V  F  R  S  S  L  G  S  L      366
TTA CTG AGC GCG GGG GGG TTA GTG TCC CCC CAA TCC GGC CCC GAC GCC GCG GTG TTC CGC AGC TCG CTG GGG TCC CTG     5006

 L  Y  W  P  G  V  R  A  L  L  D  R  D  C  R  V  A  A  R  Y  A  G  R  M  T  Y      392
CTG TAC TGG CCC GGG GTG CGC GCG CTG CTG GAC CGC GAC TGT CGC GTG GCC GCC CGC TAT GCC GGC CGC ATG ACG TAC     5084

 L  A  T  G  A  L  L  A  R  F  N  P  D  A  V  R  C  V  L  T  R  E  A  A  F  L      418
CTG GCC ACC GGG GCC CTG CTC GCC CGC TTC AAT CCC GAC GCC GTC CGG TGC GTT TTG ACG CGG GAG GCC GCC TTC CTG     5162

 G  R  V  L  D  V  L  A  V  M  A  E  Q  T  V  Q  W  L  S  V  V  V  G  A  R  L      444
GGG CGC GTG CTG GAT GTG CTG GCG GTG ATG GCG GAG CAG ACG GTC CAG TGG CTC TCG GTG GTC GTG GGG GCG CGC CTG     5240

 H  P  H  V  H  H  P  A  F  A  D  V  A  R  E  E  L  F  R  A  L  P  L  G  S  P      470
CAC CCG CAC GTG CAC CAC CCC GCC TTT GCG GAC GTG GCG CGG GAG GAG CTG TTT CGC GCC CTG CCC CTG GGA AGC CCC     5318

 A  V  V  G  A  E  H  E  A  L  G  D  T  A  A  R  R  L  L  A  N  S  G  L  N  A      496
GCG GTC GTG GGG GCC GAG CAC GAG GCG CTG GGC GAC ACC GCA GCG CGC CGG CTG CTC GCC AAC AGC GGG CTC AAC GCC     5396

 V  L  G  A  A  V  Y  A  L  H  T  A  L  A  T  V  T  L  K  Y  A  R  A  C  G  D      522
GTG CTG GGC GCT GCG GTG TAC GCG CTG CAC ACG GCC CTG GCG ACC GTG ACC TTA AAG TAC GCC CGG GCG TGC GGG GAC     5474

 A  H  R  R  R  D  D  A  A  A  T  R  A  I  L  A  A  G  L  V  L  Q  R  L  L  G      548
GCG CAC CGG CGC CGG GAC GAC GCG GCG GCC ACG CGC GCC ATT CTG GCC GCC GGG CTC GTC CTG CAG CGG CTG CTG GGC     5552

 F  A  D  T  V  V  A  C  V  T  L  A  A  F  D  G  G  F  T  A  P  E  V  G  T  Y      574
TTT GCC GAC ACC GTG GTG GCG TGC GTG ACA CTG GCC GCG TTT GAC GGG GGA TTC ACG GCC CCC GAG GTG GGC ACG TAC     5630

 T  P  L  R  Y  A  C  V  L  R  A  T  Q  P  L  Y  A  R  T  T  P  A  K  F  W  A      600
ACC CCC CTG CGC TAC GCG TGC GTC CTC CGA GCG ACC CAG CCC CTG TAC GCG CGC ACC ACC CCC GCC AAG TTC TGG GCG     5708

 D  V  R  A  A  A  E  H  V  D  L  R  P  A  S  S  A  P  R  A  P  V  S  G  T  A      626
GAC GTC CGC GCG GCC GCG GAG CAC GTG GAT CTG CGC CCC GCC TCC TCA GCG CCC CGG GCC CCC GTG TCC GGG ACG GCA     5786

 D  P  A  F  L  L  K  D  L  E  P  F  P  P  A  P  V  S  G  G  S  V  L  G  P  R      652
GAC CCC GCC TTT CTT CTC AAG GAC CTG GAG CCC TTC CCC CCC GCC CCC GTA AGC GGC GGG TCC GTG TTG GGC CCG CGG     5864

 V  R  V  V  D  I  M  S  Q  F  R  K  L  L  M  G  D  E  G  A  A  A  L  R  A  H      678
GTC CGC GTG GTG GAC ATC ATG TCC CAG TTT AGG AAA CTG CTG ATG GGA GAC GAG GGG GCC GCC GCC CTG CGG GCG CAC     5942

                                                  ?----------------mRNA2----------------?
 V  S  G  R  R  A  T  G  L  G  G  P  P  R  P  -                                      694
GTG TCG GGG AGG CGC GCG ACC GGG CTG GGA GGC CCG CCA CGC CCA TAA GCTCCTCCCGATAAAAAGCGCCCCGATGGCCCTGGACGCGGC     6032

                                                BmF72'
                                          M  Q  R  R  T  R  G  A  S  S  L  R  L  A  R  C      16
ATAACTCCGACCGGCGGGTCCCGACCGAACGGGCGTCACC ATG CAG CGC CGG ACG CGC GGC GCG AGC TCC CTG CGG CTG GCG CGG TGC     6120

 L  T  P  A  N  L  I  R  G  D  N  A  G  V  P  E  R  R  I  F  G  G  C  L  L  P      42
CTG ACG CCT GCC AAC CTG ATC CGC GGC GAC AAC GCG GGC GTT CCC GAG CGG CGC ATC TTC GGC GGG TGT CTG CTC CCC     6198

 T  P  E  G  L  L  S  A  A  V  G  A  L  R  Q  R  S  D  D  A  Q  P  A  F  L  T      68
ACC CCG GAG GGG CTC CTT AGC GCG GCC GTG GGC GCC TTG CGG CAG CGC TCC GAC GAC GCG CAG CCG GCG TTT CTG ACC     6276

 C  T  D  R  S  V  R  L  A  A  R  Q  H  N  T  V  P  E  S  L  I  V  D  G  L  A      94
TGC ACC GAT CGC AGC GTC CGG TTG GCC GCG CGG CAA CAC AAC ACG GTT CCC GAG AGT TTG ATC GTG GAC GGG CTC GCC     6354

 S  D  P  H  Y  E  Y  I  R  H  Y  A  S  A  A  T  Q  A  L  G  E  V  E  L  P  G      120
AGC GAC CCG CAC TAC GAG TAC ATC CGG CAC TAC GCT TCG GCC GCC ACC CAG GCG CTG GGC GAG GTG GAG CTG CCC GGC     6432

 G  Q  L  S  R  A  I  L  T  Q  Y  W  K  Y  L  Q  T  V  V  P  S  G  L  D  V  P      146
GGC CAG TTG AGC CGC GCC ATC CTC ACG CAG TAC TGG AAG TAC CTG CAG ACG GTG GTG CCC AGC GGC CTG GAC GTC CCC     6510

 E  D  P  V  G  D  C  D  P  S  L  H  V  L  L  R  P  T  L  A  P  K  L  A  G  A      172
GAG GAC CCC GTG GGC GAC TGC GAC CCC AGC CTT CAC GTG CTG CTG CGG CCC ACC CTG GCG CCA AAG CTC GCT GGC GCG     6588
```

```
 H  P  V  Q  E  R  G  R  G  G  K  Y  A  A  T  V  A  G  L  R  D  A  L  H  R  I      198
CAC CCC GTT CAA GAG CGG GGC CGT GGC GGC AAG TAC GCC GCC ACC GTG GCC GGC CTG CGC GAC GCC CTC CAT CGG ATT    6666

 Q  Q  Y  M  F  F  M  R  P  A  D  P  S  R  P  S  T  D  T  A  L  R  L  N  E  L      224
CAG CAG TAC ATG TTC TTT ATG CGC CCG GCG GAC CCA AGC CGC CCC AGC ACC GAC ACC GCA CTG CGG CTC AAC GAG CTC    6744

 L  A  Y  V  S  V  L  Y  R  W  A  S  W  M  L  W  T  T  D  K  H  V  C  H  R  L      250
CTG GCC TAC GTC TCC GTG CTG TAC CGG TGG GCG TCC TGG ATG CTG TGG ACG ACG GAC AAG CAC GTA TGT CAC CGG CTG    6822

 S  P  S  N  R  R  F  L  P  L  G  G  S  P  E  A  P  A  E  T  F  A  R  H  L  D      276
AGT CCC TCC AAT CGC CGG TTC CTC CCG CTC GGC GGC AGC CCG GAG GCG CCC GCG GAA ACG TTC GCG CGC CAT CTG GAC    6900

 R  G  P  S  G  T  T  G  S  M  Q  C  M  A  L  R  A  A  V  S  D  V  L  G  H  L      302
CGG GGT CCC AGC GGC ACG ACC GGC TCC ATG CAG TGC ATG GCG CTC AGG GCG GCG GTC AGC GAC GTC CTG GGC CAC CTG    6978

 T  R  L  A  N  L  W  Q  T  G  K  R  S  G  G  T  Y  G  T  V  D  T  V  V  S  T      328
ACG CGC CTA GCC AAC CTG TGG CAG ACC GGC AAG CGG AGC GGC GGT ACG TAC GGG ACC GTG GAT ACC GTC GTA TCC ACG    7056

 V  E  V  L  S  I  V  H  H  H  A  Q  Y  I  I  N  A  T  L  T  G  Y  G  V  W  A      354
GTG GAG GTT CTG TCG ATC GTC CAT CAC CAC GCC CAG TAT ATC ATC AAC GCG ACC CTC ACC GGG TAC GGC GTC TGG GCC    7134

 T  D  S  L  N  N  E  Y  L  R  A  A  V  D  S  Q  E  R  F  C  R  T  T  A  P  L      380
ACC GAC AGC CTG AAC AAC GAG TAT CTG CGG GCC GCG GTG GAC AGC CAG GAG CGT TTC TGT CGG ACC ACC GCC CCC CTG    7212

 F  P  T  M  T  A  P  S  W  A  R  M  E  L  S  I  K  A  W  F  G  A  A  L  A  A      406
TTC CCC ACG ATG ACC GCC CCC AGC TGG GCC CGG ATG GAG TTG AGC ATC AAG GCT TGG TTC GGG GCC GCC CTG GCC GCG    7290

 D  L  L  R  N  G  A  P  S  L  H  Y  E  S  I  L  R  L  V  A  S  R  R  T  T  W      432
GAT CTG CTC CGC AAC GGG GCG CCG TCG CTC CAC TAC GAG TCC ATC CTG CGG CTC GTG GCG TCT CGC CGG ACG ACG TGG    7368

 S  A  G  P  P  P  D  D  M  A  S  G  P  G  G  H  R  A  G  G  G  T  C  R  E  K      458
TCC GCG GGG CCT CCC CCG GAC GAC ATG GCC AGC GGC CCG GGG GGG CAT CGC GCG GGG GGT GGG ACC TGT CGG GAA AAG    7446

 I  Q  R  A  R  R  D  N  E  P  P  P  L  P  R  P  R  L  H  S  T  P  A  S  T  R      484
ATT CAG CGG GCG CGG CGC GAC AAC GAG CCC CCG CCC CTC CCC CGA CCT CGC CTA CAC TCG ACC CCC GCG TCC ACC CGG    7524

 R  F  R  R  R  R  A  D  G  A  G  P  P  L  P  D  A  N  D  P  V  A  E  P  P  A      510
AGG TTC CGG AGG CGC CGC GCG GAC GGC GCG GGG CCC CCG CTT CCG GAT GCG AAC GAC CCG GTC GCC GAG CCC CCC GCT    7602

 A  A  T  Q  P  A  T  Y  Y  T  H  M  G  E  V  P  P  R  L  P  A  R  N  V  A  G      536
GCG GCC ACA CAG CCG GCC ACG TAT TAC ACG CAC ATG GGG GAG GTG CCC CCG CGC CTC CCG GCC CGT AAC GTC GCG GGA    7680

 P  D  R  R  P  P  A  A  T  C  P  L  L  V  R  R  A  S  L  G  S  L  D  R  P  R      562
CCC GAC AGG CGA CCG CCG GCG GCG ACG TGC CCC CTC CTC GTC CGG CGC GCG TCT CTG GGG AGC CTC GAT CGG CCA CGG    7758

 V  W  G  P  A  P  E  G  E  P  D  Q  M  E  A  T  Y  L  T  A  D  D  D  D  D  D      588
GTG TGG GGA CCC GCC CCG GAG GGA GAA CCC GAC CAG ATG GAA GCC ACG TAT CTG ACG GCC GAC GAC GAC GAC GAC GAC    7836

 A  R  R  K  A  T  H  A  A  S  A  R  E  R  H  A  P  Y  E  D  D  E  S  I  Y  E      614
GCC CGC CGC AAA GCC ACC CAC GCC GCC TCG GCC CGC GAA CGG CAC GCC CCC TAC GAG GAC GAC GAG TCA ATA TAC GAG    7914

 T  V  S  E  D  G  G  R  V  Y  E  E  I  P  W  M  R  V  Y  E  N  V  C  V  N  T      640
ACG GTG AGC GAG GAC GGG GGG CGT GTC TAC GAG GAA ATA CCA TGG ATG CGG GTC TAC GAA AAC GTC TGC GTG AAC ACG    7992

 A  N  A  A  P  A  S  P  Y  I  E  A  E  N  P  L  Y  D  W  G  G                      661
GCG AAT GCA GCG CCG GCC TCC CCG TAC ATT GAG GCG GAA AAT CCC CTG TAC GAC TGG GGG GGA                      8055
```

fragment 3' end-labelled at an XcyI site (position 3301). The results are shown in Figure 15. The MluI probe gave an S1-resistant band estimated to be 67 bases in size, whereas an approximately 200 base band was obtained using the SalI probe. These data are consistent and place the mRNA start site at around position 1767. The 3' end-labelled probe produced an S1-resistant band estimated to be 205 bases in size, locating the end of the mRNA at position 3512.

In the immediate 5' and 3' flanking sequences of the mRNA Vmw65 are several motifs similar to sequences implicated in the control of transcription initiation and termination, the details of which are discussed in the introduction. These are as follows:

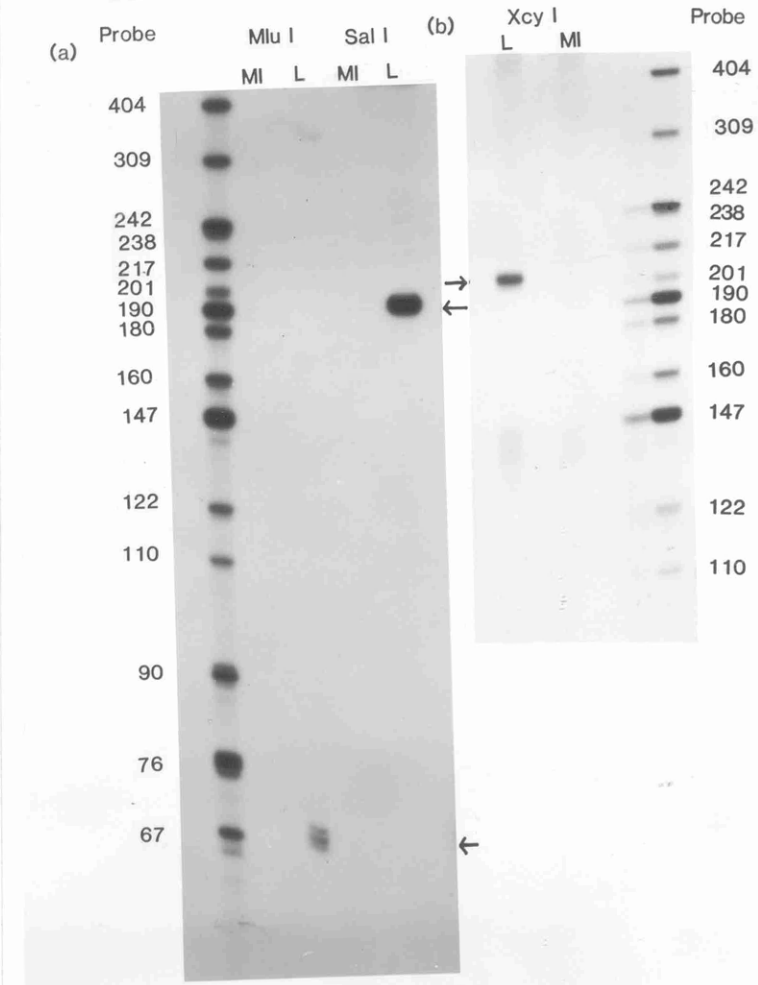
(a) TTAAAT, at position 1738, which is a reasonable candidate for the "TATA box", a sequence important for efficient and accurate initiation of transcription. It is in approximately the expected position, about -30 with respect to the cap site. Similar sequence elements have been found in many genes transcribed by RNA polymerase II.

(b) AGCCAATGT, at position 1688, an excellent match to a sequence, GGCCAATCT (the CAAT box), implicated in the transcriptional activity of globin promoters but also found in many others.

(c) GC-rich motifs similar to those found in both the SV40 21bp repeats and the HSV-1 TK gene. Specifically, CCGCCC at position 1658 and GGGCGG at position 1718, are in approximately similar places relative to the "TATA box" as those elements found in the TK promoter.

(d) On the 3' side of the coding region lies a typical polyadenylation signal sequence AATAAA, at position 3483, followed about 30bp downstream by a sequence TGCGCTTG. This is consistent with sequences (consensus YGTGTTYY, where Y=pyrimidine) found in this position in many genes and shown to be functionally important in the termination of HSV mRNA.

The first initiating ATG codon to be found after the mRNA start was at position 1946 and conforms satisfactorily with the rules of Kozak (1983; 1984). This codon initiates an open reading frame of 490 amino acids whose codon usage


(a)
Probe
Mlu I
Sal I
Ml
L
Ml
L
404
309
242
238
217
201
190
180
160
147
122
110
90
76
67
(b)
Xcy I
L
Ml
Probe
404
309
242
238
217
201
190
180
160
147
122
110

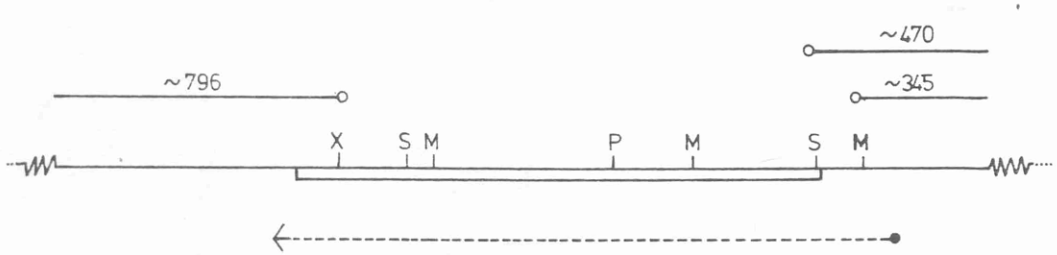

~470
~345
~796
X
S
M
P
M
S
M

Figure 15: S1 mapping of the 5' and 3' termini of the mRNA for Vmw65

Panel (a) shows the results of using a 5' end-labelled restriction fragment to map the mRNA initiation site. The probes used were an EcoRI- SalI fragment, labelled at the SalI site, and an MluI-EcoRI fragment, labelled at the MluI site. In both cases the EcoRI site is in the vector sequences. The MluI probe gives a hybrid band of about 67bp in size positioning the mRNA start at position 1767, consistent with the hybrid, of about 200bp, obtained with the SalI probe.

Panel (b) shows an equivalent experiment to map the 3' terminus of the mRNA. The probe used in this case was a 3' end-labelled HindIII-XcyI fragment. The HindIII site is in the vector. The hybrid band of about 205bp positions the end of the mRNA at 3512.

In both panels MI stands for mock-infected RNA and L is late, infected-cell, mRNA. The relative positions of the end-labelled probes used are shown below the gel.

Table 3: Predicted amino acid composition of Vmw65

RES.	NUM.	%	RES.	NUM.	%	RES.	NUM.	%	RES.	NUM.	%
Ala	58	11.8	Arg	40	8.2	Asn	12	2.4	Asp	41	8.4
Cys	6	1.2	Gln	10	2.0	Glu	29	5.9	Gly	32	6.5
His	13	2.7	Ile	9	1.8	Leu	65	13.3	Lys	4	0.8
Met	16	3.3	Phe	23	4.7	Pro	40	8.2	Ser	26	5.3
Thr	27	5.5	Trp	4	0.8	Tyr	18	3.7	Val	17	3.3

is consistent with that of other, known, HSV-1 proteins as determined by the program of Staden and McLachlan (1982).

5.2.1 The Vmw65 Protein

From the above predicted ORF, Vmw65 is composed of 490 amino acid residues and has a molecular weight of 54,342. The amino acid composition is presented in Table 3. The protein is highly charged, with the acidic residues, aspartic and glutamic acid, accounting for 14.3% of the total amino acids. Those amino acid residues with basic side chains, arginine and lysine, make up a further 9%. The predicted net acidic character is consistent with two-dimensional gel electrophoresis of HSV-1 proteins, which has shown Vmw65 to be amongst the most acidic proteins in infected cells (Haarr & Marsden, 1981).

Analysis of hydropathicity shows that there are no extensive regions of an extreme hydrophobic or hydrophilic nature (data not shown). However, a histogram (Figure 16) of the numbers of basic or acidic residues every 30 amino acids reveals that there is a distinct polarity along the peptide backbone, with acidic residues prevalent in the carboxy-terminal 80 amino acids and basic residues concentrated toward the amino-terminal region. The high leucine content may indicate that Vmw65 has a well defined secondary structure as this chemically inert amino acid is relatively hydrophobic and has a propensity to form pleated sheet.

5.2.2 Homology to Other proteins

A clear homologue to Vmw65 has been detected in the genome of VZV (Dalrymple et al, 1985). The VZV protein has a molecular weight of 46,522 (46K). Figure 17 shows a matrix homology plot between the two proteins. It can be seen that they are approximately colinear except that the HSV-1 protein has 80 amino acid residues, compared to VZV 46K, at the carboxy-terminus. Figure 18 is an optimal alignment of the two amino acid sequences, showing that approximately 35% of amino acid residues are conserved between the two

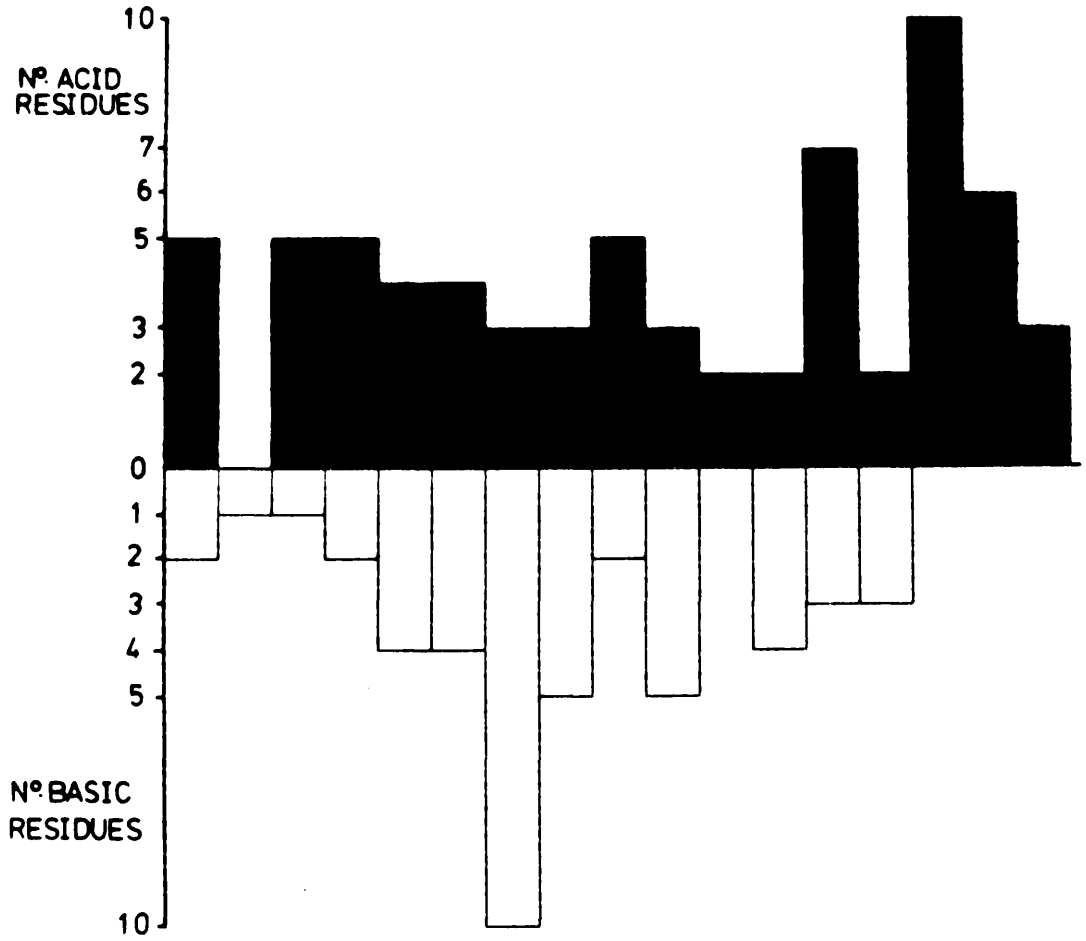

10
Nº ACID RESIDUES
7
6
5
3
2
0
1
2
3
4
5
Nº BASIC RESIDUES
10

Figure 16: Histogram of Acid/Basic Residue Distribution.

The distribution of acidic and basic residues along the polypeptide backbone of Vmw65 is plotted here. The interval is 30 amino acids. The data show that the distribution of each type of residue is not even.

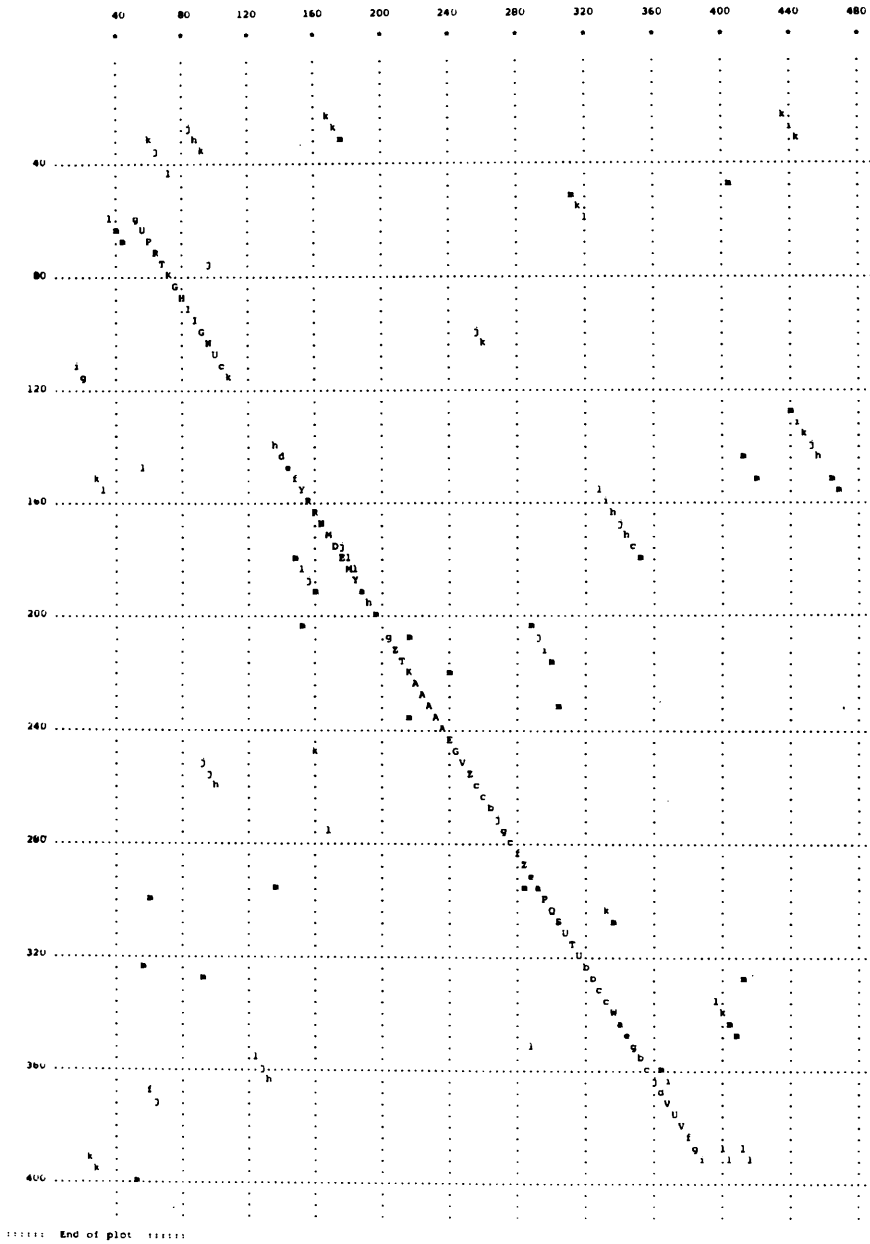

40
80
120
160
200
240
280
320
360
400
440
480
40
80
120
160
200
240
280
320
360
400
::::::: End of plot :::::::

Figure 17: A CINTHOM Plot of Homology between Vmw65 and VZV 46K.

This plot was obtained by use of the program described in section 4.9.2, using the mutation replacement matrix, a range of 12 and showing a minimum score of 25. It clearly shows colinearity between the two amino acid sequences except for the carboxy terminal 80 amino acid residues of Vmw65. HSV-1 Vmw65 lies along the X-axis and VZV 46k along the Y-axis.

First sequence: HSV1BAMF.54K

Second sequence: VZV.46K

```
   1 MD..LLVDELFADMNADGASPPPPRPA..GGPKNTPAAPPLYATGRLSQA.QLMPSPPMPVPPAALFNRLLDDL
     *   *       *              *           *       *   * *    ****    *  *   *  **
   1 MECNLGTEHPSTDTWNRSKTEQAVVDAFDESLFGDVASDIGFETSLYSHAVKTAPSPPWVASPKILYQQLIRDL

  70 GFSAGPALCTMLDTWNEDLFSALPTNADLYRECKFLSTLPSDV...VEWGDAYVPERTQIDIRAHGDVAFPTLP
      ** ** *   * ********  * * ***     **  * **   *   *    *       *  *  *  *  *
  75 DFSEGPRLLSCLETWNEDLFSCFPINEDLYSDMMVLSPDPDDVISTVSTKD..HVEMFNLTTR..GSVRLPSPP

 141 ATRDGLGLYYEALSRFFHAELRAREESYRTVLANFCSALYRYLRASVRQLHRQAHMRGRDRDLGEMLRATIADR
       **  *        *  ******* *   *   *    ***                *       **  *  *
 145 KQPTGLPAYVQEVQDSFTVELRAREEAYTKLLVTYCKSIIRYLQGTAKRTTIGLNIQNPDQKAYTQLRQSILLR

 215 YYRETARLARVLFLHLYLFLTREILWAAYAEQMMRPDLFDCLCCDLESWRQLAGLFQPFMFVNGALTVRGVPIE
     **** * *** * *****  ***   *  ** *   ** *  *       **    * *     *     * *
 219 YYREVASLARLLYLHLYLTVTREFSWRLYASQSAHPDVFAALKFTWTERRQFTCAFHPVLCNHGIVLLEGKPLT

 289 ARRLRELNHIREHLNLPLVRSAATEEPGAPLTTPPTLHGNQARASGYFMVLIRAKLDSYSSFTTSPSEAVMREH
     *  *** *  *  * *****    **   **   *       *  *     *  *** *          *   *
 293 ASALREINYRRRELGLPLVRCGLVEENKSPLVQQPSFSVHLPRSVGFLTHHIKRKLDAYAVKHPQEPRHVRADH

 363 AYSRARTKNNYGSTIEGLLDLPDDDAPEEAGLAAPRLSFLPAGHTRRLSTAPPTDVSLGDELHLDGEDVAMAHA
      *       **** *              *   ** *  *    *   *    ***   *
 367 PYAKVVENRNYGSSI.............EAMILAPPSPSEILPGDPPR....PPTCGFLTR

 437 DALDDFDLDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFTDALGIDEYGG
```

Figure 18: An Optimal Alignment of the Amino Acid Sequences of Vmw65 and VZV 46K.

The comparison was obtained by use of the optimal alignment program of Taylor (1984) - HOMOL - and the default parameters recommended by the author.

proteins. Since the VZV and HSV-1 genomes have diverged considerably, to the extent that their base compositions differ by 21 percentage points (A. Davison, personal communication), this homology is regarded as good confirmation of the correctness of the Vmw65 ORF assignment. It has not been possible to detect a homologue to Vmw65 in another herpes virus, Epstein Barr virus, whose genome has been completely sequenced. Similarly, a search of the National Biomedical Research Foundation protein database produced no matches of obvious significance.

5.3 Discussion: Vmw65

The results of Campbell et al (1984) clearly mapped the IE specific trans-activator gene to a portion of BamHI F, within which had been detected only one mRNA species, of about 1.7 kbp, encoding a protein of molecular weight 65,000 (Hall et al, 1982). Immune precipitations and hybrid-arrested translation experiments identified this protein as Vmw65, a major tegument species (Campbell et al, 1984). Linker insertions in the middle of the transcribed region, designed to disrupt potential coding sequences, destroyed the ability of the fragment to trans-activate IE promoters. It was, however, still formally possible that a minor mRNA lay across these sequences and encoded a second protein, smaller than Vmw65 and possibly lying in a different reading frame. Analysis of the sequence of BamHI F shows that this is unlikely as there are no other ORFs of typical HSV-1 composition in the area covered by the mRNA for Vmw65 (see Figures 12b and 13). The predicted 54K MW ORF, designated as the gene for Vwm65, shows a codon usage remarkably similar to both the HSV-1 TK gene coding sequences (data not shown) and to other ORFs in BamHI F (Figure 13). The S1 mapping data (Figure 15) is consistent both with the previously published map of Hall and co-workers and with the ORF prediction. The work presented here positions the 5' end of mRNA Vmw65 at 1766 and the 3' terminus at 3512, giving a total length for the transcript of 1,746 bases, remarkably close to the value determined

from Northern blots by Hall et al (1982). The evidence that Vmw65 is responsible for trans-activation of HSV-1 IE promoters is overwhelming.

The difference between the molecular weight determined by SDS-polyacrylamide gel electrophoresis and that from the predicted amino acid composition of Vmw65 is considerable. This is unlikely to be due to the reported phosphorylation of Vmw65 (Gibson & Roizman,1974; Marsden et al, 1978; McLean et al, 1982) since the in vitro translation product is of similar electrophoretic mobility (Hall et al, 1982; Campbell et al, 1984), although some in vitro translation systems can phosphorylate newly synthesised proteins (M. Frame, personal communication). It is possible that the size difference is due to the excess of negatively charged residues (resulting in less bound SDS and hence lower mobility) or to the high proline content which might result in the formation of local kinks in the peptide backbone and hence retard protein mobility.

The question of which potential initiation codon is utilised in the translation of mRNA Vmw65 has arisen with the publication of the nucleotide sequence of the analogous region in HSV-1 strain F by Pellet et al (1985). The first two ATG codons, after the 5' end of the mRNA, lie at 1945 and 1978 on the sequence of BamHI F (Figure 14). Pellet and co-workers favour M1978 as the initiating codon, presumably on the basis that an 8 bp linker insertion into a SalI site lying between the two ATGs does not affect trans-activation (Campbell et al, 1984). However, there are two lines of evidence which suggest that M1945 is at least as good a candidate and, indeed, that both AUG codons are utilised in vitro. Kozak (1983; 1984) has identified the sequence A/GNNAUGG as the optimal initiating environment for translation, although it is clear that other sequences are still functional. M1945 has the structure CCAATGG, a functional initiator in eleven cases out of 180 (Kozak, 1983) while M1978 has the structure GAGATGA*, functional in only 4 cases. The conflict with Pellet et al (1985) lies in

the fact that there is a transition in their sequence from A* to G resulting in an initiation context which is functional in 10% of the cases examined by Kozak. The statistical significance of 18/180 compared to 11/180 is doubtful and it is likely that translation initiates at M1945, thus the mRNA has a 5' untranslated leader of 179 bases and a 3' non-coding region of 97 bases. This conclusion is supported by two-dimensional gel electrophoresis of infected cell proteins (Haarr & Marsden, 1981; Marsden et al, 1983). The position of Vmw65 in the 2-D gel system has been identified using the products of in vitro translation (C. Preston, personal communication), however, just below and to the basic side, a second protein species is consistently visualised and correlates with a similar protein from infected cells. The small shift in size and charge required to arrive at this position, relative to Vmw65, is consistent with an alternative initiation at M1978, which removes about 1,000 molecular weight and four acidic residues. Polypeptides resulting from translation initiation at downstream, in phase, AUG codons have been previously described in the case of HSV-1 thymidine kinase (Preston & McGeoch, 1981; Marsden et al, 1983).

It is of some interest to compare the Vmw65 sequence reported for HSV-1 strain F (Pellet et al, 1985) with that of HSV-1 strain 17 (Dalrymple et al, 1985; this thesis). Aside from the different translation initiation points taken, there are only three changes at the amino acid level (Figure 19) and none of these fall in regions of particularly high homology with VZV 46K (Figures 18 and 19; see later). At the level of nucleic acid, there are four deletions and three insertions in strain F relative to strain 17, two of the deletions being located in the putative coding sequences of BmF33. Seventeen single base changes have occurred in 2517 bp of published sequence, of which 13 reside in the coding sequences of either Vmw65 or BmF33. Eleven of the changes fall in the Vmw65 coding region but only three result in amino acid changes. It is interesting to note that all three base changes which lead

First sequence: HSV-1 strain 17

Second sequence: HSV-1 strain F

```
1107 GTCGACGGCGCCAACCCGATCCAAGACACCCGCGCAGGGGCTGGCCAGAAAGCTGCACTTTAGCACCGCCCCCCCAAACCCCGACGCGCCATGGACCCCCCGGGTGGCCGGCT
     GTCGACG CGCCAACCCGATCCAAGACACCCGCGCAGGG CTGGCCAGAAAGCTGCACTTTAGCACCGCCCCCCCAAACCCCGACGCGCCATGGACCCCCCGGGTGGCCGGCT

     TTAACAAGCGCGTCTTCTGCGCCGCGGTCGGGCGCCTGGCGGCCATGCATGCCCGGATGGCGGCGGTCCAGCTCTGGGACATGTCGCGTCCGCGCACAGACGAAGACCTCAAC
     TTAACAAGCGCGTCTTCTGCGCCGCGGTCGGGCCGCTGGCGGCCATGCATGCCCGGATGGCGGCGGTCCAGCTCTGGGACATGTCGCGTCCGCGCACAGACGAAGACCTCAAC

     GAACTCCTTGGCATCACCACCATCCGCGTGACGGTCTGCGAGGGCAAAAACCTGCTTCAGCGCGCCAACGAGTTGGTGAATCCAGACGTGGTGCAGGACGTCGACGCGGCCAC
     GAACTCCTTGGCATCACCACCATCCGCGTGACGGTCTGCGAGGGCAAAAACCTGCTTCAGCGCGCCAACGAGTTGGTGAATCCAGACGTGGTGCAGGACGTCGACGCGGCCAC

     GGCGACTCGAGGGCGTTCTGCGGCGTCGCGCCCCACCGAGCGACCTCGAGCCCCAGCCCGCTCCGCTTCTCGCCCCAGACGGCCCGTCGAGTGAAAA CTTCCGTACCCAGAC
     GGCGACTCGAGGGCGTTCTGCGGCGTCGCGCCCCACCGAGCGACCTCGAGCCCCAGCCCGCTCCGCTTCTCGCCCCAGACGGCCCGTCGAGTGAAAAACTTCCGTACCCAGAC

     AATAAAGCACCAACAGGGGTTCATTCGGTGTTGGCGTTGCGTGCCTTTGTTTCCCAATCCGACGGGGACCGGGACTGGGTGGCGGGGGGTGGGTTGGACAGCCGCCCTCGGTT
     AATAAAGCACCAACAGGGGTTCATTCGGTGTTGGCGTTGCGTGCCTTTGTTTCCCAATCCGACGGGGACCGTGACTGGGTGGCGGGGGGTGGGTTGGACAGCCGCCCTCGGTT

     CGCCTTCACGTGACAGGAGCCAATGTGGGGGGAAGTCACGAGGTACGGGGCGGCCCGTGCGGGTTGCTTAAATGCGTGGTGGCGACCACGGGCTGTCATTCCTCGGGAACGGA
     CGCCTTCACGTGACAGGAGCCAATGTGGGGGGAAGTCACGAGGTACGGGGCGGCC GTGCGGGTTGCTTAAATGCGGGGTGGCGACCACGGGCTGTCATTCCTCGGGAACGGA

     CGGGGTTCCCGCTGCCCACTTCCCCCCATAAGGTCCGTCCGGTCCTCTAACGCGTTTGGGGGTTTT CTCTTCCCGCGCCGTCGGGCGTCCCACACTCTCTGGGCGGGCGGGG
     CGGGGTTCCCGCTGCCCACTTCCCCCCATAAGGTCCGTCCGGTCCTCTAACGCGTTTGGGGGTTTTTCTCTTCCCGCGCCGTCGG CGTCCCACACTCTCTGGGCGGGCGGGG

     ACGATCGCATCAAAAGCCCGATATCGTCTTTCCCGTATCAACCCCACCCAATGGACCTCTTGGTCGACGAGCTGTTTGCCGACATGAACGCGGACGGCGCTTCGCCACCGCCC
     ACGATCGCATCAAAAGCCCGATATCGTCTTTCCCGTATCAACCCCACCCAATGGACCTCTTGGTCGACGAGCTGTTTGCCGACATGGACGCGGACGGCGCTTCGCCACCACCC
                                                  M  D  L  L  V  D  E  L  F  A  D  M  D  A  D  G  A  S  P  P  P

     CCCCGCCCGGCCGGGGGTCCCAAAAACACCCCGGCGGCCCCCCCGCTGTACGCAACGGGGCGCCTGAGCCAGGCCCAGCTCATGCCCTCCCCACCCATGCCCGTCCCCCCCGC
     CCCCGCCCGGCCGGGGGTCCCAAAAACACCCCGGCGGCCCCTCCGCTGTACGCAACGGGGCGCCTGAGCCAGGCCCAGCTCATGCCCTCCCCGCCCATGCCCGTCCCCCCCGC
      P  R  P  A  G  G  P  K  N  T  P  A  A  P  P  L  Y  A  T  G  R  L  S  Q  A  Q  L  M  P  S  P  P  M  P  V  P  P  A

     CGCCCTCTTTAACCGTCTCCTCGACGACTTGGGCTTTAGCGCGGGCCCCGCGCTATGTACCATGCTCGATACCTGGAACGAGGATCTGTTTTCGGCGCTACCGACCAACGCCG
     CGCCCTCTTTAACCGTCTCCTCGACGACTTGGGCTTTAGCGCGGGCCCCGCGCTATGTACCATGCTCGATACCTGGAACGAGGATTTGTTTTCGGCGTTACCGACCAACGCCG
       A  L  F  N  R  L  L  D  D  L  G  F  S  A  G  P  A  L  C  T  M  L  D  T  W  N  E  D  L  F  S  A  L  P  T  N  A

     ACCTGTACCGGGAGTGTAAATTCCTATCAACGCTGCCCAGCGATGTGGTGGAATGGGGGGACGCGTACGTCCCCGAACGCACCCAAATCGACATTCGCGCCCACGGCGACGTG
     ACCTGTACCGGGAGTGTAAATTCCTATCAACGCTGCCCAGCGATGTGGTGGAATGGGGGGACGCGTACGTCCCCGAACGCGCCCAAATCGACATTCGCGCCCACGGCGACGTG
     D  L  Y  R  E  C  K  F  L  S  T  L  P  S  D  V  V  E  W  G  D  A  Y  V  P  E  R  A  Q  I  D  I  R  A  H  G  D  V

     GCCTTCCCTACGCTTCCGGCCACCCGCGACGGCCTCGGGCTCTACTACGAAGCGCTCTCTCGTTTCTTCCACGCCGAGCTACGGGCGCGGGAGGAGAGCTATCGAACCGTGTT
     GCCTTCCCGACGCTTCCGGCCACCCGCGACGGCCTCGGGCTCTACTACGAAGCGCTCTCTCGTTTCTTCCACGCCGAGCTACGGGCGCGGGAGGAGAGCTATCGAACCGTGTT
      A  F  P  T  L  P  A  T  R  D  G  L  G  L  Y  Y  E  A  L  S  R  F  F  H  A  E  L  R  A  R  E  E  S  Y  R  T  V  L

     GGCCAACTTCTGCTCGGCCCTGTACCGGTACCTGCGCGCCAGCGTCCGGCAGCTGCACCGCCAGGCGCACATGCGCGGACGCGATCGCGACCTGGGAGAAATGCTGCGCGCCA
     GGCCAACTTCTGCTCGGCCCTGTACCGGTACCTGCGCGCCAGCGTCCGGCAGCTGCACCGCCAGGCGCACATGCGCGGACGCGATCGCGACCTGGGAGAAATGCTGCGCGCCA
       A  N  F  C  S  A  L  Y  R  Y  L  R  A  S  V  R  Q  L  H  R  Q  A  H  M  R  G  R  D  R  D  L  G  E  M  L  R  A

     CGATCGCGGACAGGTACTACCGAGAGACCGCTCGTCTGGCGCGTGTTTTGTTTTTGCATTTGTATCTATTTTTGACCCGCGAGATCCTATGGGCCGCGTACGCCGAGCAGATG
     CGATCGCGGACAGGTACTACCGAGAGACCGCTCGTCTGGCGCGTGTTCTGTTTTTGCATTTGTATCTATTTTTGACCCGCGAGATCCTATGGGCCGCGTACGCCGAGCAGATG
      T  I  A  D  R  Y  Y  R  E  T  A  R  L  A  R  V  L  F  L  H  L  Y  L  F  L  T  R  E  I  L  W  A  A  Y  A  E  Q  M

     ATGCGGCCCGACCTGTTTGACTGCCTCTGTTGCGACCTGGAGAGCTGGCGTCAGTTGGCGGGTCTGTTCCAGCCCTTCATGTTCGTCAACGGAGCGCTCACCGTCCGGGGAGT
     ATGCGGCCCGACCTGTTTGACTGCCTCTGTTGCGACCTGGAGAGCTGGCGTCAGTTGGCGGGTCTGTTCCAGCCCTTCATGTTCGTCAACGGAGCGCTCACCGTCCGGGGAGT
      M  R  P  D  L  F  D  C  L  C  C  D  L  E  S  W  R  Q  L  A  G  L  F  Q  P  F  M  F  V  N  G  A  L  T  V  R  G  V

     GCCAATCGAGGCCCGCCGGCTGCGGGAGCTAAACCACATTCGCGAGCACCTTAACCTCCCGCTGGTGCGCAGCGCGGCTACGGAGGAGCCAGGGGCGCCGTTGACGACCCCTC
     GCCAATCGAGGCCCGCCGGCTGCGGGAGCTAAACCACATTCGCGAGCACCTTAACCTCCCGCTGGTGCGCAGCGCGGCTACGGAGGAGCCACGGGCGCCGTTGACGACCCCTC
       P  I  E  A  R  R  L  R  E  L  N  H  I  R  E  H  L  N  L  P  L  V  R  S  A  A  T  E  E  P  R  A  P  L  T  T  P

     CCACCCTGCATGGCAACCAGGCCCGCGCCTCTGGGTACTTTATGGTGTTGATTCGGGCGAAGTTGGACTCGTATTCCAGCTTCACGACCTCGCCCTCCGAGGCGGTCATGCGG
     CCACCCTGCATGGCAACCAGGCCCGCGCCTCTGGGTACTTTATGGTGTTGATTCGGGCGAAGTTGGACTCGTATTCCAGCTTCACGACCTCGCCCTCCGAGGCGGTCATGCGG
     P  T  L  H  G  N  Q  A  R  A  S  G  Y  F  M  V  L  I  R  A  K  L  D  S  Y  S  S  F  T  T  S  P  S  E  A  V  M  R

     GAACACGCGTACAGCCGCGCGCGTACGAAAAACAATTACGGGTCTACCATCGAGGGCCTGCTCGATCTCCCGGACGACGACGCCCCCGAAGAGGCGGGGCTGGCGGCTCCGCG
     GAACACGCGTACAGCCGCGCGCGTACGAAAAACAATTACGGGTCTACCATCGAGGGCCTGCTCGATCTCCCGGACGACGACGCCCCCGAAGAGGCGGGGCTGGCGGCTCCGCG
      E  H  A  Y  S  R  A  R  T  K  N  N  Y  G  S  T  I  E  G  L  L  D  L  P  D  D  D  A  P  E  E  A  G  L  A  A  P  R

     CCTGTCCTTTCTCCCCGCGGGACACACGCGCAGACTGTCGACGGCCCCCCCGACCGATGTCAGCCTGGGGGACGAGCTCCACTTAGACGGCGAGGACGTGGCGATGGCGCATG
     CCTGTCCTTTCTCCCCGCGGGACACACGCGCAGACTGTCGACGGCCCCCCCGACCGATGTCAGCCTGGGGGACGAGCTCCACTTAGACGGCGAGGACGTGGCGATGGCGCATG
       L  S  F  L  P  A  G  H  T  R  R  L  S  T  A  P  P  T  D  V  S  L  G  D  E  L  H  L  D  G  E  D  V  A  M  A  H

     CCGACGCGCTAGACGATTTCGATCTGGACATGTTGGGGGGACGGGGATTCCCCGGGGCCGGGATTTACCCCCCACGACTCCGCCCCCTACGGCGCTCTGGATATGGCCGACTTC
     CCGACGCGCTAGACGATTTCGATCTGGACATGTTGGGGGGACGGGGATTCCCCGGGTCCGGGATTTACCCCCCACGACTCCGCCCCCTACGGCGCTCTGGATATGGCCGACTTC
     A  D  A  L  D  D  F  D  L  D  M  L  G  D  G  D  S  P  G  P  G  F  T  P  H  D  S  A  P  Y  G  A  L  D  M  A  D  F

     GAGTTTGAGCAGATGTTTACCGATGCCCTTGGAATTGACGAGTACGGTGGGTAGGGGGCGCGACCGGACCCGCATCCCCCGTCTGGGTTTTCCCCTCCCGTCACCGGTTCGTA
     GAGTTTGAGCAGATGTTTACCGATGCCCTTGGAATTGACGAGTACGGTGGGTAGGGGGCGCGACCGGACCCGCATCCCCCGTCTGGGTTTTCCCCTCCCGTCCCCGGTTCGTA
      E  F  E  Q  M  F  T  D  A  L  G  I  D  E  Y  G  G  -

     TCCACAATAAACACGAGCACATACATTACAAAA CCTGCGGTTGTCGTCTGATTATTTGGTGGTGGGGGAAAGAACTAGCCAGGAGACGGGACCGCGCAACCAACCCACTGGG
     TCCACAATAAACACGAGCACATACATTACAAAAACCTGCGGTTGTCGTCTGATTATTTGGTGGTGGGGAAAAGAACTAGCCAGGAGACGGGACCGCGCAACCAACCCACTGGG

     GTCTGGGTTGCCGGCGTGTGTGTTAGCCGCGTCTGCGG 3627

     GTCTGGGTTGCCGGCGTGTGTGTTAGCCGCGTCTGCGG
```

Figure 19: Comparison of the Vmw65 Regions of HSV-1 strains 17 and F

The sequence published by Pellet et al (1985) for HSV-1 strain F is compared to the analogous region of strain 17. Insertion/deletion events are noted by the use of ^ or v. Mismatches are marked by !. The translation of Vmw65 is shown and the occurence of mutations which alter the encoded amino acid are marked.

to amino acid changes have occurred in the first position, whilst only t out of eight conservative base substitutions are in the first position, the remainder occurring in the third position.

The homologue to Vmw65 detected in the VZV genome, VZV 46K, has an almost co-linear amino acid sequence but is shorter by 80 amino acid residues at the carboxy terminus. It will be of interest to determine whether the VZV homologue and Vmw65 can stimulate expression from their heterologous IE genes. A preliminary experiment, in which a clone containing the entire VZV 46K gene and flanking sequences, was assayed for the ability to stimulate expression from pTKN2 (see Figure 27) in BHK cells by a standard co-transfection procedure, did not show stimulation of expression from the HSV-1 promoter (data not shown). This experiment was far from definitive as it was not determined whether the VZV gene was expressed, an important point since it is known that VZV cannot grow in BHK cells. In order to determine definitively whether the VZV 46K gene product can, or cannot, stimulate HSV-1 IE promoters, it will be necessary to replace exactly the coding sequences of Vmw65 in BamHI F with the coding sequences of VZV 46K and assay the hybrid gene in co-transfection systems. One piece of evidence which suggests that the VZV homologue may not be able to stimulate expression of IE genes is based on the fact that the VZV gene homologous to HSV-1 IE3 does not exhibit sequences corresponding to the TAATGARAT consensus in its 5' flanking regions (Davison & Scott, 1984), and therefore, either VZV 46K recognises a different upstream sequence or it has no transcriptional stimulatory function. In this respect it is interesting to speculate that the carboxy terminal 80 amino acids of Vmw65 may be involved in the process of transcription stimulation. The distribution

of acidic/basic residues (Figure 16) indicates a marked polarity to the protein molecule in which basic residues predominate near the amino terminus and acidic side chains at the carboxy terminus. It is unlikely that this acidic "tail" interacts with highly negatively charged DNA, but it may allow Vmw65 to interact with cellular DNA binding proteins. An experiment in which an 8 bp linker was inserted into the coding sequences at the SalI site at position 3179 (thus placing the final 80 amino acids of Vmw65 out of phase), resulted in the loss of stimulatory activity (C. Preston, personal communication). Possible mechanisms for the action of Vmw65 will be discussed later.

There exists a temperature sensitive mutant of HSV-2 strain HG52, called ts13, which has an altered 65K polypeptide (H. Rixon, personal communication). This lesion was mapped to 0.64-0.70 map units and is believed to lie in the HSV-2 equivalent of Vmw65 (Moss et al, 1979). The BglII I fragment of HSV-2 containing the region of interest was cloned, from both wild-type and mutant viral DNA, and tested for the ability to stimulate HSV-1 IE transcription in a short-term transfection assay. Both mutant and wild-type clones gave efficient stimulation of expression from an HSV-1 IE3 promoter and were indistinguishable at both 31°C and 38.5°C (data not shown). The same preparation of mutant plasmid DNA was subsequently shown to be able to transfer the ts13 lesion into a wild-type background, demonstrating that the cloning process had not resulted in reversion of the mutation (V.Preston, personal communication). It has recently been shown that the plasmid pMC1, containing the HSV-1 Vmw65K gene, is capable of rescuing the ts13 mutation (V. Preston, personal communication) strongly suggesting that the ts13 lesion resides in the HSV-2 equivalent of Vmw65. The implications of these findings are that it is possible to uncouple the transcription activator function of Vmw65 from its role as a structural component of the virion and that Vmw65 is an essential structural protein. It remains to be determined whether the activation function of the protein is also essential for virus viability.

There are few clues as to the mechanism by which Vmw65 is able to specifically activate IE genes, therefore the following section is speculative. Clearly, any mechanism must act to alter the specificity of RNA pol II such that IE gene promoters are at a competitive advantage compared to other promoters. Conceptually such an alteration might affect either the polymerase itself (including associated factors) or the template. Template modifications in eukaryotic gene expression have been described, for example DNA methylation (Razin & Riggs, 1980), but it is difficult to envisage how such a mechanism might operate in the case of Vmw65. For this reason only the alteration of polymerase specificity will be considered. Below are listed the most obvious candidates, together with known examples where one exists. Most of the detailed analyses of transcription modulation have been performed in prokaryotic systems but may provide possibilities to be investigated.

(i) Chemical modification of E.coli RNA polymerase, by the addition of an adenosine-5'-diphospho-5''-β-D-ribosyl residue, has been shown to occur during infection by the bactriophage T4 (Seifert et al, 1969; Rohrer et al, 1975). The enzyme involved is phage encoded and packaged into mature virus particles. The first polymerase modifications occur within 30 seconds of infection and are probably due to transfer of the enzyme from the phage into the cell. Alteration of the polymerase molecule affects its ability to bind the σ factor, which is essential for mRNA initiation. There is no evidence for an enzymatic activity, of any kind, associated with Vmw65.

(ii) Vmw65 might act by substituting for a host cell encoded polymerase subunit. The precident for such a mechanism again comes from T4. It is likely that phage genes code for proteins which become part of the E.coli polymerase complex (Stevens, 1970; 1972; 1974). In Bacillus subtilis there exist alternate "σ factors" for initiation at promoters active during sporulation, and this cellular system is exploited by the Bacillus phage SPO1 which encodes its own σ factors (Losick & Pero, 1981). In E.coli it is

likely that a novel factor is involved in the transcription of heat shock genes (Grossman et al, 1984). Presumably these substitute factors act to increase the affinity of RNA polymerase for a small sub-set of promoters, at the expense of others, a situation with some relevance to the expression of HSV-1 IE genes.

(iii) Vmw65 might act as a transcription factor in much the same way as Sp1 or HSTF (see section 1.2.3). Experiments designed to detect direct binding of Vmw65 to HSV-1 IE promoter sequences, in vitro, have failed to demonstrate such an activity (M.E.M Campbell, personal communication), however it is possible that an association of Vmw65 with a host cell transcription factor could alter the specificity of this factor without Vmw65 itself being able to bind DNA. Thus the binding specificity (for the TAATGARAT sequences) could reside in the viral protein and the DNA binding activity in the cell protein. This model would predict that Vmw65 (purified from virions), when mixed with a crude uninfected cell extract, might bind DNA in vitro. It is possible that a cellular modification to Vmw65 carried by the virus, for example phoshorylation or a specific proteolytic cleavage, could alter the affinity of the protein for DNA. Alternatively, Vmw65 might act to release a chromatin-bound transcription factor into the nucleoplasm where incoming HSV-1 DNA was at a competitive advantage compared to cellular genes. Many cell non-histone proteins (NHPs), such as the high mobility group proteins (HMGs), are particularly acidic. These proteins are believed to modulate the structure of chromatin in regions where genes are being actively transcribed (reviewed in Lewin, 1980). Due to the acidity of Vmw65, and the fact that it is present in substantial quantities in the virus particle, it is tempting to propose that it acts like an NHP either in the release of "TAATGARAT binding factors" from host chromatin or in catalysing the formation of stable transcription complexes on the promoters of IE genes.

In general, proteins have been assigned to the tegument by a process of elimination rather than by direct

experimental analysis, that is, tegument species are those structural proteins which cannot be found in either capsids or the envelope (Roizman & Furlong, 1974). Perhaps it is a measure of the mystery surrounding this structure that apparently elegant and important studies have remained unpublished and relegated to communication by review article (Roizman & Furlong, 1974; Spear, 1980). These authors report experiments, involving partial degradation of viral particles by detergent, which suggest that Vmw65 (VP16) is located towards the outside of the viral tegument, in contrast to other putative tegument proteins which could be extracted only under conditions which resulted in capsid solubilisation.

The function of the viral tegument is unknown. Vmw65 comprises a significant proportion of the total protein in the tegument and is likely to play an important structural role. In certain enveloped RNA viruses, such as influenza and vesicular stomatitis virus (VSV), there exists a protein species - the matrix or M protein - which aids the specific interaction of nucleocapsids and cell membrane (reviewed in Dubois-Dalq et al, 1984). The M protein probably has two binding sites, one for nucleocapsids and a second for some component of the membrane. Two possible mechanisms can be envisaged by which the M protein and envelope might bind, either by direct interaction via a large hydrophobic domain or, indirectly, by complexing with virus encoded envelope proteins and, indeed, both mechanisms are probably used. Influenza virus encodes an M protein with a large hydrophobic domain (Gregoriades, 1980) which is likely to interact with the lipid bilayer. Vesicular stomatitis virus has been shown to have a basic M protein which apparently binds electrostatically to the cytoplasmic membrane (Rose & Galliani, 1981) and associates with the virally encoded G protein, a membrane glycoprotein. Furthermore, the VSV M protein is a negative regulator of transcription in vitro (Carrol & Wagner, 1979). Gibson (1981) has suggested that the tegument of HCMV, and by implication that of all herpesviruses, has a function analogous to the M proteins of

enveloped RNA viruses in interfacing the nucleocapsid and envelope. There is no large hydrophobic domain in the amino acid sequence of Vmw65, but, there is evidence to suggest that this protein is complexed with glycoprotein B and possibly other viral glycoproteins which reside in the envelope (M. E. M. Campbell, personal communication) which might implicate it as a candidate for an "M-type" protein. The relatively extensive conservation of amino acid sequence between Vmw65 and VZV 46K is suggestive of a strong selective pressure and may indicate that the protein has more than one structural role. Other functions one might postulate for the tegument would be important in the early stages of the infectious cycle, e.g. the tegument may provide a protective layer around the nucleocapsid or it may be responsible for the nuclear tropism of the virion. It will be interesting to dissect the tegument in detail and identify the proteins and genes involved in its synthesis.

5.4 Properties of the Other Predicted Polypeptides

Aside from the gene for Vmw65, four distinct ORFs can be seen in the sequence of BamHI F (Figure 12b). Three of these ORFs are described in detail below and have been named BamF33, BamF73 and BamF72', the fourth, marked * in figure 12b, poses a problem. This ORF does not exhibit a strongly "herpetic" codon preference, with G/C residues in the third position, furthermore, Hall *et al* (1982) could not detect a mRNA which would correspond to a gene in this region. An ORF of similar size is found, in the identical position, in the genome of VZV, but exhibits no detectable homology to the HSV-1 ORF. The balance of evidence favours the conclusion that this ORF does not correspond to an actual protein coding gene and therefore no further reference to it will be made.

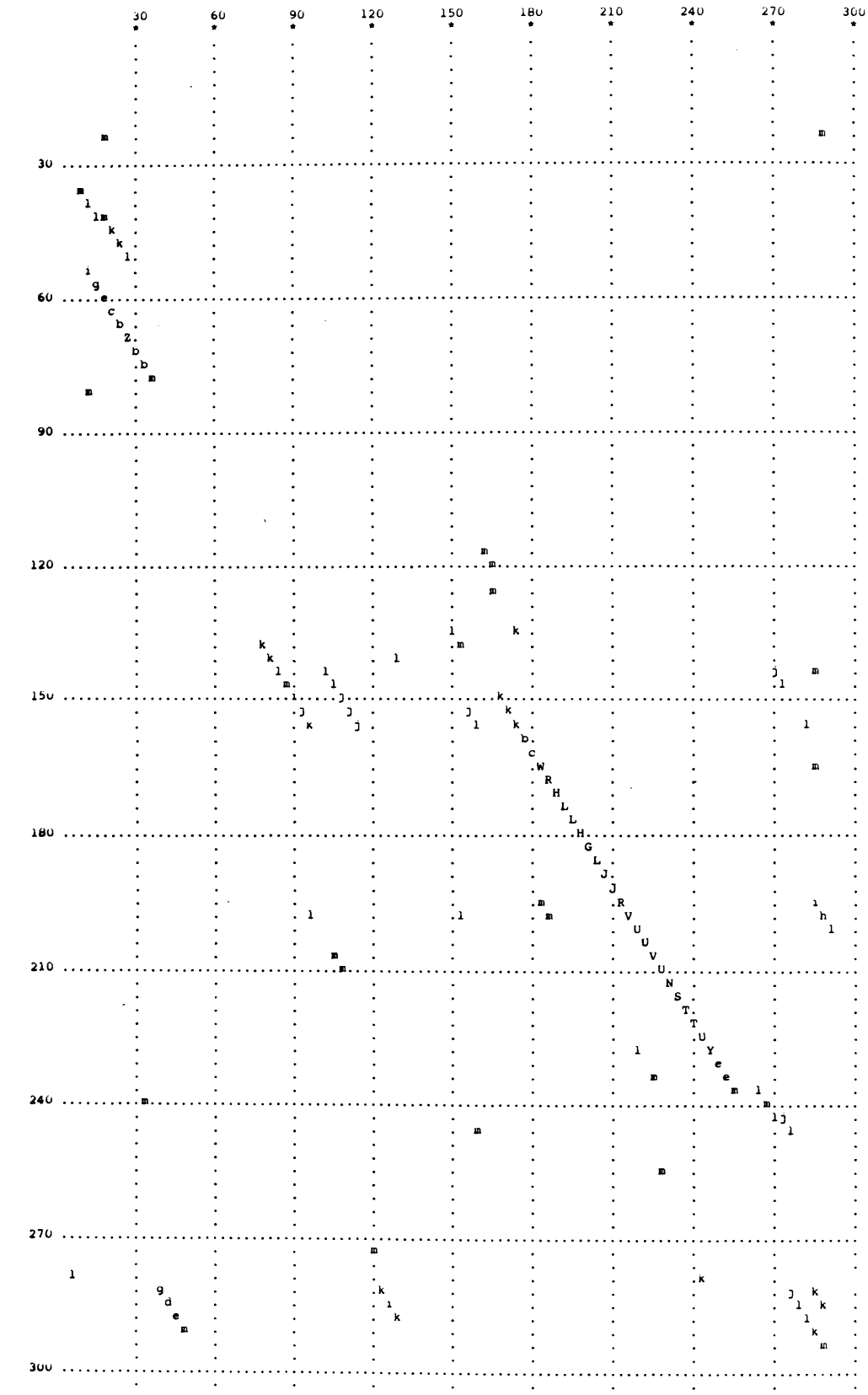
5.4.1 BmF33

The paper by Hall and co-workers (1982) places the 5' end of two mRNAs (4 and 7) at a position equivalent to 477 in the sequence of BamHI F reported here. At position 1557 there is a polyadenylation signal AATAAA, in a region

reported to contain the 3' ends of mRNA7 and mRNA6 (mRNA6 initiates upstream of BamHI F). Hall et al predicted, on the basis of in vitro translation data, that this small family of 5' and 3' co-terminal mRNAs encoded a single 42K molecular weight protein. Consistent with this view, there lies a single open reading frame, between position 634 and 1536, of 301 amino acid residues. The predicted molecular weight of this protein is therefore 32,252. The amino acid composition of BmF33 (Table 4(a)) shows that the protein is basic in nature with about 26% of its residues strongly charged. The most common amino acids are alanine (14.3%), arginine (13.0%) and proline (12.3%). BmF33 has a homologue in the genome of VZV. Figure 20 is a matrix plot of the aligned amino acid sequences, revealing a strong block of homology in a 100 amino acid stretch towards the carboxy end of each protein. An optimal alignment of the two amino acid sequences is presented in Figure 21 and the block of homology underlined.

5.4.2 BmF73

Hall et al (1982) described a large, 4.7kb, mRNA which initiates 260bp upstream of a SalI site at map unit 0.674. The equivalent site in the sequence of BamHI F is at position 4099 and would place an equivalent mRNA cap site at around 3830 (see Figure 14). In vitro translation of this mRNA resulted in the production of a 70K polypeptide. This size is consistent with an open reading frame, whose initiating methionine is at position 3909, stretching for 693 codons to give a predicted molecular weight of 73,812. Again the protein is slightly basic in nature, 23% of its residues are strongly charged and the most represented amino acids are alanine (16.2%), arginine (11.5%) and leucine (10.5%). The complete amino acid composition is given in Table 4(b). In the relevant region of the vZV genome lies an ORF of 90K. This shows only limited homology to the HSV-1 73K protein (Figures 22 and 23), most apparent at the carboxy terminus. The size difference in the two proteins can be largely explained by the presence of a repeated amino



:::::: End of plot ::::::

Figure 20: CINTHOM Plot of the relationship between HSV-1 33K and VZV 33K

The HSV-1 protein sequence lies along the X-axis and that of the VZV protein along the Y-axis.

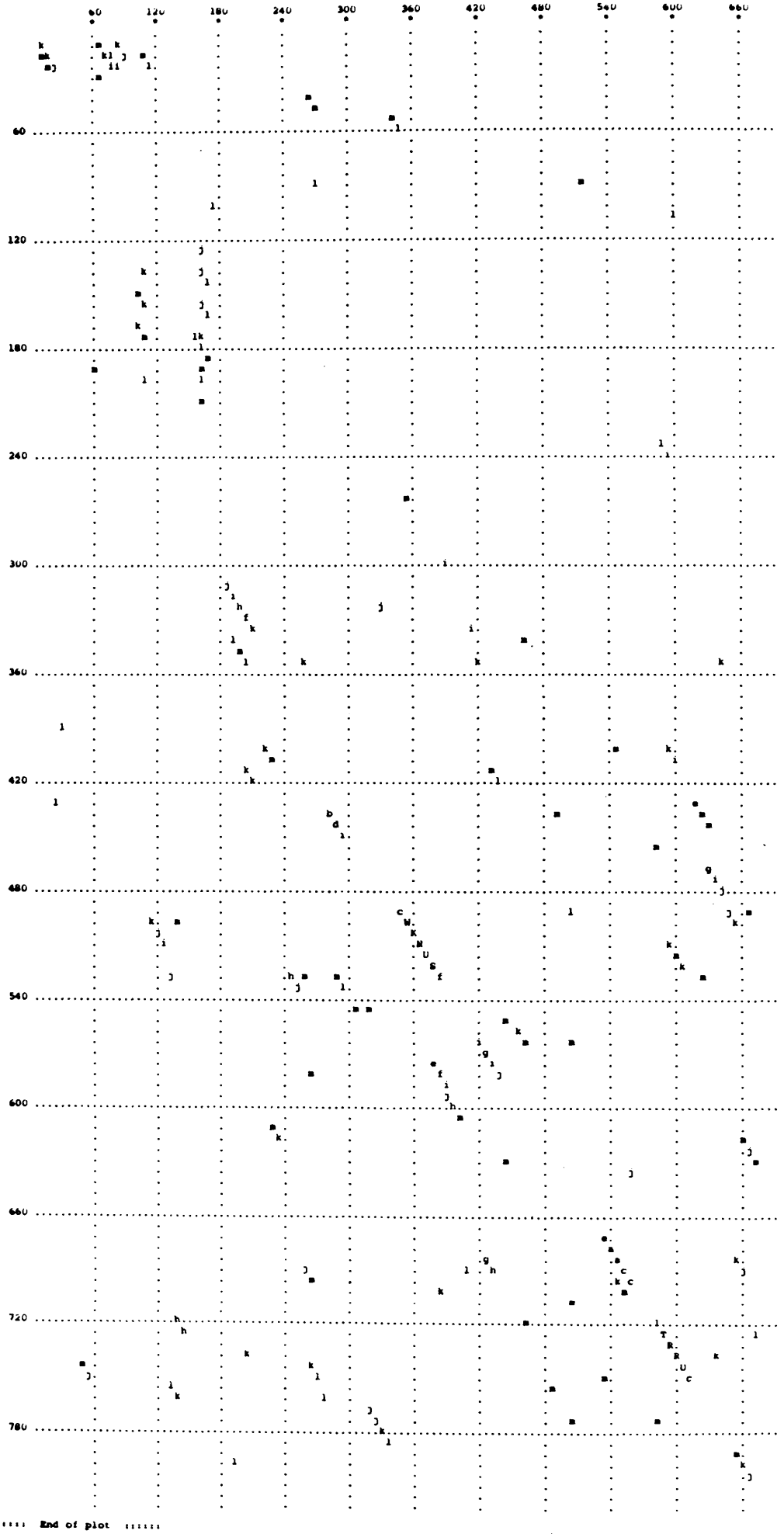
First sequence: HSV1BAMF.33K

Second sequence: VZV.33K

```
  1 M.........................TSRRSVKSGPREVPRDEYEDLYYTPSSGMASPDSPPDTSRRGALQT
    *                         * ****  **   * *    * *    *  **      *   *
  1 MASSDGDRLCRSNAVRRKTTPSYSGQYRTARRSVVVGP...PDDSDDSLGYITTVGADSP......SPVYADLY

 48 RSRQRGEVRFVQYDESDYALYGGSSSEDD..EHPEVPRTRRPVSGAVLSGPGPARAPPPPAGSGGAGRTPTTAP
           *  *   *       * ****     **    *    * *       *       *         *
 66 FEHKNTTPRVHQPNDS.......SGSEDDFEDIDEVVAAFR...EARLRHELVEDAVYENPLS.......VEKP

120 RAPRTQRVATKAPAAPAAETTRGRKSAQPESAALPDAPASTAPTRSKTPAQGLARKLHFSTAPPNPDAPWTPRV
      *   * *    *  *    *         * * * *          * *  *   *****       *
123 SRSFTKNAAVK....PKLEDSPKR........APPGAGA.........IASG..RPISFSTAPKTATSSWCGPT

194 AGFNKRVFCAAVGRLAAMHARMAAVQLWDMSRPRTDEDLNELLGITTIRVTVCEGKNLLQRANE..LVNPDVVQ
     ****** ** * *** *  **   *    **    *  **    ** ** ** ** * ***  *     *
174 PSYNKRVFCEAVRRVAAMQAQKAAEAAWNSNPPRNNAELDRLLTGAVIRITVHEGLNLIQAANEADLGEGASVS

266 DVDAATAT.............RGRSAASR....PTER..PRAPARSASRPRRPVE
          *               *   **     * *   *  ******
248 KRGHNRKTGDLQGGMGNEPMYAQVRKPKSRTDTQTTGRITNRSRARSASRTDTRK.
```

Figure 21: Optimal Alignment of the HSV-1 and VZV 33K ORFs

Parameters as recommended by Taylor(1984).



::::::: End of plot :::::::

Figure 22: CINTHOM Plot of the relationship between HSV-1 73K and VZV 90K

The HSV-1 protein sequence lies along the X-axis and that of the VZV protein along the Y-axis.

First sequence: HSV1BAMF.73K

Second sequence: VZV.90K

```
  1 M.....SAREPAGRRRRASTRPRASP.....
    *          * *  * * * **   *
  1 MQSGHYNRRQSRRQRISSNTTDSPRHTHGTRYR STNWYTHPPQILSNSETLVAVQELLNSEMDQDSSSDASDD

 22 .....VADEPAGDGVGFMGYLRAV
                  * * * *  *    *
 74 FPGYALHHSTYNGSEQNTSTSRHENRIFKLTEREANEEININTDAIDDEGEAEEGEAEEDAIDDEGEAEEGEAE

 41 FRGDDD....SELEALEEMAGDE.PPVRRRREGPRARRRRASEAPPTSHRRASRQRPGPDA.....ARS
      **    * ** *   **       *  *   * *      *       **          *
148 EDAIDDEGEAEEGEAEEDAIDDEGEAEEGEAEEGEAEEGEAEEDAIDDEGEAEEDAAEEDAIDDEGEAEEDYFS

100 QSVRGRLDDDDEVPRGPPQARQGGYL....GPVDARAILGRVGGSRVAPSPLFLEELQYEEDDYPEAVGPEDGG
     *    * *       *       *       *     * * *   *    * *
222 VSQVCSRDADEVYFTLDPEISYSTDLRIAKVMEPAVSKELNVSKRCVEPVTLTGSMLAHNGFD..ESWFAMREC

170 GARSPPKVEVLEGRVPGPELRAAFPLDRLAPQVAVWDESVRSALALGHPAGFYPCPDSAFGLS.......RVGV
     *   *   *      ** *   *  *              **      *   *   *       *
294 TRREYITV...QGLYDPIHLRYQFDTSRMTPPQILRTIPALPNMTLGELLLIFPIEFMAQPISIERILVEDVFL

237 MHFASPDNPAVFFR......QTLQQGEALAWYITG..DGILDLTDRRTKTSPAQAMSFLADAVVRLAINGWVCG
     **       *         *     * *  *  *       *         *  * *       *  *
365 DRRASSKTHKYGPRWNSVYALPYNAGKMYVQHIPGFYDVSLRAVGQGTAIWHHMILSTAACAISNRISHGDGLG

303 TRLHAEARGS....DLDDRAAELRRQFASL.TALRPVGAAAVP.LLSAGGLVSPQSGPDAAVFRSSLGSLLYWP
     * *  * *     *          *   *     *** *    ** *  **  *  *   * * ***
439 FLLDAAIRISANCIFLGRNDNFGVGDPCWLEDHLAGLPREAVPDVLQVTQLVLPNRGPTVAIMRGFFGALAYWP

371 GVRALLDRDCRVAARYAGRMTYLATGALLARFNPDAVRCVLT.REAAFLGRVLDV......LAVMAEQTVQWLS
     *          ***     **   *  *          * **       * *      *       *   *
513 ELRIAISEPSTSLVRYATGHMELAEWFLFSRTHSLKPQFTPTEREMLASFFTLYVTLGGGMLNWICRATAMYL.

438 VVVGARLHPHVHHPAFADVAREELFRALPLGSPAVVGAEHEALGDTAARRLLANSGLNAVLGAAVYALHTALAT
          *    *   *        *  *      *   **           *       *  **
586 ......AAPYHSRSAYIAVCESLPYYYIPVNSDLLCDLEVLLLGEVDLPTVCESYATIAHELTGYEAVRTAATN

512 VTLKYARACGDAHRRRDDAAATRAILAAGLVLQRLLGFADTVVACVTLAAFDGGFTAPEVGTYTPLRYACVLRA
      *   *              * * * * *** ** *          **  **           *    *
654 FMIEFA....DCYKESETDLMVSAYLGAVLLLQRVLGHANLLLLLLSGAALYGG.CSIYIPRGILDAYNTLMLA

586 TQPLYARTTPAKFWADVRAAAEHVDLRP.....ASSAPRAPVSGTADPAFLLKDLEPFPPAPVSGGSVLGPRVR
    ****  *   ** *   *      **          *         *   **   *      **
723 ASPLYAHQTLTSFWKDRDDAMQTLGIRPTTDVLPKEQDRIVQASPIEMNFRFVGLETIYPREQPIPSV....DL

655 VVDIMSQFRKLLMGDEGAAALRAHVSGRRATGLGGPPRP
      *      * *   *
793 AENLMQYRNEILGLDWKSVAMHLLRKY
```

Figure 23: Optimal Alignment of the HSV-1 73K and VZV 90K ORFs

Parameters as recommended by Taylor (1984).

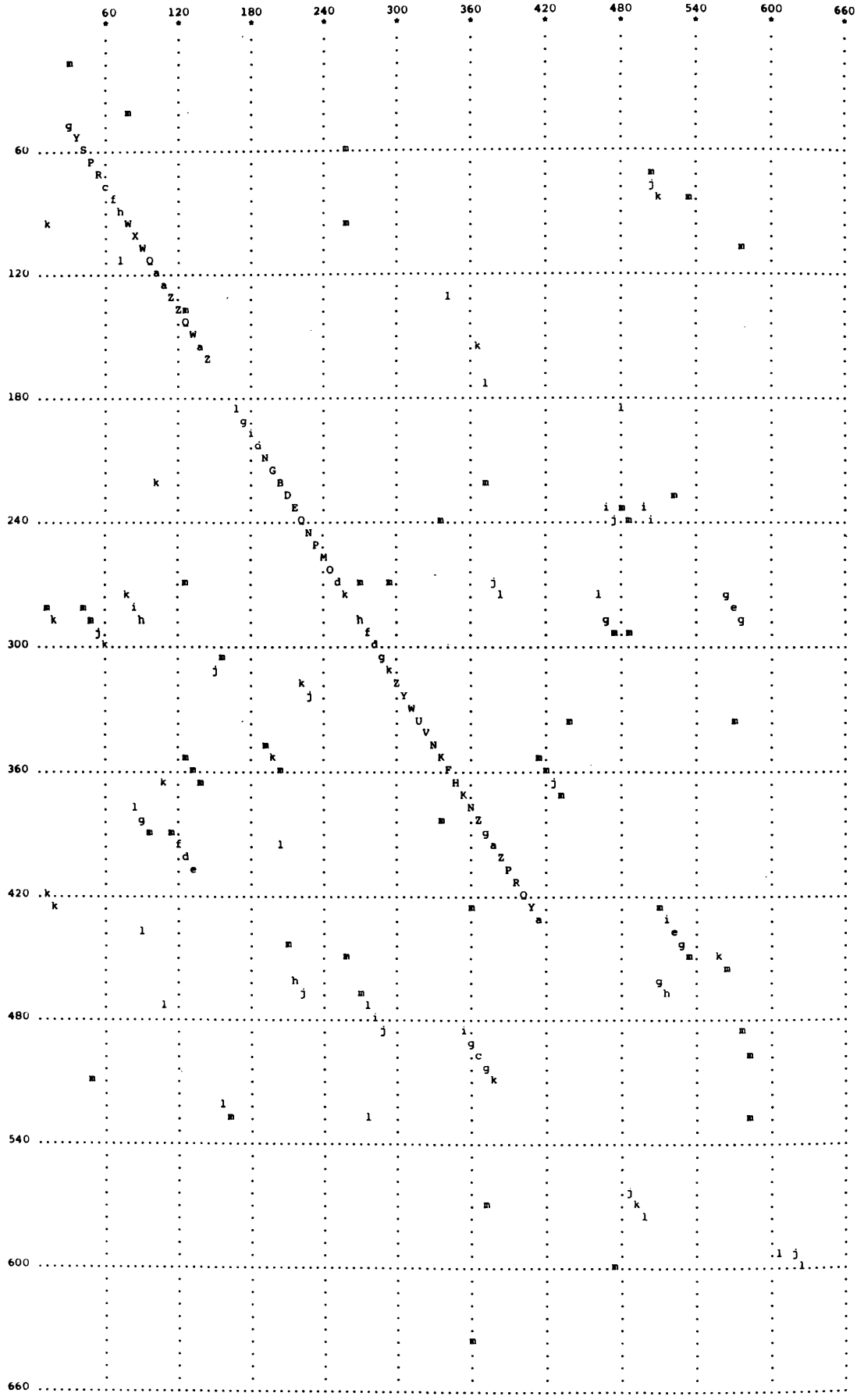
acid sequence in the VZV 90K polypeptide, which is underlined.

5.4.3 BmF72'

A prime has been used in the nomenclature to indicate that the ORF extends 3' to BamHI F. In fact 661 codons are present in BamHI F representing a molecular weight of 72,292. Hall et al (1982) describe a mRNA, labelled mRNA2, which initiates 1,250bp upstream of a PvuII site in the left hand end of EcoRI I (Figure 12a). It appears that the equivalent PvuII site in this sequence is at position 7233 (see Figure 14) and the first potential initiating codon after this start site is at 6073. When mRNA2 was translated in vitro an 85K polypeptide was detected. From this it is estimated that about 80-90% of the coding sequences for this protein are present in BamHI F. The VZV genome has a homologous gene in the predicted genomic position. A matrix comparison of the amino acid sequences (Figure 24) shows good homology for the first 420 residues and an optimal alignment reveals that there are no large blocks of conserved amino acids but rather that the homology is scattered over the entire length of the protein, tailing off towards the carboxy terminus (Figure 25).

5.5 Discussion: Other Genes in BamHI F

There are three predicted ORFs in BamHI F, beside Vmw65; an organisation consistent with the mRNA mapping and in vitro translation studies of Hall et al (1982) (see Figures 12b and 14). All the ORFs in BamHI F lie on the same strand. There is no evidence, either from codon analysis or from the presence of polyadenylation signals, for the existence of protein coding genes on the opposite strand. Unlike Vmw65, there are no studies which assign functions to these genes and any discussion of such must be speculative. The three ORFs will be dealt with in order, from right to left on the prototype genome.



:::::: End of plot ::::::

Figure 24: CINTHOM Plot of the relationship between HSV-1 72'K and VZV 73K

The HSV-1 protein sequence lies along the X-axis and that of the VZV protein along the Y-axis.

First sequence: HSV1BAMF.72K

Second sequence: VZV.73K

```
  1 M.QRRTRGASSLRLAR.................CLTPANLIRGDNAGVPERRIFGGCLLPTPEGLLSAAVGALRQ
    * *  *  **   *                    **                *   *** ***   *  ** **
  1 MFSRFARSFSSDDRTRKSYDGSYQSFNAGERDLPTPTRDWCSISQRITSERVRDGCLIPTPGEALETAVKALSE

 58 RSDDAQPAFLTCTDRSVRLAARQHNTVPESLIVDGLASDPHYEYIRHYASAATQALGEVELPGGQLSRAILTQY
      *      *  *  *    ** ***** *     * *       *       *    *  *  *       *
 75 KTDSLTSPVLQSTERHSVLLGLHHNNVPESLVVSCMSNDVHDGFMQRYMETIQRCLDDLKLSGDGLWWVYENTY

132 WKYLQTVVPSGLDVPEDPVGDCDPSLHVLLRPTLAPKLAGAHPVQERGRGGKYAATVAGLRDALHRIQQYMFFM
    * **          *    *    *   *     * *    **          *       ** **   * ** **
149 WQYLKYTTGAEVPVTSEKVNKKSKSTVLLFSSVVANKPISRHPFKSKVINSDYRGICQELREALGAVQKYMYFM

206 RPADPSRPSTDTALRLNELLAYVSVLYRWASWMLWTTDKHVCHRLSPSNRRFLPLGGSPEAPAETFARHLDRGP
    ** **  ** **  *  *  **    * *  * *   *  ** *    **       *   *      *** **
223 RPDDPTNPSPDTRIRVQEIAAYTATGYGWMLWFLDVVDARVCRHLKLQFRRIRGPRAS.VIPDDLLRRHLKTGP

280 SGTTGSMQCMALRAAVSDVLGHLTRLANLWQTGKRSGGTYGTVDTVVSTVEVLSIVHHHAQYIINATLTGYGVW
      *      * *   * * *  **       *  **    *  **      *** ** **  *  **  *
296 AVSAGTGVAFILAATTASALTALLRISVLWRKEEWRDGLNGTAAAIVAAVELITLLHHHFQYLINMMLIGYACW

354 ATDSLNNEYLRAAVDSQERFCRTTAPLFPTMTAPSWARMELSIKAWFGAALAADLLRNGAPSLHYESILRLVAS
      **  *   *  * **      *  **   **   * *  **  * *   *  *      *      * **
370 GDGGLNDPYILKALRAQGRFLYFAGQLVRTMSTHSWVVLETSTHMWFSRAVAQSILAHGGKPTKY.YAQVLAAS

428 RRTTWSAGPPPDDMASGPGGHRAGGGTCREKIQRARRDNEPPPLPRPRLHSTPASTRRFRRRRADGAGPPLPD.
     * *            *   *          *   * *        ** * *  *            *  * *
443 KRYT.............PLHLRRISEPSSVSDQPYIRFN........RLGS.PIGTGIGNLECVCLTGNYLSDD

501 ..ANDPVAEPPAAATQPATYYTHMGEVPPRLPARNVAGPDRRPPAATCPLLVRRASLGSLDRPRVWGPAPEGEP
      *   *   *   *           *  *    *      * *      **     *
495 VNASSHVINTEAPLNSIA.............PDTNRQRTSRVLVRPDTGLDVTVRKNHCLDIGHTDG.......

573 DQMEATYLTADDDDDDARRKATHAASARERHAPYED...DESIYETVSEDGGRVYEEIPWMRVYEN....VCVN
      **      *   * **      *       *   *    **       *** **   *  *      *
549 SPVDPTY.....PDHYTRIKAEYEGPVRDESNTMFDQRSDLRHIETQASLNDHVYENIPPKEVGFNSSSDLDVD

640 TAN.........AAPASP.YIEAENPLYDWGG
      *          **  *   *
618 SLNGYTSGDMHTDDDLSPDFIPNDVPVRCKTTVTFRKNTPKSHH
```

Figure 25: Optimal Alignment of the HSV-1 72'K and VZV 73K ORFs

Parameters as recommended by Taylor(1984) .

5.5.1 BmF33

This ORF lies immediately upstream of that for Vmw65. Hall et al (1982) could detect four mRNAs spanning this region, named mRNAs 3, 4, 6 and 7, which all translated to give a 42,000 molecular weight product in vitro. The mRNAs 4 and 7 were 5' co-terminal as were the other two mRNAs, while it was implied that mRNAs 6 and 7 were 3' co-terminal and that mRNAs 3 and 4 were 3' co-terminal with mRNA 5 (that for Vmw65). As only a single potential polyadenylation signal can be seen in the area in which mRNA 6 and 7 are expected to terminate, it is reasonable to assume that indeed these mRNAs terminate at the same position. Similarly, the next polyadenylation signal downstream is that for mRNA Vmw65, indicating that this mRNA is 3' co-terminal with mRNAs 3 and 4. The detected mRNAs were shown by Hall et al (1982) to fall into two temporal classes: 4 and 7 are early mRNAs, whilst 3 and 6 are late mRNAs. This might suggest that the product of BmF33 plays an important role in the replication cycle.

The ORF is comprised of 301 amino acids, whose composition is given in Table 4(a). Compared to the average amino acid composition of all non-mitochondrial human proteins (NMHP) present in the NBRF database version 2.0 (see Table 5; Chen & Barker, 1985), BmF33 is much higher in proline, arginine, and alanine - a general characteristic of HSV proteins due to the high %G+C content of HSV DNA (McGeoch, 1984). The protein is very basic, with a surplus of 14 strongly basic residues over acidic residues, and slightly more highly charged than average. An analysis of the charge distribution along the protein backbone showed that it is slightly skewed such that the centre of the molecule has fewer charged amino acid residues than either end (not shown). There are fewer residues with large bulky side chains, for example phenylalanine (4), tryptophan (2), histidine (3) and tyrosine (6), than the NMHP average, possibly suggesting a tightly packed tertiary structure.

Table 4: Amino Acid Composition Tables

(a) BmF33: Number of identified codons= 301
Approximate Molecular Weight= 32251.86

RES.	NUM.	%	RES.	NUM.	%	RES.	NUM.	%	RES.	NUM.	%
Ala	43	14.3	Arg	39	13.0	Asn	6	2.0	Asp	16	5.3
Cys	2	0.7	Gln	9	3.0	Glu	16	5.3	Gly	20	6.6
His	3	1.0	Ile	2	0.7	Leu	15	5.0	Lys	7	2.3
Met	5	1.7	Phe	4	1.3	Pro	37	12.3	Ser	26	8.6
Thr	24	8.0	Trp	2	0.7	Tyr	6	2.0	Val	19	6.3

Translation begun with base no. 634
Translation stopped at termination codon (base no.1537)

(b) BmF73: Number of identified codons= 693
Approximate Molecular Weight= 73812.33

RES.	NUM.	%	RES.	NUM.	%	RES.	NUM.	%	RES.	NUM.	%
Ala	112	16.2	Arg	80	11.5	Asn	5	0.7	Asp	44	6.3
Cys	7	1.0	Gln	17	2.5	Glu	33	4.8	Gly	64	9.2
His	13	1.9	Ile	6	0.9	Leu	73	10.5	Lys	6	0.9
Met	9	1.3	Phe	23	3.3	Pro	57	8.2	Ser	37	5.3
Thr	28	4.0	Trp	6	0.9	Tyr	14	2.0	Val	59	8.5

Translation begun with base no.3909
Translation stopped at termination codon (base no.5988)

(c) BmF72': Number of identified codons= 661
Approximate Molecular Weight= 72292.12

RES.	NUM.	%	RES.	NUM.	%	RES.	NUM.	%	RES.	NUM.	%
Ala	82	12.4	Arg	67	10.1	Asn	17	2.6	Asp	39	5.9
Cys	10	1.5	Gln	18	2.7	Glu	33	5.0	Gly	50	7.6
His	19	2.9	Ile	15	2.3	Leu	63	9.5	Lys	8	1.2
Met	12	1.8	Phe	10	1.5	Pro	60	9.1	Ser	41	6.2
Thr	45	6.8	Trp	12	1.8	Tyr	23	3.5	Val	37	5.6

Translation begun with base no.6073
Translated to 3' end--No termination codon found.

Table 5: % Amino Acid Composition of Non-Mitochondrial Human Proteins.

RES.	%	RES.	%	RES.	%	RES.	%
Ala	6.7	Arg	5.1	Asn	3.8	Asp	5.3
Cys	2.7	Gln	4.4	Glu	7.0	Gly	6.6
His	2.4	Ile	4.0	Leu	9.8	Lys	6.2
Met	2.1	Phe	4.0	Pro	5.3	Ser	8.4
Thr	5.9	Trp	1.3	Tyr	3.0	Val	6.7

5.5.2 BmF73

This ORF lies immediately downstream of Vmw65. Hall et al (1982) identified a 4.7 kbp mRNA spanning the entire left half of BamHI F and terminating in the adjacent DNA (Frink et al, 1981). This mRNA (mRNA 1) was translated to an approximately 70,000 molecular weight polypeptide upon in vitro translation. The predicted size of BmF73 is 693 codons with a molecular weight of 73,812, thus the predicted protein is larger than the in vitro translated product (however, a close scrutiny of the gels presented by Hall and co-workers suggests that the mRNA 1 translation product is closer to 73K than to 70K). If the size difference is real, it is likely to be an artifact as a result of some physico-chemical property of the SDS-protein complex. As in the case of BmF33, BmF73 is high in alanine (16.2%) and arginine (11.5%). The proline content is much closer to the average NMHP value than that of BmF33. In terms of the strongly charged amino acids, BmF73 is predicted to be slightly basic. There are no obvious outstanding features to the predicted protein. Morse et al (1978), by means of intertypic recombinants, were able to assign a protein, of molecular weight 72,500 (ICP23), to a region of the HSV-1 genome which includes BamHI F. It is tempting to speculate that they were detecting the product of BmF73 and this might allow further characterisation of this gene.

5.5.3 BmF72'

Hall and co-workers reported that a 2.5 kb mRNA (mRNA 2) initiated transcription 1,250 bp upstream of a PvuII site in the left hand end of BamHI F, within the 4.7 kb mRNA 1. This mRNA 2 translated to an 85,000 molecular weight polypeptide in vitro. In an earlier paper from Wagner's group (Frink et al, 1981) the 3' end of a 2.6 kbp mRNA was reported to map around 0.647 map units and probably consists of the 3' termini of both mRNA 1 and mRNA 2. There is a discrepancy in this region between the restriction map of Wagner's group and the sequence of BamHI F. Wagner's lab consistently report the position of a HindIII site within

the left hand end of BamHI F, however, the sequence presented here reveals no such site. Due to this inconsistency there must still be some doubt as to the exact location of the 3' ends of mRNAs 1 and 2, but it is reasonable to assume that no more than 300-400 bases of mRNA 2 lie outside BamHI F. Thus, if there is less than 100 bases of 3' untranslated sequence, there may be up to 100 amino acids still to be allocated to BmF72' which would increase the predicted molecular weight closer to the in vitro translation product of 85K.

There are 661 amino acids present in BmF72' with a total molecular weight of 72,292. The protein shows a fairly typical HSV-1 amino acid composition. The most highly represented amino acids are alanine (12.4%), arginine (10.1%), leucine (9.5%) and proline (9.1%). There is a very slight surplus of strongly charged basic residues over acidic residues. The percentage of residues with large side chains (9.7%) is close to the BMHP average of 10.7%. The protein has no exceptional characteristics.

5.5.4 Homology

VZV has been assigned to the sub-family alphaherpesvirinae, the same group as HSV-1 and -2, on the basis of its biological properties (Roizman et al, 1981). However, VZV displays a different genome structure to these two viruses. Based on the genome classification of Roizman (1982), VZV was assigned to group D, i.e. composed of two segments, L and S, the S segment having structure IR_S-U_S-TR_S (much like that of HSV) but the L segment having no large repeated region. Davison (1984) has shown that, in fact, the VZV U_L segment is flanked by short (88.5 bp) inverted repeats and that the genome has no equivalent to the HSV a sequence. Despite this observation, the VZV L and S segments are able to invert with respect to each other, but invertion of U_L occurs less frequently than that of U_S (only about 5% of VZV genomes). By convention, the prototype orientation of VZV is equivalent to the I_{SL} orientation of HSV-1 (see Figures 5 and 26). This means that when comparing the

prototype orientations of HSV-1 and VZV the region 0.645-0.695 in HSV-1 corresponds to 0.086-0.144 in VZV. These corresponding regions of the HSV-1 and VZV genomes appear to be almost exactly colinear. As can be seen from Figures 17, 18, and 20-25, all four predicted ORFs in BamHI F have homologues in the corresponding region of VZV, and furthermore the genetic organisation is the same. This information is summarised in Figure 26. It is interesting to note that the best conserved gene in the region is that for Vmw65/VZV 46K. It is reasonable to infer from this observation that the selective pressure on this locus is correspondingly greater than on the genes flanking it, consistent with this protein having an important structural role in the virion particle. The converse of this argument is that the BmF73/VZV 90K locus may be of interest when searching for genes which confer biologically unique properties on the respective viruses, as these genes show a very low degree of similarity. Draper and co-workers (1984) have compared HSV-1 and -2 in the region around 0.645 map units by means of DNA-DNA hybridisation. They found that there was strong cross-hybridisation between sequences now known to correspond to Vmw65, BmF33 and BmF73, a finding consistent with the fact that the product of these HSV-2 sequences can trans-activate HSV-1 IE promoters (data not shown). However, sequences downstream of the PvuII site at position 7237 hybridised poorly or not at all to the corresponding HSV-2 region. These sequences contain the carboxy terminal half of BmF72' and it is interesting to note that this portion of BmF72' has no homology to the corresponding region of VZV 73K. Indeed, it is likely that the HSV-1 protein is much longer than its VZV homologue.

In general the degree of conservation at the amino acid level between HSV-1 and VZV is remarkable, especially as no significant homology exists at the DNA level (data not shown). Neither the ORFs in BamHI F, nor their equivalents in VZV, exhibit obvious homology to ORFs in the EBV genome, which has been completely sequenced (Baer et al, 1984; data not shown; A. Davison, personal communication). Indeed, a

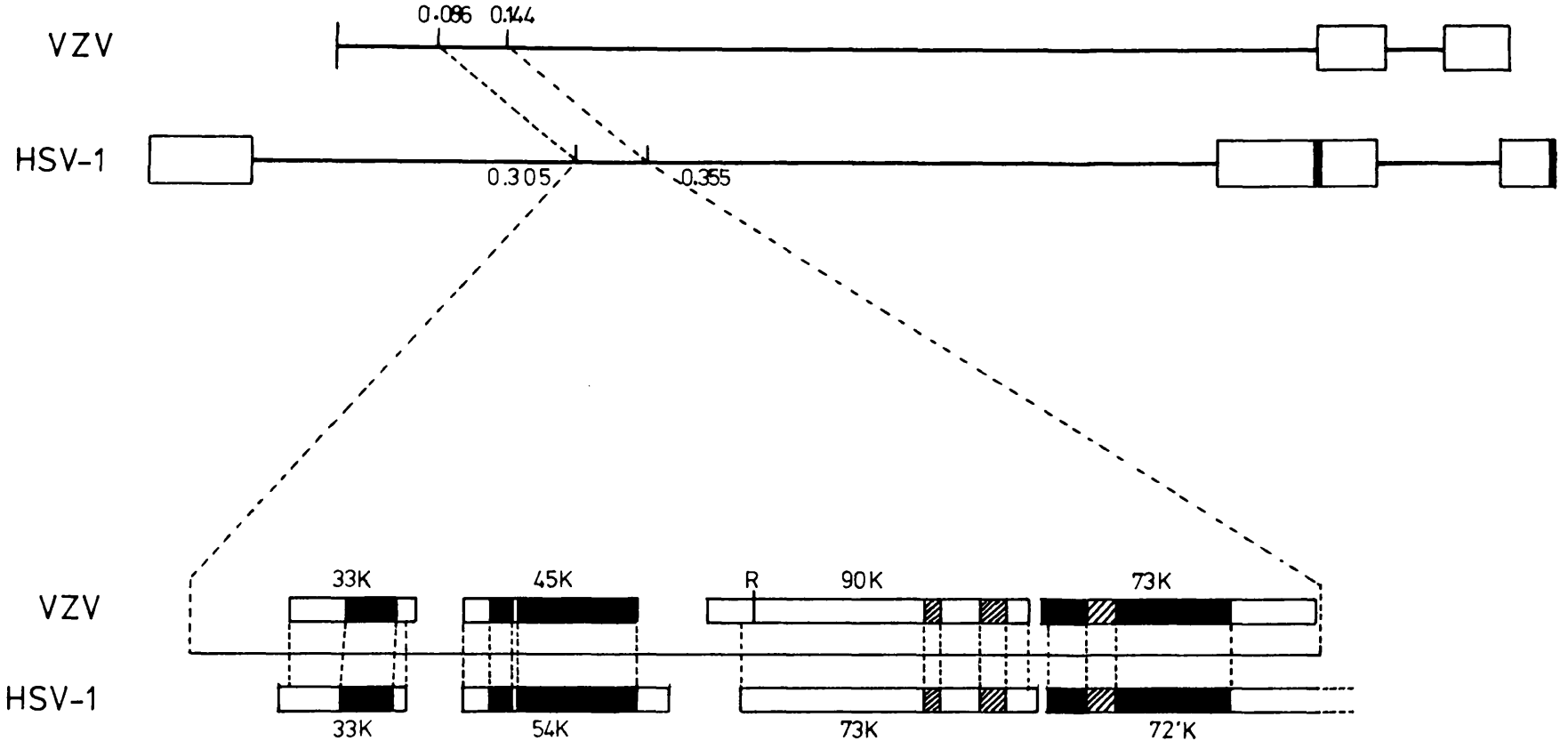

VZV
0.086
0.144
HSV-1
0.305
0.355
VZV
33K
45K
R
90K
73K
HSV-1
33K
54K
73K
72'K

Figure 26: An Overview of the Homology Between HSV-1 and VZV in the Region of BamHI F

The equivalent regions of the two viruses are aligned. Solid boxes indicate regions of "strong" homology (>35%), hatched boxes indicate regions of lower homology (25%< x <35%) and open boxes indicate areas of little or no homology.

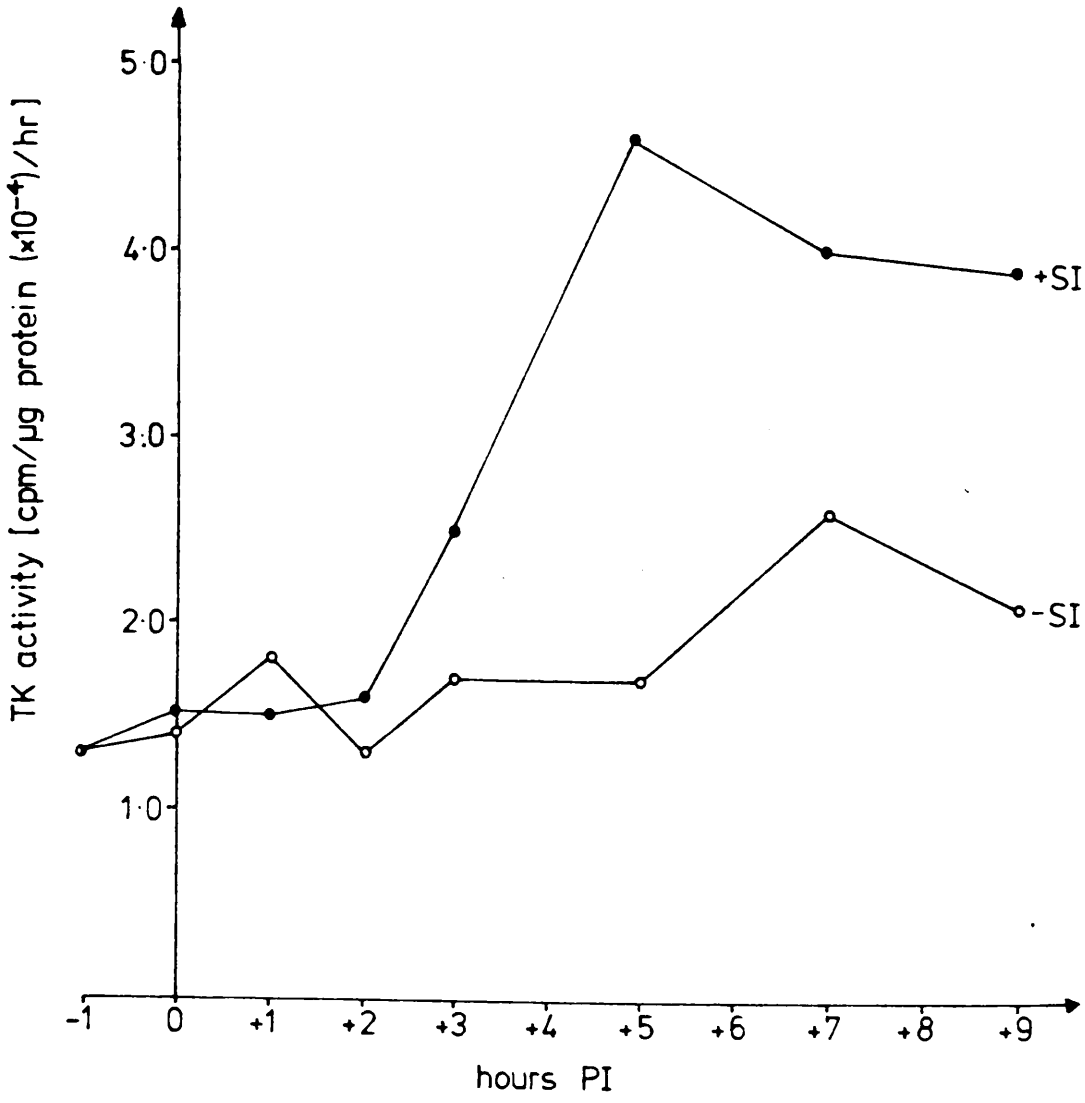
This figure is available as a loose sheet at the back of the thesis.

Table 6: Time course of pTKN2 superinfection by TK⁻ virus

TIME (hr PI)	-1	0	+1	+2	+3	+5	+7	+9
-SI	1.3*	1.4	1.8	1.3	1.7	1.7	2.6	2.1
+SI	1.3	1.5	1.5	1.6	2.5	4.6	4.0	3.9

*cpm/ ug protein ($x10^{-4}$)/ hr

Figure 28:

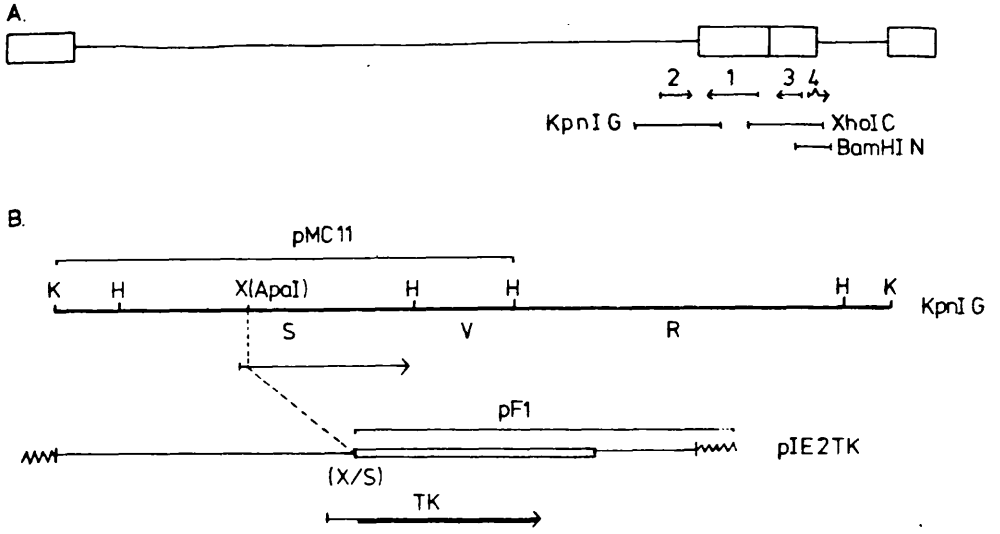


comparison of the entire VZV genome against that of EBV has revealed remarkable genome rearrangements and shows that detectable VZV/EBV homology disappears around the region of VZV 46K (A. Davison, personal communication). EBV is classified as a gammaherpesvirus (Roizman et al, 1981) and, at the genomic level, would certainly appear to be much more distantly related to either HSV-1 or VZV than the two alphaherpesviruses are to each other. It is possible that the power of gene cloning, DNA probing and sequencing could be harnessed by the viral taxonomist to rapidly assign any herpesvirus to one of the sub-families by identifying conserved and non-conserved genomic regions across the family as a whole. In this respect the BamHI F region of HSV-1, and in particular conservation of the Vmw65 gene, may be diagnostic of the alphaherpesvirinae.

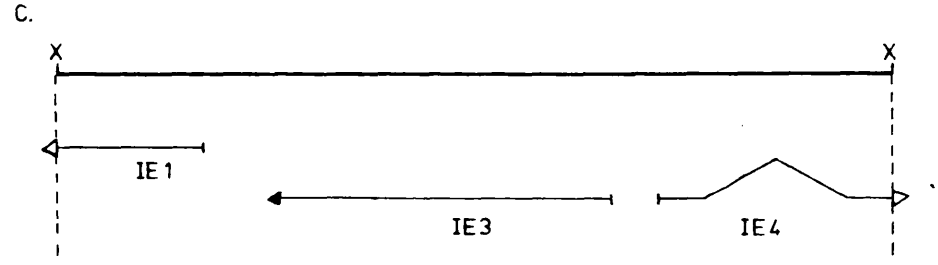
5.6 Negative Control

5.6.1 Evidence that an IE3-TK hybrid plasmid is first stimulated then down-regulated by HSV-1 superinfection.

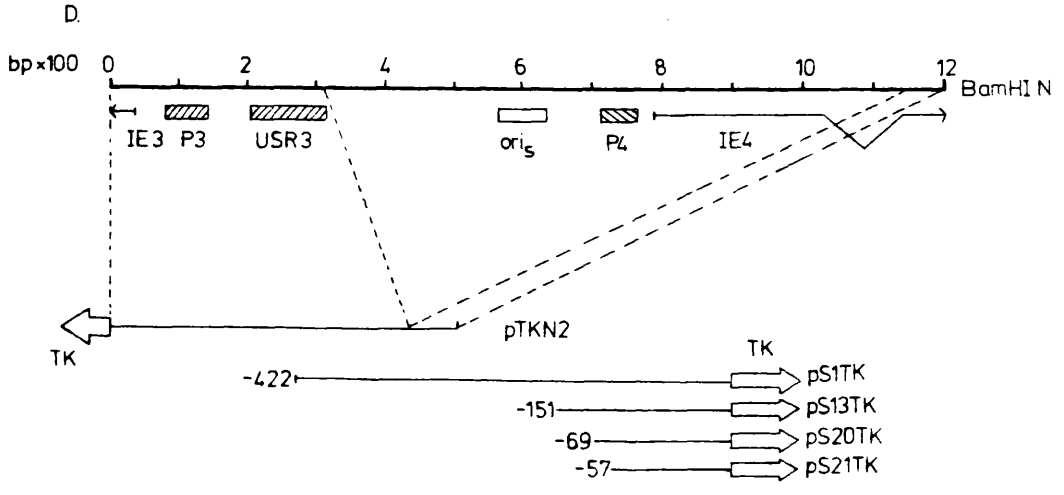
The rationale behind the design of this experiment was that plasmid DNA, containing the relevant cis-acting control elements and an assayable marker (the TK gene) might, when transfected into BHK cells, be regulated as a viral gene upon superinfection by HSV-1 TK^- virus. Petri plates of BHK C13 monolayers were therefore transfected with the plasmid pTKN2 and incubated overnight at 37°C. Plasmid pTKN2 contains the IE3 promoter sequences to -331 (see Figure 27) driving expression of the HSV-1 TK structural gene. The transfected plates were then superinfected with 20 pfu/cell of HSV-1 TK^- virus, or mock-infected, and individual plates were harvested after various time intervals. TK assays were performed on cytoplasmic extracts and calibrated to the extract protein content (as infection with live virus leads to some cell loss). The data are presented in Table 6 and Figure 28. The shape of the curve indicates that expression of TK activity from the chaemeric plasmid was first stimulated, presumably by the virion component, but that



STRUCTURE OF pIE2TK



CODING CAPACITY OF XhoI C



STRUCTURE OF pTKN2, pS1TK, S13TK, S20TK & S21TK

Figure 27: DNA Sequences Present in the Plasmids Used in Co-transfection Studies

Panel A. is a representation of the prototype HSV-1 genome showing the relative positions of the DNA fragments used in the various plasmid constructions.

Panel B. is a detail of the KpnI G fragment showing the structure of the plasmid pIE2TK. This plasmid was constructed using two other plasmids: p27 (C.M. Preston, unpublished data) and pF1 (Pelham, 1982). p27 is a derivative of pMC11 (M.E.M. Campbell, unpublished data), which is shown in the figure and consists of the left hand end of KpnI G from the KpnI site to the right most boundary of HpaI V. An ApaI site at position +60, with respect to the start of IE mRNA 2, was converted to a unique XhoI site. pF1 contains the HSV-1 TK gene in which a HaeIII site just before the mRNA start has been converted to a unique SalI site. The construction of pIE2TK involved the ligation of an XhoI-EcoRI fragment from p27 to a SalI-EcoRI fragment from pF1. In both cases the EcoRI site is in vector sequences. the XhoI/SalI hybrid site is shown.

Panel C. shows the coding capacity of XhoI C. The figure is not to scale. The salient features are that this fragment contains the entire coding and flanking sequences of IE gene 3 but severely truncated forms of the genes for IE1 and IE4. Arrows indicate the direction of transcription of each gene, hollow heads indicate that the mRNA does not fall entirely within XhoI C.

Panel D. details the sequences present in the plasmids pTKN2, pS1TK, pS13TK, pS20TK and pS21TK, the construction of these plasmids has been described previously (Cordingley et al, 1983; Preston et al, 1984). P3 and P4 indicate the extent of the promoter regions of IE gene 3 and IE gene 4 respectively. USR3 indicates the upstream enhancer sequences of IE gene 3. Ori_S is the origin of replication in the short repeat of HSV-1.

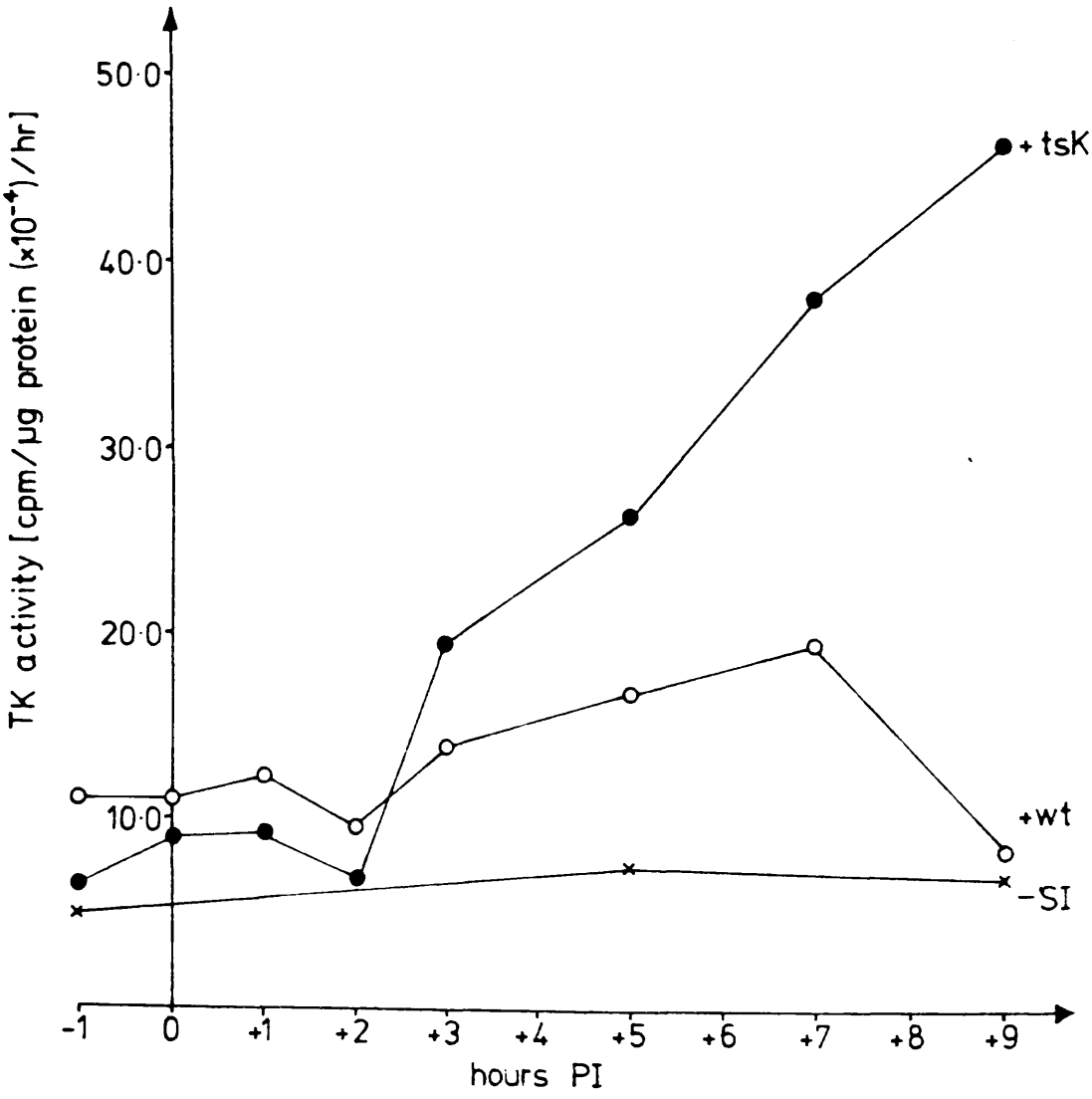
The letters indicate: K=KpnI; H=HpaI; X=XhoI; Sa=SalI. S, V and R indicate genomic HpaI fragments.

Table 7: Time course of pTKN2 superinfection by HSV-1 tsK TK⁻ virus

TIME (hr PI)	-1	0	+1	+2	+3	+5	+7	+9
tsK TK⁻	6.4*	9.0	9.4	6.9	19.6	26.4	38.0	46.4
TK⁻	11.0	11.0	12.3	9.5	13.8	16.9	19.7	8.3
-SI	4.9	-	-	-	-	7.5	-	6.8

*cpm/ ug protein (x10^{-4})/ hr

Figure 29:



this activity soon peaked (at around 5-7 hours post-infection) and then reached a plateau or even dropped. One conclusion from this experiment is that the initial "switch-on" was arrested by subsequent steps in the viral cycle. The plateau effect observed is consistent with a reduced level of transcription, a lowering of mRNA stability or a decrease in the efficiency by which the hybrid mRNA is translated into active TK. The latter two possibilities are deemed unlikely as there were only 27bp of IE3 mRNA (untranslated leader) present in the construction, therefore the working hypothesis was that the down-regulation occurred at the level of transcription.

5.6.2 Vmw175 is involved in the plateau effect.

Previous work has shown that the virus mutant tsK, which contains a temperature sensitive lesion in the IE3 gene product Vmw175, over-produces IE mRNA and proteins at the NPT. Temperature shift experiments led to the conclusion that Vmw175 was responsible for its own transcriptional autoregulation as well as for the activation of E and L genes (Preston, 1979b, Dixon & Schaffer, 1980, Watson & Clements, 1980). Using the superinfection system described in the previous experiment, it was possible to determine whether functional Vmw175 was required for the plateau effect. A similar experiment was performed, but plates of transfected cells were superinfected with either HSV-1 TK^- virus or HSV-1 TK^- tsK virus and incubated at 38.5°C, the NPT for tsK. A third set of plates was set up without superinfection. Cells were harvested at suitable times, as before, and TK assays performed on cell extracts. The data are presented in Table 7 and Figure 29. It is clear that the level of TK activity did not reach a plateau when the transfected cells were superinfected under conditions in which Vmw175 is non-functional. Two distinct possibilities exist. The first is that the mutant virus is more effective at stimulating the expression of the chaemeric plasmid. The second is that the phenomenon observed with plasmid pTKN2 accurately reflects the autoregulation which occurs during

virus infection. Evidence will now be presented to support the latter case.

5.6.3 Vmw175 can trans-activate other IE promoters, but not its own.

Several workers have described the ability of certain viral IE products to promiscuously trans-activate both viral and non-viral promoters in a co-transfection assay (see INTRODUCTION). In order to determine if this was the case for HSV-1 IE promoters, co-transfection experiments were performed in which suitable IE/TK chaemeric test plasmids (for details see Figure 27) were mixed with cloned XhoI C fragment from either wild-type or tsK HSV-1 (see Figure 27). XhoI C contains the complete coding and flanking sequences of the gene for Vmw175 and faithfully stimulates expression from early promoters. The tsK derived XhoI C fragment is active in trans-activation at 31°C but not 38.5°C (C.M.Preston, unpublished results). The co-transfection was performed at the NPT for tsK (38.5°C), and cell extracts made and assayed for TK activity.

Table 8 shows the results of these experiments. It is clear that an IE4-TK chaemeric plasmid (pS1TK), in which the TK coding sequences are transcribed from the IE4 promoter, is stimulated about two-fold by wild-type XhoI C but not at all by tsK XhoI C. A construct in which the IE2 5' flanking sequences mediate expression of the TK coding region (pIE2TK) is stimulated to an even greater extent, about five fold on average, although a slight stimulation by tsK XhoI C was noted. It is apparent, however, that pTKN2 (the prototype IE3-TK chaemeric plasmid) is not stimulated even by wild-type XhoI C, indeed there may be some dimunition in the level of TK activity in the presence of this plasmid. Two possible explanations for this result appear likely. First, there may be specific sequences which mediate the stimulatory response of Vmw175. This would predict that IE3 lacks these sequences which are present in the promoter regions of both IE2 and IE4/5. The apparent switch-off of IE3 expression by Vmw175 may be an indirect consequence of

Plasmid	XhoI C Fragment	Experiment 1	2	3	4	5	Avg.	stim	Avg. stim
pS1TK	+	12946*	29445	29123	13590	14482	19917	2.0	2.8
	tsK	7610	12921	15520	4903	5765	9344	0.9	1.2
	-	7614	19720	17068	2356	4157	10183	1	1
pIE2TK	+	20939	7215	8324			12159	4.9	8.4
	tsK	4860	2574	2589			3341	1.4	2.1
	-	1031	3025	3333			2463	1	1
pTKN2	+	12795	16411	9762			12989	0.7	0.9
	tsK	27671	20585	14985			21080	1.2	1.3
	-	31054	12860	9787			17900	1	1

*cpm/ 10^4cells/ hr

Table 8: Trans-activation of IE promoters by Vmw175

This table shows the results from transfection experiments. The type of plasmid DNA used is in column 1, the kind of XhoI C plasmid co-transfected is in column 2 (+ = wild-type, tsK = that from HSV-1 tsK, and - = pBR322 control). The next 5 columns represent the results from independent experiments. These results are summarised by presentation of the average response and the stimulation relative to cells transfected without XhoI C, calculated from this average. The final column is the average of the stimulations calculated for each individual experiment.

competition from other promoters which do respond to trans-activation. Second, IE3 might have inhibitory sequences in its 5' flanking region which prevent the Vmw175 protein from activating expression or which respond negatively to the presence of Vmw175. A qualitative comparison of the different levels of expression from the three IE promoters shows that, without trans-activation, IE2 was poorer than both IE3 and IE4/5. This difference was less pronounced in the presence of XhoI C, indeed the stimulation of the IE2 construct was more than twice that of the IE4/5 plasmid. These experiments suggest that Vmw175 is capable of modulating the expression of at least two IE genes in vivo and that this may play an important role in the viral life cycle.

5.6.4 Are specific sequences required for the trans-activation of IE4 by Vmw175 ?

A series of 5' deletion mutants of the IE4-TK plasmid construct (see Figure 27) were assayed for the ability to be stimulated by XhoI C in the co-transfection system, and the results are presented in Table 9. A deletion which removed sequences to -151 (pS13TK), with respect to the mRNA start site, showed reasonable unstimulated expression and about a four-fold increase in the presence of XhoI C. This experiment was performed only once and thus no extensive conclusions can be made. Deleting sequences to -69 (pS20TK) had no effect on either stimulated or unstimulated levels of expression, as compared to the prototype plasmid pS1TK. From this it can be inferred that there are no sequences between -422 and -69 which substantially alter the response of IE4 to Vmw175. However, deletion of a further 12bp to -57 (pS21TK) resulted in a significant drop in the absolute level of TK activity but not in the degree of stimulation, which remained two-fold. It was concluded from this result that elements which respond to Vmw175 are integral to the constitutive promoter function and may be indistinguishable from "promoter" elements. This conclusion is consistent with similar, more detailed, work on the gD promoter of HSV-1, an

Plasmid	XhoI C Fragment	Experiment 1	2	3	4	5	6	Avg.	Stim.	Avg. Stim.
pS13TK	+ tsK -	12695* 4621 3425							3.7 1.3 1	
pS20TK	+ tsK -	24841 3453 10631	10399 4029 3245	28703 3376 ND	29187 11925 20855	26291 14937 11345	21904 4403 7099			
pS20TK (cont.)	+ tsK -	28125 6983 4873	29082 20630 19219	34914 16843 21693	10701 2873 2998	9963 3312 2487		23100 8433 10438	2.2 0.8 1	2.9 1.0 1
pS21TK	+ tsK -	10918 3452 7774	21992 4906 5578	16606 3536 5808	2115 856 1580	2640 947 1332		9325 2805 4612	2.0 0.6 1	2.0 0.6 1

*cpm/ 10^4cells/ hr

Table 9: Sequences required for the activation of IE4 by Vmw175

This table shows the results from transfection experiments. The type of plasmid DNA used is in column 1, the kind of XhoI C plasmid co-transfected is in column 2 (+ = wild-type, tsK = that from HSV-1 tsK, and - = pBR322 control). The next 6 columns represent the results from independent experiments, note that 11 seperate experiments were performed with pS20TK. These results are summarised by presentation of the average response and the stimulation relative to cells transfected without XhoI C, calculated from this average. The final column is the average of the stimulations calculated for each individual experiment.

early class promoter which responds to Vmw175 (Everett, 1983; 1984).

5.6.5 The IE3 enhancer region can inhibit trans-activation by Vmw175

It has been suggested that enhancers are capable of inhibiting trans-activation by viral IE products (Green et al, 1983; Imperiale et al, 1983; Velcich & Ziff, 1984). In order to test whether the IE3 enhancer-like sequences prevented the stimulation of IE3 by XhoI C, a plasmid was constructed in which the IE3 enhancer region (-331 to -174) was placed upstream of the natural TK gene promoter. This construct was compared to pTK1 in co-transfection experiments, as described above. The results of these experiments are presented in Table 10. These data suggest that the IE3 enhancer element is not able to stimulate expression from the full TK promoter, indeed it may even slightly depress the wild-type level of expression. Furthermore, whilst the normal TK promoter responds vigorously to the trans-acting factors produced by XhoI C, when the IE3 enhancer is present in cis the stimulation is significantly reduced.

5.7 Discussion: Negative Control of IE3

The data presented in Figures 28 and 29 appear to confirm the role of Vmw175 as an autoregulatory protein and indicate that the tsK mutant of Vmw175 has lost its ability to autoregulate at the NPT. The co-transfection experiments which followed were an attempt to investigate further the mechanism of autoregulation. Two main conclusions can be arrived at from the results presented, (a) the product of IE gene 3 (Vmw175) stimulates expression from two other IE promoters - IE2 and IE4/5, (b) Vmw175 is not able to stimulate expression from its own promoter.

The stimulation of other IE promoters by Vmw175 was unexpected since it was believed that at E and L times all IE transcription is much reduced (Clements et al, 1977; Dixon & Schaffer, 1980). The results obtained here suggest

Plasmid	XhoI C Fragment	Experiment 1	2	3	4	Avg.	stim.	Avg. stim
pTK1	+	26844*	30954	22370	175576	24436	4.5	4.8
	tsK	1153	2986	4751	3096	2997	0.6	0.7
	-	3584	8142	5284	4545	5389	1	1
pIE3TK1	+	4038	3571	3114		3573	1.8	2.1
	tsK	1569	1512	1445		1509	0.7	0.9
	-	3211	1389	1337		1979	1	1

*cpm/ 10^4cells/ hr

Table 10: Role of IE3 enhancer in trans-activation

This table shows the results from transfection experiments. The type of plasmid DNA used is in column 1, the kind of XhoI C plasmid co-transfected is in column 2 (+ = wild-type, tsK = that from HSV-1 tsK, and - = pBR322 control). The next 4 columns represent the results from independent experiments. These results are summarised by presentation of the average response and the stimulation relative to cells transfected without XhoI C, calculated from this average. The final column is the average of the stimulations calculated for each individual experiment.

either that some IE genes are transcribed throughout the replicative cycle (under IE3 control just like true E and L genes) or that an E or L gene product acts to switch-off IE gene expression, an idea originally postulated by Honess and Roizman (1974; 1975). The argument will be advanced here that within the IE class there are at least three distinct types of promoter which behave very differently and that probably only IE3 is autoregulated at post-IE times.

It is now generally held that Vmw175 falls into a class of trans-activators, which act via basal promoter elements in a relatively non-specific manner and include adenovirus E1A protein and the product of the pseudorabies virus IE gene (Everett, 1983; 1984; Everett & Dunlop, 1984; Eisenberg et al, 1985). It would appear that the same mechanism of stimulation operates in the system described here. At least in the case of IE4/5, it seems likely that the stimulatory effect of Vmw175 was dependent on the function of the resident promoter element (Preston et al, 1984) and not on a Vmw175-specific responder element. Thus, deletion of 12bp between -69 and -57 resulted in a corresponding decrease in the absolute levels of expression, but did not effect the degree of stimulation relative to non-stimulated levels of expression (see Table 9). Placing this result in the context of the whole virus, it seems likely that Vmw175 acts to increase the levels of IE mRNAs for Vmw63, 68 and 12 throughout the replicative cycle, unless there is some counterbalancing effect from another viral product.

In purely qualitative terms, it is clear that the promoters of IE2, IE4/5 and IE3 are not identical. To obtain maximal, non-stimulated, expression from each promoter requires the presence of sequences extending to different positions with respect to the cap site. Thus, IE3 requires sequences to at least -331 (Cordingley et al, 1983; Herz & Roizman, 1983; Lang et al, 1984), IE4/5 shows maximal expression from only 69 bp of 5' flanking sequence (Preston et al, 1984; Table 9) and pIE2TK exhibits relatively poor expression despite the fact that it contains 5' flanking sequences extending far upstream (Table 8). From these data

it is likely that only IE3 has an upstream enhancer-like element, which drives a very poor promoter, whilst IE4/5 and IE2 have high and medium/low efficiency basal promoters respectively, but no enhancers. If these observations are correct, it is clearly unwise to assume that all IE promoters behave identically throughout the infectious cycle and, indeed, preliminary analyses indicate that there are different levels of mRNA detectable from individual IE promoters at different times post infection, including late times (F. Rixon, personal communication; Whitton, 1984).

Several workers have shown that the product of IE gene 1 (Vmw110) can trans-activate both IE and E promoters in transient expression systems (Everett, 1984; O'Hare & Hayward, 1985a; 1985b; Gelman & Silverstein, 1985). O'Hare and Hayward (1985a, b) have tested the IE3 promoter and claimed to have recreated "negative regulation of IE genes" by co-transfection of cloned sequences. I argue that one cannot treat all IE promoters alike and that it is likely that only IE3 is down-regulated by IE products. It will be of interest to see how the IE1 promoter responds to Vmw175 and how all IE promoters respond to Vmw110. The finding that autoregulation of IE3 was dominant over stimulation by Vmw110 (O'Hare & Hayward, 1985b) might be explained by postulating that Vmw175 also turns off transcription from the IE1 promoter.

It is clear, from the data in Table 8, that Vmw175 does not stimulate expression from its own promoter in the same way as it trans-activates IE2 and IE4/5. Indeed, there is some evidence from these data to suggest that Vmw175 represses IE3 activity slightly, a finding recently reported by others (O'Hare & Hayward, 1985b). Therefore, the IE3 promoter behaves in isolation in the way one would predict if the gene autoregulates its expression. Three of the many possible mechanisms which may operate to enable this autoregulation to occur are:

(a) Vmw175 might bind directly to specific negative, cis-acting, regulatory sequences as has been shown to occur in the case of SV40 large T-antigen.

(b) autoregulation might occur due to an abortive association of Vmw175 with transcription factors, freezing a potential stable complex in a non-functional state.

(c) the effect might be indirect in that Vmw175 can activate all other HSV-1 promoters, and possibly some cellular promoters, but not its own (due to a different transcription factor requirement or the presence of an enhancer, as discussed later). Thus, there would exist a competition between the trans-activated promoters and IE3 in which the latter would be at a distinct disadvantage. The end result would be an apparent down-regulation of IE3 expression.

The last model is the most attractive since it predicts that IE3 expression is not completely prevented but reaches an equilibrium based on the concentration of Vmw175. This is an important point for it is known that functional Vmw175 is required throughout the replicative cycle (Preston, 1979a, b; Dixon & Schaffer, 1980; Watson & Clements, 1980). Furthermore, it fits the observation that mutants with non-functional Vmw175 over produce IE polypeptides (Preston, 1979a, b; Dixon & Schaffer, 1980; DeLuca et al, 1984). O'Hare and Hayward (1985b) provide some evidence in support of such a model. Their experiments showed that the degree of autoregulation of IE3 by Vmw175 was dependent on the amount of plasmid introduced into the cells and, by implication, depended on the concentration of available Vmw175 protein.

The data in Table 10 suggest that the IE3 enhancer element acts to suppress stimulation of the TK promoter by Vmw175. Furthermore, the enhancer alone does not stimulate TK expression, in the absence of trans-activation. It has been shown that the SV40 enhancer differentially activates heterologous promoters, for example, it enhances human β-globin transcription about 100-fold (Treisman et al, 1983) but only weakly activates the promoters for Drosophila hsp70 (Pelham, 1982) and human α-globin (Treisman et al, 1983). It would appear that the IE3 enhancer exhibits similar properties in this respect and it may indicate a fundamental difference in the transcriptional requirements

of different promoters. Borelli et al (1984) have shown that the Ad2 E1A product (a general trans-activator) represses the enhancer functions of SV40, polyoma virus and the Ad2-E1A gene itself, and that the repression might act via the enhancer sequences themselves. A negative acting role for the E1A product has also been suggested in the transcriptional control of the Ad2 E2A promoter (Rossini, 1983; Guilfoyle et al, 1985). There is some evidence to suggest that the E1A gene itself is negatively regulated at late times. It has been shown that the level of E1A mRNA increases only modestly after the E to L shift in Ad2 infected cells (Spector et al, 1978; Shaw & Ziff, 1980), when compared with the increase in template copy number. Recently, Smith et al (1985) claim to have shown autoregulation by E1A in transient expression systems. They constructed a variety of mutated E1A genes on plasmid vectors, harbouring the SV40 origin of replication and able to replicate in COS 7 cells. One mutant, which encoded a severely truncated EIA polypeptide, appeared to over-produce E1A mRNA by about 30-fold in comparison to other plasmids. This was taken as evidence that the fully functional E1A protein acts to repress its own transcription, possibly via the E1A gene enhancer element. Stenberg and Stinski (1985) have suggested that autoregulation of the HCMV IE1 gene occurs in transient expression systems by performing almost identical experiments to Smith and co-workers. Thus, expression of a truncated IE product resulted in an increased transcription from the IE1 promoter compared to the wild-type case. In both these papers there must be some doubt as to the validity of the results due to the authors' use of replicating SV40 vector systems. It has been shown that the E1A product is able to repress transcription from the SV40 early promoter (Velcich & Ziff, 1984), which includes part of the viral origin of replication and such an effect cannot be ruled out as responsible for the apparent autoregulation of genes linked on the same plasmid. Clearly a similar phenomenon might be taking place in the case of

the HSV-1 IE3 gene. However, from the limited data in Table 10, it would appear that the entire repressive effect of Vmw175 on the IE3 promoter cannot be mediated via the enhancer element as some stimulation of TK expression is evident. It may be that the full autoregulation occurs by an interaction between Vmw175, the IE3 enhancer region and the promoter sequences. This hypothesis is consistent with the observation that the IE3 promoter alone (sequences to -108) cannot be trans-activated by Vmw 175 (data not shown).

CONCLUSIONS &

FUTURE PROSPECTS

Conclusions and Future Prospects

The two main genes studied in this work, those for Vmw65 and Vmw175, encode effectors involved in the global control of at least two gene sets in HSV-1. The most detailed work on such systems has been performed in prokaryotic organisms, but can give useful guidelines. Maas and Clark (1964) defined a unit, called the regulon, which consisted of two or more operons (transcription units) that shared a common regulator molecule. The transcription units in the regulon may be unlinked and also subject to seperate, alternative, controls. Such a general system description is obviously relevant to both IE and E gene control. The task of analysing a regulon has been described as falling into four stages (Gottesman & Neidhart, 1983):

(i) recognition of the genes and products comprising the regulon.

(ii) discovery of the regulatory element(s), that is, the genes and products.

(iii) elucidation of the molecular details of the controlling activities.

(iv) understanding the physiological significance of the coordinate regulation.

In the case of Vmw65, tasks (i) and (ii) have been accomplished. The genes recognised by this positive regulator molecule are the IE genes and some of their functions have been described. The identity of the regulator molecule as Vmw65 is clear, although the precise biochemical structure of the active protein may involve some host induced modification. Task (iii) is being actively pursued. Some of the details are already known, for example all IE genes share the TAATGARAT recognition sequence in their 5' flanking regions and presumably this acts, in some way, to determine the specificity of transcriptional activation. So far almost none of the potential mechanisms by which the Vmw65 molecule activates an IE promoter can be excluded, indeed the exciting possibility exists that a completely novel method of specific gene activation is being observed.

The effect of Vmw65 on cellular transcription is now being addressed. There is evidence that the protein is modulating IE transcription by use of an existing cellular control pathway as stimulation of chromosomally resident promoters has been detected (C.M.Preston and D.Latchman, personal communication). The isolation and identification of such genes may provide clues to the mechanism of IE gene activation. It will be interesting to see if such genes are identified by the TAATGARAT motif.

The physiological relevance (task (iv)) of coordinate stimulation of IE gene expression may be more easily approached than the detailed mechanism by which it is brought about. It is feasible to construct mutant Vmw65 genes, by established in vitro techniques, to determine their stimulatory potential by co-transfection studies and to transfer stimulation-minus mutants into the viral genome by standard recombination methods. Assuming that it is possible to disengage the structural functions of the protein from the activator role, one could produce a virus which lacked only the ability to stimulate its IE gene expression. The phenotype of such a virus would be of immense interest.

The functions of the other genes identified in BamHI F are completely unknown. Again, the powerful modern techniques of in vitro mutagenesis may allow the construction of viruses with specific defects in these genes. The predicted amino acid sequence of the proteins permits the synthesis of oligopeptides to specific regions of the predicted proteins and, in turn, these peptides can be used to obtain monospecific antibodies which often react with the whole, native, protein of interest. In these two ways the DNA sequence can provide the foundation for rigorous genetic and biochemical studies.

Much remains to be done on the problem of autoregulation by Vmw175. The experiments in this work and the published accounts of O'Hare and Hayward (1985b) suggest that the details of the phenomenon can be approached in isolation from the whole virus. Thus, co-transfection

experiments, coupled with the construction of mutations both in the coding sequences and 5' flanking region of the Vmw175 gene, may provide the answers. The role of the other IE polypeptides in transcription regulation is only now being addressed. It is probable that a complex interaction of Vmw175 with other viral proteins and host cell factors determines the function of this protein. In this case, the use of *in vitro* transcription systems and purified viral polypeptides is likely to be a fruitful direction for future studies.

LITERATURE CITED

Ackerman, S., Bunick, D., Zandomeni, R. & Weinmann, R. (1983). RNA polymerase II ternary transcription complexes generated *in vitro*. Nucleic Acids Res. 11, 6041-064.

Alford, C.A. and Britt, W.J. (1985). Cytomegalovirus., in "Virology", B.N. Fields (Ed.), Raven Press, N.Y..

Aloni, Y., Dhar, R., Laub, O., Horowitz, M. & Khoury, G. (1977). Novel mechanisms for RNA maturation: the leader sequences of simian virus 40 mRNA are not transcribed adjacent to the coding sequences. Proc. Natl. Acad. Sci., U.S.A. 74, 3686-690.

Alwine, J.C. & Khoury, G. (1980). Effect of a tsA mutation on simian virus 40 late gene expression: variations between host cell lines. J. Virol. 33, 920-25.

Anderson, K.P., Costa, R.H., Holland, L.E. and Wagner, E.K. (1980). Characterisation of herpes simplex virus type 1 RNA present in the absence of *de novo* protein synthesis. J. Virol. 34, 9-27.

Anderson, K.P., Frink, R.J., Devi, G.B., Gaylord, B.H., Costa, R.H. and Wagner, E.K. (1981). Detailed characterisation of the mRNA mapping in the HindIII fragment *k* region of the herpes simplex virus type 1 genome. J. Virol. 37, 1011-1027.

Ashburner, M. and Bonner, J.J. (1979). The induction of gene activity in *Drosophila* by heat-shock. Cell 17, 241-254.

Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Seguin, C., Tuffnell, P.S. and Barrell, B.G. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature 310, 207-211.

Brady, J., Radonovich, M., Vodkin, M., Natarajan, V., Thoren, M., Das, G., Janik, J. and Salzman, N.P. (1982). Site-specific base substitution and deletion mutations that enhance or supress transcription of the SV40 major late RNA. Cell 31, 625-633.

Breathnach, R. and Chambon, P. (1981). Organisation and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349-384.

Brinster, R.L., Chen, H.Y., Warren, R., Sarthy, A. and Palmiter, R.D. (1982). Regulation of metallothionein-thymidine kinase fusion plasmids injected into mouse eggs. Nature 296, 39-42.

Brown, D.D. (1984). The role of stable complexes that repress and activate eucaryotic genes. Cell 37, 359-365.

Brown, S.M., Ritchie, D.A. and Subak-Sharpe, J.H. (1973). Genetic studies with herpes simplex virus type 1. The isolation of temperature sensitive mutants, their arrangement into complementation groups and recombination analysis leading to a linkage map. J. Gen. Virol. 18, 329-346.

Bunick, D. Zandomeni, R., Ackerman, S. and Weinmann, R. (1982). Mechanism of RNA polymerase II-specific initiation of transcription in vitro: ATP requirement and uncapped run-off transcripts. Cell 29, 877-886.

Burke and Soll, D. (1985). Functional analysis of fractionated Drosophila Kc cell tRNA gene transcription components. J. Biol. Chem. 260, 816-823.

Busby, D.W.G., House, W. and McDonald, J.R. (1964), in "Virological techniques", Churchill, London.

Bogenhagen, D.F., Sakonju, S. and Brown, D.D. (1980). A control region in the center of the 5S RNA gene directs specific initiation of tanscription: II. The 3' border of the region. Cell 19, 27-35.

Bogenhagen, D.F., Wormington, W.M. and Brown, D.D. (1982). Stable transcription complexes of Xenopus 5S RNA genes: a means to maintain the differentiated state. Cell 28, 413-421.

Borelli, E., Hen, R. and Chambon, P. (1984). Adenovirus-2 E1A products repress enhancer-induced stimulation of transcription. Nature 312, 608-612.

Bos, J.L. and Ten Wolde-Kraamwinkel, H.C. (1983). The E1b promoter of Ad12 in mouse Ltk$^-$ cells is activated by adenovirus region E1a. EMBO J. 2, 73-76.

Bos, J.L. and Van der Eb, A.J. (1985). Adenovirus region E1A: transcription modulator and transforming agent. Trends in Biochem. Sci. 10, 310-313.

Boshart, M., Weber, F., Jahn, G., Dorsch-Hasler, K., Fleckenstein, B. and Schaffner, W. (1985). A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell 41, 521-530.

Brady, J., Bolen, J.B., Radonovich, M., Salzman, N. and Khoury, G. (1984a). Stimulation of simian virus 40 late gene expression by simian virus 40 tumor antigen. Proc. Natl. Acad. Sci., U.S.A. 81, 2040-2044.

Brady, J., Radonovich, M., Thoren, M., Das, G. and Salzman, N.P. (1984b). Simian virus 40 major late promoter: an upstream DNA sequence required for efficient in vitro transcription. Mol. Cell. Biol. 4, 133-141.

Berk, A.J. and Sharp, P.A. (1978b). The structure of adenovirus early mRNAs. Cell 14, 695-711.

Berk, A.J., Lee, F., Harrison, T., Williams, J. and Sharp, P.A. (1979). Pre-early adenovirus 5 gene product regulates synthesis of early viral messenger RNAs. Cell 17, 935-944.

Bieker, J.J., Martin, P.L. and Roeder, R.G. (1985). Formation of a rate-limiting intermediate in 5S RNA gene transcription. Cell 40, 119-127.

Bienz, M. (1985). Transient and developmental activation of heat-shock genes. Trends in Biochem. Sci. 10, 157-161.

Bienz, M. and Gurdon, J.B. (1982). The heat shock response in Xenopus oocytes is controlled at the translational level. Cell 29, 811-819.

Bienz, M. and Pelham, H.R.B. (1982). Expression of a Drosophila heat shock protein in Xenopus oocytes: conserved and divergent regulatory signals. EMBO J. 1, 1583-1588.

Biggin, M.D., Gibson, T.J. and Huang, G.F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid sequence determination. Proc. Natl. Acad. Sci., U.S.A. 80, 3963-3965.

Bishop, M. and Thompson, E. (1984). Fast computer search for similar DNA sequences. Nucleic Acids Res. 12, 5471-5474.

Blumenthal, R.M., Rice, P.J. and Roberts, R.J. (1982). Computer programs for nucleic acid sequence manipulation. Nucleic Acids Res. 10, 91-101.

Baker, C.C. and Ziff, E.B. (1981). Promoters and heterogeneous 5' termini of the messenger RNAs of adenovirus serotype 2. J. Mol. Biol. 149, 189-221.

Banerjee, A.K. (1980). 5' terminal cap structure in eukaryotic messenger ribonucleic acids. Microb. Revs. 44, 175-205.

Banerji, J., Olson, L. and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33, 729-740.

Banerji, J., Rusconi, S. & Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell 27, 299-308.

Banks, L., Purifoy, D.J.M., Hurst, P-F., Killington, R.A. and Powell, K.L. (1983). Herpes simplex virus non-structural proteins. IV. Purification of the virus-induced deoxyribonuclease and characterisation of the enzyme using monoclonal antibodies. J. Gen. Virol. 64, 2249-2260.

Batterson, W and Roizman, B. (1983). Characterisation of the herpes simplex virion-associated factor responsible for the induction of genes. J. Virol. 46, 371-377.

Batterson, W., Furlong, D. and Roizman, B. (1983). Molecular genetics of herpes simplex virus VIII. Further characterisation of a temperature-sensitive mutant, defective in release of viral DNA and in other stages of the viral reproductive cycle. J. Virol. 45, 397-407.

Baty, D., Barrera-Saldana, H.A., Everett, R.D., Vigneron, M. and Chambon, P. (1984). Mutational dissection of the 21 bp repeat region of the SV40 early promoter reveals that it contains overlapping elements of the early-early and late-early promoters. Nucleic Acids Res. 12, 915-932.

Becker, Y., Dym, H. and Sarov, I. (1968). Herpes simplex virus DNA. Virology 36, 184-192.

Ben-Porat, T. and Kaplan, A.S. (1962). The chemical composition of herpes simplex and pseudorabies viruses. Virology 16, 261-266.

Ben-Zeev, A. and Becker, Y. (1977). Requirement of host cell RNA polymerase II in the replication of herpes simplex virus in α-amanitin sensitive and resistant cell lines. Virology 76, 246-253.

Ben-Zeev, A., Asher, Y. and Becker, Y. (1976). Synthesis of herpes simplex virus-specified RNA by an RNA polymerase II in isolated nuclei in vitro. Virology 71, 302-311.

Benoist, C. and Chambon, P. (1980). Deletions covering the putative promoter region of early mRNAs of simian virus 40 do not abolish T antigen expression. Proc. Natl. Acad. Sci. U.S.A. 77, 3865-3869.

Benoist, C. and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. Nature 290, 304-310.

Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980). The ovalbumin gene sequence of putative control regions. Nucleic Acids Res. 8, 127-142.

Berg, P.E., Yu, J-K., Popovic, Z., Schumperli, D., Johansen, H., Rosenberg, M. and Anderson, W.F. (1983). Differential activation of the mouse β-globin promoter by enhancers. Mol. Cell. Biol. 3, 1246-1254.

Berk, A.J. and Sharp, P.A. (1978a). Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. 75, 1274-1278.

Busslinger, M., Portman, R., Irminger, J.C. and Birnstiel, M.L. (1980). Ubiquitous and gene-specific regulatory 5' sequences in a sea urchin histone DNA clone coding for histone protein variants. Nucleic Acids Res. 8, 957-977.

Bzik, D.J. and Preston, C.M. (1986). Analysis of DNA sequences which regulate the transcription of herpes simplex virus immediate early gene 3: DNA sequences required for enhancer-like activity and response to trans-activation by a virion polypeptide. Nucleic Acids Res. 14, 929-943.

Cabral, G.A., Courtney, R.J., Schaffer, P.A. and Marciano-Cabral, F. (1980). Ultrastructural characterisation of an early, non-structural polypeptide of herpes simplex virus. J. Virol. 33, 1192-1198.

Campbell, M.E., Palfreyman, J.W. and Preston, C.M. (1984). Identification of herpes simplex virus DNA sequences which encode a transacting polypeptide responsible for stimulation of immediate-early transcription. J. Mol. Biol. 180, 1-21.

Campo, M.S., Spandidos, D.A., Lang, J. and Wilkie, N.M. (1983). Transcriptional control signals in the genome of bovine papillomavirus type 1. Nature 303, 77-80.

Carlock, L.R. and Jones, N.C. (1981). Transformation-defective mutant of adenovirus type 5 containing a single altered E1a mRNA species. J. Virol. 40, 657-664.

Carroll, A.R. and Wagner, R.R. (1979). Role of the membrane (M) protein in endogenous inhibition of in vitro transcription by vesicular stomatitis virus. J. Virol. 29, 134-142.

Carter, A.D., Felber, B.K., Walling, M., Jubier, M-F., Schmidt, C.J. and Hamer, D.H. (1984). Duplicated heavy metal control sequences of the mouse metallothionein-I gene. Proc. Natl. Acad. Sci. U.S.A. 81, 7392-7396.

Cavalieri, R.L., Havell, E.A., Vileck, J. and Pestka, S. (1977). Induction and decay of human fibroblast interferon mRNA. Proc. Natl. Acad. Sci. U.S.A. 74, 4415-4419.

Celander, D. and Haseltine, W.A. (1984). Tissue-specific transcription preference as a determinant of cell tropsim and leukaemogenic potential of murine retroviruses. Nature 312, 159-162.

Chandler, V.L., Maler, B.A. and Yamamoto, K.R. (1983). DNA sequences bound specifically by glucocorticoid receptor *in vitro* render a heterologous promoter hormone responsive *in vivo*. Cell 33, 489-499.

Chartrand, P., Crumpacker, C.S., Schaffer, P.A. and Wilkie, N.M. (1980). Physical and genetic analysis of the herpes simplex virus DNA polymerase locus. Virology 103, 311-326.

Chartrand, P., Stow, N.D., Timbury, M.C. and Wilkie, N.M. (1979). Physical mapping of *paa*^{r} mutations of herpes simplex viruses types 1 and 2 by intertypic marker rescue. J. Virol. 31, 265-276

Chen, H.R. and Barker, W.C. (1985). The protein identification resource and its applications. Trends in Genet. 1, 221-223.

Ciliberto, G., Castagnoli, L., Melton, D.A. and Cortese, R. (1982a). Promoter of a eukaryotic tRNA^{pro} gene is composed of three non-contiguous regions. Proc. Natl. Acad. Sci. U.S.A. 79, 1195-1199.

Ciliberto, G., Traboni, C. and Cortese, R. (1982b). Relationship between the two components of the split promoter of eukaryotic tRNA genes. Proc. Natl. Acad. Sci. U.S.A. 79, 1921-1925.

Clements, J.B., McLauchlan, J. and McGeoch, D.J. (1979). Orientation of herpes simplex virus type 1 immediate-early mRNAs. Nucleic Acids Res. 7, 77-91.

Clements, J.B., Watson, R.J. and Wilkie, N.M. (1977). Temporal regulation of herpes simplex virus type 1 transcription: location of transcripts on the viral genome. Cell 12, 275-285.

Cochet, M., Chang, A.C.Y. and Cohen, S.N. (1982). Characterization of the structural gene and putative 5' regulatory sequences for human propiomelanocortin. Nature 297, 335-339.

Cohen, G.H. (1972). Ribonucleotide reductase activity of synchronised KB cells infected with herpes simplex virus. J. Virol. 9, 408-418.

Conrad, S. and Botchan, M. (1982). Isolation and characterisation of human DNA fragments with nucleotide sequence homologies with the simian virus 40 regulatory region. Mol. Cell. Biol. 2, 949-965.

Corces, V., Pellicer, A., Axel, R. and Meselson, M. (1981). Integration, transcription and control of a Drosophila heat shock gene in mouse cells. Proc. Natl. Acad. Sci. U.S.A. 78, 7038-7042.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980). Promoter sequences of eukaryotic protein-coding genes. Science 209, 1406-1414.

Cordingley, M.G. and Preston, C.M. (1981). Transcription and translation of the herpes simlplex virus type 1 thymidine kinase gene after microinjection into Xenopus laevis oocytes. J. Gen. Virol. 54, 409-414.

Cordingley, M.G., Campbell, M.E. and Preston, C.M. (1983). Functional analysis of a herpes simplex type 1 promoter: identification of far-upstream regulatory sequences. Nucleic Acids Res. 11, 2347-2365

Costa, R.H., Draper, K.G., Kelly, T.J. and Wagner, E.K. (1985). An unusual spliced herpes simplex virus type 1 transcript with sequence homology to Epstein-Barr virus DNA. J. Virol. 54, 317-328.

Costanzo, F., Campadelli-Fiume, G., Foa-Tomasi, L. and Cassai, E. (1977). Evidence that herpes simplex virus DNA is transcribed by cellular RNA polymerase B. J. Virol. 21, 996-1001.

Courtney, R.J. and Powell, K.L. (1975). Immunological and biochemical characterisation of polypeptides induced by herpes simplex virus type 1 and 2. In "Oncogenesis and Herpesviruse II", part I, pp 63-73. Eds. G. de Thé, M.A. Epstein and H. zur Hausen. IARC Scientific Publications, no 11, Lyon.

Cowan, K., Tegtmeyer, P. and Anthony, D.D. (1973). Relationship of replication and transcription of simian virus 40 DNA. Proc. Natl. Acad. Sci., U.S.A. 70, 1927-1930.

Culotta, V.C., Wildes, R.J. and Sollner-Webb, B. (1985). Eucaryotic transcription complexes are specifically associated in large sedimentable structures: rapid isolation of polymerase I, II and III transcription factors. Mol. Cell. Biol. 5, 1582-1590.

Dalrymple, M.A., McGeoch, D.J., Davison, A.J. and Preston, C.M. (1985). DNA sequence of the herpes simplex virus type 1 gene whose product is responsible for transcriptional activation of immediate early promoters. Nucleic Acids Res. 13, 7865-7879.

Das, G.C. and Salzman, N.P. (1985). Simian virus 40 early promoter mutations that affect promoter function and autoregulation by large T antigen. J. Mol. Biol. 182, 229-239.

Davison, A.J and Wilkie, N.M. (1983). Inversion of the two segments of the herpes simplex virus genome in intertypic recombinants. J. Gen. Virol. 64, 1-18.

Davison, A.J. (1984). Structure of the genome termini of varicella-zoster virus. J. Gen. Virol. 65, 1969-1977.

Davison, A.J. and Scott, J.E. (1984). DNA sequence of the major inverted repeat in the varicella-zoster virus genome. J. Gen. Virol. 66, 207-220.

Davison, A.J. and Wilkie, N.M. (1981). Nucleotide sequences of the joint between the L and S segments of herpes simplex virus types 1 and 2. J. Gen. Virol. 55, 315-331.

Davison, A.J., Marsden, H.S. and Wilkie, N.M. (1981). One functional copy of the long terminal repeat gene specifying the immediate-early polypeptide IE 110 suffices for a productive infection of human foetal lung cells by herpes simplex virus. J. Gen. Virol. 55, 179-191.

Davison, B.L., Egly, J-M., Mulvihill, E.R. and Chambon, P. (1983). Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature 301, 680-686.

Davison, M-J., Preston, V.G. and McGeoch, D.J. (1984). Determination of the sequence alteration in the DNA of the herpes simplex virus type 1 temperature-sensitive mutant tsK. J. Gen. Virol. 65, 859-863.

De The, G., Geser, A., Day, N.E., Tukei, P.M., Williams, E.H., Beri, D.P., Smith, P.G., Dean, A.G., Bornkamm, G.W., Fearino, P. and Henle, W. (1978). Epidemiological evidence for causal relationship between Epstein-Barr virus and Burkitt's lymphoma from Ugandan prospective study. Nature 274, 756-761.

De Villiers, J. and Schaffner, W. (1981). A small segment of polyoma virus DNA enhances the expression of a cloned β-globin gene over a distance of 1400 base pairs. Nucleic Acids Res. 9, 6251-6264.

De Villiers, J., Olson, J., Tyndall, C., and Schaffner, W. (1982). Transcriptional "enhancers" from SV40 and polyoma virus show a cell type preference. Nucleic Acids Res. 10, 7965-7976.

DeFranco, D., Sharp, S. and Söll, D. (1981). Identification of regulatory sequences contained in the 5' flanking region of Drosophila lysine tRNA genes. Journal of Biological Chemistry 256, 12424-12429.

Deininger, P.L. (1983). Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Anal. Biochem. 129, 216-223.

Delius, H. and Clements, J.B. (1976). A partial denaturation map of herpes simplex virus type 1 DNA: Evidence for inversions of the unique DNA regions. J. Gen. Virol. 33, 125-133.

DeLuca, N.A., Courtney, M.A. and Schaffer, P.A. (1984). Temperature-sensitive mutants in herpes simplex virus type 1 ICP4 permissive for early gene expression. J. Virol. 52, 767-776.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387-395.

Dhar, R., Subramanian, K.N., Pan, J. and Weissman, S.M. (1977). Structure of a large segment of the genome of simian virus 40 that does not encode known proteins. Proc. Natl. Acad. Sci. U.S.A. 74, 827-831.

Dierks, P., van Öoyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissmann, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell 32, 695-706.

Dierks, P., Van Öoyen, A., Mantei, N. and Weissman, C. (1981). DNA sequences preceding the rabbit β-globin gene are required for formation in mouse L cells of β-globin RNA with the correct 5' terminus. Proc. Natl. Acad. Sci. U.S.A. 78, 1411-1415.

Dimaio, D. and Nathans, D. (1982). Regulatory mutants of simian virus 40: effect of mutations at a T antigen binding site on DNA replication and expression of viral genes. J. Mol. Biol. 156, 531-548.

Dingerman, T., Burke, D.J., Sharp, S., Schaack, J. and Söll, D. (1982). The 5' flanking sequences of Drosophila $tRNA^{arg}$ genes control their in vitro transcription in a Drosophila cell extract. J. Biol. Chem. 257, 14738-14744.

Dixon, R.A.F. and Schaffer, P.A. (1980). Fine-structure mapping and functional analysis of temperature-sensitive mutants in the gene encoding the herpes simplex virus type 1 immediate early protein VP175. J. Virol. 36, 189-203.

Draper, K.G., Frink, R.J., Devi, G.B., Swain, M., Galloway, D. and Wagner, E.K. (1984). Herpes simplex virus types 1 and 2 homology in the region between 0.58 and 0.68 map units. J. Virol. 52, 615-623.

Dubois-Dalcq, M., Holmes, K.V. and Rentier, B. (1984). Assembly of enveloped RNA viruses. Springer-Verlag, Wien-New York.

Dudler, R. and Travers, A.A. (1984). Upstream elements necessary for optimal function of the hsp70 promoter in transformed flies. Cell 38, 391-398.

Durnam, D.M. and Palmiter, R.D. (1981). Transcriptional regulation of the mouse metallothionein-I gene by heavy metals. J. Biol. Chem. 256, 5712-5716.

Dutia, B.M. (1983). Ribonucleotide reductase induced by herpes simplex virus has a virus-specified constituent. J. Gen. Virol. 64, 513-521.

Dynan, W.S. and Tjian, R. (1983a). Isolation of transcription factors that discriminate between different promoters recognised by RNA polymerase II. Cell 32, 669-680.

Dynan, W.S. and Tjian, R. (1983b). The promoter specific transcription factor Spl binds to upstream sequences in the SV40 early promoter. Cell 35, 79-87.

Dynan, W.S. and Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature 316, 774-778.

Easton, A.J. and Clements, J.B. (1980). Temporal regulation of herpes simplex virus type 2 transcription and characterisation of virus immediate-early mRNAs. Nucleic Acids Res. 8, 2627-2645.

Eddy, B.E., Borman, G.S., Grubbs, G.E. and Young, R.D. (1962). Identification of the oncogenic substance in rhesus monkey kidney cell cultures as simian virus 40. Virology 17, 65-75.

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980). The structure and evolution of the β-globin gene family. Cell 21, 653-668.

Eisenberg, S.P., Coen, D.M. and McKnight, S.L. (1985). Promoter domains required for expression of plasmid-borne copies of the herpes simplex virus thymidine kinase gene in virus infected mouse fibroblasts and microinjected frog oocytes. Mol. Cell. Biol. 5, 1940-1947.

El Kareh, A., Murphy, A.J.M., Fichter, T., Efstratiadis, A. and Silverstein, S. (1985). "Transactivation" control signals in the promoter of the herpesvirus thymidine kinase gene. Proc. Natl. Acad. Sci., U.S.A. 82, 1002-1006.

Engelke, D.R., Ng, S-Y., Shastry, B.S. and Roeder, R.G. (1980). Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell 19, 717-728.

Ephrussi, A., Church, G.M., Tonegawa, S. and Gilbert, W. (1985). B lineage-specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science 227, 134-140.

Epstein, M.A. (1962). Observations on the fine structure of mature herpes simplex virus and on the composition of its nucleoid. J. Exp. Med. 115, 1-9.

Everett, R.D. (1983). DNA sequence elements required for regulated expression of the HSV-1 glycoprotein D gene lie within 83 bp of the RNA capsites. Nucleic Acids Res. 11, 6647-6666.

Everett, R.D. (1984). A detailed analysis of an HSV-1 early promoter: sequences involved in trans-activation by viral immediate-early gene products are not early-gene specific. Nucleic Acids Res. 12, 3037-3056.

Everett, R.D. and Dunlop, M. (1984). Trans-activation of plasmid-borne promoters by adenovirus and several herpes group viruses. Nucleic Acids Res. 12, 5969-5978.

Everett, R.D., Baty, D. and Chambon, P. (1983). The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter. Nucleic Acids Res. 11, 2447-2464.

Feldman, L.T., Imperiale, M.J. and Nevins, J.R. (1982). Activation of early adenovirus transcription by the herpes virus immediate-early gene: evidence for a common cellular control factor. Proc. Natl. Acad. Sci. U.S.A. 79, 4952-4956.

Fenwick, M.L. and Roizman, B. (1977). Regulation of herpesvirus macromolecular synthesis. VI. Synthesis and modification of viral polypeptides in enucleated cells. J. Virol. 22, 720-725.

Fenwick, M.L. and Walker, M.J. (1979). Phosphorylation of a ribosomal protein and of virus-specific proteins in cells infected with herpes simplex virus. J. Gen. Virol. 45, 397-405.

Ferguson, B., Jones, N. and Rosenberg, M. (1984). Adenovirus Ela gene product expressed at high levels in Escherichia coli is functional. Science 224, 1343-1346.

Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van der Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G. and Ysebert, M. (1978). Complete nucleotide sequence of SV40 DNA. Nature 273, 113-120.

Fire, A., Samuels, M. and Sharp, P.A. (1982). Interactions between RNA polymerase II, factors, and template leading to accurate transcription. J. Biol. Chem. 259, 2509-2516.

Fitzgerald, M. and Shenk, T. (1981). The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell 24, 251-260.

Flint, S.J. (1982). Expression of adenoviral genetic information in productively infected cells. Biochem. Biophys. Acta 651, 175-208.

Frame, M.C., Marsden, H.S. and Dutia, B.M. (1985). The ribonucleotide reductase induced by herpes simplex virus type 1 involves minimally a complex of two polypeptides (136K and 38K). J. Gen. Virol. 66, 1581-1587.

Fried, M., Griffiths, M., Davies, B., Bjursell, G., La Mantia, G. and Lania, L. (1983). Isolation of cellular DNA sequences that allow expression of adjacent genes. Proc. Natl. Acad. Sci. U.S.A. 80, 2117-2121.

Friedman, R.L., Manly, S.P., McMahon, M., Kerr, I.M. and Stark, G.R. (1984). Transcriptional and post-translational regulation of interferon-induced gene expression in human cells. Cell 38, 745-755.

Frink, R.J., Anderson, K.P. and Wagner, E.K. (1981). Herpes simplex virus type 1 HindIII fragment L encodes spliced and complementary mRNA species. J. Virol. 39, 559-572.

Fromm, M and Berg, P. (1982). Deletion mapping of DNA required for SV40 early promoter *in vivo.* J. Mol. Appl. Genet. 1, 457-481

Fromm, M. and Berg, P. (1983). Simian virus 40 early- and late-region promoter functions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations. Mol. Cell. Biol. 3, 991-999.

Gaffney, D., McLauchlan, J., Whitton, J.L. and Clements, J.B. (1985). A modular system for the assay of transcription regulatory signals: The sequence TAATGARAT is required for herpes simplex virus immediated early gene activation. Nucleic Acids Res. 13, 7847-7863.

Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981). Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature 294, 626-631.

Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979). Organisation and sequences at the 5' end of a cloned complete ovalbumin gene. Nature 278, 428-434.

Gaynor, R.B. and Berk, A.J. (1983). *Cis*-acting induction of adenovirus transcription. Cell 33, 683-693.

Gaynor, R.B., Hillman, D. and Berk, A.J. (1984). Adenovirus early region 1A protein activates transcription of a non-viral gene introduced into mammalian cells by infection or transfection. Proc. Natl. Acad. Sci. U.S.A. 81, 1193-1197.

Geisse, S., Scheidereit, C., Westphal, H.M., Hynes, N.E., Groner, B. and Beato, M. (1982). Glucocorticoid receptors recognise DNA sequences in and around murine mammary tumour virus DNA. EMBO J. 1, 1613-1619.

Gelman, I.H. and Silverstein, S. (1985). Identification of immediate early genes from herpes simplex virus that transactivate the virus thymidine kinase. Proc. Natl. Acad. Sci., U.S.A. 82, 5265-5269.

George, D.G., Barker, W.C. and Hunt, L.T. (1986). The protein identification resource (PIR). Nucleic Acids Res. 14, 11-15.

Ghosh, P. and Lebowitz, P. (1981). Simian virus 40 mRNAs contain multiple 5' termini upstream and downstream from a Hogness-Goldberg sequence: a shift in 5' termini during the lytic cycle is mediated by large T antigen. J. Virol. 40, 224-240.

Ghosh, P.K., Lebowitz, P., Frisque, R.J. and Gluzman, Y. (1981). Identification of a promoter component involved in positioning the 5' termini of the simian virus 40 early mRNAs. Proc. Natl. Acad. Sci., U.S.A. 78, 100-104.

Gibson, W. (1981). Structural and nonstructural protein of strain colburn cytomegalovirus. Virology 111, 516-537.

Gibson, W. and Roizman, B. (1974). Proteins specified by herpes simplex virus. X. Staining and radiolabelling properties of B capsid and virion proteins in polyacrylamide gels. J. Virol. 13, 155-165.

Gidoni, D., Dynan, W.S. and Tjian, R. (1984). Multiple specific contacts between a mammalian transcription factor and its cognate promoters. Nature 312, 409-413.

Gilbert, W. (1978). Why genes in pieces ?. Nature 271, 501.

Gillies, S.D., Morrison, S.L., Oi, V.T. and Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33, 717-728.

Godowski, P.J. and Knipe, D.M. (1985). Identification of a herpes simplex virus function that represses late gene expression from parental genomes. J. Virol. 55, 357-365.

Ginsberg, A.M., King, B.O. and Roeder, R.G. (1984). Xenopus 5S gene transcription factor, TFIIIA: characterisation of a cDNA clone and measurement of RNA levels throughout development. Cell 39, 479-489.

Gitschier, J., Wood, W.I., Goralka, T.M., Wion, K.L., Chen, E.Y., Eaton, D.H., Velgar, G.A., Capon, D.J. and Lawn, R.M. (1984). Characterisation of the human factor VIII gene. Nature 312, 326-330.

Gluzman, Y., Sambrook, J.F. & Frisque, R.J. (1980). Expression of early genes of origin-defective mutants of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. 77, 3898-3902.

Goding, C., Jalinot, P., Zajchowski, D., Boeuf, H. and Kedinger, C. (1985). Sequence-specific trans-activation of the adenovirus EIIa early promoter by the viral EIV transcription unit. EMBO J. 4, 1523-1528.

Godowski, P.J. and Knipe, D.M. (1983). Mutations in the major DNA-binding protein gene of herpes simplex virus type 1 result in increased levels of viral gene expression. J. Virol. 47, 478-486.

Goodbourn, S., Zinn, K. and Maniatis, T. (1985). Human β-interferon gene expression is regulated by an inducible enhancer element. Cell 41, 509-520.

Gottesman, S. and Neidhardt, F.C. (1983). Global control systems., in "Gene Function in Prokaryotes", Eds. Beckwith, J., Davies, J. and Gallant, J.A., Cold Spring Harbor Laboratory.

Graham, F.L. and Van der Eb, A.J. (1973a). A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52, 456-467.

Graham, F.L., Veldheisen, G. and Wilkie, N.M. (1973b). Infectious herpesvirus DNA. Nature New Biology 245, 265-266.

Grantham, R., Gautier, C., Mercier, R. and Pave, A. (1980). Codon catalog usage and the genome hypothesis. Nucleic Acids Res. 8, r49-r62.

Green, M.R., Treisman, R. and Maniatis, T. (1983). Transcriptional activation of cloned human β-globin genes by viral immediate-early gene products. Cell 35, 137-148.

Gregoriades, A. (1980). Interaction of influenza M protein with viral lipid and phosphatidyl-choline vesicles. J. Virol. 36, 470-479.

Gregory, S.P. and Butterworth, P.H.W. (1983). A comparison of the promoter strengths of two eukaryotic genes in vitro reveals a region of DNA that can influence the rate of transcription in cis over long distances. Nucleic Acids Res. 11, 5317-5326.

Grosschedl, R. and Birnsteil, M.L. (1980). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Natl. Acad. Sci., U.S.A. 77, 1432-1436.

Grosschedl, R. and Birnstiel, M.L. (1982). Delimitation of far upstream sequences required for maximal in vitro transcription of an H2A histone gene. Proc. Natl. Acad. Sci., U.S.A. 79, 297-301.

Grossman, A.D., Erickson, J.W. and Gross, C.A. (1984). The htpR gene product of E.coli is a sigma factor for heat-shock promoters. Cell 38, 383-390.

Grosveld, G.C., de Boer, E., Shewmaker, C.K. and Flavell, R.A. (1982). DNA sequences necessary for transcription of the rabbit β-globin gene *in vivo*. Nature 295, 120-126.

Grosveld, G.C., Shewmaker, C.K., Jat, P. and Flavell, R.A. (1981). Localization of DNA sequences necessary for transcription of the rabbit β-globin gene *in vitro*. Cell 25, 215-226.

Grummt, I. (1982). Nucleotide sequence requirements for specific initiation of transcription by RNA polymerase I. Proc. Natl. Acad. Sci. U.S.A. 79, 6908-6911.

Grummt, I., Roth, E. and Paule M.R. (1982). Ribosomal RNA transcription *in vitro* is species specific. Nature 296, 173-174.

Gruss, P., Dhar, R. and Khoury, G. (1981). Simian virus 40 tandem repeated sequences as an element of the early promoter. Proc. Natl. Acad. Sci. U.S.A. 78, 943-947.

Guarente, L., Lalonde, B., Gifford, P. and Alani, E. (1984). Distinctly regulated tandem upstream activation sites mediate catabolite repression of the CYC1 gene of *S. cerevisiae*. Cell 36, 503-511.

Guarente, L., Yocum, R.R. and Gifford, P. (1982). A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. Proc. Natl. Acad. Sci. U.S.A. 79, 7410-7414.

Guilfoyle, R.A., Osheroff, W.P. and Rossini, M. (1985). Two functions encoded by adenovirus early region 1A are responsible for the activation and repression of the DNA-binding protein gene. EMBO J. 4, 707-713.

Haar, L. and Marsden, H.S. (1981). Two dimensional gel analysis of HSV type 1 induced polypeptides and glycoprotein processing. J. Gen. Virol. 52, 77-92.

Haegeman, G. and Fiers, W. (1978). Characterisation of the 5'-terminal cap structures of late simian virus 40-specific mRNA. J. Virol. 25, 824-830.

Hager, L.J. and Palmiter, R.D. (1981). Transcriptional regulation of mouse liver metallothionein-I gene by glucocorticoids. Nature 291, 340-342.

Halbert, D.N., Cutt, J.R. and Shenk, J. (1985). Adenovirus early region 4 encodes functions required for efficient DNA replication, late gene expression, and host cell shut-off. J. Virol. 56, 250-257.

Hall, L.M., Draper, K.G., Frink, R.J., Costa, R.H. and Wagner, E.K. (1982). Herpes simplex virus mRNA species mapping in EcoRI fragment I. J. Virol. 43, 594-607.

Halliburton, I.W. (1980). Intertypic recombinants of herpes simplex viruses. J. Gen. Virol. 48, 1-23.

Hamm, G.H. and Cameron, G.N. (1986). The EMBL data library. Nucleic Acids Res. 14, 5-9.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166, 557-580.

Hanas, J.S., Bogenhagen. D.F. and Wu, C-W. (1983a). Cooperative model for the binding of Xenopus transcription factor A to the 5S RNA gene. Proc. Natl. Acad. Sci. U.S.A. 80, 2142-2145.

Hanas, J.S., Bogenhagen. D.F. and Wu, C-W. (1984). DNA unwinding ability of Xenopus transcription factor A. Nucleic Acids Res. 12, 1265-1276.

Hanas, J.S., Hazuda, D.J., Bogenhagen. D.F., Wu, F.Y-H. and Wu, C-W. (1983b). Xenopus transcription factor A requires zinc for binding to the 5S RNA gene. J. Biol. Chem. 258, 14120-14125.

Handa, H., Kingston, R.E. and Sharp, P.A. (1983). Inhibition of adenovirus early region IV transcription in vitro by a purified viral DNA binding protein. Nature 302, 545-547.

Hansen, U., Tenen, D.G., Livingston, D.M. and Sharp, P.A. (1981). T antigen repression of SV40 early transcription from two promoters. Cell 27, 603-612.

Harris-Hamilton, E. and Bachenheimer, S.L. (1985). Accumulation of herpes simplex virus type 1 RNAs of different kinetic classes in the cytoplasm of infected cells. J. Virol. 53, 144-151.

Hartzell, S.W., Byrne, B.J. and Subramanian, K.N. (1984a). Mapping of the late promoter of simian virus 40. Proc. Natl. Acad. Sci., U.S.A. 81, 23-27.

Hartzell, S.W., Byrne, B.J. and Subramanian, K.N. (1984b). The simian virus 40 minimal origin and the 72 base pair repeat are required simultaneously for efficient induction late gene expression with large tumor antigen. Proc. Natl. Acad. Sci. U.S.A. 81, 6335-6339.

Hay, N. and Aloni, Y. (1985). Attenuation of late simian virus 40 mRNA synthesis is enhanced by the agnoprotein and is temporally regulated in isolated nuclear systems. Mol. Cell. Biol. 5, 1327-1334.

Hay, N., Skolnik-David, H. and Aloni, Y. (1982). Attenuation in the control of SV40 gene expression. Cell 29, 183-193.

Hay, R.T. and Hay, J. (1980). Properties of herpesvirus-induced "immediate-early" polypeptides. Virology 104, 230-234.

Hay, R.T. and Hay, J. (1981). Characteristics of chromatin preparations from herpes simplex virus infected cells. Biochem. Biophys. Acta 655, 71-81.

Hayward, G.S., Jacob, R.J., Wadsworth, S.C. and Roizman, B. (1975). Anatomy of herpes simplex virus DNA: evidence for four populations of molecules that differ in the relative orientations of their long and short components. Proc. Natl. Acad. Sci. 72, 4243-4247.

Hearing, P. and Shenk, T. (1983). The adenovirus type 5 E1A transcriptional control region contains a duplicated enhancer element. Cell 33, 695-703.

Hearing, P. and Shenk, T. (1985). Sequence-independent autoregulation of the adenovirus type 5 E1A transcription unit. Mol. Cell. Biol. 5, 3214-3221.

Hen, R., Borelli, E., Sassone-Corsi, P., and Chambon, P. (1983). An enhancer element is located 340 base pairs upstream from the adenovirus-2 E1A capsite. Nucleic Acids Res. 11, 8747-8760.

Hen, R., Sassone-Corsi, P., Corden, J., Gaub, M.P. and Chambon, P. (1982). Sequences upstream from the T-A-T-A box are required *in vivo* and *in vitro* for efficient transcription from the adenovirus serotype 2 major late promoter. Proc. Natl. Acad. Sci., U.S.A. 79, 7132-7136.

Herbommel, P., Bourachot, B. and Yaniv, M. (1984). Two distinct enhancers with different cell specificities co-exist in the regulatory region of polyoma. Cell 39, 653-662.

Herz, C. and Roizman, B. (1983). The α-promoter regulator-ovalbumin chimeric gene resident in human cells is regulated like an authentic α4 gene after infection with herpes simplex virus 1 mutants in α4 gene. Cell 33, 145-151.

Hill, T.J. (1985). Herpes simplex virus latency., in "The Herpesviruses", vol.3, pp175-240., B. Roizman (Ed.), Plenum Press.

Hiscott, J., Cantell, K. and Weismann, C. (1984a). Differential expression of human interferon genes. Nucleic Acids Res. 12, 3727-3746.

Hiscott, J., Ryals, J., Dierks, P., Hofmann, V. and Weismann, C. (1984b). The expression of human interferon alpha genes. Phil. Trans. R. Soc. Lond. B 307, 217-226.

Hofstetter, H., Kressmann, A. and Birnsteil, M.L. (1981). A split promoter for a eukaryotic tRNA gene. Cell 24, 573-585.

Honess, R.W. and Roizman, B. (1974). Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. J. Virol. 14, 8-19.

Honess, R.W. and Roizman, B. (1975). Regulation of herpesvirus macromolecular synthesis: sequential transition of polypeptide synthesis requires functional viral polypeptides. Proc. Natl. Acad. Sci. U.S.A. 72, 1276-1280.

Hopp, T.P. and Woods, K.R. (1981). Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci., U.S.A. 78, 3824-3828.

Horikoshi, M., Sekimizu, K. and Natori, S. (1984). Analysis of the stimulatory factor of RNA polymerase II in the initiation and elongation complex. J. Biol. Chem. 259, 608-611.

Hu, S-L. and Manley, J.L. (1981). DNA sequence required for initiation of transcription in vitro from the major late promoter of adenovirus 2. Proc. Natl. Acad. Sci., U.S.A. 78, 820-824.

Hynes, N.E., Kennedy, N., Rahmsdorf, U. and Groner, B. (1981). Hormone-responsive expression of an endogenous proviral gene of mouse mammary tumour virus after molecular cloning and gene transfer into cultured cells. Proc. Natl. Acad. Sci. U.S.A. 78, 2038-2042.

Ihara, S., Feldman, L., Watanabe, S. and Ben-Porat, T. (1983). Characterisation of the immediate early functions of pseudorabies virus. Virology 131, 437-454.

Imperiale, M.J. and Nevins, J.R. (1984). Adenovirus 5 E2 transcription unit: an E1A-inducible promoter with an essential element that functions independently of position or orientation. Mol. Cell. Biol. 4, 875-882.

Imperiale, M.J., Feldman, L.T. and Nevins, J.R. (1983). Activation of gene expression by adenovirus and herpesvirus regulatory genes acting in trans and by a cis-acting adenovirus enhancer element. Cell 35, 127-136.

Jacob, R.J., Morse, L.S. and Roizman, B. (1979). Anatomy of herpes simplex virus DNA. XII. Accumulation of head-to-tail concatamers in nuclei of infected cells and their role in the generation of the four isomeric arrangements of viral DNA. J. Virol. 29, 448-457.

Jamieson, A.T. and Subak-Sharpe, J.H. (1974). Biochemical studies on the herpes simplex virus specific deoxypyrimidine kinase activity. J. gen. Virol. 24, 481-492.

Jay, G., Nomura, S., Anderson, C.W. and Khoury, G. (1981). Identification of the SV40 agnoprotein product: a DNA binding protein. Nature 291, 346-349.

Johansson, K., Persson, H., Lewis, A.M., Pettersson, U., Tibbetts, C. and Phillipson, L. (1978). Viral DNA sequences and gene products in hamster cells transformed by adenovirus type 2. J. Virol. 27, 628-639.

Johnston, S.A. and Hopper, J.E. (1982). Isolation of the yeast regulatory gene GAL4 and analysis of its dosage effects on the galactose/melibiose regulon. Proc. Natl. Acad. Sci., U.S.A. 79, 6971-6975.

Jolly, D.J., Esty, A.C., Subramani, S., Friedmann, T. and Verma, I.M. (1983). Elements in the long terminal repeat of murine retroviruses enhance stable transformation by thymidine kinase gene. Nucleic Acids Res. 11, 1855-1872.

Jones, K.A. and Tjian, R. (1984). Essential contact residues within SV40 large T antigen binding sites I and II identified by alkylation-interference. Cell 36, 155-162.

Jones, K.A. and Tjian, R. (1985). Spl binds to promoter sequences and activates herpes simplex virus "immediate-early" gene transcription in vitro. Nature 317, 179-182.

Jones, K.A., Yamamoto, K.R. and Tjian, R. (1985). Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell 42, 559-572.

Jones, N. and Shenk, T. (1979a). An adenovirus type 5 early gene function regulates expression of other early viral genes. Proc. Natl. Acad. Sci. U.S.A. 76, 3665-3669.

Jones, N. and Shenk, T. (1979b). Isolation of adenovirus type 5 host range deletion mutants defective for transformation of rat embryo cells. Cell 17, 683-689.

Jones, P.C. and Roizman, B. (1979). Regulation of herpesvirus macromolecular synthesis. VIII. The transcription program consists of three phases during which both extent of transcription and accumulation of RNA in the cytoplasm are regulated. J. Virol. 31, 299-314.

Jongstra, J., Reudelhuber, T.L., Oudet, P., Benoist, C., Chae, C-B., Jeltsch, J-M., Mathis, D.J. and Chambon, P. (1984). Induction of altered chromatin structures by simian virus 40 enhancer and promoter elements. Nature 306, 708-714.

Kao, H-T. and Nevins, J.R. (1983). Transcriptional activation and subsequent control of the human heat shock gene during adenovirus infection. Mol. Cell. Biol. 3, 2058-2065.

Karin, M. (1985). Metallothioneins: proteins in search of function. Cell 41, 9-10.

Karin, M. and Herschman, H.R. (1979). Dexamethasone stimulation of metallothionein synthesis in HeLa cell cultures. Science 204, 176-177.

Karin, M., Andersen, R.D., Slater, E., Smith, K. and Herschman, H.R. (1980). Metallothionein mRNA induction in HeLa cells in response to zinc or dexamethasone is a primary induction response. Nature 286, 295-297.

Karin, M., Haslinger, A., Holtgreve, H., Cathala, G., Slater, E. and Baxter, J.D. (1984a). Activation of a heterologous promoter in response to dexamethasone and cadmium by metallothionein gene 5' flanking DNA. Cell 36, 371-379.

Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krautner, P., Westphal, H.M. and Beato, M. (1984b). Characterisation of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-II_A gene. Nature 308, 513-519.

Keir, H.M. and Gold, E. (1963). Deoxyribonucleic acid nucleotidyltransferase and deoxribonuclease from cultured cells infected with herpes simplex virus. Biochim. Biophys. Acta 72, 263-276.

Keir, H.M., Subak-Sharpe, H., Shedden, W.I.H., Watson, D.H. and Wildy, P. (1966). Immunological evidence for a specific DNA polymerase produced after infection by herpes simplex virus. Virology 30, 154-157.

Keller, J.M. and Alwine, J.C. (1984). Activation of the SV40 late promoter: direct effects of T antigen in the absence of viral DNA replication. Cell 36, 381-389.

Keller, J.M. and Alwine, J.C. (1985). Analysis of an activatible promoter: sequences in the simian virus 40 late promoter required for T-antigen-mediated trans-activation. Mol. Cell. Biol. 5, 1859-1869.

Khoury, G. and May, E. (1977). Regulation of early and late simian virus 40 transcription: overproduction of early viral RNA in the absence of T antigen. J. Virol. 23, 167-176.

Kieff, E.D., Bachenheimer, S.L. and Roizman, B. (1971). Size, composition and structure of the deoxyribonucleic acid of herpes simplex virus subtypes 1 and 2. J. Virol. 8, 125-132.

Kit, S. and Dubbs, D.R. (1963). Acquisition of thymidine kinase activity by herpes simplex virus infected mouse fibroblast cells. Biochem. Biophys. Acta 11, 55-59.

Kit, S. and Dubbs, D.R. (1977). Regulation of herpesvirus thymidine kinase activity in LMTK$^-$ cells transformed by ultra violet light irradiated herpes simplex virus. Virology 76, 331-340.

Klemperer, H.G., Haynes, G.R., Shedden, W.I.H. and Watson, D.H. (1967). A virus-specific thymidine kinase in BHK 21 cells infected with herpes simplex virus. Virology 31, 120-128.

Kneale, G.G. and Bishop, M.J. (1985). Nucleic acid and protein sequence databases. Comp. Appl. Biosci. 1, 11-17.

Kohorn, B.D. and Rae, P.M.M. (1982). Non-transcribed spacer sequences promote *in vitro* transcription of *Drosophila* ribosomal DNA. Nucleic Acids Res. 10, 6879-6886.

Konarska, M.M., Padgett, R.A. and Sharp, P.A. (1984). Recognition of cap structure in splicing *in vitro* of mRNA precursors. Cell 38, 731-736.

Kozak, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. Microb. Revs. 47, 1-45.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. 12, 857-872.

Kozak, M. and Roizman, B. (1974). Regulation of herpesvirus macromolecular synthesis: nuclear retention of nontranslated viral RNA sequences. Proc. Natl. Acad. Sci. U.S.A. 71, 4322-4326.

Kriegler, M. and Botchan, M. (1983). Enhanced transformation by a simian virus 40 recombinant virus containing a Harvey murine sarcoma virus long terminal repeat. Mol. Cell. Biol. 3, 325-339.

Krippl, B., Ferguson, B., Rosenberg, M. and Westphal, H. (1984). Functions of purified E1A protein microinjected into mammalian cells. Proc. Natl. Acad. Sci., U.S.A. 81, 6988-6992.

Kristie, T. and Roizman, B. (1984). Separation of sequences defining basal expression from those confering α gene recognition within the regulatory domains of herpes simplex virus 1 α genes. Proc. Natl. Acad. Sci. U.S.A. 81, 4065-4069.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132.

Laimins, L.A., Khoury, G., Gorman, C., Howard, B. and Gruss, P. (1982). Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci., U.S.A. 79, 6453-6457.

Lang, J.C., Spandidos, D.A. and Wilkie, N.M. (1984). Transcriptional regulation of a herpes virus immediate-early gene is mediated through an enhancer-type sequence. EMBO J. 3, 389-395.

Larson, D., Bradford-Wilcox, J., Young, L.S. and Sprague, K.U. (1983). A short 5' flanking region containing conserved sequence is required for silkworm alanine tRNA gene activity. Proc. Natl. Acad. Sci. U.S.A. 80, 3416-3420.

Lassar, A.B., Martin, P.L. and Roeder, R.G. (1983). Transcription of class III genes: formation of pre-initiation complexes. Science 222, 740-748.

Laughon, A. and Gesteland, R.F. (1982). Isolation and preliminary characterisation of the GAL4 gene, a positive regulator of transcription. Proc. Natl. Acad. Sci., U.S.A. 79, 6827-6831.

Learned, R.M., Smale, S.T., Haltiner, M.M. and Tjian, R. (1983). Regulation of human ribosomal RNA transcription. Proc. Natl. Acad. Sci., U.S.A. 80, 3558-3562.

Lebowitz, P. and Ghosh, P.K. (1982). Initiation and regulation of simian virus 40 early transcription in vitro. J. Virol. 41, 449-461.

Leff, T., Elkaim, R., Goding, C.R., Jalinot, P., Sassone-corsi, P., Perricaudet, M., Kedinger, C. and Chambon, P. (1984). Individual products of the adenovirus 12S and 13S EIa mRNAs stimulate viral EIIa and EIII expression at the transcriptional level. Proc. Natl. Acad. Sci., U.S.A. 81, 4381-4385.

Lemaster, S. and Roizman, B. (1980). Herpes simplex virus phosphoproteins: II. Characterisation of the virion protein kinase and of the polypeptides phosphorylated in the virion. J. Virol. 35, 798-811.

Lengyel, P. (1982). Biochemistry of interferons and their actions. Ann. Rev. Biochem. 51, 251-282.

Leung, W-C., Dimock, K., Smiley, J. and Bachetti, S. (1980). Herpes simplex virus thymidine kinase transcripts are absent from both nucleus and cytoplasm during infection in the presence of cycloheximide. J. Virol. 36, 361-365.

Levinson, B., Khoury, G., Vande Woude, G. and Gruss, P. (1982). Activation of the SV40 genome by the 72 base pair tandem repeats of Moloney sarcoma virus. Nature 295, 568-572.

Lewin, B. (1980). "Gene expression 2: Eucaryotic chromosomes", 2nd Ed., John Wiley & Sons, Inc..

Lewis, M.K. and Burgess, R.R. (1982). Eukaryotic RNA polymerases., in "The Enzymes: Nucleic Acids part B", vol. XV, P.D. Boyer (Ed.).

Lis, J.T., Simon, J.A. and Sutton, C.A. (1983). New heat shock puffs and β-galactosidase activity resulting from transformation of Drosophila with an hsp70-lacZ hybrid gene. Cell 35, 403-410.

Long, E.O. and Dawid, I.B. (1980). Repeated genes in eukaryotes. Ann. Rev. Biochem. 49, 727-764.

Losick, R. and Pero, J. (1981). Cascades of sigma factors. Cell 25, 582-584.

Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951). Protein measurement with the Folin phenol reagent. J. Biol. Chem. 193, 265-275.

Luciw, P.A., Bishop, J.M., Varmus, H.E. and Capecchi, M.R. (1983). Location and function of retroviral and SV40 sequences that enhance biochemical transformation after microinjection of DNA. Cell 33, 705-716.

Maas, W.K. and Clark, A.J. (1964). Studies on the mechanism of repression of arginine biosynthesis in Escherichia coli. II. Dominance of repressibility in diploids. J. Mol. Biol. 8, 365-370.

Mackem, S. and Roizman, B. (1982a). Differentiation between α-promoter and regulator regions of herpes simplex virus 1: the functional domains and sequence of a movable regulator. Proc. Natl. Acad. Sci. U.S.A. 79, 4917-4921.

Mackem, S. and Roizman, B. (1982b). Structural features of the herpes simplex virus α gene 4, 0, and 27 promoter-regulatory sequences which confer α regulation on chimeric thymidine kinase genes. J. Virol. 44, 939-949.

Mackem, S. and Roizman, B. (1982c). Regulation of α genes of herpes simplex virus: the α27 gene promoter-thymidine kinase chimera is positively regulated in converted L cells. J. Virol. 43, 1015-1023.

MacPherson, I and Stoker, M. (1962). Polyoma transformation of hamster cell clones-an investigation of genetic factors affecting cell competence Virology 16, 147-151

Maitland, N.J. and McDougall, J.K. (1977). Biochemical transformation of mouse cells by fragments of herpes simplex virus DNA. Cell 11, 233-241.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory.

Maniatis, T., Jeffrey, A. and van de Sande, H. (1975). Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. Biochem. 14, 3787-3794.

Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Gefter, M.L. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci., U.S.A. 77, 3855-3859.

Maroteaux, L., Chen, L., Mitrani-Rosenbaum, S., Howley, P.M. and Revel, M. (1983a). Cycloheximide induces expression of the human interferon β_1 gene in mouse cells transformed by bovine papilloma virus-interferon β_1 recombinants. J. Virol. 47, 89-95.

Maroteaux, L., Kahana, C., Mory, Y., Revel, M. and Howley, P.M. (1983b). Sequences involved in the regulated expression of the human interferon β_{-1} gene in recombinant SV40 DNA vectors replicating in monkey cells. EMBO J. 2, 325-332.

Marsden, H.S., Crombie, I.K. and Subak-Sharpe, J.H. (1976). Control of protein synthesis in herpes virus-infected cells: Analysis of the polypeptides induced by wild type and sixteen temperature-sensitive mutants. J. Gen. Virol. 31, 347-372.

Marsden, H.S., Haarr, L. and Preston, C.M. (1983). Processing of herpes simplex virus proteins and evidence that translation of thymidine kinase mRNA is initiated at three seperate AUG codons. J. Virol. 46, 434-445.

Marsden, H.S., Lang, J., Davison, A.J., Hope, R.G. and McDonald, D.M. (1982). Genomic location and lack of phosphorylation of the HSV immediate-early polypeptide IE12. J. Gen. Virol. 62, 17-27.

Marsden, H.S., Stow, N.D., Preston, V.G., Timbury, M.C. and Wilkie, N.M. (1978). Physical mapping of herpes simplex virus-induced polypeptides. J. Virol. 28, 624-642.

Mathis, D.J. and Chambon, P. (1981). The SV40 early region TATA box is required for accurate *in vitro* initiation of transcription. Nature 290, 310-315.

Matsui, T., Segall, J., Weil, P.A. and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. 255, 11992-11996.

Maxam, A. and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65, 499-559.

Mattaj, I.W., Lienhard, S., Jiricny, J. and De Robertis, E.M. (1985). An enhancer-like sequence within the Xenopus U2 gene promoter facilitates the formation of stable transcription complexes. Nature 316, 163-167.

Maul, G.G. and Hamilton, T.H. (1967). The intranuclear localisation of two DNA-dependent RNA polymerase activities. Proc. Natl. Acad. Sci., U.S.A. 57, 1371-1378.

McGeoch, D.J. (1984). The nature of animal virus genetic material., in "The Microbe 1984: part 1 Viruses", Eds. Mahy, B.W.J. and Pattison, J.R., Cambridge University Press.

McGeoch, D.J., Dolan, A., Donald, S. and Rixon, F.J. (1985). Sequence determination and genetic content of the short unique region in the genome of herpes simplex virus type 1. J. Mol. Biol. 181, 1-13.

McKay, R. and DiMaio, D. (1981). Binding of an SV40 T antigen-related protein to the DNA of SV40 regulatory mutants. Nature 289, 810-813.

McKnight, S.L. (1980). The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucleic Acids Res. 8, 5949-5964.

McKnight, S.L. (1982). Functional relationships between transcriptional control signals of the thymidine kinase gene of herpes simplex virus. Cell 31, 355-365.

McKnight, S.L. and Gavis, E.R. (1980). Expression of the herpes thymidine kinase gene in Xenopus laevis oocytes: an assay for the study of deletion mutants constructed in vitro. Nucleic Acids Res. 8, 5931-5948.

McKnight, S.L. and Kingsbury, R. (1982). Transcription control signals of a eukaryotic protein coding gene. Science 217, 316-324.

McKnight, S.L., Gavis, E.R. and Kingsbury, R. (1981). Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell 25, 385-398.

McKnight, S.L., Kingsbury, R.C., Spence, A. and Smith, M. (1984). The distal transcription signals of the herpesvirus TK gene share a common hexanucleotide control sequence. Cell 37, 253-262.

McLauchlan, J., Gaffney, D., Whitton, J.L. and Clements, J.B. (1985). The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. Nucleic Acids Res. 13, 1347-1368.

McLean, C., Buckmaster, A., Hancock, D., Buchan, A., Fuller, A. and Minson, A. (1982). Monoclonal antibodies to three non-glycosylated antigens of herpes simplex virus type 2. J. Gen. Virol. 63, 297-305.

Mellon, P., Parker, V., Gluzman, Y., Maniatis, T. (1981). Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 host-vector system. Cell 27, 279-288.

Melton, D.W., Konecki, D.S., Brennand, J. and Caskey, C.T. (1984). Structure, expression and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci., U.S.A. 81, 2147-2151.

Mercola, M., Goverman, J., Mirell, C. and Calame, K. (1985). Immunoglobulin heavy chain enhancer requires one or more tissue-specific factors. Science 227, 266-270.

Messing, J. and Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of a double digest restriction fragment. Gene 19, 269-276.

Messing, J., Crea, R. and Seeburg, P.H. (1981). A system for shotgun DNA sequencing. Nucleic Acids Res. 9, 309-321.

Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977). Filamentous coliphage M13 as a cloning vehicle: insertion of a Hind III fragment of the lac regulatory region in M13 replicative form in vitro. Proc. Natl. Acad. Sci., U.S.A. 74, 3642-3646.

Metzler, D.W. and Wilcox, K.W. (1985). Isolation of herpes simplex virus regulatory protein ICP4 as a homodimeric complex. J. Virol. 55, 329-337.

Miesfeld, R. and Arnheim, N. (1984). Species-specific rDNA transcription is due to promoter-specific binding factors. Mol. Cell. Biol. 4, 221-227.

Miller, J., McLachlan, A.D. and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. 4, 1609-1614.

Minson, A.C., Wildy, P., Buchan, A. and Darby, G. (1978). Introduction of the herpes simplex virus thymidine kinase gene into mouse cells using virus DNA or transformed cell DNA. Cell 13, 581-587.

Mirault, M-E., Southgate, R. and Delwart, E. (1982). Regulation of heat shock genes: a DNA sequence upstream of Drosophila hsp70 genes is essential for their induction in monkey cells. EMBO J. 1, 1279-1285.

Mishima, Y., Financsek, I., Kominami, R. and Muramatsu, M. (1982). Fractionation and reconstitution of factors required for accurate transcription initiation of mammalian ribosomal RNA genes: identification of a species-dependent initiation factor. Nucleic Acids Res. 10, 6659-6670.

Mishoe, H., Brady, J.N., Radonovich, M. and Salzman, N.P. (1984). Simian virus 40 guanine-cytosine rich sequences function as independent transcriptional control elements in vitro. Mol. Cell. Biol. 4, 2911-2920.

Mocarski, E.S. and Roizman, B. (1981). Site specific inversion of the herpes simplex virus genome: domain and structural features. Proc. Natl. Acad. Sci. U.S.A. 78, 7047-7051.

Mocarski, E.S. and Roizman, B. (1982). Structure and role of the herpes simplex virus DNA termini in inversion, circularisation and generation of virion DNA. Cell 31, 89-97.

Mocarski, E.S., Post, L.E. and Roizman, B. (1980). Molecular engineering of the herpes simplex virus genome: insertion of a second L-S junction into the genome causes additional genome inversions. Cell 22, 243-255.

Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1981). Resolving the functions of overlapping viral genes by site specific mutagenesis at a mRNA splice site. Nature 295, 380-384.

Moore, C.L. and Sharp, P.A. (1984). Site-specific polyadenylation in a cell-free reaction. Cell 36, 581-591.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P. (1981). The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. 9, 6047-6068.

Morse, L.S., Pereira, L., Roizman, B. and Schaffer, P.A. (1978). Anatomy of herpes simplex virus (HSV) DNA X. Mapping of viral genes by analysis of polypeptides and functions specified by HSV-1 x HSV-2 recombinants. J. Virol. 26, 389-410.

Moss, H., Chartrand, P., Timbury, M.C. and Hay, J. (1979). Mutants of herpes simplex virus type 2 with temperature sensitive lesions affecting virion thermostability and DNase activity: Identification of the lethal mutation and physical mapping of the nuc$^-$ lesion. J. Virol. 32, 140-146.

Moss, T. (1982). Transcription of cloned Xenopus laevis ribosomal DNA microinjected into Xenopus oocytes, and the identification of an RNA polymerase I promoter. Cell 30, 835-842.

Mulvihill, E.R., LePennec, J-P. and Chambon, P. (1982). Chicken oviduct progesterone receptor: location of specific regions of high-affinity binding in cloned DNA fragments of hormone-responsive genes. Cell 24, 621-632.

Mulvihill, E.R., LePennec, J-P. and Chambon, P. (1982). Chicken oviduct progesterone receptor: location of specific regions of high affinity binding in cloned DNA fragments of hormone-responsive genes. Cell 24, 621-632.

Murthy, S.C.S., Bhat, G.P. and Thimmappaya, B. (1985). Adenovirus EIIA early promoter: transcriptional control elements and induction by the viral pre-early EIA gene, which appears to be sequence independent. Proc. Natl. Acad. Sci., U.S.A. 82, 2230-2234.

Myers, R.M., Rio, D.C., Robbins, A.K. and Tjian, R. (1981). SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. Cell 25, 373-384.

Nahmias, A.J. and Dowdle, W.R. (1968). Antigenic and biological differences in herpesvirus hominis. Prog. Med. Virol. 10, 110-159.

Nahmias, A.J. and Roizman, B. (1973). Infection with herpes simplex viruses 1 and 2. N. Eng. J. Med. 289, 667-675.

Neiderman, J., Miller, G., Pearson, H. and Pagano, J. (1976). Patterns of excretion of Epstein-Barr virus in saliva and other oropharyngeal sites during infectious mononucleosis. N. Eng. J. Med. 294, 1355-1359.

Neuberger, M.S. (1983). Expression and regulation of immunoglobulin heavy chain gene transfected into lymphoid cells. EMBO J. 2, 1373-1378.

Nevins, J.R. (1981). Mechanism of activation of early viral transcription by the adenovirus EIA gene product. Cell 26, 213-220.

Nevins, J.R. and Winkler, J.J. (1980). Regulation of early adenovirus transcription: a protein product of early region 2 specifically represses region 4 transcription. Proc. Natl. Acad. Sci., U.S.A. 77, 1893-1897.

Nir, U., Cohen, B., Chen, L. and Revel, M. (1984). A human IFN-$_1$ gene deleted of promoter sequences upstream from the TATA box is controlled post-transcriptionally by ds RNA. Nucleic Acids Res. 12, 6979-6993.

O'Hare, P. and Hayward, G.S (1985a). Evidence for a direct role for both the 175,000- and 110,000-molecular weight immediate-early proteins of herpes simplex virus in the *trans* activation of delayed-early promoters. J. Virol. 53, 751-760.

O'Hare, P. and Hayward, G.S (1985b). Three trans-acting regulatory proteins of herpes simplex virus modulate immediate-early gene expression in a pathway involving positive and negative feedback regulation. J. Virol. 56, 723-733.

Padgett, R.A., Grabowski, P.J., Konarska, M.M. and Sharp, P.A. (1985). Splicing messenger RNA precursors: branch sites and lariat RNAs. Trends in Biochem. Sci. 10, 154-157.

Palfreyman, J.W., MacLean, J.B., Messeder, E. and Sheppard, R.C. (1984). Successful use of oligopeptides as immunogens in the preparation of antisera to immediate-early gene products of herpes simplex virus type 1. J. Gen. Virol. 65, 865-874.

Parker, C.S. and Roeder, R.G. (1977). Selective and accurate transcription of the Xenopus laevis 5S RNA genes in isolated chromatin by purified RNA polymerase III. Proc. Natl. Acad. Sci., U.S.A. 74, 44-48.

Parker, C.S. and Topol, J. (1984a). A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA binding activity. Cell 36, 357-369.

Parker, C.S. and Topol, J. (1984b). A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp70 gene. Cell 37, 273-283.

Paule, M.R. (1981). Comparative subunit composition of the eukaryotic nuclear RNA polymerases. Trends in Biochem. Sci. 6, 128-131.

Payvar, F., DeFranco, D., Firestone, G.L., Edgar, B., Wrange, O., Okret, S., Gustafsson, J-A. and Yamamoto, K.R. (1983). Sequence-specific binding of glucocorticoid receptor to MTV DNA at sites within and upstream of the transcribed region. Cell 35, 381-392.

Payvar, F., Wrange, O., Carlsdet-Duke, J., Okret, S., Gustafsson, J-A. and Yamamoto, K.R. (1981). Purified glucocorticoid receptors bind *in vitro* to a cloned DNA fragment whose transcription is regulated by glucocorticoids *in vivo*. Proc. Natl. Acad. Sci., U.S.A. 78, 6628-6632.

Pelham, H.R.B. (1982). A regulatory upstream promoter element in the *Drosophila* Hsp 70 heat-shock gene. Cell 30, 517-528.

Pelham, H.R.B. and Bienz, M. (1982). A synthetic heat-shock promoter element confers heat-inducibility on the herpes simplex virus thymidine kinase gene. EMBO J. 1, 1473-1477.

Pelham, H.R.B. and Brown, D.D. (1980). A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. Proc. Natl. Acad. Sci., U.S.A. 77, 4170-4174.

Pellet, P.E., McKnight, J.L.C., Jenkins, F.J. and Roizman, B. (1985). Nucleotide sequence and predicted amino acid sequence of a protein encoded in a small herpes simplex virus DNA fragment capable of *trans*-inducing α genes. Proc. Natl. Acad. Sci., U.S.A. 82, 5870-5874.

Pereira, L., Wolff, M.H., Fenwick, M. and Roizman, B. (1977). Regulation of herpesvirus macromolecular synthesis V. properties of a polypeptides made in HSV-1 and HSV-2 infected cells. Virology 77, 733-749.

Perricaudet, M., Akusjarvi, G., Virtanen, A. and Petterson, U. (1979). Structure of two spliced mRNAs from the transforming region of human subgroup C adenoviruses. Nature 281, 694-696.

Persson, R.H., Bacchetti, S. and Smiley, J.R (1985). Cells that constitutively express the herpes simplex virus immediate-early protein ICP4 allow efficient activation of viral delayed-early genes in trans. J. Virol. 54, 414-421.

Pitha, P.M., Ciufo, D.M., Kellum, M., Raj, N.B.K., Reyes, G.R. and Hayward, G.S. (1982). Induction of human β-interferon synthesis with poly(rI.rC) in mouse cells transfected with cloned cDNA plasmids. Proc. Natl. Acad. Sci., U.S.A. 79, 4337-4341.

Plummer, G., Waner, J.L., Phuangsab, A. and Goodheart, C.R. (1981). Type 1 and type 2 herpes simplex viruses: serological and biological differences. J. Virol. 5, 51-59.

Poffenberger, K.L., Tabares, E. and Roizman, B. (1983). Characterisation of a viable, non-inverting herpes simplex virus 1 genome derived by insertion of sequences at the L-S component junction. Proc. Natl. Acad. Sci., U.S.A. 80, 2690-2694.

Ponce de Leon, M., Eisenberg, R.J. and Cohen, G.H. (1977). Ribonucleotide reductase from herpes simplex virus (type 1 and 2) infected and uninfected KB cells: Properties of the partially purified enzymes. J. Gen. Virol. 36, 163-173.

Post, L.E. and Roizman, B. (1981). A generalised technique for deletion of specific genes in large genomes: α gene 22 of herpes simplex virus 1 is not essential for growth. Cell 25, 227-232.

Post, L.E., Mackem, S. and Roizman, B. (1981). Regulation of genes of herpes simplex virus: expression of chimeric genes produced by fusion of thymidine kinase with α gene promoters. Cell 24, 555-565

Preston V.G. and Fisher, F.B. (1984). Identification of the herpes simplex virus type 1 gene encoding the dUTPase. Virology 138, 58-68.

Preston, C.M. (1979a). Control of herpes simplex type 1 mRNA synthesis in cells infected with wild-type virus or the temperature-sensitive mutant tsK. J. Virol. 29, 275-284.

Preston, C.M. (1979b). Abnormal properties of an immediate-early polypeptide in cells infected with the herpes simplex virus type 1 mutant tsK. J. Virol. 32, 357-359.

Preston, C.M. and Cordingley, M.G. (1982). mRNA- and DNA-directed synthesis of herpes simplex virus-coded exonuclease in Xenopus laevis oocytes. J. Virol. 43, 386-394.

Preston, C.M. and McGeoch, D.J. (1981). Identification and mapping of two polypeptides encoded within the herpes simplex virus type 1 thymidine kinase gene sequences. J. Virol. 38, 593-605.

Preston, C.M. and Notarianni, E. (1983). Poly(ADP-ribosyl)ation of a herpes simplex virus immediate early polypeptide. Virology 131, 492-501.

Preston, C.M. and Tannahill, D. (1984). Effects of orientation and position on the activity of a herpes simplex virus immediate early gene far-upstream region. Virology 137, 439-444.

Preston, C.M., Cordingley, M.G. and Stow, N.D. (1984). Analysis of DNA sequences which regulate the transcription of a herpes simplex virus immediate early gene. Virology 50, 708-716.

Preston, V.G. (1981). Fine-structure mapping of herpes simplex virus type 1 temperature-sensitive mutations within the short repeat region of the genome. J. Virol. 39, 150-161.

Preston, V.G., Davison, A.J., Marsden, H.S., Timbury, M.C., Subak-Sharpe, J.H. and Wilkie, N.M. (1978). Recombinants between herpes simplex virus types 1 and 2: analyses of genome structures and expression of immediate-early polypeptides. J. Virol. 28, 499-517.

Prives, C., Gilboa, E., Revel, M. and Winocour, E. (1977). Cell-free translation of simian virus 40 early messenger RNA coding for viral T antigen. Proc. Natl. Acad. Sci., U.S.A. 74, 457-461.

Proudfoot, N.J. (1979). Eukaryotic promoters. Nature 279, 376.

Proudfoot, N.J. and Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic mRNA. Nature 263, 211-214.

Pustell, J. and Kafatos, F.C. (1982). A high speed, high capacity homology matrix: zooming through SV40 and polyoma. Nucleic Acids Res. 10, 4765-4782.

Queen, C. and Baltimore, D. (1983). Immunoglobulin gene transcription is activated by downstream sequence elements. Cell, 33, 741-748.

Quinlan, M.P. and Knipe, D.M. (1985). Stimulation of expression of a herpes simplex virus DNA-binding protein by two viral functions. Mol. Cell. Biol. 5, 957-963.

Raab-Traub, N., Hood, R., Yang, C-S., Henry, B. and Pagano, J.S. (1983). Epstein-Barr virus transcription in nasopharyngeal carcinoma. J. Virol. 48, 580-590.

Raj, N.B.K. and Pitha, P.M. (1983). Two levels of regulation of β-interferon gene expression in human cells. Proc. Natl. Acad. Sci., U.S.A. 80, 3923-3927.

Razin, A. and Riggs, A. (1980). DNA methylation and gene function. Science 210, 604-610.

Reddy, V.B., Ghosh, P.K., Lebowitz, P., Piatak, M. and Weissman, S.M. (1979). Simian virus 40 early mRNAs: I. genomic localisation of 3' and 5' termini and two major splices in mRNA from transformed and lytically infected cells. J. Virol. 30, 279-296.

Renkawitz, R., Schultz, G., Von der Ahe, D. and Beato, M. (1984). Sequences in the promoter region of the chicken lysozyme gene required for steroid regulation and receptor binding. Cell 37, 503-510.

Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L. and Luskey, K.L. (1984). HMGCoA reductase: a negatively regulated gene with unusual promoter and 5' untranslated regions. Cell 38, 275-285.

Ricciardi, R.P., Jones, R.L., Cepko, C.L., Sharp, P.A. and Roberts, B.E. (1981). Expression of early adenovirus genes requires a viral encoded acidic polypeptide. Proc. Natl. Acad. Sci., U.S.A. 78, 6121-6125.

Richards, J.C., Hyman, R.W. and Rapp, F. (1979). Analysis of the DNAs from seven varicella-zoster virus isolates. J. Virol. 32, 812-821.

Richards, R.I., Heguy, A. and Karin, M. (1984). Structural and functional analysis of the human metallothionein-I_A gene: Differential induction by metal ions and glucocorticoids. Cell 37, 263-272.

Ringold, G., Yamamoto, K.R., Bishop, J.M. and Varmus, H.E. (1977). Glucocorticoid-stimulated accumulation of mouse mammary tumour virus RNA: increased rate of synthesis of viral RNA. Proc. Natl. Acad. Sci., U.S.A. 74, 2879-2883.

Rio, D., Robbins, A., Myers, R. and Tjian, R. (1980). Regulation of SV40 early transcription in vitro by purified tumour antigen. Proc. Natl. Acad. Sci., U.S.A. 77, 5706-5710.

Rio, D.C. and Tjian, R. (1983). SV40 T antigen binding site mutations that affect autoregulation. Cell 32, 1227-1240.

Rixon, F.J. and McGeoch, D.J. (1984). A 3' co-terminal family of mRNAs from the herpes simplex virus type 1 short region: Two overlapping reading frames encode unrelated polypeptides one of which has a highly reiterated amino acid sequence. Nucleic Acids Res. 12, 2473-2487.

Roeder, R.G. and Rutter, W.J. (1970). Specific nucleolar and nucleoplasmic RNA polymerases. Proc. Natl. Acad. Sci., U.S.A. 65, 675-682.

Rohrer, H., Zillig, W. and Mailhammer, R. (1975). ADP-ribosylation of DNA-dependent RNA polymerase of Escherichia coli by an NAD^+:protein ADP-ribosyltransferase from bacteriophage T4. Eur. J. Biochem. 60, 227-238.

Roizman, B. (1982). The family herpesviridae: general description, taxonomy and classification, in "The Herpesviruses" vol.I, ppl-23. Ed. B. Roizman, Plenum Press, N.Y..

Roizman, B. and Furlong, D. (1974). The replication of herpesviruses. In "Comprehensive Virology" vol. 3, pp229-403. Ed. H. Fraenkel-Conrat and R.R. Wagner. Plenum Press, N.Y.

Roizman, B., Carmichael, L.E., Deinhardt, F., De The, G., Nahmias, A.J., Plowright, F., Sheldrick, P., Takahashi, M. and Wolf, K. (1981). Herpesviridae: Definition, provisional nomenclature and taxonomy. Intervirology 16, 201-217.

Rose, J.K. and Gallione, C.J. (1981). Nucleotide sequences of the mRNAs encoding the vesicular stomatitis virus G and M proteins determined from cDNA clones containing the complete coding regions. J. Virol. 39, 519-528.

Rosenberg, M. and Court, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. 13, 319-353.

Rosenthal, N., Kress, M., Gruss, P. and Khoury, G. (1983). BK viral enhancer element and a human cellular homolog. Science 222, 749-755.

Rossini, M. (1983). The role of adenovirus early region 1A in the regulation of early regions 2A and 1B expression. Virology 131, 49-58.

Ruet, A., Camier, S., Smagowicz, W., Sentenac, A. and Fromageot, P. (1984). Isolation of a class C transcription factor which forms a stable complex with tRNA genes. EMBO J. 3, 343-350.

Ryals, J., Dierks, P., Ragg, H. and Weissman, C. (1985). A 46-nucleotide promoter segment from an IFN-α gene renders an unrelated promoter inducible by virus. Cell 41, 497-507.

Sacks, W.R., Greene, C.C., Aschman, D.P. and Schaffer, P.A. (1985). Herpes simplex virus type 1 ICP27 is an essential regulatory protein. J. Virol. 55, 796-805.

Sakonju, S. and Brown, D.D. (1982). Contact points between a positive transcription factor and the Xenopus 5S RNA gene. Cell 31, 395-405.

Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980). A control region in the center of the 5S RNA gene directs specific initiation of transcription: I. The 5' border of the region. Cell 19, 13-25.

Sakonju, S., Brown, D.D., Engelke, D., Ng, S-Y., Shastry, B.S. and Roeder, R.G. (1981). The binding of a transcription factor to deletion mutants of a 5S ribosomal tRNA gene. Cell 23, 665-669.

Sambrook, J., Botchan, M., Gallimore, P., Ozanne, B., Pettersson, U., Williams, J. and Sharp, P.A. (1974). Viral DNA sequences in cells transformed by simian virus 40, adenovirus type 2 and adenovirus type 5. Cold Spring Harbor Symp. Quant. Biol. 39, 615-632.

Samuels, M., Fire, A. and Sharp, P.A. (1982). Seperation and characterisation of factors mediating accurate transcription by RNA polymerase II. J. Biol. Chem. 257, 14419-14427.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980). Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143, 161-178.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci., U.S.A. 74, 5463-5467.

Sassone-corsi, P., Dougherty, J.P., Wasylyk, B. and Chambon, P. (1984). Stimulation if in vitro transcription from heterologous promoters by the simian virus 40 enhancer. Proc. Natl. Acad. Sci., U.S.A. 81, 308-312.

Saunders, P.G. (1981). Analysis of mutations produced at random and in specific regions of the HSV genome. Ph.D. Thesis, University of Glasgow.

Sassone-Corsi, P., Hen, R., Borelli, E., Leff, T. and Chambon, P. (1983). Far upstream sequences are required for efficient transcription from the adenovirus-2 E1A transcription unit. Nucleic Acids Res. 11, 8735-8745.

Sassone-corsi, P., Wildeman, A. and Chambon, P. (1985). A trans-acting factor is responsible for the simian virus 40 enhancer activity in vitro. Nature 313, 458-463.

Sawadogo, M. and Roeder, R.G. (1984). Energy requirement for specific transcription initiation by the human RNA polymerase II system. J. Biol. Chem. 259, 5321-5326.

Scheidereit, C., Geisse, S., Westphal, H.M.and Beato, M. (1983). The glucocorticoid receptor binds to defined nucleotide sequences near the promoter of mouse mammary tumour virus. Nature 304, 749-752.

Schirm, S., Weber, F., Schaffner, W. and Fleckenstein, B. (1985). A transcriptional enhancer in the herpesvirus saimiri genome. EMBO J. 4, 2669-2674.

Schlesinger, M.J., Ashburner, M. and Tissieres, A. (Eds.). "Heat Shock: from bacteria to man.". Cold Spring Harbor Laboratory, 1982.

Schroder, C.H., DeZazzo, J., Knopf, K.W., Kaerner, H.C., Levine, M. and Glorioso, J. (1985). A herpes simplex virus type 1 mutant with a deletion in the polypeptide-coding sequences of the ICP4 gene. J. Gen. Virol. 66, 1589-1593.

Schwartz, R.M. and Dayhoff, M.O., in "Atlas of Protein Sequence and Structure.", M.O. Dayhoff, (Ed.), vol.5, Suppl.3, pp.353-358. National Biomedical Research Foundation, Washington, D.C., 1979.

Searle, P.F., Davison, B.L., Stuart, G.W., Wilkie, T.M., Norstedt, G. and Palmiter, R.D. (1984). Regulation, linkage and sequence of mouse metallothionein I and II genes. Mol. Cell. Biol. 4, 1221-1230.

Searle, P.F., Stuart, G.W., and Palmiter, R.D. (1985). Building a metal-responsive promoter with synthetic regulatory elements. Mol. Cell. Biol. 5, 1480-1489.

Sears, A.E., Halliburton, I.W., Meignier, B., Silver, S. and Roizman, B. (1985). Herpes simplex virus 1 mutant deleted in the α22 gene: Growth and gene expression in permissive and restrictive cells and establishment of latency in mice. J. Virol. 55, 338-346.

Segall, J., Matsui, T. and Roeder, R.G. (1980). Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III. J. Biol. Chem. 255, 11986-11991.

Seifert, W., Qasba, P., Walter, G., Palm, P., Schachner, M. and Zillig, W. (1969). Kinetics of the alteration and modification of DNA-dependent RNA polymerase in T4 infected E.coli cells. Eur. J. Biochem. 9, 319-324.

Sekimizu, K., Yokoi, H. and Natori, S. (1982). Evidence that stimulatory factor(s) of RNA polymerase II participates in accurate transcription in a HeLa cell lysate. J. Biol. Chem. 257, 2719-2721.

Sentenac, A. (1985). Eukaryotic RNA polymerases. CRC Crit. Rev. Biochem. 18, 31-90.

Sergeant, A., Bohmann, D., Zentgraf, H., Weiher, H. and Keller, W. (1984). A transcription enhancer acts in vitro over distances of hundreds of base pairs on both circular and linear templates but not on chromatin-reconstituted DNA. J. Mol. Biol. 180, 577-600.

Shalloway, D., Kleinberger, T. and Livingston, D.M. (1980). Mapping of SV40 DNA replication origin region binding sites for the SV40 T antigen by protection against exonuclease III digestion. Cell 20, 411-422.

Sharp, S., DeFranco, D., Dingermann, T., Farrell, P. and Söll, D. (1981). Internal control regions for transcription of eukaryotic tRNA genes. Proc. Natl. Acad. Sci., U.S.A. 78, 6657-6661.

Shaw, A.R. and Ziff, E.B. (1980). Transcripts from the adenovirus-2 major late promoter yield a single early family of 3' coterminal mRNAs and five late families. Cell 22, 905-916.

Shein, H.M. and Elders, J.F. (1962). Transformation induced by simian virus 40 in human renal cell cultures. I. Morphology and growth characteristics. Proc. Natl. Acad. Sci. U.S.A. 48, 1164-1170.

Sheldrick, P. and Berthelot, N. (1974). Inverted repetitions in the chromosome of herpes simplex virus. Cold Spring Harbor Symp. Quant. Biol. U.S.A. 39, 667-678.

Sheldrick, P., Laither, M., Lando, D. and Ryhner, M.L. (1973). Infectious DNA from herpes simplex virus infectivity of double and single stranded molecules. Proc. Natl. Acid. Sci. U.S.A. 70, 3621-3625.

Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980). E.coli RNA polymerase interacts homologously with two different promoters. Cell 20, 269-281.

Skinner, J.A., Ohrlein, A. and Grummt, I. (1984). In vitro mutagenesis and transcriptional analysis of a mouse ribosomal promoter element. Proc. Natl. Acad. Sci., U.S.A. 81, 2137-2141.

Slattery, E., Dignam, J.D., Matsui, T. and Roeder, R.G. (1983). Purification and analysis of a factor which suppresses nick-induced transcription by RNA polymerase II and its identity with poly(ADP-ribose) polymerase. J. Biol. Chem. 258, 5955-5959.

Sleigh, M.J., Topp, W.C., Hanich, R. and Sambrook, J.F. (1978). Mutants of SV40 with an altered small t protein are reduced in their ability to transform cells. Cell 14, 79-88.

Smale, S.T. and Tjian, R. (1985). Transcription of herpes simplex virus tk sequences under the control of wild-type and mutant human RNA polymerase I promoters. Mol. Cell. Biol. 5, 352-362.

Smiley, J.R., Fong, B.S. and Leung, W-C. (1981). Construction of a double jointed herpes simplex viral DNA molecule: inverted repeats are required for segment inversion and direct repeats promote deletions. Virology 113, 345-362.

Smiley, J.R., Swan, H., Pater, M.M., Pater, A. and Halpern, M.E. (1983). Positive control of the herpes simplex virus thymidine kinase gene requires upstream DNA sequences. J. Virol. 47, 301-310.

Smith, D.H., Kegler, D.M. and Ziff, E.B. (1985). Vector expression of adenovirus type 5 Ela proteins: Evidence for Ela autoregulation. Mol. Cell. Biol. 5, 2684-2696.

Spaete, R.R. and Mocarski, E.S. (1985). Regulation of cytomegalovirus gene expression: α and β promoters are trans activated by viral functions in permissive human fibroblasts. J. Virol. 56, 135-143.

Spaete, R.R., and Frenkel, N. (1982). The herpes simplex virus amplicon: a new eukaryotic defective-virus cloning-amplifying vector. Cell 30, 295-304.

Spear, P.G. (1980). Composition and organisation of herpesvirus virions and properties of some of the structural proteins., in "Oncogenic Herpesviruses", vol.I, F.Rapp (Ed.), CRC Press.

Spear, P.G. and Roizman, B. (1972). Proteins specified by herpes simplex virus. V. purification and structural proteins of the herpes virion. J. Virol. 9, 143-159.

Spector, D.J., McGrogan, M. and Raskas, H.J. (1978). Regulation of the appearance of cytoplasmic RNAs from region 1 of the adenovirus 2 genome. J. Mol. Biol. 126, 395-414.

Spradling, A.C. and Rubin, G.M. (1982). Transposition of cloned P elements into Drosophila germ line chromosomes. Science 218, 341-347.

Sprague, K.U., Larson, D. and Morton, D. (1980). 5' flanking sequence signals are required for activity of silkworm alanine tRNA genes in homologous in vitro transcription systems. Cell 22, 171-178.

Staden, R. (1979). A strategy of DNA sequencing employing computer programs. Nucleic Acids Res. 6, 2601-2610.

Staden, R. (1980). A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res. 8, 3673-3694.

Staden, R. and McLachlan, A.D. (1982). Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Res. 10, 141-156.

Stein, R. and Ziff, E.B. (1984). HeLa cell β-tubulin gene transcription is stimulated by adenovirus 5 in parallel with early genes by an Ela-dependent mechanism. Mol. Cell. Biol. 4, 2792-2801.

Stenberg, R.M. and Stinski, M.F. (1985). Autoregulation of the human cytomegalovirus major immediate-early gene. J. Virol. 56, 676-682.

Stevens, A. (1970). An isotopic study of DNA-dependent RNA polymerase of E.coli following T4 infection. Biochem. Biophys. Res. Commun. 41, 367-373.

Stevens, A. (1972). New small polypeptides associated with DNA-dependent RNA polymerases of Escherichia coli after infection with bacteriophage T4. Proc. Natl. Acad. Sci., U.S.A. 69, 603-607.

Stevens, A. (1974). Deoxyribonuleic acid dependent ribonucleic acid polymerases from two T4 phage-infected systems. Biochem. 13, 493-503.

Stinski, M.F. and Roehr, J.J. (1985). Activation of the major immediate early gene of human cytomegalovirus by cis-acting elements in the promoter-regulatory sequence and by virus-specific trans-acting components. J. Virol. 55, 431-441.

Stow, N.D. (1982). Localization of an origin of DNA replication within the TR_S/IR_S repeated region of the herpes simplex virus type 1 genome. EMBO J. 1, 863-867.

Stow, N.D. and McMonagle, E.C. (1983). Characterization of the TR_S/IR_S origin of DNA replication of herpes simplex virus type 1. Virology 130, 427-438.

Stow, N.D. and Wilkie, N.M. (1976). An improved technique for obtaining enhanced infectivity with herpes simplex virus type 1 DNA. J. gen. Virol. 33, 447-458.

Stow, N.D., McMonagle, E.C. and Davison, A.J. (1983). Fragments from both termini of the herpes simplex virus type 1 genome contain signals required for the encapsidation of viral DNA. Nucleic Acids Res. 11, 8205-8220.

Struhl, K. (1984). Genetic properties and chromatin structure of the yeast gal regulatory element: an enhancer-like sequence. Proc. Natl. Acad. Sci., U.S.A. 81, 7865-7869.

Stuart, G.W., Searle, P.F., Chen, H.Y., Brinster, R.L. and Palmiter, R.D. (1984). A 12 base pair DNA motif that is repeated several times in metallothionein gene promoters confers metal regulation to a heterologous gene. Proc. Natl. Acad. Sci., U.S.A. 81, 7318-7322.

Svensson, C. and Akusjarvi, G. (1984). Adenovirus 2 early region 1A stimualtes expression of both viral and cellular genes. EMBO J. 3, 789-794.

Swanstrom, R.I., Pivo, K. and Wagner, E.K. (1975). Restricted transcription of the herpes simplex virus genome occurring early after infection and in the presence of metabolic inhibitors. Virology 66, 140-150.

Tavernier, J., Gheysen, D., Duernick, F., Van der Heyden, J. and Fiers, W. (1983). Deletion mapping of the inducible promoter of human IFN-β gene. Nature 301, 634-636.

Taya, Y., Devos, R., Tavernier, J., Cheroutre, H., Engler, G. and Fiers, W. (1982). Cloning and structure of the human immune interferon-γ chromosomal gene. EMBO J. 1, 953-958.

Taylor, P. (1984). A fast homology program for aligning biological sequences. Nucleic Acids Res. 12, 447-455.

Taylor, P. (1986). A computer program for translating DNA sequences into protein. Nucleic Acids Res. 14, 437-441.

Tegtmeyer, P. (1972). Simian virus 40 deoxyribonucleic acid synthesis: the viral replicon. J. Virol. 10, 591-598.

Tegtmeyer, P., Schwartz, M., Collins, J.K. and Rundell, K. (1975). Regulation of tumor antigen synthesis by simian virus 40 gene A. J. Virol. 16, 168-178.

Tenen, D.G., Haines, L.L. and Livingston, D.M. (1982). Binding of an analog of the simian virus 40 T antigen to wild-type and mutant viral replication origins. J. Mol. Biol. 157, 473-492.

Thimmappaya, B. and Weissman, S.M. (1977). The early region of SV40 DNA may have more than one gene. Cell 11, 837-843.

Tjian, R. (1978). The binding site on SV40 DNA for a T-antigen-related protein. Cell 13, 165-179.

Tjian, R., Fey, G. and Graessmann, A. (1978). Biological activity of purified simian virus 40 T antigen proteins. Proc. Natl. Acad. Sci. U.S.A. 75, 1279-1283.

Tooze, J. (1980). "Molecular Biology of Tumor Viruses: DNA Tumor Viruses." 2nd Ed., part2. Cold Spring Harbor Laboratory.

Treisman, R. and Maniatis, T. (1985). Simian virus 40 enhancer increases number of RNA polymerase II molecules on linked DNA. Nature 315, 72-75.

Treisman, R., Green, M. and Maniatis, T. (1983). *Cis* and *trans*-activation of globin gene transcription in transient assays. Proc. Natl. Acad. Sci. U.S.A. 80, 7428-7432.

Tsai, S-Y., Tsai, M-J. and O'Malley, B.W. (1981). Specific 5' flanking sequences are required for faithful initiation of *in vitro* transcription. Proc. Natl. Acad. Sci., U.S.A. 78, 879-883.

Tsuda, M. and Suzuki, Y. (1981). Faithful transcription of fibroin gene in a homologous cell-free system reveals an enhancing effect of 5' flanking sequence far upstream. Cell 27, 175-182.

Twigg, A.J. and Sherratt, D. (1980). Trans-complementable copy-number mutants of plasmid ColE1. Nature 283, 216-218.

Tyndall, C., La Mantia, G., Thacker, C.M., Favaloro, J. and Kamen, R. (1981). A region of the polyoma virus genome between the replication origin and late protein coding sequences is required in cis for both early gene expression and viral DNA replication. Nucleic Acids Res. 9, 6231-6251.

Tyndall, C., La Mantia, G., Thacker, C.M., Favaloro, J. and Kramer, R. (1981). A region of the polyoma virus genome between the replication origin and late protein coding sequences is required in cis for both early gene expression and viral DNA replication. Nucleic Acids Res. 9, 6231-6250.

Ucker, D.S., Ross, S.R. and Yamamoto, K.R. (1981). mammary tumor virus DNA contains sequences required for its hormone-regulated transcription. Cell 27, 257-266.

Ueno, K., Sekimizu, K., Mizuno, D. and Natori, S. (1979). Antibody against a stimulatory factor of RNA polymerase II inhibits nuclear RNA synthesis. Nature 277, 145-146.

Valerio, D., Duyvesteyn, M.G.C., Dekker, B.M.M., Weeda, G., Berkvens, T.M., Van der Voorn, L., Van Ormondt, H. and Van der Eb, A.J. (1985). Adenosine deaminase: characterisation and expression of a gene with a remarkable promoter. EMBO J. 4, 437-443.

Varmuza, S.L. and Smiley, J.R. (1985). Signals for site-specific cleavage of HSV DNA: Maturation involves two seperate cleavage events at sites distal to the recognition sequences. Cell 41, 793-802.

Velcich, A. and Ziff, E. (1985). Adenovirus Ela proteins repress transcription from the SV40 early promoter. Cell 40, 705-716.

Veldman, G.M., Lupton, S. and Kamen, R. (1985). Polyoma virus enhancer contains multiple redundant sequence elements that activate both DNA replication and gene expression. Mol. Cell. Biol. 5, 649-658.

Wadsworth, S., Jacob, R.J. and Roizman, B. (1975). Anatomy of herpes simplex virus DNA II. Size composition and arrangement of inverted terminal repeats. J. Virol. 5, 1487-1497.

Wagner, E.K. (1984). Regulation of HSV transcription. J. Inv. Derm. 83, 48s-52s.

Wagner, E.K. (1985). Individual HSV transcripts: Characterisation of specific genes., in "The Herpesviruses", vol.3, pp45-104., B. Roizman (Ed.), Plenum Press.

Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teolue, R. and Chambon, P. (1980a). Specific *in vitro* transcription of conalbumin gene is drastically decreased by a single-point mutation in TATA box homology sequence. Proc. Natl. Acad. Sci., U.S.A. 77, 7024-7028.

Wasylyk, B., Kedinger, C., Corden, J., Brison, O. and Chambon, P. (1980b). Specific *in vitro* initiation of transcription on conalbumin and ovalbumin genes and comparison with adenovirus 2 earlyand late genes. Nature 285, 367-373.

Wasylyk, B., Wasylyk, C., Augereau, P. and Chambon, P. (1983a). The SV40 72bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. Cell 32, 503-514.

Wasylyk, B., Wasylyk, C., Matthes, H., Wintzerith, M. and Chambon, P. (1983b). Transcription from the SV40 early-early and late-early overlapping promoters in the absence of DNA replication. EMBO J. 2, 1605-1611.

Watson, R.J. and Clements, J.B. (1978). Characterization of transcription-deficient temperature-sensitive mutants of herpes simplex virus type 1. Virology 91, 364-379.

Watson, R.J. and Clements, J.B. (1980). Identification of a herpes simplex virus type 1 function continuously required for synthesis of early and late virus RNAs. Nature 285, 329-330.

Watson, R.J., Preston, C.M. and Clements, J.B. (1979). Separation and characterization of herpes simplex virus type 1 immediate-early mRNAs. J. Virol. 31, 42-52.

Weaver, R.F. and Weissmann, C. (1979). Mapping of RNA by a modification of the Berk-Sharp procedure: The 5' termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. 7, 1175-1193.

Weber, F. and Schaffner, W. (1985). Simian virus 40 enhancer increases RNA polymerase density within the linked gene. Nature 315, 75-77.

Weber, F., De Villiers, J. and Schaffner, W. (1983). An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell 36, 983-992.

Weeks, D.L. and Jones, N.C. (1983). E1A control of gene expression is mediated by sequences 5' to the transcriptional starts of the early viral genes. Mol. Cell. Biol. 3, 1222-1234.

Weeks, D.L. and Jones, N.C. (1985). Adenovirus E3-early promoter: sequences required for activation by E1A. Nucleic Acids Res. 13, 5389-5402.

Weidle, U. and Weissman, C. (1983). The 5' flanking region of a human IFN-α gene mediates viral induction of transcription. Nature 303, 442-446.

Weiher, H., Konig, M. and Gruss, P. (1983). Multiple point mutations affecting the simian virus 40 enhancer. Science 219, 626-631.

Weil, P.A., Luse, D.S., Segall, J. and Roeder, R.G. (1979b). Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell 18, 469-484.

Weil, P.A., Segall, J., Harris, B., Hg, S. and Roeder, R.G. (1979a). Faithful transcription of eukaryotic gene by RNA polymerase III in systems reconstituted with purified DNA templates. J. Biol. Chem. 254, 6163-6173.

Weller, S.K., Spadaro, A., Schaffer, J.E., Murray, A.W., Maxam, A.M. and Schaffer, P.A. (1985). Cloning, sequencing, and functional analysis of ori_L, a herpes simplex virus type 1 origin of DNA synthesis. Mol. Cell. Biol. 5, 930-942.

Weller, T.H., Witton, H.M. and Bell, E.J. (1958). The etiologic agents of varicella and herpes zoster: isolation, propagation and cultural characteristics in vitro. J. Exp. Med. 108, 843-868.

Whitton, J.L. (1984). Structural elements involved in herpes simplex virus immediate-early transcription. Ph.D. Thesis, University of Glasgow.

Whitton, J.L. and Clements, J.B. (1984). Replication origins and a sequence involved in coordinate induction of the immediate-early gene family are conserved in an intergenic region of herpes simplex virus. Nucleic Acids Res. 12, 2061-2079.

Whitton, J.L., Rixon, F.J., Easton, A.J. and Clements, J.B. (1983). Immediate-early mRNA2 of herpes simplex viruses types 1 and 2 is unspliced: conserved sequences around the 5' and 3' termini correspond to transcription regulatory signals. Nucleic Acids Res. 11, 6271-6287.

Wigler, M., Silverstein, S., Lee, L., Pellicer, A., Cheng, Y. and Axel, R. (1977). Transfer of purified HSV TK gene to cultured mouse cells. Cell 11, 223-232.

Wilbur, W.J. and Lipman, D. (1983). Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci., U.S.A. 80, 726-730.

Wilcox, K.W., Kohn, A., Sklyanskaya, E. and Roizman, B. (1980). Herpes simplex virus phosphoproteins I. Phosphate cycles on and off some viral polypeptides and can alter their affinity for DNA. J. Virol. 33, 167-182.

Wildeman, A.G., Sassone-corsi, P., Grundstrom, T., Zenke, M. and Chambon, P. (1984). Stimulation of *in vitro* transcription from the SV40 early promoter by enhancer involves a specific *trans*-acting factor. EMBO J. 3, 3129-3133.

Wildy, P., Russell, W.C. and Horne, R.W. (1960). The morphology of herpesvirus. Virology 12, 204-222.

Wilkie, N.M. (1976). Physical maps for herpes simplex virus type 1 DNA for restriction endonucleases HindIII, HpaII and XbaI. J. Virol. 20, 222-233.

Winberg, G. and Shenk, T. (1984). Dissection of overlapping functions within the adenovirus type 5 E1A gene. EMBO J. 3, 1907-1912.

Wohlrab, F. and Francke, B. (1980). Deoxyribopyrimidine triphosphatase activity specific for cells infected with herpes simplex virus type 1. Proc. Natl. Acad. Sci. U.S.A. 77, 1872-1876.

Wu, C. (1984). Two protein-binding sites in chromatin implicated in the activation of heat shock genes. Nature 309, 229-234.

Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. Nature 289, 751-758.

Yee, S-P., Rowe, D.T., Tremblay, M.L., McDermott, M. and Branton, P.E. (1983). Identification of human adenovirus early region 1 products by using antisera against synthetic peptides corresponding to the predicted carboxy termini. J. Virol. 46, 1003-1013.

Zajchowski, D.A., Boeuf, H. and Kedinger, C. (1985). The adenovirus-2 early EIIa transcription unit possesses two overlapping promoters with different sequence requirements for EIa-dependent stimulation. EMBO J. 4, 1293-1300.

Zinn, K., DiMaio, D and Maniatis, T. (1983). Identification of two distinct regulatory regions adjacent to the human β-interferon gene. Cell 34, 865-879.

Zipser, D., Lipsich, L. and Kwoh, J. (1981). Mapping of functional domains in the promoter region of the herpes thymidine kinase gene. Proc. Natl. Acad. Sci., U.S.A. 78, 6276-6280.

DNA sequence of the herpes simplex virus type 1 gene whose product is responsible for transcriptional activation of immediate early promoters

M.A.Dalrymple, D.J.McGeoch+, A.J.Davison+ and C.M.Preston+

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

Received 29 August 1985; Accepted 2 October 1985

ABSTRACT

Previous work has shown that transcriptional activation of herpes simplex virus type 1 (HSV-1) immediate early genes is mediated by a protein species (Vmw65) present in the tegument of infecting virions. This paper describes DNA sequence analysis and mRNA mapping of the Vmw65 gene in HSV-1 strain 17. The Vmw65 coding region was identified as a 490 codon sequence encoding a polypeptide of molecular weight 54,342 and characterised by a high proportion of charged amino acid residues. A homologue to Vmw65 was detected in the genome of varicella-zoster virus, another human herpesvirus. Apart from its role in trans-activation, Vmw65 is a major constituent of the virion. Its possible significance in virus structure is discussed.

INTRODUCTION

Herpes simplex virus type 1 (HSV-1) is a large, enveloped, icosahedral virus (1). It has a linear, double-stranded DNA genome of approximately 155 kilobase pairs (kbp) (2, 3) which is transcribed and replicated in the cell nucleus. At least three classes of genes have been distinguished in a temporal cascade, based on their requirements for de novo protein synthesis and viral DNA replication before transcriptional activation (4, 5, 6, 7). These classes have been named immediate early (IE), early (E) and late (L). The IE group consists of five genes named IE1(2 copies), 2, 3 (2 copies), 4 and 5. These genes are the first to be expressed after infection and do not require prior viral protein synthesis for their transcription (4, 5). Polypeptide Vmw175, the product of IE3, is required continuously for the expression of all E and L genes (8, 9). There is growing evidence that other IE proteins are involved in modulating transcription of the later temporal classes, probably in conjunction with Vmw175 (10, 11, 12).

The 5' flanking sequences which control expression of HSV-1

IE genes have been studied by several groups. In particular, IE3 and 4/5 (IE mRNAs 4 and 5 have identical 5' termini and upstream sequences) possess efficient promoters, including upstream regions with enhancer-like activity (13, 14, 15, 16, 17, 18, 19). Furthermore, transcription from all IE genes is strongly stimulated by a structural component of the herpes simplex virion (13, 15, 18, 20). Functional analysis of the IE gene regulatory regions (18, 21) located an AT rich, cis-acting, DNA sequence essential for the stimulatory response. This element is represented by a consensus sequence TAATGARATTC (R=purine) present in one or more copies far upstream of all IE genes, both in HSV-1 and HSV-2 (14, 20, 22, 23, 24). Thus, the available evidence suggests that stimulation of HSV IE gene transcription is mediated by a virion component which interacts, either directly or indirectly, with specific DNA sequences located far upstream from the mRNA 5' terminus.

Studies by Batterson and Roizman (25) have shown that the virion component is trans-acting and probably consists of one or more polypeptides located outside the nucleocapsid. Campbell, Palfreyman and Preston (26) applied a novel strategy to identify the HSV-1 gene or genes encoding the specific stimulatory factor. Cloned restriction fragments of the HSV-1 genome were transfected into tissue culture cells, together with a plasmid containing the HSV-1 thymidine kinase (TK) gene under the control of the IE3 promoter. Genomic fragments which stimulated the levels of TK activity were identified. Initial experiments located the IE specific trans-activator to the EcoRI I region (map coordinates 0.635-0.721). By a process of subcloning and insertion mutagenesis the authors were able to exclude all mRNAs known to map in EcoRI I except one (27). Hybrid arrested in vitro translation and immunoprecipitation studies identified the gene responsible as that for Vmw65, a major virion tegument phosphoprotein (28, 29). The tegument is an amorphous layer of protein, which lies between the viral nucleocapsid and envelope, of unknown function.

We report here the complete nucleotide sequence of the gene encoding Vmw65. The gene maps to a 2.6kbp fragment located in the long unique region of the genome, between map coordinates

0.669 and 0.685. In this region there is only one potential open reading frame consistent with our mapping data and analysis of codon usage. Furthermore, the 5' and 3' flanking regions exhibit typical promoter and termination sequence elements.

MATERIALS AND METHODS

Plasmids

The sequence reported here is derived from the sequence of the BamHI F fragment of HSV-1 strain 17, cloned into the BamHI site of pAT153 (30) and named pGX158. The plasmid was propagated in Escherichia coli DH1 (31). Plasmid DNA was prepared by established techniques and banded by CsCl gradient ultracentrifugation (32).

DNA Sequence Analysis

The dideoxy sequencing methodology was used (33, 34). Plasmid pGX158 was digested with BamHI and the vector sequences separated by agarose gel electrophoresis. The insert fragment was electroeluted from the gel slice. Purified fragment was self-ligated,using T4 DNA ligase, and then sonicated using a Dawe probe sonicator (35). Sonicated DNA was end repaired by incubation with T4 DNA polymerase and all four deoxynucleoside triphosphates. The end repaired fragments were then size fractionated by a second round of gel electrophoresis and electroelution. Purified fragments in the size range 300-600 base pairs were cloned into SmaI cleaved and phosphatase-treated M13mp8 vector (36). Clones were transfected into E.coli JM101, suitable plaques picked and single stranded template isolated and sequenced.

Computing

The sequence data were managed with the database system of Staden (37), implemented by Dr.P.Taylor on a DEC PDP 11/44 machine running under the RSX-11M operating system.

Open reading frames were identified using the program of Blumenthal et al (38) and codon usage evaluated by the method of Staden and McLachlan (39). Amino acid sequence hydrophobicities were examined with the parameters of Kyte and Doolittle (40). Sequence homologies were assessed by use of a matrix comparison program (41) and an optimal alignment program (42).

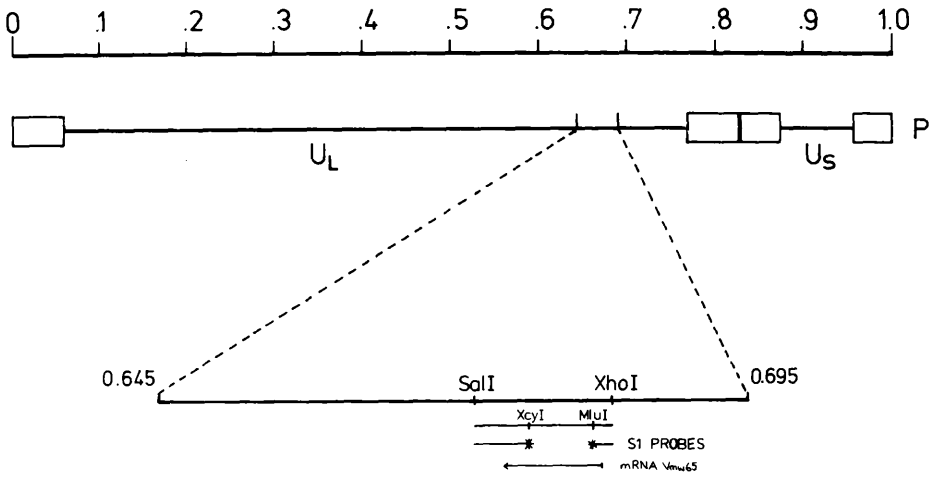


Figure 1: Location of the BamHI F fragment in the HSV-1 genome

A diagrammatic representation of the HSV-1 prototype genome. U_L and U_S are the long and short unique regions of the genome. The open boxes represent repeated sequences. BamHI F is located between 0.645 and 0.695 map units. The sequence presented is marked as a line between SalI and XhoI sites. Also shown are the positions of the end-labelled probes used to determine the 5' and 3' termini of the mRNA for Vmw65.

mRNA Mapping

Total cytoplasmic RNA was extracted from BSC-1 cells at 6h after infection with 20 plaque forming units of wild-type HSV-1 strain 17 per cell, or mock infected, as described by Preston (43).

To map the 5' terminus of the mRNA, 10ug of RNA was hybridised with a 5' end-labelled MluI or SalI fragment derived from the plasmid pMC1 (26), a 2.6kb sub-clone of pGX158 which contains the complete coding and flanking sequences of the Vmw65 gene and whose sequence is reported here. The 3' end of the mRNA was mapped by hybridisation to a 3' end labelled XcyI fragment. The respective probes are shown in Figure 1. The full procedure has been described previously (44).

RESULTS

Location of the Gene for Vmw65

The gene for Vmw65 lies between map coordinates 0.669 and

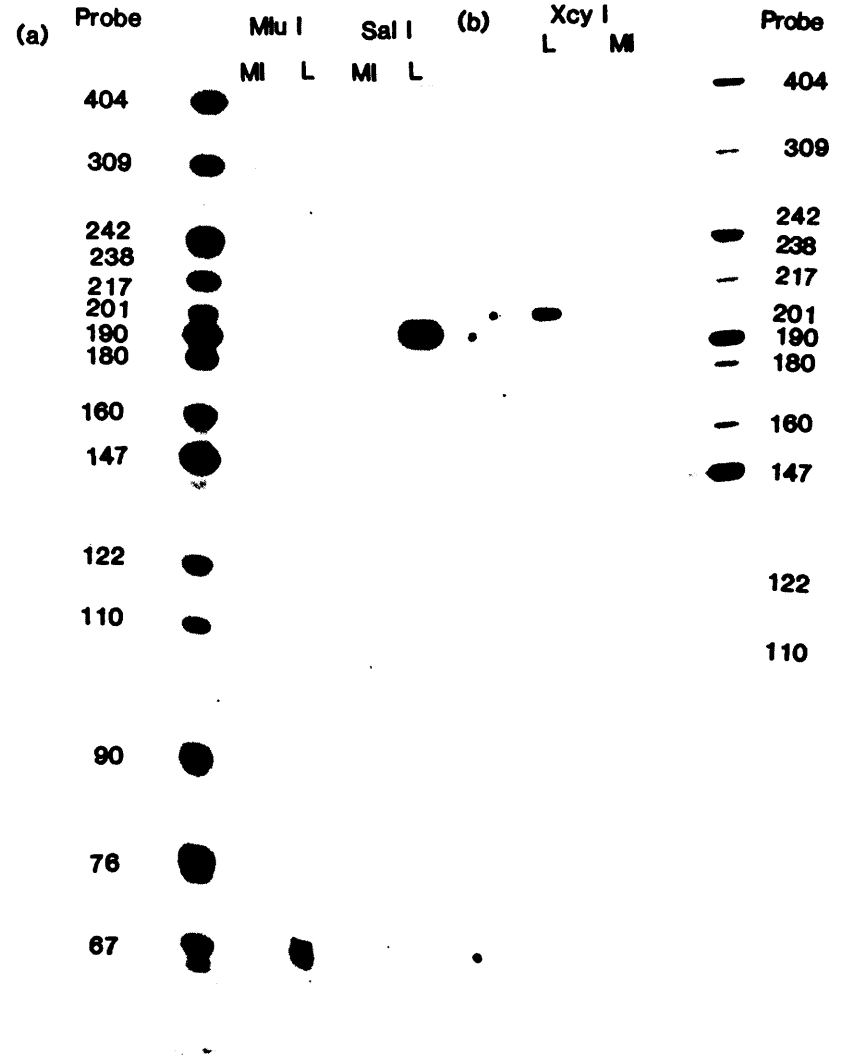


Figure 2: S1 mapping of the 5' and 3' termini of the mRNA for Vmw65

Panel (a) shows the results of using 5' end-labelled restriction fragments to map the mRNA initiation site. The probes used are indicated above the lanes. The hybrid bands obtained are discussed in the text.

Panel (b) shows an equivalent experiment to map the 3' terminus of the message. In both panels MI stands for mock-infected RNA and L is late, infected-cell, mRNA.

```
GAGCCCCAGCCCGCTCCGCTTCTCGCCCCAGACGGCCCGTCGAGTGAAAACTTCCGTACCCAGACAATAAAGCACCAACAGGGGTTCATTCGGTGTTGGCGTTGCGTGCCTTTG  114

TTTCCCAATCCGACGGGGACCGGGACTGGGTGGCGGGGGGTGGGTTGGACAGCCGCCCTCGGTTCGCCTTCACGTGACAGGAGCCAATGTGGGGGGAAGTCACGAGGTACGGGG  228

                                                    O--------->
CGGCCCGTGCGGGTTGCTTAAATGCGTGGTGGCGACCACGGGCTGTCATTCCTCGGGAACGGACGGGGTTCCCGCTGCCCACTTCCCCCCATAAGGTCCGTCCGGTCCTCTAAC  342

                                                                                                                    0
GCGTTTGGGGGTTTTCTCTTCCCGCGCCGTCGGGCGTCCCACACTCTCTGGGCGGGCGGGGACGATCGCATCAAAAGCCCGATATCGTCTTTCCCGTATCAACCCCACCCA  453

 M  D  L  L  V  D  E  L  F  A  D  M  N  A  D  G  A  S  P  P  P  P  R  P  A  G  G  P                              28
ATG GAC CTC TTG GTC GAC GAG CTG TTT GCC GAC ATG AAC GCG GAC GGC GCT TCG CCA CCG CCC CCC CGC CCG GCC GGG GGT CCC        537

 K  N  T  P  A  A  P  P  L  Y  A  T  G  R  L  S  Q  A  Q  L  M  P  S  P  P  M  P  V                              56
AAA AAC ACC CCG GCG GCC CCC CCG CTG TAC GCA ACG GGG CGC CTG AGC CAG GCC CAG CTC ATG CCC TCC CCA CCC ATG CCC GTC        621

 P  P  A  A  L  F  N  R  L  L  D  D  L  G  F  S  A  G  P  A  L  C  T  M  L  D  T  W                              84
CCC CCC GCC GCC CTC TTT AAC CGT CTC CTC GAC GAC TTG GGC TTT AGC GCG GGC CCC GCG CTA TGT ACC ATG CTC GAT ACC TGG        705

 N  E  D  L  F  S  A  L  P  T  N  A  D  L  Y  R  E  C  K  F  L  S  T  L  P  S  D  V                              112
AAC GAG GAT CTG TTT TCG GCG CTA CCG ACC AAC GCC GAC CTG TAC CGG GAG TGT AAA TTC CTA TCA ACG CTG CCC AGC GAT GTG        789

 V  E  W  G  D  A  Y  V  P  E  R  T  Q  I  D  I  R  A  H  G  D  V  A  F  P  T  L  P                              140
GTG GAA TGG GGG GAC GCG TAC GTC CCC GAA CGC ACC CAA ATC GAC ATT CGC GCC CAC GGC GAC GTG GCC TTC CCT ACG CTT CCG        873

 A  T  R  D  G  L  G  L  Y  Y  E  A  L  S  R  F  F  H  A  E  L  R  A  R  E  E  S  Y                              168
GCC ACC CGC GAC GGC CTC GGG CTC TAC TAC GAA GCG CTC TCT CGT TTC TTC CAC GCC GAG CTA CGG GCG CGG GAG GAG AGC TAT        957

 R  T  V  L  A  N  F  C  S  A  L  Y  R  Y  L  R  A  S  V  R  Q  L  H  R  Q  A  H  M                              196
CGA ACC GTG TTG GCC AAC TTC TGC TCG GCC CTG TAC CGG TAC CTG CGC GCC AGC GTC CGG CAG CTG CAC CGC CAG GCG CAC ATG       1041

 R  G  R  D  R  D  L  G  E  M  L  R  A  T  I  A  D  R  Y  Y  R  E  T  A  R  L  A  R                              224
CGC GGA CGC GAT CGC GAC CTG GGA GAA ATG CTG CGC GCC ACG ATC GCG GAC AGG TAC TAC CGA GAG ACC GCT CGT CTG GCG CGT       1125

 V  L  F  L  H  L  Y  L  F  L  T  R  E  I  L  W  A  A  Y  A  E  Q  M  M  R  P  D  L                              252
GTT TTG TTT TTG CAT TTG TAT CTA TTT TTG ACC CGC GAG ATC CTA TGG GCC GCG TAC GCC GAG CAG ATG ATG CGG CCC GAC CTG       1209

 F  D  C  L  C  C  D  L  E  S  W  R  Q  L  A  G  L  F  Q  P  F  M  F  V  N  G  A  L                              280
TTT GAC TGC CTC TGT TGC GAC CTG GAG AGC TGG CGT CAG TTG GCG GGT CTG TTC CAG CCC TTC ATG TTC GTC AAC GGA GCG CTC       1293

 T  V  R  G  V  P  I  E  A  R  R  L  R  E  L  N  H  I  R  E  H  L  N  L  P  L  V  R                              308
ACC GTC CGG GGA GTG CCA ATC GAG GCC CGC CGG CTG CGG GAG CTA AAC CAC ATT CGC GAG CAC CTT AAC CTC CCG CTG GTG CGC       1377

 S  A  A  T  E  E  P  G  A  P  L  T  T  P  P  T  L  H  G  N  Q  A  R  A  S  G  Y  F                              336
AGC GCG GCT ACG GAG GAG CCA GGG GCG CCG TTG ACG ACC CCT CCC ACC CTG CAT GGC AAC CAG GCC CGC GCC TCT GGG TAC TTT       1461

 M  V  L  I  R  A  K  L  D  S  Y  S  S  F  T  T  S  P  S  E  A  V  M  R  E  H  A  Y                              364
ATG GTG TTG ATT CGG GCG AAG TTG GAC TCG TAT TCC AGC TTC ACG ACC TCG CCC TCC GAG GCG GTC ATG CGG GAA CAC GCG TAC       1545

 S  R  A  P  T  K  N  N  Y  G  S  T  I  E  G  L  L  D  L  P  D  D  D  A  P  E  E  A                              392
AGC CGC GCG CGT ACG AAA AAC AAT TAC GGG TCT ACC ATC GAG GGC CTG CTC GAT CTC CCG GAC GAC GAC GCC CCC GAA GAG GCG       1629

 G  L  A  A  P  R  L  S  F  L  P  A  G  H  T  R  R  L  S  T  A  P  P  T  D  V  S  L                              420
GGG CTG GCG GCT CCG CGC CTG TCC TTT CTC CCC GCG GGA CAC ACG CGC AGA CTG TCG ACG GCC CCC CCG ACC GAT GTC AGC CTG       1713

 G  D  E  L  H  L  D  G  E  D  V  A  M  A  H  A  D  A  L  D  D  F  D  L  D  M  L  G                              448
GGG GAC GAG CTC CAC TTA GAC GGC GAG GAC GTG GCG ATG GCG CAT GCC GAC GCG CTA GAC GAT TTC GAT CTG GAC ATG TTG GGG       1797

 D  G  D  S  P  G  P  G  F  T  P  H  D  S  A  P  Y  G  A  L  D  M  A  D  F  E  F  E                              476
GAC GGG GAT TCC CCG GGG CCG GGA TTT ACC CCC CAC GAC TCC GCC CCC TAC GGC GCT CTG GAT ATG GCC GAC TTC GAG TTT GAG       1881

 Q  M  F  T  D  A  L  G  I  D  E  Y  G  G  -                                                                       491
CAG ATG TTT ACC GAT GCC CTT GGA ATT GAC GAG TAC GGT GGG TAG GGGGCGCGACCGGACCCGCATCCCCCGTCTGGGTTTTCCCCTCCCGTCACCGGT  1980

                                                 --------X
TCGTATCCACAATAAACACGAGCACATACATTACAAAACCTGCGGTTGTCGTCTGATTATTTGGTGGTGGGGGAAAGAACTAGCCAGGAGACGGGACCGCGCAACCAACCCACT  2094

GGGGTCTGGGTTGCCGGCGTGTGTTAGCCGCGTCTGCGGGCCTGTCGTGTAGATTCGAAACCACGGACGGGTGATTGTGTCGCAGGGCGGCCCGCGTATAAAGGCGAGAGCG  2208

CGGGACCGTTTCCGCATTTGGCCGGGGGCTGGGGCGGCGGGTAGCCTTCGCGGGAGATACTGCGTTTTTTTGCGCCGGCCCCGTCGCTCCCGTCCATTCCCATCGCGAGGGGGT  2322

CCGGCGGCACCTACCCCGGCCTCCATCCGCGCTGTGGGGTCTTTTTCTTTTTTGGGGGGTAGCGGACATCCGATAACCCGCGTCTATCGCCACCATGTCGGCTCGCGAACCCGC  2436

GGGGCGCAGGAGGCGCGCATCCACCCGCCCCCGCGCCTCGCCCGTGGCGGACGAGCCAGCGGGCGATGGGGTGGGGTTCATGGGGTACCTGCGTGCGGTGTTCCGCGGGGATGA  2550

CGACAGCGAGCTAGAGGCTCTGGAGGAGATGGCGGGCGACGAGCCGCCCGTGCGCCGTC                                                        2609
```

0.685 on the prototype HSV-1 genome, in a region bounded by SalI and XhoI sites (see figure 1 and reference 26). The length of this fragment of the genome is 2609 base pairs. Approximately 30,000 characters were present in the database and greater than 98.5% of the sequence was determined on both strands. The base composition is 66.0% G+C.

We have accurately located the 5' and 3' termini of the single, 1.7kb, mRNA which is known to map entirely within this region (27). The results are presented in figures 2 and 3. The 5' end of the mRNA was mapped using two independent probes end-labelled at SalI and MluI sites respectively. The MluI probe gave a hybrid band of about 67bp in size, consistent with the hybrid of approximately 200bp obtained using the SalI probe. These data position the mRNA start site at approximate position 274. The mRNA 3' terminus was mapped using a 3' end-labelled XcyI probe which gave an S1 hybrid band of about 205bp, positioning the end of the mRNA at approximately 2020. Note that the sequence presented is numbered right to left (5' to 3') with respect to the genomic orientation.

In the 5' flanking regions it is possible to discern various sequence motifs which are similar to sequences implicated in control of transcription. These are as follows:

(a) TTAAAT, at position 246, which is a reasonable candidate for the "TATA box" (45), a sequence important for efficient and accurate initiation of RNA polymerase II transcription. It is in approximately the correct position, about -30 with respect to the cap site.

(b) AGCCAATGT, at position 196, an excellent match to a sequence, GGCCAATCT, implicated in the transcriptional activity of globin promoters (46, 47) but also found in many others (48).

Figure 3: The sequence of the gene for Vmw65

We present here 2609 base pairs of DNA sequence including the gene for Vmw65. The initiation site of the transcript for this gene is shown as O-----. The terminus of the mRNA is marked as -----X. Canonical sequences implicated in the control of transcription in other systems, and referred to individually in the text, are underlined. The single letter amino acid code is used to indicate the extent of the open reading frame for Vmw65.

(c) GC rich motifs similar to those found in both the SV40 21bp repeats and the HSV-1 TK gene (49, 50, 51). Specifically, CCGCCC at position 167 and GGGCGG at position 226, are in approximately similar places relative to the "TATA box" as those elements found in the TK promoter.

On the 3' side of the coding region lies a typical polyadenylation signal sequence AATAAA (52), at position 1991, followed about 30bp downstream by a sequence TGCGCTTG. This is consistent with a consensus (YGTGTTYY, where Y=pyrimidine) found in this position in many genes and shown to be functionally important in the termination of HSV mRNA (53,54).

The first initiating ATG codon to be found after the mRNA start was at position 454 and conforms satisfactorily with the rules of Kozak (55). This codon initates an open reading frame of 490 amino acids whose codon usage was consistent with that of other, known, HSV-1 proteins as determined by the program of Staden and McLachlan (39).

The DNA sequence of Fig.3 also contains what we consider to be the termini of the genes flanking the gene for Vmw65. At position 66 is a polyadenylation signal, presumably that for two, 3' co-terminal, messages coding for a 42,000 molecular weight polypeptide located upstream of Vmw65 (27). Hall and co-workers also indicate that a 70,000 molecular weight polypeptide should be translated from a mRNA initiating soon after the end of the mRNA for Vmw65. We believe that this protein begins at position 2417 and there is a reasonable "TATA" box at 2193.

The Vmw65 Protein

From the above predicted open reading frame, Vmw65 is composed of 490 amino acid residues and has a molecular weight of 54,342. The amino acid composition is presented in Table 1. The protein is highly charged, with the acidic residues, aspartic and glutamic acid, accounting for 14.3% of the total amino acids. Those amino acid residues with strongly basic side chains, arginine and lysine, make up a further 9%. The predicted net acidic character is consistent with two-dimensional gel electrophoresis of HSV-1 proteins, which has shown Vmw65 to be amongst the most acidic proteins in infected cells (56, 57).

Table 1: Predicted amino acid composition of Vmw65

Residue	No.	%	Residue	No.	%	Residue	No.	%
Ala	58	11.8	Arg	40	8.2	Asn	12	2.4
Asp	41	8.4	Cys	6	1.2	Gln	10	2.0
Glu	29	5.9	Gly	32	6.5	His	13	2.7
Ile	9	1.8	Leu	65	13.3	Lys	4	0.8
Met	16	3.3	Phe	23	4.7	Pro	40	8.2
Ser	26	5.3	Thr	27	5.5	Trp	4	0.8
			Val	17	3.5			

Analysis of hydropathicity shows that there are no extensive regions of an extreme hydrophobic or hydrophilic nature (data not shown). However, a histogram (figure 4) of the numbers of basic or acidic residues every 30 amino acids reveals that there is a distinct polarity along the peptide backbone, with acidic residues prevalent in the carboxy-terminal 80 amino acids and

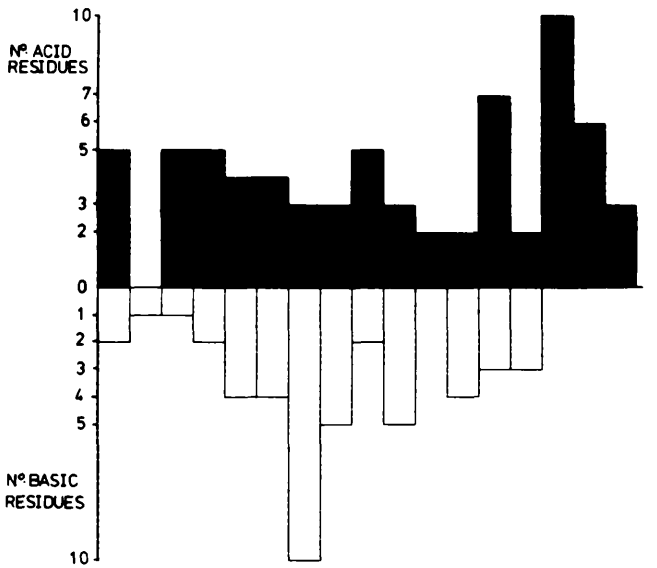


Figure 4: Histogram of acid/basic residue distribution
The distribution of acidic and basic residues along the polypeptide backbone is plotted. The interval is 30 amino acids.

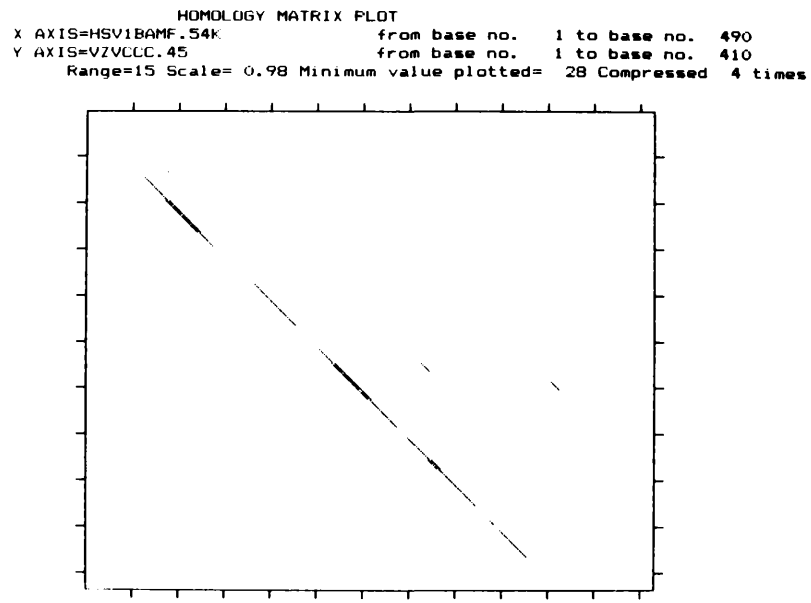


Figure 5: A dot matrix plot of homology between Vmw65 and VZV 45K

This plot was obtained by use of the program of Pustell and Kafatos (41) with the parameters indicated in the figure. Graphical output was programmed by Philip Taylor. It clearly shows colinearity between the two amino acid sequences except for the carboxy terminal 80 amino acid residues of Vmw65.

basic residues concentrated near the amino-terminal region. The relatively high leucine content may indicate that Vmw65 has a well defined secondary structure.

Homology to other proteins

We have detected a clear homologue to Vmw65 in the genome of varicella-zoster virus (VZV), another human herpesvirus (A.Davison, manuscript in preparation). The VZV protein has a molecular weight of 45,000 (45K). Figure 5 shows a dot matrix homology plot between the two proteins. It can be seen that they are approximately colinear except that the HSV-1 protein has 80 amino acid residues at the carboxy-terminus which the VZV 45K lacks. Figure 6 is an optimal alignment of the two amino acid sequences. This indicates that approximately 35% of amino acid residues are conserved between the two proteins. Since the VZV

```
  1 MD  LLVDELFADMNADGASPPPPRPA  GGPKNTPAAPPLYATGRLSQA QLMPSPPMPVPPAALFNRLLDDL
    *   *     *          *          *       *   * *   ****    *  *   *  **
  1 MECNLGTEHPSTDTWNRSKTEQAVVDAFDESLFGDVASDIGFETSLYSHAVKTAPSPPWVASPKILYQQLIRDL

 70 GFSAGPALCTMLDTWNEDLFSALPTNADLYRECKFLSTLPSDV   VEWGDAYVPERTQIDIRAHGDVAFPTLP
    ** ** *   * ********  * * ***     **  * **   *   *     *       *  * * *  *
 75 DFSEGPRLLSCLETWNEDLFSCFPINEDLYSDMMVLSPDPDDVISTVSTKD  HVEMFNLTTR  GSVRLPSPP

141 ATRDGLGLYYEALSRFFHAELRAREESYRTVLANFCSALYRYLRASVRQLHRQAHMRGRDRDLGEMLRATIADR
      **  *       *  ******* *   *   *    ***                *      **  *  *
145 KQPTGLPAYVQEVQDSFTVELRAREEAYTKLLVTYCKSIIRYLQGTAKRTTIGLNIQNPDQKAYTQLRQSILLR

215 YYRETARLARVLFLHLYLFLTREILWAAYAEQMMRPDLFDCLCCDLESWRQLAGLFQPFMFVNGALTVRGVPIE
    **** * *** * *****  ***  *  ** *   ** *  *        **    * *     *     * *
219 YYREVASLARLLYLHLYLTVTREFSWRLYASQSAHPDVFAALKFTWTERRQFTCAFHPVLCNHGIVLLEGKPLT

289 ARRLRELNHIREHLNLPLVRSAATEEPGAPLTTPPTLHGNQARASGYFMVLIRAKLDSYSSFTTSPSEAVMREH
    *  *** *  *  * *****    **   **   *       *  *     *  *** *        *   *
293 ASALREINYRRRELGLPLVRCGLVEENKSPLVQQPSFSVHLPRSVGFLTHHIKRKLDAYAVKHPQEPRHVRADH

363 AYSRARTKNNYGSTIEGLLDLPDDDAPEEAGLAAPRLSFLPAGHTRRLSTAPPTDVSLGDELHLDGEDVAMAHA
     *     **** *            *  ** *  *    *   *    ***   *
367 PYAKVVENRNYGSSI            EAMILAPPSPSEILPGDPPR    PPTCGFLTR

437 DALDDFDLDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFTDALGIDEYGG
409
```

Figure 6: An optimal alignment of the amino acid sequences of Vmw65 and VZV 45K

Obtained using the optimal alignment program of Taylor (42) and using the default parameters recommended by the author. The top line is the amino acid sequence of Vmw65.

and HSV-1 genomes have diverged considerably, to the extent that their base compositions differ by 21 percentage points, we regard this homology as a good confirmation of the correctness of our Vmw65 reading frame assignment. We have been unable to detect a homologue to Vmw65 in another herpesvirus, Epstein Barr virus, whose genome has been completely sequenced (58).Similarly, a search of the National Biological Research Foundation protein database produced no matches of obvious significance.

DISCUSSION

The results of Campbell and co-workers (26), using transfection of plasmid clones, preclude the possibilities either that the tegument's structural integrity is involved in

IE gene stimulation or that other viral proteins play an essential part. Since no other open reading frames of significant length, with acceptable codon usage, were detected on either strand we conclude that the polypeptide described here is alone responsible for specific trans-activation of HSV-1 IE promoters. The difference between the molecular weight determined by SDS-polyacrylamide gel electrophoresis and that obtained by sequence analysis cannnot be due to phosphorylation since the in vitro translation product is of similar electrophoretic mobility (26). We believe that it is more likely that the large number of charged residues result in aberrant electrophoretic mobility. The first AUG codon may not be the sole initiating codon in vivo, since experiments involving hybrid-arrest translation, in vitro, followed by two-dimensional gel electrophoresis have detected a minor polypeptide which probably results from initiation of translation at the second AUG codon, 33bp further downstream (C.M.Preston and L.Haarr, unpublished observations). This smaller polypeptide is capable of trans-activation, since plasmids possessing an 8bp oligonucleotide insert at the SalI site between the two ATG codons are active in the co-transfection assay (26). Polypeptides resulting from translation initiation at downstream, in phase, AUG codons of the HSV thymidine kinase mRNA have been previously described (57,59).

The VZV 45K polypeptide is homologous to HSV Vmw65 but lacks the carboxy-terminal 80 amino acids, which are relatively rich in acidic residues. It will be interesting to determine whether the VZV homologue and Vmw65 can stimulate expression of the heterologous IE genes. The VZV gene homologous to HSV-1 IE3 does not exhibit sequences corresponding to the TAATGARATTC consensus in its 5'-flanking region (60). Therefore, either the VZV 45K protein recognises a different upstream sequence or it may lack the ability to stimulate transcription.

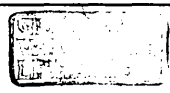
The function of the viral tegument is unknown. Vmw65 comprises a significant proportion of the total protein in the tegument and is likely to play an important structural role. The tegument may be instrumental in the maturation of virus particles by directing the envelopment of nucleocapsids. Two

possible mechanisms by which this could be brought about are either by direct interaction with the lipid component of the envelope or, indirectly, by complexing with virally encoded envelope proteins. There is no obvious sequence evidence, in the form of a large hydrophobic domain, for direct membrane interaction. It is known that Vmw65 is tightly complexed with glycoprotein B and possibly other viral glycoproteins which reside in the envelope (M.Campbell, personal communication). Another possible role for the protein may be in maintaining the structural integrity of the tegument. Our analysis of the distribution of acidic/basic residues indicates a potential polarity to the protein molecule. Such a polarity might allow Vmw65 to be arranged, in the tegument, in an ordered fashion and result in a structure held in place by hydrogen bonding interactions. It is thought that the tegument is especially sensitive to dehydration (61) which is compatible with such a mechanism. Other functions we can postulate for Vmw65 would be important in the early stages of infection, for example the tegument might provide a protective layer around the nucleocapsid or it may play a role in the virion's nuclear tropism.

In conclusion, we present here the DNA and implied protein sequences of a gene from HSV-1 whose product specifically stimulates transcription from HSV-1 IE promoters. The predicted protein is notably acidic. It shows homology to an open reading frame in the VZV genome, but no significant homology to any other known amino acid sequence. The sequence will allow detailed molecular genetic analyses to define functionally important regions of the polypeptide. Further immunological and biochemical experiments may provide insights into how this protein specifically stimulates HSV IE transcription and into the mechanisms of promoter recognition.

ACKNOWLEDGEMENTS

We thank Professor J. H. Subak-Sharpe for reading the manuscript and Dr. D. Bzik for his help and interest in the work. Thanks are also due to Dr. A. Coulson and Dr. P. Taylor for their help in setting up protein database searches. M.D. was a recipient of a Medical Research Council Research Training Award.



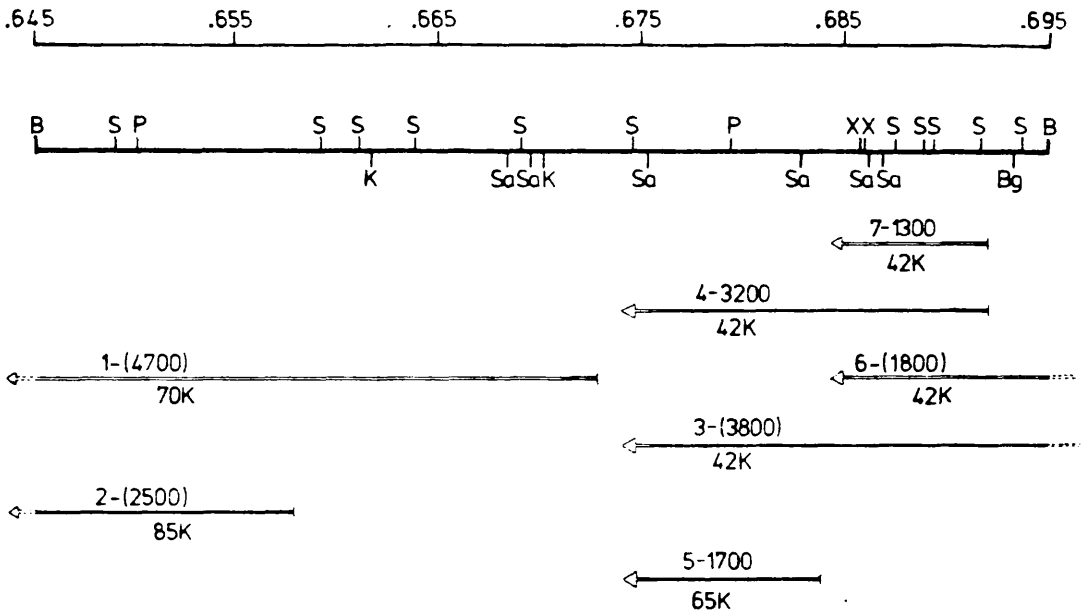
+ Members of the MRC Virology Unit.

LITERATURE CITED

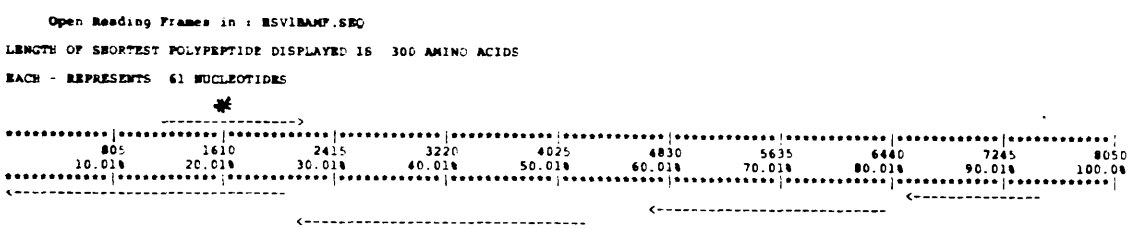
1. Wildy, P., Russell, W.C. and Horne, R.W. (1960) Virology 12, 204-224.
2. Becker, Y., Dym, H. and Sarov, I. (1968) Virology 36, 184-192.
3. Graham, B.J., Ludwig, H., Bronson, D.L., Benyesh-Melnick, M. and Biswal, N. (1972) Biochem. Biophys. Acta 259, 13-23.
4. Honess, R.W. and Roizman, B. (1974) J. Virol. 14, 8-19.
5. Swanstrom, R.I., Piro, K. and Wagner, E.K. (1975) Virology 66, 140-150.
6. Clements, J.B., Watson, R.J. and Wilkie, N.M. (1977) Cell 12, 275-285.
7. Jones, P.C. and Roizman, B. (1979) J. Virol. 31, 299-314.
8. Preston, C.M. (1979) J. Virol. 29, 275-285.
9. Watson, R.J. and Clements, J.B. (1980) Nature 285, 329-330.
10. Everett, R.D. (1984) EMBO J. 3, 3135-3141.
11. O'Hare, P. and Hayward, G.S. (1985) J. Virol. 53, 751-760.
12. Quinlan, M.P. and Knipe, D.M. (1985) Mol. Cell. Biol. 5, 957-963.
13. Post, L.E., Mackem, S. and Roizman, B. (1981) Cell 24, 555-565.
14. Mackem, S. and Roizman, B. (1982) Proc. Natl. Acad. Sci. USA 79, 4917-4921.
15. Cordingley, M.G., Campbell, M.E.M. and Preston, C.M. (1983) Nucleic Acids Res. 11, 2347-2365.
16. Lang, J.C., Spandidos, D. and Wilkie, N.M. (1984) EMBO J. 3, 389-395.
17. Preston, C.M. and Tannahill, D. (1984) Virology 137,439-444.
18. Preston, C.M., Cordingley, M.G. and Stow, N.D. (1984) Virology 50, 708-716.
19. Puga, A., Gomez-Marquez, J., Brayton, P.R., Cantin, E.M., Long, L.K., Barbacid, M. and Notkins, A.L. (1985) J. Virol. 54, 879-881.
20. Mackem, S. and Roizman, B. (1982) J. Virol. 44, 939-949.
21. Kristie, T.M. and Roizman, B. (1984). Proc. Natl. Acad. Sci. USA 81, 4065-4069.
22. Murchie, M.J. and McGeoch, D.J. (1982) J. Gen. Virol. 62, 1-15.
23. Whitton, J.L., Rixon, F.J., Easton, A.E. and Clements, J.B. (1983) Nucleic Acids Res. 11, 6271-6287.
24. Whitton, J.L. and Clements, J.B. (1984) Nucleic Acids Res. 12, 2061-2079.
25. Batterson, W. and Roizman, B. (1983) J. Virol. 46, 371-377.
26. Campbell, M.E.M., Palfreyman, J.W. and Preston, C.M. (1984) J. Mol. Biol. 180, 1-19.
27. Hall, L.M., Draper, K.G., Frink, R.J., Costa, R.H. and Wagner, E.K. (1982) J. Virol. 43, 594-607.
28. Marsden, H.S., Stow, N.D., Preston, V.G., Timbury, M.C. and Wilkie, N.M. (1978) J. Virol. 28, 624-642.
29. Lemaster, S. and Roizman, B. (1979) J. Virol. 35, 798-811.
30. Twigg, A.J. and Sherratt, D. (1980) Nature 283, 216-218.
31. Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.

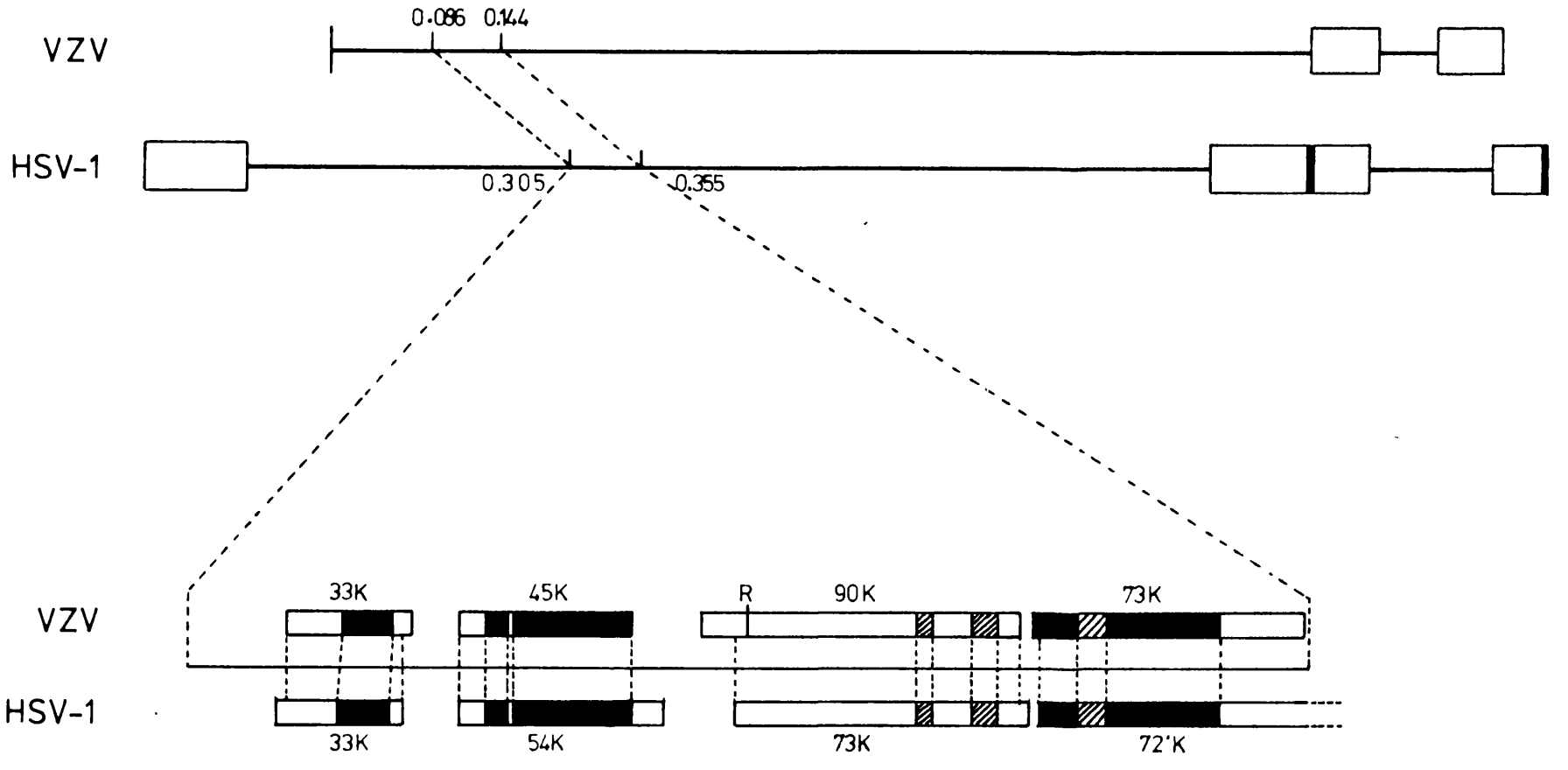
32. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory CSH.
33. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
34. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
35. Deininger, P.L. (1983) Anal. Biochem. 129, 216-223.
36. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
37. Staden, R. (1982) Nucleic Acids Res. 10, 4731-4751.
38. Blumenthal, R.M., Rice, P.J. and Roberts, R.J. (1982) Nucleic Acids Res. 10, 91-101.
39. Staden, R. and McLachlan, A.D. (1982) Nucleic Acids Res. 10, 141-156.
40. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
41. Pustell, J. and Kafatos, F.C. (1982) Nucleic Acids Res. 10, 4765-4782.
42. Taylor, P. (1984) Nucleic Acids Res. 12, 447-456.
43. Preston, C.M. (1977) Virology 78, 349-353.
44. Rixon, F.J. and Clements, J.B. (1982) Nucleic Acids Res. 10, 2241-2256.
45. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science (Wash.) 209, 1409-1414.
46. Dierks, P., Van Ooyen, A., Mantel, N. and Weissmann, C. (1981) Proc. Natl. Acad. Sci. USA 78, 1411-1415.
47. Grosveld, G.C., de Boer, E., Shewmaker, C.K. and Flavell, R.A. (1982) Nature 295, 120-126.
48. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142.
49. Everett, R.D., Baty, D. and Chambon, P. (1983) Nucleic Acids Res. 11, 2447-2464.
50. McKnight, S.L., Gavis, E.R., Kingsbury, R. and Axel, R. (1981) Cell 25, 385-398.
51. McKnight, S.L., Kingsbury, K.C., Spence, A. and Smith, M. (1984) Cell 37, 253-262.
52. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260.
53. Taya, Y., Devos, R., Tavernier, J., Cheroutre, H., Engler, G. and Fiers, W. (1982) EMBO J. 1, 953-958.
54. McLauchlan, J., Gaffney, D., Whitton, J.L. and Clements, J.B. (1985) Nucleic Acids Res. 13, 1347-1368.
55. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
56. Haarr, L. and Marsden, H.S. (1981) J. Gen. Virol. 52, 77-92.
57. Marsden, H.S., Haarr, L. and Preston, C.M. (1983) J. Virol. 46, 434-445.
58. Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Seguin, C., Tuffnell, P.S. and Barrell, B.G. (1984) Nature 310, 207-211.
59. Preston, C.M. and McGeoch, D.J. (1981) J. Virol. 38, 593-605.
60. Davison, A.J. and Scott, J.E. (1985) J. Gen. Virol. 66, 207-220.
61. Vernon, S.K., Lawrence, W.C., Long, C.A., Rubin, B.A. and Sheffield, J.B. (1982) J. Ultrastruct. Res. 81, 163-171.

(a)


.645
.655
.665
.675
.685
.695
B S P S S S S S P XX S SS S S B
K Sa Sa K Sa Sa Sa Bg
7-1300
42K
4-3200
42K
1-(4700)
70K
6-(1800)
42K
3-(3800)
42K
2-(2500)
85K
5-1700
65K

(b)


Open Reading Frames in : HSV1BAMF.SEQ
LENGTH OF SHORTEST POLYPEPTIDE DISPLAYED IS 300 AMINO ACIDS
EACH - REPRESENTS 61 NUCLEOTIDES
805 1610 2415 3220 4025 4830 5635 6440 7245 8050
10.01% 20.01% 30.01% 40.01% 50.01% 60.01% 70.01% 80.01% 90.01% 100.0%


VZV
0.086
0.144
HSV-1
0.305
0.355
VZV
33K
45K
R
90K
73K
HSV-1
33K
54K
73K
72'K

Codon Usage Table of the Coding Sequences of Vmw65

TTT	Phe	12	2.4%	TCT	Ser	3	0.6%	TAT	Tyr	3	0.6%	TGT	Cys	3	0.6%
TTC	Phe	11	2.2%	TCC	Ser	6	1.2%	TAC	Tyr	15	3.1%	TGC	Cys	3	0.6%
TTA	Leu	1	0.2%	TCA	Ser	1	0.2%	TAA	---	0	0.0%	TGA	---	0	0.0%
TTG	Leu	12	2.4%	TCG	Ser	6	1.2%	TAG	---	1	0.2%	TGG	Trp	4	0.8%
CTT	Leu	3	0.6%	CCT	Pro	2	0.4%	CAT	His	3	0.6%	CGT	Arg	6	1.2%
CTC	Leu	16	3.3%	CCC	Pro	21	4.3%	CAC	His	10	2.0%	CGC	Arg	19	3.9%
CTA	Leu	8	1.6%	CCA	Pro	4	0.8%	CAA	Gln	1	0.2%	CGA	Arg	2	0.4%
CTG	Leu	25	5.1%	CCG	Pro	13	2.6%	CAG	Gln	9	1.8%	CGG	Arg	11	2.2%
ATT	Ile	4	0.8%	ACT	Thr	0	0.0%	AAT	Asn	1	0.2%	AGT	Ser	0	0.0%
ATC	Ile	5	1.0%	ACC	Thr	17	3.5%	AAC	Asn	11	2.2%	AGC	Ser	10	2.0%
ATA	Ile	0	0.0%	ACA	Thr	0	0.0%	AAA	Lys	3	0.6%	AGA	Arg	1	0.2%
ATG	Met	16	3.3%	ACG	Thr	10	2.0%	AAG	Lys	1	0.2%	AGG	Arg	1	0.2%
GTT	Val	1	0.2%	GCT	Ala	5	1.0%	GAT	Asp	11	2.2%	GGT	Gly	3	0.6%
GTC	Val	8	1.6%	GCC	Ala	26	5.3%	GAC	Asp	30	6.1%	GGC	Gly	9	1.8%
GTA	Val	0	0.0%	GCA	Ala	1	0.2%	GAA	Glu	6	1.2%	GGA	Gly	7	1.4%
GTG	Val	8	1.6%	GCG	Ala	26	5.3%	GAG	Glu	23	4.7%	GGG	Gly	13	2.6%

Average Base Composition of the Coding Sequences of Vmw65

	Y	T	C	A	G	R	ALL
NUMBER	782	271	511	243	448	691	1473
PERCENT	53.1	18.4	34.7	16.5	30.4	46.9	100.0