ANALYSIS OF THE TRANSCRIPTS OF BOVINE PAPILLOMAVIRUS TYPE 4.

Alasdair C. Stamps.

Thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy.

Beatson Institute for Cancer Research, Glasgow.

December, 1987.

CONTENTS

ACKNOWLEDGEMENTS

Much credit is due for my work on this thesis to the members of Group R6 of the Beatson Institute for their help and support, in particular my supervisor, Dr. M. Saveria Campo, and also Dr. Kenneth T. Smith.

I would also like to acknowledge the provision of tumour samples by Glasgow University Veterinary School; synthesis of a cDNA library by Anne Sproul (Beatson Institute); and synthesis of oligonucleotides by Dr.Veer Math of Glasgow University Biochemistry Department.

Many thanks also to my fellow Ph.D. students at the Beatson Institute, particularly John Chester for helpful discussions.

Finally, I am overwhelmingly indebted to Virginia Greenwood, BVMS(Hons), MRCVS, for continuous support, patience and tolerance over the past year.

ABBREVIATIONS

A	adenine
bp	base pairs
BSA	bovine serum albumin
C	cytosine
cm	centimetre
cDNA	complementary DNA
DNA	deoxyribonucleic acid
ds	double stranded
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	~~deoxy~~thymidine triphosphate
dNTP	deoxynucleotide triphosphate (in reference to the four above).
dd-	dideoxy-
DEPC	diethylpyrocarbonate
EDTA	ethylenediaminetetra-acetic acid
G	guanine
g	grammes
HEPES	N-2-hydroxyethyl piperazine-N'-2-ethane sulphonic acid
hr	hour
kb	kilobase pairs
kD	kiloDaltons
M	molar
mM	millimolar
M_r	molecular weight

mm	millimetres
ml	millilitres
mg	milligrammes
min	minutes
mRNA	messenger RNA
MOPS	3-(N-morpholino)propane sulphonic acid
nm	nanometres
ng	nanogrammes
nM	nanomolar
nt	nucleotides
$O.D._{260nm}$	optical density at wavelength 260 nm
orf	open reading frame
pers. comm.	personal communication
poly(A)$^+$	polyadenylated
RNA	ribonucleic acid
tRNA	transfer RNA
sec	seconds
SDS	sodium dodecyl sulphate
T	thymine
Tris	tris(hydroxymethyl)aminomethane
TEMED	N,N,N'N'-tetramethylethylenediamine
V	volts
% v/v	volume in ml per 100 ml water
% w/v	weight in g per 100 ml water
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
µg	microgrammes
µl	microlitres
µM	micromolar

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The transcripts of bovine papillomavirus type 4 (BPV4) were analysed through cDNA cloning and sequencing. Two cDNA's of particular interest were analysed. cDNA 7E11 was found to represent the 1 kb major transcript of BPV4. Primer extension and S1 nuclease protection mapping showed that this transcript had a multiple initiation site within the E6 orf region, with the major start site at nt870. cDNA and RNA sequencing showed that the transcript was spliced from nt1016 to nt3376. The polyadenylated 3' terminus was at nt4034, defining the early polyadenylation site of BPV4.

Sequencing of cDNA's and RNA revealed alterations to the published genomic sequence (Patel et al., 1987) which changed the genomic organisation of BPV4. cDNA 7E11 was shown to encode a fusion orf in which the E4 orf was linked to an upstream ATG codon. Thus, as in other papillomaviruses, the major transcript of BPV4 encodes one of the E4 proteins (Doorbar et al., 1986). Orf homology analysis revealed that the BPV4 E4 orf is most like those of the carcinogenic HPV's 16, 18, 33 and 8. Since the E4 proteins are thought to be involved in the subversion of normal keratinocyte differentiation, this may be a significant finding.

In addition, rare transcripts represented by cDNA Q were mapped by RNA sequencing and S1 nuclease protection. A 1kb transcript, which may be spliced, extends 5' to nt2834 and minor 5' termini at nt3063, nt3093 and nt3151 may have represented transcripts encoding the putative E2 repressor of BPV4.

INTRODUCTION : Contents.

1.1. Papillomaviruses: general.

1.1.1. Properties of papillomaviruses.

Papillomaviruses, a genus of the family Papovaviridae, are the aetiological agents of papillomas, the limited epithelial and, in some cases, fibroepithelial proliferations of the skin or mucosa in mammals and birds commonly known as warts (Pfister, 1984; Giri & Danos, 1986).

Virions extracted from productive papillomas are symmetrical icosahedral units approximately 55 nanometers (nm) in diameter according to electron microscopic (e.m.) visualisation. They are composed of 72 capsomeres arranged around a double-stranded (ds) circular molecule of DNA of average size 8000 base pairs (bp) which exists in both the supercoiled and relaxed (nicked) form. Viral DNA in benign papillomas is episomal and exists at high copy numbers per cell. Acetylated cellular histone proteins have been found associated with bovine papillomaviruses types 1 and 2 (BPV1 and BPV2) and with human papillomavirus type 1 (HPV1) DNA's in a chromatin-like structure; these proteins comigrate with histones H2a, H2b, H3 and H4 in sodium dodecyl sulphate polyacrylamide gel electrophoresis (Pfister, 1984).

1.1.2. Papillomavirus types: host and tissue specificity.

Papillomaviruses are known to infect a wide range of mammals: humans, cattle, sheep, goats, deer, elk, horses, rabbits,

monkeys, pigs, opossums, elephants and Mastomys natalensis. A papillomavirus affecting the harvest mouse, Micromys minutus, has recently been discovered. One avian papillomavirus has been described (Pfister, 1984; Sundberg et al., 1987; Moreno-Lopez et al., 1984).

More than forty different types, and several subtypes of human papillomaviruses have been characterised so far (Table 1). A new type is defined if it has less than 50% DNA sequence homology with known papillomavirus types by liquid hybridisation analysis. Groups are defined which have less than 1% homology with other groups. This essentially arbitrary distinction appears to hold good with respect to site and histopathology of the lesions induced by different papillomaviruses (Pfister, 1984).

The more recently discovered HPV's have not yet been classified into groups, so it is more convenient to divide them into two general categories: those infecting skin and those infecting mucosa (Table 1b).

Six distinct types of bovine papillomaviruses have been characterised (Table 2); these are divided into two subgroups: A: fibropapillomaviruses, and B: true epitheliotropic papillomaviruses (Jarrett, 1985).

Although two members of the former group can infect heterologous hosts , e.g. BPV1 induces equine sarcoids and tumours in the hamster (Lancaster and Olson, 1982), papillomaviruses are truly host specific in that all known intergeneric infections are non-productive. They are also strikingly site specific. Infections even of heterologous

Table 1. Classification of HPV's.

GROUP	TYPE	LESION
A	HPV1	verruca plantaris
B	HPV2	verruca vulgaris
	HPV3	verruca plana
	HPV10	verruca plana; E.V.
C	HPV4	verruca vulgaris
D	HPV5	E.V. & immunosuppressed patients; skin carcinoma
	HPV8	E.V.; skin carcinoma
	HPV9	E.V.
	HPV12	E.V.
	HPV14	E.V.
	HPV15	E.V.
	HPV17	E.V.; skin carcinoma
E	HPV6	condyloma acuminatum; ano-genital carcinoma
	HPV11	condyloma acuminatum; laryngeal papilloma; CIN
	HPV13	focal epithelial hyperplasia
F	HPV7	"butchers' warts"
G	HPV16	condyloma acuminatum; CIN; genital cancer
	HPV18	Morbus-Bowen disease; genital cancer

from Pfister, 1984.

E.V. = Epidermodysplasia verruciformis
CIN = Cervical intraepithelial neoplasia

Table 1b. The relationship between HPV type and lesion site.

HPV TYPE	SITE OF LESION
	SKIN
HPV 1,2,3,4,7,10,31,33	common cutaneous warts
HPV 3,5,8,9,10,12,14,15, 17,19-29	epidermodysplasia verruciformis
	MUCOSAL TISSUE
HPV13	buccal cavity
HPV 6,11,30	larynx; genitals
HPV 16,18,33	genitals (cervix, penis)

from Campo, 1987.

Table 2. Classification of bovine papillomaviruses.

SUBGROUP	TYPE	LESION
A	BPV1	fibropapilloma of paragenital area
	BPV2	fibropapilloma of skin
	BPV5	fibropapilloma of teat
B	BPV3	papilloma of skin
	BPV4	papilloma of upper alimentary tract
	BPV6	papilloma of udder

from Jarrett, 1985.

sites within the same animal are usually non-permissive if at all successful (Jarrett et al., 1984). This specificity suggests a narrow-margined permissive system for papillomaviruses, an hypothesis supported by studies of their infective cycle .

1.1.3. Biology of papillomavirus infection.

The molecular biology of papillomaviruses may be more easily put in the context of their effects on the host by a description of the virus-productive infective cycle.

Papillomavirus infection may be experimentally induced by inoculating scarified epithelium with virus (Jarrett, 1985). In at least BPV1, BPV2 and Shope papillomavirus, epithelial trauma is an absolute requirement for successful infection, as apparently is epithelial contact with mesenchyme (Kreider and Bartlett, 1981). Presumably this accounts for the low infectivity of some papillomaviruses (Pfister, 1984). Consequent of successful inoculation are the five stages of papillomavirus infection defined for the cottontail rabbit papillomavirus (CRPV) in the domestic rabbit by Kreider (1980; Table 3).

Stage 0 is the "latency period" during which no apparent changes occur on the macroscopic level: the length of this varies between species, being from 3 to 18 months in man and from 3 to 8 weeks in rabbits. Molecular events taking place during this time are unknown; however it has been shown that cloned BPV1 DNA can reside in cultured cells without viral

Table 3. Stages in the progression of CRPV-induced tumours in the domestic rabbit.

STAGE	EVENT	DAYS POST-INFECTION	% PAPILLOMAS AFFECTED
0	Latency	0 - 14	100
I	Papillomas appear	14 - 21	100
II	Exponential growth	36 - 60	100
	Regression	60 - 90	10 - 40
III	Persistence	Indefinite	20 - 30
IV	Carcinoma	350 - 450	40 - 60

from Kreider and Bartlett, 1981

Table 4. Abattoir survey of alimentary papillomatosis in cattle.

CATTLE	INCIDENCE OF PAPILLOMAS,%	MORE THAN 3 PAPILLOMAS,%	MORE THAN 1 SITE,%
Normal	19	0.025	4
From high cancer farms	80	55	-
Cancerous cattle	96	90	65

from Jarrett et al., 1980

replication taking place (Meneguzzi et al., 1984), and latent papillomavirus infections are well documented. Of particular interest in this respect was a report by Maitland et al. (1987) in which HPV16 DNA was shown to exist in 41 % of biopsies of normal oral epithelium. This followed reports by the same group in which 33 % of normal cervical tissue biopsies were found to contain HPV16 DNA (Cox et al., 1986; Meanwell et al., 1987); in both cases the DNA was in episomal form. All these samples came from patients with no clinical history of HPV16 infection. Thus papillomavirus DNA may remain latent in infected cells for considerable periods before some undefined event(s) initiate its vegetative life cycle

It is assumed that transformation of one or more basal cells occurs at some time during the latent period, followed by proliferation of the epidermis resulting in the clinical appearance of papillomas at Stage 1.

Papillomas then enter a stage of exponential growth due to hyperproliferation of keratinocytes (Stage 2). In situ hybridisation of wart sections with tritiated papillomavirus cRNA probes suggested that extensive viral replication occurs in the stratum granulosum or keratohyalin layer, and may commence in the stratum spinosum (Orth et al., 1971); viral capsid antigens only appear in the stratum granulosum where virions may be seen in the nucleus, and in the stratum corneum, where they form crystalline arrays embedded in keratin (Pfister, 1984). Productively infected keratinocytes at this stage contain large numbers of viral genomes, as many

as 100,000 genome equivalents per cell (Campo *et al.*, 1985b). The mature papilloma consists of a number of tubular subunits, attached at their base and extending into tapered fronds with highly keratinised tips.

Mature papillomas either regress within a time period dependent on species and type of virus, or persist. The factor(s) which predispose a papilloma to regress are unknown, but regressors are characterised by massive infiltration of leukocytes into the basal and suprabasal layers, accompanied by a 50% reduction in tritiated thymidine uptake, which alone would effect the regression of the tumour. Persisting papillomas are said to be in Stage 3 (Kreider, 1980).

Fibropapillomas have a similar aetiology in the epithelium, but their initial site of proliferation is in the subjacent fibroblasts; fibroblastic proliferation precedes that of the epithelium, which tends to be less markedly hyperplastic than in purely epitheliotropic virus infection (Jarrett, 1985).

It is apparent from these studies of benign papillomas and from their strict host- and tissue specificity that the sole permissive system for papillomaviruses is differentiating epithelium. This has been confirmed by attempts to produce virions in heterologous hosts and tissues and *in vitro* (Butel, 1971; Pfister, 1984). Basal cells are the presumed site of initial infection and transformation, but viral replication and synthesis of viral capsid proteins only occur in the *stratum granulosum* , i.e. keratin-synthesising cells, and full differentiation is required for virion assembly.

Malignant progression (Stage 4) was first observed by Rous and Beard (1935) in the Shope papilloma-carcinoma complex of domestic rabbits. 40-60% of persisting papillomas undergo malignant conversion 12-15 months post infection, developing into invasive and highly metastatic carcinomas.Since then several more papillomaviruses have been implicated in carcinogenesis (Smith and Campo, 1985).

1.1.4. Papillomaviruses and cancer: the Shope papilloma-carcinoma complex in domestic rabbits.

The Shope, or cottontail rabbit papillomavirus (CRPV) was discovered by Richard Shope in 1933 in his investigations of horn-like papillomas on cottontail rabbits (Shope and Hurst, 1933). CRPV displays a relaxed host specificity, although intergeneric cross-infections are non-permissive. However, as further work demonstrated, neoplastic progression of Shope papillomas occurs in domestic rabbits , as outlined above. CRPV was thus one of the first mammalian oncogenic viruses discovered, and the Shope papilloma-carcinoma complex of domestic rabbits was the first model of oncogenesis by a papillomavirus (Kreider and Bartlett, 1981).
Study of the model established several features peculiar to the system:

1). In papillomas, the viral DNA exists in an episomal, monomeric form.
2). Cancers are normally non-permissive, but viral capsid

antigens can be detected in carcinoma sections by immunofluorescence methods.

3). In the transplantable carcinomas VX2 and VX7, derived from Shope papillomavirus-induced malignancies, the viral DNA is integrated into the host cell genome (Georges et al. 1984). This may be the general case for CRPV-induced carcinomas and metastases, although it may not preclude the existence of episomal forms (Wettstein and Stevens, 1982).

4). Rous and Beard (1935) discovered that the application of tar to rabbits' ears induced generalised hyperplasia and papillomatosis. Rous and Kidd (1937) found that subsequent intravenous injection of CRPV resulted in the development of carcinomas in the tarred areas. Application of tar to established Shope papillomas accelerated the onset and increased the frequency of malignant progression. These latter findings show the synergistic effects of two tumour-inducing agents, a phenomenon encountered also in other papillomavirus-carcinoma systems.

1.1.5. Human papillomavirus-associated cancer.

A rare genetic disorder in humans, characterised by a lesion of the cell-mediated immune system, is manifested by the disease Epidermodysplasia verruciformis (EV) in which the patient develops generalised papillomatosis caused by HPV's of Group D (Pfister, 1984 . Table 1). Papillomas caused by HPV's 5, 8 and 17 are known to progress to carcinomas in about 30% of EV sufferers, mainly in sunlight-exposed areas

of the skin (Orth et al., 1980). Pre-malignant, virus-producing lesions have been observed progressing to Bowen's carcinoma; although the cancers contain viral DNA, they are not virion productive (Jablonska et al., 1972); however, the presence of HPV17-specific transcripts in an EV skin carcinoma points to a possible role for the viral DNA in the oncogenic process (Yutsudo and Hakura, 1987). This situation appears to parallel the Shope papilloma-carcinoma system and may indicate a common mechanism of transformation.

Immunosuppressed patients, e.g. renal allograft recipients, present a similar case to that of EV: an increased incidence of warts and the association of these with carcinomas in areas exposed to sunlight . Frequency of malignant progression is related to duration of sunlight exposure in both cases, implicating ultraviolet light in the role of co-carcinogen (Boyle et al., 1984)). In a study of one hundred and twenty renal allograft recipients of varying graft duration, 48% were found to have warts, with 87% of those with graft survival time of more than 5 years having warts. Six of the latter patients with multiple warts also had skin or anogenital carcinomas; the skin cancers occurred in light-exposed sites. HPV16 DNA was found in the case of a vulvar carcinoma. The EV-like plaque lesions on one patient were found to contain HPV5 DNA (Rudlinger et al., 1987). The latter case is of great interest as it may be indicative of latent infection by EV papillomaviruses of the general population at a sub-clinical level. The large number of types of EV papillomaviruses in this rare and widely geographically

scattered condition would lead one to expect a more general incidence of infection.

Papillomavirus research has proliferated with the discovery that cervical intraepithelial neoplasia (CIN) and cervical cancer are associated with papillomavirus infection (Gissmann et al., 1984). HPV's 16 and 18 are most frequently involved in genital cancer; HPV's 6 and 11 are also found, although they usually infect the external genitalia and HPV11 additionally causes laryngeal papillomas. More recently, HPV's 31 (Lorincz et al., 1986) and 33 (Beaudenon et al., 1986) have also been discovered in a small percentage of CIN's.

HPV's 16 and 18 are usually found to be integrated into the host genome in cancers and metastases (Shirasawa et al., 1986; Georges et al., 1984; Pater and Pater, 1985); however, other cancer associated HPV's are not e.g. HPV's 6, and in some cases 16 (Gissmann, 1984) and EV associated HPV's 5, 8 and 17 (Ostrow et al., 1982; Yutsudo et al, 1985).

No co-carcinogen has yet been associated with oncogenesis in genital HPV's, although women who smoke have an increased risk of cervical cancer. Sexual promiscuity also leads to a higher incidence of genital cancer (Gissmann, 1984)

Four common statements are emergent from the above examples in association with oncogenic papillomaviruses:

1). Cancers can be shown to progress from benign, papillomavirus-induced lesions;

2). Genetic predisposition and/or environmental co-carcinogens are involved in malignant progression.

3). Cancers are non-permissive for viral production.

4). Viral DNA persists in the malignant cells and in some cancers it is integrated into the host cell genome.

1.1.6. BPV1 in transformed fibroblasts as a model for papillomavirus-associated oncogenesis.

As has already been mentioned, the study of papillomaviruses has until recently been retarded by the lack of a permissive *in vitro* system. However, the advent of new molecular biological techniques has enabled considerable advances to be made in the field , such that the biochemical activities of some of the early proteins are now known, as is the nature of some of their transcriptional control.

No permissive *in vitro* system has yet been found for papillomaviruses, but three non-permissive systems have yielded much information:

1. cells directly transformed by papillomaviruses and/or their DNA, in particular BPV1 in mouse fibroblast cell lines NIH 3T3 and C127 (Dvoretsky *et al.*, 1980);

2. cell lines explanted from HPV- and CRPV-infected carcinomas;

3. production and expression of molecularly cloned papillomavirus DNA and their subgenomic fragments in *E. coli*, and transfection of these DNA's into fibroblasts.

The fact that BPV1 infects fibroblasts in its natural host makes it an obvious candidate for *in vitro* mouse fibroblast transformation. BPV1 virions and DNA have been shown also to

be capable of non-productively transforming primary cell cultures of bovine origin (Pfister, 1984), but since mouse NIH3T3 and C127 cells form a well-characterised and universally utilised molecular biological system, it is from these that most of the current information on BPV1 has been obtained.

Mouse fibroblasts transfected *in vitro* by BPV1 DNA form foci, grow in soft agar and are tumourigenic in nude mice, i.e. they conform to the phenotypic criteria of transformation. As such, they provide a system in which to study not only transformation activity but also such functions as regulation of transcription, translation and replication of the viral genome. These characterisations have progressed through the use of complementation studies on the partially deleted genome, and the directing of transcription of subgenomic fragments and cDNA clones by heterologous promoters. The use of well-characterised assays, e.g. the chloramphenicol acetyl transferase (CAT) assay, and inhibition assays of transcription and translation, has defined *trans*- and *cis*-acting sequences and their functions.

BPV1, because of the ease with which it transforms fibroblasts, and because it has been studied for longer than most papillomaviruses, is the best characterised papillomavirus in this system, and has become the reference point for other papillomavirus research. It must, however, be borne in mind that its specificity for fibroblasts sets BPV1 apart from most papillomaviruses including all the human ones, in spite of their homology; also, studies on *in vitro*

cell cultures cannot precisely reflect the *in vivo* mammalian system.

This limitation is most apparent in view of the fact that BPV1 is not oncogenic in its natural site of infection. Cells explanted from BPV1-induced fibropapillomas do not exhibit the transformed phenotype in culture; however, hamsters injected with BPV1 DNA develop fibrosarcomas from which transformed cells may be recovered (Pfister, 1984). Oncogenesis in heterologous host cells has been investigated *in vitro* by a number of workers: of particular interest are those experiments carried out on untransformed primary cells. Amtmann and Sauer (1982a) demonstrated the induction of replication and transcription of quiescent BPV1 DNA in mouse embryo fibroblasts using the tumour promoters 12-O-tetradecanoyl phorbol-13-acetate (TPA) and 12-O-retinoyl phorbol-13-acetate (RPA). The cells subsequently attained a transformed phenotype.

Transgenic mice in which tandem repeats of a transforming fragment of the BPV1 genome were integrated into the host chromosomes developed epithelial proliferations over the entire skin surface; the tumours were mitotically active but not metastatic, nor were tumour cells transformed in phenotype (Lacy *et al.*, 1986). This model may represent an early stage in carcinogenesis, or merely a non-productive infection in an heterologous host.

Latency of transformation has also been shown using molecularly cloned BPV1 DNA in immortalised C127 mouse

fibroblasts, in which a single application of TPA was sufficient to induce transformation (Meneguzzi et al., 1984). These types of experiments may reflect the apparent two-stage progression to malignancy seen in vivo in other papillomaviruses. Other groups, however, report the transformation of both primary and immortalised cells by papillomavirus DNA alone. "Single-hit" transformation of NIH 3T3 fibroblasts by molecularly cloned BPV's at a rate of 200-300 transformants per µg of DNA was demonstrated by Campo and Spandidos (1983). Cloned HPV16 DNA was also capable of transforming NIH 3T3 cells, but the transformation efficiency of this supposedly highly oncogenic papillomavirus was surprisingly low at only 27 transformants per µg (Yasumoto et al., 1986). Recently Matlashewski and co-workers (1987) demonstrated the cooperation of the activated ras oncogene with HPV16 sequences in the transformation of baby rat kidney primary cells. The HPV16 DNA was under the control of Moloney murine leukaemia virus regulatory sequences. Integrated HPV16 DNA alone did not effect reproducible transformation of these cells. Cloned HPV16 DNA under control of the Moloney murine tumour virus promoter with a neomycin resistance selection marker transformed human foreskin keratinocytes to an immortal, tumourigenic phenotype under selection conditions (Pirisi et al., 1987). This finding would appear to have a closer relationship to the natural system, which occurs in human genital epithelial cells.

These experiments and others, in spite of the superficial differences in their findings, may eventually lead to a

unified theory of papillomavirus-mediated oncogenesis. One such theory is that papillomavirus infection may expand a target for oncogenic events, and have no other role in progression of the cell in question to malignancy (M. S. Campo, personal communication). Before this question may be addressed, however, a comprehensive characterisation of papillomavirus functions is required. The steps which have been taken towards such an understanding will now be reviewed.

1.2. The molecular biology of the papillomaviruses.

1.2.1 Genome organisation of the papillomaviruses.

So far, no *in vitro* permissive system has been found for any of the papillomaviruses, even in human keratinocytes capable of partial differentiation in culture (Butel, 1971; Laporta and Taichman, 1982; Burnett and Gallimore, 1983; Pirisi *et al.*, 1987)). Hence the examination of their genetic structure has been directed along two routes: (i) analysis of transcripts both *in vivo* and in cultured cells transformed by molecularly cloned papillomavirus DNA; (ii) DNA sequence analysis. The elucidation of the sequences of several papillomaviruses represents a considerable advance in papillomavirus research as current knowledge of the functions of DNA sequences may be extended to them and point to pertinent aspects for further investigation.

The genomic organisations of HPV1, HPV16, CRPV and BPV1 are presented as typical examples of papillomaviruses (Fig. 1) These data intimate another set of statements peculiar to the papillomaviruses:

1. All open reading frames (orf's) of significant size are confined to one strand of the DNA duplex. This is supported by the finding that transcription is unidirectional (Amtmann and Sauer, 1982b).
2. All three reading frames are utilised, and several orf's overlap or are contained within another orf.

3. The genome can be functionally divided into three sections:

(i) that containing orf's coding for the two viral capsid proteins (Engel *et al*, 1983), termed the Late region by analogy to the L-region of polyomavirus and SV40 (Tooze, 1980);

(ii) the "transforming region", so named because a fragment comprising 69% of the BPV1 genome, containing these orf's, was found sufficient to transform mouse fibroblasts *in vitro* (Lowy *et al*., 1980): this is termed the Early region. The 3' ends of the early and late regions are signified by one or two polyadenylation signals, AATAAA;

(iii) a non-coding region of approximately 1 kb which recent work has shown to direct most of the complex transcriptional control of papillomaviruses.

4. Considerable homologies exist in some orf's of different viruses at the amino acid level (Giri and Danos, 1986): the L1 capsid gene is particularly well conserved, as is the E1 orf; however

5.the L2 capsid gene is widely divergent, except at its 3' end, indicating a possible tissue recognition function; similarly the non-conservation of the E4, E5, E6 and E7 orf's is suggestive of adaptations to divergent intracellular conditions in different tissue types.

Such hints at gene function demanded more detailed examination, with the result that a large body of data has been collected on the molecular biology of papillomaviruses in *in vitro* culture systems.

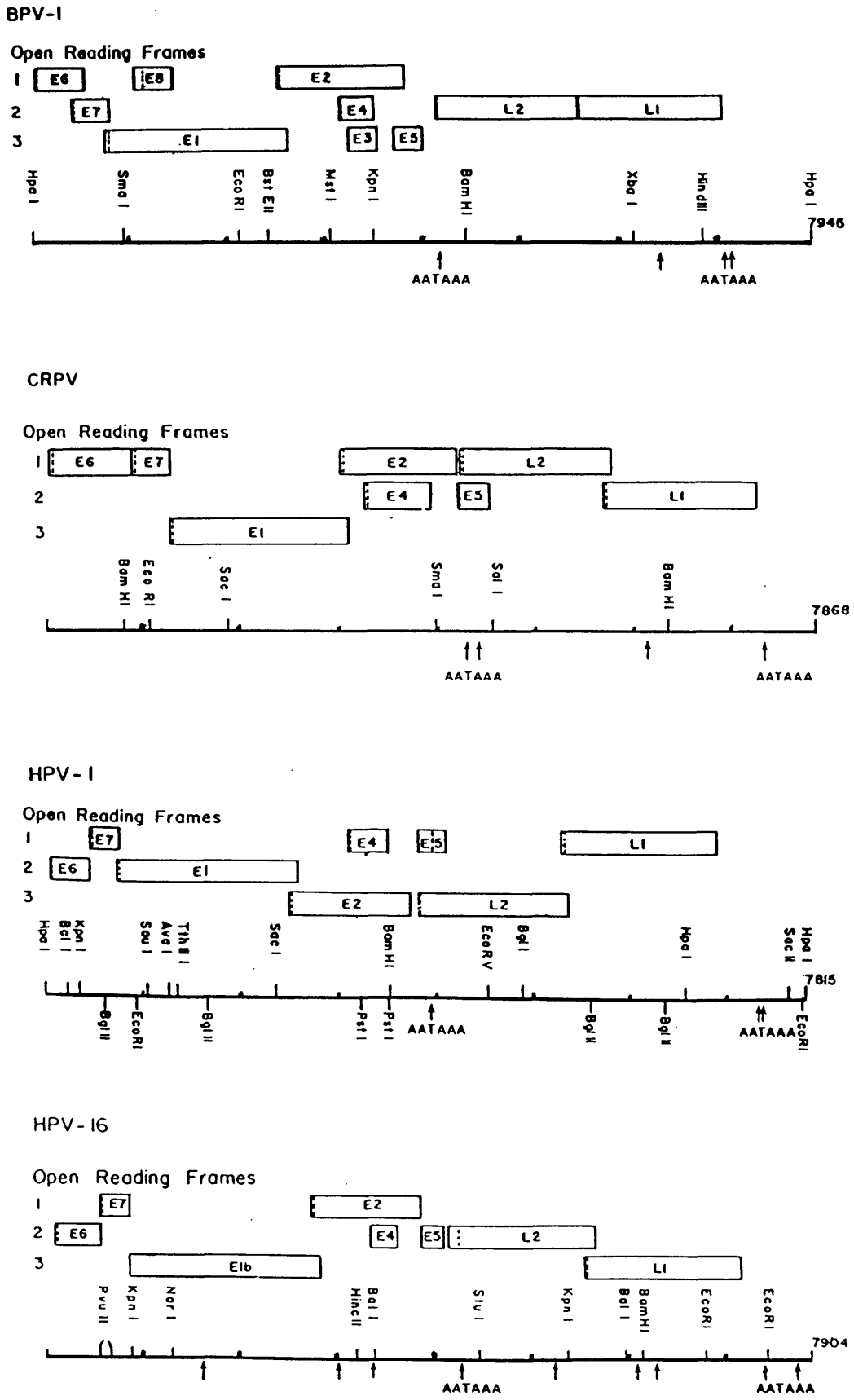
Figure 1.

The format used in this diagram to represent papillomavirus genome organisation is maintained in all other such diagrams in this thesis. DNA is represented by an unbroken horizontal line. Boxes represent orf's in the three translational frames. Vertical broken lines indicate the position of the first ATG codon in each orf. E = early; L = late.

Polyadenylation signals are shown in this figure by vertical arrows. Restriction sites are also included.

Numbers refer to nucleotide positions on the BPV4 genome.

Figure 1. Genomic organisations of BPV1, CRPV, HPV1 and HPV16.



Broker and Botchan. 1986.

1.2.2. The state of papillomavirus DNA and its transcription in vivo and in vitro.

1.2.2.(i) BPV1 DNA in fibropapillomas and transformed cells.

BPV1 DNA replicates in bovine fibroblasts and keratinocytes *in vivo* and in mouse C127 fibroblasts *in vitro* independently of the host cell genome. It is maintained as a circular episome which in BPV1-transformed C127 cells (ID13) exists at a copy number of about 150 per cell (Law *et al.*, 1981). Initial experiments aimed at functional analysis tested the transforming activity of various subgenomic fragments cloned into plasmids and showed that a cloned fragment of BPV1 DNA representing 69 % of the genome (Fig. 2) was capable of transforming C127 cells (Lowy *et al.*, 1980). Plasmids containing the fragment remained episomal in the transformed cells. Sequence analysis later established that the eight early orf's were encoded in this fragment (Chen *et al.*, 1982), and this led to the designation "early" or "transforming" region for these orf's.

1.2.2.(i)(a). The transcriptional organisation of BPV1 in fibropapillomas.

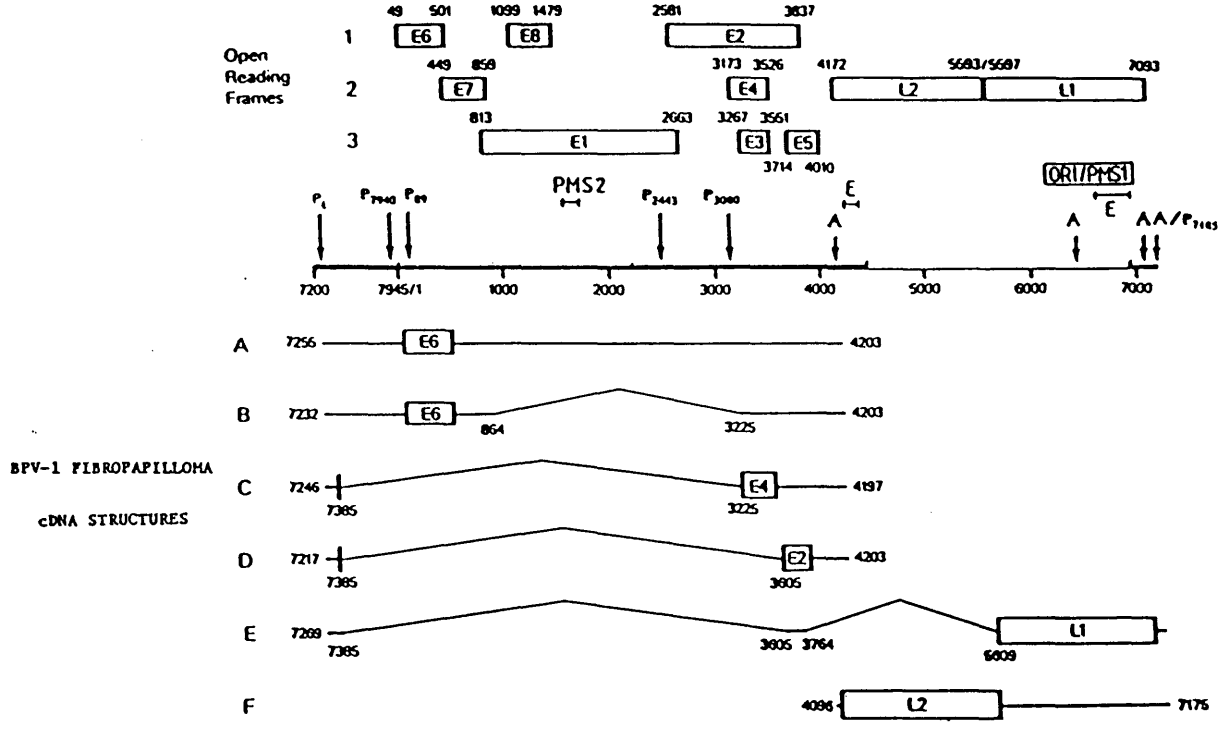
Northern blot analyses of BPV1-specific transcripts in bovine fibropapillomas using single-stranded BPV1 DNA probes showed that transcription occurred only on one strand. Eight size

classes of RNA were found in the polyadenylated fraction, of 8.00; 6.70; 4.05; 3.80; 3.70; 1.70; 1.15 and 1.05 kb respectively. It was also shown that the 8.0; 6.7; 3.8 and 1.7 kb species contained sequences hybridising to the late region of BPV1 and that the 3' termini of these transcripts were near the late polyadenylation signal. The 8.0 kb species hybridised to fragments from all regions of the genome. The other RNA's hybridised to the early region only and terminated exclusively at the early polyadenylation site. In view of the large number of RNA splicing signals scattered throughout the BPV1 genome, it was considered that the larger transcripts were probably nuclear precursors (Engel et al., 1983).

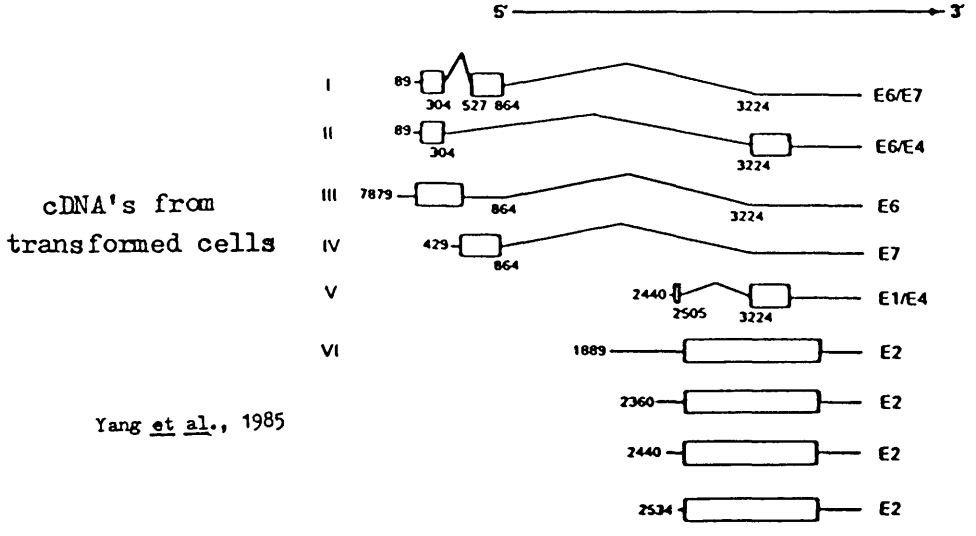
The analysis of a cDNA library produced from BPV1 fibropapilloma mRNA is presented in Figure 2 (Baker and Howley, 1987). It is immediately apparent from this map that a complex splicing pattern is directed by viral sequences, in some cases joining parts of different orf's to create new orf's. Another characteristic of papillomaviruses demonstrated by this analysis is the low abundance of their transcripts: in spite of high viral genome copy numbers in papilloma cells, viral transcripts accounted for only 1.3 % of polyadenylated RNA in this analysis (Baker and Howley, 1987).

All the cDNA classes isolated had 5' ends around nt7250 in the non coding region, except for class F. The most abundant BPV1 transcript in fibropapillomas was represented by cDNA's of class C (Fig. 2). Sequence analysis showed that a fusion

Figure 2. Genomic and transcriptional organisation of BPV1.



C.C.Baker and P.M.Howley 1987



Yang et al., 1985

LEGEND

▭	orf
———	DNA
A	polyadenylation site consensus sequence
P	promoter
PMS	plasmid maintenance sequence
E	transcriptional enhancer
ORI	origin of replication
▬▬	69T region

Diagonal lines in cDNA's represent splice junctions

orf was created by the splice, with the leader exon from the non-coding region contributing an ATG plus one other codon, joined onto a truncated E4 orf. The E5 orf was also encoded by this cDNA.

Class D cDNA's specified a splice junction between the same leader and the 3' 78 codons of the E2 orf. The 3' terminus of both classes of cDNA was at the early polyadenylation site.

A late transcript, represented by cDNA's of class E, was formed by the splicing of the leader exon described above to a short exon containing sequences transcribed from the within the E2 orf; this was in turn spliced onto the L1 orf. Two major orf's were encoded by this cDNA: analysis of sequences flanking the ATG codons of each orf according to rules defined by Kozak (1986) showed that an ATG lying within the L1 orf was in the most favourable configuration for translation initiation. However, another fusion orf beginning with the two codons from the leader exon continued through the middle exon and terminated in a short open reading frame in the L1 exon, but not in the L1 reading frame. This orf specified a putative 72 amino acid protein. Although the translation initiator sequence is a weak one, the elaborate splicing pattern seen in this cDNA would seem to argue in favour of some function for the fusion protein; otherwise it may serve as an unusually large leader for the L1 orf.

The 5' terminus of the class F cDNA, encoding the L2 orf, was not close to any known promoter and was considered to be a truncated copy of the transcript it represents.

The two other cDNA classes defined in this analysis,

represented transcripts derived from the early region, both having 3' termini at the early polyadenylation site, nt4203. Class A cDNA's were colinear with the BPV1 genome, whereas those of class B were spliced from the beginning of the E1 orf to the splice acceptor in the middle of both the E2 and E4 orf's.

1.2.2.(i)(b). Transcriptional promoters of BPV1.

The cDNA's described above did not account for all the size classes of BPV1 RNA reported by Engel et al. (1983). Furthermore, all but one showed transcriptional initiation from one small area of the genome, whereas sequence analysis located several other potential promoters (Chen et al., 1982). Therefore, further transcriptional analysis was carried out on polyadenylated RNA from fibropapillomas using S1-nuclease protection and primer extension (Baker and Howley, 1987). These experiments showed that the majority of mRNA's were transcribed from a variety of sites between nt7214 and nt7254, implying that transcriptional initiation from this promoter is not precisely directed in vivo. However, when polyadenylated RNA from the fibromatous part only of the fibropapillomas was used in this analysis, no transcripts could be found which had initiated from this promoter. Since the expression of late viral products is confined to the outer differentiated layers of epithelial cells (Amtmann and Sauer, 1982b; Pfister, 1984), this promoter was termed the major late promoter, P_L. Two tandem

copies of a sequence 73 % homologous to the SV40 late promoter sequence, GGTACCTAACC, were found upstream of the "late" transcription initiation sites. This promoter is important in SV40 for the specification and efficient utilisation of the major late RNA start site (Brady et al., 1984). Since many of the transcripts initiating at this site in BPV1 are confined to "early" region sequences, the re-definition of this region as the transforming region might be more appropriate.

Mapping of transcriptional start sites directed by other putative promoters was also presented in this report (Baker and Howley, 1987). An additional start site was found at nt7185. Initiation mapped to a single nucleotide and a TATA homology was found overlapping the late polyadenylation signal 25 nucleotides upstream, but transcripts initiating here were far less abundant than those starting in P_L, particularly in RNA from the fibromatous part of the tumour. This minor late promoter was designated P_{7185}.

Transcriptional initiation sites were also mapped at nt89, and a cluster between nt7934 and nt7939. TATA consensus sequences were found about 30 nucleotides upstream of these two sites and were termed P_{89} and P_{7940}, respectively.

Two other start sites were found within the transforming region. In one, the main cluster of termini was seen between nt3072 and nt3080, with minor signals scattered as far upstream as nt3012. The other start site was mapped to nt2443 (Ahola et al., 1987). The promoters for these start sites were designated P_{3080} and P_{2443}, respectively. No TATA box

sequences could be found for P_{3080}, but a TATA homology exists at nt2414 which could direct P_{2443} transcription. Mapping of transcripts initiated at P_{2443} showed that, while some were colinear with the genome, 90 % were spliced from nt2505 to nt3225. The 3' termini of transcripts directed by all promoters except P_L was at the transforming region polyadenylation site.

1.2.2.(i)(c). Differential utilisation of BPV1 promoters in fibroblasts and epithelial cells of bovine fibropapillomas.

Quantitative analyses of transcription from the various promoters of BPV1 carried out separately on RNA isolated from the fibromatous and epithelial parts of a bovine fibropapilloma revealed differences between levels of promoter utilisation in the two types of cell (Baker and Howley, 1987).

Apart from the non-utilisation of P_L in bovine fibroma cells, it was also found that the level of transcription from P_{89} and P_{7940} in these cells was approximately ten times that in the epithelial part of the tumour. Transcription from P_{2443} was reduced in epithelial cells to approximately 30 % of the level quantified in fibromatous tissue. Transcripts initiated at P_{2443} and P_{3080} were the most abundant viral transcripts found in fibroma cells, while in the epithelial portion of the tumour, P_L-directed transcripts were 10- to 100-fold more abundant than those initiated at any other viral promoter.

These data have important implications for the fibroblast model for papillomavirus functions. It appears that transcription from P_L is activated only in epithelial cells and must therefore be dependent upon the interaction, either direct or indirect, of keratinocyte-specific factors with viral promoters. The transcription of the capsid encoding orf's L1 and L2 is also dependent on failure to respond to the early transcription termination signal. Since the majority of classes of P_L-directed transcripts terminate here, transcription of L1 and L2 may be due to the attenuation of some termination mechanism operating at this site. This could be effected through titration of the termination factor(s) due to rapid increase in viral copy number in the suprabasal layers of the epithelium, or by the expression of a cellular factor specific to keratinocytes. Failure to express late transcripts, and the increased level of transcription from the other promoters, may be a feature of an early, fibroblastic phase of the life cycle of BPV1, and *in vitro* studies in fibroblasts have to be judged within these parameters. In particular, progression to malignancy is a feature of persistent, long-term papillomatosis in the Shope system (Kreider and Bartlett, 1981) and therefore less likely to be associated with an early event in the viral life cycle. Transformation of cells *in vitro* may not necessarily be brought about by the same mechanisms operating *in vivo*.

1.2.2.(i)(d). Transcriptional organisation of BPV1 in transformed C127 cells.

BPV1-specific transcripts in ID13 cells (Law et al., 1981) are a subset of those found in fibropapillomas and represent 0.02 % of polyadenylated RNA, according to analysis of a cDNA library (Yang et al., 1985a). The primary difference accounting for these discrepancies is the absence of transcripts initiated at the P_L transcriptional start sites. Mapping of BPV1 mRNA's in ID13 cells has been done by cDNA analysis (Yang et al., 1985a), electron microscopy of DNA:RNA hybrids, S1-nuclease protection analysis and primer extension (Stenlund et al., 1985; Ahola et al., 1987; Baker and Howley, 1987). Results of these analyses are in good agreement and are therefore represented by the cDNA map in the lower part of Figure 2.

Again, differential splicing of viral transcripts results in junctions between parts of different orf's with potential to code for new orf's. 90 % of BPV1 transcripts in ID13 cells were found to be transcribed from a region around P_{2443}; their 5' ends were highly heterogeneous, with the major start sites occurring between nt2436 and nt2447, and minor signals mapping around nt2410 (Ahola et al., 1987). 90 % of these were of the spliced species described under the previous heading (Baker and Howley, 1987), and represented by class V cDNA's (Fig. 2), and the remainder were unspliced. The only orf of significant size in these transcripts was the E5 orf, which contained an internal AUG codon following a leader of

nearly 800 nucleotides. The colinear transcripts of P_{2443} also encode the E5 orf, but also the full-length E2 orf. Transcripts directed by the P_{89} promoter were represented by approximately 4 % of cDNA's, comprising classes I and II. Class I cDNA's encoded an E6/E7 fusion orf while those of class II contained an E6/E4 fusion; these fusions were created by differential splicing from the same splice donor in the middle of the E6 orf. cDNA's encoding the full-length E6 orf (class III) were found in very low abundance in this analysis, as were those encoding the E7 orf. The apparent start sites represented by these molecules did not conform to mapped coordinates (Baker and Howley, 1987; Ahola et al., 1987), so these were probably truncated cDNA's.

Quantitative analysis of BPV1 transcripts in ID13 cells showed that levels of viral promoter utilisation in these cells were equivalent to those in the fibromatous part of fibropapillomas (Baker and Howley, 1987). No capsid protein encoding transcripts have been found in ID13 cells, although P_{7185} was found to be active at a very low level (Baker and Howley, 1987), and transcription from P_L could only be detected after treatment of the cells with cycloheximide (Kleiner et al., 1986). According to cDNA analysis, transcripts initiated at P_L have the potential to direct the synthesis of a variety of proteins from different orf's, each presumably essential to the propagation and possibly persistence of the virus in epithelial cells. Non-utilisation of this promoter may be responsible for the transformed phenotype of ID13 cells, perhaps due to overexpression of

transcripts suppressed in keratinocytes. The untransformed nature of fibroblasts in fibropapillomas, however, indicates that some feature of C127 cells or their culture system must act synergistically with the virus in the induction of the transformed phenotype.

The absence of P_L-directed transcripts in ID13 cells presents an opportunity to investigate the functions of their translation products by supertransfection of the appropriate cDNA under control of constitutive promoters. cDNA's, or subgenomic fragments containing specific orf's, may be used in this way in two experimental approaches aimed at characterisation of papillomaviruses:

1). as functional assays for the encoded proteins, and in complementation experiments with deletion mutants;

2). in the analysis of interactions between viral proteins and cis-acting regulatory regions.

In addition, sub-genomic fragments and cDNA's may be used in prokaryotic vectors for the production in bacteria of viral peptides against which antisera may be raised, providing a means of isolating the protein in question from cells, or of localising it in cells and tumour sections by in situ fluorescence attachment.

1.2.2.(ii). Transcriptional organisation of CRPV in papillomas and the VX2 carcinoma.

As an animal model for papillomavirus-associated oncogenesis, the Shope complex in domestic rabbits presents an apparently

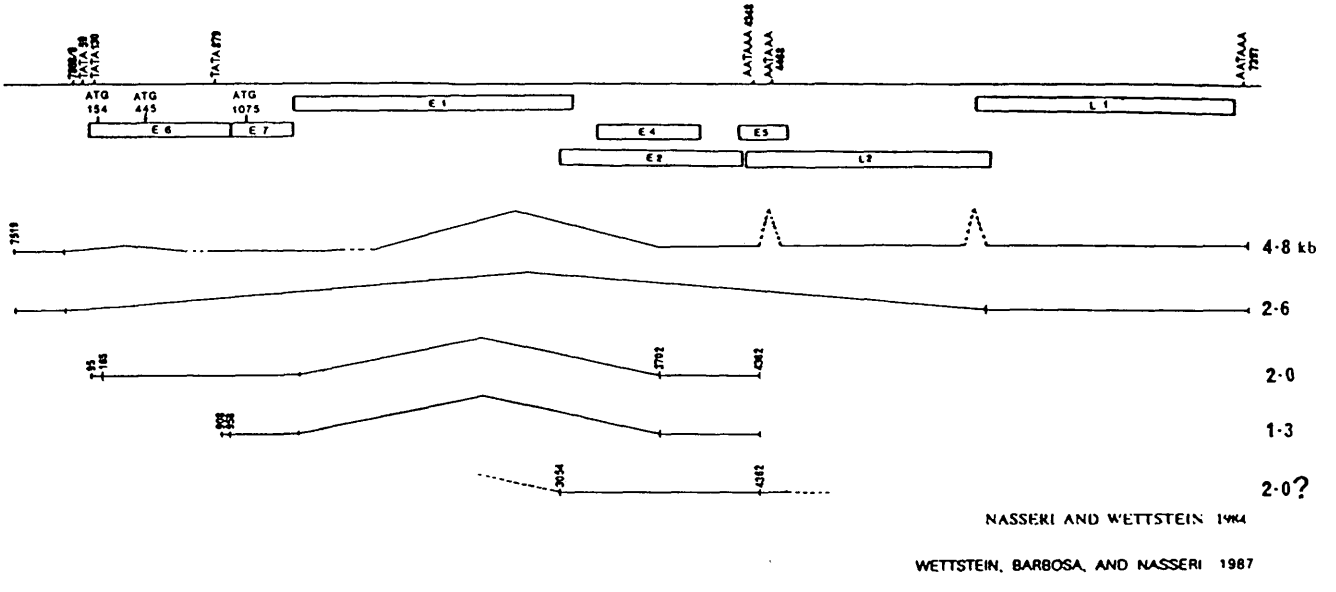
ideal system. However, the molecular biological characterisation of CRPV has lagged behind that of BPV1, in spite of the availability of the VX2 carcinoma, transplanted between generations of domestic rabbits for over 30 years (1.1.4.).

As has already been mentioned, CRPV DNA, although episomal in benign papillomas, exists entirely in integrated form in VX2 carcinoma cells. Most of the 10-20 copies of the genome in malignant cells are arranged in a head-to-tail tandem configuration; the viral integration coordinate is unknown (Georges et al., 1984).

Transcriptional analysis of CRPV in both the benign and malignant contexts was carried out by R-loop analysis (Georges et al., 1984) and nuclease protection assays (Nasseri and Wettstein, 1984; Danos et al., 1985; Wettstein et al., 1987). The results of these experiments are summarised in Figure 3.

Northern blot hybridisation analysis defined 5 transcript size classes in virion-productive papillomas of cottontail rabbits. As in transcripts of BPV1, all CRPV transcripts terminated at either the early or the late polyadenylation site. Transcription was shown to be unidirectional. Two transcripts hybridising to the late region of CRPV, of 4.8 and 2.6 kb, and a message containing sequences from the E2/E4 region, of 0.9 kb, were not found in the non-productive VX2 carcinoma or domestic rabbit papillomas induced by CRPV (Nasseri and Wettstein, 1984). The late region exons of the 4.8 and 2.6 kb messages were found to be spliced to distal 5'

Figure 3. Transcriptional organisation of CRPV in papillomas and carcinomas.



exons; like their P_L-directed counterparts in BPV1, their transcription initiated at the 5' end of the non coding region, and the leader exon was spliced from within this region to another exon within the transforming region.The second exons of these transcripts have not yet been mapped, but are thought to be in the E6-E7 region (Wettstein et al., 1987). In the 2.6 kb species, the second exon was spliced onto an L1-encoding terminal exon, while the main body of the 4.8 kb mRNA may consist of multiple exons transcribed from the E2/E4, L1 and L2 orf's (Nasseri and Wettstein, 1984).

Transcripts of 2.0 and 1.3 kb were found in all CRPV-induced papillomas and carcinomas. In papillomas, the 2.0 kb message was more abundant than the 1.3 kb, whereas in the VX2 carcinoma, the 1.3 kb transcript predominated. In addition, a minor 1.25 kb species was found in VX2 RNA.

Detailed S1-nuclease mapping showed that the 2.0 kb transcript consisted of two exons: a leader transcribed from the beginning of the E6 orf and spliced from a position downstream of the E7 orf to an exon transcribed from within the E2/E4 region, terminating at the first polyadenylation site of the transforming region (Danos et al., 1985; Wettstein et al., 1987). Two transcriptional start sites were found at nt95 and nt165; these sites were 35 nucleotides downstream of two TATA boxes and straddled the first ATG codon of the E6 orf, the major start site being downstream. The E6 orf of CRPV is unusually large by comparison with other papillomaviruses (Giri et al., 1985), so it may be that

the truncated form encoded by the major 2.0 kb leader exon represents the "consensus" E6. The significance of the larger exon has not been established; there is no difference in its transcription level between benign and malignant tumours.

The 1.3 kb transcript, expressed at an elevated level in the VX2 carcinoma, has two transcriptional start sites. The major start site was mapped to nt908, at the 3' end of E6, and utilised in all CRPV-induced lesions. A minor start site at nt958, also at the 3' end of E6, was defined only in the VX2 carcinoma, and accounts for the 1.25 kb RNA species (Danos et al., 1985). The leader exon of both species was shown to be spliced onto the E2/E4 exon utilising the same splicing signals as the 2.0 kb transcript (Wettstein et al., 1987). The significance of the differential start site is unknown, as the first AUG codon of both transcripts is the same, within the E7 orf.

In addition, a rare transcript containing an exon coding for the full-length E2 orf was found. In RNA extracted from a Shope carcinoma, this exon terminated at either of the two early polyadenylation sites. However, in polyadenylated RNA from a cottontail rabbit papilloma, this exon was shown to extend 3' to the early polyadenylation signals in a sub-population of molecules. It was also shown by neutral agarose gel analysis of mung bean nuclease protection assays that this exon was spliced to approximately 1.25 kb of upstream sequences, and may therefore be another 2.6 kb transcript.

The current transcriptional map of CRPV in the Shope

papilloma carcinoma system thus demonstrates, as in the case of BPV1, the "switching-off" of a viral transcriptional promoter in an heterologous cell system, accompanied by failure to transcribe past the early polyadenylation signals and an alteration in the steady-state levels and positional accuracy of transcriptional initiation from another promoter. Most notable is the fact that only three CRPV-specific transcripts are found in domestic rabbit papillomas and carcinomas, at least two of which have as their first orf the E6 or E7 orf's. This may implicate the translation products of these orf's in the oncogenic process and/or maintenance of the malignant state, particularly in the VX2 carcinoma in which these sequences have presumably been transcribed for over thirty years. The above studies, although detailed and to an extent repetitious, do not represent a functional study, which could only be effectively instituted using cDNA's to express translational products of the transcripts in question.

1.2.2.(iii) HPV's in papillomas and carcinoma cell lines.

The bovine model for papillomavirus infection has allowed rapid progress in the characterisation of BPV1 *in vivo*. Ethical restraint has retarded such studies in the case of HPV's, but the establishment of several cell lines from genital cancers containing HPV DNA has provided potential for the investigation of papillomavirus-associated oncogenesis. Like BPV1, HPV DNA in benign papillomas and condylomata

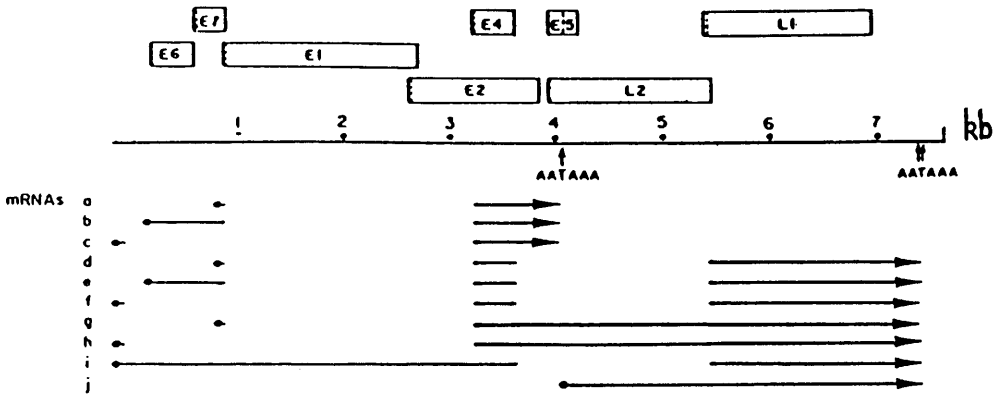
exists as a multicopy episome, but in most cancers and cancer-derived cell lines at least some of the viral DNA is integrated into host chromosomes.

1.2.2.(iii)(a) Transcription of HPV's 1, 6b and 11 in benign tumours.

Viral transcription in plantar warts containing HPV1, in a vulvar condyloma containing HPV11 and in an anal condyloma containing HPV 6b was studied by R-loop analysis, the results of which are reproduced in Figure 4 (Chow *et al.*, 1987a,b). Although precise details of transcriptional start sites and splice junctions cannot be mapped by electron microscopy of R-loops, these exhaustive analyses provide a more comprehensive examination of papillomavirus transcripts and define an emerging pattern of transcription in benign epithelial tumours. Furthermore, they enabled the selection of subgenomic fragments from which antisense probes specific for each class of transcript could be synthesised. These probes were used to localise transcripts in the various differential layers of condylomata by *in situ* hybridisation. The transcriptional organisation of HPV's 1, 6b and 11 was found to be in many ways similar to the overall map of BPV1 transcription in fibropapillomas. The 3' ends of all transcripts occurred near either the early or late polyadenylation signal, and transcription was shown to be unidirectional. Nearly all R-loops demonstrated splicing. Transcription appeared to be initiated from three main

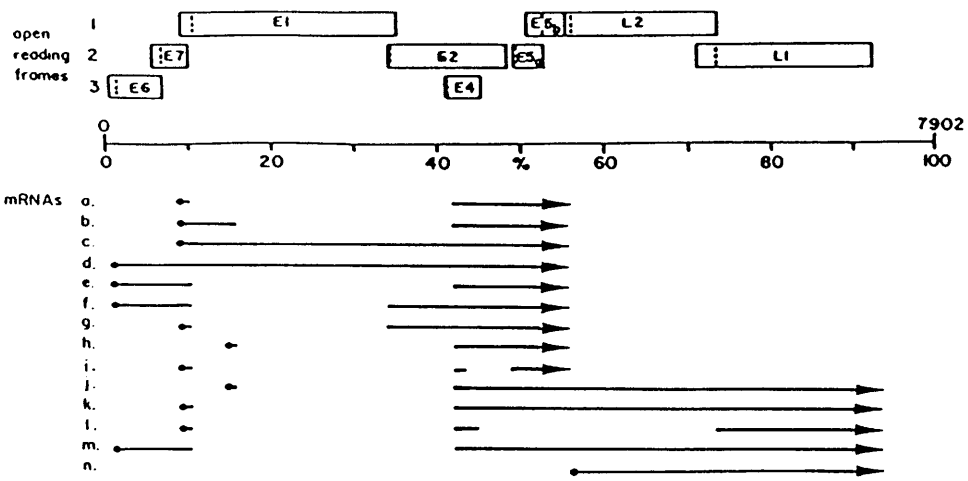
Figure 4. Transcriptional organisation of HPV's 1, 6b and 11 in benign tumours.

HPV-1



CHOW ET AL. 1987b

HPV-6b and HPV-11



CHOW ET AL. 1987a

LEGEND

→ polyadenylated 3' end

●— 5' terminus

promoters: in HPV1, these occurred in the non coding region, at the beginning of the E6 orf and within the E7 orf; in HPV's 6b and 11, they were located in the non coding region, in the overlapping region of the E7 and E1 orf's and within the E1 orf. Many transcripts had small, distal leader exons transcribed from these promoters spliced onto the main exon(s). The single exception was a transcript spanning the late region which initiated at a point immediately 5' to the L2 orf, similarly to cDNA's of class F from BPV1 transcripts. The large majority of R-loops represented a single mRNA species, class (a), containing sequences transcribed from the E2/E4 region spliced to a distal upstream leader of about 120 nucleotides, transcribed from the E7/E1 region. A cDNA library synthesised from RNA from a condyloma containing HPV11 yielded cDNA's representing only this transcript. Sequence analysis revealed a fusion orf consisting of four codons following an ATG from the 5' end of E1, joined to 85 codons form the E4 orf; the E5a and E5b orf's were also encoded by these cDNA's. A high degree of amino acid sequence conservation was observed between the fusion orf and a predicted corresponding fusion orf in HPV6b (Nasseri et al., 1987). Antisense probes from the E4-E5 region of HPV6b displayed a broad area of hybridisation to condyloma sections, particularly in the middle and outer differentiated layers (Chow et al., 1987a).

HPV1 RNA's of class (b) and HPV6b and 11 RNA's of class (e) bear similarity to class III cDNA's of ID13 cells (Fig. 2; Yang et al.,1985a); again, it could not be determined whether

these rare species were partially spliced intermediates of an analogue of ID13 class I transcripts, although no R-loops representing E6/E7 splices were found (Chow et al., 1987a). Transcripts of HPV1 classes (d-f) and HPV6b and 11 class (m) were similar to BPV1 class E. A 320 nucleotide stretch transcribed from the E2/E4 region was spliced 3' to an L1 exon, and 5' either to short leaders from the non coding region (HPV1) and the E7/E1 overlap (HPV's 1, 6b and 11), or to a longer exon containing the E6 and E7 orf's (HPV1). This apparent conservation of splicing adds more weight to the argument for a function encoded by the fusion orf in the BPV1 class E cDNA [1.2.2.(i)(a)]. Similar leader exons were found in transcripts terminating at the early polyadenylation site. Antisense probes from the E6 and E7 orf's localised higher concentrations of transcripts containing these sequences to the more differentiated strata of condyloma sections, and L1 and L2 probes only gave signals in the superficial layers, showing that they are late transcripts in the temporal as well as notional sense.

Downstream exons of the transforming region transcripts of HPV's 6b and 11 contained both the full-length E2 orf and a truncated E2 beginning within the overlapping region of the E4 orf. All contained both E5 orf's; a rare transcript was internally spliced in the E2/E4 overlapping region. By contrast, all the corresponding HPV1 transcripts found in plantar wart RNA encoded only the 3'-E2/E4 region in their 3' exons, and terminated before the first AUG codon of the single E5 orf.

Colinear transcripts and those containing long uninterrupted exons were extremely rare in these analyses and thought to be nuclear intermediates of mature transcripts. As was the case with BPV1, transcription from the E1 region was rarely seen, and an antisense DNA probe from this region did not show a detectable level of hybridisation to the cytoplasm of condyloma cells, although nuclear signals were obtained (Chow et al., 1987a,b). These could represent unspliced precursors of other mRNA's or undegraded by-products of splicing, or precursors of highly labile E1 messages.

Apart from the differences noted above, transcription of HPV's in benign tumours parallels, at least superficially, that of BPV1 in fibropapillomas. This might be expected in view of their very similar genomic organisations. The most notable difference between the transcript maps of the HPV's themselves was that the E5 orf of HPV1 was not transcribed in any of the early transcripts, whereas all the early mRNA's of the cancer-causing HPV's 6b and 11 contained the respective E5a and E5b orf's. Whether the E5 orf confers carcinogenic potential is therefore an important topic for functional analysis, although HPV1 DNA has been shown to transform mouse cells in culture (Watts et al., 1984).

In situ hybridisation analysis, although it identified the L1 and L2-encoding transcripts as truly late messages, only demonstrated a general increase in the levels of other transcripts towards the superficial layers of the benign tumour, which may only reflect the amplification of the viral template rather than any regulatory activity. Again, the

conclusion is that true functional dissection can only be achieved through manipulation of papillomavirus-specific cDNA's.

1.2.2.(iii)(b). Transcription of integrated HPV16 DNA in carcinoma cell lines.

(1). The state of HPV16 DNA in cervical carcinomas and their derived cell lines.

HPV16 DNA transforms both NIH3T3 fibroblasts (Yasumoto et al., 1986) and human foreskin keratinocytes (Pirisi et al., 1987); it is found in 62 % of cervical carcinomas and shows a prevalence in cervical dysplasias proportional to the severity of the lesions (Shirasawa et al., 1986). In all cases, the viral DNA is integrated into the host chromosomes, invariably with rearrangements and deletions of the viral genome. The presence of integrated HPV16 in precancerous lesions (Shirasawa et al., 1986; Schneider-Manoury et al., 1987), together with the almost universal presence of HPV DNA in cervical carcinomas (95 %, using hybridisation probes of known type: Gissmann et al., 1987), are strong evidence for a causal link between papillomaviruses and cancer.

Two carcinoma cell lines containing HPV16 DNA have been characterised: the CaSki and SiHa cell lines. CaSki cells contain approximately 600 copies per cell of the viral genome, predominantly as head-to-tail tandem repeats of the intact genome, and one from which the non coding region has been deleted. By contrast, the SiHa cell line contains only a

single copy of the HPV16 genome, which has a 215 bp deletion of the E2 orf. The integration site coincides with the ends of this deletion, creating fusion orf's with human DNA sequences. These run into the E2/E4 region from the left-flanking human sequences, and out of the 5' part of the E2 orf into the right-flanking sequences (Baker et al., 1987) The integration of viral DNA in cell lines and carcinomas does not preclude the existence of the episome, which has been found in some tumour samples (Smotkin and Wettstein, 1986); however, virions are never produced by these cancers, and the absence of episomal papillomavirus DNA in most cervical cancers (Gissmann et al., 1987) indicates that it has no function in the malignant state.

(2). Transcription of HPV16 in SiHa and CaSki cells.

Hybridisation of strand-specific subgenomic probes to RNA extracted from SiHa cells showed that transcription of the HPV16 genome in these cells was unidirectional and confined to the region of the E6, E7 and E1 orf's, although the transcript sizes of 4.6 and 4.1 kb indicated that 3' flanking cellular sequences were also transcribed and presumably spliced onto the viral sequences.

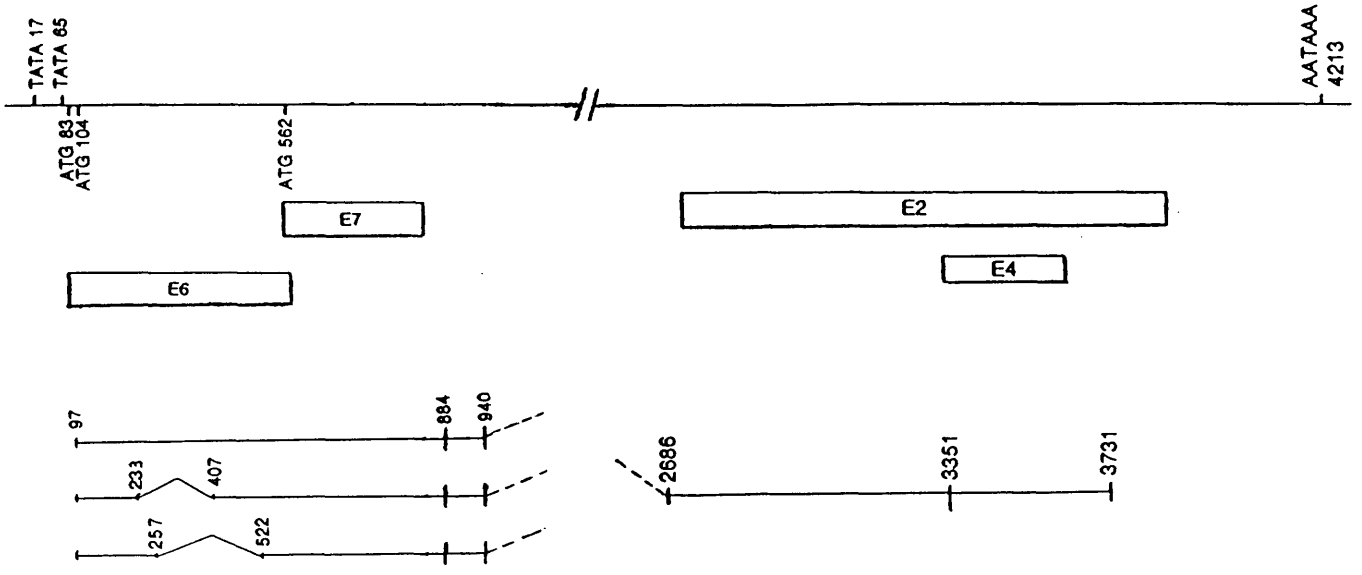
The CaSki cell line contained four size classes of HPV RNA: a 3.6 kb band hybridising to probes from the E6, E1 and E2 orf's; 1.5 and 1.0 kb bands encoding E6, E7 and E2 sequences and a 0.65 kb species hybridising to E2 sequences. No late region transcription was detectable in either case (Baker et al., 1987).

Exon mapping of HPV16 transcripts in CaSki cells was carried out by Smotkin and Wettstein (1986; Fig.5). Their identification of major transcripts of 4.5 and 1.5 kb, and a minor species of 2.3 kb, was at variance with the later report (Baker et al., 1987), although the same methods were used.

5' exons were mapped to the E6-E7 region of HPV16; these transcripts had a common initiation site at nt97, just downstream of the first ATG of the E6 orf. A minor exon contained both the E6 and E7 orf's, and two other transcripts were found to initiate in this region. A major transcript initiating in this region was spliced from nt232 internally within the E6 orf to nt407. Another minor transcript contained a splice junction between nt257 and nt522, just upstream of the E7 orf. Two 3' ends were mapped downstream of these splice junctions, at nt884 and nt940, both 3' to the E7 orf.

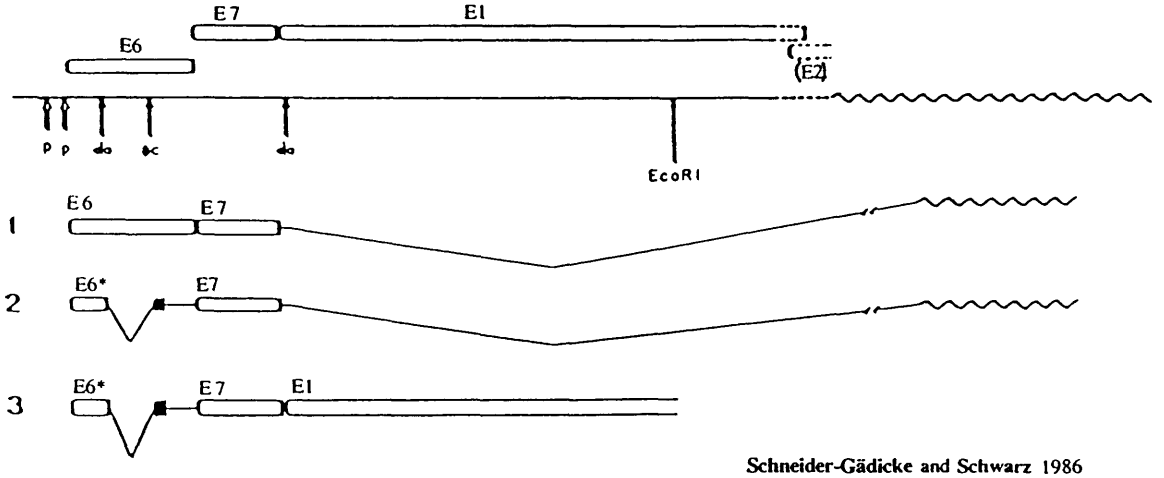
These transcripts were shown by non-denaturing gel analysis of S1-nuclease protection assays to be spliced to exons mapping to the E2/E4 region. The 5' ends of these exons were mapped to a minor site upstream of the E2 orf and a major site within the overlapping region. The 3' end of both was identical and occurred within the E2 orf, downstream of the E4 orf. Since no polyadenylation signal was found near this position, and the maximum estimated size of the full transcripts was short of those established by Northern analysis, it was assumed that this corresponded to a viral/host DNA junction and that transcription proceeded

Figure 5. Transcriptional organisation of HPV18 DNA in CaSki cells.



Smotkin and Wettstein, 1986.

Figure 6. Transcriptional organisation of HPV18 DNA in HeLa, C4-1 and SW 756 cells.



Schneider-Gädicke and Schwarz 1986

LEGEND

~~	host sequences
do	splice donor site
ac	splice acceptor site
p	TATA promoter

along human sequences until a termination signal was reached. In a separate analysis of RNA from a tumour containing episomal as well as integrated HPV16 DNA, these 3' exons terminated at the early polyadenylation site.

These analyses showed the extent of disablement of HPV16 in malignant tumours. Similarities in the splicing arrangements to those found in BPV1 and other HPV transcript maps indicates that the CaSki transcripts are a subset of those in benign tissues; it also appears that they are only derived from viral DNA at an integration junction. This precludes the transcription of the 3' end of the transforming region, containing the E5 orf and the 3' end of the E2 orf, and of the late region, although the latter is not transcribed even from episomal HPV16 DNA in cervical carcinoma cells. Only one promoter is active, directing transcription from just within the E6 orf. An unusual internal splice is seen in sequences coding for the E6 orf, and the transcription of only E6, E7 and E2/E4 sequences may indicate an involvement of these in the progression to and/or maintenance of the malignant state, particularly in view of the finding that integrated HPV16 sequences are found in precancerous dysplasias.

1.2.2.(iii)(c). Transcription of HPV18 DNA in carcinoma cell lines.

Three established cervical carcinoma-derived cell lines were found to contain integrated HPV18 DNA: HeLa, C4-1 and SW756. The viral integration pattern is similar to that of HPV16 in

SiHa cells, although several copies of the HPV18 genome exist at different integration sites in these cell lines. Integration is again within the E2 orf, and the 3' end of the transforming region and part of the late region are deleted in HeLa and C4-1 cells, indicating that these parts of the viral genome at least are no longer required in cancerous cells (Schwarz et al., 1985).

Polyadenylated RNA from HeLa, C4-1 and SW756 cells hybridised only to probes from the E6-E7-E1 region of HPV18, even in RNA species of up to 6.5 kb. Analysis of a cDNA library synthesised from these RNA fractions revealed a transcription pattern very similar to that of HPV16 in CaSki cells (Fig. 6; Schneider-Gadicke and Schwarz, 1986). Primer extension analysis showed that transcription initiated around a position 30 nucleotides downstream of a TATA box, and near to the first ATG codon of the E6 orf, which is in the optimal translation initiation context ACCATGG according to Kozak rules (Kozak, 1986).

Three classes of cDNA were isolated (Fig. 6), distinguished by their differential splicing patterns. cDNA's of class 1 consisted of a 5' exon spanning the whole of the E6 and E7 orf's, spliced from the fifth codon of the E1 orf to cellular sequences downstream of the integration site. In cDNA's of class 2, the E1 splice site was also utilised but in addition, internal sequences from the E6 orf were spliced out. This created a novel orf resulting from the in-frame fusion of the 5' part of E6, including its translation initiator, with the 3' end of another orf within E6 but in a

different reading frame, 180 nucleotides downstream. The new orf, designated E6*, encoded 57 amino acids in a sequence distantly homologous to those of epidermal growth factors.

Class 3 cDNA's also contained the E6* orf, but were not spliced within E1 and terminated at an Eco R1 restriction endonuclease site; this termination was presumably due to a cloning artifact.

Analysis of viral/human fusion sequences in the cDNA's showed an E1-human fusion orf of only 24 codons, and the maximum size of any cellular orf was 87 codons. As with viral-human sequences analysed in a genomic library of SiHa cells (Baker et al., 1987), no homologies were found between the human sequences represented in the cDNA's and any known human gene. However, E6- and E6*-specific DNA probes hybridised to a variety of RNA's in HeLa, C4-1 and SW756 cells, predominantly the 1.5 kb size class of which the majority were of the E6* type in HeLa cells, but also to larger transcripts up to 6.5 kb in size. Thus transcription from the E6 proximal promoter may continue well into cellular sequences in these cell lines, and the expression of endogenous genes by the viral promoter cannot yet be ruled out.

As with the VX2 carcinoma, it appears that viral integration and the malignant phenotype are associated with the transcription and presumably expression of sequences derived from the E6 and E7 orf's of the resident viral DNA. Co-transcribed viral sequences have an unknown significance at this stage.

The conservation of transcriptional organisation of different

papillomas in cancers affecting different species suggests either that it is a feature of the conserved viral integration site, or that the E6 and E7 orf's and their spliced derivatives have some association with progression to and/or maintenance of the malignant phenotype. The latter interpretation is supported by the findings that the E6* orf has some homology to epidermal growth factors; if it functions as such, "viral transformation" may occur by clonal expansion of a target cell for a subsequent oncogenic event, due to the activity of the E6* product.

In ID13 cells, similar E6- and E7-containing transcripts are transcribed at an increased level from the BPV1 genome [1.2.2.(i)(b)], and the transcriptional organisation of HPV16 in cervical carcinoma cells containing both integrated and episomal viral DNA is the same as in CaSki cells. This transcriptional pattern may be governed by cellular factors expressed only in the transformed state, and may be part of a cascade of transforming events. On the other hand, the persistence of the viral genome may be merely incidental, having played an early role in oncogenesis such as the expansion of a cellular target as mentioned above.

Whatever the case, the isolation of cDNA's of the transcripts presents exciting potential for the functional analysis of HPV18 in transformed cells.

1.2.3. Transcriptional control in papillomaviruses: cis- and trans-acting factors.

1.2.3.(i) A transcriptional enhancer element in the late region of BPV1.

The first report of enhancer activity in BPV1 was by Campo et al. (1983; Fig. 2): a subgenomic fragment 3' to the early region and 2.5 kb away from the nearest downstream promoter was found to enhance the expression of the vector-linked herpes simplex virus thymidine kinase gene in an orientation independent manner. Detailed mapping located the enhancer to a 40 bp fragment, nt4394 to nt4433, and showed that it consisted of two essential domains separated by an octamer (Weiher and Botchan, 1984). These domains contained consensus sequences for enhancer activity. The enhancer exhibited three fold greater activity in bovine cells than in mouse fibroblasts (Spandidos and Wilkie, 1983), indicating that cell-specific factors were involved in its activation. The effect of this enhancer in vivo is unknown. BPV1 mutants carrying deletions of the 3' enhancer element do not show impaired transformation efficiency in mouse C127 cells (Howley et al., 1985). It is possible, therefore, that it drives transcription from P_L and the putative upstream promoter of transcriptional initiation at nt4096 in response to a keratinocyte-specific factor.

1.2.3.(ii). A viral trans-acting transcriptional enhancer.

A factor which activated transcription from the non coding region (ncr) of BPV1 was mapped to the E2 orf (Spalholz et al., 1985; Yang et al., 1985b). Termination mutants were engineered into a cDNA encoding the E2, E3, E4 and E5 orf's and expressed under control of the SV40 early promoter. Enhancer activation in the ncr was detected by the simultaneous transfection into CV1 cells of this construct and a derivative of the pA10CAT vector in which the BPV1 ncr was cloned upstream of the enhancer-deleted SV40 early promoter, driving the transcription of the chloramphenicol acetyl transferase (CAT) gene. When a construct containing the intact E2 orf was transfected, CAT activity in the cells was restored to levels expressed by the intact SV40 early enhancer/promoter. This enhancer activation was also seen in C127 cells with the E2 orf expressed by a suitable promoter. Similar experiments using the HPV16 E2 orf and ncr demonstrated the same trans-activation function (Phelps and Howley, 1987). Quantitation of CAT gene transcription by primer extension also demonstrated that enhancer activation operated at the level of transcription in both cases (Spalholz et al., 1985; Phelps and Howley, 1987).

1.2.3.(iii). E2-responsive elements in the ncr of BPV1.

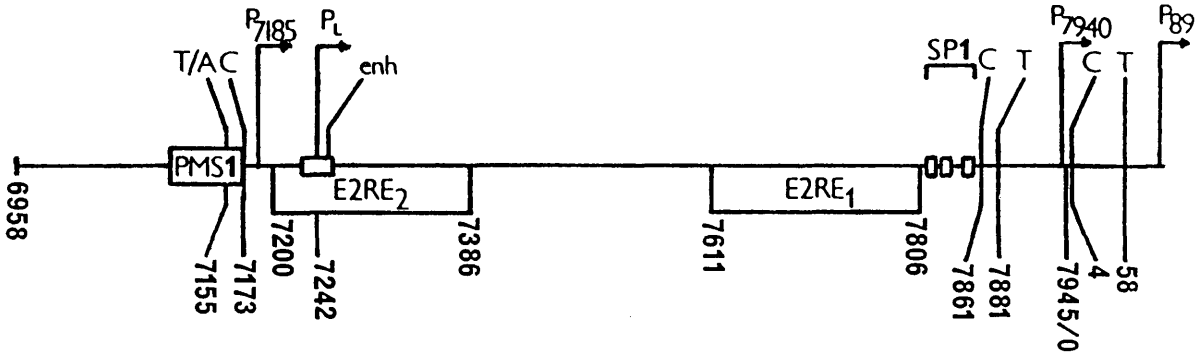
It was clear from the above experiments that the translational product(s) of the E2 orf of papillomaviruses

acted on the viral ncr to enhance transcription. Experiments were therefore carried out on the BPV1 system to dissect this function. Transcriptional regulatory elements acting in *cis* were defined by two separate experimental approaches.

One approach was to study the effect of progressive deletions from either end of the ncr on its responsiveness to E2 *trans*-activation, using the CAT assay as above (Spalholz *et al.*, 1987). Two E2-responsive elements (E2RE's) were found. The first of these was an orientation-independent element located in a fragment spanning nt7611 - nt7806, which was termed $E2RE_1$ (Fig. 7). Further deletion of this fragment from nt7806 to nt7768 decreased its E2-activated enhancer activity to approximately 50 % of the original level, and a further deletion of only 7 nucleotides abolished enhancer activity altogether. Deletions from the other end of the element to nt7656 also reduced CAT expression to basal levels in this assay, showing that two separate but dependent sequences were required for enhancer activity. Sequence analysis of $E2RE_1$ revealed a repeated motif, $ACCN_6GGT$, which occurred in two paired arrangements, each located around the deletion sites which impaired enhancer activity. Deletion from nt7806 to nt7768 removed one of these motifs, showing a requirement for two intact pairs of motifs for full $E2RE_1$ activity. Deletion of a pair of motifs at either end of the element eliminated enhancer activity.

The second E2-responsive element, $E2RE_2$, was mapped to a region between nt7200 and nt7386 (Fig. 7). Although deletion of this element did not affect the overall enhancer activity

Figure 7. Non coding region of the BPV1 genome.



LEGEND

SP1	SP1 binding site consensus sequence
C	CAAT box
T	TATA box
A	polyadenylation signal
P	transcription initiation site
enh	enhancer core sequence
E2RE	E2 protein responsive element

of the ncr, in isolation from $E2RE_1$ it sustained transcriptional enhancement in this assay at 20 % of the maximum level. This may indicate that it does not act on the SV40 promoter when distanced from it by the ncr. $ACCN_6GGT$ motifs were also found in $E2RE_2$, and deletion of one was shown to reduce its enhancer activity.

The responsiveness of the ncr's of HPV's to their own or heterologous E2 orf products has also been examined (Chow et al., 1987a; Phelps and Howley, 1987). Trans-activation by E2 orf's was demonstrated not only in their own ncr's but in ncr's of other papillomaviruses also. The ncr's of HPV's 1, 6, 7, 11, 16 and 18 and CRPV all responded to the E2 trans-acting factors of HPV's 1 and 11, and BPV1. In a reciprocal experiment, it was shown that HPV16 E2 trans-activated the BPV1 ncr E2RE's (Phelps and Howley, 1987) No E2RE activation was observed when the BPV1 ncr was transfected in the pA10CAT vector into HeLa, CaSki or SiHa cells. E2RE activation was observed in these cells, however, when the BPV1 E2 orf was co-transfected under a constitutive promoter, showing that it was the absence of an E2 trans-activator, rather than a cellular repressor, which prevent activation of the BPV1 ncr. It was therefore assumed that the ncr's of the integrated HPV genomes in these cells must also be inactive.

The cross-reactivities displayed by E2 orf's of different papillomaviruses may result from the conservation of amino acid sequence seen in this orf, particularly towards the 3' and 5' ends (Giri and Danos, 1986). This also imparts some

significance to the $ACCN_6GGT$ motif, also conserved in the ncr's of papillomaviruses.

This significnace was defined in another experimental approach aimed at locating the site of E2 enhancer activation. The 3' three-quarters of the E2 orf of BPV1, including the first ATG, was expressed in bacteria in a prokaryotic expression vector. The purified fusion polypeptide was used to raise an antiserum in rabbits, which was used to immunoprecipitate radiolabelled subgenomic fragments bound to the E2 protein (Androphy et al., 1987). DNA fragments thus precipitated were analysed for molecular weight and thereby identified. Fragments from the ncr bound to the fusion polypeptide; these contained sequences from what were later defined as the E2RE's. The finding of the repeated $ACCN_6GGT$ motif prompted the use of a double-stranded oligonucleotide containing the sequence ACCGTCTTCGGT in competition assays with the labelled subgenomic fragments. The unlabelled oligonucleotide competed with the labelled fragments for binding to the E2 fusion protein. In addition, the use of higher concentrations of protein identified a subgenomic fragment, nt2904 - nt3259, which also bound to the E2 product. This fragment contained a similar E2RE motif.

It can be concluded from these experiments that the enhancer activation function encoded by the E2 orf is associated with the binding of a protein product of the orf to conserved motifs near the main promoters of the viral genome. Enhancer activation takes place through an increase in transcription; the mechanism of activation is unknown. The requirement for

two separate pairs of binding motifs may indicate cooperative binding of an E2 dimer or multimer. Experiments have not been carried out to ascertain the number of binding sites on the E2 protein. Palindromic symmetry of the motif means that there are up to 8 binding sites in $E2RE_1$; the deletion of one or two of these resulted in drastically reduced enhancer activity, favouring the cooperative model.

The presence of a sequence in $E2RE_1$ which resembles the polyomavirus α enhancer (Herbomel et al., 1984), also invokes the interaction of cellular factors with this element. The position of this sequence between the two interdependent pairs of E2-binding motifs suggests that binding of the E2 trans-activator might alter the tertiary structure of the element in a way which allows the interaction of such a cellular factor. E2 enhancer activation may also be mediated by interactions of the E2 protein itself with cellular factors. The fact that the E4-overlapping region of E2 is not conserved between papillomaviruses infers a possible cell-specific function such as this.

1.2.3.(iv). The E2-encoded transcriptional repressor.

Treatment of ID13 cells with cycloheximide led to a 10 fold increase in levels of viral transcripts after only one hour (Kleiner et al., 1986). BPV1 mutants containing a stop codon immediately upstream of the second ATG of the E2 orf responded in similar fashion in cycloheximide-treated C127 cells, and this was thought to negate the role of the E2 orf

in this phenomenon. The translation inhibitor had no effect, however, on papillomavirus transcription in HeLa, C4-1 or SiHa cells. These data indicated that viral transcription in ID13 cells is under the control of a labile protein repressor not found in the cell lines.

Genetic analysis established several phenotypic classes of E2 mutants in BPV1, suggesting that more than one function was encoded by the orf (Rabson et al., 1986). Furthermore, the utilisation of the two early region transcriptional start sites, P_{2443} and P_{3080} in BPV1, and of different splice acceptors sites in the E2 orf in all papillomaviruses so far characterised, would lead to the expression of both the full-length E2 orf and a truncated version. In BPV1, the second ATG of E2, at nt3091, is both downstream of P_{3080} and in an optimal context for translation according to Kozak rules. The common 3' end of both putative polypeptides is not transcribed in any of the cell lines used in the translation inhibition experiments (Kleiner et al., 1986).

A BPV1 cDNA containing nt845 - nt4203 of the genomic sequence was cloned into a plasmid under control of the SV40 early enhancer/promoter and cotransfected into the E2 activator-BPV1 ncr CAT assay system described previously (Spalholz et al., 1987). E2 enhancer activation was reduced 15 fold in this experiment. When this cDNA was expressed in C127 cells in a cotransfection with BPV1 DNA, the transformation efficiency of BPV1 was decreased 5 to 10 fold in a non-cytotoxic manner (Lambert et al., 1987).

This cDNA contained the full-length orf's E1-E5 and E8. The

insertion of translation termination signals in all of these orf's demonstrated that only the 3' part of the E2 orf was required to be intact to retain the repressor function. Primer extension analysis localised transcriptional initiation to P_{3080} in cells transfected only with the cDNA construct. Deletion of the promoter in the same construct rendered it unable to repress transcription. A trans-acting factor was therefore responsible for the observed repressor activity.

On the evidence presented by these results, the protein encoded by the 3' region of the E2 orf was termed the E2 transcriptional repressor (E2TR).

1.2.3.(v). Control of BPV1 transcription by E2 proteins.

The $E2RE_1$ was shown by primer extension analysis to act on P_{7940} and P_{89} (Spalholz et al., 1987). Transcripts initiating at these positions encode the E6, E7 and possibly E1 orf's (Stenlund et al., 1985) and the E6/E7 and E6/E4 fusion orf's (Yang et al., 1985a), and hence variations in enhancer activation will have a pleiotropic effect on viral functions. Indeed, this has been demonstrated in BPV1 mutants with deletions of the E2 orf, which have reduced transforming efficiency and viral copy number in C127 cells (Sarver et al., 1984; Rabson et al., 1986), and by the low level of transcription of papillomavirus DNA integrated into host chromosomes within the E2 orf (Smotkin and Wettstein, 1986; Baker et al., 1987).

The E2 *trans*-activator effects enhancer activation through binding to discrete sequences upstream of the promoters involved. Thus it may influence transcription from P_{3080}, since it binds to a subgenomic fragment from this region, enhancing expression of E2TR. E2TR is thought to act through the E2RE's; if so, it may have a critical regulatory function in the expression of the papillomavirus early region, including its own expression.

Preliminary results reported by Lambert and co-workers (1987) suggested that E2TR exhibits the same DNA binding specificity as the full-length E2 protein. Repression of transcription may therefore be due to competition for the E2-binding sites by the shared domain of the full-length E2 activator and E2TR. This would also suggest that the 5' domain of the E2 enhancer activator, which is absent in E2TR, mediates the activator function. Transcriptional activation could also be modulated through oligomerisation of activator and repressor molecules, reducing the cooperative effect of several activators binding together.

The phenomenon of transcriptional regulation by antagonists encoded by the same orf is also seen in the adenovirus E1a gene, in which differential splicing of transcripts leads to the expression of a transcriptional activator 46 amino acids larger than a transcriptional repressor translated from the same orf (Lillie *et al.*, 1986). The mechanism of regulation is not in this case through direct binding to *cis* elements, but a functional domain of 49 amino acids specific to the larger protein has been shown to indirectly activate the

E1a-inducible promoter (Lillie *et al.*, 1987). The dissection of protein into functional "cassettes" was demonstrated by the synthetic fusion of a domain of the yeast GAL4 *trans*-activator gene with the DNA-binding domain of the *E.coli* lexA gene. The hybrid orf translation product activated transcription from a promoter near an *E. coli* lexA operator sequence cloned in yeast (Brent and Ptashne, 1985). The activation of the early promoters in BPV1 is analogous to the activation of promoters by early gene products seen in other viruses, e.g. the SV40 late promoter by the large T antigen (Brady and Khoury, 1985); the activation of expression of the E2 and E3 genes of adenovirus by the E1a transactivator (Imperiale *et al.*, 1983); and the induction of herpes simplex virus delayed early genes by immediate early gene products (O'Hare and Hayward, 1984). This analogy is reinforced by the presence of three SP1 binding site core sequences between E2RE1 and the two promoters upon which it acts (Fig. 7; Baker and Howley, 1987).

The E2 *trans*-activator may therefore be an immediate early protein of papillomaviruses which activates viral early functions including its own modulator, E2TR. The abolition of this transcriptional control of viral DNA in carcinomas may be a factor involved in oncogenesis.

1.2.4. Control of BPV1 replication in C127 cells.

Replication of the extrachromosomal BPV1 genome in mouse fibroblasts is complex and involves factors preventing

integration and maintaining copy number at a steady level. On transfection of BPV1 DNA, either uncloned or contained in a plasmid vector, transient replication of the episome occurs at a higher rate than that of cellular chromosomes. Following this, viral replication at mitosis maintains a copy number of approximately 150 viral genomes per cell (Lusky and Botchan, 1986). How such regulation is acheived is a difficult problem to address; also, the replication characteristics of BPV1 *in vitro* do not reflect the amplification of papillomavirus DNA seen in differentiating keratinocytes. A number of factors have been characterised.

1.2.4.(i) Plasmid maintenance and viral replication: cis- and trans-acting factors.

Two regions of the BPV1 genome were found to be *cis*-essential for the maintenance and replication of the episome (Lusky and Botchan, 1984; 1986). These regions , termed plasmid maintenance sequences 1 and 2 (PMS1 and PMS2) supported the extrachromosomal replication of plasmids into which they were cloned, in the presence of the BPV1 genome which supplied a *trans*-acting factor. PMS2 was situated between nt1515 and nt1655 in the E1 orf; PMS1 was found to consist of a region homologous to PMS2 between nt7116 and nt7234, and an orientation-independent transcriptional enhancer between nt6706 and nt6793 (Fig. 2). Electron microscopy of Cairns replication intermediates in subgenomic fragments localised the origin of replication of BPV1 to the PMS1 region (Waldeck

et al., 1984).

The gene encoding the trans-acting factor involved in plasmid maintenance and replication was located by analysis of deletion mutants of BPV1. It was found that all viable mutants in the E1 orf failed to replicate in C127 cells and integrated into the cellular genome, and that these mutants had lesions of the 3' end of the E1 orf. It was therefore concluded that the 3' end of E1 encoded factor(s) responsible for plasmid maintenance and replication which acted in trans through the PMS's (Lusky and Botchan, 1986).

1.2.4.(ii) Factors affecting maintenance of copy number.

A number of molecularly cloned deletion mutants were examined which affected the copy number of BPV1 in stably transformed C127 cells. Low copy number mutants which were deleted in either E6 or E7, after transfection initially transformed C127 cells and transiently replicated to approximately 150 copies per cell. This was followed by a gradual decline in copy number down to 1 to 5 copies per cell, accompanied by reversion of the transformed phenotype (Berg et al., 1986a,b). Both these classes of mutants could be fully complemented by the simultaneous transfection of a BPV1 class I cDNA (Fig. 2; Yang et al., 1985a) cloned into an expression vector, i.e. by expression of the E6/E7 fusion orf. However, although supertransfection of the cDNA construct into established cell lines containing the low copy number mutants resulted in the restoration of the transformed phenotype, the

resident viral DNA was only slightly amplified, to 16 to 32 copies per cell. Supertransfection of these cell lines with wild type BPV1 DNA had the same effect; in this case, the incoming DNA was not amplified as might have been expected of the intact genome.

From these experiments it was inferred that the E6/E7 gene had a role in maintaining viral genome copy numbers acheived by transient replication of the incoming DNA, either by a direct interaction or by activating the transcription of gene(s) involved in maintenance of copy number (Berg et al., 1986a). Studies of BPV1 replication imply that copy number is established in cultured cells by a precisely directed and ordered series of events in two replication phases: amplification followed by maintenance. Since transcription of the E6/E7 fusion orf is itself controlled by the E2 regulatory system (1.2.3.), if the E6/E7 protein is a transcriptional enhancer then the apparent two-stage replication phenomenon is a further indication that the papillomavirus life cycle is governed by a temporal cascade of viral transcription.

1.2.4.(iii). A trans-acting repressor of replication is encoded by the E1 orf.

Failure of supertransfected wild type BPV1 DNA to replicate at all in low copy number cell lines prompted a further investigation of replication control (Berg et al., 1986b). Transcription of the low copy number mutants was extremely

low and therefore could not account for a *trans*-acting repressor of replication. Therefore the repression factor had to be endogenous to the incoming DNA, or a cellular factor induced by the resident viral DNA.

Transient replication assays of plasmid-cloned BPV1 were carried out on a range of *E1*$^-$ mutants. Replicated DNA was distinguished from unreplicated by transfecting methylated plasmids: unreplicated DNA was then sensitive to Dpn 1 restriction endonuclease digestion, whereas replicated, unmethylated DNA was not cut by this enzyme. Co-transfected mutants were distinguished by using different cloning sites in the plasmid vector, or different sizes of plasmid. Certain BPV1 mutants were lethal to C127 cells. These *trans*-complementable mutants contained frameshifts and deletions covering the 5' end of the E1 orf up to the splice donor site at nt1234, and probably therefore represented a transcriptional exon.

It was found that these mutants could replicate in C127 cell lines with resident *E6*$^-$ or *E6/E7*$^-$ low copy number mutants; if, however, they were transfected into these cells simultaneously with wild type BPV1 DNA, no replication of the incoming plasmids occurred. Transient replication assays of the E1 mutants alone in C127 cells were complicated by the lethality of the mutant, but they were considered in this report to have wild type replication characteristics, indicating that their endogenous replication factors were unaffected.

Finally, transient replication of wild type BPV1 DNA

supertransfected into ID13 cells was analysed. The incoming DNA exhibited transient replication but was maintained at only 30 copies per cell, while the resident viral DNA remained at the wild type level (Lusky and Botchan, 1984). From these data it was proposed that the 5' part of the E1 orf encodes a trans-acting negative regulator of replication. This was supported by the finding that this region of the BPV1 genome could act in trans to repress runaway replication of the SV40 origin by SV40 T antigen activity (Roberts and Weintraub, 1986).

1.2.4.(iv). A model for BPV1 replication in C127 cells.

The following model was proposed by Berg and co-workers (1986b):

Replication of BPV1 in C127 cells can be divided into two phases. In the first phase, the expression of a replication factor encoded by the 3' end of E1 (Lusky and Botchan, 1986), by the incoming plasmid, induces replication to a high copy number. This amplification phase is ended by the expression of a negative regulatory factor, encoded by the 5' end of the E1 orf, termed the replication modulator. The concerted action of both factors results in the maintenance of copy number; this is now the maintenance phase. Unpublished evidence that expression of the E1 genes was directed by the PMS1 enhancer/promoter, and that this enhancer was activated by products of the E6 and E6/E7 orf's, was cited in this report. As viral transcription of the low copy number mutants

was low, these mutants may produce only very low levels of the replication protein and modulator.

In supertransfections of ID13 cells with wild type BPV1 DNA, it might be expected that the incoming plasmid would not undergo transient replication because maintenance conditions already prevail in the cell. Instead, a short amplification phase occurs followed by maintenance of the incoming plasmid at the new, lower than wild type level. The resident viral DNA, however, remains at the same copy number. The transient replication of the incoming DNA may be caused by an excess pool of replication factors expressed by the resident viral DNA, until maintenance levels of replication protein and modulator are expressed by the new plasmid. Low copy number cell lines would have too small a pool of replication factors to effect transient replication of incoming wild type BPV1 DNA, and in this case the new DNA enters directly into maintenance phase without amplifying. Mutants lacking the negative replication factor would escape immediate entry into maintenance phase in this situation and would amplify due to production of their own replication factor.

This model has an obvious inconsistency in that it does not explain why the resident DNA in the supertransfection experiments did not undergo transient replication along with the incoming plasmid. The experiments of Roberts and Weintraub (1986) demonstrated that the activity of the E1 modulator protein was dependent on the linkage in <u>cis</u> of the BPV1 PMS1 and its proximal enhancer to the SV40 origin of replication.

It was therefore suggested that the repressor bound to the BPV1 sequences and in this way limited replication of each plasmid to once per cell cycle, implying that each episome was somehow "marked" after replication and did not re-enter the replicating pool. However, densitometric labelling of newly synthesised DNA in ID13 cells demonstrated that BPV1 episomes replicated in a random and non-selective manner throughout S-phase (Gilbert and Cohen, 1987), rendering this hypothesis impracticable. The modulatory effect of the E1 repressor may therefore have more to do with rate than selectivity of replication.

Whatever the mechanism of replication control, it is linked to very fine transcriptional modulation and may be effected by very low levels of replication factors. The fact that E1-encoding transcripts of papillomaviruses have yet to be fully characterised may reflect their low abundance or may indicate that they have very short half lives.

The E1 replication protein is also essential for plasmid maintenance. If malignant progression of papillomas is related to the integration of the viral genome, then a defect in expression of this factor may be one of the oncogenic events.

How these analyses of replication control relate to the <u>in vivo</u> situation is not known. Cultured fibroblasts probably reflect an early stage in the life cycle of BPV1 and as such are not a model for the apparently uncontrolled replication of BPV1 which occurs in the more differentiated cells of papillomas. Again, the interplay of cellular factors must be

invoked: it may be that differentiation-specific cellular factors somehow switch off the replication modulator. High levels of replication are also associated with the activation of the P_L promoter. Since this is close to PMS1 and the replication origin, it seems that this region of the genome is activated in differentiating keratinocytes.

1.2.5. The transforming functions of BPV1.

Problems encountered in the dissection of the transforming functions of papillomaviruses are compounded by two factors: the considerable overlap of open reading frames makes deletion analysis complicated, and the abolition of one function may affect the expression of other orf's encoding true transforming proteins.
The transformation functions of BPV1 are currently thought to reside in the E6 and E5 orf's.

1.2.5.(i) The E6 orf.

Since the E6-encoding transcripts are among the major mRNA species in carcinomas containing papillomavirus DNA, this orf is an obvious candidate for a transformation assay. BPV1 deletion mutants lacking the E6, E7, E8 and most of the E1 orf's transformed C127 cells with an efficiency only 6 to 8 percent of wild type (Sarver et al., 1984). In a converse experiment, a class III cDNA (Fig. 2; Yang et al., 1985a) containing the full-length E6 orf, was cloned into an expression vector under the control of SV40 transcriptional regulatory elements, and transformed C127 cells with an efficiency 3 % of wild type. The E7 and E6/E7 orf's when cloned into the same vector did not transform cells. Clearly the E6 orf, although it effected transformation in isolation from other viral sequences, required some cooperative function mediated by other viral factors to show wild type

activity. Cotransfection of the E6 construct with plasmids expressing a class VI cDNA demonstrated this cooperative effect. Class VI cDNA's themselves had intrinsic transforming efficiencies of up to 50 % of wild type; this efficiency increased to approach that of wild type in the cotransfection experiments (Yang et al., 1985a).

Experiments on the transforming activity of HPV16 DNA in baby rat kidney primary cells also demonstrated a cooperative effect (Matlashewski et al., 1987). Transformation by HPV16 DNA alone was infrequent, and only seen in conditions selective for transfected cells. Cotransfection of the activated Ha-ras oncogene, however, resulted in reproducible transformation in the absence of selection. Deletion analysis of HPV16 defined a region spanning the 3' end of E6 to a point within E7 which was absolutely required for transformation in this assay.

Antiserum to a BPV1 E6-ß-galactosidase fusion protein produced in bacteria was used to isolate a 15.5 kiloDalton (kD) protein from C127 cells expressing the E6 orf under control of the Moloney murine sarcoma virus (MMSV) promoter. The E6 protein was located in the nuclear and non-nuclear membrane fractions of these cells (Androphy et al., 1985). The amino acid sequence predicted for this protein contains four repetitions of the motif Cys-X-X-Cys, conserved in other papillomaviruses and evidently part of a 33 amino acid domain which has undergone evolutionary duplication (Giri and Danos, 1986). This arrangement is characteristic of viral nucleic acid binding proteins, including the SV40 and polyomavirus T

antigens, as is the high percentage (17 %) of basic residues. If the E6 orf does encode an enhancer activator (Berg et al., 1986b), its nuclear protein product may indeed bind to an enhancer sequence. It may therefore effect transformation by activating, and functioning cooperatively with, an oncogene. The alteration of E6 function in the internally spliced E6* orf has not been tested.

1.2.5.(ii). The E5 orf.

The intrinsic transformation function of BPV1 class VI cDNA's in C127 cells (Yang et al., 1985a) was mapped by analysis of deletion and translation termination mutants of the orf's of this region of the viral genome (Schiller et al., 1986). It was found that transforming activity could only be abolished by lesions affecting the 3' half of the E5 orf.

The E5 orf encodes a polypeptide of 44 amino acids, beginning at the first methionine. The amino acid sequence is 33 % leucine, and includes an unbroken stretch of 27 hydrophobic residues. Antisera to a synthetic peptide corresponding to the 29 C-terminal amino acids of E5 precipitated a 7 kD protein from the membrane fraction of C127 cells transformed by the E5 orf contained in a MMSV expression vector. Non-denaturing electrophoretic analysis of the immunoprecipitated protein also revealed a 15 kD form, presumably a homodimer which could be formed by the cross-linking of the Cys-X-Cys motif conserved in the E5 orf's of BPV's 1 and 2, HPV's 6 and 11 and the deer

papillomavirus. Pulse-chase analysis showed that the protein had a half-life of only two hours (Schlegel and Wade-Glass, 1987).

The E5 orf therefore encodes the smallest transforming polypeptide yet found. Although it occurs downstream of other orf's in papillomavirus transcripts, it appears to be preferentially translated due to a favourable AUG sequence context (Ahola *et al.*, 1987). Expression of E5 in transformed C127 cells and in BPV1-induced sarcomas in hamsters is quite high, yet the tumour bearing animals do not have antibodies to the protein, presumably because it is deeply buried in the cell membranes (Schlegel and Wade-Glass, 1987).

The E6 and E5 orf's are now established in the literature as transforming genes. However, neither acts *in vivo* to transform cells to malignancy in the majority of papillomavirus infections, even though transcripts encoding these genes are present in benign lesions. A true appraisal of their function awaits further analysis, as do other factors synergistic in the transformation of their host cells.

1.2.6. E4 orf expression in HPV1-induced papillomas.

Antisera against a bacterially synthesised E4 protein reacted with a variety of proteins in HPV1-induced warts (Doorbar et al., 1986). A 16/17 kD doublet and a 10/11 kD doublet accounted for up to 30 % of total cellular protein in warts, and minor species included a single 45 kD protein and doublets of 32/34 kD and 21/23 kD. Several isoelectric variants were found for each pair. The large amount of E4 protein in papillomas correlates with the relatively high level of transcripts encoding E4 and E4-fusion orfs in papillomas and condylomas induced by HPV's 1, 6b and 11 [class (a), Fig. 4; Chow et al., 1987a,b] and in fibropapillomas induced by BPV1 (class C, Fig. 2; Baker and Howley, 1987).

A cDNA representing the major transcript of HPV11 contained an E4 fusion orf potentially expressed by the transcript (Nasseri et al., 1987). The similarity between this cDNA and the class C major cDNA's of BPV1 (Fig. 2; Baker and Howley, 1987) and the major transcripts of HPV6b and HPV1 (Fig. 4; Chow et al., 1987a,b) indicate that the E4 protein is encoded by the major papillomavirus transcript in benign tumours.

These findings were confirmed in a more detailed analysis of E4 proteins in skin warts containing HPV1 (Breitburd et al., 1987). To establish the site of production of each of the E4 protein variants, warts were sectioned parallel to the skin surface and the proteins in each section analysed. The differentiated layers present in each section were recognised

by the pattern of keratins present in the lysate. The 17 kD protein was expressed at a high level in the basal layer, and at progressively decreasing levels in the differentiating layers. As expression of the 17 kD variant decreased, expression of the 16 kD protein increased, reaching its highest level in the middle layers of the warts but decreasing again in the superficial layers. Expression of both these species was barely detectable in the outer layers. The 10/11 kD doublet appeared in the middle layers and accumulated as the 16/17 kD proteins decreased, becoming the most abundant protein in the superficial layers. The other E4 variants of 45 kD, 32/34 kD and 21/23 kD were found in the middle layers by Western blot analysis.

The 17 kD protein was considered to be the full-length E4 gene product on the basis of predicted molecular weight, and the 16 kD protein a post-translational modification thereof. The 10/11 kD doublet may be a cleavage product of the larger polypeptide, or may be translated from a differentially spliced RNA (Fig. 4; Chow et al., 1987b). The latter explanation would correlate with the activation of a late promoter in the more differentiated layers of the papilloma, as is seen in BPV1 in fibropapillomas (Baker and Howley, 1987). The larger E4 proteins could result from dimerisation and polymerisation of the smaller species.

Papillomas are characterised by two populations of differentiating cells. In productively infected cells, vegetative viral replication takes place and eosinophilic inclusions accumulate; in other cells, no inclusions are

seen, and vegetative viral replication is never detected at any stage of differentiation (Croissant et al., 1985). Immunoblotting of wart sections showed that the 16/17 kD polypeptides only occurred in cells which had inclusions in the cytoplasm, and apparently aggregated in the inclusions. The non-productively infected cells appeared to undergo normal terminal differentiation and produced keratins known to be specific for this process. By contrast, the cells containing the E4 proteins did not express these keratins at any stage of their differentiation.

The functions of the E4 proteins are as yet unknown. Their existence as doublets, and their capacity to accumulate into granules are similar to the properties of filaggrin, a component of keratohyalin granules (Harding and Scott, 1985). Together with these characteristics, their great abundance and the absence of differentiation-specific keratins in cells containing E4 proteins, suggest that they are involved in the subversion of epithelial differentiation to favour viral replication and maturation. The E4 orf's of different papillomaviruse are not well conserved with regard to amino acid sequence and this has been taken to indicate tissue specificity (Giri and Danos, 1986). This may explain why productive infection by papillomaviruses is highly tissue specific.

1.2.7. Summary.

When the results of genomic, transcriptional and *in situ* immunohistochemical analyses are taken together, an emerging pattern of the molecular biology of papillomavirus infection can be seen, which reflects its biological aspects and opens up a large field of potential research.

As was suspected on the basis of histological data, the late or capsid genes are only expressed in the most differentiated layers of benign papillomas, and their transcription appears to be dependent on a "wart-specific" promoter. This implies viral interaction with a factor present only in keratin-producing cells. The smaller E4 proteins of HPV1a are also concentrated in the outer epidermal layers; this may correlate with the activation of P_L, producing a spliced transcript expressing an E4 fusion orf, smaller than the full E4 orf

A replication factor also required for maintenance of the episomal state has been mapped to the 3' end of the E1 orf. A replication repressor function has been mapped to the 5' end of this orf (Roberts and Weintraub, 1986). Extensive viral replication occurs in differentiated epithelial cells *in vivo*: this indicates a requirement to overcome the replication repressor.

The role of some of the early genes is unclear. E6 is an accepted transforming gene, present and transcribed in most cancers with integrated papillomavirus sequences; however its sequence is not conserved so it may not have the same

function in all papillomaviruses. The cys-x-x-cys motif is not confined to E6 sequences of oncogenic papillomaviruses (Giri and Danos, 1986), so cellular transformation is not necessarily mediated through E6 DNA binding properties.

Enhancer/promoter activation or repression is likely to be an important role in a differentially regulated virus. The E2 orf's of all papillomaviruses encode transcriptional enhancers, and at least in BPV1, a transcriptional repressor is translated from the same orf, suggesting a modulated system of transcriptional control. In addition, the E6/E7 orf may encode an activator of transcription, involved in replication control in the early stages of papillomavirus infection.

Papillomavirus transcripts are in most cases differentially spliced, which is to be expected for the expression of the large amount of genetic information contained in a very compact genome. Transcripts starting from the same 5' site may splice to completely different sites thousands of base pairs apart on the genome, and be expressed in different levels. The effect of variable 5' untranslated leaders in papillomavirus transcripts is unknown but must surely be significant. The simplest explanation would be that these sequences modulate translation. Translation of polycistronic messages is also modulated by the proximal sequence context, and the expression of downstream orf's may therefore occur preferentially in certain transcripts. Leader sequences may also have some bearing on differential splicing of papillomavirus transcripts.

Finally, the expression of an orf may provide a lot of data on its function and enable *in vivo* proteins to be isolated and biochemically analysed; however, these orf's may never actually be translated in their genomic form from wild type viral transcripts (e.g. E6/E7, E1/E4 fusions in BPV1), with the result that the isolated proteins may be different to their expected form (e.g. E4 of HPV1a). The production and cloning of cDNA's of mRNA's from productive papillomas circumvents this potential problem in papillomavirus analysis while providing further data on the transcripts and making possible more accurate *in vitro* modelling of viral functions.

1.3. BPV4 and bovine upper alimentary squamous carcinoma: a large animal model for papillomavirus infection and oncogenic function.

1.3.1 BPV4 in bovine upper alimentary squamous papillomas and carcinomas.

The high incidence of squamous carcinoma of the upper alimentary tract of cattle in certain areas of Scotland prompted an investigation which provided possibly the best known model for the study of papillomavirus-associated oncogenesis.

Invasive, highly metastatic squamous cell carcinomas occur at five specific sites in the bovine upper alimentary tract: the tongue, soft palate and oropharynx, oesophagus, oesophageal groove and cardia. Epithelial papillomas are found at the same sites and all histological gradations have been found from benign to malignant tumours (Jarrett et al., 1978b).

In a study of 7746 cattle slaughtered at Glasgow abattoir (Table 4), it was found that 19% had alimentary papillomas at one or more of the above sites. The incidence of alimentary cancer was 2-5%, one of the highest known. Farms of origin of the affected cattle were located; the geographical distribution of these was entirely within that of the bracken fern, Pteridium aquilinum, in Scotland and all had a history of outbreaks of bracken poisoning in their herds (Jarrett et al., 1978a; 1980).

Virions identical in e.m. morphology to papillomaviruses were

extracted from these papillomas (Jarrett et al., 1980). The purified virus was subjected to molecular biological analysis (Campo et al., 1980). The size (approximately 7200 bp) and restriction map of the viral DNA differed from those of BPV's 1 and 2, the only other BPV's known at that time; nor did it hybridise to the DNA of these viruses even under non-stringent hybridisation conditions. No serological cross-reaction to other virions was found, and it was concluded that this was a new, epitheliotropic BPV. It was later designated BPV4 (Campo et al., 1981).

Experimental injection of BPV4 into the oesophagus and palate of calves resulted in the development of papillomas histologically identical to those in the field. A typical pattern of epithelial papillomavirus infection was presented: following a 3-week latency, Stage 1 lesions appeared as small, white plaques. Viral replication occurred at a high level in the immediate sub-keratin layer (i.e. the mucosal equivalent of the stratum granulosum); viral genome copy number was as high as 100,000 in these cells (Campo et al., 1985b). Electron microscopy of the keratin layer revealed crystalline arrays of virions embedded in keratin. Stage 2 papillomas, if they had not been abraded by the passage of food, had long, hyperkeratinised fronds; they varied in size between a few millimetres and as much as 30 centimetres across. Viral replication was minimal in mature papillomas and viral content was decreased. 8 months after infection, it was difficult to find any viral DNA in the tumours, even though their appearance had not changed; they regressed

completely within 12 months (Jarrett *et al.*, 1980; Jarrett, 1985).

In experimental animals injected with BPV4 and fed a bracken containing diet, a different scenario emerged . The number of both initial and persisting papillomas was greatly increased, reflecting identically the situation in the field. The persisting papillomas lasted six years in this report, during which some animals had to be euthanased due to the development of urinary bladder cancers, which were found *post mortem* to contain BPV2 DNA; in one case, however, a carcinoma of the upper alimentary tract developed, thus reproducing the syndrome encountered in the field (Campo and Jarrett, 1987).

The toxicity and carcinogenicity of bracken fern is well documented. It has a strong immunosuppressant effect as it contains factors toxic to rapidly dividing cells, inducing aplasia in bone marrow with concomitant leucopaenia and thrombocytopaenia (Evans, W. *et al.*, 1982). A bracken diet is also known to cause chronic haematuria and bladder tumours in cattle (Ushijima *et al.*, 1983). In guinea pigs fed on bracken, those which survive the initial haematuria go on to develop urinary bladder carcinomas and carcinomas and adenomas of the small intestine. Even rats are susceptible to the carcinogenic effects of bracken, developing similar cancers to the above animals (Pamucku *et al.*, 1980)

Quercetin, a flavonoid product of bracken fern, has been identified from 40 plant flavonoids as the most mutagenic in *Salmonella typhimurium* (MacGregor and Jurd, 1978). When included in the diet of rats, it induces the same

distribution of cancers as bracken (Pamucku et al., 1980).

On the basis of these studies, Campo and Jarrett (1986) included other experimental groups in their study. Animals immunosuppressed by the drug azathioprine also contracted BPV2-associated cancers of the urinary bladder and intestine, and, when infected with BPV4, developed a high number of persisting upper alimentary papillomas.

The identical progress of these experimental animals to those in the abattoir survey shows that it is the ingestion of bracken fern over a protracted period which is responsible for amplified and persistent BPV4-induced papillomatosis, and probably the malignant progression of the benign lesions.This may be due entirely to its immunosuppressive effects, but the action of a tumour promoter cannot be ruled out.

The BPV4-bracken syndrome therefore appears entirely analogous to that of the Shope complex and EV-associated carcinoma. It is a large-scale *in vivo* model system which lends itself to investigation of the interactions between papillomavirus and environmental co-factors in oncogenesis.

1.3.2. BPV4 as a carcinogenic papillomavirus.

1.3.2.(i). Transforming activity and state of BPV4 DNA in cells.

The oncogenic potential of BPV4 is, by currently accepted criteria, beyond doubt. It induces malignant sarcomas when inoculated into hamsters (Moar *et al.*, 1986), and molecularly

cloned BPV4 DNA transforms NIH 3T3 fibroblasts *in vitro* to anchorage independence, ability to grow in soft agar and aggressive tumourigenicity in nude mice (Campo and Spandidos, 1983). A cloned subgenomic fragment of 3.5 kb was also found to be capable of transforming NIH 3T3 at efficiencies approaching that of the intact viral genome. This fragment was later shown to correspond in BPV4 to the transcribed region of HPV's in carcinoma cell lines (M. S. Campo, unpublished data). C127 cells exhibit a partially transformed morphology when transfected with cloned BPV4 DNA, but are still contact-inhibited and anchorage dependent. Treatment of such cells with TPA induces full transformation. However, no cells of bovine origin have yet been transformed by BPV4 DNA (Smith *et al.*, 1987).

BPV4 DNA exists in papillomas as a typical high copy number episome (Campo *et al.*, 1980); similarly in NIH 3T3 cells the cloned viral DNA persists as a monomeric episome, but only at 3 to 30 copies per cell (Campo and Spandidos, 1983). In TPA treated C127 cells BPV4 DNA was isolated in high molecular weight structures . Southern analysis of restriction endonuclease digests of cell DNA suggested that the viral DNA existed as head-to-tail tandem repeats, having lost the plasmid DNA. The presence of an intense extra 4 kb band in an *Eco R1* digest, and the low intensity of a wild type band, indicated that the viral DNA had undergone integration or rearrangement (Smith *et al.*, 1987). Recent work has demonstrated that BPV4 DNA is integrated into the cell chromosomes in these cells (K. T. Smith, unpublished data).

In secondary transformants, the faint wild type Eco R1 band had disappeared, but the 4 kb band was still present. The restriction pattern of BPV4 DNA in mouse tumours induced by these cells was identical (Smith et al., 1987).

Continued culture of BPV4-transformed C127 cells results in the disappearance of the viral DNA according to current detection sensitivities of 0.1 genome per cell (K. T. Smith, unpublished data). This heretofore unreported phenomenon may model the in vivo situation: although BPV4 DNA was found in papillomas, no papillomavirus DNA of any type could be detected in 68 of 70 bovine upper alimentary squamous carcinomas associated with BPV4 virion-productive papillomas (Campo et al., 1985a). Interestingly, cells explanted from alimentary papillomas exhibit a transformed phenotype and lose the resident BPV4 DNA on prolonged in vitro culture (Campo et al., 1985b).

These data led to the conclusion that BPV4 genetic information is not required for maintenance of the transformed phenotype, and suggest that viral transformation is an early event in the papilloma-carcinoma system. The suspected involvement of bracken as an oncogenic promoter of papilloma cells is supported by observation of the effect of TPA on C127 cells partially transformed by BPV4 DNA. Transcription in these cells is at a barely detectable level, but after TPA treatment both the level of transcription and the number of viral transcripts increases (Smith et al., 1987). Activation of these viral transcripts by the tumour promoter may be an event in the oncogenic process.

1.3.2.(ii) Possible mechanisms of BPV4-mediated carcinogenesis.

The characterisation of BPV4 so far has identified it as an oncogenic papillomavirus with an infective aetiology similar to those of HPV's 16 and 18 in human cervical cancer, CRPV in domestic rabbit squamous carcinomas and BPV1 in transformed cells. A minimum transforming region has been defined in NIH 3T3 cells. Transformation of C127 cells by BPV4 DNA can only be acheived in the presence of a tumour promoter: this may reflect the apparent dependence of malignant progression *in vivo* on factors contained in bracken.

However, the loss of integrated BPV4 from transformed C127 cells without phenotypic alteration, and the absence of BPV4 DNA in the vast majority of papilloma-associated bovine alimentary carcinomas represents a departure from current dogma and prompts a re-examination of the role of papillomaviruses in oncogenesis. Several formal hypotheses have been submitted (Campo *et al.*, 1985b) :

1. Papillomaviruses are initiators. The virus directs cellular transformation as an early event and therefore acts as an initiator, causing one or more heritable changes in cellular DNA. This is supported by the transformed phenotype seen in cultured cells explanted from papillomas. Bracken supplies a carcinogen and/or promoter causing full progression to malignancy , and an immunosuppressant allowing expansion of both initiated cells (thus amplifying the probability of a second, transforming event) and malignant

cells without interference from host cell-mediated defences. Finally, the loss of cell differentiation leads to non-permissivity and viral DNA rejection.

2. Indirect action. Papilloma cells may produce growth factors, or growth factor receptors which stimulate the expression of cellular proto-oncogenes in neighbouring non-infected cells. An increase in epidermal growth factor receptors has been observed in bovine cells transformed by BPV2 (Smith et al., 1987). Proliferation of cells due to stimulation by growth factors may increase the chances of a second transforming "hit" by a carcinogen. This type of event has been observed in experiments on the haemopoietic system (Wyke et al., 1986) in which lesions were induced in non-infected cell lineage by v-src expression in neighbouring retrovirus-infected stromal cells in bone marrow cell cultures.

3. The virus acts as a promoter. Papillomavirus infection causes proliferation of epithelial cells, clonally expanding initiated cells.

Whatever the mechanism of transformation may be, it appears that in the case of BPV4 at least, the viral DNA is not required for the maintenance of the transformed phenotype. This finding favours models in which early oncogenic events are important in malignant progression, which may be due to a statistical effect such as the amplification of the target site of later oncogenic events.

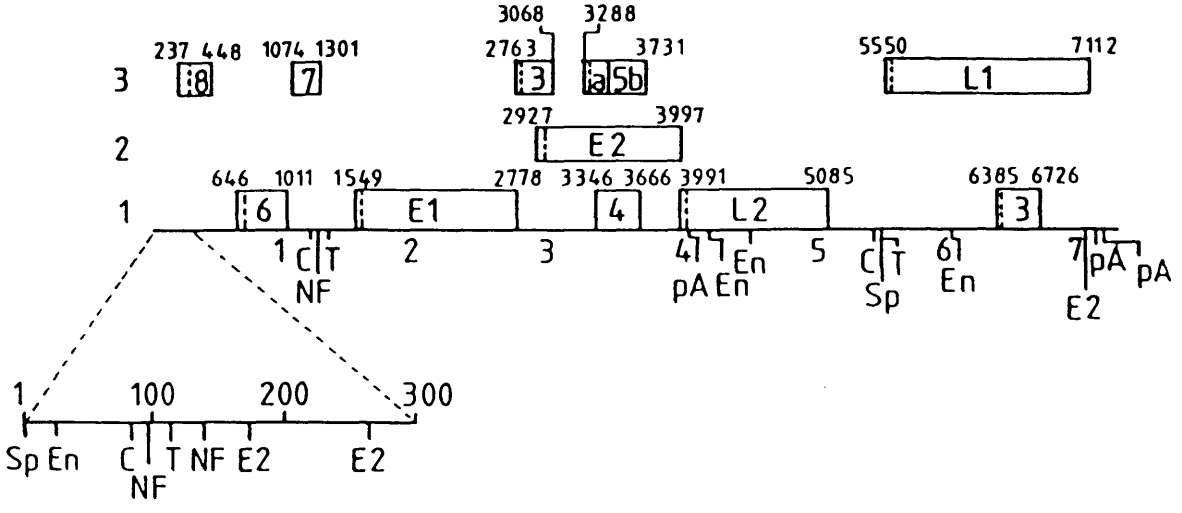
1.3.2.(iii). The BPV4 system as a large animal model for papillomavirus-associated oncogenesis.

The information presented above shows BPV4 to be a typical carcinogenic papillomavirus, and its study has several advantages over other models. It is in a controllable animal system which provides a large amount of material for experimentation, and is the only such system presently studied in which a papillomavirus in its natural host, infecting its permissive site, is associated with papillomas which progress to malignancy. This represents a more accurate model than, for instance, the Shope system in domestic rabbits. BPV4 is also a purely epitheliotropic papillomavirus, unlike BPV1, but is capable of transforming mouse fibroblasts *in vitro* and is known also to cause cancer in an heterologous host. Like some of the HPV's associated with genital cancer, BPV4 affects mucosal epithelium, and strong homologies are found between the coding sequences of these viruses (Patel *et al.*, 1987).

1.3.3. The genomic organisation of BPV4.

The entire nucleotide sequence of the 7261 bp BPV4 genome is now known (Fig. 8; Patel *et al.*, 1987). Eleven major orf's, all present on one strand of the DNA, have been assigned numbers by amino acid sequence homology with those of other sequenced papillomaviruses. Examination of the genomic sequence also identified putative regulatory sequences.

Figure 8. Genomic organisation of BPV4.



DNA is represented by a continuous line and is numbered in kilobases. The boxes represent the ORFs in the three translational frames. The map position of the ORFs is indicated in nucleotides. E, early; L, late. The ATG codons are represented by a dashed vertical line. The area of the map encompassing the first 300 nucleotides has been expanded. pA, Polyadenylation sites; C, CAAT boxes; T, TATA boxes; En, enhancer; Sp, Sp1 binding sites; NF, NF1 binding sites; E2, E2 binding sites.

Patel et al. 1987

Table 5. Transcriptional control sites and potential regulatory signals in BPV4.

Polyadenylation sites*	nt		nt		nt
ATTAAA	4009	CA	4028	G/T cluster	4040
AATAAA	7151				
AATAAA	7187	CA	7209	G/T cluster	7212

Splice sites†	Sequence	Nucleotide position
Donor	TAG.GTAAGT	204
	AAG.GTACTA	1016
	CAG.GTAGAT	1350
	CAG.GTGGAA	1779
	AAG.GTGTAG	2188
	AAG.GTAAAA	2922
	AAG.GCTAGT	2932
	AAG.GTGCTG	3770
	AAG.GTGTCA	3962
	CAG.GTTGAG	5033
	AAG.GTCAGA	5225
	CAG.GTTAAA	6908
Acceptor	CTTTGACAAG.A	1553
	$(T)_2C(T)_3CT(C)_3TCTGTTACCAG.T$	3376
	GTTCGTGCAG.C	4028
	$CC(T)_nCTACATCCCAG.T$	5541
	TGTTTTACAG.A	5595

Site		
Sp1 binding site	ACCCGCCTTG	1
	TCCCGCCTGA	5481
	$(3'^{C}_{A}CCCGCC^{CCG}_{TTA}\ 5')$	
E2 binding site	ACCGattGCGGTCG	176
	ACCGaatCGGGTGC	267
	ACCGaaaACGGTC	7046
	$(ACCGnnn^{C}_{T}CGGTGC)$	
NF1 binding site	TGGCcga..GCCAA	96
	TGGCttttgGCCA	142
	TAGAtccgcGCCAA	1276
	$(TGG^{C}_{A}nnnnnGCCAA)$	
5′ enhancer	$CTGTGTAAAG(N)_{15}CACACC$	24
	$(NTGTGG^{AAA}_{TTT}G)$	
3′ enhancer	GCTACACACC	4217
	ACACATC	4519
	3′ CCGACACAT 5′	6066
	$(GGTACACACC(N)_3CCG(N)_{11}CCACATC)$	
Transcription promoter	TGTACCTAACA	213
	(GGTACCTAACC)	
	CAAA	86
	TATAA	115
	CAAT	1242
	TATAA	1316
	CAAT	5466
	TATAA	5491

* The polyadenylation sites are numbered from the first nucleotide.

† The splice sites are numbered from the nucleotide preceding the splice in the donor or following it in the acceptor, and the dots indicate the splicing nucleotides.

The sequences in parentheses are the consensus sequences of the binding sites of Sp1 (Kadonaga *et al.*, 1986), NF1 (Hennighausen & Fleckenstein, 1986) and E2 (Androphy *et al.*, 1987); the enhancer core sequence (Sassone-Corsi & Borrelli, 1986); the BPV-1 3′ enhancer (Weiher & Botchan, 1984), and the first domain of the SV40 late promoter (Brady *et al.*, 1982). All the sequences are shown 5′ to 3′, except where indicated. The lower case letters indicate variable nucleotides.

from Patel et al., 1987.

1.3.3.(i) Potential regulatory elements.

The most notable feature of the BPV4 genome is the presence of two short non coding regions rather than the single ncr found in all other papillomavirus genomes so far sequenced. One occurs in the same position as in the other viruses, between the 3' end of the L1 orf at nt7112 and the beginning of the early region at nt237. This is designated ncr1 (Fig.8) and is only 385 bp long, compared to the average 1000 bp papillomavirus ncr. The other non coding region, ncr2, lies between the L2 and L1 orf's (nt5085 - nt5550) and is therefore 464 bp in length.

Transcriptional control signals are clustered in the ncr's, but also occur elsewhere in the genome (Table 5). Three polyadenylation signals are present: one at nt4009 at the 3' end of the transforming region and two at the 3' end of the late region, at nt7151 and nt7187. Three TATA boxes are found, one at nt115 in ncr1, one at nt1316 between the E6 and E7 orf's, and one in ncr2 at nt5491. Those in the ncr's are preceded by a CAAT box 29 and 25 nucleotides upstream, respectively; another CAAT box lies 74 nucleotides upstream of the TATA box at nt1316. Between the CAAT and TATA sequences lie single nuclear factor I (NF1) binding sites in the first two instances; a SP1 binding site is found between the same sequences in ncr2. There is also an SP1 binding site at nt1. An 11 nucleotide sequence starting at nt213 is closely homologous to the late promoter of SV40. This may represent P_L of BPV4, by analogy to BPV1.

Enhancer core sequences (Sassone-Corsi and Borrelli, 1986) are found at nt24, nt4217 and nt4519. Three E2 binding site consensus sequences (Androphy et al., 1987) are found at nt7046, nt176 and nt267; these are not paired as in BPV1, but the spacing of the latter two is similar to that in $E2RE_1$ (Spalholz et al., 1987). It is interesting to note that the E2 and NF1 binding sites are inverted forms of one another. Splice donor and acceptor sites are scattered throughout the genome. There are 12 potential donor sites but only 5 acceptor sequences.

1.3.3.(ii). The late region.

The late region of BPV4 differs from those of other known papillomaviruses (Fig. 1) in two respects: the two major orf's, L1 and L2, neither abut nor overlap but are separated by ncr2; and the L2 orf is some 400 nucleotides shorter than the average. In addition, an orf of 341 nucleotides, containing an ATG, is found within the coding sequences of the L1 orf: this is designated L3. A similarly sized orf is found in the deer papillomavirus, overlapping the 3' end of the L2 orf (Groff and Lancaster, 1985).
The amino acid sequence of L1 shows a high degree of homology with that of other papillomaviruses, as do the first 100 amino acids of L2. This is consistent with previous comparisons between papillomavirus orf's (Danos et al., 1984).

1.3.3.(iii) The early region.

The minimal transforming region of BPV4 is indicated in Fig.8. The early region of BPV4 comprises 55% of the genome, as compared with 69% in BPV1, and most of the early orf's are shorter than in other papillomaviruses. In particular, the E1 orf is 500 to 800 nucleotides shorter than other E1 orf's, but this does not affect viral replicative functions *in vivo*. It is significant that BPV4 DNA used for cell transfections was cloned into the plasmid pAT153 (Twigg and Sherratt, 1980) at the unique Bam H1 restriction site in the 3' half of the E1 orf; this may explain the low copy number in NIH 3T3 cells, and the integration of the viral DNA in C127 cells.

Homology between the amino acid sequences of E2 of BPV4 and those of other papillomaviruses is greatest among the C-terminal 100 residues, varying from 28.8 % with HPV16 to 45.6 % with CRPV. This would be expected if the E2 orf of BPV4 regulates viral transcription in the same way as has been defined for all other papillomaviruses so far characterised.

The other early orf's are widely divergent among papillomaviruses, and numbers were assigned with some difficulty to those of BPV4. BPV4 contains an E3 orf, which has only previously been identified in BPV1. The E8 orf contains an unusual peptide motif, Cys-X_4-Cys-X_3-Cys-X_2-Cys, thought to be a potential metal binding domain (Berg, 1986). The E5 orf is split into two separate orf's, as in HPV6, except that the two are contiguous and separated only by a

stop codon. The E5b orf contains no ATG; nor do the E4 or E7 orf's.

1.3.4. Transcriptional organisation of BPV4.

The transcriptional organisation of BPV4 in productive papillomas and transformed C127 cells is shown beneath the genome map in Fig. 9.

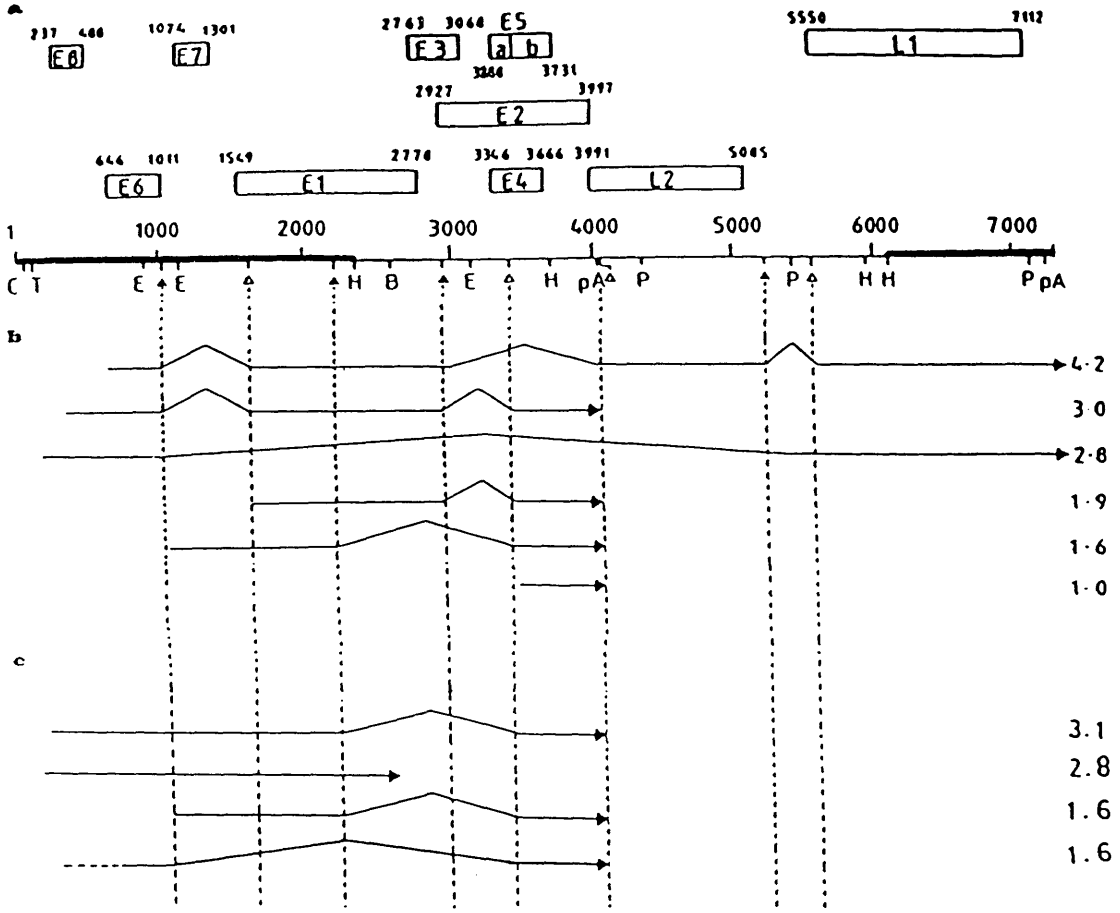
1.3.4.(i) BPV4 transcripts in productive papillomas.

BPV4-specific RNA from upper alimentary papillomas was mapped by Northern hybridisation and S1-nuclease protection analysis (Smith et al., 1986). Seven size classes were identified in Northern blots: three major classes of 4.2, 2.8 and 1.0 kb, and two minor classes of 3.0 and 1.6 kb were found in all RNA preparations; two other transcripts of 3.6 and 1.9 kb were extremely rare and occasionally absent. The 1.0 kb size class was very much the most abundant in all preparations.

Probing of Northern blots with subgenomic fragments showed that the 4.2 and 2.8 kb major transcripts hybridised to the late region, and all transcripts hybridised to at least one part of the early region.

The splicing organisation was elucidated using S1-nuclease mapping in conjunction with the results of Northern analysis. The 4.2 and 2.8 kb transcripts terminated at nt45, 119 nucleotides downstream of the nearest polyadenylation site; the 3.6 kb transcript was too rare to be mapped, but all the

Figure 9. Transcriptional organisation of BPV4.



(*a*) Genomic organization of BPV-4; (*b*) viral RNAs in oesophageal papillomas; (*c*) viral RNAs in BPV-4-transformed cells.

Smith et al. 1987

Restriction enzyme sites.

H - Hind III

B - Bam HI

E - Eco RI

P - Pst I

LEGEND.

→ Polyadenylated 3' end

▲ Splice donor site

Δ Splice acceptor site

Bold line represents transforming region of BPV4.

other transcripts terminated at the early polyadenylation site. Transcripts which originated from one strand only of the BPV4 genome were found, as predicted by the genomic analysis. The splicing arrangement was unusual in comparison to the transcriptional organisation of the other papillomaviruses. The 4.2 kb transcript was spliced three times; according to the predicted splice sites in the genomic sequence, splice junctions would have joined the following pairs of sites: nt1016 to nt1553; nt2932 to nt4028 and nt5225 to nt5595. A number of translatable orf's would be present in this transcript. From the 5' end at nt654, these include the E6 orf; a fusion of E1 to 4 codons from the end of the E6 region; a fusion of L2 to 13 codons from the E3 region; the L3 orf and the L1 orf. Which of these is expressed is a matter of speculation as none of the available AUG codons appears to be in a favourable translational initiation context over any of the others. Only the expression of a cDNA of this transcript, or any other viral transcript would provide any opportunity to analyse its translational control. The 2.8 kb transcript contained two exons, one from nt204 to nt1016 spliced to another from nt5595 to the late termination site. It encodes the L1 and L3 orf's, and the E8 and E6 orf's could also be translated from it.

The major early transcript of the 1.0 kb size class was mapped as a single exon from nt3376 to nt4030. A short distal exon such as those found in the transcripts of other papillomaviruses may have been present in the 1.0 kb transcript. The coincidence of the mapped 5' end and a splice

acceptor site, and the small size of the transcript (655 nucleotides) add weight to this argument. Also the sequence of this transcript does not contain an AUG codon in front of an orf of significant size, whereas a 5' exon could fuse a short leader orf onto the E2, E4 or E5a orf's. The abundance of this transcript indicates that it encodes the E4 protein of BPV4, similarly to the class (a) transcript of HPV1 (Doorbar et al., 1986; Breitburd et al., 1987; Chow et al., 1987b).

The exon represented by the 1.0 kb transcript was common to all the early region transcripts. The 3.0 kb transcript contained three exons, the first two of which had the same coding capacity as the first two exons of the 4.2 kb late transcript, although the start site of this transcript was at nt354. Apart from the E6 orf and the E1 fusion orf, this transcript encodes a fusion orf consisting of the first 52 codons of the E3 orf, after the first AUG, spliced to the 3' three quarters of the E2 orf, and another between the 13 codons from the E3 region and the E5a orf. As the full-length E2 orf is not encoded by any of the mapped transcripts, the E3/E2 may be a fusion analogue of the E2 trans-activator.

The 1.9 kb species consisted of two exons; it was initiated at nt1153 and spliced from nt2932 to nt3376, the acceptor site of the common downstream exon. The 1.6 kb message initiated at nt1142 and was spliced from nt2188 to the downstream exon. This transcript, which was presumably directed by the same promoter as the 1.9 kb transcript, could express a fusion of the 5' part of E1 with the E5a orf.

The coding capacities of these transcripts are only assumed according to the consensus splice signals seen in the genomic sequence. Analysis of cDNA's of transcripts is required if these splice junctions are to be properly defined and if functional analyses are to be carried out.

1.3.4.(ii). Viral transcription in BPV4-transformed cells

The level of viral transcription in BPV4-transformed NIH 3T3 cells and partially transformed C127 cells was extremely low, and only one size class of 1.6 kb could be detected (Smith et al., 1987). Treatment of these cells with cycloheximide for 2.5 hours resulted in a greatly increased level of transcription. Three more size classes of RNA appeared in Northern blots of RNA from BPV4-containing C127 cells treated in this way, of 4.7, 3.1 and 2.8 kb, as well as an increased level of the 1.6 kb transcript. The 4.7 and 3.1 kb size classes also appeared in the cycloheximide treated NIH 3T3 cells.

The amplification of viral transcription on treatment with an inhibitor of protein synthesis was also seen in ID13 cells (Kleiner et al., 1986), and may in the case of BPV4 also be due to degradation of the labile E2TR (Lambert et al., 1987). TPA treatment of the BPV4-transfected C127 cells resulted in a similar transcriptional activation in fully transformed cells, in which no further increase in transcriptional levels could be induced by treatment with cycloheximide (Smith et al., 1987). As has already been mentioned , the viral DNA in

TPA-treated cells may be integrated. The integration of HPV's 16 and 18 and of CRPV in their host cell chromosomes results in the disablement of the E2 orf; if the integration of BPV4 DNA follows a similar pattern, then the E2 repressor would be disabled, leading to an increase in transcription.

The transcripts of transformed C127 cells were roughly mapped by hybridisation to subgenomic probes. Assuming that the same promoters and splice sites are utilised in cultured cells as *in vivo*, the approximate transcriptional map is shown in Fig.9. The 4.7 kb transcript was not polyadenylated and was therefore considered to be a nuclear precursor. No structural transcripts were found, as is typical of papillomavirus infections of cells *in vitro*. The 1.0 kb transcript was not found in detectable amounts in cell lines, indicating that it is also a wart-specific transcript and lending weight to the argument that it encodes the E4 transcript.

1.3.5. Functional analysis of BPV4.

The functional dissection of BPV4 has yet to be addressed. Characterisation of other papillomaviruses, particularly BPV1, and the similarities of papillomaviruses in terms of genomic and transcriptional organisation, puts us in the fortunate position of being able to direct our analysis more precisely at areas of possible functional significance. Previous studies on papillomaviruses also show that differential splicing of viral transcripts often renders direct assessment of genomic orf functions inapplicable to

the *in vivo* situation. Functional analyses of relevance to the papillomavirus life cycle can in these cases only be carried out using cDNA's synthesised from viral RNA's.

The aim of the project described in the remainder of this thesis was to produce a cDNA library from RNA from BPV4-induced bovine upper alimentary papillomas, and to isolate and analyse cDNA's which would provide details of the splicing patterns and 5' termini of viral transcripts, and a suitable system for the functional dissection of BPV4.

SECTION 2

MATERIALS AND METHODS

MATERIALS AND METHODS: Contents.

2.1. MATERIALS

2.1.1. Chemicals

Analytical reagent grade chemicals were used. Chemicals were purchased from the following companies :

Amersham International plc.

Oligo-$(dT)_{17}$, universal sequencing primer, reverse sequencing primer, M13 sequencing kit and radiolabelled nucleotides as listed below :

Nucleotide	Catalogue no	Specific activity
[α^{32}P]-dATP	PB10204	>111 TBq mmol^{-1}
[α^{32}P]-dCTP	PB10205	>111 TBq mmol^{-1}
[γ^{32}P]-dATP	PB10218	>185 TBq mmol^{-1}
[α^{35}S]-dATPαS	SJ304	>22.2 TBq mmol^{-1}

Gibco-BRL Ltd (BRL Ultrapure Reagents).

Caesium chloride, guanidinium isothiocyanate, guanidinium-HCl oligo(dT)-cellulose, agarose, low melting-point agarose, urea

BCL.

Tris(hydroxymethyl)aminomethane (Tris).

Pharmacia

100 mM "minimum diphosphate" deoxynucleotide triphosphates : dATP, dCTP, dGTP, dTTP.

10mM dideoxynucleotide triphosphates:

ddATP, ddCTP, ddGTP, ddTTP.

Unless otherwise stated, all other chemicals were obtained from BDH Ltd., or Sigma Chemical Company.

2.1.2. Enzymes.

Restriction enzymes were obtained from either BCL or BRL with the exception of Bbv1 which was supplied by New England Biolabs.

Other enzymes were purchased from :

BCL

Deoxyribonuclease-1, Escherichia coli DNA polymerase-1 (DNA pol-1), Klenow fragment of DNA pol-1, S1-nuclease, T4-polynucleotide ligase.

BRL

Bacterial alkaline phosphatase (BAP), cloned MMTV-reverse transcriptase, ribonuclease-H (RNase-H), T4-DNA polymerase, T4-polynucleotide kinase.

Bio-Rad Laboratories

AMV-reverse transcriptase.

Pharmacia

AMV-reverse transcriptase.

Sigma Chemical Company

Lysozyme, ribonuclease-A (RNase-A).

2.1.3. Other materials.

BRL: Nucleic acid size markers-

Lambda DNA Hind III digest

ΦX174 DNA Hae III digest

RNA ladder

Eastman-Kodak Company: X-Omat XAR5 film.

PALL Process Filtration Ltd.: Biodyne nylon transfer membrane

Schleicher and Schuell: Elutip-d columns

2.1.4. Centrifugation

Preparative ultracentrifugation was carried out in a Kontron Instruments Centrikon T-2070. A Du Pont Instruments Sorvall RC-5B centrifuge was used for large-scale preparations, and a MSE Mistral 4L low speed centrifuge for pelleting bacteria. An Eppendorf 5414 "microfuge" was used for general small-scale bench-top centrifugation.

2.1.5. Stock solutions

T.E. buffer

10 mM Tris-HCl (pH 8.0)

1 mM EDTA

10 x E buffer

36 mM Tris

30.5 mM NaH_2PO_4 This gave a stock solution of pH 7.5

10 mM EDTA

10 x TEA

40 mM Tris

20 mM sodium acetate

20 mM NaCl — The pH of this buffer was adjusted to

20 mM EDTA — 8.15 with acetic acid.

10 x TBE

0.89 M Tris

0.89 M boric acid — This gave a stock solution of pH 8.3

25 mM EDTA

20 x SSC

3 M NaCl

0.3 M sodium citrate

Large volume stock solutions (i.e. ≥100 ml) were prepared in deionised, purified water. Small volume stocks, e.g. enzyme buffers, were prepared using reagent grade ultrapure water.

Phenol/chloroform

25:25:1 (v/v) T.E.-saturated phenol: chloroform: propan-2-ol. Distilled phenol was obtained from Rathburn Chemicals.

Formamide (Fluka AG)

Formamide was deionised by stirring with Bio-Rad AG501-X8(D) mixed bed ion-exchange resin for 30 min, filtered and stored at -20°C.

2.1.6. Media.

L-Broth:

1% (w/v) bacto tryptone (Difco)

0.5% (w/v) bacto yeast extract (Difco)

170 mM NaCl

L-Agar

1.5% (w/v) agar (Difco) in L-Broth, autoclaved and cooled to 42°C before the addition of any required antibiotic, etc.

2.1.7. DNA vectors and bacterial strains.

Plasmids pUC8, pUC13 (Vieira and Messing, 1982); pAT153 (Twigg and Sherratt, 1980) were used. pUC plasmids were obtained from Pharmacia.

Bacteriophage M13mp18 and M13mp19 (Norrander et al., 1983) were used as cloning vectors for DNA sequencing. M13 vectors were obtained from BRL

The pUC plasmids and recombinants were propagated in E. coli strain JM83; Plasmid pAT153 and the recombinant pBV4 were propagated in E. coli strain HB101. M13 recombinants were grown in E. coli strain JM101.

2.1.8. Oligonucleotides.

Oligonucleotides were synthesised at the University of Glasgow Biochemistry Department by Dr. V. Math.

2.1.9. Biological samples.

Bovine upper alimentary papillomas were supplied by the University of Glasgow Veterinary School.

2.2. METHODS

2.2.1. General

2.2.1.(i) Sterilisation

All glassware and most solutions were sterilised by autoclaving for 20 min. in a pressure cooker, using 3M autoclave tape as an indicator. Solutions containing heat-labile components were filtered through Gelman Acrodisc filter discs of pore size 0.2 µm. Glassware used for RNA work was first siliconised by immersion in 2% (v/v) Baysilone siliconising fluid (Bayer) for 15 min.,followed by two washes in water. It was then rendered RNase-free, as were all solutions used for RNA manipulation, by the addition of 0.1% (v/v) DEPC with vigourous shaking. After several hours' treatment, traces of DEPC were removed by autoclaving.

2.2.1.(ii) Storage

DNA samples were dissolved in T.E. (pH 8.0) and RNA in water. DNA samples which would be used within a few days of preparation were stored at 4°C. RNA, and samples of DNA which were not required immediately, were stored at -20°C. Small-volume (≥1 ml) stock solutions, e.g. restriction enzyme buffers, were frozen at -20°C. All frozen samples were stored in small aliquots to avoid repetitive thawing and re-freezing. All enzymes used in this project were stored at -20°C. Papillomas were plunged into liquid nitrogen upon collection, and stored at -70°C thereafter. Bacterial stocks

were also stored at -70°C.

2.2.1.(iii) General procedures in nucleic acid work.

Steps were taken to ensure the minimum possible exposure of nucleic acids to degradative conditions.

General procedures (1): extraction and purification.

Extraction procedures for the purification or concentration of DNA common to most of the protocols detailed in this thesis are described below.

Phenol extraction

This was used to remove contaminating proteins, or enzymes not required for the next step in a protocol. An equal volume of phenol/chloroform (2.1.5.) was added to the DNA solution, thoroughly dispersed by vortexing for 15 s. and centrifuged for 2 min. in the microfuge. The aqueous top layer was then removed to a clean tube.

Ether extraction

Phenol/chloroform extractions were always followed by an ether extraction step to ensure the removal of any traces of phenol. Two volumes of diethyl ether were added and dispersed by shaking. The ether phase was removed after brief centrifugation.

Butanol extraction

When a large solution volume precluded the effective recovery

of DNA by precipitation, this volume was reduced by butanol extraction. An equal volume of butan-2-ol was added and dispersed by shaking. After brief centrifugation, the upper layer of water-saturated butanol was removed and further extractions performed as required.

Ethanol precipitation

DNA was precipitated from solution by the addition of one-tenth volume of 3 M sodium acetate followed by two volumes of absolute ethanol. The ethanol was added straight from storage at -20°C, mixed by inversion and left on dry ice (-70°C) for 5 to 10 min. The DNA precipitate was then pelleted by centrifugation in the microfuge for 15 min. This procedure is effective in sedimenting a high percentage even of nanogram quantities of DNA (Focus 7 no.4, 1985). After centrifugation, the supernatant was poured off carefully and excess salt removed from the pellet by washing in 1 ml of 70% (v/v) ethanol. After a further 5 min. centrifugation, the supernatant was again poured off and the pellet lyophilised. Where very small (ng) quantities of DNA were to be precipitated, yeast tRNA was added as a carrier to ensure good recovery.

TCA precipitation

To calculate the incorporation of radiolabelled nucleotides into newly-synthesised DNA, two small aliquots of the synthesis reaction were dried onto 5 mm squares of Whatman 3MM filter paper. One was assayed immediately by

scintillation counting; the other was washed twice in ice-cold 5% (w/v) TCA, once in absolute ethanol and dried before scintillation counting. DNA precipitates in ice-cold TCA and is therefore deposited on the filter paper while free nucleotides are washed away: hence comparison of figures obtained from scintillation counting gives an estimate of the incorporation of radioactive nucleotide.

General procedures (2): Quantitation of nucleic acids.

Optical density (O.D.) readings of solutions of DNA in T.E. (pH 8.0), or RNA in water, were carried out at a wavelength of 260 nm. Approximate concentrations were calculated according to the following data (Maniatis et al., 1982).

The approximate concentrations of solutions of various types of nucleic acids giving an $O.D._{260nm}$ of 1.0 in a quartz cuvette of lightpath 1 cm are:

DNA	50 $\mu g\ ml^{-1}$
RNA	40 $\mu g\ ml^{-1}$
oligodeoxynucleotides	20 $\mu g\ ml^{-1}$

2.2.1.(iv) Handling of bacteria

Recombinant bacterial work was carried out under Category 1 containment.

Frozen stocks of bacteria.

Bacteria were grown in liquid culture from single colonies obtained by streaking a bacterial culture onto an agar plate.

1 ml of a 5 ml overnight culture was then mixed with 1 ml of 40% (v/v) glycerol in growth medium. 0.5 ml aliquots were placed in sterile screw-cap microfuge tubes, chilled on ice and then flash-frozen in liquid nitrogen. Stocks were stored at -70°C. New cultures were inoculated with scrapings of frozen cells from the top of the glycerol stock (Hanahan, 1985).

2.2.2. Collection of tumour material.

Bovine upper alimentary papillomas were collected from the affected animal as soon as possible after slaughter. They were plunged directly into liquid nitrogen to prevent degradation of DNA and RNA by endogenous nucleases. Papillomas from different sites were segregated in order to study possible variations in their content of BPV4-specific DNA and RNA.

2.2.3. Nucleic acid preparation.

2.2.3.(i) Preparation and separation of DNA and RNA from tumour material.

The method of Chirgwin et al. (1979) was chosen. This minimises contact between RNA and endogenous RNases by the rapid disruption of the latter in a concentrated solution of the chaotropic agent guanidinium isothiocyanate. The presence of 2-mercaptoethanol reduces protein disulphide bonds,

rendering the nucleases inactive. DNA and RNA are then separated by centrifugation through 5.7 M caesium chloride.

For each preparation, approximately 1 g of frozen papillomas was ground up in 3 ml of 4 M guanidinium isothiocyanate , 0.1 M 2-mercaptoethanol, 25 mM sodium citrate, 0.5% (w/v) sodium N-lauryl sarcosine (pH 7.0), in a pestle and mortar. The addition of a little sand assisted in the grinding of the tissue. The resulting suspension was centrifuged at 4 000 x g_{max} for 5 mins to sediment the sand and remaining solid tissue. The supernatant was set aside while the grinding up process was repeated twice on the remaining tissue. Supernatants were pooled and incubated at 37^{o}C for a few hours. This incubation increases RNA yields, presumably through greater disruption of proteins.

The tissue lysate was then carefully layered onto 4 ml of filtered, DEPC-treated 5.7 M caesium chloride, 25 mM sodium acetate (pH 5.0), of refractive index 1.3990, in a 14 x 95 mm polyallomer tube. It was then covered by a layer of liquid paraffin and centrifuged at 190 000 x g_{max} in a Kontron TST41.14 rotor for 24 hr at 20^{o}C.

After centrifugation, the DNA could be seen as thick strands in the caesium chloride, while the RNA was in the form of a gelatinous pellet at the bottom of the tube. The guanidinium isothiocyanate layer was discarded, and the DNA precipitated from caesium chloride by adding 3 volumes of 70% (v/v) ethanol and storing at 4^{o}C overnight. This prevented the precipitation of caesium chloride which would occur under normal ethanol precipitation conditions. The DNA precipitate

was recovered by centrifugation at 16 000 x g_{max} for 10 mins at 4 ^{o}C. The pellet was resuspended in T.E. (pH 8.0), phenol/chloroform extracted twice and ethanol precipitated. After resuspension, $O.D._{260nm}$ was read and yield calculated. The RNA pellet was resuspended in 1 ml of 7.5 M guanidinium-HCl, 25 mM sodium citrate, 5 mM dithiothreitol (pH 7.0) to destroy any co-sedimented RNase activity, then precipitated by the addition of 0.025 volumes of 1 M acetic acid and 0.5 volumes of absolute ethanol, at -20^{o}C. The RNA precipitate was centrifuged in the microfuge for 15 min., dried, and resuspended in 0.5ml of water. A small sample was kept for an O.D. reading, and the rest frozen at -20^{o}C.

2.2.3.(ii) Isolation of polyadenylated RNA

The polyadenylated fraction [poly(A)$^{+}$RNA] of total RNA extracted from papillomas was separated by binding to an oligodeoxythymidylic acid-cellulose [oligo(dT)-cellulose) column (Aviv and Leder, 1972). Oligo(dT)-cellulose was first equilibrated with "binding buffer +" (BB+): 0.5 M NaCl, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.5% (w/v) SDS (BRL recommended buffer). Fines were removed by aspiration. A 1 cm x 2 cm column was poured and rinsed with 0.1 M NaOH, followed by neutralisation with BB+. About 1 mg of total RNA in 0.5 ml of water was added to 0.5 ml of 2 x BB+ and denatured by incubation at 65^{o}C for 10 min. It was then loaded straight onto the column, as cooling on ice caused precipitation of the SDS, and run through. The eluant was

collected and run through the column a second time. This eluant, the poly(A)$^-$ fraction, was stored at -20^{o}C. The column was then washed with 6 column volumes of BB+, followed by 8 volumes of BB-. BB- was of the same composition as BB+ except that it contained no SDS. It was used to wash out SDS which would otherwise inhibit the enzymatic reactions to which the RNA was later subjected. This step was found to be more effective if the buffers were both heated to 37^{o}C prior to loading, as SDS precipitated in the column at room temperature on cooler days. Poly(A)$^+$RNA was eluted with 2 ml of water, and ethanol precipitated, without washing, after reading O.D.$_{260nm}$. It was resuspended in 1 M LiCl and precipitated by the addition of 7 volumes of absolute ethanol. This effected the removal of any remaining traces of SDS. After resuspension in water at a concentration of approximately 1 µg µl^{-1}, poly(A)$^+$RNA was divided into 5 µg aliquots and stored at -20^{o}C.

2.2.3.(iii). Large-scale preparation of plasmid DNA (plasmid preps).

Large-scale preparations yielding mg quantities of plasmid DNA were carried out according to a method described by Maniatis et al. (1982), based on the small-scale preparation method of Birnboim and Doly (1979).

The principles of the technique are as follows: a buffered suspension of lysozyme-treated bacterial cells is brought to pH 12.0 - 12.5, and the cells lysed completely, by the

addition of alkali and SDS. At this pH, most of the high M_r DNA is denatured, while supercoiled, circular plasmid DNA remains double-stranded . Upon neutralisation of the solution, the chromosomal DNA renatures in a complex network and is precipitated along with protein-SDS complexes. The precipitated DNA and protein are then removed by centrifugation. Plasmid DNA remaining in the supernatant is then precipitated by addition of isopropanol, resuspended in a solution of caesium chloride and ethidium bromide dye and subjected to isopycnic centrifugation. Supercoiled plasmid DNA binds more dye than does open circular or linear DNA, and therefore bands further down the density gradient (Radloff et al., 1967).

The protocol used was as follows: 5 ml of L-Broth containing selective antibiotic (in this case, ampicillin at 50 $\mu g\ ml^{-1}$) was inoculated with plasmid-transformed bacteria. This was incubated at $37^{o}C$ in an orbital incubator for 8 hr., then added to 500 ml of L-Broth plus antibiotic. The 500 ml culture was shaken at $37^{o}C$ for 16 hr., decanted into Sorvall centrifuge bottles and centrifuged at 4 000 x g_{max} for 10 min at $4^{o}C$. The supernatant medium was discarded and the sedimented cells resuspended in 10 ml of 50 mM glucose, 25 mM Tris-HCl (pH8.0), 10 mM EDTA, 2 mg ml^{-1} lysozyme (Solution I) This suspension was left at room temperature for 5 min., after which 20 ml of ice cold, freshly prepared 0.2 M NaOH, 1% (w/v) SDS (Solution II) was added. This was mixed rapidly but gently by inversion to ensure efficient buffering of the alkali and thorough lysis of the cells, while minimising the

shearing of chromosomal DNA. After 10 min. on ice, 15 ml of 3 M potassium acetate (pH 4.8) (Solution III) was added and mixed by rapid inversion. The lysate was left on ice for a further 10 min. and then centrifuged at 4 000 x g_{max} for 10 min. at 4°C to sediment the precipitated DNA and protein. The supernatant was filtered through sterile gauze into 0.6 volumes of propan-1-ol, mixed and left at room temperature for 15 min. to precipitate the plasmid DNA. This was recovered by centrifugation at 16 000 x g_{max} for 10 min at room temperature, and the DNA pellet washed with 70% (v/v) ethanol and air-dried.

The plasmid DNA was then resuspended in 6 ml of T.E. (pH 8.0) containing 600 µg ml^{-1} of ethidium bromide. Exactly 6 g of caesium chloride was dissolved in this; adherence to this measurement obviated the use of a refractometer. This solution was placed in a Kontron 13.5 ml thin-wall polyallomer tube, covered by a layer of liquid paraffin and centrifuged in a Kontron TFT65.13 fixed-angle rotor at 96 000 x g_{max} for 60 - 72 hr. at 20 °C.

Centrifugation developed a density gradient of caesium chloride in which two separate bands of DNA formed, about 0.5 cm apart, halfway down the tube. Ethidium bromide staining was normally sufficiently heavy to enable visualisation of the bands by daylight illumination; in preps giving a low-yield, plasmid bands were visualised using ultra-violet illumination. The lower band, containing supercoiled plasmid DNA, was recovered and the ethidium bromide removed immediately by several extractions with

caesium chloride-saturated propan-1-ol, as light-activated ethidium bromide induces single-stranded breaks in DNA. In order to remove caesium chloride, the solution was dialysed against two changes of T.E. (pH 8.0) for 20 hr. It was then ethanol precipitated, resuspended in T.E. (pH 8.0), and quantitated by measurement of $O.D._{260nm}$. Routinely, 4 -5 mg of pAT153- or pUC-derived plasmid was recovered from a 500 ml bacterial culture.

2.2.3.(iv). Preparation of test samples of plasmid DNA.

The "mini-prep" method of Birnboim and Doly (1979) was used. 5 ml bacterial cultures were set up as described in 2.2.3.(iii) and incubated overnight. 1.5 ml of each culture was aliquoted into a 1.5 ml microfuge tube and centrifuged for 5 min. All centrifugation in this protocol was carried out in the Eppendorf microfuge. The pelleted cells were resuspended in 100 µl of Solution I and left for 5 min. at room temperature. 200 µl of ice-cold, freshly prepared Solution II was added, mixed by inversion and left on ice for 5 min. Next 150 µl of Solution III was added and mixed by light vortexing. After another 5 min. on ice, the precipitated material was removed by centrifugation for 5 min., and the plasmid-containing supernatant was extracted twice with phenol/chloroform, once with diethyl ether, and finally precipitated by the addition of 2 volumes of absolute ethanol. The precipitated plasmid DNA was centrifuged down after 5 min. at room temperature, washed in 70% ethanol,

lyophilised and resuspended in 50 µl of T.E. (pH 8.0) containing 20 µg ml^{-1} of pre-boiled RNase-A. Mini-preps were generally used for restriction analysis, and the incubation period of these was sufficient to ensure the degradation of all contaminating RNA. However, this method yields plasmid DNA in a sufficiently pure form to allow sequence analysis, so mini-preps for plasmid sequencing were pre-incubated at 37 ^{o}C for 30 mins, followed by phenol/chloroform and ether extractions and ethanol precipitation.

2.2.4. Gel electrophoresis.

All gel electrophoresis techniques were carried out as described by Maniatis et al. (1982).

2.2.4.(i). Agarose gel electrophoresis of DNA.

Vertical gels 10 cm in height and 3 mm in thickness were cast using combs with 4 mm teeth to make loading wells. Agarose was melted in 1 x E buffer; Low Melting Point (LMP) agarose for separation and extraction of DNA fragments was melted in 1 x TEA buffer. The same respective buffers were used for electrophoresis. The percentage of agarose in the gel was selected to suit the size range of fragments to be separated. Loading buffer was 30% (w/v) Ficoll 400 (Pharmacia), 0.025% (w/v) Orange E dye in 1 x E ; one-fifth volume was added to 30 µl restriction digests. Electrophoresis was in all cases carried out at 5 V cm^{-1}, with recirculation of buffer

effected by a peristaltic pump. DNA fragments were stained by soaking the gel in a solution of 0.4 μg ml^{-1} ethidium bromide in electrophoresis buffer for 20 min., and visualised and photographed on an ultra-violet transilluminator.

2.2.4.(ii). Denaturing (alkaline) agarose gels.

Single- stranded nucleic acid molecules adopt different conformations due to secondary structure caused by intra-strand base pairing. This affects electrophoretic mobility to the extent that these molecules cannot be accurately sized unless they are fully denatured. Denaturing agarose gels were therefore used for size analysis of first- and second-strand reaction products of radiolabelled cDNA synthesis.

A 1.4% agarose gel was prepared in 50 mM NaCl, 1 mM EDTA, then soaked for at least 30 min. in alkaline electrophoresis buffer (30 mM NaOH, 1 mM EDTA). Samples were ethanol precipitated and resuspended in 0.5% (w/v) Ficoll 400, 0.025% Orange E in electrophoresis buffer before loading. After running at 5 V cm^{-1} for 2.5 hr., the gel was soaked in 7% TCA for 30 min. and washed with water. It was then placed on Whatman 3MM filter paper, covered with a sheet of cellophane and dried on a slab gel dryer for autoradiography.

2.2.4.(iii). Agarose gel electrophoresis of RNA.

RNA secondary structure affects its mobility under electrophoresis and leads to inaccurate sizing of individual transcript types. It is therefore always run in a denatured conformation in electrophoretic gels.

1% (w/v) agarose gels containing 2.2 M formaldehyde were used for this purpose. The gel running buffer stock was 0.2 M morpholinopropanesulphonic acid (MOPS) (pH 7.0), 50 mM sodium acetate, 1 mM EDTA, diluted ten times before use. Gels were prepared by melting agarose in water, cooling to 60 ^{o}C and adding 10 x MOPS buffer and formaldehyde to give the appropriate final concentrations. 20 µg of total RNA in 4.5 µl of water was added to 2 µl of 10 x MOPS buffer, 3.5 µl of formaldehyde and 10 µl of formamide and denatured by incubation at 55 ^{o}C for 15 min. before loading. RNA gels were run at 10 V cm^{-1} for 3 hr. RNA size markers were stained by first soaking the gel in 0.1 M ammonium acetate for 1hr, then in 0.4 µg ml^{-1} ethidium bromide in water for 30 min. followed by destaining in water for at least 2 hrs. They were visualised and photographed in the usual way.

2.2.4.(iv). Extraction of DNA fragments from LMP-agarose gels

The method of Weislander (1979) was used, as described by Maniatis et al. (1982). The required band was located by U.V. visualisation, excised and melted at 69 ^{o}C in an equal volume of 1 x TEA buffer. After equilibration at 37 ^{o}C for 10 min.,

the melt was extracted with an equal volume of ice-cold, T.E.-saturated phenol. The use of ice-cold phenol for this step was found to greatly improve the efficiency of the extraction procedure. Two phenol/chloroform extractions were carried out followed by ether extraction and ethanol preciptation. DNA purified in this way was suitable for nick translation, 5' end-labelling and ligation to other DNA molecules.

2.2.4.(v). Polyacrylamide gel electrophoresis.

Polyacrylamide gels were poured between two 48 cm x 20 cm glass plates, one of which was "notched", held apart by spacers of required thickness and sealed with waterproof tape. The notched plate was siliconised using 2% dimethyldichlorosilane in 1,1,1-trichloroethane to facilitate its removal after electrophoresis. A 50% (w/v) stock solution was prepared, containing acrylamide and N,N'-methylene bisacrylamide in the ratio 19:1 by weight, and gels were made up by adding this stock and 10 x TBE to water to give the required percentage of acrylamide in 1 x TBE. Polymerisation was brought about by the addition of 100 µl of TEMED and 200 µl of freshly prepared 10% (w/v) ammonium persulphate per 50 ml.

2.2.4.(vi). Thin denaturing polyacrylamide "sequencing" gels.

Single-stranded DNA fragments are separated with high resolution on 0.4 mm denaturing polyacrylamide gels, which are also easily dried down for autoradiography. Nucleic acid sequencing reactions are therefore run on these gels.
Sequencing gels in this project contained 6% (w/v) acrylamide and 8 M urea and were run at 1400 V; the high concentration of urea and the high running temperature (60 ^{o}C) ensured that DNA sequencing fragments were maintained in the denatured state during the run. However, the high temperature also presented potential problems such as cracked plates and uneven sequencing ladders due to temperature differentials across the plates. These were circumvented by clamping aluminium sheets to both sides of the gel. Gels were dried down onto Whatman 3MM paper after covering with cellophane and without fixing, and autoradiographed in X-Ray cassettes.

2.2.4.(vii). Non-denaturing polyacrylamide gels for strand separation.

When a double-stranded (ds) DNA fragment is denatured and run into a non-denaturing gel, in most cases the separated strands migrate at slightly different rates, eventually resulting in their clear separation and enabling their separate recovery. The differential migration rate is thought to be due to differences in intra-strand duplex conformations (Maniatis et al., 1982).

Radioactively labelled single strands of DNA were separated on 6% (w/v) polyacrylamide gels of thickness 1.5 mm. The running temperature was kept low by applying a low voltage (6 V cm^{-1}) in order to prevent denaturation of the strands once they were in the gel. DNA samples were resuspended in 80% (v/v) formamide containing marker dyes, denatured by boiling for 3 min. and loaded into 1 cm wells. After running, the top plate was removed and the gel covered by plastic sheet and autoradiographed. High specific activity labelling allowed for autoradiographic exposures of only 2 min., during which time alignment marks were made for use in the extraction procedure.

2.2.4.(viii). Recovery of DNA from strand-separating gels.

To extract single strands of DNA from a polyacrylamide gel, an aligned autoradiograph of the gel was used as a template to cut out a slice of gel containing the radioactive band of interest. Gel slices were then chopped up finely and eluted into 250 µl of 0.5 M ammonium acetate, 1 mM EDTA in a sealed 1 ml pipette tip with a plug of siliconised glass wool at the end. After overnight incubation at 37 ^{o}C, the liquid was drained out through the glass wool, which filtered out particles of polyacrylamide. A further 250 µl of elution buffer was flushed through the gel, and eluant volume was reduced to about 25 µl by butanol extraction. The DNA was then ethanol precipitated. Recovery was 90 - 100% in most cases.

2.2.5. Restriction endonuclease digestion.

Restriction endonucleases were used according to the manufacturers' specifications. Latterly, suitable buffers were supplied by BRL and BCL with each enzyme, and these were used where appropriate. Up to 10 µg of DNA was digested in a 30 µl reaction volume. Double digests were carried out where an optimal buffer for both enzymes could be prepared; otherwise DNA samples were phenol/chloroform extracted and ethanol precipitated between digests.

2.2.6. Southern blot hybridisation analysis.

Restriction endonuclease digests of genomic DNA from papillomas, or of plasmids containing cDNA inserts, were analysed after agarose gel electrophoresis by the method of Southern (1975). This involves the transfer of DNA from the gel to nylon membrane in such a way that the precise distribution pattern of fragments in the gel is preserved on the membrane. Hybridisation to a specific, radiolabelled probe is then used to identify and map restriction fragments.

2.2.6.(i). Preparation of agarose gels for Southern transfer.

Gel markers were photographed next to a ruler for accurate sizing of bands appearing in Southern hybridisation analysis. The gel was then placed in 0.25 M HCl for 15 min.: this

brought about partial depurination of the DNA, which assists in the transfer of large molecules. It was then soaked in 1.5 M NaCl, 0.5 M NaOH for 20 min. to neutralise the acid and denature the DNA, followed by 0.3 M Tris-HCl (pH 7.5), 0.2 M NaCl for 45 min. to neutralise the alkali.

The gel was then laid on top of a 2.5 cm pile of Whatman 3MM paper saturated in 20 x SSC. About 1 l of SSC was poured around the paper to act as a reservoir. PALL Biodyne nylon membrane was cut to the exact size of the gel and layered onto it, excluding air bubbles. The buffer was soaked up through the gel by placing paper towels on top of it. By this process, DNA travels up through the gel with the buffer and becomes trapped in the membrane in the same distribution pattern as in the gel. After about 16 hr., the DNA was baked onto the membrane at 80 ^{o}C for 1 hr.

2.2.6.(ii) Radiolabelling of DNA probes by nick-translation.

DNase-1 in sufficiently low concentrations induces single-stranded "nicks" in ds DNA. These nicked fragments serve as primers for *E. coli* DNA polymerase-1 (pol-1), which digests away the 5' end of the downstream fragment and resynthesises the strand onto the 3' end of the primer, using the intact complementary strand as template (Kelly *et al.*, 1970). By supplying the enzyme with radioactively labelled nucleotides, the newly synthesised strands become labelled (Rigby *et al.*, 1977) and therefore detectable by autoradiography when hybridised to homologous sequences

immobilised on a Southern blot.

The method of Maniatis et al. (1982) was followed. 1 µg of DNA probe was nick-translated in 50 mM Tris-HCl (pH 7.2), 10 mM $MgSO_4$, 0.1 mM dithiothreitol, 50 µg ml^{-1} BSA, containing 20 µM dGTP and dTTP and 1.85 MBq each of [α^{32}P]-dATP and [α^{32}P]-dCTP. DNase-1 at a final concentration of 1 ng ml^{-1} and 1 unit of pol-1 were added to start the reaction which was incubated at 16 ^{o}C for 1 hr. Incubation at this temperature ensures more uniform labelling of the DNA as it minimises "snapback synthesis" in which the polymerase transfers to a displaced strand. Incorporation of label was measured by the TCA precipitation method. This was routinely around 50%, giving a specific activity in excess of 10^8 dpm μg^{-1}.

The labelled probe was separated from unincorporated radioactive nucleotides using an Elutip-d column. The nick-translation reaction was loaded onto the equilibrated column in 1 ml of 0.2 M NaCl, 20 mM Tris-HCl (pH 7.5), 1 mM EDTA. The column was flushed through with this buffer, leaving bound DNA which was then eluted in 0.7 ml of 1 M NaCl 20 mM Tris-HCl (pH 7.5), 1 mM EDTA.

2.2.6.(iii). Pre-hybridisation and hybridisation of Southern blots.

Liquid hybridisation of complementary DNA strands is optimised at (T_m-25) ^{o}C, where T_m is the temperature at which half the DNA molecules in a solution are denatured. T_m is

also related to $[Na^+]$, the concentration of sodium ions in the solution, and to %GC, the percentage of dG and dC residues in the DNA molecule, according to the following formula (Marmur and Doty, 1962):

$$T_m = 16.6 \log [Na^+] + 0.41 (\%GC) + 81.5\ ^oC$$

Hybridisation of BPV4 DNA, which contains about 40% dG and dC residues, was carried out in 2 x SSC, i.e. at a sodium ion concentration of 0.33 M. T_m under these conditions was 90 ^oC, so hybridisation temperature was set at 65 ^oC.

Radiolabelled DNA also binds to non-homologous sites on nylon and other transfer membranes, so these sites must blocked by other ligand-binding molecules prior to hybridisation of the probe (Denhardt, 1966). Southern blots were pre-hybridised overnight in plastic bags containing the hybridisation medium (2 x SSC, 10 x Denhardt's solution), at the hybridisation temperature in a shaking water bath. 10 x Denhardt's contained 0.2% (w/v) each of Ficoll-400, bovine serum albumin and polyvinylpyrollidone. Calf thymus DNA, sheared by sonication and denatured by boiling, was added at a rate of 200 $\mu g\ ml^{-1}$ to bind non-homologous DNA sites. When particular sequences of DNA were required to be blocked, an excess quantity of unlabelled DNA of homologous sequence was denatured and added at the pre-hybridisation stage.

Probes were denatured by boiling and added directly to bags containing Southern blots in hybridisation medium. Hybridisation was carried out for up to 24 hr.

2.2.6.(iv). Pre-hybridisation and hybridisation of Northern blots.

Hybridisation of DNA probes to RNA blots is carried out in the presence of formamide which favours the formation of complementary DNA:RNA hybrids (Casey and Davidson, 1977).

The method of Thomas (1980) was used. The hybridisation medium was 5 x SSC, 5 x Denhardt's, 50 mM sodium phosphate (pH 6.5), 0.1% (w/v) SDS, 50% (v/v) formamide. 200 $\mu g\ ml^{-1}$ of sonicated, denatured calf thymus DNA was added immediately before pre-hybridisation. Hybridisation was carried out as previously described at 42 ^{o}C in a shaking water bath.

2.2.6.(v). High-stringency washing of hybridised blots.

Southern and Northern blots were washed immediately after hybridisation in 2 x SSC, 0.1% (w/v) SDS, with agitation, to remove free probe. This wash was repeated and the blots were then subjected to a "high-stringency" wash in 0.1 x SSC, 0.1% (w/v) SDS at 55 ^{o}C in a shaking water bath (protocol supplied by PALL Process Filtration Ltd.). This wash, at T_m - 12 ^{o}C, was carried out to remove any non-specifically bound probe. Two 15 min. washes at this temperature were normally sufficient to remove non-specific signals, but in exceptional cases with Southern blots, more persistent contamination could be removed in the same solution at 65 ^{o}C without significant loss of signal.

Washed blots were dried by pressing between sheets of Whatman 3MM paper, and wrapped in plastic sheeting for autoradiography in X-Ray cassettes.

Blots could be stripped of labelled probe for re-hybridisation by incubating in 10 mM sodium phosphate (pH 6.5), 50% (v/v) formamide at 65 ^{o}C for 1hr, followed by washing in 2 x SSC, 0.1% (w/v) SDS at room temperature for 15 min. The membrane could then be pre-hybridised and hybridised with the relevant probe.

2.2.7. Synthesis of cDNA from poly(A)$^{+}$RNA.

cDNA was synthesised from oligo(dT)$_{17}$-primed poly(A)$^{+}$RNA isolated from papillomas, using the method of Gubler and Hoffman (1983). RNA-dependent DNA-polymerase (reverse transcriptase) catalyses the synthesis of DNA complementary to an RNA template by sequential addition of nucleotides onto a heteroduplexed DNA primer. The use of oligo(dT) as a primer ensures the reverse transcription of all messenger RNA's as it hybridises to poly(A) tails. Once a complementary strand of DNA has been synthesised (cDNA first strand synthesis), the hybridised RNA molecule is "nicked" at regular intervals using RNase-H, a RNA endonuclease specific for RNA:DNA hybrids. The nicked fragments provide suitable 3'-hydroxyl primers for pol-1, which is then used to synthesise the second strand of cDNA. Second strand synthesis is therefore closely analogous to the nick-translation method. T4-DNA polymerase is used to fill in recessed 3' ends, producing a

flush-ended cDNA molecule suitable for cloning.

2.2.7.(i). First-strand synthesis of cDNA.

The following protocol was used (A. Hall, Chester Beatty Institute, London, pers. comm.):

10 µg poly(A)$^+$RNA in water was heated to 90 ^oC for 10 min. to denature secondary structure, and immediately cooled on ice. The denatured RNA and 5 µg of oligo(dT)$_{17}$ were added to a solution such that the final concentrations of all components were: 50 mM Tris-HCl (pH 8.3 at 42 ^oC), 140 mM KCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, with each of the four dNTP's at 2 mM. The solution was kept on ice while 120 units of reverse transcriptase was added and mixed, and one-tenth volume was removed to another tube containing 0.37 MBq of [α^{32}P]-dCTP. The reactions were then incubated at 42 ^oC for 1 hr. EDTA was then added to a concentration of 50 mM to stop the reactions, which were then phenol/chloroform extracted, ether extracted and ethanol precipitated. The radioactive sample was divided in half: one was assayed for radioisotope incorporation by TCA precipitation, and the other was run on an alkaline agarose gel next to radiolabelled size markers to obtain a record of size distribution of first-strand synthesis products.

N.B. A pilot synthesis was first performed using globin mRNA (BRL) to check reagents before limited stocks of papilloma RNA were used.

2.2.7.(ii). Second-strand synthesis of cDNA.

Non-radioactive first-strand reaction products were resuspended in 40 mM Tris-HCl (pH 7.5), 200 mM KCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, with dNTP's each at 0.1 mM. 1.48 MBq of [α^{32}P]-dCTP was added, followed by 1 unit of RNase-H and 25 units of pol-1. The reaction was mixed and incubated at 12 ^{o}C for 1 hr., then at 22 ^{o}C for 1 hr.
The reaction was terminated by the addition of EDTA to 100 mM followed by phenol/chloroform extraction, ether extraction and ethanol precipitation. On resuspension, a tenth aliquot of second-strand reaction products was kept aside for analysis as above. An estimate of the amount of cDNA synthesised was calculated using radioisotope incorporation data and the median size of second strands as shown by autoradiography of the alkaline agarose gel.

2.2.7.(iii). Conversion of ds cDNA to flush-ended DNA using T4-DNA polymerase.

cDNA molecules synthesised by the Gubler and Hoffman method may be unsuitable for blunt-end ligation due to incomplete second strand synthesis leaving protruding 5' ends corresponding to the 3' end of the original RNA molecule. The 3' recess may be filled in using T4-DNA polymerase, which in the presence of all four dNTP's synthesises DNA complementary to the protruding strand onto the recessed 3'-hydroxyl group (Maniatis et al., 1982).

Products of the second strand synthesis reaction were resuspended in 33 mM Tris-acetate (pH 7.9), 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM dithiothreitol and dNTP's at 0.1 mM. 4 units of T4-DNA polymerase were added and mixed and the reaction was incubated at 37 ^{o}C for 10 minutes. The reaction was terminated by the addition of EDTA to 50 mM, phenol/chloroform and ether extracted and ethanol precipitated. The reaction time was short because as the concentration of free nucleotides decreases, the 3'-5' exonucleolytic activity of T4-DNA polymerase becomes dominant and, together with its 5'-3' nuclease, the enzyme starts to degrade the DNA.

2.2.8. Construction of a cDNA library in pUC13.

A cDNA library was constructed by the cloning of flush-ended cDNA into the plasmid vector pUC13 (Vieira and Messing, 1982).

2.2.8.(i). Preparation of the vector.

pUC13 was linearised using the restriction endonuclease Sma1 which induces a double-stranded break in the multiple cloning site (Figure 10). In the ligation step, recircularisation of the vector is favoured over ligation to heterologous molecules; this was prevented by removal of the 5'-phosphate groups of the linearised vector, so that only ligations between 3'-hydroxyl groups of the vector and 5'-phosphate

groups of the cDNA took place.

The vector was therefore prepared as follows:

1 µg of Sma1-linearised pUC13 was resuspended in 50 µl of 50 mM Tris-HCl (pH 8.8) and 50 units of bacterial alkaline phosphatase added. After incubation at 60 ^{o}C for 30 min., EDTA was added to 10 mM and the solution subjected to two phenol extractions followed by ether extraction and ethanol precipitation. Rigourous adherence to this method ensured that no contaminating phosphatase activity was carried through to the ligation reaction.

2.2.8.(ii). Ligation of vector to cDNA.

T4-DNA ligase catalyses the formation of phosphodiester bonds between 5'-phosphate and 3'-hydroxyl groups on adjacent molecules of DNA, accompanied by hydrolysis of ATP (Weiss et al., 1968). Hence the phosphatased vector is unable to religate, ensuring that only recombinant molecules may be propagated in bacteria. A large excess of vector over cDNA favours reactions between heterologous molecules and minimises end-to-end ligations of cDNA.

Pilot ligations were set up using 10 ng of Hae III-restricted (flush-ended) BPV4 DNA with 100 ng of phosphatased vector to check reagents. The following control reactions were also performed: 100 ng phosphatased vector plus T4-DNA ligase; 100ng vector without enzyme; 1 ng of non-restricted vector without ligase. The amount of phosphatased vector required to give a 5 molar excess over cDNA molecules was calculated

using the data from 2.2.7.(ii).

Ligations were carried out according to the method described by Maniatis et al. (1982). The DNA to be ligated was mixed in the smallest possible volume of 50 mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM spermidine, 1 mM ATP and 100 μg ml^{-1} BSA. 10 units of T4-DNA ligase were added per μg of vector and the reaction was incubated at 4 ^{o}C overnight. Ligated DNA was phenol/chloroform and ether extracted and ethanol precipitated.

2.2.9. Preparation and transformation of competent bacterial cells.

2.2.9.(i). Preparation of competent E. coli strain JM83.

Plasmid DNA may be introduced into E. coli cells which have been treated with ice-cold calcium chloride solution. This competence increases with up to 24 hr. of such treatment (Dagert and Erlich, 1979).

A variation of the method of Dagert and Erlich was used, incorporating some aspects of the simple transformation procedure of Hanahan (1985) (A. Sproul, Beatson Institute, pers. comm.)

E. coli strain JM83 (Messing, 1979), the host bacterium for pUC plasmids, was grown from glycerol stock in L-Broth overnight. 1 ml of this culture was then inoculated into 100 ml of SOB medium: 2% (w/v) bacto tryptone, 0.5% bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$

(Hanahan, 1985). This culture was incubated at 37 °C until it reached an $O.D._{550nm}$ of between 0.3 and 0.5. 50 ml of this culture was then cooled on ice and centrifuged at 1000 x g_{max} in the MSE 4L centrifuge for 7.5 min. at 4 °C. The supernatant was discarded and the sedimented bacteria carefully resuspended by swirling in 10 ml of ice-cold 10 mM $MgSO_4$. This, and all other steps, was carried out on ice or at 4 °C. The resuspended bacteria were left for 20 min., then sedimented by centrifugation at 1000 x g_{max} for 5 min. The sedimented bacteria were resuspended in 10 ml of ice-cold 50 mM $CaCl_2$, again as gently as possible. After one hour they were again centrifuged for 5 min. and this time resuspended in 4 ml of ice-cold 50 mM $CaCl_2$. The competent cells were then stored at 4 °C to await transformation.

2.2.9.(ii). Transformation of competent E. coli strain JM83.

DNA may be introduced into competent bacterial cells by incubation with the cells on ice followed by heat-shocking the cells. The simple transformation procedure of Hanahan (1985) was used.

Ligated recombinant plasmids and control samples were resuspended in water to a concentration of 1 ng μl^{-1} and 10 µl placed in plastic bijoux (Sterilin) on ice. 200 µl of competent cells was added, mixed thoroughly and left on ice for 40 min. The cells were then heat-shocked by placing the bijoux in a 42 °C waterbath for 90 s., and immediately returned to ice. 800 µl of SOB medium containing 20 mM

glucose was added to the heat-shocked cells which were then incubated at 37 ^{o}C for a "recovery period" of 1 hr. The recovery period allowed the expression of antibiotic resistance before exposing the cells to ampicillin, and the presence of glucose enhanced the recovery of cells from heat-shock.

Since prolonged calcium chloride treatment increases competence of E. coli, pilot transformations were carried out on fresh competent cells, which could then be used the next day at greater efficiency if the competence was indicated.

2.2.10. cDNA library screening.

2.2.10.(i). Selection of recombinants.

pUC13 contains two selection markers (Figure 10): the ampicillin resistance gene ß-lactamase, and a modified ß-galactosidase gene, which complements a defective gene in the host cell, JM83. Ampicillin resistance is used to select for transformants. Active ß-galactosidase cleaves 5-bromo-4-chloro-3-indolyl-ß-D-galactoside (X-gal) to give a blue colour. Since insertions in the multiple cloning site of the pUC plasmids convert them to lac^{-}, transformants containing recombinant plasmid produce white colonies when grown on medium with the chromogenic substrate, while intact pUC containing transformants are blue.

Transformed cells were spread onto 14 cm L-Agar plates containing 50 µg ml^{-1} ampicillin and 0.2 µg ml^{-1} X-gal and

grown overnight at 37 °C.

2.2.10.(ii). Screening of recombinants by Grunstein-Hogness colony-lift analysis.

A replica of the colony distribution on an agar plate may be made on a disc of transfer membrane by layering a sterile disc onto the plate and carefully peeling it off. The transferred colonies are then lysed *in situ* and their DNA baked onto the membrane for Southern hybridisation analysis (Grunstein and Hogness, 1975).

PALL nylon membrane filter discs were used to analyse pUC transformants. Dry, sterile discs were carefully layered onto the plates. Alignment marks were made for later identification of clones before peeling off the membranes. It was found that colonies were apparently transferred to the membrane in their entirety, but a few hours' incubation of the plates at 37 °C resulted in satisfactory regrowth from remaining cells. Transferred colonies were lysed by placing the dics, colony-side up, on Whatman 3MM paper saturated with 1.5 M NaCl, 0.5 M NaOH for 5 min. The discs were then neutralised on 3 M sodium acetate in the same fashion, air dried and baked at 80 °C for 1 hr.

The membranes were then analysed by hybridisation to radiolabelled BPV4 DNA as described in 2.2.6.(iii) and 2.2.6.(v). Alignment markers were radioactively labelled, and recombinants containing BPV4 sequences were identified by realigning autoradiographs of hybridised discs with the

original plates. BPV4-positive colonies were inoculated into 5 ml L-Broth + ampicillin using sterile toothpicks and cultured overnight for glycerol stocks and mini-preps.

2.2.11. Sequencing of DNA and RNA.

2.2.11.(i). DNA sequencing in bacteriophage M13 vectors.

The principles of "dideoxy" sequencing (Sanger et al., 1977) and methods for cloning and sequencing DNA in M13 vectors are clearly described in the M13 cloning and sequencing manual supplied by Amersham International. These protocols were strictly adhered to.

2.2.11.(ii). DNA sequencing in pUC plasmids.

An application of the dideoxy sequencing technique to double-stranded plasmid DNA (Chen and Seeburg, 1985) was used to sequence BPV4-positive cDNA inserts in pUC13. M13 universal primer or reverse sequencing primer is added to plasmid DNA denatured by treatment with alkali; it anneals to the complementary sites in the plasmid DNA on acidification, while the two strands of plasmid DNA remain denatured. After a precipitation step, the insert is sequenced in the normal way.

2.2.11.(iii). RNA sequencing.

5'-^{32}P labelled specific oligonucleotide primers were used to sequence BPV4 transcripts in poly(A)$^+$RNA isolated from papillomas. The method entailed dideoxy chain termination of first-strand cDNA synthesis (Geliebter, 1987).

5 ng of a 5'-^{32}P labelled oligonucleotide was annealed to 10 µg of poly(A)$^+$RNA in 10 µl of 10 mM Tris-HCl (pH 8.3), 250 mM KCl. The specificity of this annealing was optimised by denaturing the RNA at 80 ^{o}C then incubating for 45 min. at a temperature dependent upon the composition of the 17 3'-most bases of the oligonucleotide. Effective annealing temperatures, T_a, were calculated using the following "rule-of-thumb" equation:

$$T_a = 4(G + C) + 2(A + T) - 5 \ ^{o}C,$$

where G, C, A and T are numbers of guanosine, cytosine, adenosine and thymidine residues, respectively, in the 17mer. 2 µl aliquots of annealed primer/RNA were added to 3.3 µl of 24 mM Tris-HCl (pH 8.3), 16 mM $MgCl_2$, 8 mM dithiothreitol, 0.8 mM dGTP and dATP, dCTP, dTTP each at 0.4 mM, containing 5 units of AMV reverse transcriptase (Pharmacia), in each of four microfuge tubes labelled A, C, G and T. 1 µl of 1 mM ddATP was added to tube A, 1 µl of 1 mM ddCTP to tube C, 1 µl of 1 mM ddGTP to tube G and 1 µl of 2mM ddTTP to tube T. The reactions were incubated at 50 ^{o}C for 45 min. and stopped by the addition of 2 µl of formamide containing marker dyes.

They were then boiled for 3 min. and run on a sequencing gel at 48 W for 2 hr.

2.2.12. Primer extension analysis.

Primer extension analysis is a means of precisely mapping the 5' end of a transcript by synthesising cDNA onto a radiolabelled single-stranded DNA primer specific to that RNA (Jones et al., 1985).

2.2.12.(i). 5' end labelling of DNA.

5'-dephosphorylated DNA was γ^{32}P-phosphate labelled using T4-polynucleotide kinase (Chaconas and van de Sande, 1981). This enzyme catalyses the transfer of the γ -phosphate group of ATP to the dephosphorylated 5'-end of a DNA strand. Radiolabelling was acheived using [γ-^{32}P] dATP. Transfer to recessed 5' ends is poor unless they are partially denatured by heating. Since some of the ds DNA fragments used in this project as primers had recessed 5' termini, a two-stage buffer system was used in which the heat-labile components were added after the denaturation step (based on Maniatis et al., 1982):

Kinase buffer 1	Kinase buffer 2
20 mM Tris-HCl (pH 9.5)	0.5 M Tris-HCl (pH 9.5)
1 mM spermidine	0.1 M $MgCl_2$
0.1 mM EDTA	50 mM dithiothreitol
	50% (v/v) glycerol

5'-dephosphorylated DNA (2.2.8(i)) was resuspended in 2.5 µl of 10 mM Tris-HCl (pH 8.0). The quantity of ds DNA used was varied according to its size, e.g. 0.1 µg of 0.1 kbp, 0.2 µg of 0.2 kbp, etc. 0.1 µg of oligonucleotides was used. The use of excessive amounts led to low specific activity labelling as labelled nucleotide was converted to the diphosphate. 17.5 µl of Kinase buffer 1 was added to the resuspended DNA which was then heated at 90 ^{o}C for 2 mins and promptly immersed in ice/water. 2.5 µl of Kinase buffer 2 and 4.625 MBq of [γ^{32}P]-dATP in a volume of 2.5 µl. 20 units of T4-polynucleotide kinase were then added and mixed, and the reaction was incubated at 37 ^{o}C for 30 min. The reaction was terminated by the addition of EDTA to 25 mM. If the labelled DNA was to be strand-separated, 25 µg of yeast tRNA was added as carrier and the solution was ethanol precipitated, resuspended in 80% (v/v) formamide and run on a strand separating gel (2.2.4.(vii)). For use as a hybridisation probe, it was separated from radioactive nucleotide in an Elutip-d column.

2.2.12.(ii). Annealing of primers and RNA.

Hybridisation of single-stranded (ss) DNA primers to poly(A)$^{+}$RNA was carried out in 0.4 M NaCl in the presence of 80% formamide. Formamide has the effect of lowering T_m of nucleic acid duplexes, allowing the use of lower temperatures which have a less damaging effect on RNA, and RNA:DNA hybrids are more stable in the presence of formamide (Casey and

Davidson, 1977) The T_m of such hybrids is higher than that of the corresponding DNA duplex, T_m^f, which is related to formamide concentration and nucleotide frequency by the following formula:

$$T_m^f = T_m - (0.5 \times \text{mol } f^{rn}GC + 0.75 \times \text{mol } f^{rn}AT) \times \%\text{formamide},$$

(Thomas et al., 1976) where mol f^{rn}AT/GC is the mole fraction of dG, dC or dA, dT in the DNA molecule, and T_m is dependent on salt concentration [2.2.6.(iii)]. Therefore if hybridisations are carried out at the T_m^f of the DNA primer, RNA:DNA hybrids should predominate over intrastrand duplexes and ds DNA duplexes formed with any contaminating complementary strand.

Approximately 2000 cpm or 5 ng of radiolabelled, single-stranded DNA primer was lyophilised with 5 µg of poly(A)$^+$RNA and 20 µg of yeast tRNA. This was resuspended in 10 µl of 40 mM PIPES (pH 6.4), 1 mM EDTA, 80% (v/v) formamide (Ghosh et al., 1981) and sealed in glass capillaries. A control reaction was carried out in the same buffer with primer and tRNA only. The sealed capillaries were heated to 90 ^oC for 7 min. to denature the DNA and RNA, then immediately submerged in labelled test tubes in a water bath at the predetermined hybridisation temperature. After incubation, the capillaries were carefully broken and their contents flushed out with 200 µl of 0.2 M sodium acetate and ethanol precipitated.

2.2.12.(iii). Synthesis and analysis of primer extended cDNA.

First-strand cDNA was synthesised from the annealed primers using 200 units of cloned MMTV-reverse transcriptase at 42 ^{o}C (d'Alessio et al., 1987) in the reaction buffer supplied by BRL, with dNTP's at 0.5 mM. In order to increase full-length synthesis products, reactions times were extended up to 2 hr and reactions were chased every 30 min. with 200 units of fresh enzyme. The reaction was terminated by the addition of EDTA to 50 mM, and the solution was phenol/chloroform and ether extracted and ethanol precipitated in the presence of yeast tRNA. Primer extension products were then resuspended in formamide containing dye markers and run on a sequencing gel next to radiolabelled size markers. Gels were analysed by autoradiography in the usual way.

2.2.13. S1-nuclease protection analysis.

Primer extension and S1-mapping complement one another in the fine mapping of 5' ends of RNA transcripts. In S1-nuclease protection analysis (Berk and Sharp, 1977), a 5'-^{32}P labelled genomic restriction fragment from the region of interest is strand-separated and the individual strands hybridised to poly(A)$^{+}$RNA. The heteroduplexes are then digested with S1-nuclease, which degrades ss DNA but not ds DNA or RNA:DNA hybrids. Analysis of the reaction products enables 5' ends of RNA to be mapped according to the length of the protected fragments.

2.2.13.(i). Hybridisation of ss probes and RNA.

5' end labelled single-stranded probes were hybridised to poly(A)$^+$RNA or control tRNA as described in 2.2.12.(ii). Optimum hybridisation temperatures could be more accurately calculated if primer extension had previously been used to map the transcript.

2.2.13.(ii). S1-nuclease digestion.

Capillaries containing the annealed probe or appropriate control were flushed out with 235 µl of ice-cold 250 mM NaCl, 30 mM sodium acetate (pH 4.6), 1 mM $ZnSO_4$ and 200 µg ml^{-1} sonicated salmon sperm DNA. 100 units of S1-nuclease were added per reaction, and the digests incubated at 37 ^oC for 1 hr. In the case of highly A/T-rich probes, this temperature was reduced to 18 ^oC. This reduces "breathing" and hence inappropriate degradation of the A/T-rich regions (Pelham, 1982).

Reactions were terminated by phenol/chloroform extraction, then ether extracted and ethanol precipitated. Reaction products were run on a sequencing gel next to radioactive size markers and analysed by autoradiography as previously described.

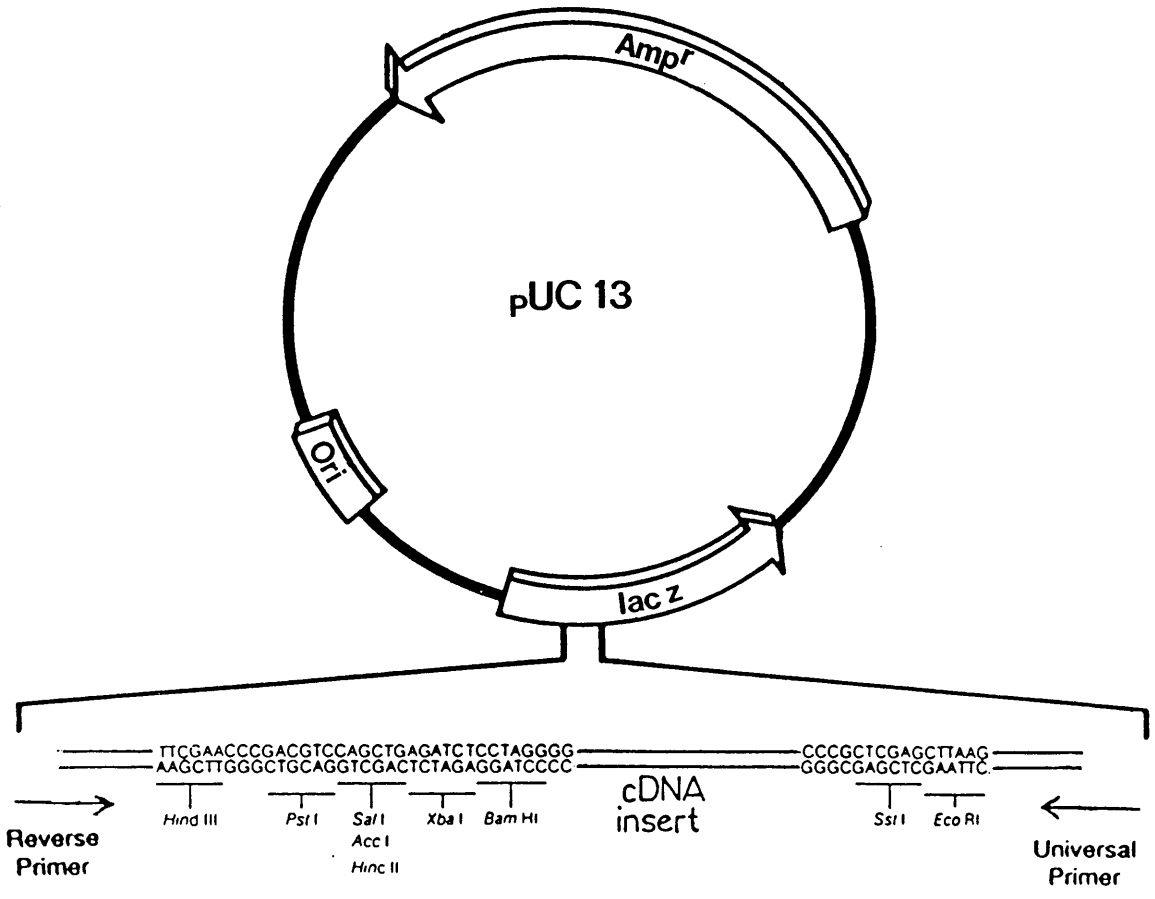


Figure 10. The multiple cloning site of pUC13.

The pUC13 vector, showing the positions of cDNA's blunt-end ligated into the SmaI restriction site.

SECTION 3

RESULTS AND DISCUSSION

RESULTS AND DISCUSSION: Contents.

3. RESULTS AND DISCUSSION.

In this section, results obtained in this project are described in appropriate subsections and a discussion of each result included under the same subsection. A summary of results and a concluding discussion are presented at the end of the section.

3.1. Isolation and analysis of DNA and RNA from papillomas.

3.1.1. Collection and selection of bovine upper alimentary papillomas.

Bovine upper alimentary papillomas were supplied by Glasgow Veterinary School Pathology Department. Bovines with multiple upper alimentary papillomatosis from farms around the Glasgow area were slaughtered at the Veterinary School. Papillomas were excised and plunged into liquid nitrogen, care being taken to include the peduncle of the papilloma while minimising the amount of surrounding normal tissue. This ensured maximum recovery of the full range of viral transcripts expressed in each differential layer of the epithelium.

For the purposes of isolating DNA and RNA, papillomas from each affected site were subdivided according to size: those $\leq$6 mm in diameter (0.4 g) were kept separate from those >6 mm (Fig. 11). These distinctions were made because it had been found previously that smaller papillomas yielded

proportionally more BPV4-specific RNA (M. S. Campo, pers. comm.). This may be due to an increased proportion among the larger papillomas of regressing tumours containing reduced quantities, or no BPV4 DNA, in which viral transcriptional levels are low or absent (Jarrett, 1985). Larger papillomas also have a larger proportion of keratinised fronds over the solid core, and hence a smaller proportion of the less differentiated tissue in which viral transcription takes place.

A single case with multiple site infection was supplied; this was designated AOP 298. This animal had multiple papillomas on the tongue, oesophageal wall and cardiac area of the rumen. Papillomas from different sites were stored and processed separately in order to detect any possible variation in the viral DNA content and transcriptional pattern at the different sites of infection. They were labelled AOP 298tong, AOP 298oes and AOP 298rum, respectively.

In most cases, papillomatosis was confined to the upper oesophagus (Table 3.1), and it was usual to collect up to 200 papillomas from this site. Oesophageal papillomas had a typical "onion" shape, the main body of the tumour having a larger girth than the basal peduncle, tapering into long, keratinised fronds. Papillomas from this site extended an average 1 cm from the basal peduncle into the oesophagus and were usually arranged in straight, longitudinal lines. This pattern may be due to viral infection occurring in lines of scarified epithelium caused by the passage of abrasive

material ingested by the animal.

Papillomas at other sites (tongue, cardiac area of rumen) were less abundant in case AOP 298 at 20 to 50 per site; tongue papillomas were flatter and more rounded in appearance, having no keratinised fronds, presumably due to mechanical abrasion. Papillomas from the rumen were in general larger and often the tubules appeared loosely bunched and could easily be separated. It was considered that these were at a more advanced stage of BPV4 infection and may have been regressing.

3.1.2. Isolation of DNA and RNA from papillomas.

It was important for the purposes of this project to obtain a preparation of papilloma RNA of the highest possible quality. It was often difficult to obtain tumour samples immediately after the slaughter of the animal, and in addition, papilloma tissue is exceptionally tough and difficult to break up, and took some time to homogenise in guanidinium isothiocyanate buffer. These factors resulted in the inevitable degradation of some or all of the RNA in wart samples.

RNA was extracted from papillomas of seven different cases; with the AOP 298 samples, this was a total of nine extractions. DNA was recovered from cases AOP 30/10, AOP3/11, AOP 298 and AOP A.

The guanidinium isothiocyanate/ caesium chloride density gradient centrifugation technique (Chirgwin *et al.*, 1979) yielded between 0.22 and 0.66 mg of DNA and between 0.47 and

1.2 mg of RNA per g of papilloma tissue (Table 3.1.).

3.1.3. Southern blot hybridisation analysis of papilloma DNA

Papilloma DNA was subjected to Southern blot hybridisation analysis in order to demonstrate the presence and integrity of the BPV4 episome. 10 µg samples of AOP A DNA were run on a 1% agarose gel either undigested (Fig. 12(1), lane A), or digested with restriction endonucleases (lanes B, C, D). After Southern transfer, a radiolabelled BPV4 genomic DNA probe was hybridised to the blot. This was isolated from plasmid pBV4 (Fig. 13). An autoradiograph of the blot is shown next to the gel in Figure 12(1); sizes of the hybridisation bands were calculated using a computer analysis program (Chuck Buckler Programs, Bethesda).

The autoradiograph showed the typical restriction pattern of BPV4 DNA (Campo et al., 1980). In the undigested sample, the most intense band corresponds to supercoiled Form I viral DNA (Fig 12(1), lane a). The faint band of lower mobility represents relaxed Form II DNA and an even fainter band between the two, linearised Form III DNA. Digestion with restriction endonuclease BamHI, in this case a partial digest, resulted in the 7.2 kb band of increased intensity in lane (c). Digestion with EcoRI and HindIII resulted in the typical hybridisation pattern (Campo et al., 1980) seen in lanes (b) and (d).

10 µg samples of the other DNA preparations were also run on a 1% agarose gel , transferred to nylon membrane and

hybridised to the BPV4 DNA probe (Fig. 12(2)). These were run undigested (Fig. 12(2), lanes 1), or after digestion with restriction enzymes BamHI (lanes 2) and EcoRI (lanes 3). In each case, the results showed that no major deletions or rearrangements had occurred in the BPV4 DNA contained in the papillomas, and little variation was seen in viral DNA content between cases, except for AOP 298tong which had an increased content (Fig. 12(2), lanes c).

3.1.4. Northern blot hybridisation analysis of papilloma RNA

20 µg of each RNA preparation was electrophoresed on a 1% agarose gel containing 2.2 M formaldehyde and transferred to nylon membrane. A nick-translated probe of BPV4 genomic DNA was then hybridised to the immobilised RNA. Autoradiographs of the resulting Northern blots are presented in Figure 14, and the calculated sizes of the hybridisation bands in each lane are tabulated in Table 3.1.

The highest quality RNA preparation was from AOP A papillomas (Fig. 14, lane e). Nine distinct hybridisation bands can be seen , of 6.4, 5.4, 4.3, 3.6, 2.8, 2.3, 2.0, 1.6 and 1.2 kb, respectively. Some of the bands are quite broad, indicating that more than one transcript size is represented, particularly the 1.2 kb band.

The major transcripts, indicated by the more intense bands, are represented by the 6.4, 5.4, 4.3, 2.8, 2.0 and 1.2 kb size classes. The last four of these were considered to correspond to the 4.2, 2.8, 1.9 and 1.0 kb transcripts

reported previously (Smith et al., 1986). The 3.6 and 1.6 kb minor transcripts have also been identified previously. The larger 6.4 and 5.4 kb size classes have not been reported. They may represent further mRNA's of BPV4 or may just be unspliced nuclear intermediates. No previous report of the 2.3 kb major size class has been made, although the high intensity of this hybridisation band indicates that it is a mature mRNA.

Variations in the overall intensity of autoradiographic bands occur for different RNA samples in Figure 14. In some cases, such as AOP 30/10 and AOP 3/11, this was due to degradation of the RNA, giving diffuse bands with high background. Preparation AOP X contained no BPV4-specific RNA (Fig. 14, lane d) ; since only 290 µg of total RNA was recoverd from 1g of tissue in this instance, this may have been due to a failure in the preparative technique.

In preparations AOP 298tong, AOP 298rum and AOP 299, there are marked reductions in the relative intensity of the 6.4, 5.4, 4.3 and 2.8 kb bands. The 4.3 and 2.8 kb size classes contain late transcripts (Smith et al., 1986) and the size of the 6.4 and 5.4 kb species indicates that they must contain sequences from the late region of BPV4. The papillomas used for these preparations were predominantly of the larger size gradation in AOP 298tong and AOP 299 and exclusively large in AOP 298rum. Therefore a larger proportion of older papillomas may have been present in these cases, and the reduction in late transcripts may relate to the decrease in virion production seen in older, regressing papillomas (Jarrett,

1985). Low virion production may also be a feature of the site of infection in the tongue and rumenal papillomas. The late transcripts of HPV's 1, 6b and 11 were found to be extremely rare in warts and condylomata, and this was similarly taken to correlate with low virion productivity in these lesions (Chow et al., 1987a,b).

The larger number of viral transcript size classes found in this study is similar to that of other papillomaviruses in vivo. BPV1-specific transcripts in fibropapilloma RNA occupy eight size classes (Engel et al., 1983) and a large number of differentially spliced mRNA's have been found (Baker and Howley, 1987; Ahola et al., 1987); fourteen different viral transcripts have been mapped in condylomata containing HPV6b and HPV11 (Chow et al., 1987a).

The main objective of screening the papilloma RNA preparations in this manner was to select RNA suitable for cDNA synthesis and transcript mapping. Preparation AOP A was considered the most suitable and a further 2 g of papilloma tissue was processed, giving an accumulated yield of 3072 µg of total RNA.

3.1.5. Isolation of polyadenylated RNA.

Poly(A)$^+$RNA was separated from total AOP A RNA using an oligo(dT) cellulose column. Measurement of optical density at 260nm showed that 120 µg of poly(A)$^+$RNA was recovered (Table 3.1). This was 3.9% by weight of total RNA, close to the value of 3% obtained by Aviv and Leder (1972) for poly(A)$^+$RNA

from rabbit reticulocyte total RNA. This was taken as an indication that the papilloma mRNA had been purified sufficiently for further analysis.

1 µg of the polyadenylated RNA was subjected to Northern blot hybridisation analysis, using the BPV4 genomic probe, to confirm the integrity of the isolated transcripts (Fig. 43, lane A; Table 3.1). The results of this analysis show a dramatic reduction in the relative intensities of the 6.4, 5.4, 4.3 and 2.0 kb size classes, while those of the 3.6, 2.8, 1.6 and 1.2 kb are maintained. The 2.3 kb size class is no longer present. This is somewhat surprising when the Northern blot of total AOP A RNA is compared. It is unlikely that selective degradation of certain transcripts occurred, so the apparent variation in their abundance may have been due to poor transfer onto the membrane in the Northern blot, or low retention in the oligo(dT) column due to short or fully degraded poly(A) tails.

Figure 11. Bovine oesophageal papillomas.

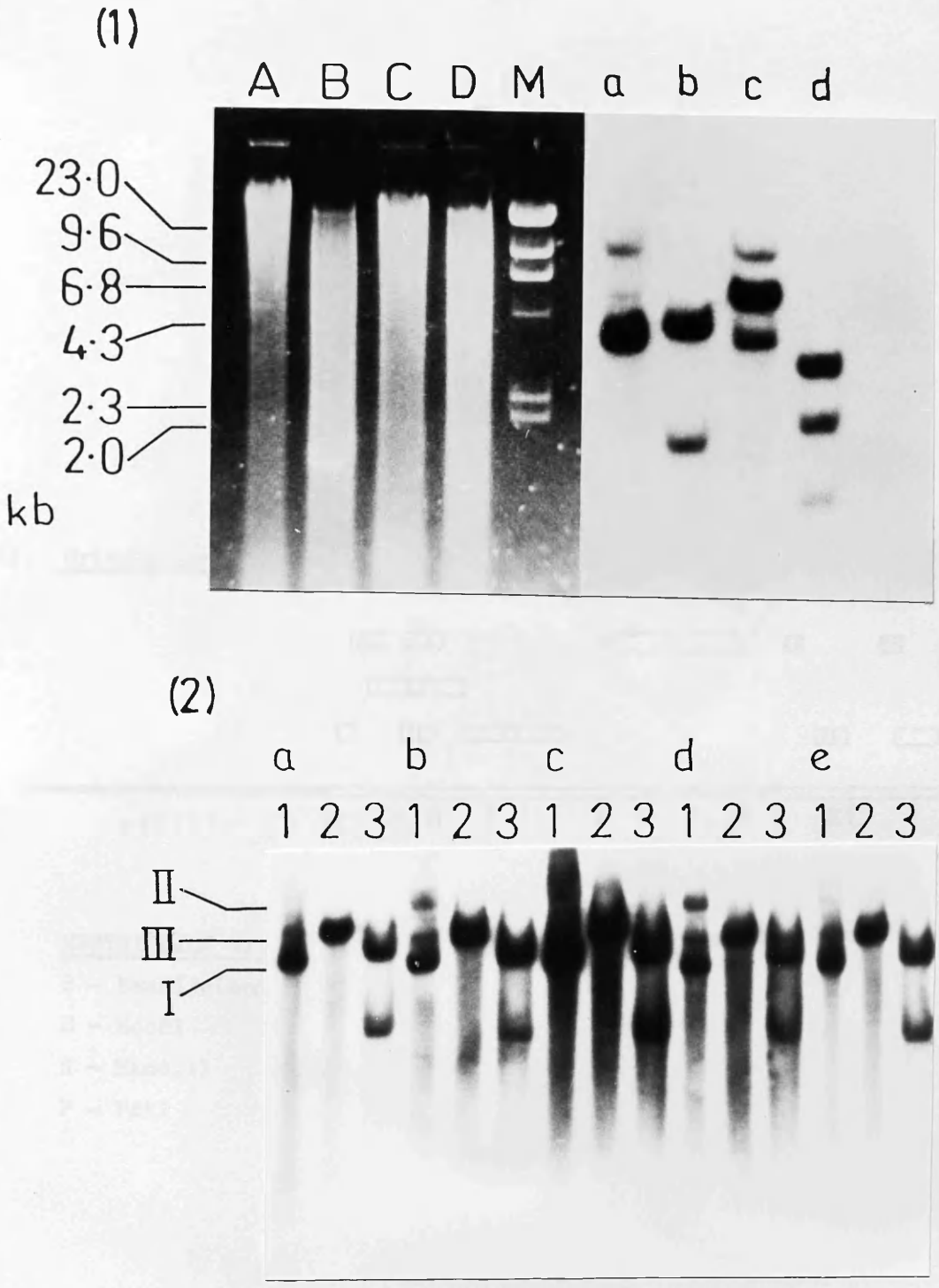


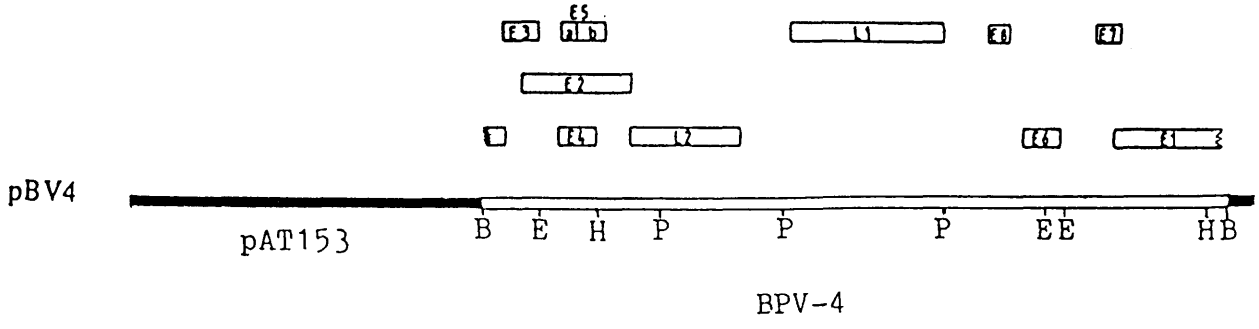
Figure 12. Southern blot hybridisation of papilloma DNA.

1). LEFT: Ethidium bromide stained 1% agarose gel containing AOP A DNA - A: undigested; B: EcoRI-digested; C: BamHI-digested; D: HindIII-digested. M: size markers- HindIII-digested λ DNA.

RIGHT: DNA was transferred to nylon membrane and probed with BPV4 DNA. The hybridised blot was autoradiographed for 1.5 hr. (lanes a-d).

2). Southern blot of papilloma DNA hybridised to BPV4 DNA and autoradiographed for 1.5 hr: Lanes 1: undigested; Lanes 2: BamHI-digested; Lanes 3: EcoRI-digested. DNA was extracted form papillomas of cases- (a): AOP 30/10; (b): AOP 3/11; (c): AOP 298tong; (d): AOP 298oes; (e): AOP 298rum.

Figure 13. Orientation of the BPV4 genome in pBV4.



RESTRICTION SITES

B - BamHI cloning site (nt2601).

E - EcoRI

H - HindIII

P - PstI

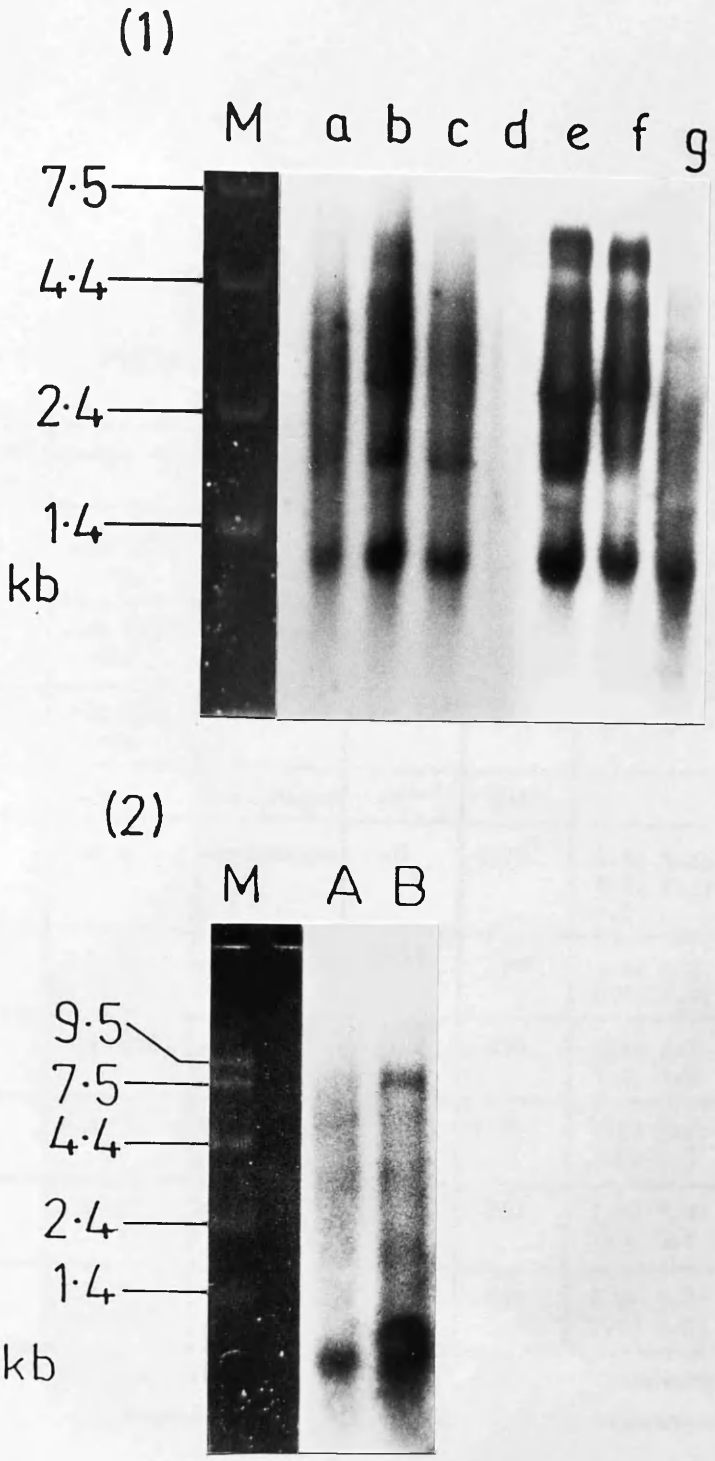


Figure 14. Northern blot hybridisation of papilloma RNA.

20 µg Total RNA from papillomas was run on a 2.2 M formaldehyde, 1% agarose gel, transferred to nylon membrane and probed with BPV4 DNA. The blot was exposed to X-Ray film for 48 hr.
Lanes M: RNA size markers (BRL).
(a): AOP 298tong; (b): AOP 298oes; (c): AOP 298rum; (d): AOP X; (e): AOP A; (f): AOP B; (g): AOP 299

2). A: AOP 30/10; B: AOP 3/11.

Table 3.1. Analysis of RNA from papillomas.

Fig.13, lane:	Sample No.	Site	Size of pap./mm	RNA/μg*	RNA size classes/kb	Rel. amt. RNA	Rel. amt. DNA
a	AOP 298 tong	tongue	all	880	6.4; 4.3; 2.8; 2.0; 2.1	+	++
b	AOP 298 oes	oesophagus	≥6	470	6.4; 5.4; 4.3; 3.6; 2.8; 2.0; 1.6; 1.2	++	+
c	AOP 298 rum	rumen	>6	665	4.3; 3.6; 2.0; 1.2	+	+
d	AOP X	oesophagus	>6	290	-	-	nd
e	AOP A	oesophagus	≥6	3075[1]	6.4; 5.4; 4.3; 3.6; 2.8; 2.3; 2.0; 1.6; 1.2	+++	+
f	AOP B	oesophagus	≥6	320[2]	6.4; 5.4; 4.3; 3.6; 2.8; 2.3; 1.6; 1.2	+++	+
g	AOP 299	oesophagus	all	480	5.4; 4.3; 3.6; 2.8; 1.6; 1.2	++	nd
A	AOP 30/10	oesophagus	≥6	1260	7.9; 5.4; 3.6; 3.1; 1.8; 0.8	+	+
B	AOP 3/11	oesophagus	≥6	830	7.9; 5.4; 3.6; 3.1; 1.8; 0.8	+	+
	AOP A poly(A)+	-	-	120	6.4; 5.4; 4.3; 3.6; 2.5; 2.0; 1.6; 1.1		

nd - not determined

* recovered from 1 g of tissue

[1] recovered from 3 g of tissue

[2] recovered from 0.5 g of tissue

3.2. Construction and screening of a cDNA library.

Mapping of BPV4-specific transcripts by Northern blot hybridisation and S1-nuclease protection analyses provided information on the organisation of transcriptional exons and the orf's contained in their sequences (Smith *et al.*, 1986). The transcriptional organisation of BPV4 was shown to be complex, with differential splicing producing polycistronic transcripts containing potential fusions between different orf's. A number of possible transcriptional initiation sites were also identified (Fig. 9). This work made it clear that the expression *in vitro* of subgenomic fragments containing separate orf's would in many cases not constitute a functional appraisal of BPV4 genes and that the only way to dissect the functions of *in vivo* gene products would be to synthesise cDNA's from viral transcripts from papilloma RNA and express these in an *in vitro* system. A cDNA library would also provide precise information on the positions of splice junctions and could be used in the fine mapping of transcriptional initiation sites.

3.2.1. Synthesis of cDNA.

cDNA was synthesised from oligo$(dT)_{17}$-primed poly$(A)^+$RNA isolated from total AOP A RNA. The method of Gubler and Hoffmann was chosen as it ensures better representation of 5' terminal sequences of mRNA's in the ds cDNA than do previous techniques relying on the action of S1-nuclease to

degrade "turn around" loops arising in first-strand synthesis (Fig. 15).

Aliquots of first and second strand synthesis reactions of the globin mRNA pilot experiment and of AOP A mRNA were electrophoresed on 1.4% alkaline agarose gels. Autoradiographs of the gels are shown in Figure 16.

In the globin mRNA pilot experiment, it can be seen that full-length products of approximately 600 nucleotides were obtained in both the first and second strand reactions and that these accounted for the vast majority of reaction products (Fig. 16(a)). Similar intensities are seen for bands in both lanes as equivalent fractions of the original mRNA input were loaded onto each. Incorporation of radioisotope into TCA-precipitable material was measured by scintillation counting and it was calculated from this that approximately 50% of the globin mRNA had been reverse transcribed into single stranded cDNA, and that virtually all of this had been converted into double stranded (ds) DNA in the second strand reaction. Since this represented optimum efficiency for this protocol (A. Hall, Chester Beatty Institute, London, personal communication), cDNA was synthesised from 10 µg of AOP A poly(A)$^+$RNA using the same reagents.

Calculations based on radiolabel incorporation data gave an estimate of only 20% conversion of mRNA into ds cDNA. The reduction in conversion efficiency occurred mainly in the first strand reaction, as second strand synthesis was approximately 90% efficient. The reasons for failure to reverse transcribe efficiently may have been:

1). the presence of biological and chemical contaminants in the mRNA sample inhibiting reverse transcriptase activity, such as RNA-binding proteins and traces of SDS remaining from the column chromatography step;

2). premature termination of first strand synthesis on a proportion of transcripts due to RNA secondary structure.

Alkaline agarose gel analyses of first and second strand reaction products are shown in Figure 16. According to these autoradiographs, the majority of first strand reaction products were between 150 and 2000 nucleotides in length, the greatest intensity of the radioactive signal being between 500 and 1200 nucleotides (Fig. 16(b)). The ds cDNA had much the same distribution, although an increase in higher molecular weight species was seen (Fig. 16(c)). Gubler and Hoffmann (1983) noted that a small percentage of globin cDNA molecules of twice the normal size appeared after second strand synthesis using their method; this was thought to be due to priming of DNA pol-1 by the hairpin-loop structure produced in first strand synthesis, leading to slightly reduced cloning efficiency as the hairpin is not degraded in this protocol.

Since the conversion of AOP A mRNA to ds cDNA appeared to cover a size range which would include at least some viral transcripts, and was satisfactory according to radiolabel incorporation data, the second strand reaction products were "blunt-ended" for cloning using T4-DNA polymerase.

3.2.2. Preparation of the vector.

Plasmid vector pUC13 was linearised with restriction endonuclease SmaI and treated with phosphatase to prevent religation. Test ligation reactions were set up using blunt ended HaeIII fragments of BPV4 DNA with appropriate controls as described in Section 2.2.8.(ii). The ligated DNA was used to transform newly prepared competent E. coli JM83 cells which were then spread onto 1.5% L-agar plates containing ampicillin and X-gal, and incubated overnight at 37 ^{o}C. Blue and white colonies on each plate were counted; results are presented in Table 3.2(a)

A minimal number of colonies grew on the cut vector control plates, showing that the vector had been efficiently prepared. According to the colony count on the uncut vector control plate , the transformation efficiency was calculated at 2.19×10^{6} transformants per µg of vector, an order of magnitude below the maximum efficiency reported by Dagert and Ehrlich (1979). 20% of the test ligation colonies were blue, but since the ligation control plate produced only 15 blue colonies, this was thought to be due to the expression of fusion proteins from orf's cloned in frame with ß-galactosidase in the vector. This preparation of competent JM83 cells was therefore considered suitable for transformation by pUC13-cloned cDNA.

3.2.3. Transformation of E. coli cells with cloned cDNA.

Blunt ended cDNA was ligated to the phosphatased pUC13 vector and used to transform the competent cells approximately 24 hours after their preparation. The entire cDNA preparation was resuspended in 100 µl of water and three ligation reactions were set up containing 0.5, 1.0 and 2.5 µl of this solution and 50 ng of vector; each of these ligations was used to transform 200 µl of competent cells. The transformations were spread onto selective medium plates as above in aliquots of one-fifth of the total volume, and incubated overnight.

Results are shown in Table 3.2(b). The transformation efficiency of the cells was increased, at 3.49×10^6 colonies per µg of pUC13. Optimum ligation efficiency was obtained when 1 µl of the cDNA solution was used in the ligation reaction , giving a predicted library size of 1.47×10^5 recombinants. The remainder of the cDNA was ligated to pUC13 and introduced into competent cells at an average transformation efficiency of 2×10^6 colonies per µg of vector. 14 cm L-Agar plates were dried for 30 mins. under a Category I hood and the whole of each transformation spread onto a single plate. Between 2500 and 3000 colonies grew on each plate, and a total of approximately 107000 colonies were obtained, representing 10700 recombinants per µg of input mRNA. This was considerably less than the cloning efficiency of 300000 clones per µg of mRNA reported by Gubler and Hoffmann; the lowered transformation efficiency of the

bacterial cells prepared in this project, and the reduced conversion efficiency of the cDNA synthesis account for this low yield. Also, blunt end cloning, while it was chosen to reduce the number of manipulations carried out from RNA isolation to recombinant transformation, is less efficient than other more elaborate methods such as polynucleotide tailing. Nevertheless, a cDNA library of over 10^5 clones is of a useful size and should be representative of rare transcripts such as those of BPV4.

In addition, glycerol stocks of JM83 cells transformed by pUC8-cloned cDNA produced from bovine oesophageal papilloma mRNA were plated out. These constituted the remaining part of a cDNA library synthesised by A. Sproul (Beatson Institute) to provide the cDNA's used in mapping the 3' ends of viral transcripts (Smith et al., 1986). The viability of these stocks was considerably reduced and a total of only 2670 colonies was obtained.

3.2.4. Screening of the cDNA library.

The two cDNA libraries were screened by the "colony-lift" method of Grunstein and Hogness (1975). Membrane discs were hybridised to nick-translated BPV4 genomic DNA, using unlabelled pUC DNA at a concentration of 20 $\mu g\ ml^{-1}$ in the prehybridisation to block non-specific hybridisation to vector sequences and thus reduce the number of false positives obtained. The use of calf thymus DNA in the pre-hybridisation solution also minimised non-specific

hybridisation to cDNA inserts containing bovine sequences. The washed membrane discs were autoradiographed: a typical autoradiograph is shown in Figure 17.

Two BPV4-positive colonies were found in the cDNA library supplied by A. Sproul. These were designated 10-1 and 7E11. Hybridisation of BPV4 was observed to 19 colonies in the cDNA library produced from AOP A mRNA: these were designated A to S. Autoradiographs were matched up to corresponding plates and the positive colonies picked and grown up in mini-prep cultures. All of these were white colonies, indicating that no orf's were cloned in frame with the ß-galactosidase gene. Part of each mini-prep was used to make glycerol stocks and plasmid DNA was isolated from the remainder.

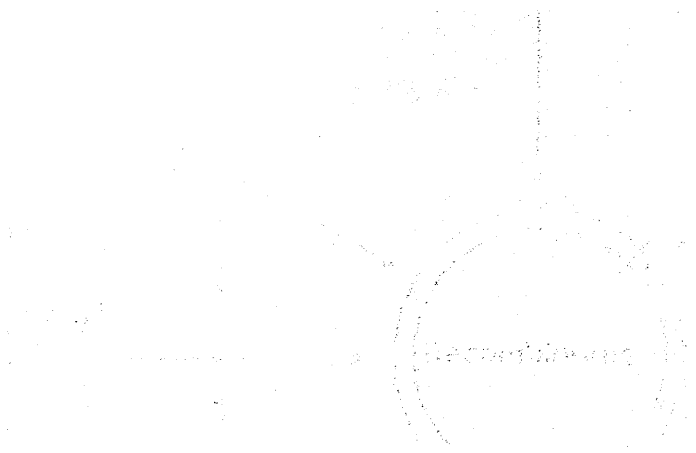
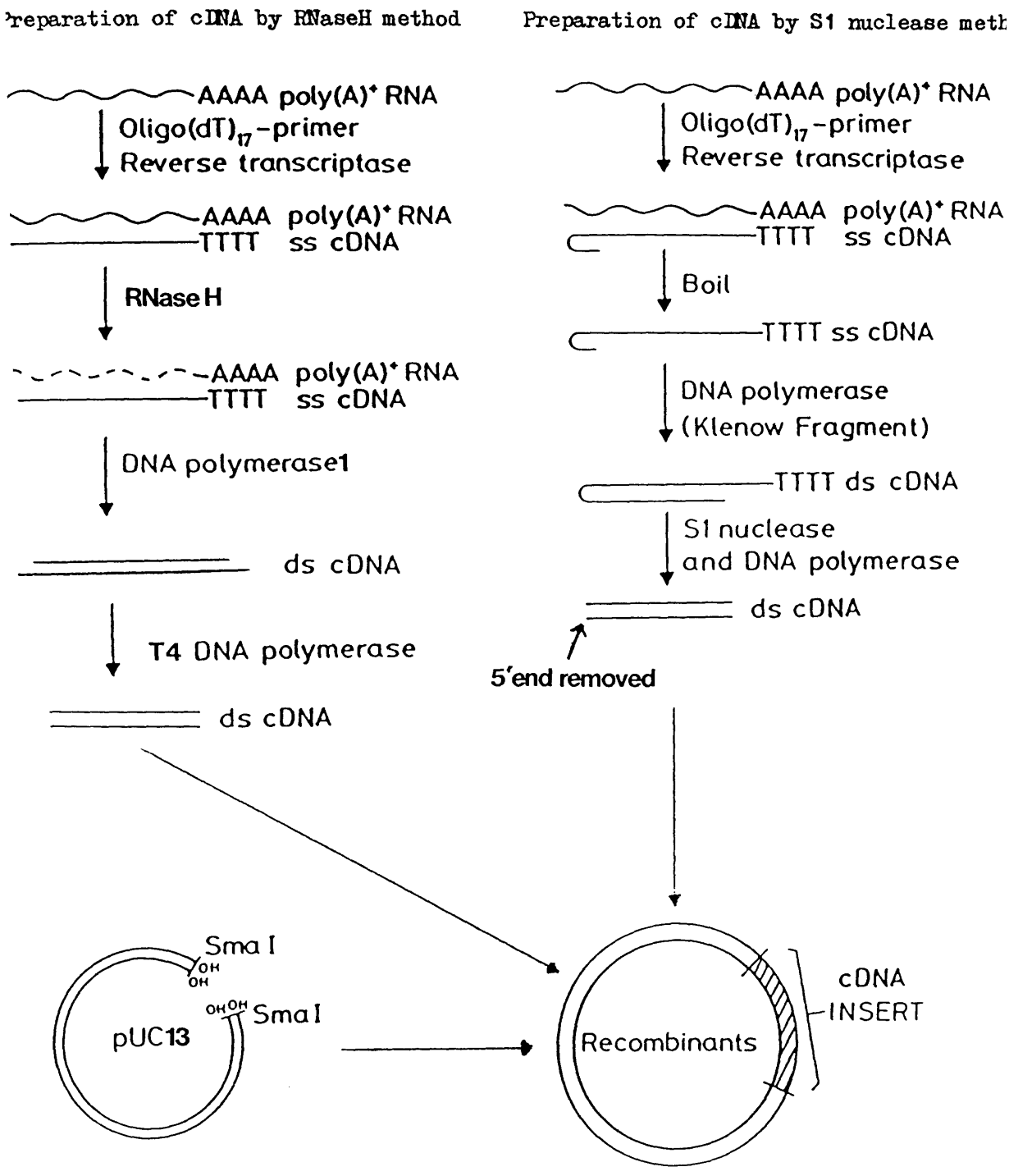


Figure 15. Alternative schemes for cDNA synthesis and cloning.



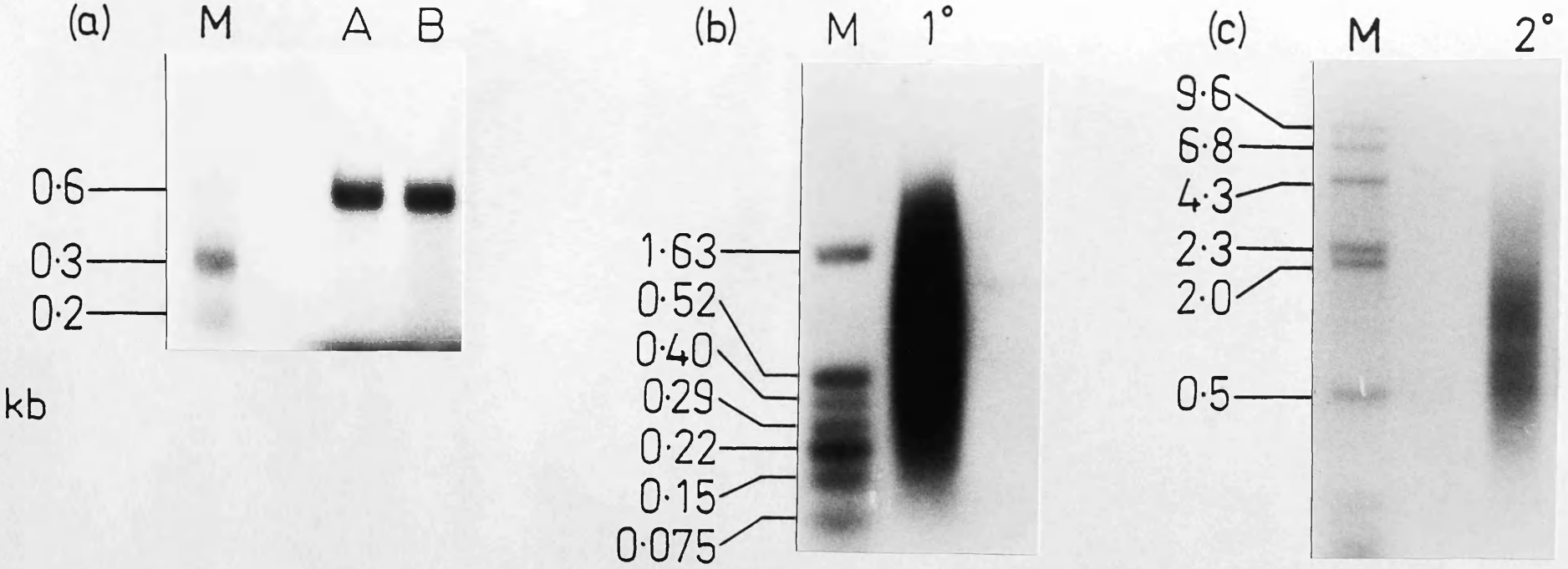


Figure 16. Monitoring of cDNA synthesis.

cDNA synthesised in the first and second strand reactions was electrophoresed on alkaline agarose gels, which were then dried and autoradiographed for 16 hrs.

(a) Globin mRNA pilot cDNA synthesis - M: size markers (Sau96I-digested pAT153 DNA); A: first strand synthesis; B: second strand synthesis.

(b) First strand synthesis of cDNA from AOP A poly(A)$^+$RNA - M: size markers (HinfI-digested pAT153 DNA); 1°: first strand cDNA

(c) Second strand synthesis of cDNA from poly(A)$^+$RNA - M: size markers (HindIII-digested DNA); 2°: second strand cDNA.

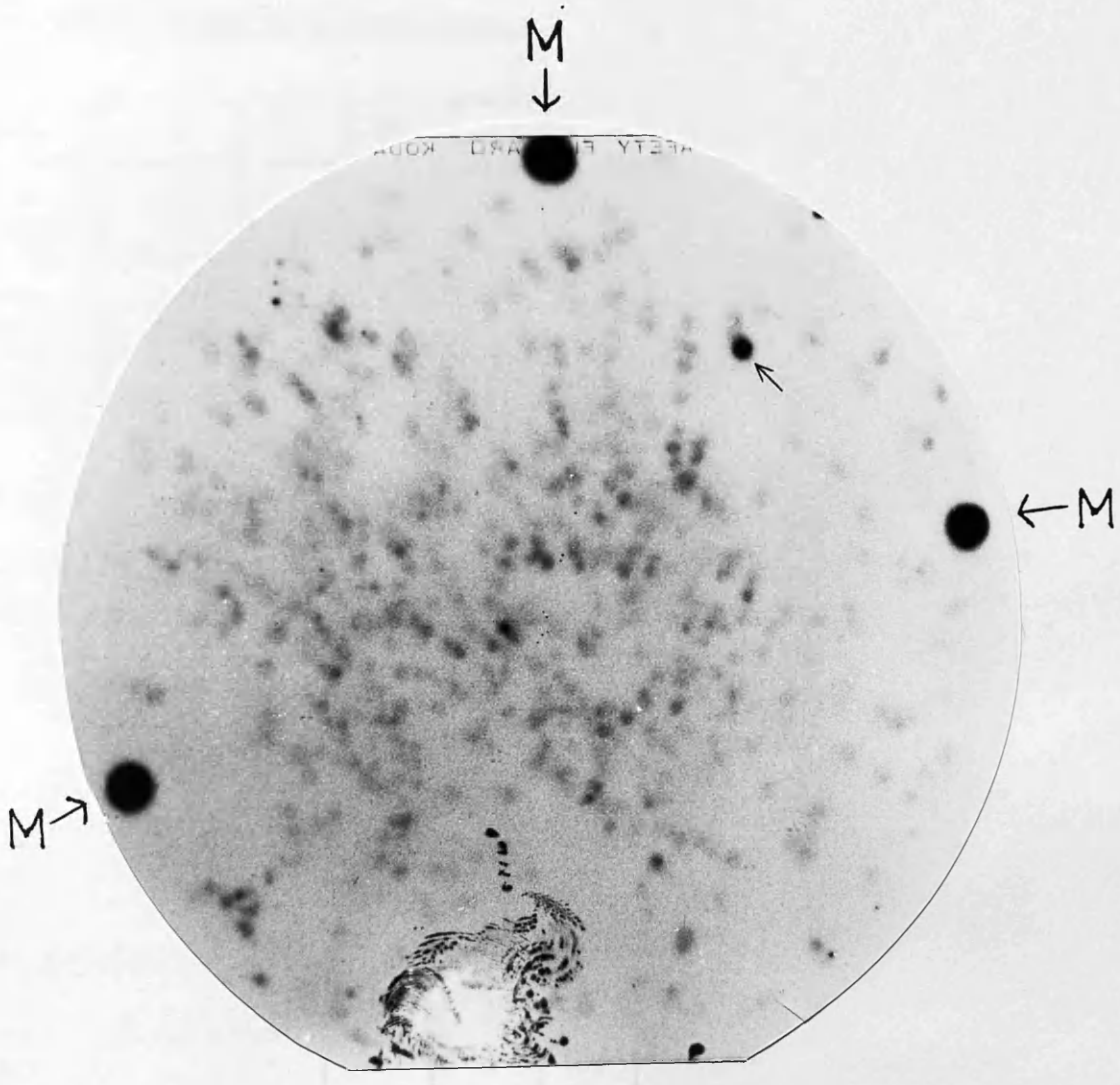


Figure 17. Grunstein-Hogness colony lift hybridised to BPV4 DNA.

Competent E. coli JM83 cells transformed with cloned cDNA's were spread on 14 cm selective medium plates, and the resulting colonies analysed by the Grunstein-Hogness colony lift technique. An autoradiograph of a typical filter disc probed with BPV4 DNA is shown. A positive colony is indicated by a small arrow.
M: radioactive alignment markers.

Table 3.2.(a) Pilot transformations.

Plate	T	LC	CC	UC
Blue colonies	110	15	7	217
White colonies	420	1	0	2
TOTAL	530	16	7	219

T: 5 ng linearised, phosphatased pUC13 plus 1 ng HaeIII-cut BPV4 DNA.

LC: Ligation control. 5 ng linearised, phosphatased vector plus ligase.

CC: linear vector control. 5 ng linearised, phosphatased vector.

UC: uncut vector control. 0.1 ng intact pUC13 DNA.

Table 3.2.(b) cDNA transformations.

Plate	1	2	3	UC
Blue colonies (ave)	37	43	78	345
White colonies (ave)	110	265	647	4
TOTAL*	132	293	710	349

1: 0.5 µl cDNA in ligation reaction.

2: 1.0 µl cDNA in ligation reaction .

3: 2.5 µl cDNA in ligation reaction.

UC: 0.1 ng intact pUC13.

* Totals corrected for ligation control (Table 3.4.(a)).

3.3. Analysis of cDNA's.

cDNA's were hybridised to BPV4 DNA in a Southern blot analysis to confirm that they contained BPV4 sequences. A selection were then mapped to regions of the viral genome by hybridisation to restriction fragments of BPV4 DNA. The cDNA's were then sequenced, and their sequences analysed using the Microgenie computer analysis program (Queen and Korn, 1984).

3.3.1. Southern blot hybridisation analysis.

3.3.1.(i). Hybridisation to total BPV4 DNA.

Plasmid DNA from each of the colonies A to S was digested with restriction endonucleases BamHI and EcoRI to separate cDNA inserts from vector sequences (see pUC multiple cloning site, Fig. 10). cDNA inserts were given the same designation as the colonies from which they were isolated (A to S; 7E11; 10-1); the corresponding plasmid recombinants were distinguished by the prefix "p" (pA to pS; p7E11; p10-1). The digests were run out on a 1% agarose gel, visualised by ethidium bromide staining and U.V. transillumination and photographed (Fig. 18(a)). The DNA was then transferred to nylon membrane and hybridised to the BPV4 DNA probe (Fig. 18(b)).

One plasmid, pS, contained no cDNA insert. On reexamination of the plate from which it had been picked, no other colonies

were found near enough to have caused any error in picking colony S. It was therefore considered either that hybridisation to the colony had been spurious and that the failure to produce ß-galactosidase was due to inappropriate religation of the vector causing a frameshift within the multiple cloning site, or that the cDNA insert was so small as to have run off the gel.

Strong hybridisation to all the inserts was seen except for the smaller fragment of cDNA P (Fig. 18(b), lane p). It was therefore thought that pP contained two cDNA's ligated end to end, one containing BPV4 sequences and the other, cellular sequences. In some lanes hybridisation was observed with bands of higher molecular weight than the inserts. The larger of these bands had the same mobility as linear pUC DNA, and probably represented hybridisation of contaminating plasmid DNA in the probe. The smaller band in most cases had no detectable counterpart in the gel photograph and probably represented small quantities of undigested, supercoiled plasmid.

3.3.1.(ii). Mapping of cDNA's on the viral genome.

Southern blot hybridisation experiments were carried out to localise cDNA's to subgenomic fragments of BPV4 DNA. Five cloned subgenomic restriction fragments of BPV4 were already available; these were: EcoRI - BamHI (nt1142 -nt2601); BamHI-HindIII (nt2602 - nt3705); HindIII - HindIII (nt3706-nt5954); HindIII - EcoRI (nt5955 - nt904) and EcoRI - EcoRI

(nt905 - nt1142). However, on the basis of known positions of transcriptional exons (Smith et al., 1986) and orf's of BPV4 (Patel et al., 1987), it was considered that these fragments did not reflect the transcriptional and genetic subdivisions of the viral genome. It was decided that the use of PstI fragments of BPV4 would enable more accurate localisation of cDNA's (Fig. 20).

Rather than clone the 9 PstI fragments into plasmids, it was reasoned that the longer cDNA inserts could be classified by hybridising their respective plasmids to a Southern blot of PstI digested BPV4 DNA. Each cDNA class could then be hybridised back onto the Southern blot containing all the cDNA inserts. This was carried out on the assumption that BPV4 cDNA's could only have one of two 3' termini with respect to the original mRNA, based on the transcriptional organisation.

cDNA's of $\geq$300 bp were chosen for classification: these were cDNA's G, J, L, M, N, and Q. cDNA's 7E11 and 10-1 were also included. To ensure efficient nick translation, the appropriate plasmid DNA's were used. 10 µg aliquots of AOP A papilloma DNA were digested with PstI, run on a 1% agarose gel and transferred onto nylon membrane. Papilloma DNA was used as it eliminated the problems caused by plasmid contamination in the previous experiment and provided a plentiful supply of immediately available BPV4 DNA.

The plasmid recombinant DNA's were radiolabelled by nick-translation and hybridised to the Southern blots. Some of the resulting autoradiographs are presented in Figure 19.,

and the results are summarised in the table in Figure 20. Autoradiographs of blots probed with pG, pJ, pL and pM were all identical to that obtained using pQ (Fig. 19, lane D). These cDNA's hybridised to the 1317 bp PstI fragment (nt3017-nt4334) which contains the early polyadenylation signal at nt4009 and the splice acceptor of the 3' terminal exon of the early region at nt3376. Since the 1.0 kb major transcript was mapped to this region (Smith et al., 1986), it was not surprising that the majority of cDNA's hybridised to this PstI fragment.

Plasmid p7E11 hybridised to the 1317 bp fragment, but also to the 230 bp band (Fig. 19, lane B). The 230 bp PstI fragment extends from nt917 - nt1147, far upstream of the 1317 bp fragment; it was therefore considered either that the p7E11 insert consisted of two cDNA's ligated end to end, or that it represented a previously unidentified cDNA.

Plasmid pN (Fig. 19, lane C) hybridised to the double band containing the PstI fragments from nt7126 - nt917 containing the E8 orf, most of the E6 orf and the late polyadenylation site at nt45 (Smith et al., 1986), and from nt4334 to nt5403, containing most of the L2 orf. Since cDNA N was 300 bp in length, it seemed unlikely that it derived from a late transcript as the 3' terminus of these is only 190 nucleotides downsteam of the PstI site at nt7126, and the cDNA did not hybridise to the adjacent 1723 bp PstI fragment. Since no other polyadenylation consensus signals could be found in the sequences of either of the 1 kb fragments, it was concluded that cDNA N was a truncated cDNA.

Plasmid p10-1 (Fig. 19, lane D) hybridised to the 450 bp band representing two adjacent PstI fragments from nt1147 -nt1578 and from nt1579 to nt2051. The region encompassed by these two fragments includes the 3' end of the E7 orf and the 5' one third of the E1 orf.

As the two fragments lie across two overlapping exons representing different transcriptional interpretations of the E1 orf (Fig. 20), it was important to map cDNA 10-1 more precisely. A pAT153-derived plasmid containing the EcoRI - BamHI subgenomic fragment (nt1142 - nt2601) was digested with these two enzymes in one reaction, and in a triple digest with XhoI in another. Restriction endonuclease XhoI cuts BPV4 at nt1624 (Fig. 21). These digests were run on a 1% agarose gel (Fig. 21, lanes A and B), transferred to nylon membrane and hybridised first to pUC DNA (Fig. 21, lanes 2(a) and (b)). The radiolabelled plasmid DNA hybridised only to the 3.4 kb band representing pAT153, showing that results would not be confused by other plasmid bands. Radiolabelled p10-1 was then hybridised to the same blot. Hybridisation was observed to the 1.45 kb EcoRI - BamHI fragment (Fig. 21, lane 1(a)), but only to the 0.97 kb XhoI - BamHI fragment (Fig. 21 lane 1(b)). These results led to the conclusion that cDNA 10-1 represented sequences between nt1624 and nt2051 at the 5' end of the E1 orf, and that it was another truncated cDNA, as no polyadenylation signals are found in this region of BPV4.

3.3.2. Sizing of cDNA inserts.

Plasmids p7E11, pN, pQ and p10-1 were digested with BamHI and EcoRI and run on 1.5% agarose gels next to size markers to obtain more accurate calculations of their insert sizes (Fig. 22). The results confirmed that the 231 bp cDNA N could not represent a late transcript as it did not hybridise to the 1732 bp PstI fragment of BPV4, and also indicated that the 775 bp cDNA 7E11 could represent sequences extending upstream of a putative splice junction at nt3376.

The existence of truncated species meant that cDNA's would not fall into two simple categories defined by the early and late polyadenylation sites, and it was evident that the most practical way of analysing the cDNA's was by sequencing.

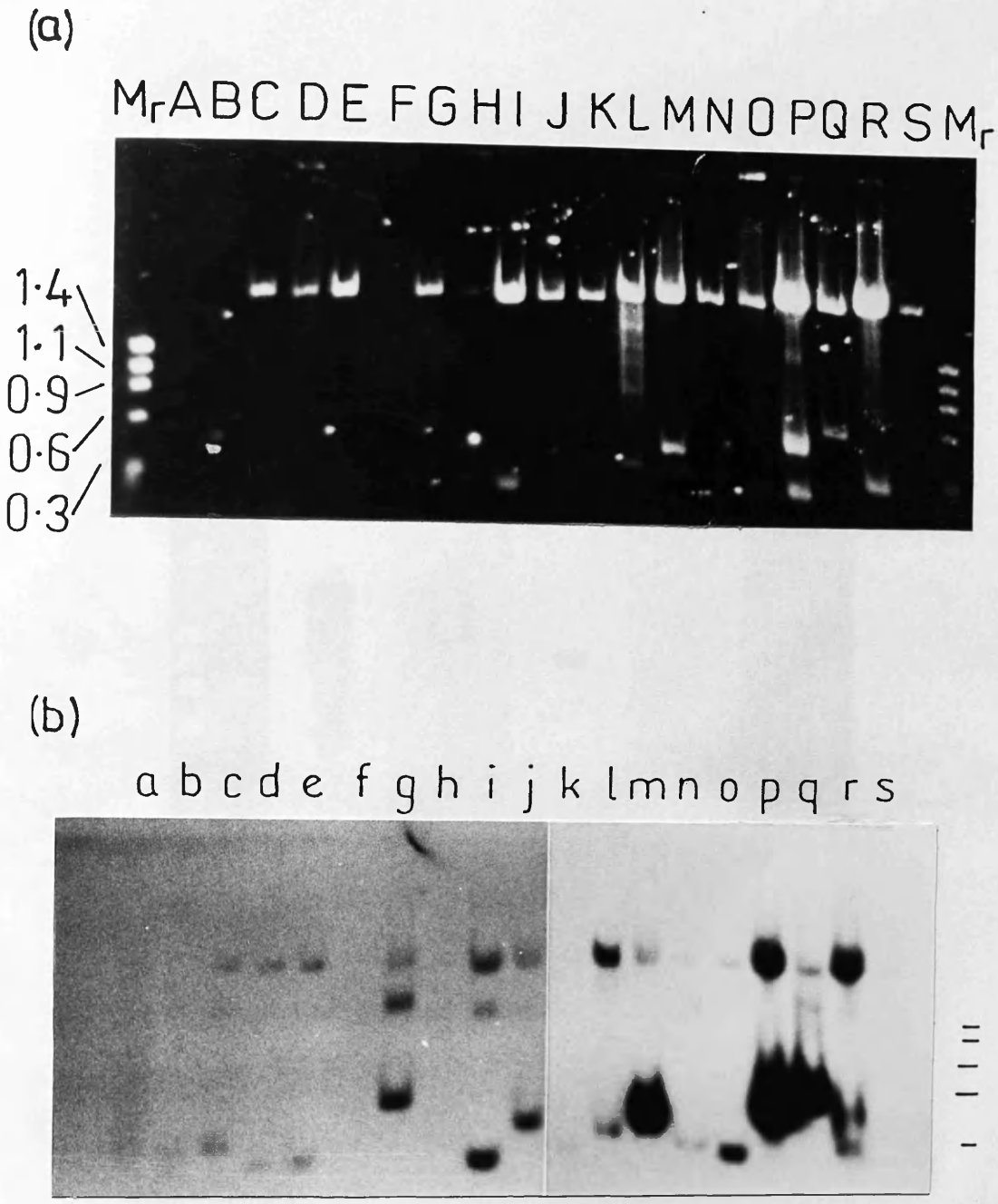


Figure 18. Southern blot hybridisation analysis of BPV4-positive cDNA inserts.

(a) Ethidium bromide stained 1% agarose gel containing BamHI + EcoRI digested plasmid DNA isolated from BPV4-positive recombinant bacterial clones. Lanes Mr: size markers (HaeIII-digested ØX174 DNA); Lanes A - S: pA to pS DNA

(b) Southern blot of the above gel, hybridised to a BPV4 DNA probe and autoradiographed for 16 hours. Lanes (a - s) correspond to lanes A - S.

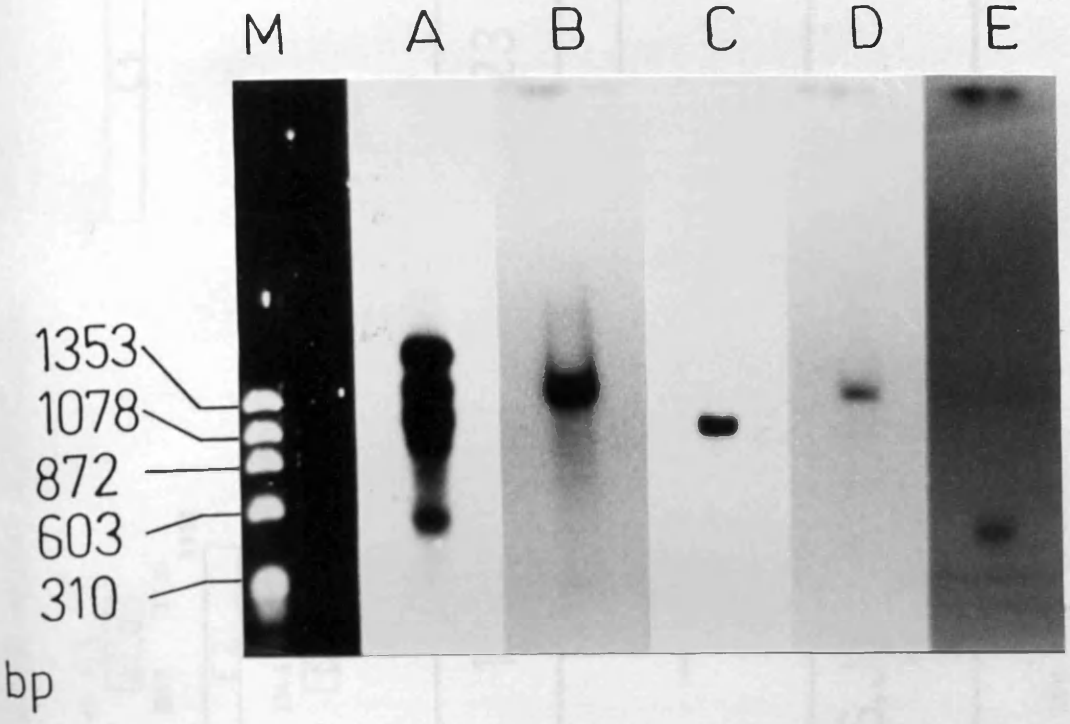


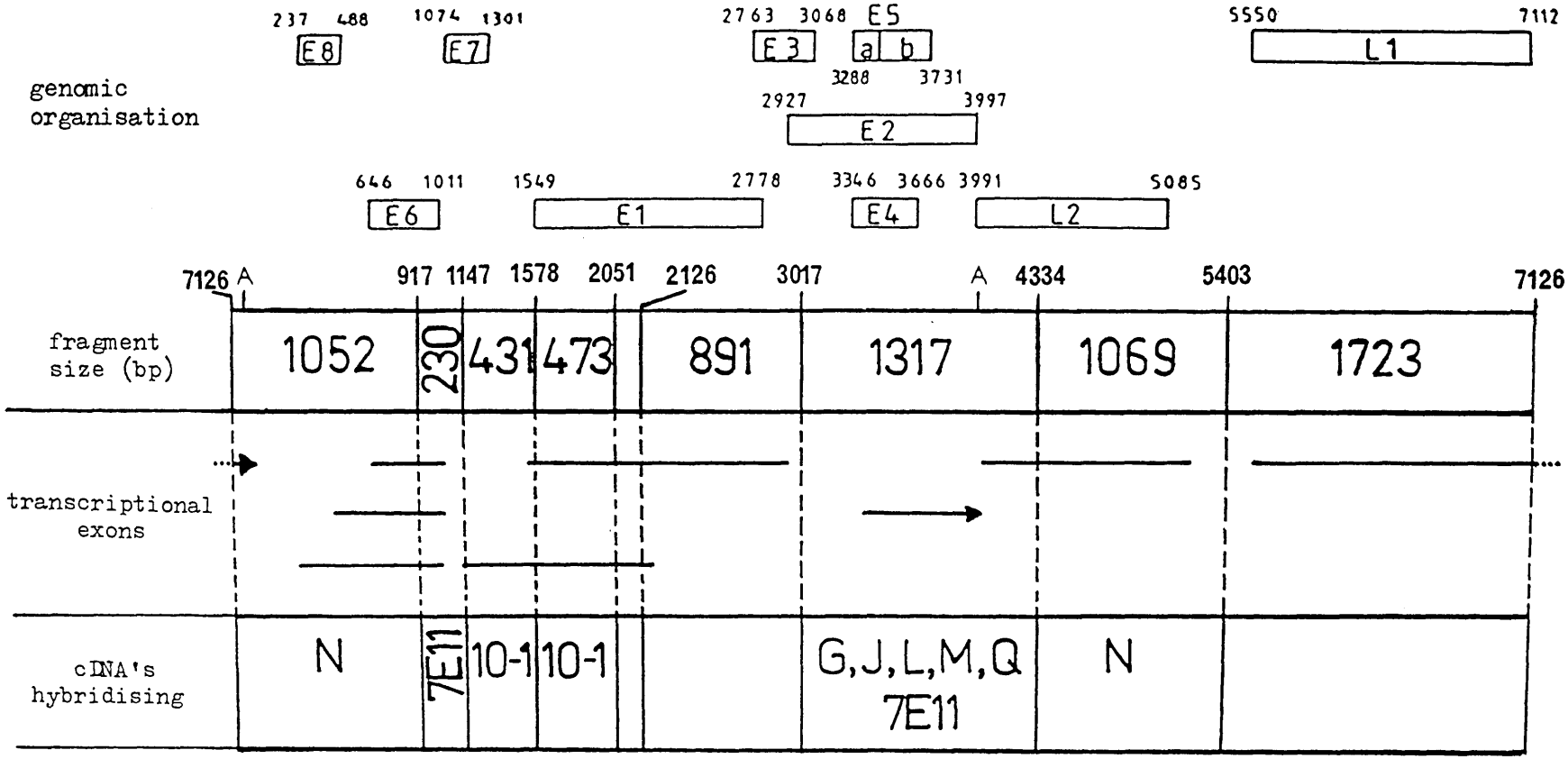
Figure 19. Mapping of cDNA inserts by Southern blot hybridisation.

10 µg aliquots of PstI-digested AOP A DNA were run on a 1% agarose gel and transferred onto nylon membrane. The Southern blots were hybridised to the following probes:

A: BPV4 DNA (control)
B: p7E11
C: pN
D: pQ
E: p10-1

Lane M: size markers in ethidium bromide stained gel (HaeIII-digested ØX174 DNA).

Figure 20. Hybridisation of cDNA's to PstI fragments of BPV4 DNA.



Numbers above the table refer to nucleotide positions in the BPV4 genome.

A - polyadenylation site ⟶ polyadenylated RNA 3' terminus

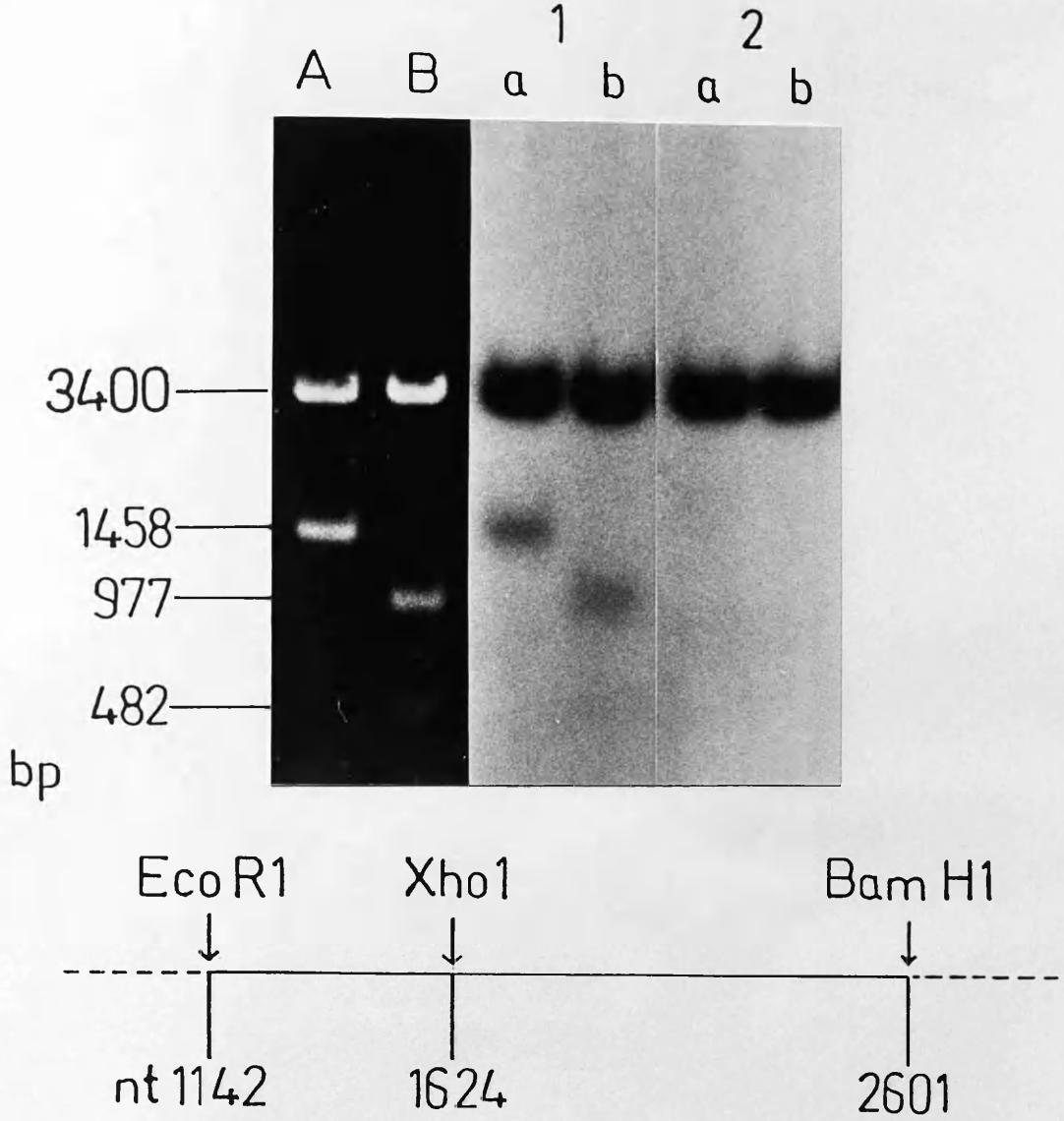


Figure 21. Mapping of cDNA 10-1 by Southern blot hybridisation.

1 µg aliquots of a pAT153-derived plasmid containing the EcoRI - BamHI subgenomic fragment were digested with:-

A: EcoRI and BamHI
B: EcoRI, BamHI and XhoI.

These digest were run on a 1% agarose gel, transferred to nylon membrane and hybridised to:-

1). p10-1; and 2). pUC13 DNA. Lanes (a) and (b) correspond to lanes A and B in the gel.

The subgenomic fragment of BPV4 used in this analysis and the nucleotide positions of the relevant restriction sites are shown below the photograph.

Figure 22. Sizing of cDNA's 10-1, 7E11, N and Q.

Ethidium bromide stained 1.5% agarose gels of recombinant plasmid DNA digested with EcoRI and BamHI:

(a) 1: p10-1; M: size markers (HinfI-digested pAT153 DNA)

STD LEN	DIST	PRED LEN	DEVIATION
1.63	50.500	1.63	-0.001
0.52	79.000	0.52	0.004
0.40	84.500	0.39	0.003
0.29	89.500	0.30	-0.010
0.22	94.000	0.22	-0.000
0.15	98.500	0.15	0.003

UNKNOWN FRAGMENTS:

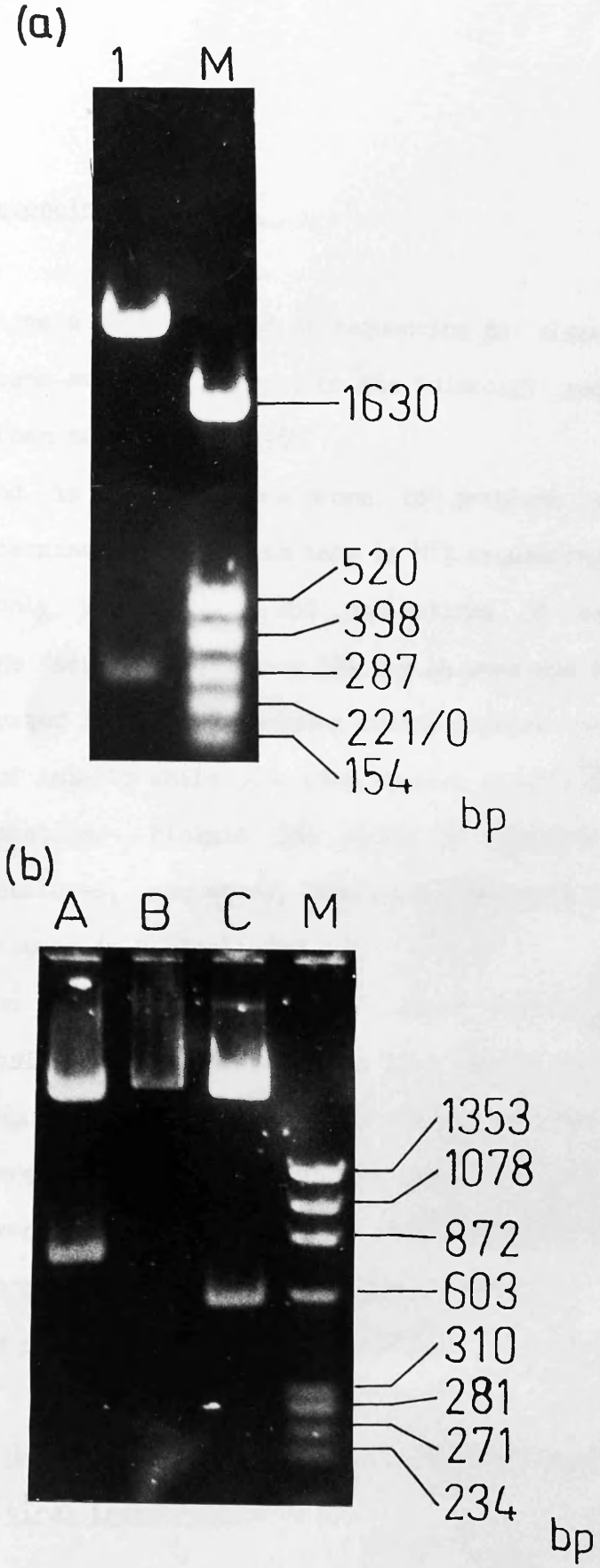
FRAGMENT	DISTANCE	PREDICTED LENGTH
1(i)	38.00	2.755
1(ii)	88.75	0.311

(b) A: p7E11; B: pN; C: pQ; M: size markers (HaeIII-digested ΦX174 DNA).

STD LEN	DIST	PRED LEN	DEVIATION
1.35	15.500	1.36	-0.011
1.09	17.500	1.10	-0.006
0.87	20.000	0.85	0.017
0.60	24.000	0.59	0.015
0.31	30.000	0.33	-0.022
0.27	32.000	0.27	-0.001
0.23	33.750	0.23	0.004
0.19	35.250	0.19	0.000
0.12	38.750	0.12	0.007

UNKNOWN FRAGMENTS:

FRAGMENT	DISTANCE	PREDICTED LENGTH
A	21.00	0.775
B	33.50	0.231
C	24.25	0.572


(a)
1
M
1630
520
398
287
221/0
154
bp
(b)
A
B
C
M
1353
1078
872
603
310
281
271
234
bp

3.3.3. Sequencing of cDNA's.

At this time a simple method of sequencing ds plasmid DNA inserts became available, based on the "dideoxy" sequencing protocol (Chen and Seeburg, 1985).

This method is somewhat more prone to problems such as premature termination artifacts than is M13 sequencing and in addition only yields 150 - 250 nucleotides of sequence. However, the fact that mini-prep DNA may be used and that ^{32}P is incorporated into the sequencing strand enables the rapid screening of inserts while providing precise details of their genomic location. Plasmid DNA could be isolated from bacterial cultures, sequenced, run on a sequencing gel and autoradiographed in a single day.

This system was ideally suited to the rapid analysis of the 21 BPV4-positive cDNA's. Those less than 400 bp in length could be sequenced end to end by this method, but the larger inserts were cloned into M13 vectors for full sequencing. Sequences were recorded and compared to BPV4 genomic sequence on the Microgenie computer program (Queen and Korn, 1984). The results of these analyses are summarised in Table 3.3.

3.3.3.(i). General; abundance of cDNA's and representation of viral transcripts.

The sequences of all of the cDNA's except cDNA P were >90% homologous to particular regions of the BPV4 genome. The sequence autoradiograph and homology comparison of cDNA I are

shown in Figure 23. The sequence of cDNA P contained no significant homologies to BPV4 or plasmid DNA, so the strong hybridisation of the viral genomic probe to this insert could not be explained. The different classes of cDNA obtained will be discussed separately.

In the AOP A cDNA library, therefore, a total of 17 BPV4 cDNA clones were isolated from 107000 clones, representing only 0.016%. Of these, 14 contained sequences exclusively from the 3' terminal exon of the early region, representing the 3' ends of the early mRNA's, confirming Southern hybridisation data for cDNA's G, J, L and M. cDNA N contained sequences from the leader exons of the 4.2, 2.8 and 3.0 kb transcripts, and cDNA Q represented an unspliced transcript, not previously mapped, from the E2/E4/E5 region. The 3' ends of the late transcripts were not represented, and this reflected the low abundance of these species in the Northern hybridisation analysis of the poly(A)$^+$RNA (Figure 43). These rarer transcripts may not have been represented in the cDNA library due to the low transformation efficiency of the bacterial cells used, and the decreased cloning efficiency resulting from blunt end ligation.

cDNA libraries of papillomavirus mRNA in general yield low numbers of cDNA's. A figure of 0.02% was obtained for ID13 cells (Yang et al., 1985a), although for fibropapilloma RNA, the yield of BPV1 cDNA's was 1.3%. These reflect the low level of viral transcription.

The BPV4 cDNA's were in general short (100 - 200 bp) and most were 3' truncated, having no poly(A) tail. This may have been

due to a combination of incomplete second strand synthesis and degradation by the exonucleolytic activities of T4-DNA polymerase. In addition, the colony-lift screening method, while indispensible for the screening of large numbers of recombinants, is subject to the limitation that not all positive clones may be detected.

While the figure of 0.016% is by no means an accurate one for the abundance of BPV4 transcripts in papilloma mRNA, it does reflect the low abundance of viral transcripts common to infections of cells with all papillomaviruses so far characterised.

3.3.3.(ii). Identification of the early polyadenylation site of BPV4.

The majority of cDNA's (18 out of 20) contained sequences from the 3' terminal exon of the early region (Table 3.3). None of these contained splice junctions except cDNA 7E11; this will be discussed in detail under subsequent headings. Nearly all had 3' ends (with respect to the original transcript) near the polyadenylation signal, ATTAAA, at nt4009. However, only three, cDNA's 7E11, G and J, had poly(A) tails. The sequence of the 3' end of cDNA 7E11 is shown in Figure 24, demonstrating the polyadenylation signal and the poly(A) attachment site. Sequence homology with BPV4 breaks down at nt4038 at the end of a run of four adenosine residues preceded by a G and followed by a cluster of G and T residues. This polyadenylation site configuration is compared

to the known polyadenylation sites of other papillomaviruses in Figure 25.

Eukaryotic RNA polyadenylation is usually directed by the motif AAUAAA at a CA dinucleotide 10 to 25 nucleotides downstream followed by a G/U cluster (Fitzgerald and Shenk, 1981).

This rule applies to the two early polyadenylation signals of CRPV (Giri et al., 1985; Danos et al., 1985), but in BPV1, the early poly(A) attachment site is a G residue directly following a CA dinucleotide (Yang et al., 1985a), and in HPV11, the early polyadenylation signal is AGUAAA and two polyadenylation sites are used, both being G residues within the G/U cluster, rather than tha CA dinucleotide upstream (Dartmann et al., 1986; Nasseri et al., 1987). The polyadenylation signal for the early transcripts of BPV4 is AUUAAA (Smith et al., 1986), known to substitute for the canonical sequence (Birnsteil et al., 1985) and the poly(A) addition site, as shown by cDNA's 7E11, G and J, is the G residue at nt4034, 25 nucleotides downstream of the first A of the polyadenylation signal at nt4009.

This polyadenylation site is unusual in that the G residue at nt4034 is chosen rather than the CA dinucleotide 5' adjacent to it. cDNA's H and I end at nt4029 and cDNA B at nt4028 (Table 3.3); this may indicate that a second site at nt 4028, the CA dinucleotide, may be polyadenylated, but no poly(A) tails have been found attached at this site.

S1-nuclease protection mapping placed the 3' end of BPV4 early region transcripts at nt4030, but the sequence of

cDNA's produced in this project allows the precise assignment of the early polyadenylation site to nt4034. The spatial arrangement of this site 25 nucleotides downstream of the polyadenylation signal is consistent with canonical eukaryotic RNA polyadenylation signalling, as is the position of the G/T cluster on the genome.

With this study, the early polyadenylation site of four papillomaviruses have been identified. These appear to conform to the eukaryotic pattern of mRNA precursor polyadenylation and cleavage at sites directed by signals contained in the sequence. However, it is also apparent that utilisation of the late viral polyadenylation site occurs only in differentiated host epithelial cells. In BPV1, this appears to be linked to the activation of a specific promoter, P_L (Baker and Howley, 1987). It is not known whether this is due to transcriptional termination upstream of the late polyadenylation signal in the early transcripts, or to an attenuated mechanism determining the site of polyadenylation and cleavage of RNA continuously transcribed all the way round the circular genome.

Nevins and Darnell (1978) in their analysis of the nuclear processing of adenovirus type 2 transcripts, showed that transcription extended well beyond the polyadenylation sites and that concerted polyadenylation and cleavage at a single site determined the 3' end of each mature mRNA. It was also shown that polyadenylation preceded splicing. This has been taken to indicate that polyadenylation site selection may influence the subsequent splicing pattern of mRNA precursors.

In the case of differential expression of immunoglobulin type M μ-heavy chain genes in maturing lymphocytes, the same transcript is polyadenylated at different sites, leading to differential splicing which gives rise to two different mRNA's expressing membrane-bound and secreted forms of the IgM heavy chain (Early et al., 1980; Kemp et al., 1983). Tissue-specific expression of calcitonin and calcitonin gene-related peptide also occurs from the same transcriptional unit by differential polyadenylation and splicing (Amara et al., 1984). In both of these instances, differential expression is brought about by splicing to exons downstream of the first polyadenylation site, which is spliced out. This appears to be the case in the late region transcripts of BPV1 (Fig. 2; Baker and Howley, 1987), HPV1 (Fig. 4; Chow et al., 1987b) and BPV4 (Fig. 9; Smith et al., 1986), so it may be that polyadenylation occurring at the late polyadenylation site directs the splicing out of the early site, thus preventing cleavage at this site. Only about 20% of P_L-directed transcripts extend beyond the early polyadenylation site; this may be related to a mechanism influencing choice of polyadenylation site on nascent transcripts. The leader sequence of the late transcripts may be involved in differential polyadenylation since transcripts from all other BPV1 promoters terminate at the early polyadenylation site.

3.3.3.(iii). Sequence analysis of cDNA 10-1.

Parts of the plasmid sequencing autoradiographs of cDNA 10-1 are shown in Figure 26, and the sequence analysis in Figure 27.

cDNA 10-1 is 288 bp in length and, according to plasmid sequencing, 100% homologous to BPV4 from nt1158 to nt1845, placing it at the 5' end of the E1 orf (Table 3.3). The 5' end of cDNA 10-1 is 7 nucleotides downstream of the acceptor site at nt1553 which is utilised in the 4.2 and 3.0 kb transcripts. It also coincides with the 5' end of the 1.9 kb transcript, indicated by minor 5' termini a few nucleotides downstream of the 5' terminus of the exon in the S1-nuclease protection assays of Smith et al.(1986). cDNA 10-1 may therefore represent the 5' end of a full length cDNA which is 3' truncated, since there is no polyadenylation consensus sequence in this region.

A possible TATA homology, (G)TATTAG (Breathnach and Chambon, 1981) occurs in an A/T-rich region at nt 1529, 29 nucleotides upstream of the 5' end of cDNA 10-1, although no other known promoter/enhancer sequences can be found in this region.

A splice donor site consensus sequence (Nevins, 1983) occurs at nt1779 which is not utilised in this cDNA, nor in any BPV4 transcript mapped so far.

Putative translation products of cDNA 10-1 in all three reading frames of the coding strand are shown beneath the sequence in Figure 27. The cDNA encodes the first 90 amino acids of the E1 orf, from the first methionine. The presence

of an A residue three nucleotides upstream of this ATG codon makes this a strong translational initiator (Kozak, 1986). Assuming that the E1 orf is functionally divided in the same way as that of BPV1 (Berg et al., 1986), cDNA 10-1 may therefore encode part of the replication modulator.

The E1 orf is encoded in the 4.2, 3.0 and 1.9 kb transcripts of BPV4; the splice junction of the 1.6 kb transcript removes the 3' end of E1. This suggests that the functional division of this orf is effected by the presence or absence of the 3' end, in the same way as the E1a gene of adenovirus (Lillie et al., 1987). If the 3' truncated E1 orf gene product does indeed modulate replication, then the paucity of the 1.6 kb transcript in papilloma RNA extracts may indicate that it is not produced in the more differentiated cell layers where viral DNA replication occurs in an apparently uncontrolled fashion. Testing of this function could be carried out using cDNA 10-1 in a suitable expression vector, but appropriate replication mutants of BPV4 would first have to be constructed. The fact that pBV4 (Fig. 13) exists in stably transformed NIH 3T3 cells at a copy number of only 3 to 30 episomes per cell (Campo and Spandidos, 1983) suggests that the replication factor has been disabled by the effective deletion of the E1 orf downstream of the BamHI cloning site at nt2601, and that the 5' end of E1 encodes a replication modulator controlling copy number in these cells. Present theories regarding papillomavirus replication relate only to cultured fibroblasts, however, and meaningful functional analyses await the development of an in vitro differentiating

epithelial system supporting vegetative propagation of papillomaviruses.

3.3.3.(iv) Sequence analysis of cDNA N.

Parts of the plasmid sequencing autoradiographs of cDNA N are shown in Figure 28; analysis of the full sequence is presented in Figure 29.

cDNA N is 205 bp in length and has perfect homology with BPV4 DNA from nt633 to nt837 (Table 3.3). It is a 3' truncated cDNA, since no polyadenylation signals exist in this region of the genome. The 5' end of cDNA N does not map close to any previously defined mRNA 5' termini, but a TATA consensus signal, (G)TAAATAG (Breathnach and Chambon, 1981), occurs 51 nucleotides upstream at nt582, so it could be part of a near full-length cDNA.

The putative translation products of cDNA N are shown beneath the sequence in Figure 29. It encodes the first 41 codons of the E6 orf, from the first methionine codon at nt715. The 5' flanking sequences of the ATG codon, from nt703 to nt715, constitute an almost perfect palindrome, capable of forming a hairpin-loop structure which could conceivably interfere with translation of the mRNA. Experiments on mutants of the preproinsulin gene showed that the translational efficiency of eukaryotic ribosomes was not affected by the presence of a stable hairpin structure constructed around the ATG codon (Kozak, 1986), and it appears that translational initiation is modulated only by nucleotides at certain key positions

around the initiator codon (Kozak, 1986), with the sequence ACCATGG being the most efficient initiator. The A residue three nucleotides upstream of the E6 ATG codon makes this a strong initiator.

The E6 orf is contained in the leader exons of tha 4.2, 3.0 and 2.8 kb transcripts, and it was thought at first that the expression of the downstream orf's would be strongly inhibited by this. However, it has been shown that the eukaryotic ribosome is capable of reinitiating translation in downstream orf's of polycistronic transcripts. The efficiency of each initiation event is affected only by the sequence context of the respective ATG codons (Kozak, 1986). The polycistronic transcriptional organisation is a feature of the papillomaviruses and may be a result of economical transcriptional control packaging in a discrete region of the viral genome.

The E6 region of BPV4 DNA has few restriction sites, making the cloning of the E6 orf difficult. cDNA N, although truncated at the 3' end, could be useful for the synthesis of E6 peptides for antiserum production.

3.3.3.(v). Sequence analysis of cDNA 7E11.

(a). Plasmid sequencing.

cDNA 7E11 was too long to be sequenced by the plasmid sequencing method. Nevertheless, initial plasmid sequencing analyses provided a good deal of information on the cDNA insert. Sequence autoradiographs showing part of the 5' end

of cDNA 7E11 are shown in Figure 30(2) and (3); clearer sequencing of the 3' end was obtained by M13 sequencing, which is shown in Figure 24.

Sequencing of the 3' end defined the early polyadenylation site of BPV4 (3.3.3.(ii)). Sequencing of the 5' end confirmed the Southern hybridisation data (Fig. 20): a stretch of DNA with perfect homology to nt943 to nt 1016 at the 3' end of the E6 region was joined to sequences from the E2/E4/E5 region, starting at nt 3376. The true length of the insert was therefore 754 bp, including the 21-nucleotide poly(A) tail (Table 3.3).

Since a transcriptional 5' terminus had been mapped around nt 3375 (Smith et al., 1986), and since many of the cDNA's sequenced were truncated at both ends, it was considered possible that the insert of p7E11 was a cloning artifact consisting of two cDNA's ligated end to end. However, three lines of evidence argued in favour of this being a single cDNA:

1). a similar E6/E4 splice junction was seen in BPV1 cDNA's produced from ID13 cell mRNA (Fig. 2;Yang et al., 1985a), and splice junctions from the E6/E7 region to the E2/E4 region had been mapped in CRPV transcripts (Nasseri and Wettstein, 1984);

2). a splice donor site was known to exist at nt1016, and an acceptor site at nt3376; the sequence of cDNA 7E11 across the junction conformed to that of a canonical splice junction (Fig. 31; Nevins, 1983);

3). electron microscopy (e.m.) of R-loops formed by

poly(A)$^+$RNA from bovine oesophageal papillomas annealed to pBV4 DNA (Fig. 13), presented several examples of mRNA species which hybridised to two widely separated regions of the BPV4 genome. The larger of the two loops mapped between nt3700 and nt4100, and the smaller loop around nt1000. Because of the position of the cloning site in the early region of the viral DNA, the position of this exon was actually downstream of the larger R-loop, and this complicated interpretation of the e.m. data (by kind permission of L. Coggins, Beatson Institute).

Although evidence was circumstantial in the first two instances, it was strongly supported by the e.m. data and it was therefore considered that cDNA 7E11 represented a single transcript. This was confirmed beyond doubt by subsequent experiments and supported by studies of papillomavirus transcriptional organisations published later.

The positions on the viral genome of the exons represented by cDNA 7E11 are shown in Figure 33.

No promoter/enhancer sequences exist in the vicinity of the 5' end of cDNA 7E11 on the viral genome and it was therefore considered that the cDNA was not full-length. Preliminary analysis of the plasmid squencing data showed that putative translation products 3' to the splice junction included the 3' ends of the E2, E4 and E5a orf's. Sequences upstream of the splice junction contained the 22 carboxy-terminal codons of the E6 orf, terminating just before the junction, and the junction created two fusion orf's. One of these contained no ATG codons, but the other was a fusion of four codons

including an ATG at nt1004 to the 3' end of the E5b orf. The translation product of this orf would have been only 23 amino acids, according to the sequence of BPV4. The first ATG of the E2 reading frame of cDNA 7E11 would have been at nt3443, downstream of the sequenced region.

The coding potential of this cDNA therefore consisted of a 23 amino acid protein encoded by a fusion orf, and the 3' half of the E2 orf. Since the cDNA apparently did not represent the full length of a transcript, it was thought that the E6 orf might also be encoded by this mRNA as the nearest known transcriptional start site was at nt654.

The cDNA was sequenced along its entire length by M13 sequencing to confirm that it consisted solely of the two exons of BPV4.

(b). M13 sequencing.

cDNA 7E11 was isolated from pUC8 vector sequences using restriction endonucleases EcoRI and SalI. The insert was separated from the vector on a 1% LMP agarose gel, purified and cloned into bacteriophage M13mp18 and mp19, which contain these cloning sites in opposite orientations relative to the universal primer sequence (Norrander et al., 1983). Due to the length of the cDNA insert, it was subcloned in two fragments using the HindIII site corresponding to nt3705 in the BPV4 genome; this enabled overlapping sequences of both strands to be obtained over the entire length of the cDNA (see Fig. 31(a) for sequencing strategy). Parts of the sequencing autoradiographs of cDNA 7E11 are shown in Figure 30; analysis of the full sequence is shown in Figure 32.

The full sequence of cDNA 7E11 was compared to the genomic sequence of BPV4 (Patel et al., 1987). The splice junction was confirmed by the sequence of the antisense strand (Fig. 31(b)), and the sequence of cDNA 7E11 was 100% homologous to that of BPV4 from nt3460 to the polyadenylation site at nt4034; however, eight mismatches were found between nt3412 and nt3460 (Fig. 31(b)). These were confirmed by repeated sequencing in both directions, showing that the mismatched sequence was not an artifact caused by secondary structure in the sequencing reaction products. They are indicated by boxes next to the sequence autoradiographs (Fig. 30) and by asterisks in the computer analysis (Fig. 31(b)). They consist of :

1). a point deletion at nt3413;

2). inversion of the CA dinucleotide at nt3430;

3). inversion of the TA dinucleotide at nt3443;

4). inversion of the AC dinucleotide at nt3446 and

5). a point insertion after nt3459, of a G residue.

It was at first uncertain whether these mismatches represented reading errors made by the reverse transcriptase or DNA pol-1 in the first or second strand synthesis of the cDNA, or whether they represented faithful transcripts of the viral DNA resident in the papilloma cells. The implications of such alterations to the published sequence and genome organisation and to the coding potential of the transcript represented by cDNA 7E11 are far-reaching and will therefore be discussed under a subsequent heading (3.5).

3.3.3.(vi). Sequence analysis of cDNA Q.

(a). Plasmid sequencing.

cDNA Q was again too long to be sequenced along its entire length by plasmid sequencing, which located the 5' and 3' ends at nt3235 and nt3788, respectively (Figs. 33; 34; Table 3.3). Although it was truncated at the 3' end, this 553 bp cDNA was of immediate interest as it contained contiguous sequences crossing the splice acceptor site at nt3376 and extending 150 nucleotides upstream. The 5' end was also positioned 50 nucleotides downstream of a consensus TATA sequence, (G)TATTAT (Breathnach and Chambon, 1981), at nt3185, and a CAAT sequence (McKnight and Tjian, 1986), was found at nt3156. However, the CAAT - TATA promoter system is not an absolute requirement for transcriptional initiation (Nevins, 1983), as evidenced by the late promoter of SV40 (Brady et al., 1982) and by recent work on BPV1 transcription (Ahola et al., 1987; Baker and Howley, 1987), so the possibility that the transcript represented by cDNA Q initiated much further upstream was not ruled out. The coding potential of the cDNA sequences (Fig. 36) was restricted to the E5a orf.

(b). M13 sequencing

cDNA Q was sequenced along its entire length in the M13mp18 and mp19 vectors. According to the genomic sequence, a KpnI site was conveniently positioned, at nt3411, for subcloning cDNA Q in two fragments. The cDNA was, however, resistant to digestion by this enzyme and was therefore cloned intact into

the M13 vectors using the EcoRI and SalI cloning sites.

Parts of a sequencing autoradiograph showing the sense sequence is shown in Figure 35(1). It emerged from the sequence data thus obtained that cDNA Q contained the same mismatched sequences as cDNA 7E11 when compared with the genomic sequence.

(c) M13 sequencing using a synthetic oligonucleotide primer.

Since the region of mismatched sequences was over 300 nucleotides into the antisense sequence, difficulty was experienced in obtaining clear sequencing autoradiographs. It was therefore decided to use an oligonucleotide to prime the sequencing reaction from a point closer to the mismatched region. The primer sequence was chosen from nt3551 to nt3535 of the non-coding strand of BPV4 DNA. This was designated primer 140 (Table 3.6). The sequence was chosen as it had minimal complementarity to all other parts of the genome and was rich in G+C residues (65%) which form stronger interstrand bonds than A and T, but did not contain palindromic sequences which would inhibit annealing to the template due to the formation of secondary structure. These factors ensured specific annealing of the primer to its complementary sequence in the single stranded template.

Primer 140 was used in place of the M13 universal primer in the sequencing reaction. The resulting antisense sequencing autoradiograph is shown in Figure 35(2). The mismatched sequences were confirmed. The point deletion at nt3413 destroys the KpnI site GGTACC by removing the T residue, explaining the failure of this enzyme to cut cDNA Q. Primer

140 was also used to confirm the sequence of the corresponding region in cDNA 7E11.

The isolation of two cDNA's from separate libraries which presented the same variations on the published genomic sequence strongly suggested that the variations were at the genomic level. As was previously mentioned (3.3.3.(v).), these apparent changes caused major alterations to the genomic organisation of BPV4, so a further experiment was performed to confirm the altered sequence.

3.4. RNA sequence analysis using primer 140.

In order to demonstrate that the sequences of cDNA's 7E11 and Q were faithful copies of BPV4 mRNA sequences, sequencing of poly(A)$^+$RNA from AOP A papillomas was performed according to the protocol of Geliebter (1987). This method is identical in principle to Sanger "dideoxy" sequencing (Sanger *et al.*, 1977): purified mRNA template is annealed to a radioactively labelled DNA primer with exact specificity for the point from which sequence is required. First strand cDNA is then synthesised in the presence of each of the four dideoxynucleotide triphosphate (ddNTP) analogues in separate reactions; the reaction temperature of 50^{o}C minimises RNA secondary structure. When run out on a sequencing gel, these give a radioactive sequencing ladder complementary to the original RNA template.

The sequence of primer 140 and its annealing temperature for RNA sequencing (Geliebter, 1987; section 2.2.11.(iii)) are

shown in Table 3.6.

Synthetic oligonucleotides do not possess a 5' phosphate group, so primer 140 was 5' ^{32}P labelled to high specific activity using polynucleotide kinase. It was then annealed to AOP A poly(A)$^{+}$RNA, the sequencing reaction performed, and reaction products run out on a sequencing gel. It was found that the recommended ddNTP concentrations only gave readable sequence information up to about 100 nucleotides upstream of the primer. These concentrations were halved in subsequent experiments, and the autoradiograph of such a sequence is shown in Figure 37.

The sequence clearly shows the alterations to the BPV4 DNA sequence found in cDNA's Q and 7E11. The implications of these alterations to the viral genome organisation and the expression of its encoded genes is discussed in the next subsection.

The RNA sequence was readable up to the splice acceptor site at nt3376, where bands in the C track become faint and finally disappear. This phenomenon was not overcome by reducing the ddCTP concentration in the C reaction and therefore represents the limit of the technique. The sequence was therefore impossible to read accurately beyond this point, but by interpolating C residues in the appropriate places, it was found that the sequence of the first 12 nucleotides 5' to the splice acceptor site matched that in cDNA 7E11 more closely than the corresponding contiguous sequence in the genome. This sequence did not match the upstream sequence of any other splice donor site in the BPV4

genome. This confirmed the splice junction of cDNA 7E11 and moreover suggested that it represents a transcript which is many times more abundant than any other transcript containing the early region 3' terminal exon.

Table 3.3 Sequence analysis of BPV4 cDNA's.

cDNA	Length/bp	Position on genome*	ORF's
A	275	3709 - 3983	3' end E2
B	201	3828 - 3983	"
C	267	3729 - 3995	"
D	122	3851 - 3972	"
E	113	3761 - 3873	"
F	76	3916 - 3991	"
G	323	3745 - 4034 + $(A)_{33}$	"
H	109	3921 - 4029	"
I	103	3927 - 4029	"
J	285	3755 - 4034 + $(A)_5$	"
K	152	3843 - 3994	"
L	243	3781 - 4023	"
M	487	3523 - 4009	3' end E5a, E4, E2
N	205	633 - 837	5' end E6
O	188	3830 - 4017	3' end E2
P	-375	-	-
Q	554	3235 - 3788	E4, E5a, E5b, 3' end E2
R	190	3805 - 3994	3' end E2
7E11	754	943 - 1016; 3376 - 4034 + $(A)_{21}$	3' end E6, E2, E4, E5a; whole E5b; E5a fusion
10-1	288	1558 - 1845	5' end E1

* numbers indicate nucleotide positions on the BPV4 genome.

Figure 23. Sequence of cDNA I

cDNA I was sequenced by plasmid sequencing from either end of the multiple cloning site of pUC13.

(a) Sequencing autoradiographs of pI. Parts of the sequence are shown next to the photograph. Vector sequences are boxed. Numbers refer to nucleotide positions in BPV4.

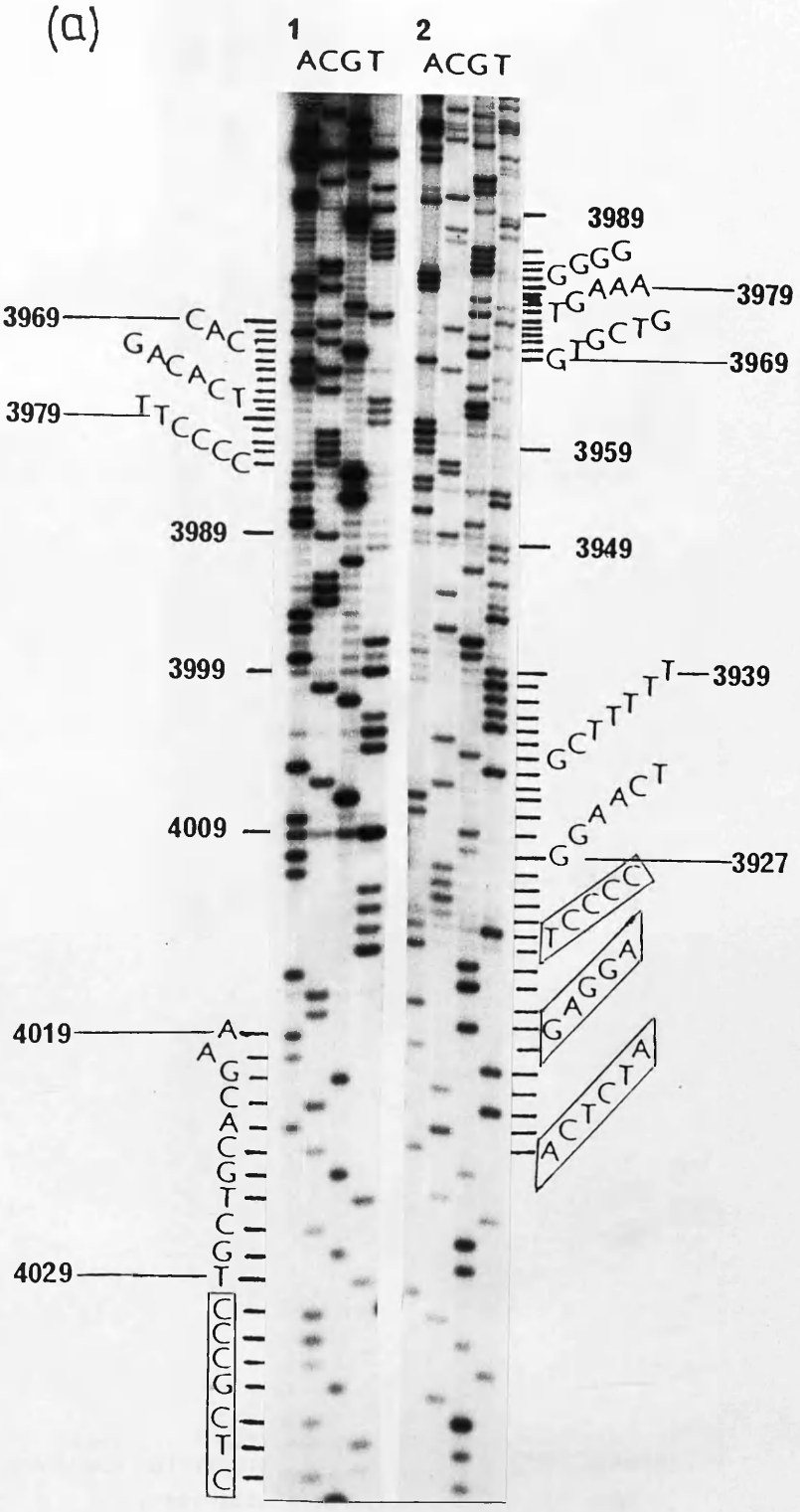
1). Plasmid sequencing using universal primer;

2). Plasmid sequencing using reverse sequencing primer.

The overlapping regions of sequence are shown near the top of the sequencing gel.

(b) Comparison of the cDNA sequence with BPV4 DNA:

1). Antisense strand
2). Sense strand.



(b)

1) bpv4

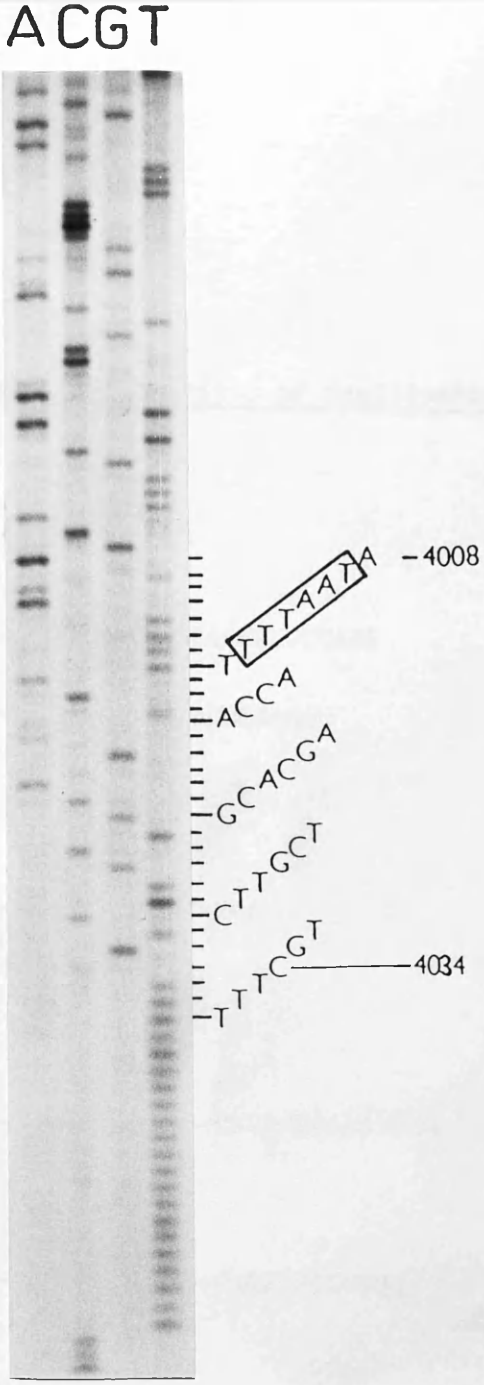
```
4031        4021        4011        4001        3991        3981        3971        3961
TTTCGTC TTGCTGCACG AACCATTTTA ATAGCATTTG CTATAACCCG TCAAGAGCCC CTTTCACAGC ACTGACACCT
CTCGCC CTGCTGCACG AACCATTTTA ATAGCATTTG CTATAACCCG TCAAGAGCCC CTTTCACAGC AC------
*   *  *
```

cDNA insert

2) bpv4

```
3929        3939        3949        3959        3969        3979        3989        3999
GAGCAAAGGA ACTGCTTTTT GGCTTCTGTT CGATTACCAA AAGGTGTCAG TGCTGTGAAA GGGGCTCTTG ACGGGTTATA
CTAGAGGATCCCCGGA ACTGCTTTTT GGCTTCTGTT CGATTACCAA AAGGTGTCAG TGCTGTGAAA GGGGCTCTTG -- ------
 * ***
```

cDNA insert

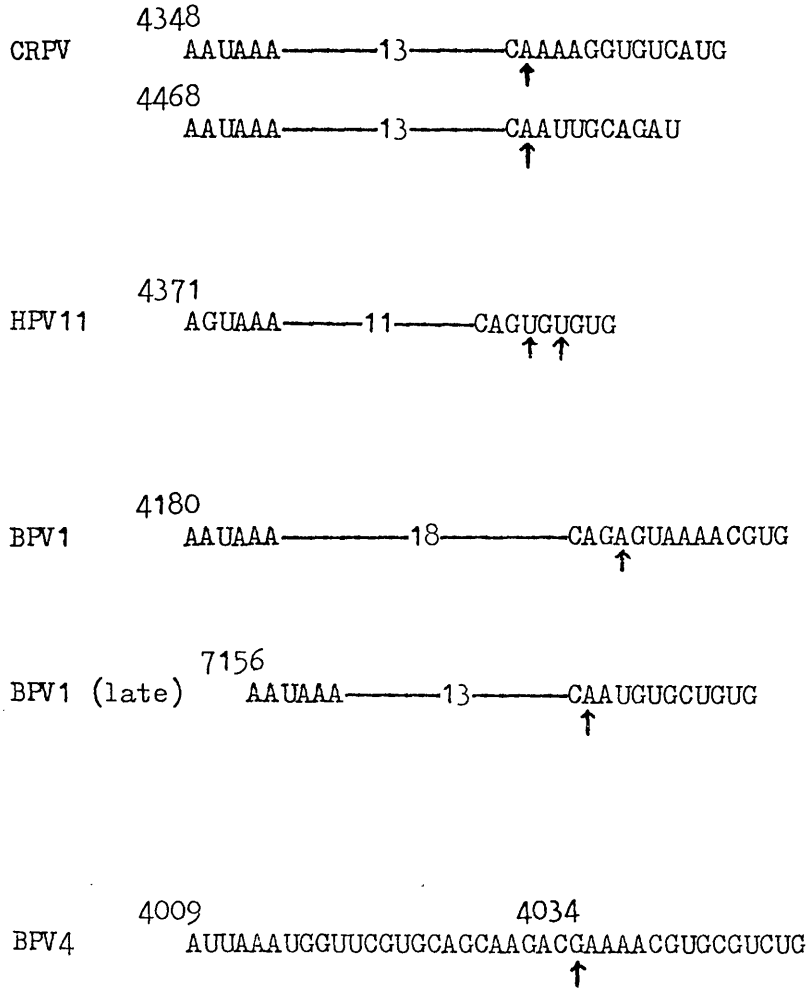



```
          4041        4031        4021        4011        4001
bpv4  CTCAGACGCA CGTTTTCGTC TTGCTGCACG GACCATTTTA ATAGCATTTG
7e11          T TTTTTTCGTC TTGCTGCACG AACCATTTTA ATAG
              * **
```

Figure 24. The early polyadenylation site of BPV4.

M13 sequencing of cDNA 7E11, showing the poly(A) tail and polyadenylation site. The early polyadenylation signal, ATTAAA, is boxed. A comparison of the cDNA sequence with that of BPV4, antisense strand, is shown below the photograph.

Figure 25. Early polyadenylation sites of papillomaviruses.



Numbers refer to nucleotide positions on the BPV4 genome.
Poly(A) addition sites are indicated by vertical arrows.

Figure 26. Sequencing of cDNA 10-1.

(a) p10-1 was sequenced in both directions using the plasmid sequencing method with:
1) universal primer;
2) reverse sequencing primer.

(b) Comparison of the sequence of cDNA 10-1 with the genomic sequence of BPV4.

Figure 27. Sequence and putative translation products of cDNA 10-1.

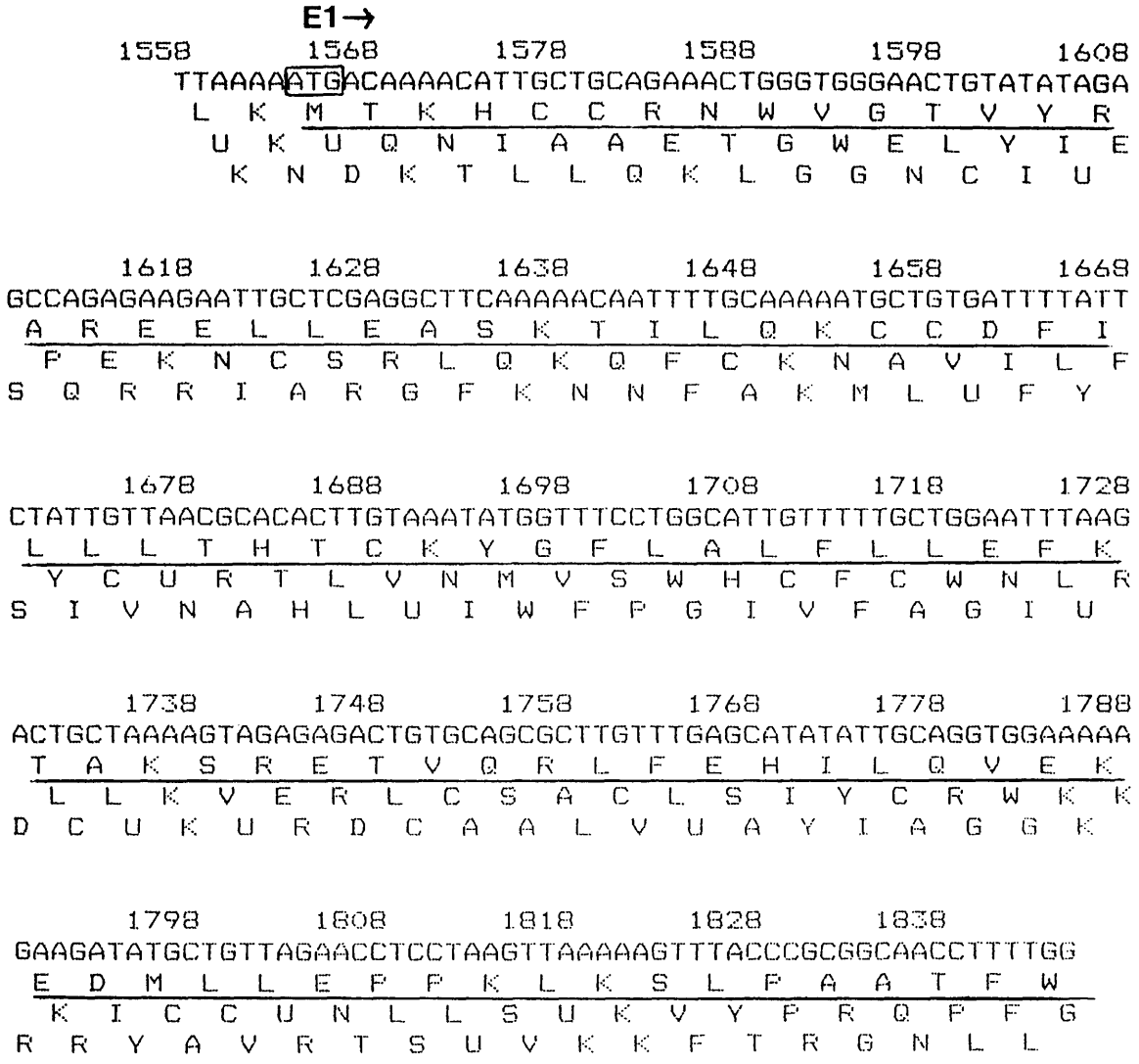


Figure 28. Sequencing of cDNA N.

(a) pN was sequenced in both directions using the plasmid sequencing method with:

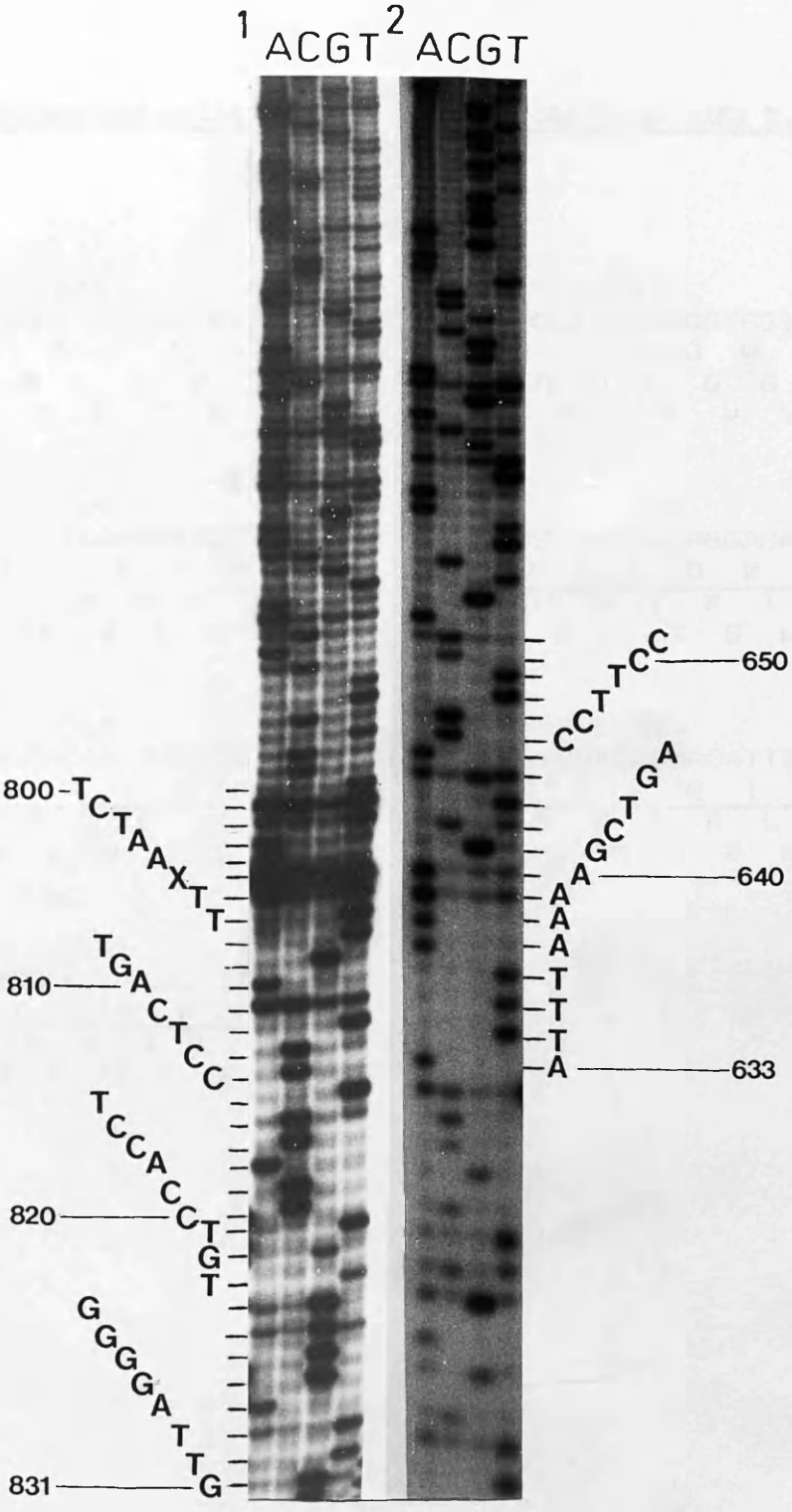
1) universal primer;

2) reverse sequencing primer.

(b) Comparison of the sequences of the 5' and 3' ends of cDNA with the genomic sequence of BPV4.

* = mismatched sequence

(a)



(b)

1)

```
bpv4
          831         821         811         801         791         781         771
TGCAAAAGGG TTAGGGGTGT CCACCTCCTC AGTTTCAATC TCCTCTTCAC AATGCAAGTT AATTGGACTA ATTG
  - ----G TTAGGGGTGT CCACCTCCTC AGTTTXAATC TCCTCTTCAC AATGCAAGTT AATTGGACTA ATTG
cdnaN                                     *
```

2)

```
bpv4
          642         652         662         672         682         692
GGCATTTAAAAGC TGACCTTCCA GTCTTAATTG CAGTAGGCGC CTAAGAGGGT GGTGGTGGTA TAAGTTC
   ATTTAAAAGC TGACCTTCCA GTCTTAATTG CAGTAGGCGC CTAAGAGGGT GGTGGTGGTA
cdnaN
```

Figure 29 Sequence and putative translation products of cDNA N.

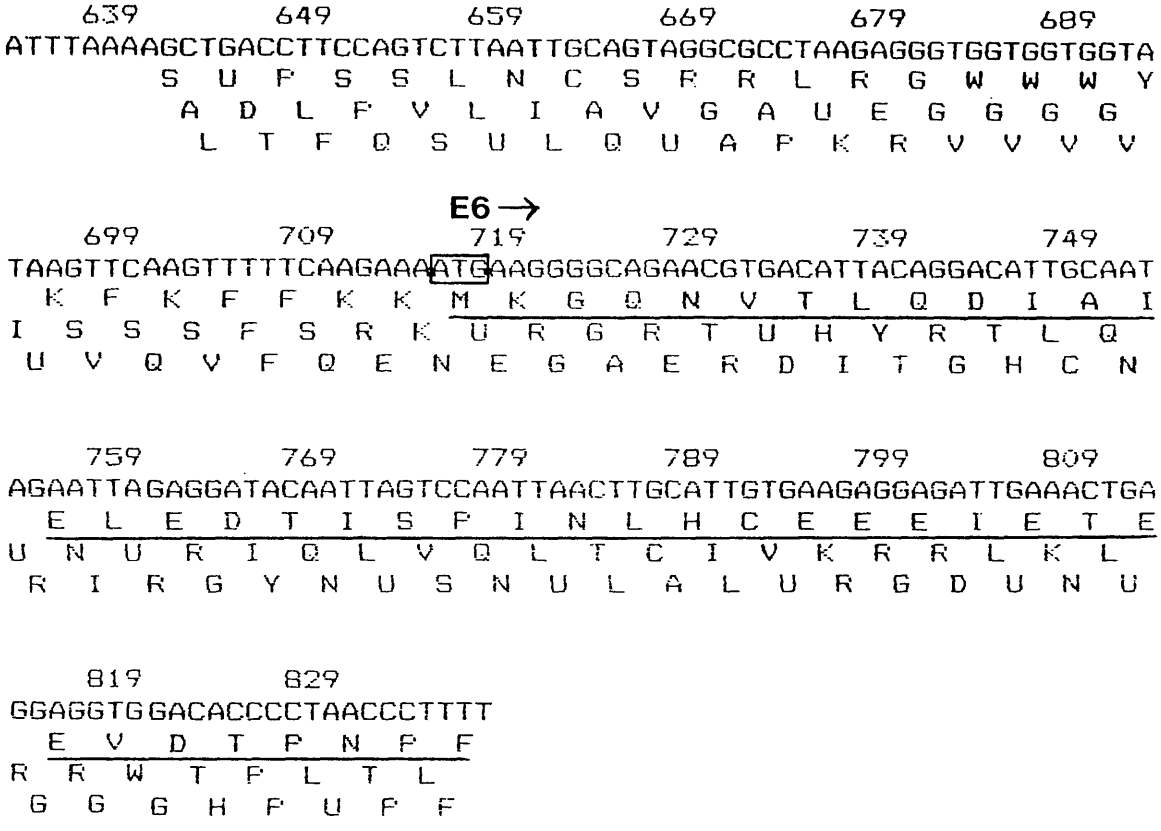


Figure 30. Sequencing of cDNA 7E11.

cDNA was cloned into M13 vectors and sequenced along its entire length. For sequencing strategy, see Figure 31. Selected autoradiographs are shown. Numbers refer to nucleotide positions in BPV4.

1) Sequencing of antisense strand of cDNA 7E11, cloned into M13 phage mp18, using primer 140 (Table 3.6). Sequence mismatches with the BPV4 genomic sequence are boxed. The splice junction is shown (nt1016/3375)

2) Sequencing of the sense strand of cDNA 7E11, cloned into M13mp19. Mismatched sequences are boxed.

3) Sequencing of the sense strand of cDNA 7E11 across the splice junction.

Figure 31. (overleaf).

(a) Sequencing strategy of cDNA 7E11 in M13 vectors.

The bold line represents plasmid DNA; boxes indicate the multiple cloning site and all numbers refer to the nucleotide sequence of BPV4. Diagonal lines show the position of the splice junction.

(b) Homology between cDNA 7E11 sequences and BPV4 genomic sequence.

Mismatches are indicated by asterisks.

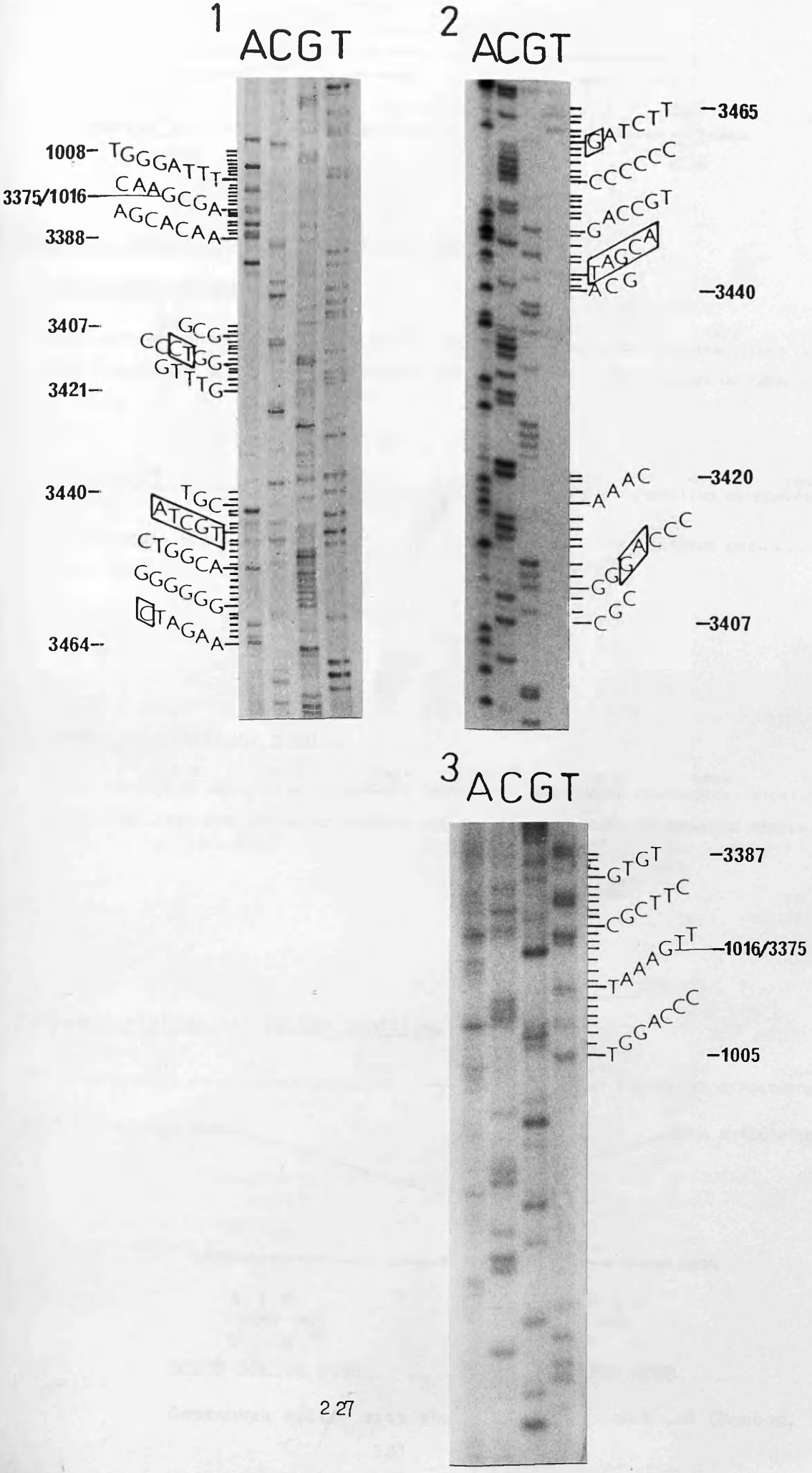
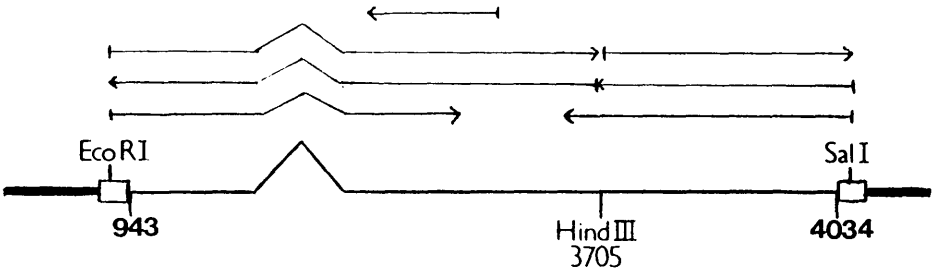

1
ACGT
1008
3375/1016
3388
3407
3421
3440
3464
2
ACGT
3465
3440
3420
3407
3
ACGT
3387
1016/3375
1005

Figure 31.

(a) Sequencing strategy of cDNA 7E11 in M13 vectors.

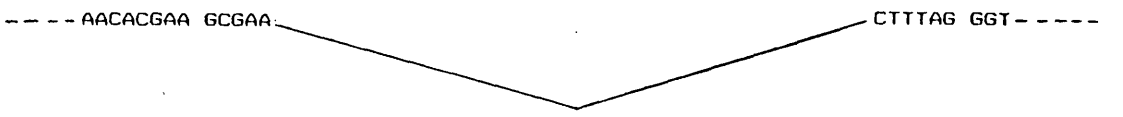


(b) Homology between cDNA 7E11 and BPV4 sequences.

1) Antisense orientation.

```
                3461        3451        3441        3431        3421        3411
bpv4  CTCTGGAAGA T GGGGGGACG GTCGTCATCG TCCTCGGGCT GGGTCCCCGG TTTGGGTACC CGCG----
7e11  CTCTGGAAGA TCGGGGGGACG GTCTGCTACG TCCTCGGGCG TGGTCCCCGG TTTGGGT CC CGCG----
                  *           ** **       * *              *
```

```
                3381        3371        3361                          1021        1011        1001
----- AACACGAA GCGAACTGGT AACAGAGGGA ——//—— AAGCACGGTA GTACCTTTAG GGTCCATTTC
```

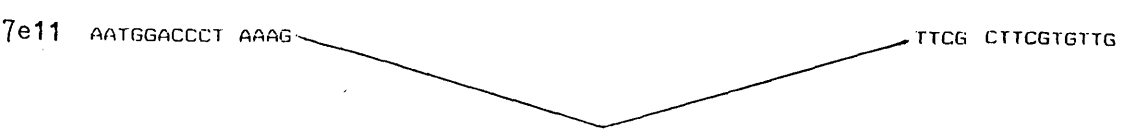


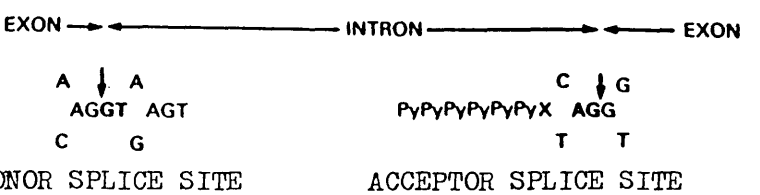
2) Sense orientation: exon 2.

```
                3409        3419        3429        3439        3449        3459        3469
bpv4  AGGACAACGC GGGTACCCAA ACCGGGGACC CAGCCCGAGG ACGATGACGA CCGTCCCCCC ATCTTCCAGA
7e11  -----CGC GGG ACCCAA ACCGGGGACC ACGCCCGAGG ACGTAGCAGA CCGTCCCCCCG ATCTT----
                *           **          ** **              *
```

3) Sense orientation: splice junction.

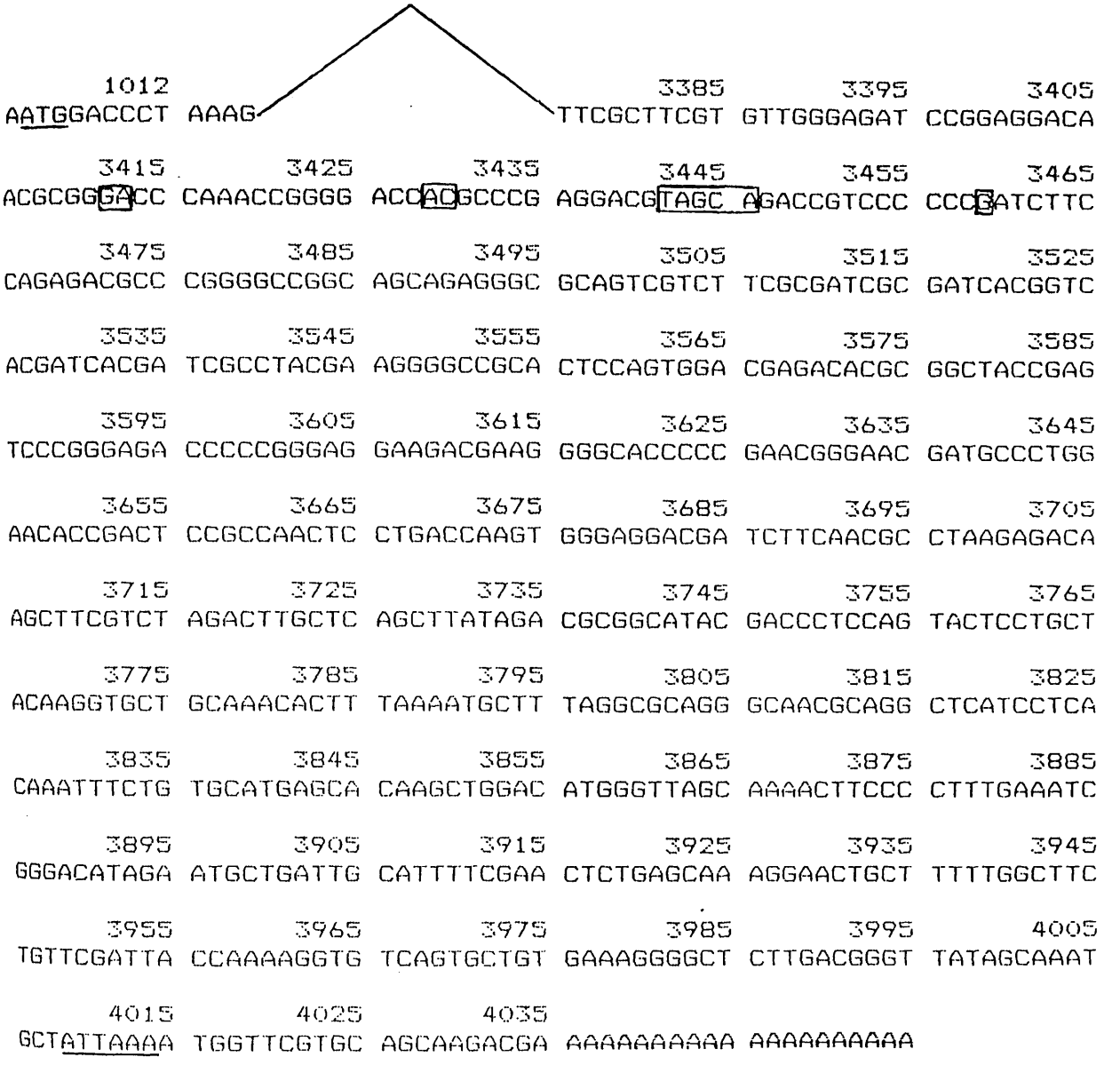
```
                1012        1022        1032                          3369        3379        3389
bpv4  AATGGACCCT AAAGGTACTA CCGTGCTTGA ——//—— CTCCCTCTGT TACCAGTTCG CTTCGTGTTG
```





Consensus splice site sequences (Breathnach and Chambon, 1981).

Figure 32. The nucleotide sequence of cDNA 7E11.



Numbers refer to nucleotide positions on the BPV4 genome. Diagonal lines indicate the splice junction. Boxed sequences are mismatches with the published BPV4 sequence (Patel et al., 1987).
Underlined sequences:

ATG - methionine codon;
ATTAAA - polyadenylation signal.

Figure 33. The positions of cDNA's relative to the genomic and transcriptional map of BPV4.

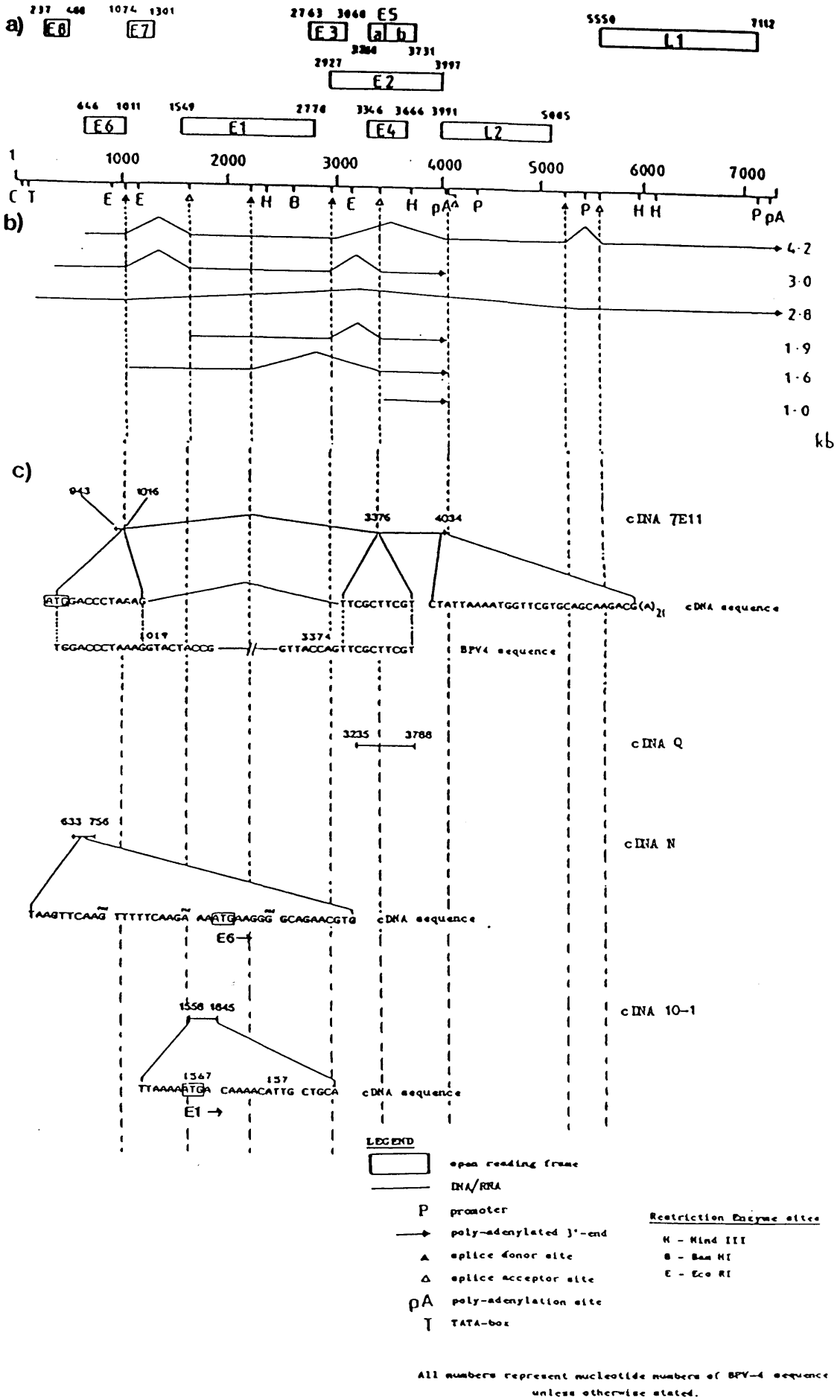


Figure 34. Plasmid sequencing of cDNA Q.

(a) Plasmid sequencing of pQ, using:

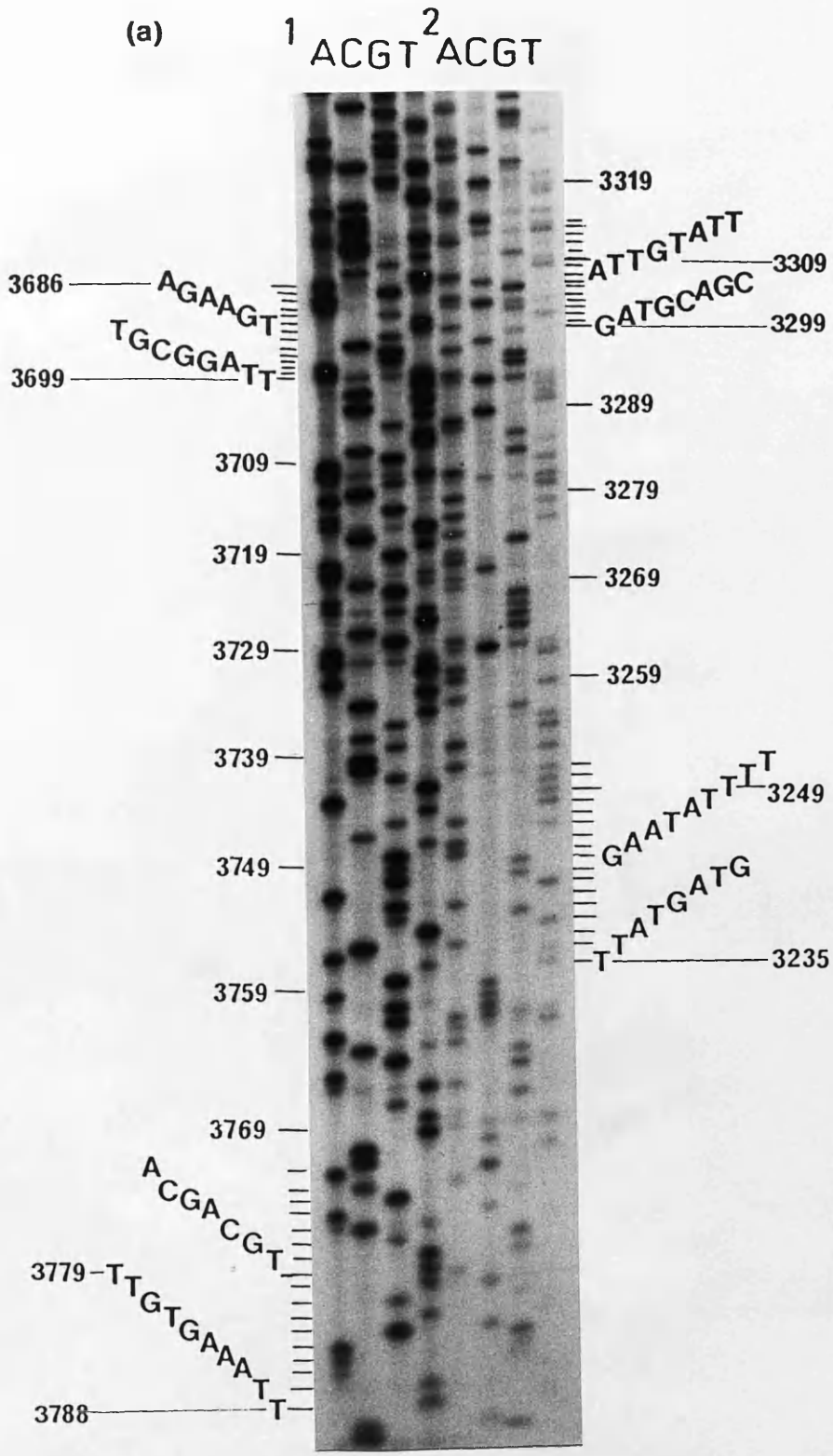
1) universal primer;

2) reverse sequencing primer;

the sequence of cDNA Q at the cloning site is shown next to the photograph.

(b) Comparison of the 3' and 5' ends of cDNA Q to the genomic sequence of BPV4.

(a)



(b)

1)

```
bpv4  3781        3771        3761        3751        3741        3731        3721        3711
TTTTAAAGTG TTTGCAGCAC CTTGTAGCAG GAGTACTGGA GGGTCGTATG CCGCGTCTAT AAGCTGAGCA AGTCTAGACG-----
 TTAAAGTG TTTGCAGCAC CTTGTAGCAG GAGTACTGGA GGGTCGTATG CCGCGTCTAT AAGCTGAGCA AGTCTAGACG-----

        3701        3691        3681
---AAGCTTGTCT CTTAGGCGTT GAAGATCGTC
-- AAGCTTGTCT CTTAGGCGTT GAAGA-----
```

2)

```
bpv4  3239        3249        3259        3269        3279        3289        3299        3309        3319
TTGGATTATG ATGGAATATT TTATATTGAT AATCAGGGGA ACAAGATATA TTATGTGAAC TTTCAGGACG ATGCAGCATT GTATTCTAAT
     TTATG ATGGAATATT TTATATTGAT AATCAGGGGA ACAAGATATA TTATGTGAAC TTTCAGGACG ATGCAGCATT GTATT-----
```

Figure 35. M13 sequencing of cDNA Q.

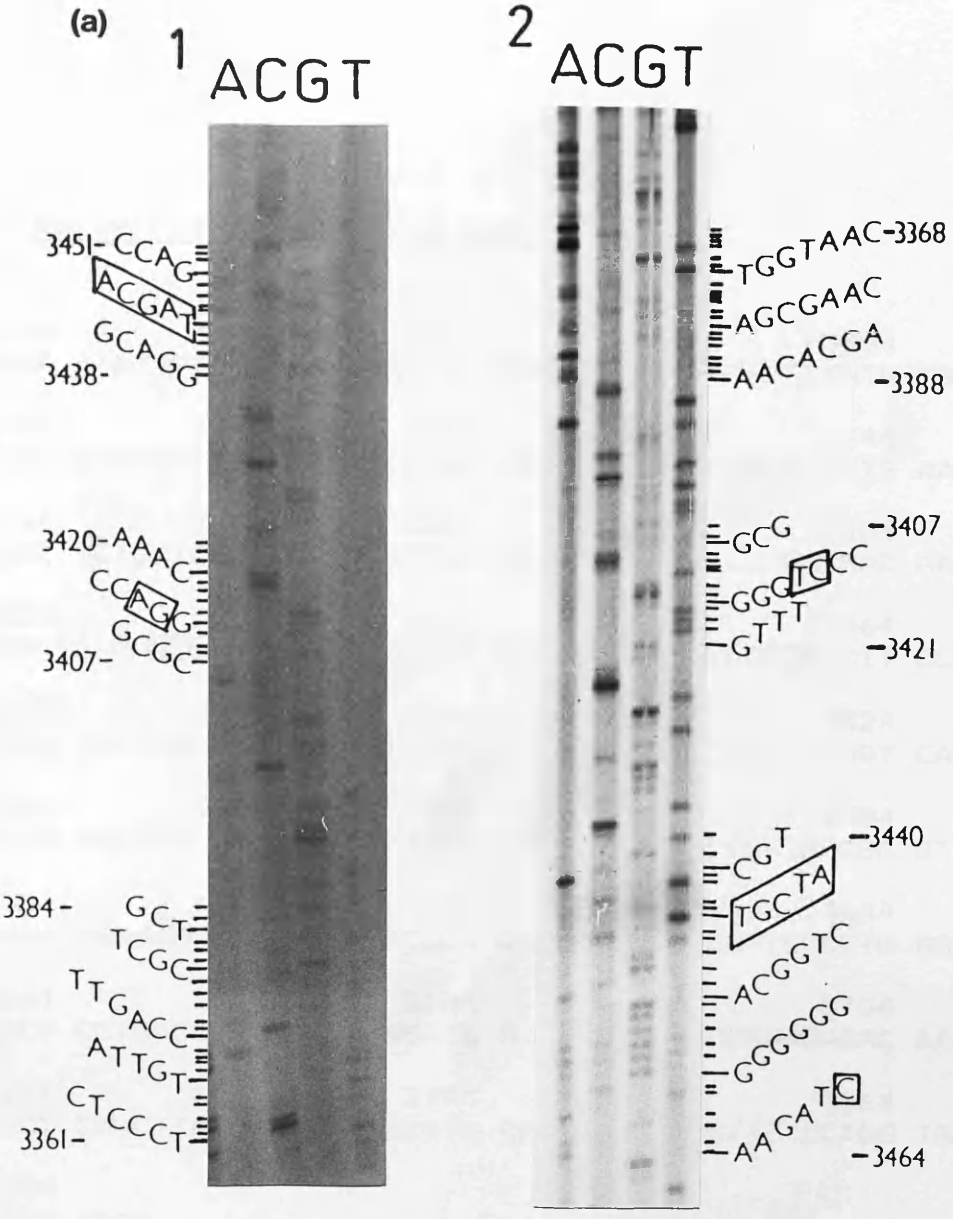
(a) Sequencing gel autoradiographs containing the region of mismatched sequences are shown:

1) Sequencing of the sense strand of cDNA Q in M13mp19.

2) Sequencing of the antisense strand of cDNA Q in M13mp18, using primer 140 (Table 3.6).

Sequence mismatches with the genomic sequence are boxed.

(b) Comparison of the sequences of cDNA Q with the BPV4 genomic sequence. Sequence mismatches are denoted by asterisks.



(b)

1)

```
                3369        3379        3389        3399
bpv4  TCCCTCTGT TACCAGTTCG CTTCGTGTTG GGAGATCCGG ----
Q     TCCCTCTGT TACCAGTTCG CTTCG-- - ---------- - - -

                3409        3419        3429        3439        3449        3459
---AGGACAACGC GGGTACCCAA ACCGGGGACC CAGCCCGAGG ACGATGACGA CCGTCCCCCC

---------CGC GGG ACCCAA ACCGGGGACC ACGCCCGAGG ACGTAGCAGA CC-- ---
                 *                   **            ** **
```

2)

```
                3461        3451        3441        3431        3421        3411
bpv4 CTCTGGAAGA T GGGGGGACG GTCGTCATCG TCCTCGGGCT GGGTCCCCGG TTTGGGTACC CGCG----

Q    ----AAGA TCGGGGGGACG GTCTGCTACG TCCTCGGGCG TGGTCCCCGG TTTGGGT CC CGCG----
               *             ** **           * *                  *

                3381        3371        3361
----CCAACACGAA  GCGAACTGGT AACAGAGGGA

------AACACGAA GCGAACTGGT AAC----
```

Figure 36. The nucleotide sequence of cDNA Q.

```
      3244        3254        3264        3274        3284        3294
TTATGATGGA ATATTTTATA TTGATAATCA GGGGAACAAG ATATATTATG TGAACTTTCA

      3304        3314        3324        3334        3344        3354
GGACGATGCA GCATTGTATT CTAATTCTGG CATGGGGCAA GTGCATTTTG AAAGCAAAGT

      3364        3374        3384        3394        3404        3414
TCTTTCTCCC TCTGTTACCA GTTCGCTTCG TGTTGGGAGA TCCGGAGGAC AACGCGGGAC

      3424        3434        3444        3454        3464        3474
CCAAACCGGG GACCACGCCC GAGGACGTAG CAGACCGTCC CCCCGATCTT CCAGAGACGC

      3484        3494        3504        3514        3524        3534
CCGGGGCCGG CAGCAGAGGG CGCAGTCGTC TTCGCGATCG CGATCACGGT CACGATCACG

      3544        3554        3564        3574        3584        3594
ATCGCCTACG AAGGGGCCGC ACTCCAGTGG ACGAGACACG CGGCTACCGA GTCCCGGGAG

      3604        3614        3624        3634        3644        3654
ACCCCCGGGA GGAAGACGAA GGGGCACCCC CGAACGGGAA CGATGCCCTG GAACACCGAC

      3664        3674        3684        3694        3704        3714
TCCGCCAACT CCTGACCAAG TGGGAGGACG ATCTTCAACG CCTAAGAGAC AAGCTTCGTC

      3724        3734        3744        3754        3764        3774
TAGACTTGCT CAGCTTATAG ACGCGGCATA CGACCCTCCA GTACTCCTGC TACAAGGTGC

      3784
TGCAAACACT TTAA
```

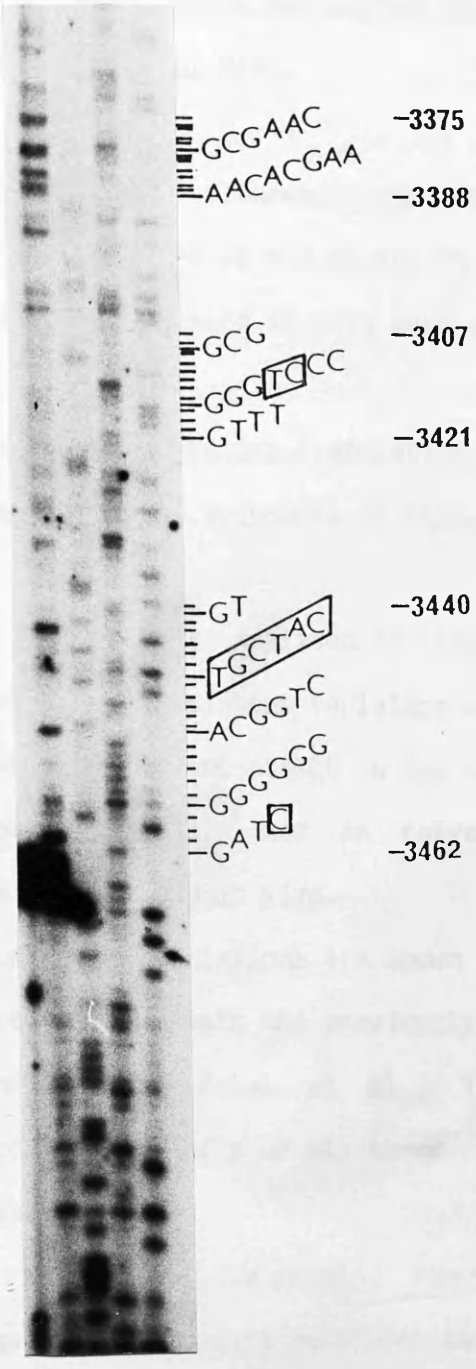
Numbers refer to nucleotide positions on the BPV4. genome.
Boxed sequences are mismatched with the genomic sequence.

Figure 37. RNA sequencing using primer 140.

(a) Sequencing autoradiograph of BPV4 transcripts in the region of mismatched sequence. These are shown boxed alongside the photograph.

(b) The RNA sequence is compared to the genomic sequence. Mismatched sequences are indicated by asterisks. The sequence of cDNA 7E11 across the splice junction is also compared.

(a)



(b)

bpv4

```
      3461        3451        3441        3431        3421        3411        3401        3391
CTCTGGAAGA T GGGGGGACG GTCGTCATCG TCCTCGGGCT GGGTCCCCGG TTTGGGTACC CGCGTTGTCC TCCGGATCTC
        GA TCGGGGGGACG GTCTGCTACG TCCTCGGGCG TGGTCCCCGG TTTGGGT CC CGCGTTGTCC TCCGGATCTC
           *             ** **           * *                 *

      3381        3371        3361
CCAACACGAA GCGAACTGGT AACAGAGGGA

CCAACACGAA GCGAACTTTA (G)GGT  AT----
                  **   *** ******

          7e11 CTTTA GGGTCCATTi
                       *   **
```

3.5. Proposed amendment to the nucleotide sequence and genome organisation of BPV4.

The evidence in favour of alterations in the sequence and changes in the nomenclature of the E4 and E5 orf's of BPV4 will be presented and discussed in this subsection.

3.5.1. Comparison of the putative translation products of the genomic and cDNA/RNA sequences of BPV4.

The sequencing experiments analysed in the previous two subsections revealed a consistent variation on the sequence of BPV4 DNA between nt3412 and nt3460 in two cDNA clones from independent cDNA libraries, and in polyadenylated RNA transcribed from this region *in vivo*.

The described sequence variations are shown in the coding strand in Figure 38(b) beneath the previously defined BPV4 sequence in this region (Patel *et al.*, 1987). Putative translation products and orf's of all three reading frames are shown with each sequence.

Reading from 5' to 3' (left to right), the first mismatch between the sequences is a point deletion corresponding to nt3413. The absence of a T residue causes a frameshift such that reading frame 1 (rf1) reads into rf2, rf2 into rf3 and rf3 into rf1. The joining of rf1 with rf2 results in the loss of a valine codon in rf1 and a tyrosine codon in rf2, and their replacement by an aspartate codon. The result of the frameshift is that the E4 orf reads into the E2 orf at this

point, E2 reads into E5a and E5a into E4.

Inversion of the CA dinucleotide at nt3430 causes codon changes in all three rf's. A glutamine codon in rf1 of the genomic sequence is changed to a threonine in rf3 of the cDNA/RNA sequence; similarly a serine of rf2 is changed to arginine in rf1 and a proline in rf3 to histidine in rf2.

The most extensive mismatch occurs from nt3443 to nt3447: the sequence ATGAC in the genomic sequence is changed to TAGCA in the cDNA/RNA sequence. The effect of these changes on the coding capacity of this region is considerable, since a methionine codon of the E2 orf in rf2 of the genomic sequence is changed to a translation termination codon, TAG, in rf1 of the cDNA/RNA sequence, and a termination codon, TGA, in rf3 at the end of the E5a orf is changed to a serine codon in rf2. In addition, a threonine codon in rf2 in the genomic sequence is changed to glutamine in rf1, next to the termination codon, and two aspartate codons in rf1 are changed to valine and alanine in rf3.

Finally, the point insertion after nt3459 causes another frameshift in reverse orientation to the first, restoring the rf's of the cDNA/RNA sequence to their original context. The insertion of a G residue alters a histidine codon in rf3 to arginine in rf2, before the frameshift, and creates an extra aspartate codon in rf3.

The overall effects of these changes on the genome organisation are as follows: the three orf's encoded in this region change frame over a stretch of 16 codons. The E4 orf is divided by a stop codon at nt3443 into two short orf's of

32 and 74 codons, respectively, of which the smaller, 5' orf has no ATG codon and the larger only has one at codon 65. The E5 orf is no longer split by the stop codon at nt3444 and is now 148 codons in length, having an ATG at codon 5. The E2 orf remains continuous with alterations in the 16 codon frameshift region.

The altered coding potential of the cDNA/RNA sequence therefore represents a considerable divergence from the genome organisation of Patel et al. (1987). Agreement between sequences of two cDNA clones produced from RNA from papillomas of two different cases of bovine oesophageal papillomatosis suggested that these represented at least a common variant of BPV4. However, these changes affect the E4 orf, the translation products of which are thought to have a central role in virion production in benign tumours (Doorbar et al., 1986; Breitburd et al., 1987). Since this orf in the cDNA/RNA sequence was split and could not be expressed by any known transcript of BPV4, it was considered that a reappraisal of the nomenclature of the E4 and E5 orf's might be required. Further evidence for this was provided by an analysis of the coding capacity of cDNA 7E11 and comparisons between this cDNA and those obtained from the transcripts of other papillomaviruses. It was therefore decided to suspend the original judgement on the nomenclature of these orf's and to temporarily designate the now continuous E5 orf, X, and the split E4 orf, Y (Figure 38).

3.5.2. The coding capacity of cDNA 7E11.

Neither of the short Y orf's encoded by the cDNA/RNA sequence contains an ATG codon in a suitable position for the translation of more than 10 amino acids, while the X orf could potentially be translated into a 143 amino acid protein. In HPV1a-induced papillomas the E4 translation product is the most abundant viral protein (Doorbar et al., 1986). It may therefore be encoded by the major transcript of HPV1 (Fig. 4; Chow et al., 1987b) in which a short leader exon mapping to the 3' end of the E7 orf is spliced onto the 3' terminal exon of the early region. Similarly spliced major transcripts have been mapped in HPV11- and HPV6-induced condylomata (Chow et al., 1987). cDNA's of the major transcript of HPV11 show that the E4 orf is expressed as a fusion with a four codon orf from the 5' end of the E1 region (Fig. 39(b); Nasseri et al., 1987). In BPV1-induced fibropapillomas, the major viral transcript in the epithelial part of the tumour encodes an E4 fusion orf in which the first two amino acids are encoded by sequences from the ncr (Fig. 2; Baker and Howley, 1987).

The coding capacity of cDNA 7E11 is shown in Figure 39(a). The X orf is fused to a short orf in the 5' exon, which has an ATG codon at nt1004; thus it encodes a 123 codon fusion orf consisting of the 3' 119 codons of the X orf attached to 4 codons from the end of the E6 region, from the first ATG. The 3' half of the E2 orf cannot be expressed, as was thought earlier (3.3.3.(v)), as there is no methionine codon at

position 3443 in the cDNA, although a 53 codon polypeptide corresponding to the 3' end of the E2 orf could be translated from a very weak initiator codon at nt3839, downstream of the X orf. No other ATG-containing orf's are encoded by cDNA 7E11, but the possibility of the E6 orf being encoded in the transcript represented by cDNA 7E11 had not at this stage been excluded.

The splicing and putative translational organisations of cDNA 7E11 therefore have very convincing similarities to those of the major transcripts of BPV1, HPV11 and probably HPV's 6 and 11, suggesting that the encoded fusion orf is the BPV4 homologue of those found in the other viruses. This appears likely in view of the highly conserved genomic and transcriptional organisations of the papillomaviruses.

On the basis of this evidence, it was considered that the X orf specified by the cDNA/RNA sequence was similar to the E4 orf of other papillomaviruses. The assignment of orf numbers in this part of the genome is difficult due to low homology at both the nucleic acid and protein level between the E4 and E5 orf's of papillomaviruses (Danos et al., 1984). The E5a and E5b orf's of BPV4 were named by their similarity to the split E5 orf's of HPV's 6 and 11, and the E4 orf due to its length of approximately 100 codons and lack of an ATG codon, as in HPV's 16 and 33. In view of the changes specified by the cDNA/RNA sequence, it seems likely that these orf's were inappropriately assigned.

This conjecture was examined by analysing the degree of homology between the putative translation products of the X

and Y orf's and the E4 and E5 orf's of other papillomaviruses.

3.5.3. Reassignment of the E4 and E5 orf's of BPV4 by homology and alignment with those of other papillomaviruses.

3.5.3.(i) Homology of X and Y with E4 and E5 orf's of other papillomaviruses.

Putative translation products of the X and the larger Y orf's defined by the cDNA/RNA sequence were compared to the E4 and E5 orf's of other papillomaviruses by amino acid sequence homology and alignment, using the function "Protein alignment by similarity" in the Microgenie computer analysis program (Queen and Korn, 1984). This function calculates the alignment of two amino acid sequences giving the maximum number of of identical or similar matched residues, e.g arginine to arginine, histidine or lysine. Homology is given as a percentage of the length of the longer sequence in the analysis. The results of the homology comparisons are presented in Table 3.5.

As has already been mentioned, the E2/E4 region of papillomaviruses is poorly conserved between different types (Danos et al., 1984). This was reflected in the low percentage homology values obtained for both the X and Y orf's.

For orf X, the highest homology was found with the E4 orf of

HPV8. This was surprising in view of the exceptional nature of the HPV8 E4 orf which at 229 codons is more than twice the average length of 90 codons. The E4 orf of HPV8 encodes protein sequences homologous to the Epstein-Barr virus nuclear antigen 2, primarily as a result of the high (22%) proline content (Fuchs et al., 1986). The alignment of the two orf's is, however, poor throughout their entire length (not shown) and the higher homology may be due to the larger size of the HPV8 sequence increasing the possibility of similar amino acids being encoded. The degree of homology between the E4 orf of HPV8 and those of other papillomaviruses is extremely low (Fuchs et al., 1986).

Of greater interest are the homology data resulting from comparisons between the BPV4 X orf and the E4 orf's of HPV16 (Seedorf et al., 1985), HPV18 (O. Danos, pers. comm.) and HPV33 (Cole and Streeck, 1986), and between the X fusion orf of cDNA 7E11 and the E1-E4 fusion orf of the major transcript of HPV11 (Nasseri et al., 1987). These homologies were 22.1%, 22.1%, 20.3% and 20.1%, respectively, considerably higher than the homologies with the E5 orf's of the same viruses, which range from 10.1% to 13.5%. It is significant that, like BPV4, all of these viruses infect mucosal epithelium (Campo and Jarrett, 1987) and, in common with HPV8, are associated with cancer (Gissmann et al., 1984).

Homology between the X fusion orf of cDNA 7E11 and a putative E1-E4 fusion orf of HPV6b (Nasseri et al., 1987, based on Schwarz et al., 1983) was slightly lower at 17.9%; homology between the BPV4 X orf and the E4 orf's of BPV1 (Chen et al.,

1982) , CRPV (Giri et al., 1985) and HPV1a (Danos et al., 1982) was lower at between 17.6% and 16.1%; nevertheless, homologies with the E5 orf's of these skin-specific papillomaviruses was again lower than with their E4 orf's at 10.1 % to 14.2 %.

Comparisons of the larger of the two Y orf's of the cDNA/RNA sequence with the E4 and E5 orf's of the other papillomaviruses in general gave very low percentage homology values. The highest values were obtained with the E4 orf's of BPV1, CRPV and HPV1 at 17.5%, 14.6% and 17.4%. However, comparisons of the values obtained with the E4 and E5 orf's of each papillomavirus did not yield a consistent pattern: for instance, homology between the BPV4 Y orf and the HPV16 E4 and E5 orf's was 7.3% and 11.3%, respectively, while for the HPV33 these figures were 12.9% and 8.5%, respectively. For CRPV, the two figures were identical at 14.6%.

Therefore the X orf specified by the cDNA/RNA sequence of BPV4 is , with the exception of HPV1, more homologous to the E4 orf's of other papillomaviruses than to their E5 orf's.

The larger Y orf of BPV4 in general shares slightly higher homology with the E5 orf's of other papillomaviruses than with their E4 orf's, but homology values are low and no clear pattern emerges. This and the lack of suitable translational initiation codons may indicate that the split Y orf is not expressed.

3.5.3.(ii). Comparison of the X orf of BPV4 with the E4 orf's of HPV's by alignment.

The protein alignments calculated for the X orf of BPV4 and the E4 orf's of HPV's 16, 18, 33 and 1a, and the E1-E4 fusion orf's of HPV's 6b and 11 were used to construct the alignment plot in Figure 40. Alignments of similar or identical amino acids across four or more of the seven sequences are boxed. Although few continuous regions of homology are found, the sequence divergence of this region of the genome may have led to the conservation of only a few positions essential to structure and/or function of the encoded polypeptide. The most conserved region of the X orf is in the carboxy-terminal end, consisting of alternating hydrophobic and hydrophilic residues, but small blocks of conserved sequences occur throughout most of the orf.

The most highly conserved amino acids are arginine, proline and leucine. Presumably the leucine residues of X would form part of the hydrophobic interior of this cytoplasmic and nuclear protein (Doorbar et al., 1986; Breitburd et al., 1987). The proline residues are probably structural. The charged arginine groups may have functional significance. Out of 148 amino acids, 24 are either arginine or lysine, but the 26 glutamate and aspartate residues indicate that the protein may have an overall negative charge. This balance is the same for the X fusion orf of cDNA 7E11. In addition, the X protein could be phosphorylated at the three serine residues found in the more conserved middle to carboxy-terminal region, and at

a conserved tyrosine residue. Although the alignment of the E4 orf's is somewhat disjointed, the conservation of specific regions, particularly of potential structurally and functionally important amino acids, suggests that the X orf of BPV 4 may encode a similar protein to the E4 orf of other papillomaviruses.

3.5.3.(iii). The BPV4 X orf corresponds to the E4 orf of other papillomaviruses: discussion.

The E4 protein encoded by HPV1a has been characterised and its expression in papilloma cells analysed (Doorbar et al., 1986; Breitburd et al., 1987). A number of proteins were found which were antigenic to antisera raised against E4 epitopes. These proteins were all acidic, and up to seven phosphorylated isoelectric variants were found for one of the larger monomeric proteins, correlating with the seven potential phosphorylation sites encoded by the full E4 orf. The several forms of the HPV1 E4 protein are considered to be derived from two monomeric polypeptides of 10 or 11 kD and 16 or 17 kD. The larger is regarded as the translation product of the full E4 orf on the basis of amino acid quantitation (Doorbar et al., 1986), and is expressed in the basal layers of papillomas (Breitburd et al., 1987) i.e. at an early stage in the viral infective cycle. The 11 kD protein, arising later in differentiation of papilloma cells, is thought either to be a cleavage product of the 17 kD protein, or the product of a spliced transcript. The latter interpretation is

strongly supported by amino acid analysis of the smaller protein which diverges slightly from that of potentially corresponding fragments of the larger E4 protein, by the failure of some antisera to recognise both proteins (Doorbar et al., 1986) and by the transcriptional map of HPV1 (Chow et al., 1987b), which shows that the major transcript may be spliced to express the smaller E4 protein. The cDNA representing the major transcript of BPV1 also encodes just such a fusion orf, and it is interesting to note that this transcript initiates in the late promoter region of BPV1, indicating that it is expressed only in the more differentiated layers of the epithelium. The similarity between cDNA 7E11 and the major, E4-encoding transcripts suggests that it may encode the BPV4 homologue of the smaller E4 protein.

The X orf specified by the cDNA/RNA sequence of BPV4 shows greatest homology to the E4 orf's of papillomaviruses infecting mucosal epithelium, suggesting that the function of the encoded protein relates to tissue specificity. This has already been suggested as a reason for the poor conservation of this region of the genome. The function of the E4 proteins is as yet unknown, but it is exclusively associated with the virion-productive cells of HPV1-induced papillomas (Breitburd et al., 1987) and accounts for 20% of wart protein content. Productively infected wart cells do not produce keratins normally associated with terminally differentiating epithelial cells, and it has therefore been suggested that the E4 protein is involved in the subversion of cellular

differentiation mechanisms to favour virion replication and maturation. The differentiation pathway of mucosal epithelial cells differs from that of skin epithelium and a different set of keratins are expressed; if the hypothesised function of the E4 orf is in fact the case, the E4 orf's of mucosa-specific papillomaviruses would be expected to show specialised adaptations to their host tissue type. Therefore the closer homology of the X orf of BPV4 to the E4 orf's of these viruses may reflect conservation of these specialised sequences. The E4 orf's of HPV's 16, 18 and 33 are all around 40% homologous to one another, so the decreased homology of the BPV4 orf relative to this may be due to the differences in the mucosal cells between host species.

The X orf of BPV4 shows less homology to the E5 orf's of other papillomaviruses. Of these, only the E5 orf of BPV1 has been characterised (Schiller et al., 1985; Schlegel and Wade-Glass, 1987). The amino acid sequence of this transforming orf is widely divergent from those of other papillomaviruses. It expresses a 7 kD, 44 amino acid protein from an ATG codon well within the orf: this protein is highly hydrophobic, consisting of 33% leucine, and is localised in cell membranes. It also contains a Cys-X-Cys motif, conserved only in other fibropapillomavirus E5 orf's.

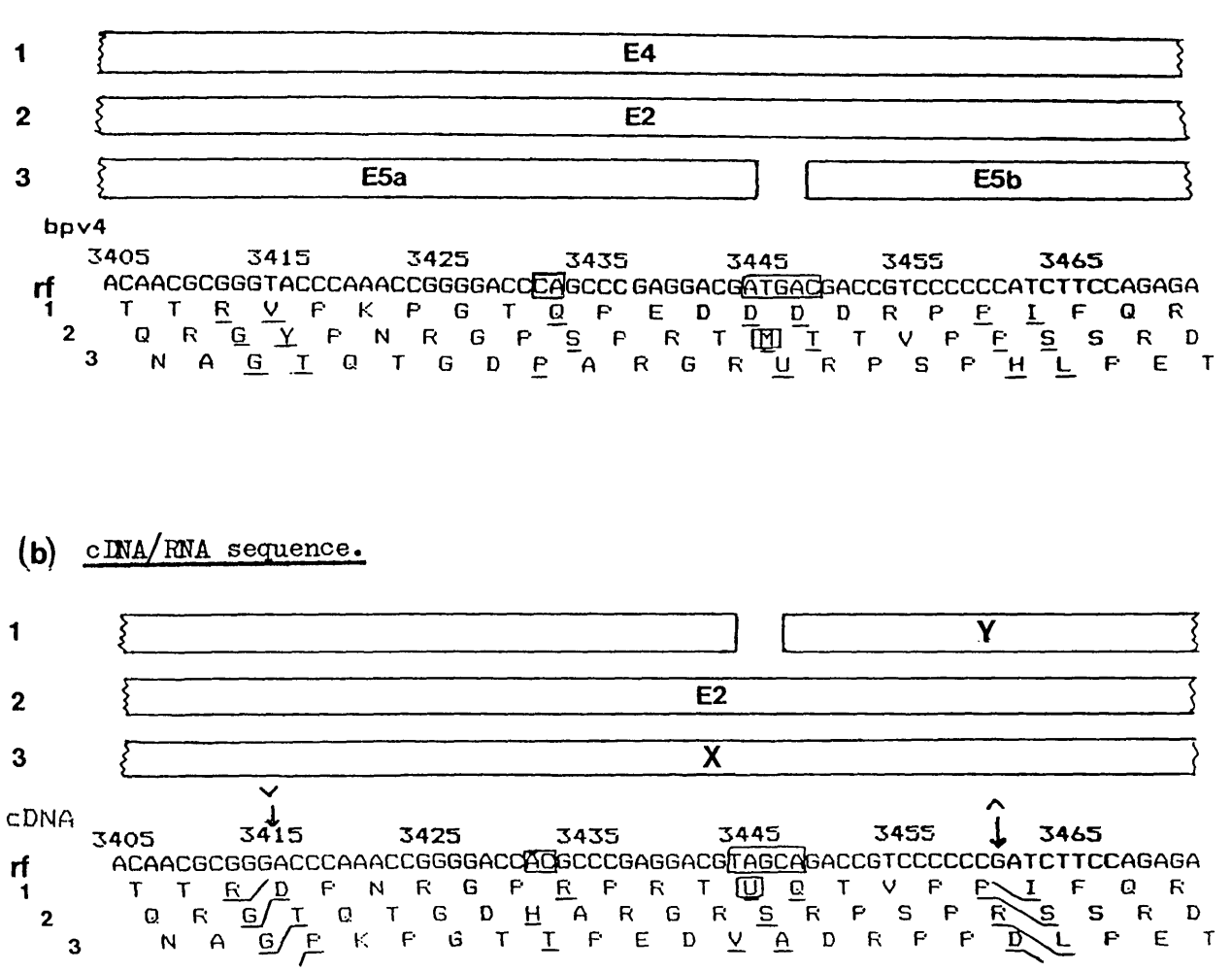
None of these features is shared by the X orf of BPV4, and the 14% homology with the E5 orf of BPV1 is confined mainly to the untranslated 5' codons of the latter. Homology to the E5 orf's of other papillomaviruses is also consistently low (Table 3.5).

3.5.3.(iv) Amendment to the sequence and genome organisation of BPV4.

On the basis of the above data and arguments stemming therefrom, I propose to amend the sequence of BPV4 DNA in the E2/E4/E5 coding region to that represented by the sequences of cDNA's 7E11 and Q and of BPV4 transcripts from this region. In addition, the temporarily assigned X orf should be redesignated E4. The absence of translational initiator codons in the split Y orf, and of potential splice acceptor sequences within the orf which could lead to its expression as a fusion orf encoded by a spliced transcript, suggest that these short orf's are not expressed. However, until further analysis can be carried out, these orf's are renamed E5a and E5b.

The proposed changes to the new E4 and E5 orf's are detailed in Figure 38; the amended genome organisation is presented beneath the published map in Figure 41.

Figure 38. Comparison of the BPV4 cDNA/RNA sequence and its putative translation products with those of the published sequence.



Amino Acid Symbols

Amino Acid	*3-letter code*	*1-letter code*
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Termination	End	U

LEGEND.

- V — point deletion
- ^ — point insertion
- frameshift
- X — altered codon
- orf
- $^{1}_{2}{}_{3}$ — reading frames

Altered sequences are boxed.
Numbers refer to nucleotide positions in the BPV4 genome.

Figure 39. (a) The coding capacity of cDNA 7E11.

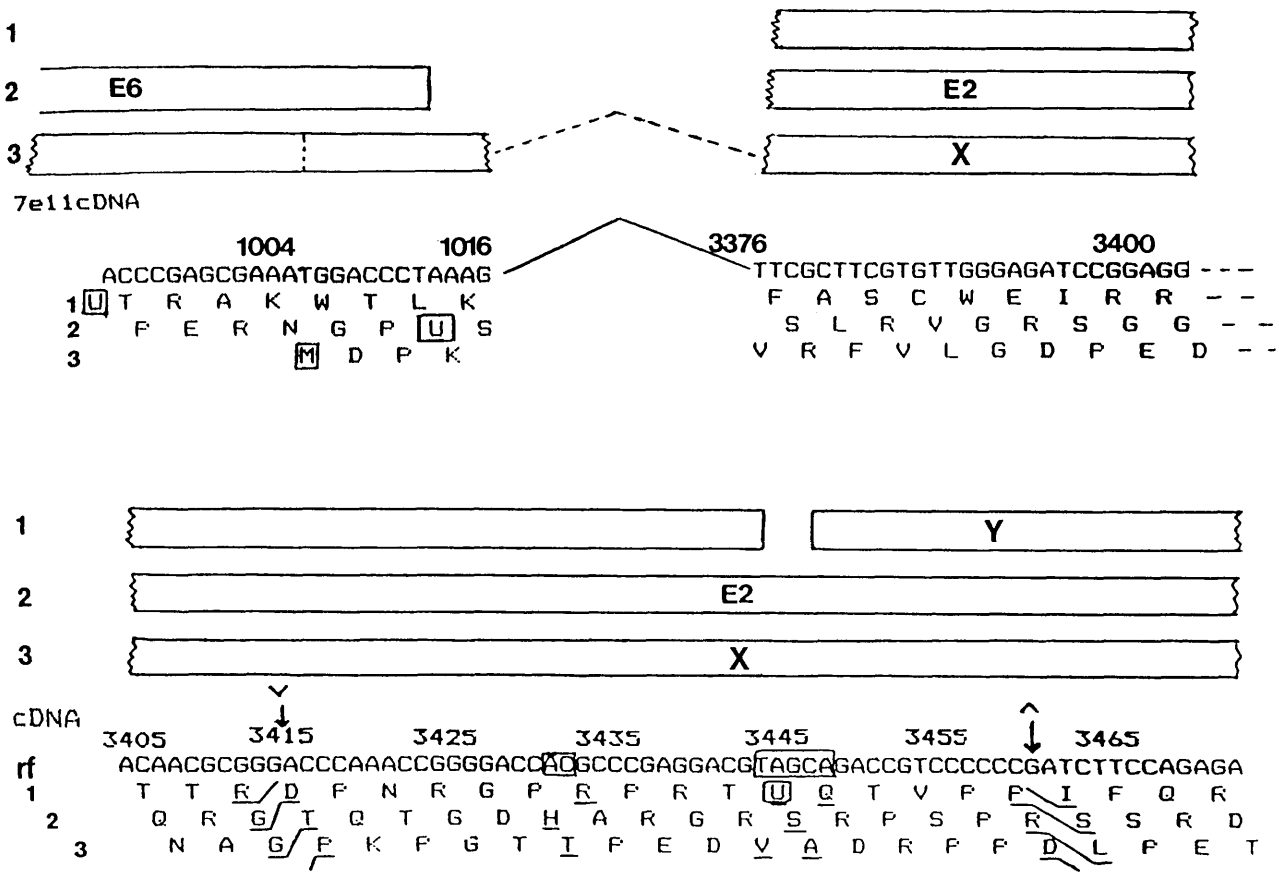
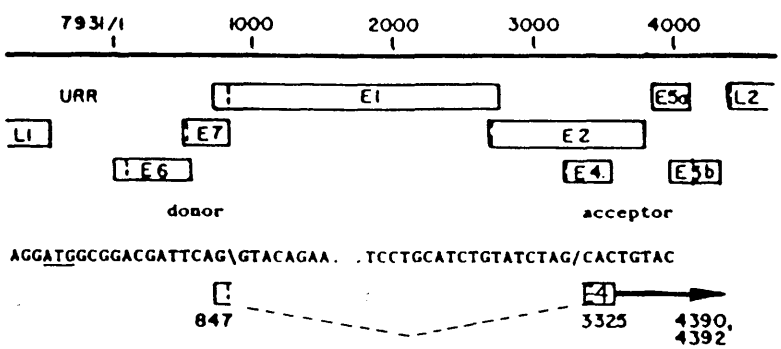


Figure 39. (b) Analysis of a cDNA clone of the major transcript of HPV11.



Nasseri et al., 1987.

LEGEND.

V	point deletion
^	point insertion
⟋	frameshift
X	altered codon
▭	orf
$^{1}_{2}{}_{3}$	reading frames

Altered sequences are boxed.
Numbers refer to nucleotide positions in the BPV4 genome.

Table 3.5 Homology between putative translation products of orf's X and Y of BPV4 and the E4 and E5 orf's of other papillomaviruses.

PV Type & orf		BPV4 X		BPV4 Y
	Number of amino acids	148		74
		1st Met	5	65
HPV8 E4	229	none	25.9 %	12.7 %
HPV16 E4	96	none	22.1 %	7.3 %
HPV16 E5	79	none	13.5 %	11.3 %
HPV18 E4	91	4	22.1 %	9.9 %
HPV18 E5	78	15	11.4 %	13.6 %
HPV33 E4	83	none	20.3 %	12.9 %
HPV33 E5	80	6	10.7 %	8.5 %
HPV11 E1^E4[1]	90	1	20.2 %[2]	13.9 %[3]
HPV11 E5a	94	4	10.1 %	13.7 %
HPV11 E5b	127	55	10.7 %	13.4 %
HPV6 E1^E4[1]	91	1	17.9 %[2]	14.8 %[3]
HPV6 E5a	103	13	10.1 %	10.7 %
HPV6 E5b	132	60	12.4 %	14.4 %
BPV1 E4	126	7	17.4 %	17.5 %
BPV1 E5	99	56	14.1 %	8.9 %
CRPV E4	212	1	16.5 %	14.6 %
CRPV E5	103	1	14.1 %	14.6 %
HPV1 E4	91	9	16.1 %	17.4 %
HPV1 E5	92	51	14.2 %	15.1 %

[1] E1^E4 fusion orf created by splice junction of major transcript.

[2] Homology with X fusion orf of cDNA 7E11.

[3] Homology between BPV4 Y and HPV E4.

Homologies calculated using protein alignment function of "Microgenie" computer analysis program (Queen and Korn, 1984).

Figure 40. Alignment of papillomavirus E4 orf's with the X orf of BPV4.

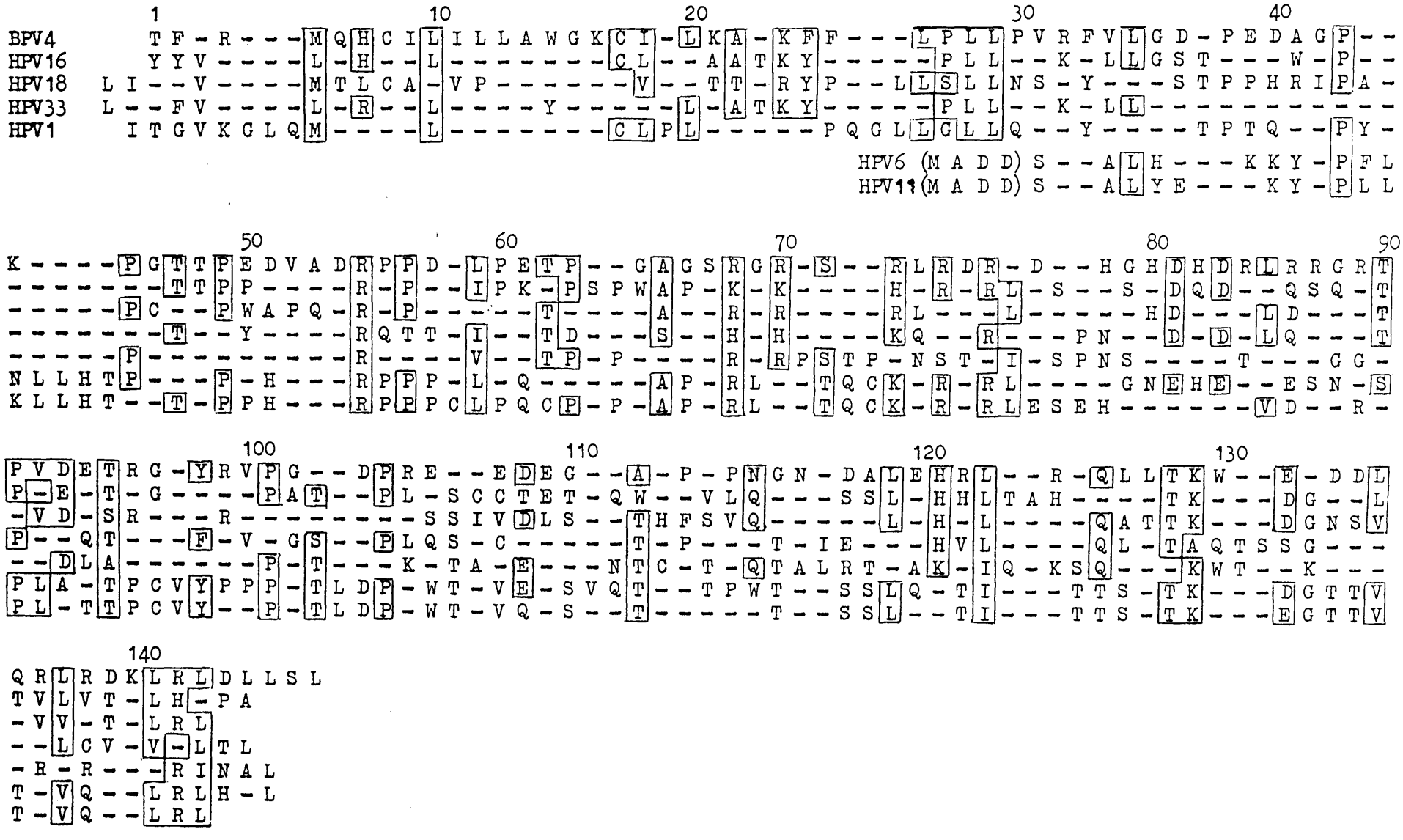
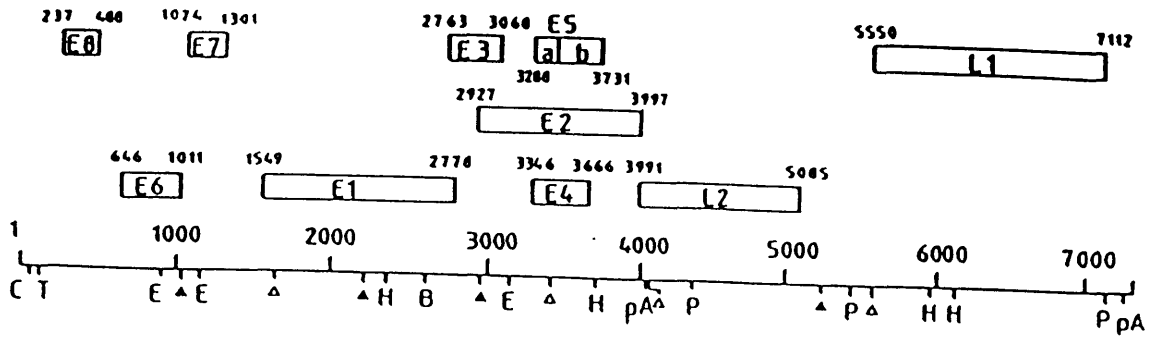
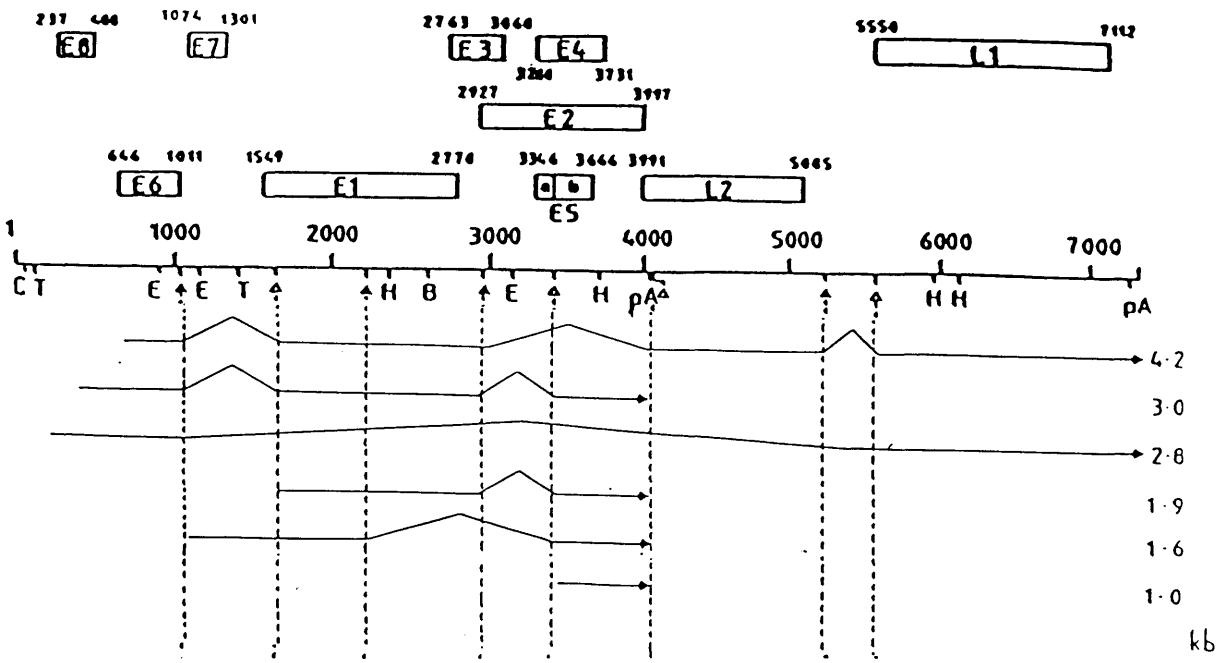


Figure 41. Amendment to the genomic organisation of BPV4.

(a) BPV4 genomic organisation according to Patel et al., 1987.



(b) Amended genomic organisation according to cDNA/RNA sequences.



3.6. Mapping of the transcript(s) represented by cDNA 7E11.

The data presented above on cDNA 7E11 show that it encodes a fusion orf of what will from now on be referred to as the E4 orf of BPV4. The splicing of distal exons of the BPV4 early region represented by this cDNA was at first thought to be a cloning artifact formed by the ligation of two cDNA's. However, the existence of a BPV4 transcript with the same splicing arrangement has been proven by R-loop mapping (L. Coggins, Beatson Institute), and RNA sequencing of transcripts in the 3' terminal exon of the early region suggested that the cDNA may represent the major 1kb mRNA. The suggestion that this encodes the E4 orf of BPV4 is supported by similarities of the splicing arrangement and coding capacity of cDNA 7E11 to the E4 fusion orf encoding major transcripts of BPV1, HPV11 and probably HPV's 6b and 1a.
It was therefore surprising that this transcript had not been mapped in the transcriptional analysis of BPV4 (Smith et al., 1986). However, these analyses were carried out by Northern blot hybridisation and S1-nuclease protection of subgenomic restriction fragments chosen without the benefit of genomic sequence information, as a result of which the small EcoRI fragment from nt905 to nt1142, encompassing the 5' exon of cDNA 7E11, was not included in the analyses. The S1 nuclease mapping probe for this region was the HindIII - EcoRI fragment from nt6106 to nt904, which would not have been protected by a short leader exon utilising the splice donor site at nt1016.

It was therefore undertaken to map the 5' end of the transcript represented by cDNA 7E11. This was carried out by Northern blot hybridisation analysis, primer extension and S1 nuclease protection mapping.

3.6.1. Northern blot hybridisation analysis.

Four different DNA fragments were used to map the transcript represented by cDNA 7E11 by hybridisation to Northern blots. 20 µg aliquots of total RNA from AOP 298 oesophageal papillomas were run on a 1% agarose gel containing 2.2 M formaldehyde and transferred onto nylon membrane. The membrane was cut into strips corresponding to the RNA-containing lanes of the gel, and radioactive DNA probes were hybridised to each strip under stringent hybridisation conditions. The results of these experiments are presented and discussed below.

3.6.1.(i). Hybridisation of pG to papilloma RNA.

cDNA G represents nt3745 to nt4034 of the BPV4 genome (Table 3.3) and was therefore used as a probe representing the 3' exon of cDNA 7E11. The same experiment had been carried out previously using a cDNA probe of similar length and sequence (Smith et al., 1986), and identical results were obtained. In the present study, 1 µg of pG DNA was nick-translated and hybridised to one of the Northern blot strips described above. Plasmid DNA was used as it would be nick-translated

more efficiently than the cDNA insert which was only 323 bp long, compared to the average single stranded length of nick-translated DNA of 400 nucleotides (Rigby et al., 1977), but would not interfere with hybridisation under stringent conditions.

An autoradiograph of the hybridised Northern blot is shown in Figure 42. The strongest hybridisation band was in the 1 kb size class and, as expected from the transcriptional map, weak hybridisation was seen to the 1.6, 1.9 and 3.0 kb transcripts. Weak hybridisation could also be detected to the 4.2 kb size class; this also came up in the same experiment performed by Smith et al. (1986). No significant homologies are found between the sequences of the 3' terminal exon of the early region and the exons of the 4.2 kb transcript, so this hybridisation band may represent a second 4.2 kb transcript or an unspliced mRNA intermediate.

3.6.1.(ii). Hybridisation of p7E11 to papilloma DNA.

1 µg of nick-translated p7E11 DNA was hybridised to a Northern blot. An autoradiograph of the blot is shown in Figure 43, lane B. Hybridisation to the 1 kb band was particularly intense and there appeared to be another, 0.8 kb band directly beneath it. In addition, hybridisation occurred to the 1.6, 2.0, 2.7, 3.6 and 4.3 kb bands presumably corresponding to the 1.6, 1.9, 2.8, 3.6 and 4.2 kb transcripts. Hybridisation to the 2.7 and 4.2 kb transcripts would have been through the 3' exon of cDNA 7E11. The 5.4 kb

band could also be detected: this experiment therefore indicated that transcripts represented by the 5.4 and 3.6 kb size classes must contain sequences found in cDNA 7E11. It was also an indication that more than one transcript may be represented by the 1 kb size class.

3.6.1.(iii). Hybridisation of the 5' exon of cDNA 7E11 to papilloma RNA.

The hybridisation pattern of intact cDNA 7E11 to BPV4 transcripts was as expected from the cDNA sequence and viral transcriptional organisation. The hybridisation pattern of the 3' exon was also shown using cDNA G.

In order to establish which transcripts contained sequences from the 5' exon of cDNA 7E11, it was separated from the 3' exon and used to probe a Northern blot.

The strategy used to separate the 5' exon from the cDNA is shown in Figure 44. The restriction enzyme SinI cuts cDNA 7E11 at the position corresponding to nt1007 on the genome, 9 nucleotides upstream of the splice junction. By digesting p7E11 with this enzyme and EcoRI, which cuts at the adjacent vector cloning site, a 73 bp fragment consisting of the 5' exon of cDNA 7E11 could be isolated. Unfortunately, however, two other SinI fragments of 58 bp and 53 bp from the downstream part of cDNA 7E11 were also produced by this digestion. To circumvent ambiguities produced by possible contamination of the 5' exon probe by these fragments, the SinI digest was carried out first and the small fragments

separated on a 4% LMP agarose gel, leaving the 5' exon probe still attached to higher molecular weight plasmid sequences (Fig. 44(a)). The higher molecular weight fragments were then isolated from the gel and digested with EcoRI to separate the cDNA 5' exon. This was isolated from a further 4% LMP agarose gel (Fig. 44(b)).

As the 73 bp fragment was too small to be nick-translated, it was then phosphatased and 5'-^{32}P labelled (2.2.12.(i)). The radioactive probe was separated from free nucleotide using a Pharmacia "Nick-Column" and hybridised to one of the Northern blot strips.

An autoradiograph of the hybridised blot is shown in Figure 43, lane C. The probe hybridised to the 1.0 kb RNA size class, producing an intense band on the autoradiograph. Faint hybridisation bands could also be seen corresponding to the 2.8 and 4.2 kb size classes.

This was a pivotal experiment in the analysis of the transcript represented by cDNA 7E11 as it showed that:

1). the exons represented by cDNA 7E11 were part of the major transcript(s) of BPV4;

2). the 5' exon hybridised to transcripts of only 1 kb; and

3). since no polyadenylation signals occur between the 5' exon of cDNA 7E11 and the early polyadenylation signal at nt4034, this exon is part of a 1 kb transcript containing a splice junction which joins sequences separated by at least 2 kb in the viral genome.

Taken together with the exon mapping experiments of Smith et al., 1986), the results of these experiments led to the

conclusion that cDNA 7E11 represented the 1kb major transcript of BPV4.

A further Northern blot hybridisation experiment was performed to map approximately the 5' terminus of the 1 kb mRNA.

3.6.1.(iv). Hybridisation of pN to papilloma RNA.

The 3' exon represented by cDNA 7E11 is 659 nucleotides in length. With the addition of a poly(A) tail of up to 200 nucleotides, this leaves approximately 150 nucleotides unaccounted for in a 1 kb mRNA. Similarities between cDNA 7E11 and maps of the major transcripts of BPV1 in fibropapillomas (Baker and Howley, 1987), HPV11 in condylomata (Nasseri et al., 1987) and HPV's 6b and 1a in benign tumours (Chow et al., 1987a,b), also indicated that the leader exon of the 1 kb transcript of BPV4 would be short.

As a preliminary experiment to map the 5' end of the 1 kb transcript, hybridisation of subgenomic restriction fragments lying upstream of the position of cDNA 7E11 was considered. However, computer analysis of this region revealed few restriction enzyme sites, so this approach was not practical. cDNA N contains sequences from nt633 to nt837 of BPV4; failure of this DNA to hybridise to the 1 kb band in a Northern blot would therefore indicate that the 5' end of the 1 kb transcript was downstream of nt837.

pN was radiolabelled by nick-translation and hybridised to a

Northern blot strip. The resulting autoradiograph is shown in Figure 43, lane E. A discrete hybridisation band of 1.1 kb appears in this autoradiograph which is of much lower intensity than the 1 kb bands produced by the cDNA 7E11 probes. Hybridisation to other E6-containing transcripts was weak but did not differ from that obtained with the 73 bp 5' exon probe (lane C).

This experiment showed that the 1 kb size class of BPV4 transcripts contained sequences upstream of nt837. Since hybridisation of cDNA N occurred in a discrete autoradiographic band representing transcripts slightly longer than 1 kb, it was concluded that only a proportion of a heterogeneous 1 kb size class of transcripts contained sequences upstream of nt837. The low intensity of the hybridisation band indicated that this proportion was small.

More precise information on the 5' ends of the 1 kb transcripts was sought using the more sensitive mapping techniques of primer extension and S1 nuclease protection analysis.

3.6.2. Primer extension of probes derived from cDNA 7E11.

The distance of the 5' terminus of a RNA transcript from a specific nucleotide position on the transcript can be determined by reverse transcribing the RNA from a radiolabelled specific antisense primer and running the resulting single stranded cDNA on a sequencing gel. The position of the RNA 5' terminus may then be calculated by

subtracting the size of the extension product from the nucleotide position of the 5' end of the primer.
The 5' end of the 1 kb mRNA represented by cDNA 7E11 was mapped using two such primers.

3.6.2.(i). Primer extension using the 168 bp BbvI fragment of cDNA 7E11.

Mapping of the transcripts of BPV4 poses the obvious problem that with the differential splicing pattern, a single primer may be specific to more than one transcript. Advantage was therefore taken of the splice junction in cDNA 7E11 to produce a primer specific to transcripts containing the same splicing arrangement. The restriction enzyme BbvI cuts BPV4 DNA at nt957 and nt3484. The same sites in cDNA 7E11 are 60 nucleotides 5' and 108 nucleotides 3' to the splice junction, respectively, producing a restriction fragment ideally suited to primer extension mapping of the 1 kb transcript (Fig. 45). In selecting this primer, it was taken into consideration that sequences upstream of the splice junction could also hybridise to other transcripts, e.g the 4.2, 3.0 and 2.8 kb mRNA's, containing the same sequences (Smith et al., 1987), leading to the erroneous assignment of 5' termini. The nucleotide composition of the BbvI primer was such that sequences 3' to the splice junction were G+C-rich, while those 5' to it tended to contain more A+T residues. The preponderance of 3' sequences increased the theoretical optimum annealing temperature of the primer to RNA in 80%

formamide (Table 3.6), favouring hybridisation of the 3' sequences. This meant that nucleation steps in the annealing process would be more likely to take place in the downstream region, allowing subsequent annealing in the 5' exon, while non-specific annealing of the 5' sequences would be minimised.

p7E11 DNA was digested with BbvI to produce a 168 bp fragment spanning the splice junction. This was separated from the other digestion products on a 2% LMP agarose gel (Fig. 45(b)) and purified. Approximately 0.1 µg of the primer was then phosphatased, 5' ^{32}P labelled and run on a 6% polyacrylamide strand separating gel (Fig. 45(c)). Three bands were seen on autoradiography of the gel, designated primers 1, 2 and 3. A high percentage of reannealing of the two strands normally occurs in strand separating gels, and the ds DNA usually has a higher mobility than the single strands which are presumably irregular in topology due to intrastrand duplex formation and therefore retarded to greater degrees by the gel matrix (Maniatis et al., 1982). The ds DNA then appears as a more intense band below the single strand bands. With the BbvI fragment, however, no such distinction could be made and all three bands were isolated.

An estimated 5 ng of each primer was annealed to AOP A poly(A)$^+$RNA. Paucity of RNA stocks necessitated a further refinement to the method. Where an abundant supply of RNA is available, primers are annealed to poly(A)$^+$RNA at a range of five to ten different temperatures and the optimum annealing temperature defined empirically for each primer according to

the quality of primer extension products obtained. To carry out such a titration with the three BbvI primers would have expended a considerable amount of poly(A)$^+$RNA, so instead the optimum annealing temperature in 80% formamide and 0.4 M NaCl was calculated according to the equations in Section 2.2.6. and 2.2.12. (Marmur and Doty, 1962; Thomas et al., 1976), using the base composition indicated by the cDNA sequence (Table 3.6). Due to the high (63.7%) G+C content of this primer, the theoretical annealing temperature was approximately 54^{o}C. Annealing was carried out at three temperatures to check the accuracy of the calculation: these were 56.5^{o}C, 54^{o}C and 49^{o}C. Control reactions containing the appropriate probe with tRNA were set up.

Primer extension was carried out using the annealed probes and the reaction products fractionated on a sequencing gel next to a sequencing T track of M13 DNA. An autoradiograph of this gel is shown in Figure 46.

The autoradiographic exposure time of 5 days was an indicator of the low abundance of viral RNA relative to cellular mRNA's. Extension products were seen above each primer; however, those of primers 2 and 3 (Fig. 46, lanes 2+ and 3+) may have been non-specific as bands were also seen in the control reactions (lanes 2- and 3-). Those of primer 1 (lanes 1+) extended well beyond the non-specific tRNA extension products (lane 1-). Of the test reactions , best results were obtained with the pre-calculated annealing temperature. This finding was also reported by other workers (N. Lowndes; J. Chester, Beatson Institute). In order to conserve poly(A)$^+$RNA

stocks, all further experiments involving the annealing of a primer to RNA were carried out at the calculated optimum annealing temperature.

The length of the major primer extension products obtained for primer 1 was used to calculate their position on the genomic sequence. These positions are shown in Figure 47. The large number of extension products at first suggested that premature terminations had occurred due to RNA secondary structure. However, it had already been shown through Northern blot hybridisation analysis that more than one transcript existed in the 1 kb size range, so these termination bands were mapped. It was found that the most intense bands, i.e. the majority of primer extension products, mapped between nt869 and nt905 on the BPV4 genome, with the most intense band corresponding to nt870. Bands representing shorter primer extension products were also found in the control lane. Fainter bands mapped to positions between nt845 and nt847 and the longest primer extension product to nt793.

Examination of the genomic sequence in this region revealed no promoter/enhancer sequences of known type. The palindromic sequence CTTGCGAGCAAG around nt870 could have formed a hairpin stucture by intrastrand base pairing in the RNA molecule, preventing readthrough by the reverse transcriptase and hence forcing a premature termination at this point. However, palindromes of this length in the multiple cloning site of pUC13 were read through easily using the same enzyme in the plasmid sequencing method, and no such sequences were

found around the other termini in this experiment. The absence of a TATA box has been demonstrated for transcriptional start sites in SV40 (Brady et al., 1982) and adenovirus type 2 (Baker et al., 1979), in each of which the RNA 5' end is heterogeneous. Moreover, it has been shown that deletion of the TATA box of the SV40 early promoter does not prevent transcriptional initiation but causes it to occur at multiple specific sites, i.e. the TATA sequence is only required for precisely directed transcriptional initiation (Benoist and Chambon, 1981; Mathis and Chambon, 1981).

These primer extension data corroborated the results of Northern blot hybridisation in that a small proportion of reaction products extended beyond nt837, the 3' end of the cDNA N probe. These were predominantly confined to the single band mapping to nt792 at the furthest extent of primer extension.

The transcriptional initiation site at nt654 was not represented by any bands in this autoradiograph or in a longer exposure, affirming the specificity of the primer.

3.6.2.(ii). Primer extension and RNA sequencing using a synthetic oligonucleotide primer.

Although continued experimentation with the BbvI primer may have led to finer mapping of the 1 kb transcript represented by cDNA 7E11, the low yield of primer did not justify the exceptional cost of this enzyme. Instead, a synthetic oligonucleotide was obtained, designated primer 142 (Table

3.6), which consisted of 24 nucleotides centred on the splice junction of cDNA 7E11, from the antisense strand. The use of a synthetic oligonucleotide had the advantage that high labelling efficiency was acheived, reducing autoradiographic exposure times and therefore background.

Primer 142 was 5' ^{32}P labelled and annealed to AOP A poly(A)$^{+}$RNA at a calculated optimum temperature of 45.5^{o}C. A control reaction was included. The annealed primer was then extended with reverse transcriptase, and at the same time a RNA sequencing reaction was set up using the same primer. The two sets of reactions were run side by side on a sequencing gel next to radioactive size markers. An autoradiograph of the gel, at 36 hrs exposure, is shown in Figure 48, and the sequencing ladder has been expanded in Figure 49.

The primer extension products in this autoradiograph map to positions shown on the genomic sequence in Figure 47, in which the proportional intensity of each of the major autoradiographic bands is indicated by the size of the arrow under the corresponding nucleotide position. These results concurred with those of the BbvI primer extension, and there were in addition three faint bands mapping to nt606, nt690 and nt722. This may have been due to lowered specificity of the oligonucleotide primer for the 1 kb mRNA, resulting in priming of cDNA synthesis from other transcripts. The results of all the primer extension experiments described in this subsection will be discussed under a subsequent heading.

The RNA sequence allowed direct determination of the nucleotide positions of the 5' ends of mRNA's and confirmed

that these transcripts were not spliced. The major termination site was shown to be nt870; upstream of this site, the RNA sequence becomes faint and was only readable on a longer autoradiographic exposure, which again showed that no splice junctions occurred for at least another 30 nucleotides upstream. The reaction temperature of 50°C in this protocol was selected to reduce RNA secondary structure without compromising the efficiency of the reaction (Geliebter, 1987). The RNA sequencing experiment therefore provided good evidence that the termination band at nt870 represents a genuine RNA 5' terminus, and the major transcriptional start site of the heterogeneous 1 kb class of BPV4 mRNA's. Other transcriptional start sites identified by RNA sequencing were at nt882, nt886, and nt898-nt901.

3.6.2.(iii). Primer extension using primer 142 at higher reaction temperatures.

Steps were taken which would increase the number of full-length primer extensions if premature terminations were being mapped in the experiments described above. It was considered that the use of higher temperature incubations in the reverse transcription reaction would reduce any RNA secondary structure inhibiting the reaction and allow a greater degree of full-length primer extension.

Radioactive primer 142 was annealed to BPV4 mRNA as before and the reverse transcription reaction initiated at 42°C. A second aliquot of reverse transcriptase was added after 30

min. and the incubation temperature increased to 50°C. After a further 30 min., another aliquot of enzyme was added and the temperature increased to 55°C. The reaction was terminated after another 30 min., and the reaction products fractionated on a sequencing gel. An autoradiograph of this gel is shown in Figure 50, and the sizes of the major bands are shown beneath the figure legend.

The genomic positions of the major termini were no different to those in the previous experiment, and the relative intensities of the bands representing them appeared unchanged. The results of the RNA sequencing experiment indicated that primer extension products mapping 5' to nt870 were produced by reverse transcription through the palindromic sequence around this position; if secondary structure in this region were responsible for the low level of readthrough products, this level would be expected to increase at the higher reaction temperature due to a reduction in secondary structure. Since this was not the case, it was concluded that nt870 represented a genuine mRNA 5' terminus.

3.6.3. S1 nuclease protection mapping of the E6 region of BPV4.

Primer extension experiments have the limitation that they do not map splice junctions and may produce spurious 5' ends due to internal termination of reverse transcription. The presence of splice junctions downstream of nt870 was ruled

out by RNA sequencing, but the existence of the 5' termini mapped in the primer extension experiments required confirmation using an alternative approach.

Protection of single stranded DNA from S1 nuclease digestion is afforded by hybridisation of the DNA to its complementary RNA strand. S1 nuclease protection of 5' end labelled, single stranded antisense DNA from the region mapped by primer extension is a complementary experimental system by which the positions of mRNA 5' termini may be confirmed. Protected fragments are resolved on a sequencing gel and mapped in the same way as primer extension products.

Previous S1 nuclease mapping of the E6 region was carried out using a probe extending upstream from the EcoRI site at nt904 (Smith et al., 1986); the greater proportion of protected fragments representing the major 1 kb transcript would therefore have been <35 nucleotides in length, leading to considerable losses in the precipitation step after S1 nuclease digestion and confusion with non-specific degradation products at the bottom of the sequencing gel.

3.6.3.(i). Isolation of a single stranded antisense DNA probe

A S1 nuclease protection assay for the E6 region of BPV4 was set up. The DNA probe was a SspI - Sau96I restriction fragment from nt592 to nt1007. The 5' end of the antisense strand was at nt1009, only seven nucleotides upstream of the splice donor site utilised in cDNA 7E11, ensuring protection of the labelled 5' end in an RNA hybrid duplex while

maximising the length of the shortest protected fragments. The strategy used to isolate the probe is shown in Figure 51. pBV4 DNA was digested with restriction endonucleases SspI and Sau96I and the 418 bp fragment separated from other digestion products by fractionation on a 2% LMP agarose gel (Fig. 51(b)). This was extracted from the gel and approximately 0.4 µg of the fragment phosphatased and 5' end labelled with ^{32}P. The two strands were separated on a polyacrylamide strand separating gel . The position of the separated strands in the gels was identified by autoradiography (Fig. 51(c)); they were extracted and designated probe 1 and 2.

Probes 1 and 2 were annealed to 5 µg AOP A poly(A)$^+$RNA in 80% formamide, 0.4 M NaCl. The optimum annealing temperature was calculated to be 42^{o}C using the base composition of nt654 to nt1009, representing the sequence from the 5' end of the antisense probe to the previously mapped transcriptional start site at nt654 (Smith et al., 1986). Control reactions containing only probe plus tRNA carrier were also set up.

3.6.3.(ii) S1 nuclease mapping of annealed probes.

Annealed probes were subjected to S1 nuclease digestion and the reaction products fractionated on a sequencing gel. An autoradiograph of the gel is shown in Figure 52.

Protected fragments were found for the reaction containing probe 1 plus mRNA only (Fig. 52, lane 1+). A small proportion of the probe was fully protected, confirming that transcripts initiate at positions upstream of nt592. A small amount of

full length probe appeared in the control reaction in lane 1-, but this band was of much lower intensity than that in lane 1+. Similarly, a faint band representing fully protected probe was seen in lane 2+, but this was thought to be due to contamination with probe 1. No protected bands were apparent in the control lane for probe 2 (not shown).

The sizes of the major protected fragments in lane 1+ are tabulated in the legend to Figure 52. The genomic positions of the 5' ends of these fragments are shown in Figure 53. These positions confirm the primer extension map of this region, except for the terminations at nt606 and nt690.

The most intense bands on the autoradiograph mapped to nt871/872, and within the error limits of the technique, corresponded to the major 5'end at nt870. The S1 assay has the advantage that RNA secondary structure is minimised by annealing the entire length of the transcript to a DNA strand under conditions favouring RNA:DNA interactions. This experiment therefore provided convincing evidence that the major 5' terminus of the 1 kb mRNA represented by cDNA 7E11 is at nt870.

The relative intensities of the other S1 bands also reflect those of the primer extension autoradiographs.

S1 nuclease protection assays sometimes produce spurious bands mapping to A+T-rich regions of the genomic sequence. This is thought to be due to "breathing" of the weaker A-T hydrogen bonding at the reaction temperature, allowing S1 nuclease attack on momentarily single stranded DNA (Pelham, 1982). However, examination of the genomic sequence in the

region of the 5' ends mapped by S1 nuclease protection revealed no A+T-rich regions (Fig. 47) and the precise correlation with primer extension data confirms the heterogeneous nature of the 5' end of the 1 kb major transcript.

3.6.4. Discussion: Mapping of the 1 kb major transcript of BPV4.

Primer extension and S1 nuclease protection data are shown relative to the orf's encoded by the mapped region in Figure 53.

The two exons represented by cDNA 7E11 were mapped by Northern blot hybridisation analysis to the 1 kb major size class of BPV4 transcripts. Two primers specific to the cDNA sequence were used to map the multiple 5' ends of the transcript, and a single stranded probe specific to the region delineated by the primer extension data confirmed these sites by S1 nuclease protection. The exact nucleotide position of some of the 5' termini was found by RNA sequencing. The major transcript of BPV4 therefore has a heterogeneous start site between nt777 and nt901. The number of nucleotides of genomic sequence represented in the 1 kb transcripts is between 775 and 900.

The termini mapped by primer extension at nt606 and by primer extension and S1 nuclease mapping at nt 722 both have TATA homologies 24 and 32 nucleotides upstream, respectively. These may represent other transcripts as they were not mapped

by the BbvI primer. A promoter at nt582 would direct the synthesis of an E6-encoding transcript, and may be represented by cDNA N, of which the 5' end is at nt633. The potential TATA promoter at nt691, directing transcription from nt722, would give rise to an RNA encoding all but the first three codons of the E6 orf from the ATG, and therefore incapable of expressing E6. The splicing pattern of transcript(s) initiating from this minor start site is not known, but the existence of an apparently TATA controlled minor transcriptional start site slightly upstream of a heterogeneous major start site is reminiscent of the arrangement of the late promoter region of BPV1 (Baker and Howley, 1987). Failure to map the RNA start site at nt654 (Smith et al., 1986) by either of the methods was unexpected; even autoradiographic exposures of 5 days did not show any bands of the appropriate size.

The majority of 5' termini mapped by primer extension and S1 nuclease protection analysis fell between nt776 and nt902, the most frequently used initiation site being at nt870. Assuming total specificity of the primer for transcripts spliced from nt1016 to nt 3376, the coding potential of all of these is the same, as the first ATG codon downstream of all of these sites is located in the orf which is fused to E4 by the splice junction in cDNA 7E11. The apparently haphazard distribution of 5' termini is very similar to that of the wart-specific promoter, P_L, of BPV1, which extends from nt7213 to nt7256 of the ncr, overlapped by a tandemly repeated sequence homologous to the late promoter of SV40

(Brady et al., 1984). The SV40 late promoter homology in BPV4 is at nt213 (Table 5; Patel et al., 1987); however, comparison of the late promoter region of BPV1 with the major transcriptional initiation site of BPV4 revealed a homologous motif. Two similar sequences from the BPV4 E6 region and their three homologous motifs in the BPV1 P_{7185} and P_L regions are shown in Figure 47. These sequences have the general form: TTGPyPuNTGPyG. A similar repeated sequence is found in the ncr's of HPV's 16, 18 and 33. The repetition of this motif in the heterogeneous major transcriptional start sites of BPV1 and BPV4, its confinement to this region of the genome and the existence of similar motifs in the transcriptional control regions of other papillomaviruses, suggest that it may be an alternative promoter. Moreover, this may indicate that the multiple start site of BPV4 is the late promoter. The transcriptional organisation of BPV4 is such that the E6 orf is not spliced out of transcripts initiating in the ncr, while the "cassette" system of gene expression through differential splicing common to BPV1 and HPV11 (Baker and Howley, 1987; Nasseri et al., 1987) could operate through the short orf containing an ATG four codons upstream of the universally utilised splice donor site at nt1016.

The fact that the major transcript of BPV4, represented by cDNA 7E11, encodes an E4 fusion orf thought to be associated with the expression of the small E4 proteins, supports the idea that this may be the late promoter, as these proteins are only expressed late in the viral infective cycle (Doorbar

et al., 1987; Breitburd et al., 1987).

The transcriptional initiation site of the 4.2 kb late transcript of BPV4, previously mapped at nt654, was not found in any of the mapping experiments described in this subsection. If the multiple initiation site is in fact the late promoter region of BPV4, the 4.2 kb transcript may actually be transcribed from here.

The discovery of a major multiple initiation site with a secondary TATA-controlled site a short distance upstream suggests that the transcriptional organisation of BPV4 in benign tumours is more similar to that of other papillomaviruses than was previously thought. It parallels the transcriptional organisation of BPV1 late in the viral infective cycle in epithelial cells (Baker and Howley, 1987). The splicing pattern of the major transcript, which allows for the translation of one of the forms of the E4 protein from a fusion orf, is identical to that of BPV1 transcribed from the late promoter, and that of HPV11 (Nasseri et al., 1987), and probably HPV's 1, 6 ~~and 11~~ (Chow et al., 1987a,b). In all of these viruses, the transcript encoding the L1 orf is also transcribed from the major promoter region; although purely speculative at this time, it is possible that the L1 encoding transcript of BPV4 is transcribed from the major promoter.

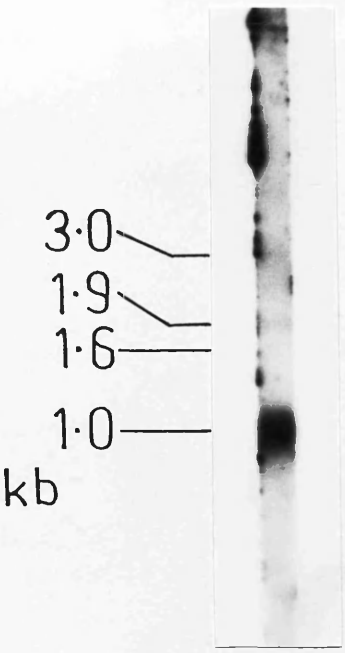


Figure 42. Mapping of the 3' exon of cDNA 7E11 by Northern blot hybridisation.

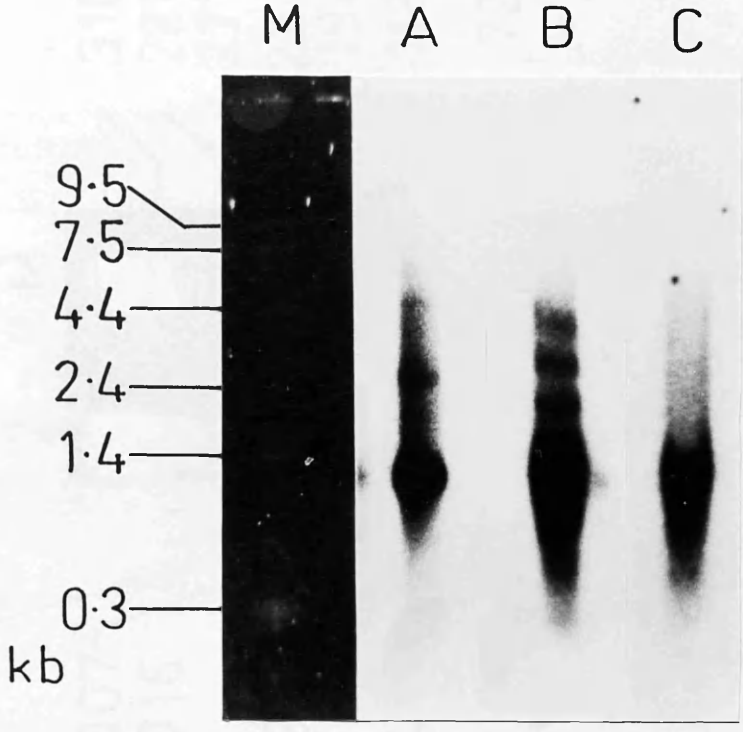
20 µg of total AOP 298oes RNA was electrophoresed on a 2.2M formaldehyde, 1% agarose gel, transferred to nylon membrane and hybridised to radioactively labelled pG DNA. The hybridised blot was autoradiographed for 48 hr.

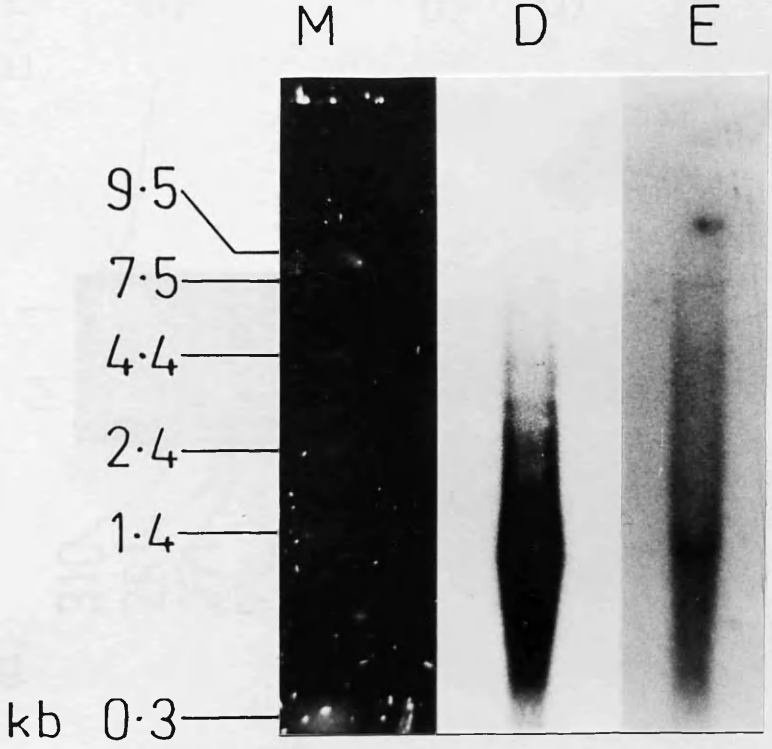
Figure 43. Mapping of the major transcript of BPV4 by Northern blot hybridisation analysis.

Northern blot strips were prepared containing 20 µg of total AOP 298oes RNA (B - E) and one of 1 µg AOP A poly $(A)^+$RNA (A). These were hybridised to the following radioactive DNA probes:

A: BPV4;
B: p7E11;
C: the 5' exon of cDNA 7E11;
D: BPV4 (control);
E: pN.

The hybridised blots were autoradiographed for 48 hr (A,B,D) or 5 days (C,E).
M: RNA size markers (BRL).


M
A
B
C
9·5
7·5
4·4
2·4
1·4
0·3
kb


M
D
E
9·5
7·5
4·4
2·4
1·4
kb
0·3

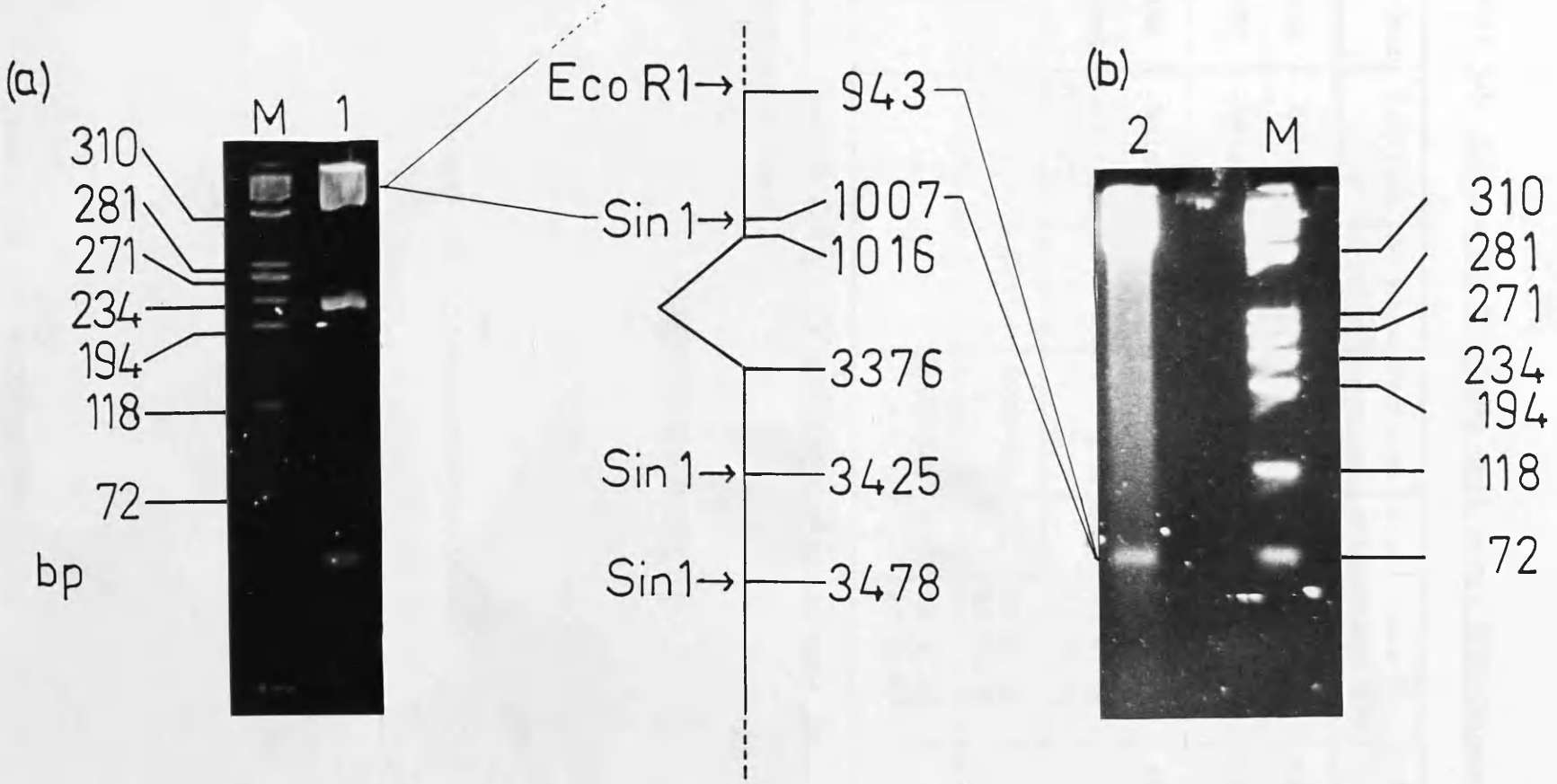


Figure 44. Strategy for isolating the 5' exon of cDNA 7E11.

(a) 20 µg p7E11 DNA was digested with Sin I and electrophoresed on a 4% LMP agarose gel. The higher M_r DNA was extracted.

(b) The extracted DNA was digested with EcoRI to release the 72 bp 5' exon, and fractionated on a further 4% LMP agarose gel. An estimated 100 ng of the 72 bp fragment was recovered.

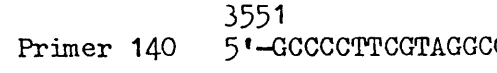
The EcoRI - SinI fragment representing the 5' exon of cDNA 7E11 is shown in the diagram. Solid lines represent cDNA; broken lines vector DNA; diagonal lines, the splice junction.

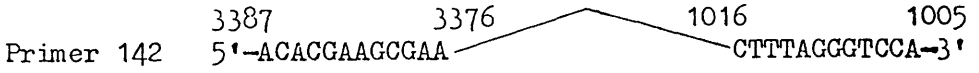
Table 3.6 Primers and S1 probes used to map BPV4 transcripts.

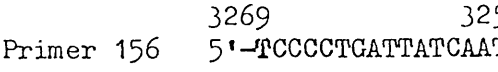
Primer	Position[1]	S1 probe	Position[1]	Size (nt)	Base composition[2]	T_a (RNA seq)	Optimum T_a (80% form.)
140	below			17		51°C	
142	below			24	GC 50% AT 50%	47°C	45.4°C
156	below			17		41°C	36.4°C
BbvI	3484–3376; 1016–957			168	GC 63.7% AT 36.3%		53.75°C
		7E11	1009–592	418	GC 43.9% AT 56.1%		41.7°C
		Q	3244–2842	406	GC 35.2% AT 64.8%		36.4°C

[1] numbers refer to nucleotide positions on the BPV4 genome (antisense strand).
[2] antisense strand.

Oligonucleotide primers.







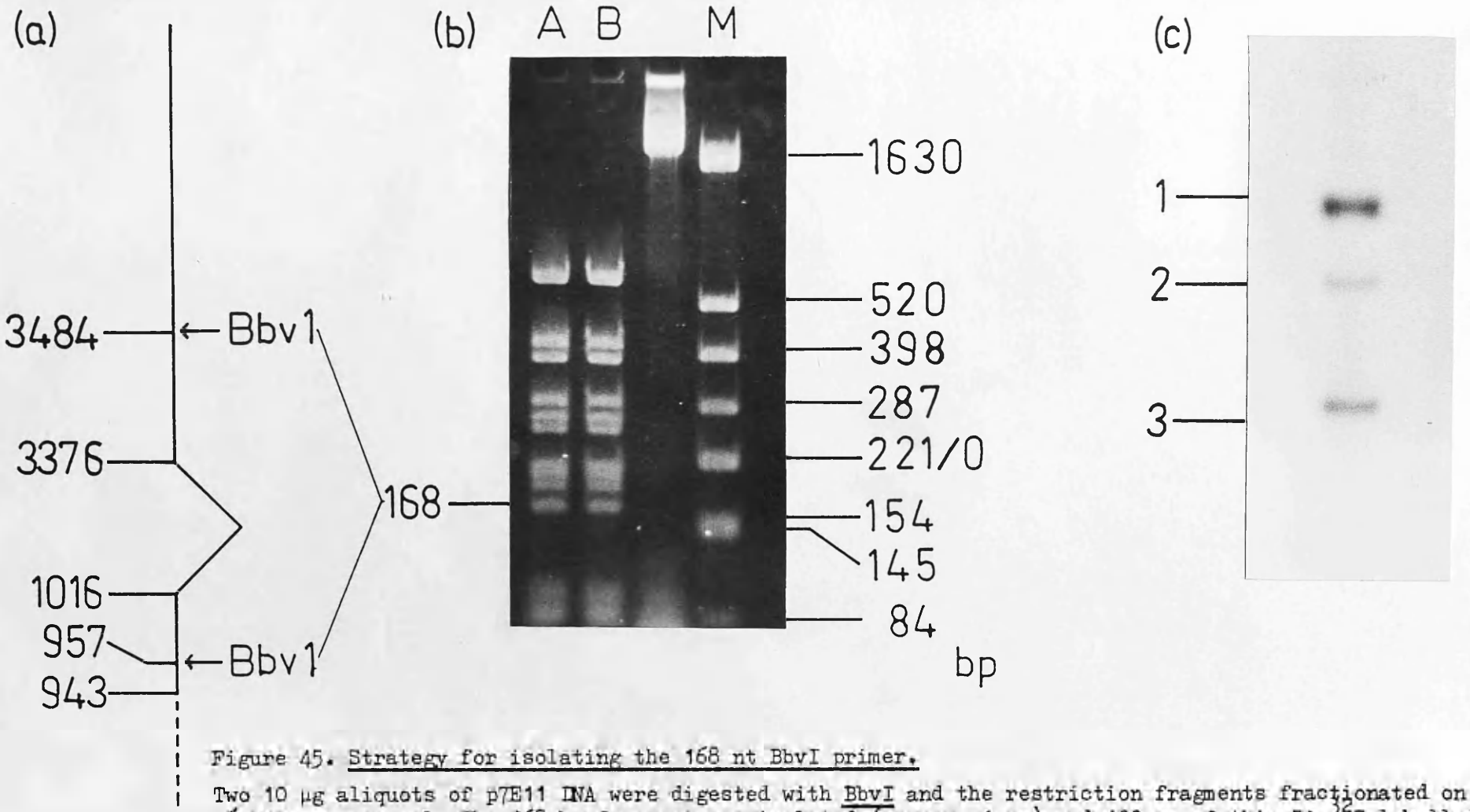


Figure 45. Strategy for isolating the 168 nt BbvI primer.

Two 10 µg aliquots of p7E11 DNA were digested with BbvI and the restriction fragments fractionated on a 2% LMP agarose gel. The 168 bp fragment was isolated (approx. 1 µg) and 100 ng of this 5' ^{32}P labelled and strand separated on a 6% neutral polyacrylamide gel.

(a) Schematic representation of cDNA 7E11, showing the positions of the BbvI sites and the splice junction.
(b) Ethidium bromide stained LMP agarose gel; Lanes A and B - BbvI fragments of cDNA 7E11; M - size markers (HinfI-digested pAT153).
(c) Autoradiograph of strand separating gel; exposure time was 2 min. The bands were designated primers 1, 2 and 3 as shown.

Figure 46. Primer extension using the 168 nt BbvI primer.

Primers 1, 2 and 3 (fig. 45 (c)) were each annealed to 5 µg AOP A poly(A)$^+$RNA at three different temperatures and extended with reverse transcriptase. tRNA controls were included. The reaction products were fractionated on a sequencing gel. An autoradiograph of 5 days' exposure is shown.

Lane M: T-track of M13 DNA;

1- : control reaction, primer 1
2- : control reaction, primer 2
3- : control reaction, primer 3

Lanes 1+ : test reactions, primer 1
2+ : test reactions, primer 2
3+ : test reactions, primer 3

Annealing reactions were performed at:

(a) 49°C
(b) 54°C
(c) 56.5°C

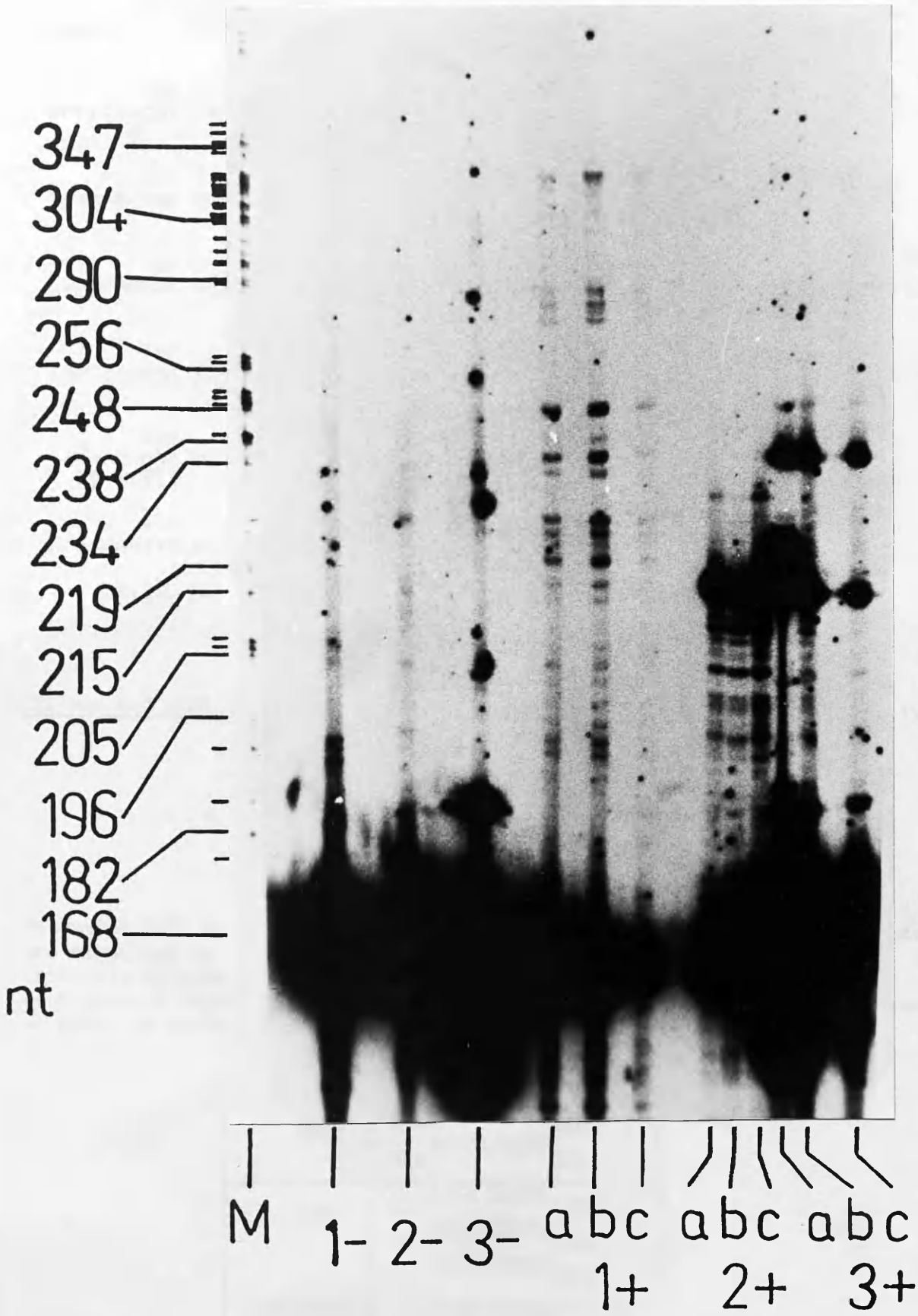
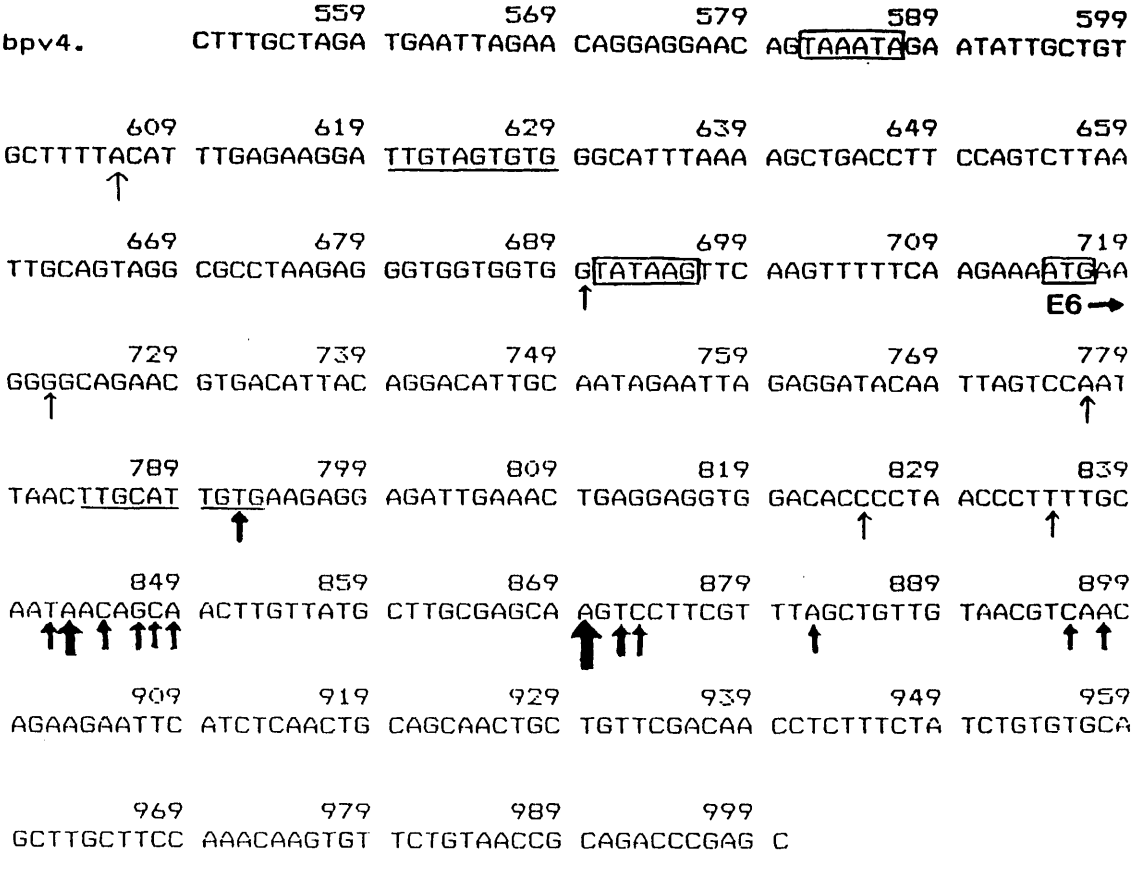
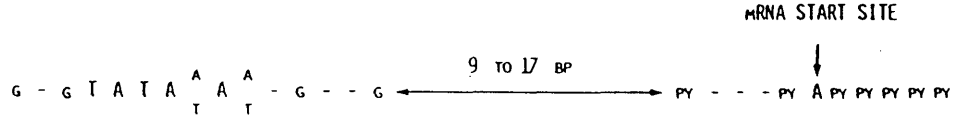

347
304
290
256
248
238
234
219
215
205
196
182
168
nt
M
1-
2-
3-
a b c
1+
a b c
2+
a b c
3+

Figure 47. mRNA 5' ends mapped in the E6 region of BPV4 by primer extension and S1 nuclease protection.



TATA box and mRNA start site consensus sequences (Breathnach and Chambon, 1981)



Potential TATA boxes and the ATG codon of the E6 orf are boxed. RNA 5' ends are signified by vertical arrows: size of arrow is proportional to the intensity of primer extension and S1 nuclease protection bands mapping to each site. A repeated sequence motif, conserved in the late promoter region of BPV1, is underlined.

CONSERVED SEQUENCE MOTIF.

BPV4	$620^{TTGTAGTGTG}_{629}$ $784^{TTGCATTGTG}_{793}$
BPV1	$7177^{ATGT\ CTGTG}_{7185}$ $7233^{TTGCAGTGCG}_{7242}$ $7255^{TTGTGCTGGC}_{7264}$
CONSENSUS	TTGPyPuNTGPyG

Figure 48. 5' ^{32}P-labelled primer 142 was annealed to 5 µg mRNA at the theoretical optimum temperature of 45.5°C. A control reaction was also set up. Primer extension and RNA sequencing using primer 142 were carried out simultaneously and the products fractionated on a sequencing gel. An autoradiograph of 36 hr. exposure is shown.

M : size markers (radiolabelled Sau96I-digested pAT153 DNA).

+ : test reaction, primer extension.

- : control reaction, primer extension

ACGT : sequencing lanes.

The RNA sequence is shown alongside the photograph.

Figure 49. (overleaf) The RNA sequencing autoradiograph is expanded. The RNA sequence is shown alongside the photograph in the region of the multiple start site, and at the overlap with the cDNA 7E11 5' end.
A comparison of this sequence with the BPV4 genomic sequence is shown below the photograph.

Figure 48. Primer extension and RNA sequencing using primer 142.

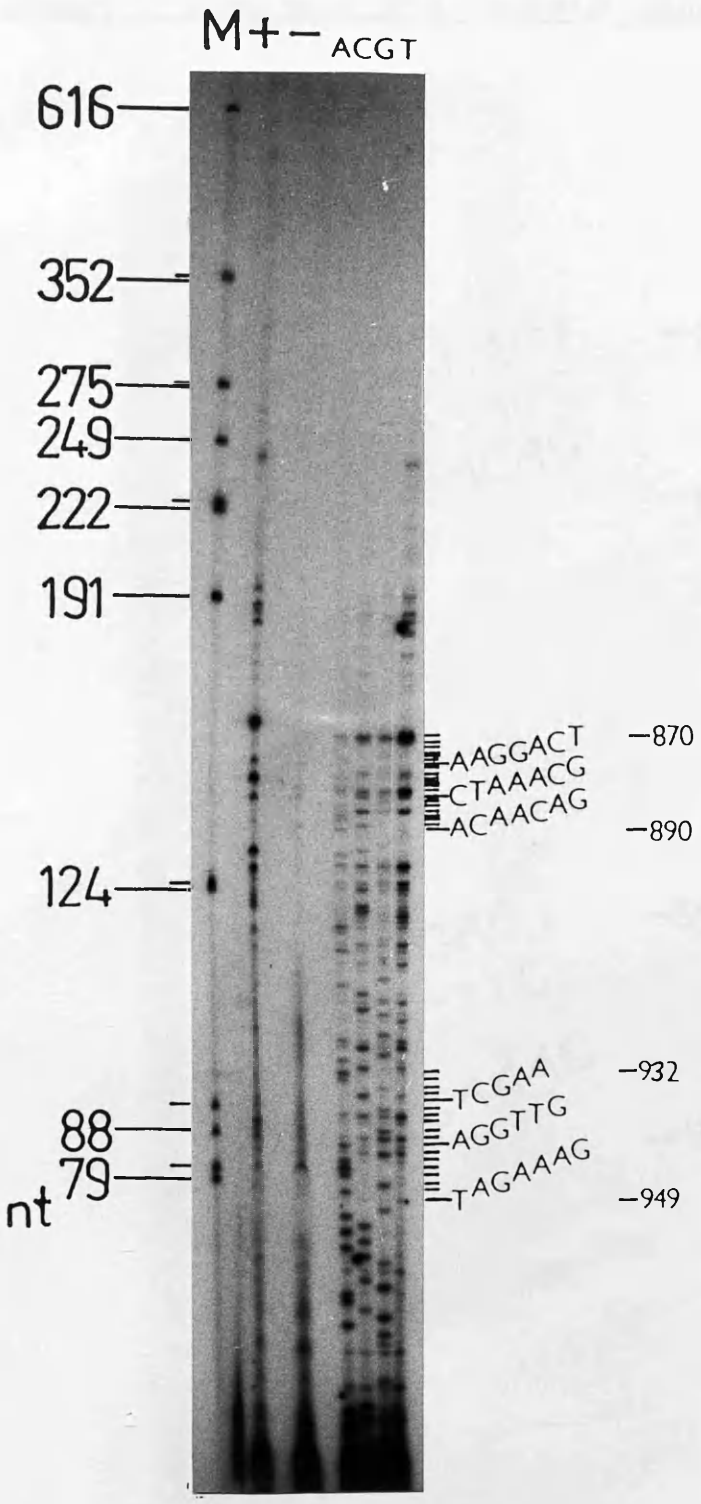
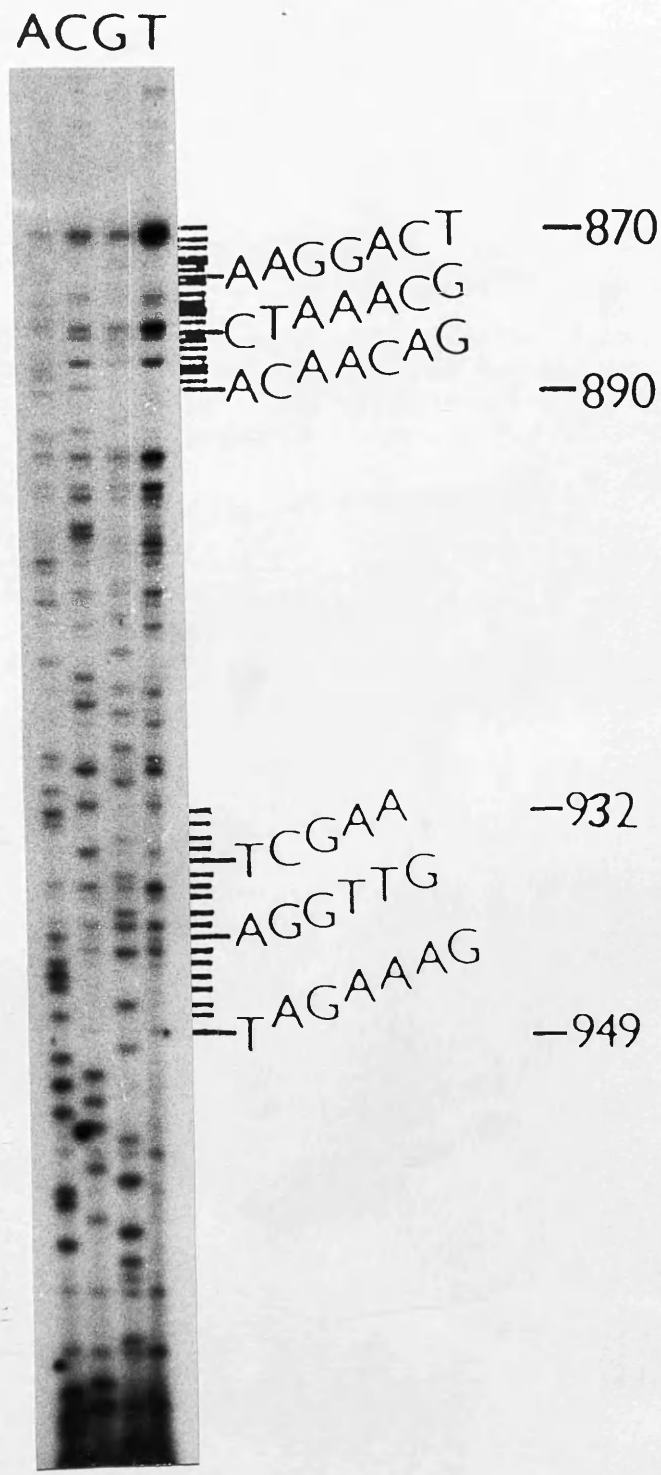


Figure 49. Sequencing of the major BPV4 transcript using primer 142.




```
bpv4   951         941         931         921         911         901         891
       CTGCACACAG  ATAGAAAGAG  GTTGTCGAAC  AGCAGTTGCT  GCAGTTGAGA  TGAATTCTTC  TGTTGACGTT
RNA    --  -TAGAAAGAG  GTTGTCGAAC  AGCAGTTGCT  GCAGTTGAGA  TGAATTCTTC  TGTTGACGTT

       881         871         861
       ACAACAGCTA  AACGAAGGAC  TTGCTCGCAA
       ACAACAGCTA  AACGAAGGAC  T
```

Figure 50. Primer extension using primer 142 at elevated reaction temperatures.

The primer extension reaction described for Figure 48 was carried out at gradually increasing temperatures up to 55°C. The products were fractionated on a sequencing gel and autoradiographed as before.

M : size markers - Sau96I-digested pAT153 DNA plus HhaI-digested pAT153 DNA.

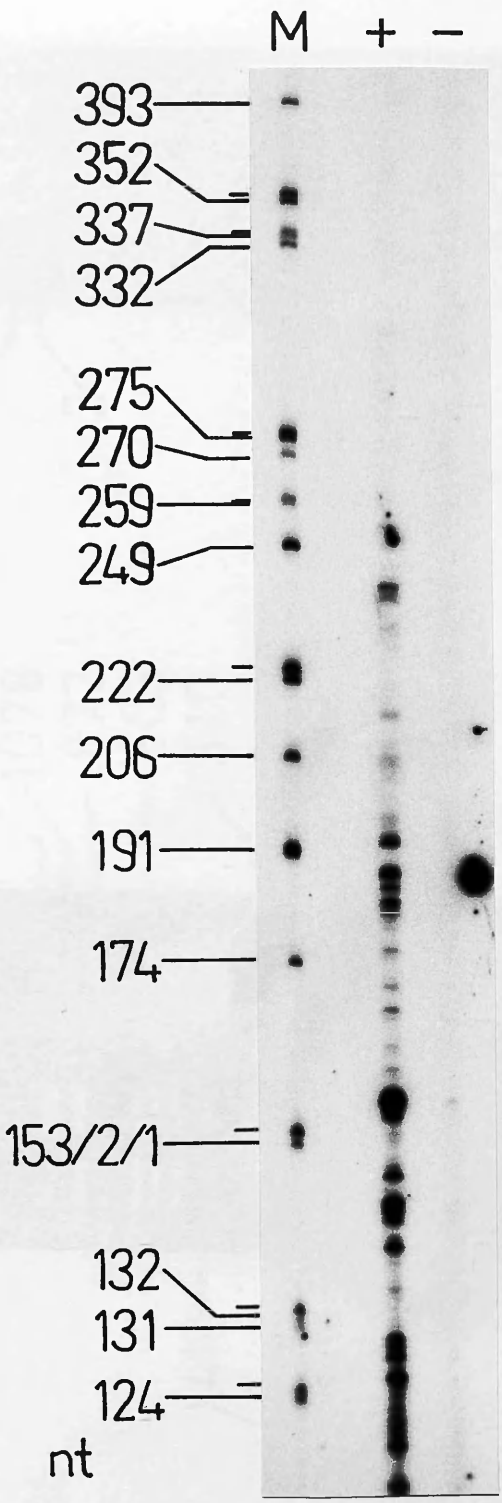
+, - : as Fig. 48.

Sizes of the major bands in this autoradiograph are tabulated below:

STD LEN	DIST	PRED LEN	DEVIATION
616.00	70.000	608.09	7.913
393.00	113.500	402.91	-9.910
352.00	129.500	355.86	-3.857
337.00	135.500	340.58	-3.579
332.00	137.500	335.73	-3.732
275.00	168.500	273.18	1.819
270.00	171.000	268.98	1.016
259.00	179.000	256.23	2.768
249.00	185.500	246.57	2.427
222.00	207.000	218.42	3.579
206.00	219.500	204.30	1.702
191.00	234.500	189.12	1.879
174.00	252.500	173.05	0.952
151.00	282.500	150.39	0.606
132.00	309.000	133.73	-1.733
131.00	312.000	132.01	-1.010
124.00	324.000	125.41	-1.412

UNKNOWN FRAGMENTS:

FRAGMENT	DISTANCE	PREDICTED LENGTH
+1	150.50	306.864
+2	173.50	264.891
+3	182.00	251.700
+4	188.50	242.310
+5	190.00	240.222
+6	191.50	238.163
+7	192.25	237.144
+8	193.50	235.461
+9	212.50	212.026
+10	219.50	204.298
+11	221.00	202.698
+12	232.00	191.529
+13	233.00	190.561
+14	237.00	186.759
+15	238.00	185.826
+16	240.50	183.525
+17	242.50	181.714
+18	243.50	180.819
+19	273.50	156.718
+20	275.00	155.638
+21	276.50	154.568
+22	277.50	153.861
+23	287.00	147.368
+24	291.00	144.749
+25	292.50	143.784
+26	294.00	142.827
+27	298.50	140.011
+28	313.00	131.442
+29	314.50	130.597
+30	316.50	129.482
+31	320.00	127.561
+32	331.00	121.766


M
+
−
393
352
337
332
275
270
259
249
222
206
191
174
153/2/1
132
131
124
nt

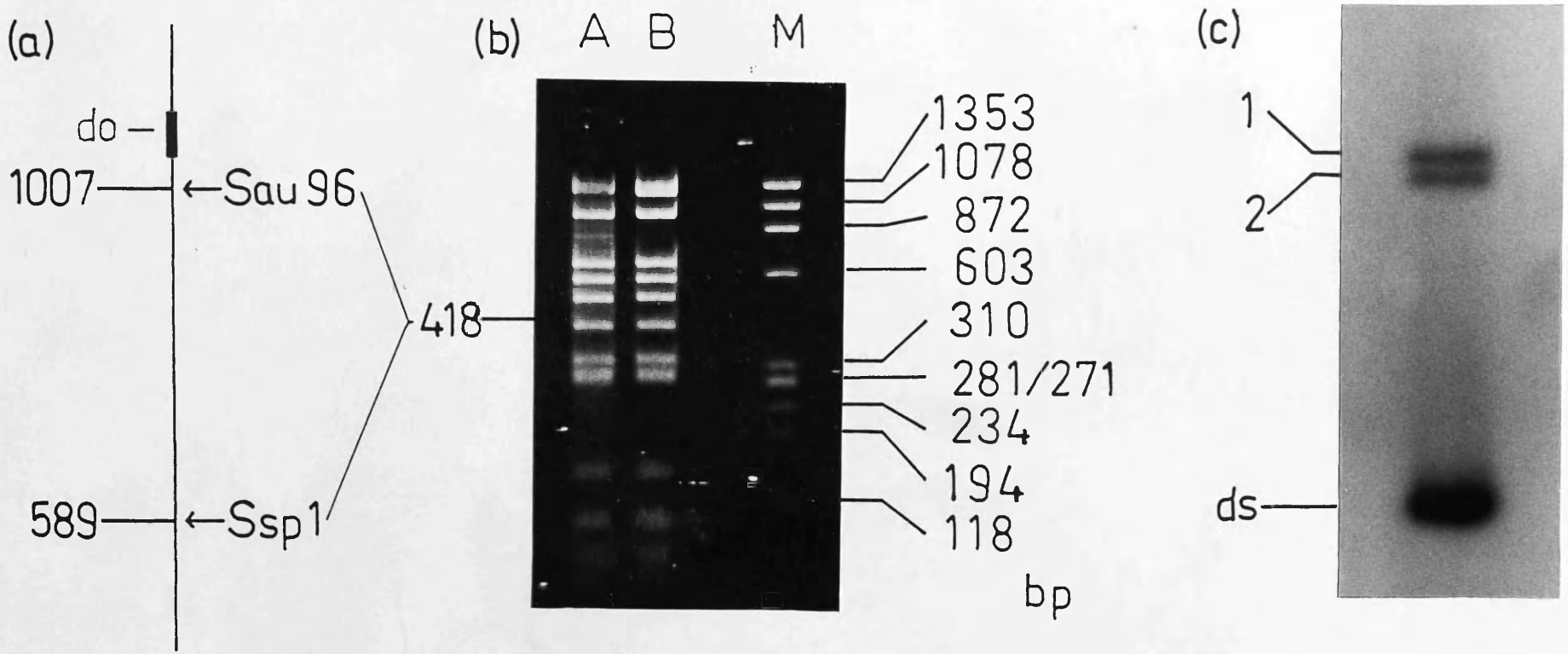


Figure 51. Strategy for isolating S1 probe 7E11.

Two aliquots of 25 µg pBV4 DNA were digested with Sau96I and SspI and fractionated on a 2% LMP agarose gel. The 418 bp fragment was isolated, 5' end labelled and strand separated on a 6% neutral polyacrylamide gel. The separated strands were designated probes 1 and 2

(a) Schematic representation of BPV4 DNA in the E6 region showing the splice donor site at nt1016 (do) and the SspI and Sau96I sites.

(b) Ethidium bromide stained 2% LMP agarose gel showing Sau96I-SspI fragments of pBV4.

(c) 2 min. autoradiographic exposure of 6% polyacrylamide strand separating gel. ds - double stranded probe DNA.

Figure 52. S1 nuclease protection analysis of probe 7E11.

Probes 1 and 2 (Fig 51) were each annealed to 5 µg AOP A poly(A)$^+$RNA at the precalculated temperature of 42°C, with appropriate tRNA controls. The annealed probes were then subjected to S1 nuclease attack at 37°C for 30 min., and the digestion products fractionated on a sequencing gel which was autoradiographed for 48 hr.

M: radioactive size markers (Sau96I-digested pAT153 DNA; HhaI-digested pAT153 DNA

1+: test reaction, probe 1
1−: control reaction, probe 1
2+: test reaction, probe 2
1 : undigested probe 1.

The sizes of the principal protected fragments are tabulated below:

STD LEN	DIST	PRED LEN	DEVIATION
616.00	63.000	610.99	5.013
393.00	96.500	400.22	-7.221
352.00	108.500	354.08	-2.082
337.00	112.500	340.74	-3.735
332.00	114.000	335.96	-3.955
275.00	137.500	273.89	1.111
270.00	140.000	268.44	1.558
259.00	145.500	257.08	1.918
249.00	151.500	245.58	3.421
222.00	169.000	216.47	5.528
206.00	178.500	202.95	3.048
191.00	190.500	187.72	3.280
174.00	195.000	182.47	-8.471
151.00	229.000	149.24	1.760
132.00	250.000	133.08	-1.082
131.00	253.000	130.99	0.014
124.00	263.500	124.01	-0.011
109.00	287.500	109.93	-0.931

UNKNOWN FRAGMENTS:

FRAGMENT	DISTANCE	PREDICTED LENGTH
1+1	132.00	286.554
1+2	160.00	230.687
1+3	169.00	216.472
1+4	193.50	184.194
1+5	203.00	173.691
1+6	210.00	166.535
1+7	216.50	160.290
1+8	220.00	157.074
1+9	242.50	138.542
1+10	245.00	136.686
1+11	258.50	127.264
1+12	260.00	126.276
1+13	261.50	125.298
1+14	282.50	112.673
1+15	285.00	111.290

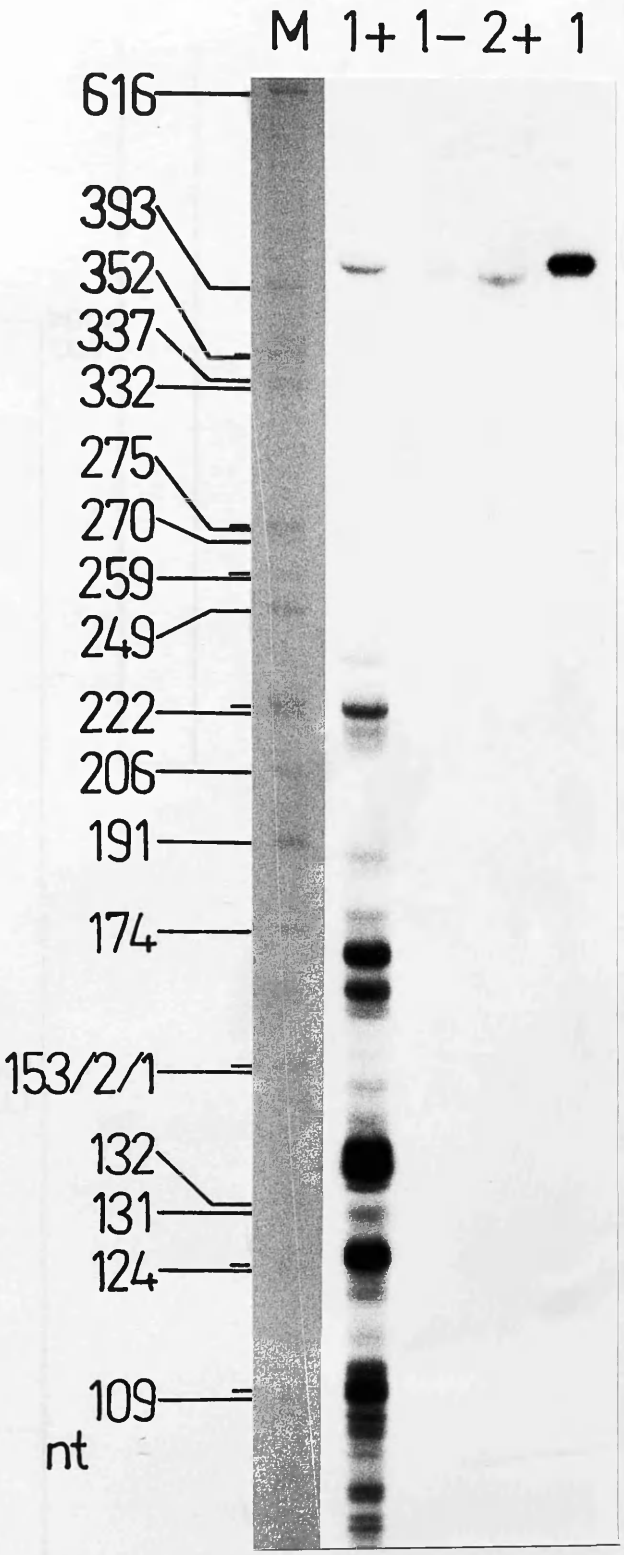
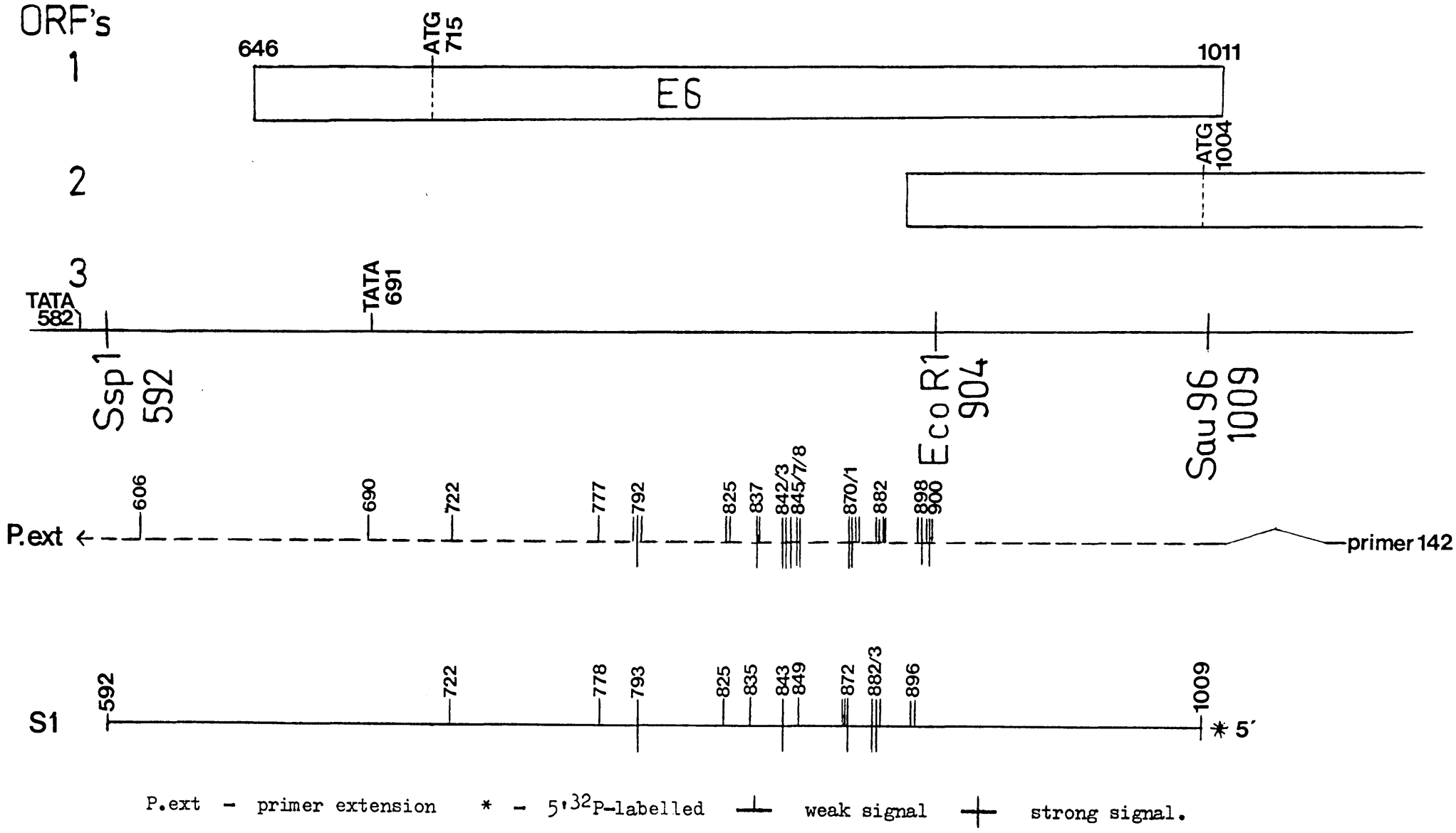

M
1+
1–
2+
1
616
393
352
337
332
275
270
259
249
222
206
191
174
153/2/1
132
131
124
109
nt

Figure 53. Positions of mRNA 5' ends mapped by primer extension and S1 nuclease protection analysis in the E6 region of BPV4.



P.ext - primer extension * - 5'^{32}P-labelled ⊥ weak signal ┼ strong signal.

3.7. Mapping of the transcript(s) represented by cDNA Q.

Sequence analysis of cDNA Q showed that it was a 554 bp truncated cDNA representing nt3235 to nt 3788 of the BPV4 genome. It therefore represented a previously unmapped transcript in which the splice acceptor site at nt3376 is not utilised. The presence in the genomic sequence of a TATA box homology 50 nucleotides upstream of the 5' end indicated that the cDNA might be nearly full length. Although it was truncated at the 3' end, it was assumed that the 3' terminus of the original transcript was at the early polyadenylation site since no splice junctions had been mapped to this part of the early region (Smith et al., 1986). The coding capacity of cDNA Q featured the full E4 orf (Fig. 33) and it was thought possible that the represented transcript could express the larger species of the E4 protein as in HPV1a (Doorbar et al., 1986). However, studies on the BPV1 E2 orf transcriptional enhancer/repressor system suggested that a similar transcriptional arrangement for the expression of the E2 orf might prevail in BPV4. Two minor transcriptional initiation sites exist within the early region of BPV1, one multiple site around nt2440 and one at nt3080 (Fig. 2; Ahola et al., 1987; Stenlund et al., 1985). A small proportion of transcripts initiating at P_{2440} are unspliced and potentially express the full E2 orf (Stenlund et al., 1985); transcripts directed by P_{3080} encode the 3' three quarters of the E2 orf, to which the short-lived E2 repressor protein was mapped (Lambert et al., 1987), and could express the repressor

protein from an ATG codon at nt3091 (Ahola et al., 1987). The first methionine codon of the BPV4 E2 orf is at nt2996, and two closely spaced ATG sequences are found in this orf at nt3049 and nt3158. Transcripts initiating downstream of these points would be translated from the ATG codon of the E4 orf at nt3300, or a spliced transcript could encode a new fusion orf. The multiple translational potential of this region prompted the mapping of the transcript(s) represented by cDNA Q.

3.7.1. Northern blot hybridisation of pQ to polyadenylated AOP A RNA.

Transcripts containing contiguous genomic sequences upstream of the splice acceptor site were mapped by Northern blot hybridisation analysis. pQ DNA, used as a probe as higher specific activity labelling could be acheived, was nick-translated and hybridised to a Northern blot of AOP298oes total RNA. This was prehybridised with unlabelled cDNA 7E11, separated from vector sequences. This would prevent hybridisation of the sequences of the radioactive probe shared by cDNA's 7E11 and Q to the 1 kb major transcript. Only sequences of cDNA Q lying upstream of the splice acceptor site would hybridise to BPV4 RNA.
After prolonged autoradiographic exposure (5 days), a faint hybridisation band of 1 kb was seen. It appeared, therefore, that more than one type of transcript was contained in the 1 kb size class. In an attempt to resolve the different 1 kb

transcripts, 5 µg of AOP A poly(A)$^+$RNA was run on an extended 1.5% agarose, 2.2 M formaldehyde gel, 14 cm in length. The use of polyadenylated RNA represented an effective 6-fold increase in the amount of mRNA in the gel. The hybridisation procedure described above was repeated on a Northern blot of the gel. An autoradiograph of the hybridised blot is shown in Figure 54.

Again, a single hybridisation band appeared, calculated to be 1.0 kb in size. cDNA Q therefore represented a rare transcript of the 1 kb size class. Allowing for a poly(A) tail, the 5' end of the transcript(s), if unspliced, could have mapped to between nt3000 and nt3200. Fine mapping of the transcript(s) was carried out to establish the positions of 5' termini.

3.7.2. RNA sequencing using a synthetic oligonucleotide primer from cDNA Q 5' sequences.

Primer extension and RNA sequencing experiments were carried out to map the 5' ends of the transcripts represented by cDNA Q. A synthetic oligonucleotide of complementary sequence to nt3235 to nt3269 of the BPV4 coding strand was obtained: this was designated primer 156 (Table 3.6). The sequence was chosen to optimise specific hybridisation, with its position well upstream of the splice acceptor site at nt3376; this position was also chosen to be about 100 nucleotides downstream of the expected RNA 5' termini so that it would also be suitable for RNA sequencing.

Primer 156 was 5' ^{32}P labelled and hybridised to AOP A poly(A)$^{+}$RNA in 80% formamide, 0.4 M NaCl at the theoretical optimum annealing temperature of 36.4 ^{o}C. A control reaction was also set up. The primer extension reaction was performed at the same time as an RNA sequencing experiment using the same primer and mRNA. The reaction products were fractionated on a sequencing gel next to radioactive size markers. The gel was autoradiographed for three days, after which only faint bands could be made out in the test lanes. An autoradiographic exposure of six weeks was carried out, following which a faint but readable sequencing ladder could be made out in the lower part of the gel (Fig. 55). Unfortunately, the primer extension lane had been contaminated by overspill from the size marker lane on loading , with the result that any primer extension data was obscured. This experiment was not repeated as there was only enough poly(A)$^{+}$RNA left for the S1 nuclease protection analysis. The sequencing ladder is shown in Figure 55 next to the shorter exposure of the size markers: these were used to identify the positions of bands in the sequencing tracks and thereby to interpolate the sequence where it was difficult to read. The sequence is shown for reference on the left hand side of the autoradiograph in Figure 55.

Relatively intense bands in all four lanes of the sequencing ladder represent primer extension terminations. Two major terminations occurred in this sequence, corresponding to nt3152 and nt3093; the sequence confirmed that no splice junctions were involved. Very faint bands were observed above

the latter terminus and a further, very faint primer extension termination mapping to nt3069, although splicing could not be ruled out in this case. The faintness of these bands may also have been as much due to their position above the sequencing limit of the technique as to the relative abundance of the transcript. The positions of these terminations are shown on the genomic sequence in Figure 56, and relative to the encoded orf's of this region and their ATG codons in Figure 59.

Examination of the genomic sequence in the region of these sites revealed no known promoter/enhancer sequences; however, this is also the case for transcripts of BPV1 initiating at nt2440 and nt3080 in the same region (Ahola et al., 1987), and for the putative late promoter of BPV4 mapped in the previous subsection. The protracted exposure time required for the autoradiograph of this RNA sequence showed that the transcripts represented by cDNA Q are extremely low in abundance in vivo. In BPV1, the abundance of transcripts initiating in the E2 region was at least 100 times lower than those from the late promoter; therefore the unspliced transcripts initiating from P_{2440} were 1000 times less abundant.

The RNA sequencing data on the BPV4 transcripts was complemented by a S1 nuclease protection experiment.

3.7.3. S1 nuclease protection analysis of sequences upstream of cDNA Q.

3.7.3.(i). Isolation of the probe.

On the basis of RNA sequencing data, a suitable subgenomic restriction fragment was chosen for S1 nuclease protection analysis which contained sequences from cDNA Q 5' to the acceptor splice site and extended upstream of the mapped primer extension terminations. A 406 bp HinfI-SspI fragment from nt2838 to nt3244 was separated from pBV4 DNA using the strategy outlined in Figure 57. pBV4 was digested with HinfI and SspI and the fragments fractionated on a 1.5% LMP gel. Unfortunately, a 396 bp HinfI-SspI fragment of the pAT153 vector comigrated with the 406 bp BPV4 fragment. It was decided to separate these out on the strand separating gel. THe upper half of the appropriate double band was excised from the agarose gel to reduce the level of contaminating plasmid DNA (Fig. 57(b)). The DNA was extracted and 1 µg was phosphatased, 5' ^{32}P labelled and run on a strand separating gel. Four single strand bands appeared as two doublets in an autoradiograph of the gel, labelled probes 1 to 4 (Fig. 57(c)). The two ds DNA bands below them showed that less plasmid DNA was present than BPV4 probe; this variation in intensity was also seen in the two single strand doublets. It was therefore assumed that the doublet having lower mobility contained viral DNA, but all of the single stranded probes were isolated from the gel and hybridised to poly(A)$^+$RNA,

with appropriate controls. Probes 1 and 2 were hybridised to the remaining AOP A mRNA, while probes 3 and 4 were hybridised to AOP 298oes poly(A)$^+$RNA, of which an estimated 10 µg had been isolated from total RNA. The base composition of the first 175 nucleotides of the antisense probe (nt3244 to nt3069) was used to calculate the optimum annealing temperature of the probe to RNA's with 5' termini at nt3069. This was 36.4^{o}C, the low temperature reflecting the high A+T content of the probe (64.8%; Table 3.6).

Since inappropriate S1 nuclease digestion could have taken place in the A+T rich regions of the probe, the S1 reaction mixes of each annealed probe was divided into two aliquots and one was digested at 37^{o}C, while the other was digested at the lower incubation temperature of 18^{o}C in an attempt to reduce possible "breathing" of the weaker A-T bonds (Pelham, 1982). Products of S1 nuclease digestion were run out on a sequencing gel next to radioactive size markers. An autoradiograph at 6 weeks' exposure is shown in Figure 58.

No protected fragments were found for probes 3 and 4 (not shown), or for any of the control reactions, which were carried out at 18^{o}C. A faint band representing full length probe was seen in the probe 2 test lane of the 18^{o}C reaction (Fig. 58, lane 2(c)). This was thought to represent contamination by probe 1 DNA: probe 1 was protected along its entire length in both test reactions (lane 1(b) and (c)). The duration of autoradiographic exposure required to obtain this band intensity showed that the represented transcript was rare. Very faint bands of lower molecular weight occurred in

the two test lanes of probe 1. These bands were difficult to reproduce in a photograph but were quite apparent on the autoradiograph. In most cases, these did not coincide between the 37°C and 18°C reactions (lanes 1(b) and (c), respectively), and therefore may have been produced by non-specific S1 nuclease attack. However, bands of identical mobility in both lanes were found which mapped to nt3071, nt3094 and nt3151. These corresponded almost exactly to the RNA sequence termini. The faint band of about 191 nucleotides in lane 2(c) is due to a fold in the dried down gel and does not represent a protected fragment.

The results of RNA sequencing and S1 nuclease protection mapping of the 5' ends of the transcripts represented by cDNA Q are shown together on the genomic sequence in Figure 56, and relative to the E2 orf and its ATG codons in Figure 59.

3.7.4. Discussion: cDNA Q represents novel BPV4 transcripts.

The results of RNA sequencing and S1 nuclease protection analysis correlate at three positions: nt3069, nt3093 and nt3151. The positions mapped by RNA sequencing are used as the map position could be identified using the genomic sequence in the latter two cases. The agreement between the two experiments indicates that these are genuine 5' termini of extremely rare transcripts.

The other sites mapped by S1 nuclease analysis are not confirmed by the RNA sequencing data. The fact that S1 nuclease reactions performed at different temperatures

produced different sizes of protected fragments in this region indicated that these did not represent RNA 5' termini but inappropriately digested DNA:RNA hybrids. Examination of the genomic DNA sequence between nt3093 and nt3151 (Fig. 56) revealed that it is highly A+T rich (62.7%), with an almost unbroken stretch of A and T residues between nt3092 and nt3111. The bands occurring in this region were therefore probably due to digestion of the DNA at the weakly bonded A-U and T-A hybrids. It might also be suggested that the terminus at nt3093 was artifactual for the same reason; however, this appeared also to be a major termination site in the RNA sequencing reaction.

No known promoter/enhancer sites are found in this region of the BPV4 genome; however, transcriptional initiation from nt2440 of BPV1 occurs in the absence of an upstream TATA box which, as has been explained previously, is not required for this process (Benoist and Chambon, 1981; Mathis and Chambon, 1981). This may explain the apparent multiple initiation site found in the present study. The coding potential of transcripts initiating at nt3069 and nt3093 in BPV4 is the same: each is within the E2 orf and upstream of the ATG codon at nt3149, which is in a favourable sequence context for translational initiation (Kozak, 1986). The RNA start site at nt3151 is 7 nucleotides upstream of the ATG codon at nt3158 which is, however, in a weak translational context. It is therefore possible that the latter type of transcript is translated predominantly from the ATG codon of the full E4 orf at nt3300. This may represent the BPV4 homologue of the

larger E4 protein of HPV1a (Doorbar et al., 1987). If its expression parallels that found in HPV1a-infected papillomas, this will be confined to the basal layers of the epithelium, perhaps explaining the low abundance of this transcript in papilloma tissue as a whole.

S1 nuclease analysis showed that the most abundant transcript containing the probe sequences extended 5' to nt2842, the limit of the probe. It might be expected from the Northern hybridisation data that sequences upstream of nt3000 could not be encoded in a transcript of only 1kb; however, the mapping experiments described above do not rule out the possibility of splicing at sites downstream of nt3253. 90% of the BPV1 transcripts initiating at P_{2440} are spliced within the early region (Baker and Howley, 1987). Such a splice junction in a BPV4 transcript could bring about the expression of the E5b orf. Failure to obtain an RNA sequencing ladder at this distance from the primer may have been due to the limitations of the technique: the intensity of RNA sequencing bands falls off sharply around 200 nucleotides upstream of the position of the primer.

Further mapping experiments will have to be carried out to identify possible splice junctions in transcripts initiating in this region.

Thus the transcripts represented by cDNA Q may express a protein corresponding to the E2 repressor of BPV1, and the full length E4 orf. Unmapped transcripts initiating upstream of nt2842 were also discovered by this method, and hybridisation of cDNA Q only to 1 kb transcripts suggests

that these are spliced within the early region.

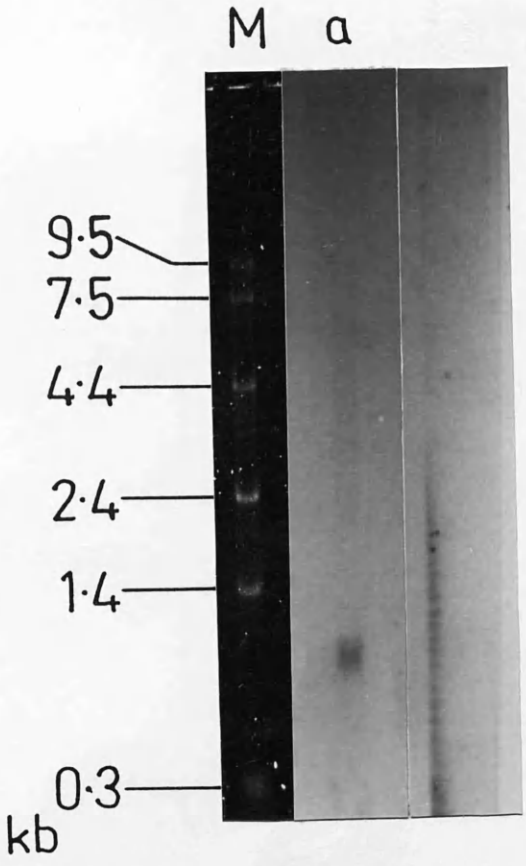


Figure 54. Northern blot hybridisation using cDNA Q as probe.

5 µg of AOP A poly(A)$^+$RNA was run on a 16 cm 2.2 M formaldehyde, 1.5% agarose gel and transferred to nylon membrane. This was prehybridised to unlabelled cDNA 7E11 DNA and then hybridised to radiolabelled pQ DNA.

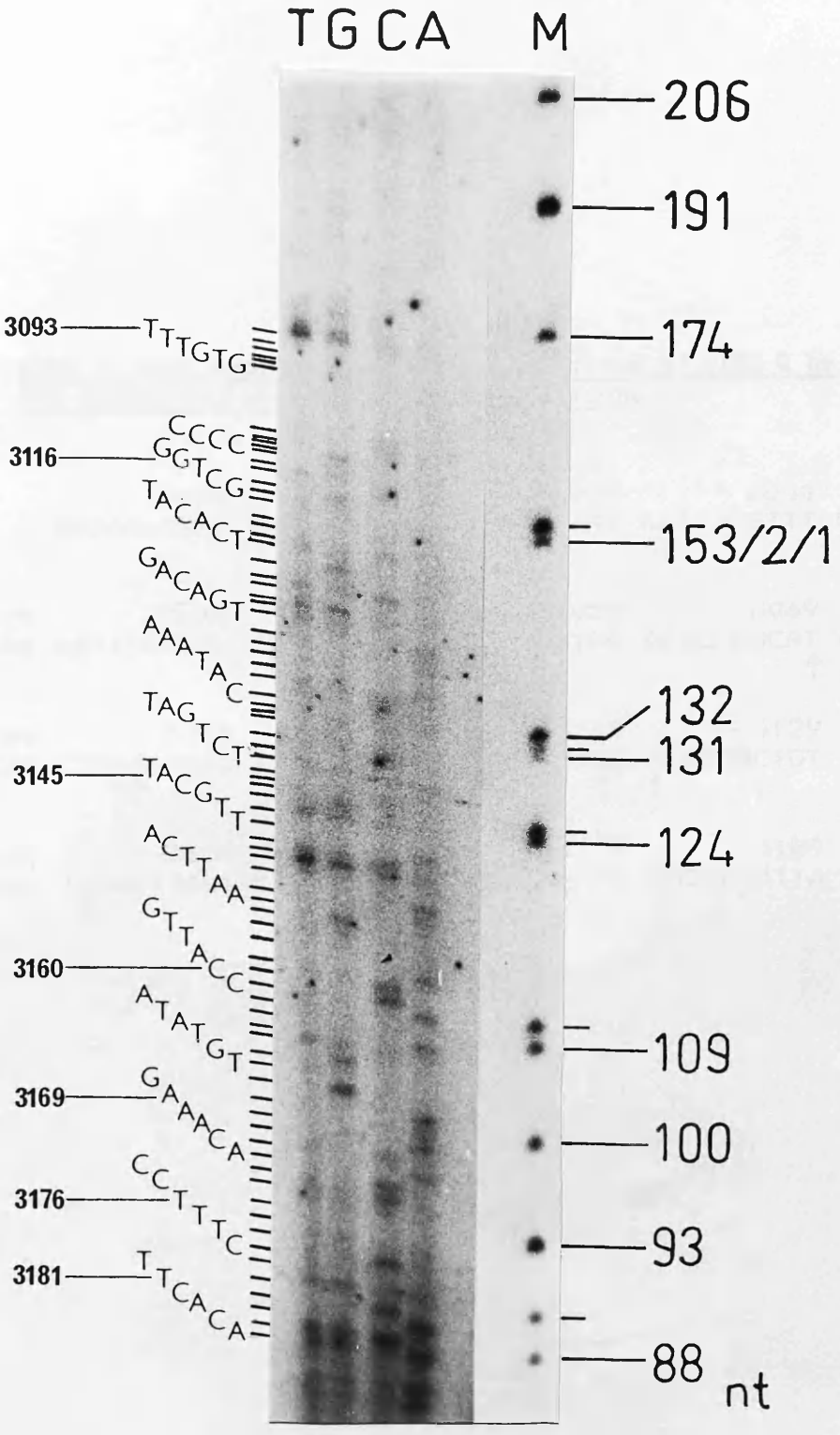
M: RNA size markers (BRL).

(a): AOP A poly(A)$^+$RNA vs. pQ DNA.

Figure 55. RNA sequencing using primer 156.

BPV4 transcripts in AOP A poly $(A)^+$RNA were sequenced using primer 156 (Table 3.6.) The faint sequence in this 6 week autoradiographic exposure was interpolated using the size markers to calculate the positions of bands on the genomic sequence, and is shown here for reference.

M : radioactive size markers (Sau96I-digested pAT153 DNA; HhaI-digested pAT153 DNA).

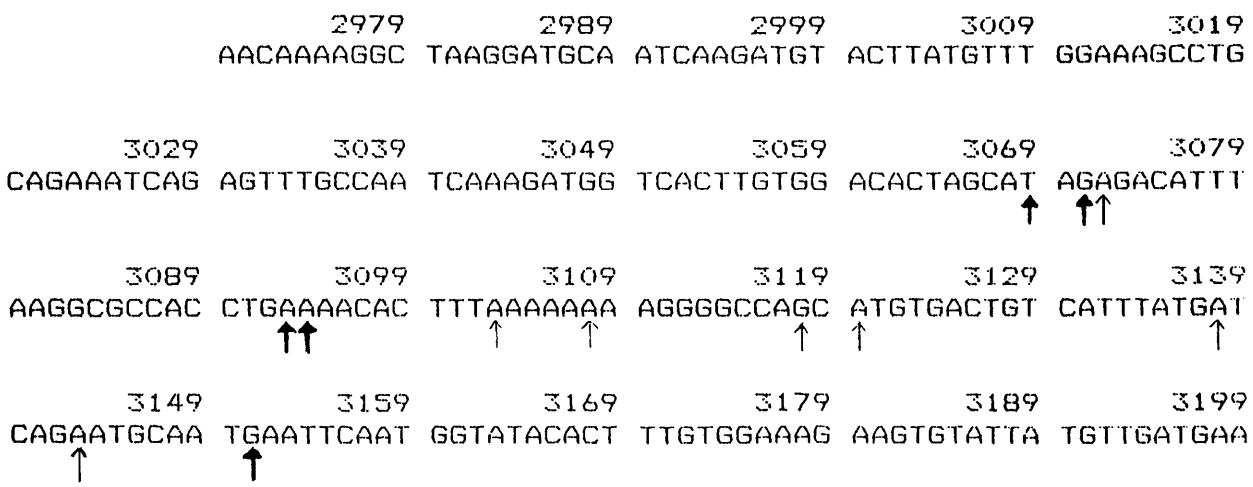


bpv4

```
      3191        3181        3171        3161        3151        3141
TTTCATCAAC ATAATACACT TCTTTCCACA AAGTGTATAC CATTGAATTC ATTGCATTCT

      3131        3121        3111        3101        3091
GATCATAAAT GACAGTCACA TGCTGGCCCC TTTTTTTTAA AGTGTTTTCA G
```

Figure 56. mRNA 5' ends mapped in the region upstream of cDNA Q by RNA sequencing and S1 nuclease protection.



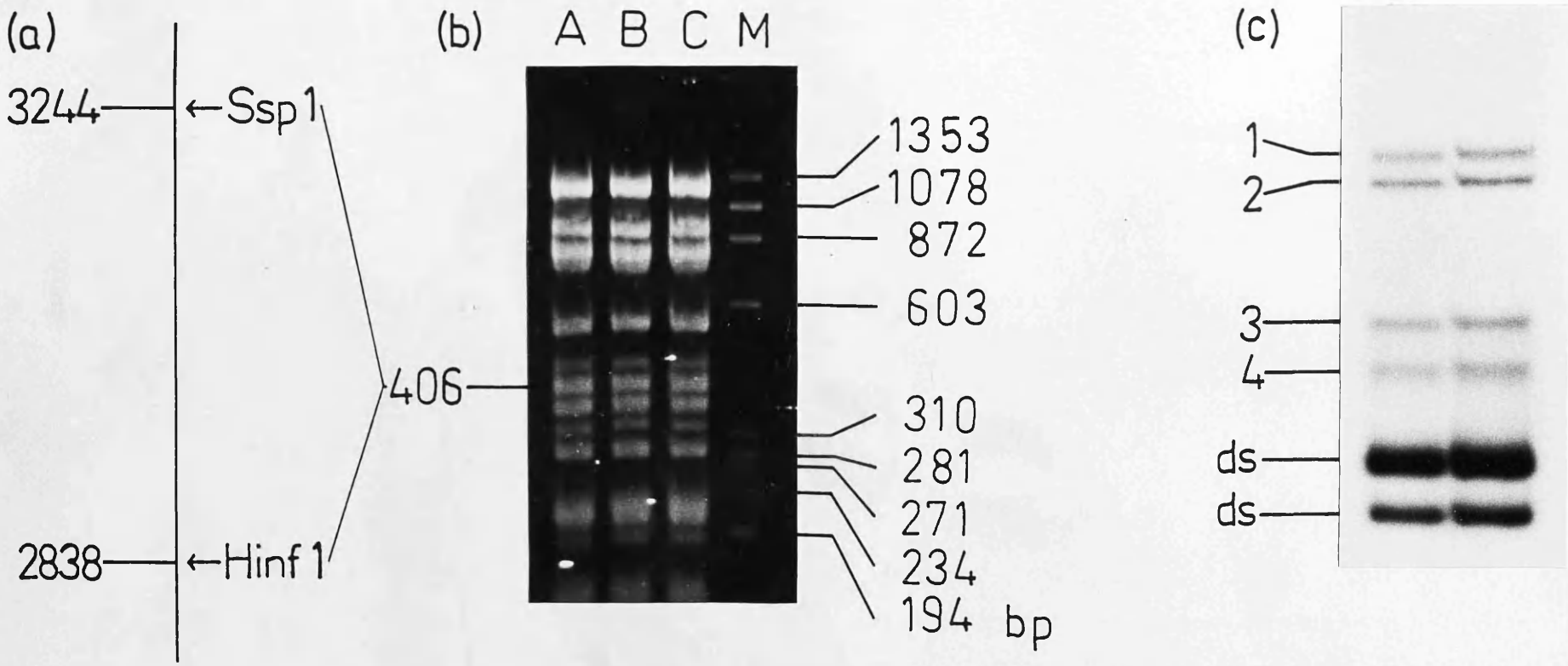


Figure 57. Strategy for isolating S1 probe Q.

Two aliquots of 25 µg pBV4 DNA were digested with HinfI and SspI. The 406 bp fragment of BPV4 was isolated together with a contaminating 396 bp vector fragment from a 2% LMP agarose gel. The extracted DNA was 5' ^{32}P-labelled and strand separated on a 6% neutral polyacrylamide gel. The single strand bands were designated probes 1 to 4.

(a) Schematic representation of BPV4 DNA showing the position of S1 probe Q between the restriction sites HinfI at nt2838 and SspI at nt3244.

(b) Ethidium bromide stained 2% LMP agarose gel showing Hinf I - SspI fragments of pBV4.

(c) 2 min. autoradiographic exposure of strand separating gel. ds - double stranded probe DNA.

Figure 58. S1 nuclease protection analysis of S1 probe Q.

Probes 1 and 2 (Fig. 57) were each annealed to 5 µg AOP A poly(A)$^+$RNA at a precalculated temperature of 36.5°C, with tRNA controls. The test reactions were divided in half: one half was digested with S1 nuclease at 37°C for 30 min., and the other at 18°C for 30 min. Controls were digested at 18°C. The reaction products were fractionated on a sequencing gel, which was autoradiographed for 6 weeks.

Lanes M: radioactive size markers (Sau96I-digested pAT153 DNA; HhaI-digested pAT153 DNA).

1: probe 1

2: probe 2

Lanes (a): undigested probe
(b): test reaction, 37°C
(c): test reaction, 18°C
(d): control reaction.

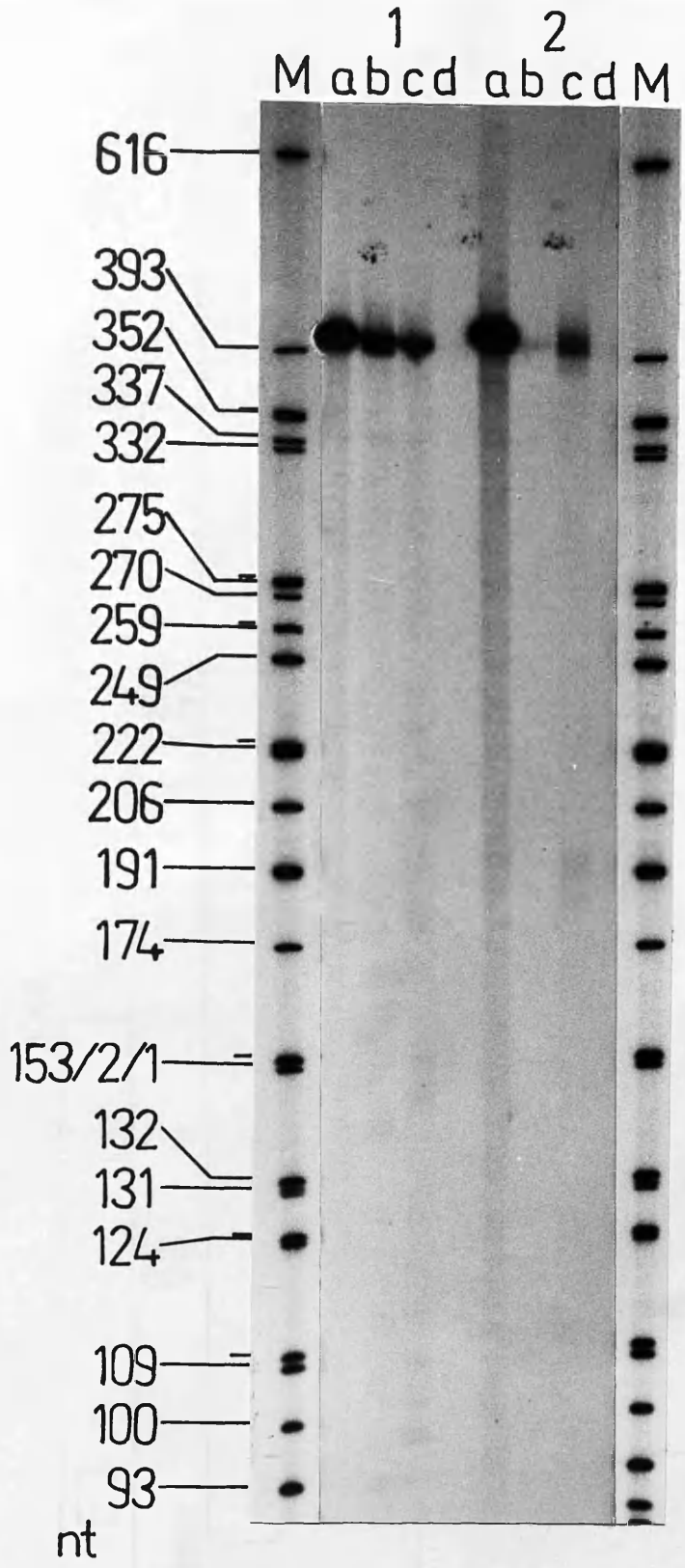

1
2
M a b c d a b c d M
616
393
352
337
332
275
270
259
249
222
206
191
174
153/2/1
132
131
124
109
100
93
nt

Figure 59. mRNA 5' ends mapped by RNA sequencing and S1 nuclease protection in the region upstream of cDNA Q in the BPV4 genome.

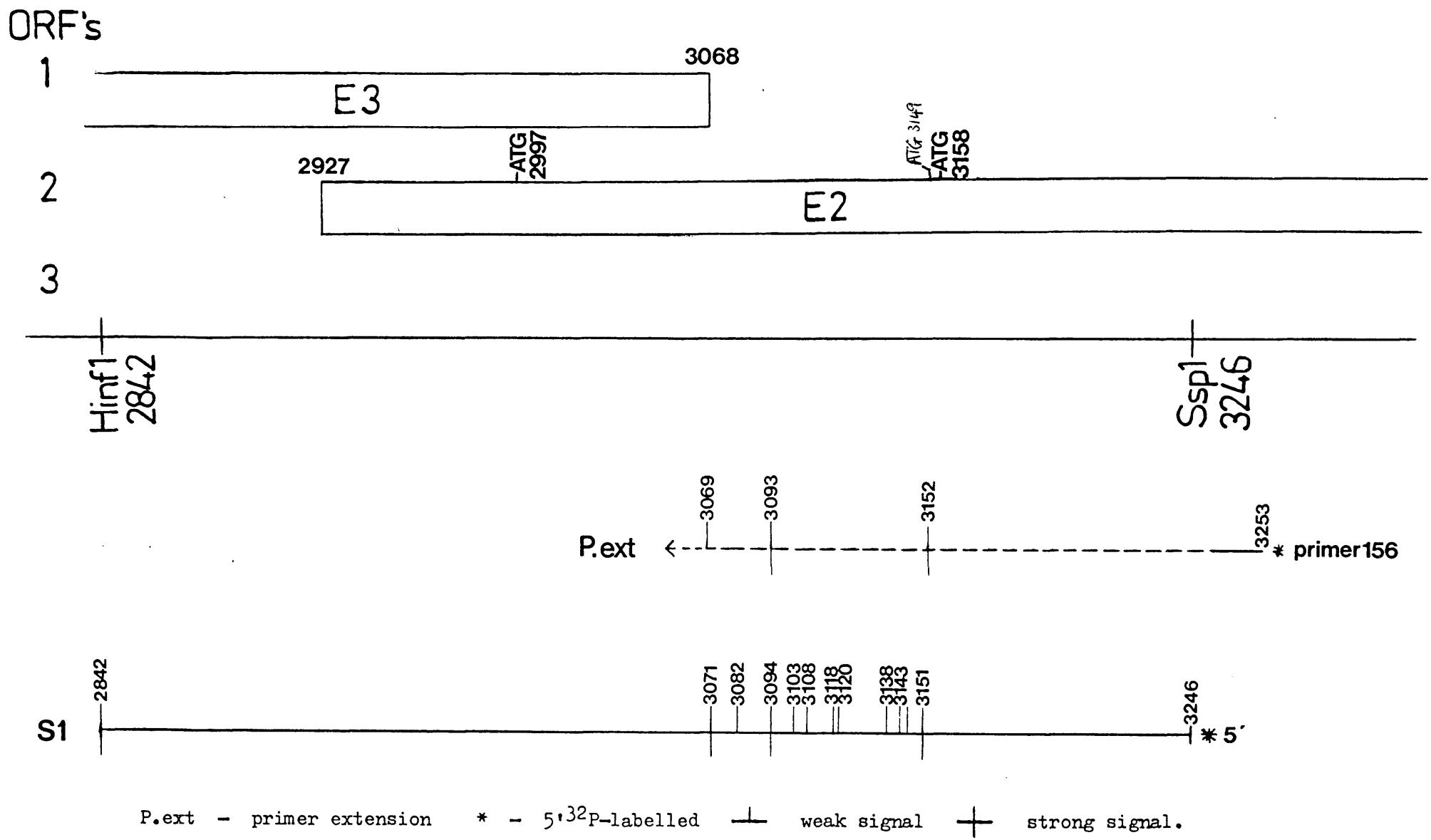


Figure 60.

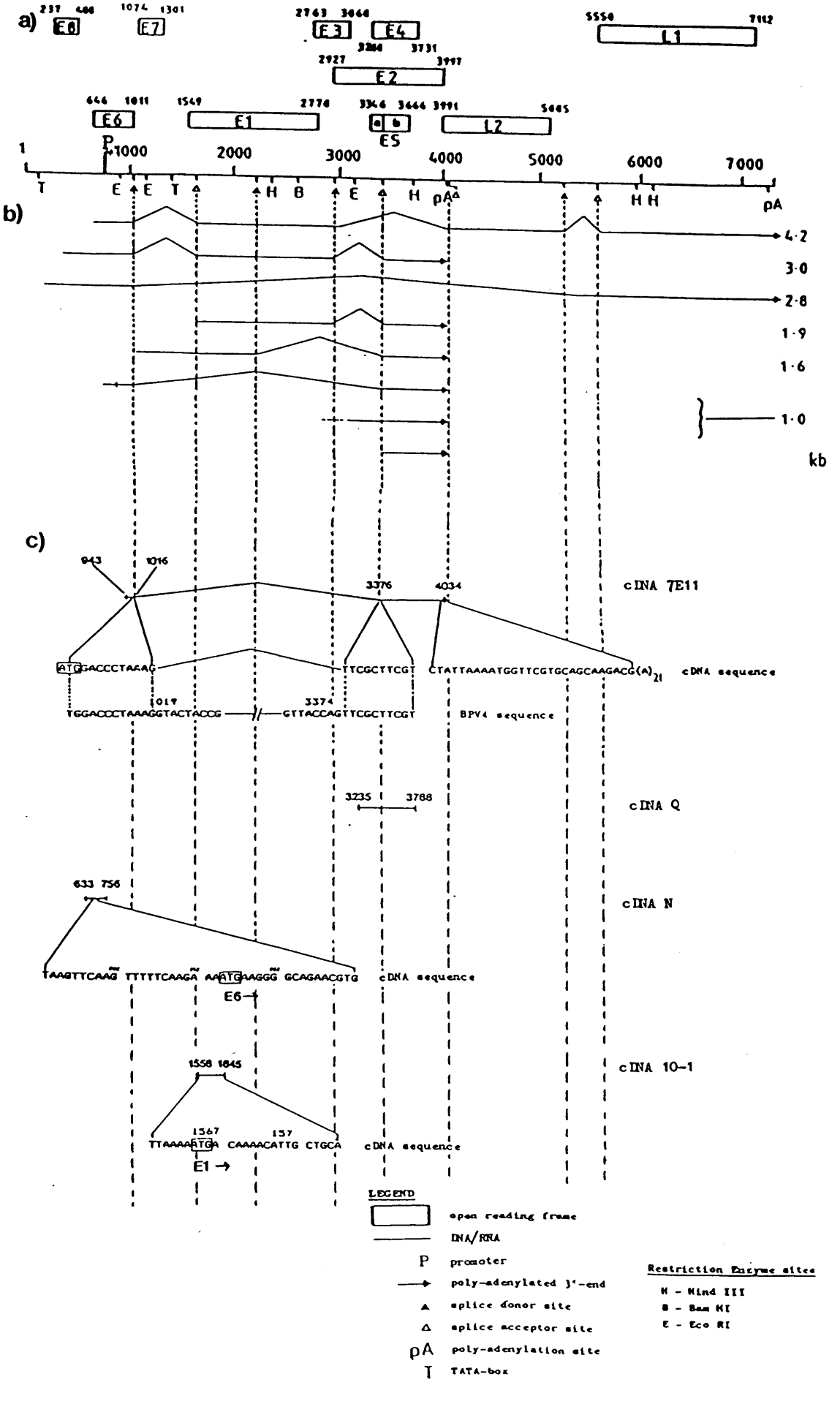
The amended genomic organisation of BPV4 is shown with alterations to the previously described E5a, E5b and E4 orf's (Patel *et al.*, 1987). In addition, the two novel transcripts, both of the 1 kb size class, are included in the transcriptional map, shown here beneath the genome. The principal cDNA's analysed in this project are also shown.

(a) Genomic organisation of BPV4

(b) Transcriptional organisation of BPV4

(c) cDNA's of BPV4 mRNA isolated from papillomas, showing positions of the splice junction of the major transcript and of the early polyadenylation site.

Figure60. Genomic and transcriptional organisation of BPV4.



All numbers represent nucleotide numbers of BPV-4 sequence unless otherwise stated.

3.8. SUMMARY AND CONCLUSIONS.

The initial objective of this project was to study the *in vivo* transcripts of BPV4 through the synthesis and analysis of cDNA's. This objective was fulfilled in the case of two previously unmapped mRNA's. Sequence analysis of cDNA's and of mRNA from papillomas also confirmed the differential utilisation of two consensus RNA splicing signals, mapped the nucleotide position of the early polyadenylation site and brought to light an error in the genomic sequence, the correction of which altered the genomic organisation of BPV4 in the region of the E4 and E5 orf's.

The results of this study will be brought together in this summary and discussed as a whole; finally, suggestions will be made for future investigations.

3.8.1. Amendment to the genomic sequence.

The sequences of two cDNA's from independent cDNA libraries, and of polyadenylated BPV4 RNA isolated from papillomas, differed from the published genomic sequence (Patel *et al.*, 1987) between nt3412 and nt3460. The agreement between the cDNA and RNA sequences argued that they represented the original viral sequence. The changes to the previous genomic sequence involved a frameshift at nt3413, which was reversed at nt3459; the abolition of the translation termination codon between what were previously the E5a and E5b orf's; and the alteration of a methionine in what had been the E2 orf to a

termination codon in the E4 orf. Effectively, the E2 orf remained open in this region while the E5 orf was now continuous and the E4 orf divided by the new termination codon.

3.8.2. The major transcript of BPV4 is homologous to the E4-encoding major transcripts of other papillomaviruses.

cDNA 7E11 was shown by Northern hybridisation analysis to represent the 1 kb major transcript of BPV4. It contained two exons, representing nt943 to nt1016 in the region of the E6 orf, joined to sequences from nt3376 to the early polyadenylation site. Primer extension and S1 nuclease protection analysis showed that this cDNA was near full-length in terms of coding potential. The first translatable orf encoded by this cDNA was a fusion orf of the now continuous E5 orf with four codons in the upstream exon, including an ATG.

The transcript represented by cDNA 7E11 was similar in every respect to the major transcripts of BPV1 (Baker and Howley, 1987), HPV11 (Nasseri et al., 1987), HPV6b and HPV1a (Chow et al., 1987a,b) in benign tumours. Those of BPV1 and HPV11 had been shown by cDNA analysis to encode fusions of the respective E4 orf with small, 2- or 4-codon orf's in the distal 5' exon, and the exon coordinates of the major transcripts of HPV6b and 1a suggested a similar arrangement in thes viruses. It was therefore considered that the continuous E5 orf specified by the BPV4 cDNA's might

represent the E4 homologue of BPV4.

3.8.3. Reassignment of the E4 and E5 orf's of BPV4.

The continuous E5 and the interrupted E4 orf's defined by BPV4 cDNA and RNA sequences were compared to the E4 and E5 orf's of other papillomaviruses by homology and alignment of their putative translation products. Homology was low, as expected for this divergent region of the papillomavirus genome (Danos et al., 1984), but it was found that the continuous E5 orf of BPV4 generally shared greater homology with the E4 orf's of other papillomaviruses than with their E5 orf's, particularly with the mucosa-specific HPV's 16, 18 and 33. The alignment plot also demonstrated isolated regions of conservation of amino acids of potential functional or structural importance.

On the basis of these homology analyses and the similarities of the major transcripts encoding these orf's, it was decided to change the designation of the BPV4 E5 orf to E4.

The interrupted orf, previously designated E4, contains no potential translational initiator codons, and according to the current transcriptional map may not be expressed. These small orf's are, for the present, renamed E5a and E5b, although they exhibit little homology to the E5 orf's of other papillomaviruses.

3.8.4. The functional significance of the E4 orf.

The E4 orf of HPV1 is translated into a number of proteins which are differentially expressed in papillomas (Breitburd et al., 1987). The abundance of these proteins and certain of their properties, such as their acidity and ability to aggregate in intracellular inclusions, and the absence in virus-propagating keratinocytes of cellular keratins normally associated with terminal differentiation, suggested that they were involved in subverting the differential metabolism of keratinocytes to favour viral reproduction. In mucosa, a different regime of keratin expression is associated with keratinocyte differentiation (Broker and Botchan, 1986), so it is perhaps not surprising that greater homology is found between the E4 orf's of mucosa-specific HPV's and the redesignated E4 orf of BPV4, which infects the same type of cell, albeit in a different host species. Relatively high homology also exists with the E4 orf of HPV11, which affects both skin and mucosa, but the highest homology (26%) is with the E4 orf of HPV8, which, although it does not infect mucosal cells, is in common with HPV's 16, 18 and 33, and BPV4, a cancer-associated papillomavirus. Although this is a long-range speculation, these data may correlate carcinogenicity as well as cell type specificity in the papillomaviruses, and since the E4 proteins are thought to be involved at a very early stage in the alteration of cell metabolism (Breitburd et al., 1987), this points to a further avenue of oncogenic research on papillomaviruses.

3.8.5. Mapping of the 1 kb major transcript.

Sequence analysis of cDNA's 7E11, G and J mapped the early polyadenylation site of BPV4 to nt4034, showing that it is homologous to the canonical eukaryotic polyadenylation signal (Fitzgerald and Shenk, 1981).

The 5' end of the major transcript of BPV4 was mapped by primer extension, RNA sequencing and S1 nuclease protection analysis, and found to be at a heterogeneous site between nt776 and nt902, with the major transcriptional initiation site at nt870. No known promoter sites were found near these positions, but a minor RNA start site at nt722 appeared to be controlled by the TATA box at nt691. The multiple transcriptional initiation site, with an upstream, TATA-controlled minor start site is very similar to the arrangement of the late promoters, P_L and P_{7185}, of BPV1 (Baker and Howley, 1987), except that the SV40 late promoter homology of BPV4 is not found in this region. The similarity of the major transcript of BPV4 to that of BPV1, and the discovery of a conserved sequence motif, TTGPyPuNTGPyG, in the heterogeneous RNA start sites of both viruses, suggested that this is the late promoter of BPV4. In addition, failure to observe the 5' end of the 4.2 kb late transcript at nt654 in S1 nuclease analysis of this region suggested that this late transcript may instead start downstream of this position in the heterogeneous transcriptional initiation site.

3.8.6. Mapping of novel transcripts in the E2 region.

cDNA Q represents previously unmapped 1 kb transcripts which do not utilise the splice site at nt3376. Northern blot hybridisation, RNA sequencing and S1 nuclease protection analysis defined minor RNA start sites at nt3069 and nt3093. These extremely rare transcripts potentially express a 5' truncated E2 orf similar to the E2 repressor of BPV1 (Lambert et al., 1987). A further rare transcript initiating at nt3151 may also express the full-length E4 orf, which in HPV1a is thought to encode the larger E4 proteins expressed early in the infective cycle (Doorbar et al., 1986; Breitburd et al., 1987). The S1 nuclease analyses also showed that a minor 1 kb transcript, slightly more abundant than the above mRNA's, initiated upstream of nt2842 and contained sequences represented by cDNA Q up to nt3244. The existence of splice junctions in these transcripts at positions downstream of nt3244 was not precluded by these analyses.

3.8.7. The revised genomic and transcriptional organisation of BPV4.

The main results described in this thesis are presented in the revised genomic and transcriptional map of BPV4 in Figure 60. The E4, E5a and E5b orf's have been redesignated, and their sequence alterations are shown in Figure 38.

The 1 kb transcript size class now has three members: the major, E4-encoding transcript initiating at the heterogeneous

start site in the E6 region; the family of transcripts represented by cDNA Q, initiating both within the E2 orf and upstream of it; and the original 1 kb transcript defined by Smith et al. (1986). However, it has yet to be shown conclusively that the latter is not simply an exon of the major transcript.

It is also possible that the 5' end of the 4.2 kb late transcript is in the heterogeneous transcriptional start site.

In addition, three new promoters have been defined: the heterogeneous start site around nt870, thought to be the late promoter of BPV4 , may be under the control of a conserved papillomavirus promoter sequence; a TATA sequence at nt691 may direct the minor transcriptional start site at nt722; and the rare transcripts represented by cDNA Q may initiate at a multiple start site controlled by a single promoter. These promoters may be analogous to P_L, P_{7185} and P_{3080} in BPV1 (Baker and Howley, 1987). It thus appears that the transcriptional organisation of BPV4 is more similar to those of other papillomaviruses than was previously thought (Smith et al., 1986).

3.9. Future work.

Further progress has been made in the transcriptional analysis of BPV4 by the work carried out in this project, and the continuation of this experimental approach would finalise some of the mapping which was left incomplete. The transcript

extending 5' to nt2842 could be mapped using an S1 probe covering this region, and perhaps sequenced using an oligonucleotide probe. Since cDNA Q hybridises to the 1 kb size class of BPV4 mRNA's, this transcript is probably spliced in the E2 region; this could be checked using a 3' end labelled S1 probe.

The 5' end of the 4.2 kb RNA could be mapped by RNA sequencing using a primer located in the second exon of this transcript. A further cDNA library would have to be synthesised to obtain late transcript cDNA's to confirm their splicing organisation.

Functional studies in cultured cells could be carried out using cDNA's already available. It would be interesting to test the effects of expressing cDNA 7E11 in cultured cells, if possible primary bovine keratinocytes. BPV4 DNA has now been recloned in a pUC vector so that the early region and ncr1 are intact and contiguous. This plasmid could be used for functional studies. In particular, cDNA 10-1 could be expressed under the control of a constitutive promoter and cotransfected with the new BPV4 plasmid to test for a replication modulatory function. cDNA Q could be "completed" as a full-length cDNA by ligation to suitable subgenomic fragments, and used in complementation tests for the E2 repressor function.

Difficulties have been experienced in cloning the E6 orf into bacterial expression vectors. cDNA N could be used instead to produce an E6 peptide for the production of antisera.

REFERENCES

Ahola, H., Stenlund, A., Moreno-Lopez, J., and Pettersson, U. (1987): J. Virol. 61 2240-2244

d'Alessio, J. M., Noon, M. C., Ley, H. L. and Gerard, G. F. (1987): Focus 9 1-4

Amara, S. G., Evans, R. M. and Rosenfeld, M. G. (1984): Mol. & Cell. Biol. 4 2151-2160

Amtmann, E. and Sauer, G. (1982a): Nature 296 675-677

Amtmann, E. and Sauer, G. (1982b): J. Virol. 43 59-66

Androphy, E. J., Schiller, J. T. and Lowy, D. (1985): Science 230 442-445

Androphy, E. J., Lowy, D. R. and Schiller, J. T. (1987): Nature 325 70-73

Aviv, H., and Leder, P. (1972): PNAS 69: 1408-1412

Baker, C. C., Herisse, T., Courtois, G., Galibert, F. and Ziff, E. (1979): Cell 18 569-580

Baker, C. C. and Howley, P. M., (1987): EMBO J. 6 1027-1035

Baker, C. C., Phelps, W. C., Lindgren, V., Braun, M. J., Gonda, M. A. and Howley, P. M. (1987): J. Virol. 61 962-971

Beaudenon, S., Kremsdorf, D., Croissant, O., Jablonska, S., Wain-Hobson, S. and Orth, G. (1986): Nature 321 246-249

Benoist, C. and Chambon, P. (1981): Nature 290 304-310

Berg, J. M. (1986): Science 232 485-487

Berg, L. J., Singh, K. and Botchan, M. R. (1986a): Mol. & Cell. Biol. 6 859-869

Berg, L. J., Lusky, M., Stenlund, A. and Botchan, M. R. (1986b): Cell 46 753-762

Berk, A. J. and Sharp, P. A. (1977): Cell 12 721-732

Birnboim, H. C. and Doly, J. (1979): NAR 7 1513-1522

Birnsteil, M. L., Busslinger, M. and Strub, K. (1985): Cell 41 349-359

Boyle, J., Briggs, J. D., Mackie, R. M. (1984): Lancet (i) 702-705

Brady. J., Radonovich, M., Thoren, M., Das, G. and Salzman, N.P. (1984): Mol. & Cell. Biol. 4 133-141

Brady, J. and Khoury, G. (1985): Mol. & Cell. Biol. 5 1391-1399

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978): PNAS 75 4853-4857

Breathnach, R. and Chambon, P. (1981): Ann. Rev. Biochem. 50 349-383

Breitburd, F., Croissant, O. and Orth, G. (1987): in Steinberg, B. M., Brandsma, J. L. and Taichman, L. B. (eds.),Cancer Cells 5/Papillomaviruses, Cold Spring Harbor Laboratory.

Brent, R. and Ptashne, M. (1985): Cell 43 729-736

Broker, T. R. and Botchan, M. R. (1986): in Botchan, M. R., Grodzicker, T. and Sharp, P. A. (eds.), Cancer Cells 4/ DNA Tumor Viruses, Cold Spring Harbor Laboratory.

Burnett, T. S. and Gallimore, P. H. (1983): J. Gen. Virol. 64 1509-1520

Butel, J. S. (1972): JNCI 48 285-299

Campo, M. S., Moar, M. H., Jarrett, W. F. H. and Laird, H. M. (1980): Nature 286 180-181

Campo, M. S., Moar, M. H., Laird, H. M. and Jarrett, W. F. H. (1981): Virology 113 323-335

Campo, M. S. and Spandidos, D. A. (1983): J. Gen. Virol. 64 549-557

Campo, M. S., Spandidos, D. A., Lang, J. and Wilkie, N. M. (1983): Nature 303 77-80

Campo, M. S., Moar, M. H., Sartirana, M. L., Kennedy, I. M. and Jarrett, W. F. H. (1985): EMBO J 4 1819-1825

Campo, M. S., Smith, K. T., Jarrett, W. F. H. and Moar, M. H. (1985): in Howley, P. M. and Broker, T. R. (eds.), Papillomaviruses: Molecular and Clinical Aspects, Alan R. Liss Inc.

Campo, M. S. and Jarrett, W. F. H. (1986): in Evered, D. and Clark, S. (eds.), CIBA Foundation Symposium 120/ Papillomaviruses, John Wiley and Sons.

Campo, M. S. (1987): Cancer Surveys 6 39-54

Campo, M. S. and Jarrett, W. F. H. (1987): in Russell, W. C. and Almond, J. W. (eds.), SGM Symposium 40: Molecular Basis of Virus Disease, Cambridge University Press.

Casey, J. and Davidson, N. (1977): NAR 4 1531-1552

Chen, E., Howley, P. M., Levinson, A. D. and Seeburg, P. H. (1982): Nature 299 529-534

Chen, E. J. and Seeburg, P. H. (1985): DNA 4 165-170

Chaconas, G. and van de Sande, J. H. (1981): Methods in Enzymology 65 75-85

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979): Biochem. 18 5294-5299

Chow, L. T., Hirochika, H., Nasseri, M., Stoler, M. H., Wolinsky, S. M., Chin, M. T., Hirochika, R., Arvan, D. S. and Broker, T. R. (1987a): in Steinberg, B. M., Brandsma, J. L. and Taichman, L. B. (eds.), Cancer Cells 5/Papillomaviruses, Cold Spring Harbor Laboratory.

Chow, L. T., Reilly, S. S., Broker, T. R. and Taichman, L. B. (1987b): J. Virol. 61 1913-1918

Cole, S. T. and Streeck, R. E. (1986): J. Virol. 58 991-995

Cox, M. F., Meanwell, C. A., Maitland, N. J., Blackledge, G., Scully, C. and Jordan, J. A. (1986): Lancet (ii) 157

Croissant, O., Breitburd, F. and Orth, G. (1985): Clin. Dermatol. 3 No.4 43-55

Dagert, M. and Ehrlich, S. D. (1979): Gene 6 23-28

Danos, O., Katinka, M. and Yaniv, M. (1982): EMBO J. 1 231-236

Danos, O., Giri, I., Thierry, F. and Yaniv, M. (1984): J. Inv. Dermatol. 83 (supp.) 7-11

Danos,O., Georges, E., Orth, G. and Yaniv, M. (1985): J. Virol. 53 735-741

Dartmann, K., Schwarz, E., Gissmann, L. and zur Hausen, H. (1986): Virology 151 124-130

Denhardt, D. T. (1966): BBRC 23 641-646

Doorbar, J., Campbell, D., Grand, R. J. A. and Gallimore, P. H. (1986): EMBO J. 5 355-362

Durst, M., Gissmann, L., Ikenburg, H. and zur Hausen, H. (1983): PNAS 80 3812-3815

Dvoretsky, I., Shober, R., Chattophadhyay, S. K. and Lowy, D.R. (1980) : Virology 103 369-375

Early, P., Rogers, J., Davis, K., Calame, K., Bond, M., Wall, R. and Hood, L. (1980): Cell 20 313-319

Engel, L. W., Heilman, C. A. and Howley, P. M. (1983): J. Virol. 47 516-528

Evans, W. C., Patel, M. C. and Koohy, Y. (1982): Proc. R. Soc. Edinb. 81B 29-64

Fitzgerald, M. and Shenk, T. (1981): Cell 24 251-260

Fuchs, P. G., Iftner, T., Weninger, J. and Pfister, H. (1986): J. Virol. 58 626-634

Geleibter, J. (1987): Focus 9 5-8

Georges, E., Croissant, O., Bonneaud, N. and Orth, G. (1984): J. Virol. 51 530-538

Ghosh, P. K., Reddy, V. B., Piatak, M., Lubowitz, P. and Weissman, S. M. (1981): Methods in Enzymology 65 580-595

Gilbert, D. and Cohen, S. (1987): Cell 50 59-68

Giri, I., Danos, O, and Yaniv, M. (1985): PNAS 82 1580-1584

Giri, I. and Danos, O. (1986): TIG 2 227-238

Gissman, L., Wolnick, L., Ikenburg, H., Koldovsky, U., Schnurch, H. G. and zur Hausen, H. (1983): PNAS 80 560-563

Gissmann, L., Boshart, M., Durst, M., Ikenburg, H., Wagner, D. and zur Hausen, H. (1984): J. Inv. Dermatol. 83 (supp.) 26-28

Gissmann, L. (1984): in Cancer Surveys 3 No.1, Oxford University Press

Gissmann, L., Durst, M., Oltersdorf, T. and von Knebel Doeberitz, M. (1987): in Steinberg, B. M., Brandsma, J. L. and Taichman, L. B. (eds.), Cancer Cells 5/Papillomaviruses, Cold Spring Harbor Laboratory.

Groff, D. E. and Lancaster, W. D. (1985): J. Virol. 56 85-8

Grunstein, M. and Hogness, D. S. (1975): PNAS 72 3961-3965

Gubler, U. and Hoffman, B. J. (1983): Gene 25 263-269

Hanahan, D. (1985): in D. Glover (ed.), DNA cloning: a practical approach, Volume 1, IRL Press, Ltd.

Harding, C. R. and Scott, I. R. (1983): J. Mol. Biol. 170 651

Hennighausen, L. and Fleckenstein, B. (1986): EMBO J. 5 1367-1371

Herbomel, P., Bourachot, B. and Yaniv, M. (1984): Cell 39 653-662

Howley, P. M., Schenborn, E. T., Lund, E., Byrne, J. C. and Dahlberg, J. E. (1985): Mol. & Cell. Biol. 5 3310-3315

Imperiale, M. J., Feldman, L. T. and Nevins, J. R. (1983): Cell 35 127-136

Jablonska, S., Dabrowski, J. and Jakubowicz, K. (1972): Cancer Res. 32 583-589

Jarrett, W. F. H., McNeil, P. E., Grimshaw, W. T. R., Selman, I. E. and McIntyre, W. I. M. (1978a): Nature 274 215-217

Jarrett, W. F. H., Murphy, J., O'Neil, B. W. and Laird, H. M. (1978b): Int. J. Cancer 22 323-328

Jarrett, W. F. H., McNeil, P. E., Laird, H. M., O'Neil, B. W., Murphy, J., Campo, M. S. and Moar, M. H. (1980): in Essex, M., Todaro, G. and zur Hausen, H. (eds.), Viruses in Naturally Occurring Cancers, Cold Spring Harbor Laboratory.

Jarrett, W. F. H., Campo, M. S., O'Neil, B. W., Laird, H. M. and Coggins, L. W. (1984): Virology 136 255-264

Jarrett, W. F. H. (1985): in R. G. Klein (ed.), Advances in Viral Oncology 5 83-102, Raven Press

Jones, K., Yamamoto, K. R. and Tjian, R. (1985): Cell 42 559-572

Kadonaga, J. T., Jones, K. and Tjian, R. (1986): TIBS 11 20-23

Kelly, R. B., Cozzarelli, N. R., Deutscher, M. P., Lehman, I. R. and Kornberg, A. (1970): J. Biol. Chem. 245 39-45

Kemp, D. J., Morahan, G., Cowman, A. F. and Harris, A. W. (1983): Nature 301 84-86

Kleiner, E., Dietrich, W. and Pfister, H. (1986): EMBO J. 5 1945-1950

Kozak, M. (1986): Cell 44 283-292

Kreider, J. W. (1980): in Essex, M., Todaro, G. and zur Hausen, H. (eds.), Viruses in Naturally Occurring Cancers, Cold Spring Harbor Laboratory

Kreider, J. W. and Bartlett, G. L. (1981): Advances in Cancer Research 35 81-110, Academic Press

Lacey, M., Alpert, S. and Hanahan, D. (1986): Nature 322 609-612

Lambert, P. F., Spalholz, B. A. and Howley, P.M. (1987): Cell 50 69-78

Lancaster, W. D., and Olson, C. (1982): Microbiol. Rev. 46 191-207

LaPorta, R. F. and Taichman, L. B. (1982): PNAS 79 3393-3397

Law, M-F., Lowy, D. R., Dvoretsky, I. and Howley, P. M. (1981): PNAS 78 2727-2731

Lillie, J. W., Green, M. and Green, M. R. (1986): Cell 46 1043-1051

Lillie, J. W., Loewenstein, P. M., Green, M. R. and Green, M. (1987): Cell 50 1091-1100

Lorincz, A. T., Lancaster, W. D. and Temple, G. F. (1986): J. Virol. 58 225

Lowy, D. R., Dvoretsky, I., Shober, R., Law, M-F., Engel, L. and Howley, P. M. (1980): Nature 287 72-84

Lusky, M. and Botchan, M. R. (1984): Cell 36 391-401

Lusky, M. and Botchan, M. R. (1986): PNAS 83 3609-3613

Lutzner, M. A. (1985): Clinics in Dermatology 3 139-144

MacGregor, J. T. and Jurd, L. (1978): Mutat. Res. 54 297-309

Maitland, N. J., Cox, M. F., Lynas, C., Primer, S., Meanwell, C. A. and Scully, C. (1987): Br. J. Cancer 56 245- 250

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982): Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory.

Marmur, J. and Doty, P. (1962): J. Mol. Biol. 5 109

Mathis, D. and Chambon, P. (1981): Nature 290 310-315

Matlashewski, G., Schneider, J., Banks, L., Jones, N., Murray, A. and Crawford, L. (1987): EMBO J. 6 1741-1746

Meanwell, C. A., Cox, M. F., Blackledge, G. and Maitland, N. J. (1987): Lancet (i) 703

Meneguzzi, G., Binetruy, B., Grisoni, M. and Cuzin, F. (1984): EMBO J. 3 365-371

Moar, M. H., Jarrett, W. F. H. and O'Neil, B. W. (1986): J. Gen. Virol. 67 187-190

Moreno-Lopez, J., Ahola, H., Stenlund, A., Osterhaus, A. and Pettersson, U. (1984): J. Virol. 51 872-875

Nasseri, M. and Wettstein, F. O. (1984): J. Virol. 51 706-712

Nasseri, M., Hirochika, R., Broker, T. R. and Chow, L. T. (1987): Virology 159 433-439

Nevins, J. R. and Darnell, J. E. (1978): Cell 15 1477-1493

Nevins, J. R. (1983): Ann. Rev. Biochem. 52 441-466

Norrander, J., Kemp, T. and Messing, J. (1983): Gene 26 101-106

O'Hare, P. and Hayward, G. S. (1984): J Virol. 52 522-531

Orth, G., Jeanteur, P. and Croissant, O. (1971): PNAS 68 1876-1880

Orth, G., Favre, M., Breitburd, F., Croissant, O., Jablonska,S., Obalek, S., Jarzabek-Chorzelska, M. and Rzesa, G. (1980): in Essex, M., Todaro, G. and zur Hausen, H. (eds.), Viruses in Naturally Occuring Cancers, Cold Spring Harbor Laboratory

Ostrow, R. S., Bender, M., Niimura, M., Seki, T., Kawashima, M., Pass, F. and Faras, A. J. (1982): PNAS 79 1634-1638

Pamukcu, A. M., Yalciner, S., Hatcher, J. F. and Bryan, G. T. (1980): Cancer Res. 40 3468-3472

Patel, K. R., Smith, K. T. and Campo, M. S. (1987): J. gen. Virol. 68 2117-2128

Pater, M. M. and Pater, A. (1985): Virology 145 313-318

Pelham, H. R. B. (1982): Cell 30 517-528

Pfister, H. (1984): Rev. Physiol. Biochem. Pharmacol. 99 112-181

Phelps, W. C. and Howley, P. M. (1987): J. Virol. 61 1630-1638

Pirisi, L., Yasumoto, S., Feller, M., Doniger, J. and DiPaolo, J. A. (1987): J. Virol. 61 1061-1066

Popescu, N., Amsbaugh, S. and DiPaolo, J. A. (1987): J. Virol. 51 1682-1685

Rabson, M. S., Yee, C., Yang, Y-C. and Howley, P. M. (1986): J. Virol. 60 626-634

Radloff, R., Bauer, W. and Vinograd, J. (1967): PNAS 57 1514-1521

Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, P. (1977): J. Mol. Biol. 113 237-251

Roberts, J. M. and Weintraub, H. (1986): Cell 46 741-752

Rous, P., and Kidd, J. G. (1938): J. Exp. Med. 67 399

Rudlinger, R., Smith, I., Bunney, M. and Hunter, J. (1986): Brit. J. Dermatol. 115 681-692

Sanger, F., Nicklen, S. and Coulson, A. R. (1977): PNAS 74 5463-5467

Sarver, N., Rabson, M. S., Yang, Y-C., Byrne, J. C. and Howley, P. M. (1984): J. Virol. 52 377-388

Sassoni-Corsi, P. and Borrelli, E. (1986): TIG 2 215-219

Schiller, J. T., Vass, W. C., Vousden, K. H. and Lowy, D. (1986): J. Virol. 57 1-6

Schlegel, R. and Wade-Glass, M. (1987): in Steinberg, B. M., Brandsma, J. L. and Taichman, L. B. (eds.), Cancer Cells 5/Papillomaviruses, Cold Spring Harbor Laboratory

Schneider-Gadicke, A. and Schwarz, E. (1986): EMBO J. 5 2285-2292

Schneider-Manoury, S., Croissant, O. and Orth, G. (1987): J. Virol. 61 3295-3298

Schwarz, E., Durst, M., Demankowski, C., Lattermann, O., Zech, R., Wolfsperger, E. Suhai, S. and zur Hausen, H. (1983): EMBO J. 2 2341-2348

Schwarz, E., Freese, U. K., Gissmann, L., Mayer, W., Roggenbuck, B., Stremlau, A. and zur Hausen, H. (1985): Nature 314 111-114

Seedorf, K., Krammer, G., Durst, M., Suhai, S. and Rowekamp, W. (1985): Virology 145 181-185

Shirasawa, H., Tomita, Y., Kubota, K., Kasai, T., Sekuja, S., Takamizawa, H. and Simizu, B. (1986): J. Gen Virol. 67 2011-2015

Shope, R. E. and Hurst, E. W. (1933): J. Exp. Med. 58 607-624

Smith, K. T. and Campo, M. S. (1985): Anticancer Res. 5 31-48

Smith, K. T., Patel, K. R. and Campo, M. S. (1986): J. gen. Virol. 67 2381-2393

Smith, K. T., Campo, M. S., Bradley, J., Gaukroger, J. and Jarrett, W. F. H. (1987): in Steinberg, B. M., Brandsma, J. L. and Taichman, L. B. (eds.), Cancer Cells 5/Papillomaviruses, Cold Spring Harbor Laboratory

Smotkin, D. and Wettstein, F. O. (1986): PNAS 83 4680-4684

Southern, E. M. (1975): J. Mol. Biol. 98 503-517

Spalholz, B., Yang, Y-C. and Howley, P. M. (1985): Cell 42 183-191

Spalholz, B., Lambert, P. F., Yee, C. L. and Howley, P. M. (1987): J. Virol. 61 2128-2137

Spandidos, D. A. and Wilkie, N. M. (1983): EMBO J. 2 1193-1199

Stenlund, A., Zabielski, J., Ahola, H., Moreno-Lopez, J. and Pettersson, U. (1985): J. Mol. Biol. 182 541-554

Sundberg, J. P., O'Banion, M. K. and Reichmann, M. E. (1987): in Steinberg, B. M., Brandsma, J. L. and Taichmann, L. B. (eds.), Cancer Cells 5/Papillomaviruses, Cold Spring Harbor Laboratory.

Swift, F., Bhat, K., Younghusband, H. B. and Hamada, H. (1987): EMBO J. 6 1339-1344

Thomas, M., White, R. L. and Davis, R. W. (1976): PNAS 73 2294-2298

Thomas, P. S. (1980): PNAS 77 5201-5205

Tooze,J. (ed.): DNA Tumor Viruses, Cold Spring Harbor Laboratory (1980)

Ushijima, J., Matsukawa, K., Yuasa, A. and Okada, M. (1983): Jpn. J. Vet. Sci. 45 593-602

Vieira, J. and Messing, J. (1982): Gene 19 259-268

Waldeck, W., Rosl, F. and Zentgraf, H. (1984): EMBO J. 3 2173-2178

Watts, S. L., Phelps, W. C., Ostrow, R., Zachow, K. R. and Faras, A. J. (1984): Science 225 634-636

Weiher, H. and Botchan, M. R. (1984): NAR 12 2901-2916

Weislander, L. (1979): Anal. Biochem. 98 305-309

Weiss, B., Jacquemin-Sablon, A., Live, T. R., Fareed, G. C. and Richardson, C. C. (1968): J. Biol. Chem. 243 4543

Wettstein, F. O., Barbosa, M. S. and Nasseri, M. (1987): Virology 159 321-328

Wettstein, F. O. and Stevens, J. G. (1982): PNAS 79 790-794

Wyke, J. A., Stoker, A. W., Searle, S., Spooncer, E., Simmons, P. and Dexter, T. M. (1986): Mol. & Cell. Biol. 6 959-963

Yang, Y-C., Okayama, H. and Howley, P. M. (1985a): PNAS 82 1030-1034

Yang, Y-C., Spalholz, B. A., Rabson, M. S. and Howley, P. M. (1985b): Nature 318 575-577

Yasumoto, S., Burkhardt, A. L., Doniger, J and DiPaolo, J. A. (1986): J Virol. 57 572-577

Yutsudo, M., Shimakage, T. and Hakura, A. (1985): Virology 144 295-298

Yutsudo, M. and Hakura, A. (1987): Int. J. Cancer 39 586-589