Studies of the Genome of Pseudorabies Virus

Gang Zhang

Thesis Submitted for the Degree of Doctor of Philosophy

Department of Biochemistry

University of Glasgow

April, 1991

Acknowledgement

I would like to express my sincere thanks to my supervisor, Dr. David P. Leader, for his very kind patience, helpful advice, and continual help, support and encouragement throughout the three and half years.

I would like to thank Professor M. D. Houslay for making the facilities of the department available to me, and Dr. John P. Goddard, my auditor, for his helpful suggestion and encouragement.

I acknowledge the British Council and the Education Committee of China to support me to study in the University of Glasgow for three years.

I thank the photographer in the Medical Illustration, Glasgow University, for her very kind and efficient help for taking the photographs.

Many thanks are given to my friends in the Department, especially Irene, June, Bertrand, Beate, Rowena and Val, for their help and friendship throughout my time in the laboratory.

I am also very grateful to my family for all their support and encouragement.

Content

Figure list		iv
Table list		vi
Abbrivation		vii
Summary		ix
Chapter 1. Introduction		
1.1.	Herpesviruses	1
1.2.	Alphaherpesviruses	5
1.2.1.	Reproductive cycle of alphaherpesvirus	5
1.2.2.	Genomes of alphaherpesviruses	7
1.3.	Protein kinases encoded by alphaherpesviruses	22
1.4	Objectives	30
Chapter 2. Materials and Methods		
2.1.	Chemicals	31
2.2.	Media and antibiotics for bacterial culture	31
2.2.1.	Liquid media	31
2.2.2.	Agar plates	32
2.2.3.	Antibiotics and inhibitors	32
2.3.	Bacteria	33
2.3.1.	Bacteria strains	33
2.3.2.	Preparation of bacterial cells competent for transformation and transfection	33
2.3.3.	Storage of bacterial cells	34
2.4.	Bacteriophages M13	34
2.5.	Cell culture and virus infection	35
2.5.1.	Virus and cell	35
2.5.2.	Cell culture	35
2.5.3.	Virus infection	35
2.6.	Vectors	36
2.6.1.	Vectors for cloning and sequencing	36
2.6.2.	Expression vectors	36
2.7.	Agarose gel electrophoresis of DNA	39

2.8.	Preparation of DNA	39
2.8.1.	Preparation of plasmid DNA	39
(a).	Small scale preparation (Mini-prep)	39
(b).	Preparation by alkaline lysis	40
(c).	Large scale preparation	41
2.8.2.	Preparation of M13 replicative form DNA	43
2.8.3.	Preparation of single-stranded bacteriophage M13 DNA	43
2.9.	Isolation of DNA fragments from low melting point agarose gel	44
2.10.	DNA subcloning	45
2.10.1.	Subcloning of restriction fragments	45
(a).	Digestion with restriction endonucleases	45
(b).	Treatment with alkaline phosphatase	45
(c).	Ligation	45
(d).	Transformation and transfection of bacterial cells	45
2.10.2.	DNase/restriction deletion subcloning	46
2.10.3.	Exonuclease III deletion subcloning	48
2.11.	Determination of nucleotide sequence	50
2.11.1.	Determination of nucleotide sequence in single-stranded M13 DNA	50
(a).	Annealing primer to template	50
(b).	Nucleotide chain extension reaction using the Klenow Fragment	51
(c).	Nucleotide chain extension reaction using *Taq* polymerase	52
(d).	Polyacrylamide buffer-gradient gel electrophoresis	53
(e).	Polyacrylamide buffer-linear gel electrophoresis	53
(f).	Autoradiography	54
2.11.2.	Determination of nucleotide sequence in double-stranded plasmid DNA	54
2.12.	Preparation of RNA from BHK cells	55
2.12.1.	Preparation of total RNA	55
2.12.2.	Preparation of polyadenylated RNA	56
2.13.	Labelling DNA	57
2.13.1.	End-labelling with [γ-^{32}P]ATP	57
2.13.2.	Nick-translation	57
2.14.	Hybridization analysis of nucleic acids	58
2.14.1.	Southern hybridization (DNA-DNA hybridization)	58
2.14.2.	Northern hybridization (RNA-DNA hybridization)	59
(a).	Agarose gel electrophoresis of RNA	59

(b). Hybridization 60
2.15. Site-directed mutagenesis using the polymerase chain reaction 60
2.16. Gel electrophoresis for protein analysis 62
2.17. Computer programs for DNA sequence analysis 62

Chapter 3. Results
3.1. Cloning and sequencing of the *Bam*HI 10 fragment of PRV genomic DNA 65
3.1.1. Cloning and identification of *Bam*HI fragments in pUC18 65
3.1.2. The nucleotide sequence of the *Bam*HI 10 fragment 70
3.2. The *pk* gene 80
3.2.1. The open reading frame 80
3.2.2. Expression of the gene *pk* in *E. coli* 86
3.3. The gene RSp40 88
3.3.1. The open reading frame 88
3.3.2. Characterization of the kinetic class of RSp40 mRNA 92
3.4. The Us/Rs junction 95

Chapter 4. Discussion
4.1. The gene *pk* 100
4.2. The gene RSp40 108
4.3. The genetic organization of the Us region of PRV 111

References 121

Figure List

Fig. 1.1.	Genomic structures of the alphaherpesviruses	3
Fig. 1.2.	Alignment of the genomes of two alphaherpesviruses	8
Fig. 1.3.	Restriction maps of the pseudorabies virus genome	20
Fig. 1.4.	Conserved amino acid residues in the catalytic domain of eukaryotic protein kinases	27
Fig. 1.5.	Deduced phylogenetic relationship of the catalytic domain of eukaryotic protein kinases	29
Fig. 2.1.	Structures of cloning and sequencing vectors	37
Fig. 2.2.	Structure of two prokaryotic expression vectors	38
Fig. 2.3.	Strategy of DNase deletion subcloning	47
Fig. 2.4.	Strategy of exonuclease deletion subcloning	49
Fig. 2.5.	Strategy of reconstruction of the *pk* gene of PRV	61
Fig. 3.1.	Selection of cloned *Bam*HI fragments of PRV DNA	66
Fig. 3.2.	Identification of the *Bam*HI 10 fragment of PRV DNA	69
Fig. 3.3.	Strategy for sequencing the *Bam*HI 10 fragment of PRV	71
Fig. 3.4.	Nucleotide sequence of the *Bam*HI 10 fragment of PRV DNA	72
Fig. 3.5.	Determination of S1-sensitive sites in pB10-BS1 DNA	75
Fig. 3.6.	Identification of S1 nuclease cleavage points by nucleotide sequencing	77
Fig. 3.7.	Alignment of possible open reading frames in the *Bam*HI 10 fragment of PRV	79
Fig. 3.8.	Nucleotide sequence of the *Eco*RV-*Bam*HI region of the *Bam*HI 10 fragment of PRV DNA	81
Fig. 3.9.	Dot-Matrix comparison of the similarity of the predicted amino acid sequence encoded by the possible open reading frame B with those of other alphaherpesvirus protein kinases	82
Fig. 3.10.	Protein-encoding potential of the open reading frame B (from nucleotides 1800 to 4008)	83
Fig. 3.11.	Hybridization analysis of the transcripts of the pk gene	85
Fig. 3.12.	Analysis by polyacrylamide gel electrophoresis of expression of the pk gene in *E. coli*	87

Fig. 3.13. Nucleotide sequence of the region between nucleotides 1 and 1807 of the *Bam*HI 10 fragment of PRV DNA 89
Fig. 3.14. Dot-Matrix comparison of the amino acid sequence encoded by the open reading frame A with that of US1 of HSV-1 and homologues in other alphaherpesviruses 90
Fig. 3.15. Protein-encoding potential of the open reading frame A (from nucleotide 1 to 1990) 91
Fig. 3.16. Effect of cycloheximide on the transcription of gene RSp40 of PRV 93
Fig. 3.17. Comparison of the *Bam*HI 10 and the *Bam*HI 12 fragments of PRV DNA 96
Fig. 3.18. Strategy for partially sequencing the of the *Bam*HI 12 fragment of PRV 97
Fig. 3.19. Nucleotide sequence of the region of the *Bam*HI 12 fragment of PRV DNA at the Us/TRs junction. 98

Fig. 4.1. Comparison of the catalytic domains of alphaherpesvirus protein kinases 102
Fig. 4.2. Comparison of the distribution of acidic and basic amino acid residues in alphaherpesvirus protein kinases 107
Fig. 4.3. Comparison of the conserved region in the predicted product of gene RSp40 with that of the homologues of alphaherpesviruses. 110
Fig. 4.4. Protein-encoding potential in the *Bam*HI 10 fragment of PRV. 113
Fig. 4.5. Comparison of the gene organization in the S segment of some alphaherpesviruses 114
Fig. 4.6. A simple model to illustrate how different gene organization in the S segment of alphaherpesviruses could have arisen by non-homologous recombination 116
Fig. 4.7. Comparison of the positions of the Us and Rs junctions in the Ka and the NIA-3 strains of PRV 118
Fig. 4.8. Model to account for the different positions of the Us and Rs junctions in individual strains of PRV 119

Table List

Table 1.1. Immdiate-early genes encoded in the HSV-1 genome. 10

Table 1.2. Genes relevant to virus DNA replication encoded in the HSV-1 genome. 12

Table 1.3. Genes related to nucleic acid metabolism encoded in the HSV-1 genome. 13

Table 1.4. Genes encoding virion proteins of HSV-1. 15

Table 1.5. Genes encoding other proteins in the HSV-1 genome. 16

Table 1.6. Homology between the genes in three alphaherpesvirus genomes. 18

Table 1.7. Properties of the alphaherpesviruse protein kinases 24

Table 3. BamHI fragments of PRV DNA cloned into pUC18. 67

Table 4.1. Amino acid differences between the protein kinases of the Ka and the NIA-3 strains of PRV. 104

Table 4.2. Homology of the genes in the S segments between three alphaherpesviruses. 112

Abbreviation

The abbreviations recommended by the *Biochemical Journal* [*Biochem. J.* (1981) **193**, 4-27.] have been used throughout this thesis with the following additions:

α-TIF	———	α-trans-inducing factor
BHK	———	baby hamster kidney fibroblasts
bp	———	base pairs
EBV	———	Epstein-Barr virus
EHV-1	———	equine herpes virus type 1
EHV-4	———	equine herpes virus type 4
HCMV	———	human cytomegalovirus
HVS	———	herpesvirus saimiri
HPLC	———	high performance liquid chromatography
HSV-1	———	herpes simplex virus type 1
HSV-2	———	herpes simplex virus type 2
HSV-PK	———	protein kinase encoded by herpes simplex virus
IPTG	———	isopropyl-β-thiogalactopyranoside
IR	———	internal repeat
IRL	———	internal long repeat
IRs	———	internal short repeat
kb	———	kilobases
nt	———	nucleotide(s)
pfu	———	plaque-forming units
PRV	———	pseudorabies virus
PRV-PK	———	protein kinase encoded by pseudorabies virus
Rs	———	short repeat
rpm	———	revolutions per min
TEMED	———	NNN'N'-tetramethylethylenediamine
TR	———	terminal repeat
TRL	———	terminal long repeat
TRs	———	terminal short repeat

ts	———	temperature sensitive mutation
Us	———	short unique
UL	———	long unique
VZV	———	varicella-zoster virus
VZV-PK	———	protein kinase encoded by varecella-zoster virus
X-gal	———	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Summary

This thesis describes the determination and analysis of the nucleotide sequence of the *Bam*HI 10 fragment in the small (S) segment of the genome of pseudorabies virus, strain Ka. The fragment was cloned into the plasmid vector pUC18, and its complete sequence determined. It was found to be 4008 nucleotides in length and to include two genes.

One gene designated *pk*, lies immediately upstream from the previously determined gene gX encoding glycoprotein X, and is predicted to encode a protein of 344 amino acid residues. The amino acid sequence predicted for this protein retains many of the motifs highly conserved in eukaryotic cellular protein kinases, and it is homologous to the predicted products of gene US3 of herpes simplex virus type 1 and gene 66 of varicella-zoster virus. The amino acid sequence of the product of the gene *pk* shares 39% and 37% identity, respectively, with those of gene US3 of herpes simplex virus type 1 and gene 66 of varicella-zoster virus in the catalytic domain, and 27% of the positions show perfect identity in all three protein kinases. In contrast to the catalytic domain, the amino acid sequences of the non-catalytic domain diverge markedly between the three viral protein kinases. Polyadenylated mRNA was isolated from cells infected with pseudorabies virus, and hybridized to DNA fragments from the genes *pk* and gX. The fragment from the gene gX hybridized to transcripts of 1.6 kb and 3 kb, and the later transcript also hybridized to the fragment from the gene *pk*. This and other evidence suggests that the protein kinase is encoded by the longer transcript, and that this is co-terminal with the shorter transcript which encodes gX.

The second gene, designated RSp40, lies in the short repeat region of the genome, and is predicted to encode a protein of 364 amino acid residues. The

predicted product of this gene is homologous to immediate-early gene US1 of herpes simplex virus type 1, gene 63 of varicella-zoster virus and ORF4 of equine herpes virus type 4. Polyadenylated mRNA was isolated from cells infected with pseudorabies virus for different times in the presence or absence of cycloheximide, and hybridized to DNA fragment from RSp40 gene, and also to one from pseudorabies virus immediate-early gene IE180. No mRNA hybridizing to the DNA fragment from RSp40 gene was synthesized in the presence of cycloheximide, in contrast to the results with IE180. This showed that, unlike its herpes simplex virus type 1 homologue, RSp40 of pseudorabies virus is not an immediate-early gene.

Between the *pk* gene and RSp40 are two reiterated sequences: R_1 and R_2. R_1 contains 11 full and apparently identical copies of a 35 nucleotide repeat, and 24 nucleotides of a twelfth copy. R_2 has 9 full and identical copies of a 10 nucleotide repeat.

The *Bam*HI 12 fragment was also analyzed, in order to identify the positions of the junctions between the short unique and short repeat regions of the genome. First, identical restriction fragments between the *Bam*HI 10 and the *Bam*HI 12 were identified, and then the nucleotide sequence of the appropriate portion of the *Bam*HI 12 fragment, which is 272 bp from the end in the Us region, was determined. The junction of the short internal repeat and short unique regions was identified as lying between nucleotides 2533 and 2534 in the *Bam*HI 10 fragment, and the junction of the short unique and the terminal repeat regions between nucleotides 251 and 252 in the *Bam*HI 12 fragment. This determination allowed comparison of the disposition of the genes in the short unique region of pseudorabies virus with those of herpes simplex virus type 1 and varicella-zoster virus and the evolution of the short unique region of different alphaherpesviruses to be considered.

CHAPTER 1
INTRODUCTION

1.1. Herpesviruses

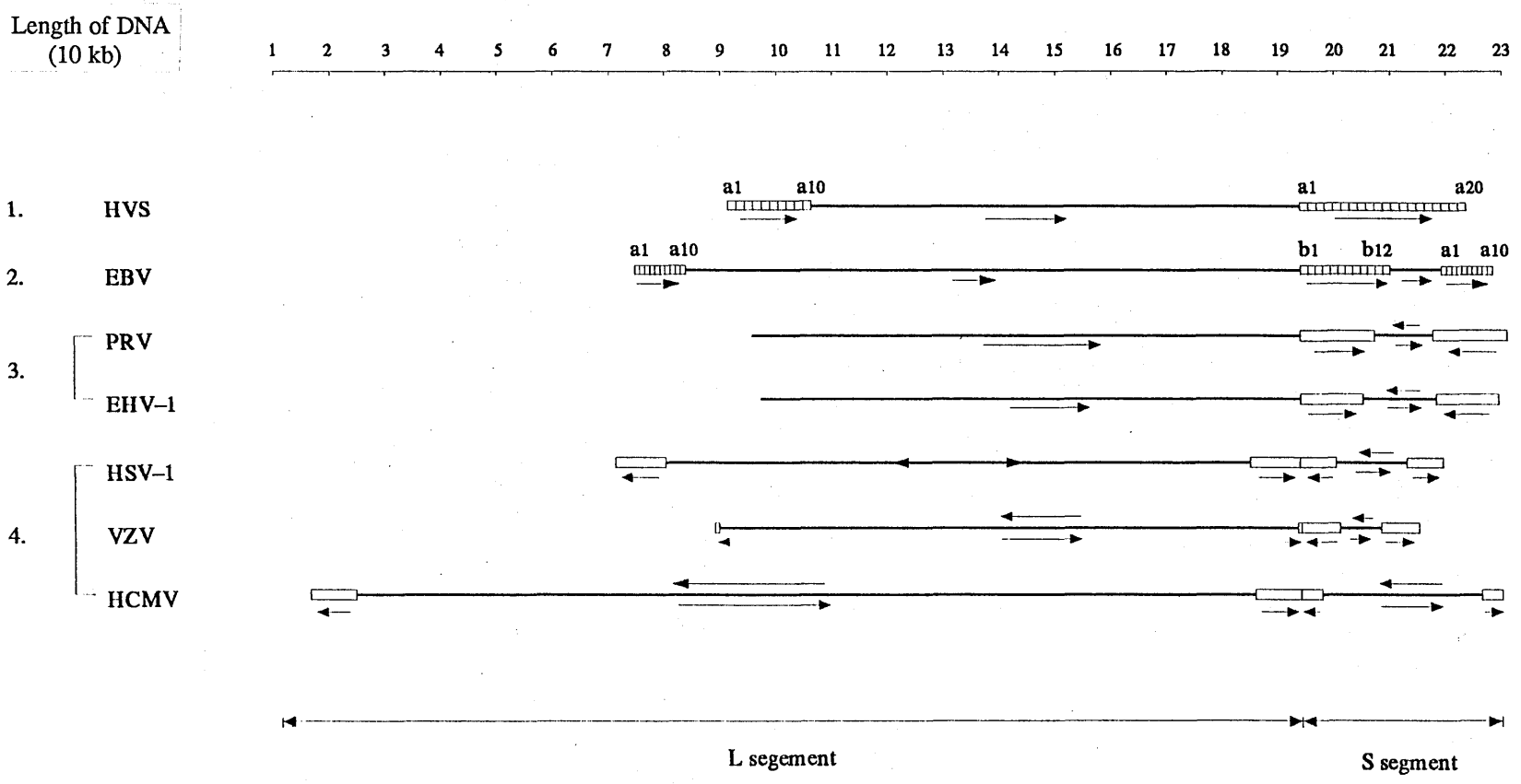
Herpesviruses, comprising the family Herpesviridae, are a group of DNA viruses which share common structural features of their virions, but infect a wide range of animal hosts (Roizman 1982). Some of them are well known as agents of infectious diseases in man, *e.g.* herpes simplex virus type 1 (HSV-1), varicella-zoster virus (VZV), and Epstein-Barr virus (EBV). Others infect animals, and it is one of these, pseudorabies virus (PRV), which is the topic of this thesis. HSV-1 causes clinical symptoms such as cold sores, and VZV is the pathogen of chicken pox in children and shingles in adults. EBV is another human virus which causes infectious mononucleosis in western countries, Burkitt's lymphoma in equatorical regions, and nasopharyngeal carcinoma in south east Asia, north and east Africa. The natural host of PRV is swine, although its host range is very wide.

Generally, the virions of herpesviruses includes four structural elements: a core, capsid, tegument, and membrane or envelope (Widly *et al.*, 1960; Roizman and Furlong, 1974). In the centre of the virion is the core, consisting of the viral DNA wrapped on a fibrillar spool, and this is covered by the capsid containing 12 pentameric and 150 hexameric capsomeres. A number of other proteins is also enclosed in the capsid, and some of them have enzyme activity. The tegument, a variable amount of material asymmetrically surrounding the capsid, contains most

of the virus structural proteins. A membrane or envelope consisting of the nuclear membrane of the host cell, in which some of viral glycoproteins are associated, encloses this entire structure. Some of the glycoproteins are crucial for the interaction of virus with its host cells (see further below).

Herpesvirus genomes released from virus particles are linear double-stranded DNA including unique and repeat regions. In contrast to the morphology of herpesviruses, the structure of viral genomes differs between individual species. On the basis of their genomic structure, herpesviruses have historically been divided into five groups (Roizman, 1982), A-E. However only four of these, *i.e.* B-E, are presented in Fig. 1.1 as the single member of group A, channel catfish virus, (Chousterman *et al.*, 1979) is now known on the basis of nucleotide sequencing, not to be a herpesvirus (Davison, A., personal communication). Viruses in group B have multiple reiterations of one set of sequences in the same orientation at both termini, *e.g.* herpesvirus saimiri (HVS) (Fleckenstein *et al.*, 1975, 1978). Members in group C have multiple reiterations of one set of sequences in the same orientation at both termini, as well as internal tandem repetitions of other sets of sequences. These tandem reiterations divide the unique region into two parts: long unique (U_L) and short unique (U_S), *e.g.* EBV (Given and Kieff, 1979; Hayward *et al.*, 1980). In group D, virus genomes are characterized by an inverted reiteration of a set of sequences from one terminus internally, as well as by reiterations of a subset of these sequences in the same orientation at the other terminus, *e.g.* PRV (Stevely, 1977; Ben-Porat *et al.*, 1979) and equine herpes virus type 1 (EHV-1) (Whalley *et al.*, 1981). These reiterated sequences are designated internal repeat (IR) and terminal repeat (TR) regions. On the basis of the orientations of the U_L and U_S regions, viruses in this group have two isomeric structures which arise from homologous recombination between the two reiterated sequences. Finally, members in group E are characterized by inverted reiterations of sets of sequences from both termini internally, as well as by

Fig. 1.1. Genomic structures of the alphaherpesviruses. The genomes presented are those of members in groups B (1), C (2), D (3) and E (4). These are herpesvirus saimiri (Fleckenstein *et al.*, 1975, 1978), EBV (Given and Kieff, 1979; Hayward *et al.*, 1980), PRV (Stevely, 1977; Ben-Porat *et al.*, 1979), equine herpes virus type 1 (EHV-1) (Whalley *et al.*, 1981), HSV-1 (Sheldrick and Berthlot, 1975; Wadswoth *et al.*, 1975), VZV (Dumas *et al.*, 1981; Davison, 1984) and human cytomegalovirus (Westrate *et al.*, 1980). The open boxes represent the repeat regions, and arrows show the orientation of both repeat and unique regions. The L and S segments are indicated for group C-E, and the scale is the length of the DNA in kb.


Length of DNA
(10 kb)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
1. HVS
a1 a10
a1 a20
2. EBV
a1 a10
b1 b12
a1 a10
3. PRV
EHV-1
4. HSV-1
VZV
HCMV
L segement
S segment

reiterations of a subset of these sequences at both termini in the same orientation, *e.g.* HSV-1 (Sheldrick and Berthlot, 1975; Wadswoth *et al.*, 1975) and VZV (Dumas *et al.*, 1981; Davison, 1984). The reiterated sequences flanking the UL region are designated long terminal repeat (TRL) and long internal repeat (IRL), whereas those at each end of Us region are short internal repeat (IRs) and short terminal repeat (TRs). The RL regions and UL region are constituents of the L segment, and the S segment contains the Rs regions and Us region. For reasons similar to those explained for group D, the viruses in group E have four isomeric forms.

On the basis of their biological properties, herpesviruses can be divided into three subclasses (Roizman, 1982): alpha (α), beta (β), and gamma (γ). Alphaherpesviruses, *e.g.* HSV-1, VZV, and PRV, have a variable host range from very wide (PRV) to very narrow (HSV-1), and a short reproductive cycle. In cell culture the viruses lyse their susceptible cells rapidly, and in nature latent infection occurs frequently, but not exclusively, in ganglia. In contrast to that of alphaherpesviruses, the host range of betaherpesviruses, *e.g.* human cytomegalovirus (HCMV), is narrow (normally restricted to a species or genus to which the host belongs), and the reproductive cycle is relatively long. Infection by the viruses in cell culture slowly produces lytic foci, and the infected cells become enlarged both *in vitro* and *in vivo*. The latent infection in nature possibly happens in secretory glands, lymphoreticular cells, kidneys, and other tissues. Gammaherpesviruses, *e.g.* EBV, specifically infect either B or T lymphocytes. *In vivo* the host range is usually limited to a family, and *in vitro* the viruses replicate in lymphoblastoid cells. Latent infections frequently establish in lymphoid tissue.

The work described in this thesis concerns the genome of PRV and therefore this introduction will concentrate on studies of alphaherpesviruses.

1.2. Alphaherpesviruses

1.2.1. Reproductive cycle of alphaherpesvirus

The reproductive cycle of alphaherpesvirus, as with other viruses, starts from entry into the host cell. The details of the pathway are not well understood, but studies with HSV-1 have indicated that this process involves specific viral-associated glycoproteins and their binding sites and receptors on the surface of target cells (Fuller and Spear, 1987). The first step of the entry is attachment of the virion onto the surface of a host cell. This involves a virus-cell interaction through both heparin-like cell-associated glycosaminoglycans (Wudunn and Spear, 1989), and the basic fibroblast growth factor receptor (Kaner *et al.*, 1990). As soon as the virus has attached to the host cell, the viral envelope fuses with the plasma membrane, and is removed (Fuller and Spear, 1987). Then, the viral capsid enters the cytoplasm of the cell (Cai *et al.*, 1988; Johnson and Spear, 1989). The capsid moves from the cytoplasm to the nucleus, and the viral DNA is injected into the nucleus through the nuclear pores (Batterson *et al.*, 1983).

On the basis of the co-ordinated regulation and sequential order of gene expression in virus infection, genes of alphaherpesviruses are classified into three kinetic classes: immediate-early (α), early (β), and late (γ) genes (Honess and Roizman, 1974). The transcription of the viral genes is performed by host RNA polymerase II (Constanzo *et al.*, 1977; Ben-Zéev and Becker, 1977). Unlike β or γ genes, the expression of α genes does not require any newly synthesized viral protein, and therefore genes in this kinetic class can be transcribed in the presence of inhibitors of protein synthesis, *i.e.* cycloheximide (Honess and Roizman, 1974). However, in HSV-1 the transcription of α genes also requires a component of the tegument, α-trans-inducing factor (α-TIF), which enters the nucleus in company with viral DNA (Roizman and Batterson, 1984). There are indications that α-TIF forms a complex with the transcription factor oct-1, and activates the transcription of

α genes, since promoter-regulatory domains of α genes contain a *cis*-acting site (for review, see Goding and O'Hare, 1989). In contrast to the situation in HSV-1, there is no requirement in PRV for a trans-inducing factor analogous to α-TIF for α gene transcription (Campbell and Preston, 1987). In HSV-1, five α genes have been identified (Honess and Roizman, 1974), whereas only one has been identified in PRV (Ihara *et al.*, 1983). The products of the α genes are required for the transcription of β and γ genes, and it is also known that the product of one of these, IE175, negatively regulates α gene expression (Kristie and Roizman, 1986a, 1986b). However, in general, the functions of α proteins are poorly understood (see further below). The expression of β genes starts very early during infection, and reaches a maximum after 5-7 h (Honess and Roizman, 1974). Examples of products of β gene are viral enzymes, such as DNA polymerase, thymidine kinase, or DNA-binding protein. The γ genes are expressed after viral DNA synthesis has been initiated, and many of their products are virion-associated structural proteins.

Immediately after infection, a striking change occurring in the host cells is that cellular DNA synthesis is shut off. Host protein synthesis is inhibited in two phases, and RNA synthesis is greatly reduced when the β and γ proteins are present (Roizman and Batterson, 1984). It has been shown that the early host shut-off of protein synthesis is dependent on a virion protein (the product of gene UL41) which decreases the stability of host (and viral) mRNA (Kwong and Frenkel, 1989).

In contrast to the linearized DNA released from virus particles, the viral DNA in infected cells is in circular form. The circularization occurs immediately after infection, and it has been observed in cells infected with both HSV-1 and PRV (Poffenberger and Roizman, 1985; Ben-Porat *et al.*, 1980). Viral DNA replication occurs by a rolling circle mechanism (Jacob and Roizman, 1977; Jacob *et al.*, 1979), and there are three origins of replication in both HSV-1 and PRV genomes; one in each of the inverted repeat regions and one in the middle of the UL region

(Roizman and Batterson, 1984; Ben-Porat and Veach, 1980). Unlike the transcription of viral DNA, virus replication depends on viral enzymes, *i.e.* the products of genes UL5, UL8, UL9, UL29, UL30, UL42 and UL52 (references see further below).

In late infection, the capsid proteins, associated with cytoskeleton, are transported into the nucleus, and viral DNA of unit length is generated and packaged into newly synthesized capsids (Bibor-Hardy *et al.*, 1982, 1985; Ben-Zéev *et al.*, 1983). The capsids are modified by assembling other viral structural proteins, and enveloped with nuclear membrane into which viral glycoproteins are inserted. The virions move though the cytoplasm, and then mature viruses are released from the cell.

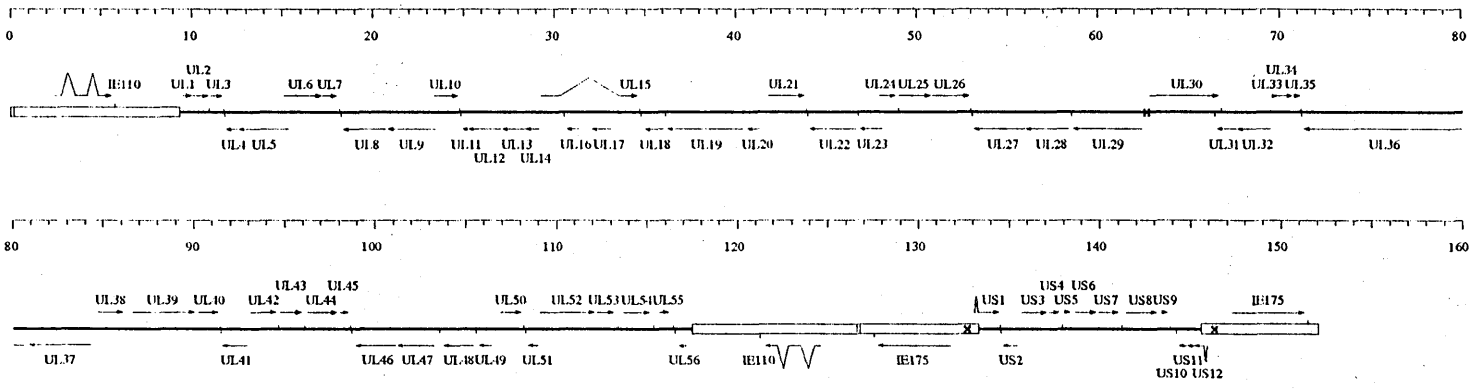
In natural infections of HSV-1, recurrent lesions frequently appear, and are often induced from the latent state by stimuli, such as fever and exposure to sunlight. The lesions occur in the skin, but the viral DNA replication takes place in the neurons, and the virus remains in the trigeminal ganglion and central nervous system (Cook *et al.*, 1974; Openshaw, 1983). Although latency of alphaherpesvirus has been studied since early this century, this complex problem, involving the host itself as well as virus gene expression, is still poorly understood (for reviews, see Hill, 1985 and Roizman and Sears, 1987).

1.2.2. Genomes of alphaherpesviruses

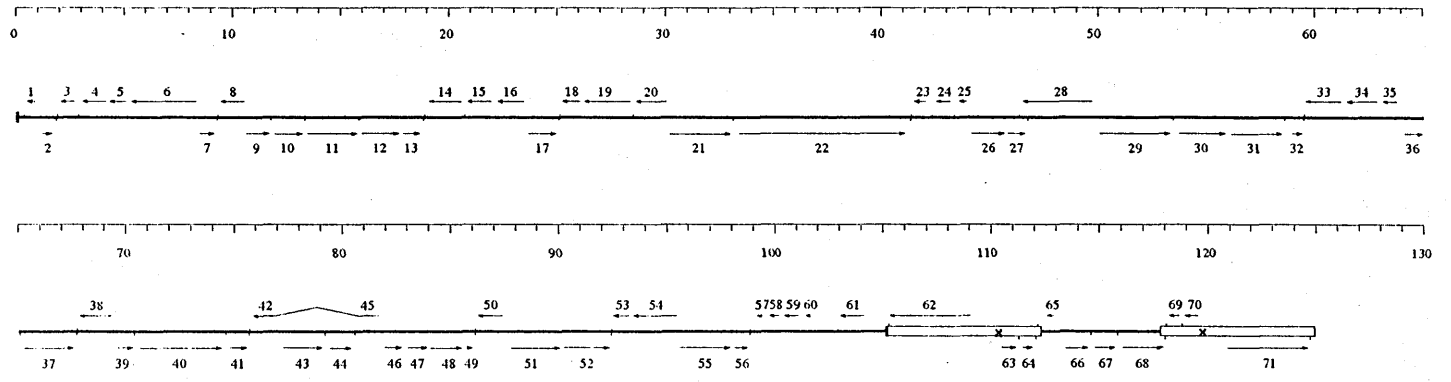
The best understood alphaherpesvirus is HSV-1, the genome of which has completely been sequenced in strain 17 (McGeoch *et al.*, 1985, 1986b, 1988a; Perry and McGeoch, 1988). Although the genome of VZV, strain Dumas has also been sequenced (Davison and Scott, 1986), much less is known about this virus. The genome of HSV-1 contains 152,260 bp (68.3% G+C), in which 72 genes encode 70 individual proteins (two genes lie in the repeat regions). The centre of the origin of replication in the UL region (ori_L) is located at nucleotides 62,475/62,476

Fig. 1.2. Alignment of the genomes of two alphaherpesviruses. The gene organizations of HSV-1 (a) and VZV (b) are shown, and the thick horizontal lines and the open boxes represent the unique and the repeat regions, respectively. The size and orientation of each proposed functional open reading frame are represented by arrows. The positions of proposed polyadenylation/processing sites are marked by short vertical lines, and introns are indicated by diagonal lines. The origins of DNA replication are denoted by 'X'. The scale is of the length of the DNA in kb.

a.


0 10 20 30 40 50 60 70 80
IE110
UL2
UL1 UL3
UL6 UL7
UL10
UL15
UL21
UL24 UL25 UL26
UL30
UL34
UL33 UL35
UL4 UL5
UL8 UL9
UL11 UL13
UL12 UL14
UL16 UL17
UL18
UL19
UL20
UL22 UL23
UL27 UL28 UL29
UL31 UL32
UL36
80 90 100 110 120 130 140 150 160
UL38 UL39 UL40
UL43 UL45
UL42 UL44
UL50 UL52 UL53 UL54 UL55
US1
US4 US6
US3 US5 US7 US8 US9
IE175
UL37
UL41
UL46 UL47 UL48 UL49
UL51
UL56
IE110
IE175
US2
US11
US10 US12

b.


0 10 20 30 40 50 60
1 3 4 5 6 8 14 15 16 18 19 20 23 24 25 28 33 34 35
2 7 9 10 11 12 13 17 21 22 26 27 29 30 31 32 36
70 80 90 100 110 120 130
38 42 45 50 53 54 5758 59 60 61 62 65 69 70
37 39 40 41 43 44 46 47 48 49 51 52 55 56 63 64 66 67 68 71

(Weller *et al.*, 1985), and those in the IRs region (*ori*$_S$) are at 131,999 and 146,234 (Stow and McMonagle, 1983). The layout of the 72 genes is presented in Fig. 1.2.a (McGeoch *et al.*, 1988a). Detailed studies of HSV-1 transcripts have showed that there are seven introns in the genome, of which four are in the coding region of the two copies of IE110 (nucleotides 2318-3084, 3750-3887, 122,482-122,619, and 123,285-124,051) (Perry *et al.*, 1986), and one is in the coding region of UL15 (nucleotides 30,047-33,634) (Costa *et al.*, 1985). The other two introns occupy the 5' non-coding regions of US1 and US12 (nucleotides 132,373-132,541 and 145,859-146,027) (Murchie and McGeoch, 1982).

Five genes, IE110, UL54, IE175, US1, and US12 encode the immediate-early proteins (Table 1.1). As mentioned above, the products of α genes play a role as transcriptional regulators. IE175 encodes protein α4 which is required for the expression of early and late genes, and also negatively regulates α gene expression (Kristie and Roizman, 1986a, 1986b). The product of IE175 has DNA-binding function (this is also true for those of IE180 of PRV and gene 62 of VZV), implying that the formation of a special complex between immediate-early protein and an promoter of α genes may be the basis of the regulation of α gene transcription (Wu and Wilcox, 1991). The expression of IE175 is continually required during infection (Watson and Clements, 1980). However, protein α47, the product of US12, is not essential for virus growth in at least some cells in culture (Mavromara-Nazos and Roizman, 1986; Longnecker and Roizman, 1986). Protein α22, encoded by US1, seems to be required late in infection, but is dispensable in some cells, although not in others (Post and Roizman, 1981; Sears *et al.*, 1985). The UL54 product, protein α27 appears to be concerned with late gene expression (Sacks *et al.*, 1985), and the product of IE110, protein α0, although much less well understood, is thought to play a role as a general trans-activator (Elkareh *et al.*, 1985; O'Hare and Hayward, 1984, 1985).

Table 1.1. Immdiate–early genes encoded in the HSV–1 genome

Gene	No. of amino acid residues	Function of gene product	Reference
IE110	775	Trans–activitor	Everett (1984); Perry *et al.* (1986)
UL54	512	Trans–activator	Watson *et al.* (1979); Sacks *et al.* (1985); Everett (1986)
IE175	1298	Trans–activator	Preston (1979); McGeoch *et al.* (1986b)
US1	420	Immediate–early proein	McGeoch *et al.* (1985)
US12	88	Immediate–early protein	Marsden *et al.* (1982); Murchie & McGeoch (1982)

The products of seven genes in the UL region are necessary and sufficient for viral DNA synthesis (Table 1.2). The products of genes UL29 and UL30 are essential for viral DNA replication (Conley *et al.*, 1981; Kops and knipe, 1988). UL30 encodes the viral DNA polymerase, and the product of UL29 is the major DNA-binding protein (Quinn and McGeoch, 1985; Gibbs *et al.*, 1985). The other DNA-binding protein is encoded by gene UL42, which is also required for viral DNA synthesis, and the product of gene UL9 specifically recognizes the viral replication origin (Wu *et al.*, 1988; McGeoch *et al.*, 1988b; Weir *et al.*, 1989). The products of UL5, UL8, and UL52 were found to form a helicase-primase complex (Crute *et al.*, 1989), suggesting that this complex may prime lagging-strand synthesis as it unwound at the viral replication fork. The complex has DNA-dependent ATPase, DNA-dependent GTPase, DNA helicase, and DNA primase activities (Dodson *et al.*, 1989), however the ATPase and helicase activities of the complex do not require the UL8 protein (Calder and Stow, 1990). Nevertheless, the product of gene UL8 is essential for viral DNA synthesis (Carmichel and Weller, 1989).

The products of some other genes in the UL region are involved in nucleotide metabolism (Table 1.3). Gene UL12 encodes a deoxyribonuclease (Draper *et al.*, 1986; McGeoch *et al.*, 1986b), and thymidine kinase is encoded by UL23 (McKnight, 1980; Wagner *et al.*, 1981). The large and small subunits of ribonucleotide reductase are the product of genes UL39 and UL40 (Draper *et al.*, 1982; Preston *et al.*, 1984; Nikas *et al.*, 1986), the product of UL2 encodes uracil-DNA glycosylase (Mullaney *et al.*, 1989; Worrad and Caradonna, 1988), and deoxyuridine triphosphatase is encoded by gene UL50 (Preston and Fisher, 1984). These enzymes are dispensable for growth of virus in cells in culture, except for the deoxyribonuclease, which appears essential (Sanders *et al.*, 1982; Goldstein and Weller, 1988; Fisher and Preston, 1986; Mullaney *et al.*, 1989). This enzyme may

Table 1.2. Genes relevant to virus DNA replication encoded in the HSV–1 genome

Gene	No. of amino acid residues	Function of gene product	Reference
UL5	882	Component of helicase–primase complex	Wu *et al.* (1988); McGeoch *et al.* (1988a); Crute *et al.* (1989)
UL8	750	Component of helicase–primase complex	Wu *et al.* (1988); McGeoch *et al.* (1988a); Crute *et al.* (1989)
UL9	851	Replication origin–binding protein	Wu *et al.* (1988); McGeoch *et al.* (1988a); Weir *et al.* (1989)
UL29	1196	Major DNA–binding protein	Conley *et al.* (1981); Quinn and McGeoch (1985)
UL30	420	Replicative DNA polymerase	Chartrand *et al.* (1979); Quinn & McGeoch (1985); Gibbs *et al.* (1985)
UL42	488	DNA–binding protein	Wu *et al.* (1988); McGeoch *et al.* (1988a); Parris *et al.* (1989)
UL52	1058	Component of helicase–primase complex	Wu *et al.* (1988); McGeoch *et al.* (1988a); Crute *et al.* (1989)

Table 1.3. Genes related to nucleic acid metabolism encoded in the HSV–1 genome

Gene	No. of amino acid residues	Function of gene product	Reference
UL2	334	Uracil–DNA glycosylase	Worrad & Caradonna (1988); Mullaney *et al.* (1989)
UL12	626	Deoxyribonuclease	Preston & Cordingley (1988); McGeoch *et al.* (1986a)
UL23	376	Thymidine kinase	McKnight (1980); Wagner *et al.* (1981)
UL39	1137	Large subunit of ribonucleotide reductase	Preston *et al.* (1984); Nikas *et al.* (1986)
UL40	340	Small subunit of ribonucleotide reductase	Draper *et al.* (1982); McLauchlan & Clements (1983)
UL50	371	Deoxyuridine triphosphatase	Preston & Fisher (1984)

possibly play a role in cleavage-packing of viral DNA into capsids (Chou and Roizman, 1989).

Another group of genes in the HSV-1 genome encode virion proteins (Table 1.4). These include UL19, encoding the major capsid protein (Costa *et al.*, 1984; Davison and Scott, 1986b), and UL48, encoding the major tegument protein and transcriptional activator (α-TIF) of immediate-early genes, as mentioned previously (Campbell *et al.*, 1984; Dalrymple *et al.*, 1985; Pellett *et al.*, 1985b; Roizman and Batterson, 1984). Some other identified virion proteins are the product of UL41, UL38, UL18 and UL26. The product of UL41 is responsible for the shut-off of host protein synthesis (Kwong and Frenkel, 1989), and the products of UL18 and UL38 are components of the capsid (Pertuiset *et al.*, 1989; Rixon *et al.*, 1990). UL38 was found to be essential for capsid assembly (Rixon *et al.*, 1990). The product of UL26 is a capsid protein which is required for encapsidation of virus DNA (Preston *et al.*, 1983). Seven genes, UL22, UL27, UL44, US4, US6, US7 and US8, encode viral glycoproteins (Table 1.4), although the functions of the glycoproteins are not clearly understood. (Gompels and Minson, 1986; McGeoch and Davison, 1986a; Bzik *et al.*, 1984; Pellett *et al.*, 1985a; Frink *et al.*, 1983; Draper *et al.*, 1984; McGeoch *et al.*, 1985, 1987; Frame *et al.*, 1986; Watson *et al.*, 1982; Longnecker *et al.*, 1987). Glycoproteins E, G and I, products of genes US8, US4 and US7, are dispensable for growth of virus in cells in culture (Longnecker and Roizman, 1986, 1987; Longnecker *et al.*, 1987). Glycoprotein C, the product of gene UL44, are also not essential for growth virus in cells in culture (Holland *et al.*, 1984), although it is responsible for the adsorption of virus onto cells (Herold *et al.*, 1991). The others genes, however, appear essential. Glycoprotein D, the product of gene US6 is found on the surface of virion, and is involved in the fusion of the viral envelope with the plasm membrane (Fuller and Spear, 1987). It is also likely to play a role in permitting release of virus from cells

Table 1.4. Genes encoding virion proteins of HSV–1

Gene	No. of amino acid residues	Function of gene product	Reference
UL18	318	Capsid protein	Pertuiset *et al.* (1989); Rixon *et al.* (1990)
UL19	1374	Major capsid protein	Costa *et al.* (1988); Davison & Scott (1986b)
UL22	838	Virion glycoprotein H	Gompels & Minson (1986); McGeoch &Davison (1986a)
UL26	653	Capsid protein	Preston *et al.* (1983)
UL27	904	Virion glycoprotein B	Bzik *et al.* (1984); Pellett *et al.* (1985a)
UL38	465	Capsid protein	Pertuiset *et al.* (1989); Rixon *et al.* (1990)
UL41	489	Virion protein	Kwong & Frenkel (1989)
UL44	511	Virion glycoprotein C	Frink *et al.* (1983); Draper *et al.* (1984)
UL48	490	Major tegument protein (α-trasn-inducing factor)	Campbell *et al.* (1984); Dalrymple *et al.* (1985); Pellett *et al.* (1985b); Roizman & Batterson (1984)
US4	238	Virion glycoprotein G	McGeoch *et al.* (1985, 1987); Frame *et al.* (1986a)
US6	394	Virion glycoprotein D	Waston *et al.* (1982)
US7	390	Virion glycoprotein I	McGeoch *et al.* (1985); Longnecker *et al.* (1987)
US8	550	Virion glycoprotein E	McGeoch *et al.* (1985)

Table 1.5. Genes encoding other proteins in the HSV–1 genome

Gene	No. of amino acid residues	Function of gene product	Reference
UL11	96	RNA–binding activity	Roller & Roizman (1990)
UL13	518	Protein kinase	Smith & Smith (1988)
US3	481	Protein kinase	McGeoch &Davison (1986b); Purves *et al.* (1987b); Frame *et al.* (1987)

(Johnson and Spear, 1989). Glycoproteins B and H, the products of genes UL27 and UL22, are also essential for growth of virus in cells in culture (DeLuca *et al.*, 1982; Gompels and Minson, 1986; Cai *et al.*, 1988), but the functions of these glycoproteins are even less clearly understood.

The genome of HSV-1 also contains genes which do not belong to any of the three groups described above (Table 1.5). One striking example is the protein kinase gene, US3 (McGeoch and Davison, 1986b; Purves *et al.*, 1987b; Frame *et al.*, 1987). The other is US11 which has been demonstrated to have sequence-specific RNA-binding activity (Roller and Roizman, 1990), but the importance of this protein in the viral reproductive cycle is not clear. Smith and Smith (1989) suggested that UL13 might encode a protein kinase, but the product of this gene has not yet been identified.

The genome of VZV, strain Dumas, contains 124,884 bp (46.02% G+C), and 70 genes in VZV genome encode 67 different proteins (Davison and Scott, 1986). The centres of the VZV origins of replication (ori_S) were found at nucleotides 110,214/110,215 and 119,682/119,683 (Stow and Davison, 1986). The layout of the 70 genes is presented in Fig. 1.2.b (Davison and Scott, 1986). Although the functions of VZV genes are much less well understood than those of the HSV-1 genes, extensive homology is found between the predicted products of the genes of the two viruses (Davison and McGeoch, 1986; McGeoch *et al.*, 1988) — see Table 1.6. In the L segment of the genome, most of the genes of HSV-1 have counterparts in VZV, except UL45 and UL56. The degree of similarity of the proteins varies widely, from 59% to only local regions of homology. VZV has five extra genes in this region: genes 1, 2, 13, 32, and 57, and does not have a counterpart to UL56 of HSV-1. Although VZV gene 61 appears to be the counterpart of IE110 of HSV-1, it occupies a position in the UL region extremely near the UL/IRL junction rather than in the RL region as in HSV-1. VZV gene 13

Table 1.6. Homology between the genes in three alphaherpesvirus genomes[i]

Gene of HSV–1	Gene of VZV	Gene of PRV
IE110	61	
UL1	60	
UL2	59	
UL3	58	
—	57	
UL4	56	
UL5	55	
UL6	54	
UL7	53	
UL8	52	
UL9	51	
UL10	50	
UL11	49	
UL12	48	
UL13	47	
UL14	46	
UL15 Exon 1	45	
Exon 2	42	
UL16	44	
UL17	43	
UL18	41	
UL19	40	
UL20	39	
UL21	38	
UL22	37	

UL23	36	tk
UL24	35	
UL25	34	
UL26	33	
—	32	
UL27	31	gII
UL28	30	*
UL29	29	
UL30	28	
UL31	27	
UL32	26	
UL33	25	
UL34	24	
UL35	23	
UL36	22	
UL37	21	
UL38	20	
UL39	19	
UL40	18	
UL41	17	
UL42	16	
UL43	15	
UL44	14	gIII
—	13	
UL45		
UL46	12	
UL47	11	
UL48	10	
UL49	9	
UL50	8	
UL51	7	
UL52	6	
UL53	5	
UL54	4	
UL55	3	

UL56	—	
—	2	
—	1	
IE175	62	IE180
US1	63	
US2	—	
US3	66	
US4	—	gX
US5	—	
US6	—	g50
US7	67	g63
US8	68	gI
US9	65	11K
US10	64	
US11	—	
US12	—	

---

i. From the data of McGeoch *et al.* (1985, 1986b, 1988a), Perry &McGoech (1988), Davison & Scott (1986), Kit *et al.* (1984), Rea *et al.* (1985), Petrovskis *et al.* (1986a, 1986b), Petrovskis and Post (1987), Robbins *et al.* (1986, 1987), Pederson & Enquist (1988) and Cheung (1989a).

–. Absent from virus genomes.

*. Sequenced in PRV, but not designated.

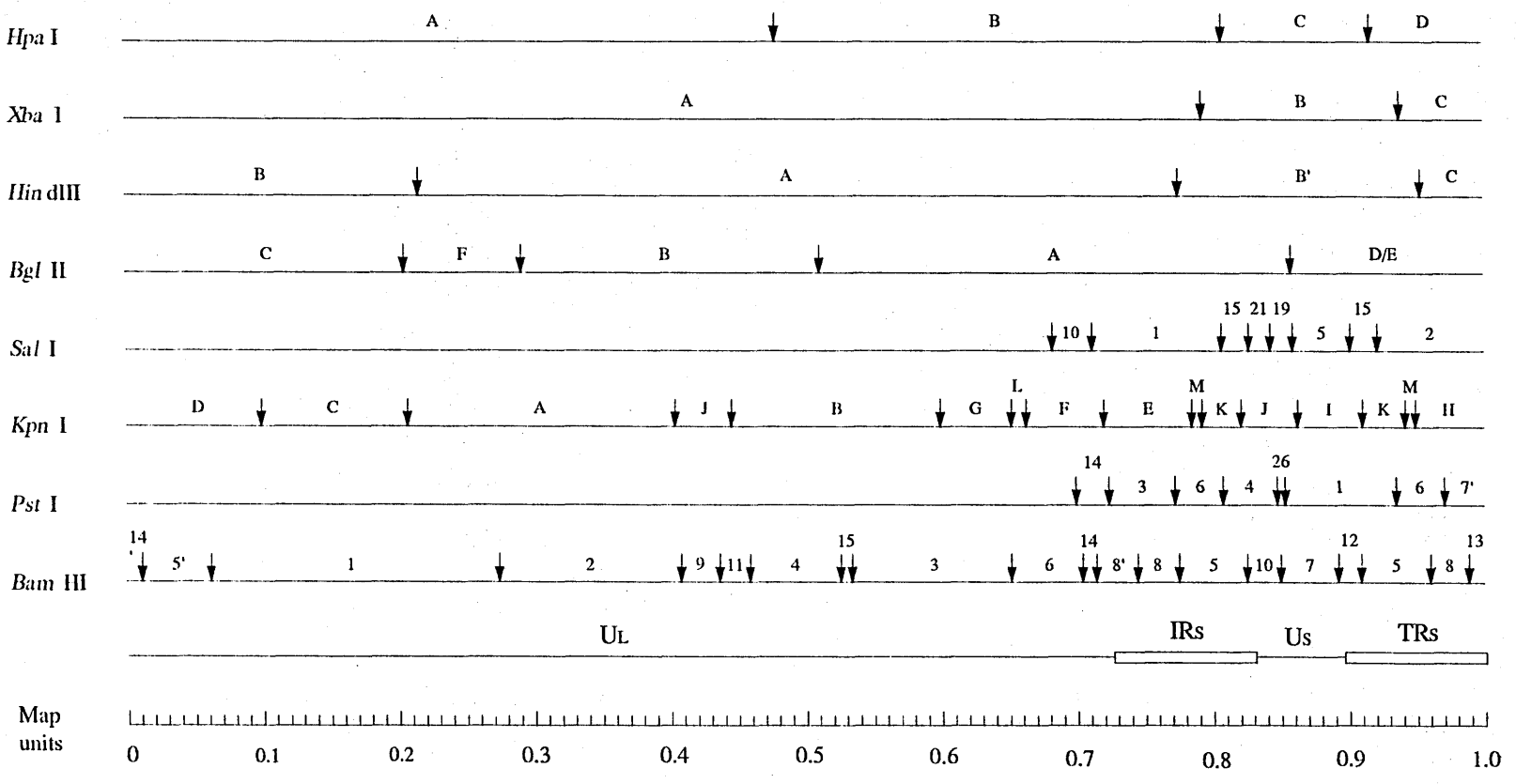
encodes a thymidylate synthetase (Thompson *et al.*, 1987), not found in HSV-1. In the S segment each gene of VZV has a counterpart of HSV-1, although the highest degree of homology of the proteins is not greater than 33%. Half of the genes of HSV-1 in the Us region lack counterparts in VZV. These are US2, US4, US5, US6, US11, and US12.

It is interesting that the relationship between the genes of the two viruses also extends to their layout in the genomes, especially in the L segment (see Fig. 1.2 and Table 1.6). In the L segment, VZV is congruent with HSV-1 in the relative positions and orientations of corresponding genes. This is also the case for most of the homologous genes in the S segments of the two viruses, except for US9 of HSV-1 and its VZV counterpart, gene 65.

Much less of the genome of PRV has been characterized than HSV-1 and VZV. The viral DNA contains about 150 kb (73% G+C). Restriction maps of PRV DNA are shown in Fig. 1.3 (Ben-Porat and Kaplan, 1985). To date, only few genes have been identified in the PRV genome, and these are mainly located in the Us or Rs region. Table 1.6 lists these genes and their homologues in HSV-1 and VZV. IE180, the immediate-early gene identified in PRV, is homologous to IE175 of HSV-1, and expression of this gene is continually required during virus infection (Ihara *et al.*, 1983). On the basis of studies using a *ts* mutant, Ahler and Feldman (1987) suggested that protein IE180 is involved in forming a transcriptional complex for the expression of early genes , but not late genes.

Six glycoprotein genes of PRV have been identified. Genes, gX, g50, g63 and gI lie in the Us region, and are homologous to US4, US6, US7, and US8 of HSV-1 (Rea *et al.*, 1985; Petrovskis *et al.*, 1986a, 1986b). Genes gII and gIII lie in the UL region, and their products are homologous to glycoproteins B and C, the products of genes UL27 and UL44 (Robbins *et al.*, 1987, 1986a). It has been shown that genes gI, gX, g63 and gIII are not essential for virus growth in cells in

Fig. 1.3. Restriction maps of the pseudorabies virus genome. Each cleavage point is indicated by a vertical arrow, and each of the fragment is marked. The maps of *SalI* and *PstI* presented are imcomplete. The structure of the genome of PRV is presented, and thick horizontal lines and open boxes represent the unique and the repeat regions respectively, which are marked.


Hpa I
A
B
C
D
Xba I
A
B
C
Hin dIII
B
A
B'
C
Bgl II
C
F
B
A
D/E
Sal I
10
1
15 21 19
5
15
2
Kpn I
D
C
A
J
B
G
L
F
E
M
K
J
I
K
M
H
Pst I
14
3
6
4
26
1
6
7'
Bam HI
14
5'
1
2
9
11
4
15
3
6
14
8'
8
5
10
7
12
5
8
13
UL
IRs
Us
TRs
Map units
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

culture (Thomsen *et al.*, 1987; Mettenleiter *et al.*, 1985; Petrovskis *et al.*, 1986c; Robbins *et al.*, 1986b; Wathen and Wathen, 1986). Nevertheless, it has been found that several of these 'non-essential' genes may have important functions. A PRV mutant defective in either gI and gIII, or g63 and gIII prevented release of virus from rabbit kidney cells (Zsak *et al.*, 1989), and was avirulent for chickens (Mettenleiter *et al.*, 1988). Although gene gIII appears not to be essential for growth of virus in cells in culture, it is required for efficient growth (Whealy *et al.*, 1988). It is likely that glycoprotein III is involved in virus adsorption through an interaction between this viral protein and a heparin-like receptor on the surface of the host cell (Mettenleiter *et al.*, 1990; Zsak *et al.*, 1990). It has also been reported that a large proportion of PRV-specific cytotoxic T lymphocytes, induced by immunization of mice and pigs with PRV, are directed against glycoprotein gIII, implying that glycoprotein gIII is a major target antigen of PRV (Zuckermann *et al.*, 1990). The function of gene gII appears complex. It is certain that, unlike glycoprotein B of HSV-1, glycoprotein II is cleaved by a specific protease (Wölfer *et al.*, 1990; Whealy *et al.*,1990), and the cleaved segments forms an oligomer jointed by disulphide linkages (Lukacs *et al.*, 1985). The function of the gene g50 is much less understood, although expression of gene g50 in Vero, HeLa, and pig kidney cells results in a reduction of virus yield (Petrovskis *et al.*, 1988).

The other PRV gene that has been identified is that of thymidine kinase, homologous to UL23 of HSV-1 (Kit *et al.*, 1984). It should be mentioned that two other regions in the PRV genome have also been sequenced, but the products of the genes in these regions have not yet been identified. The first is a region downstream of gene gI, containing a small open reading frame, homologous to gene US9 of HSV-1 (Petrovskis and Post, 1987). The second to have been sequenced is the *Sal* I 3 fragment from PRV, strain Becker, the 3' end of which overlaps 132 bp of the 5' end of the gII coding region, and contains an open reading frame, the predicted

product of which is extensively homologous to that of UL28 of HSV-1 (Pederson and Enquist, 1988).

1.3. Protein kinases encoded by alphaherpesviruses

Protein kinases are a large family of enzymes which catalyse the transfer of a phosphate group from a nucleotide triphosphate, usually ATP or GTP, to the hydroxyl group of an amino acid residue in a protein. Protein kinases can be divided for convenience into two groups on the basis of the acceptor amino acid: seryl/threonyl and tyrosyl protein kinases.

Many changes occur during infection of cells by alphaherpesviruses. One of these involves the profile of phosphorylated proteins of both viral and cellular origin (Wilcox *et al.*, 1980; Kennedy *et al.*, 1981). These proteins might, in principle, be phosphorylated either by cellular protein kinases induced or activated during virus infection, or by viral enzymes. An example of the former case is a ribosomal protein S6 kinase in mouse fibroblasts, which is activated during infection by PRV (Jakubowicz and Leader, 1987a, 1987b).

The best example of a protein kinase encoded by a virus is the similar protein kinase encoded by HSV-1 and PRV (Katan *et al.*, 1985; Purves *et al.*, 1986b). This protein kinase was initially observed as an activity in cytoplasmic extracts of BHK cells infected with PRV (Katan *et al.*, 1985, 1986). Subsequently, a similar activity was found in cells infected with HSV-1 (Purves *et al.*, 1986b). In the latter study, Purves *et al.* (1986b) found that the kinase activity in BHK cells infected with either PRV or HSV-1 was detectable 4 h after infection, and increased continually until at least 10 h after infection. The kinetics of appearance of the activity in both cases were similar to those of the viral DNA polymerase, and the activity of kinase depended on the multiplicity of infection. The chromatographic properties of the enzymes from the two sources were different, suggesting that they might be encoded by the virus genomes. It had been found that the predicted product

of US3 of HSV-1 was related to cellular protein kinases (McGeoch and Davison, 1986a), and experiments were performed to determine whether US3 encoded the activity detected in cells infected with HSV-1. First, using a US3$^-$ mutant (R7041), Purves *et al.* (1987b) demonstrated that the protein kinase found in cells infected with wild-type HSV-1 was absent from cells infected with the mutant. Second, experiments were performed using a rabbit antiserum raised against a synthetic peptide corresponding to the C-terminus of the predicted product of US3 (Frame *et al.*, 1987). It was found that the antiserum specifically reacted with a 68kDa protein though the purification of the protein kinase from the extract of cells infected with HSV-1 was incomplete. Taken together, these results showed that the protein kinase purified from cells infected with HSV-1 was the product of the gene US3.

Some properties of the protein kinases (designated PRV protein kinase, PRV-PK, and HSV protein kinase, HSV-PK) are listed in Table 1.7 (for reviews see Leader and Katan, 1988; Leader and Purves, 1988, and for original references see Katan *et al.*, 1985; Purves *et al.*, 1986a, 1986b, 1987a). PRV-PK utilizes only ATP, but not GTP, as the phosphate donor, and prefers seryl/threonyl residues of basic artificial substrates (*e.g.* protamine), rather than acidic substrates (*e.g.* casein). The kinase activity is not affected by molecules such as cyclic AMP, cyclic GMP, Ca^{2+}/calmodulin, and Ca^{2+}/phospholipid. A striking feature of both PRV-PK and HSV-PK is that the kinases are active even in the presence of 1 M potassium chloride (Purves *et al.*, 1986b). PRV-PK has been purified to homogeneity, and the enzyme is autophosphorylated when the purified PRV-PK is incubated with [γ-^{32}P]ATP in the absence of substrate (Purves *et al.*, 1987a). The enzyme is thought to be a homodimer, since the value of the molecular weight obtained from gel-exclusion chromatography is approximately twice that obtained from polyacrylamide gel electrophoresis under denaturing conditions. The molecular weight of the subunit was shown to be 38,000, and the isoelectric point

Table 1.7. Properties of the alphaherpesvirus protein kinasses

	PRV-PK	HSV-PK
Molecular weight	38,000 (dimer)	68,000 (dimer)
Isoelectric point	4.9	5.6
Target residue(s)	Ser/Thr	Ser/Thr
Phosphate donor	ATP	ATP
Artificial substrates	protamine/histone	protamine/histone
Physiological effector	none?	none?
KCl optimum (mM)	100-1000	100-1000
Autophosphorylation	Yes	Yes
Best synthetic peptide substrate	R_4ASVA	R_4ASVA

approximately 4.9. HSV-PK has not yet been purified to homogeneity. On the basis of autophosphorylation of one of the two components of the most highly purified preparation of HSV-PK, the molecular weight of the subunit and its isoelectric point were concluded to be approximately 68,000 and 5.9, respectively, and the native enzyme was concluded to be a homodimer (Purves, 1987). The substrate specificities of PRV-PK and HSV-PK were determined using artificial peptides, and both PRV-PK and HSV-PK had a preference for peptides with several arginyl residues on the N-terminal side, rather than the C-terminal side, of the phosphorylation site (Purves *et al.*, 1986a). A further more detailed study has shown that peptides of the type R_4ASVA serve as the best substrates for both PRV-PK and HSV-PK, and at least three arginyl residues on the N-terminal side of the phosphorylation site are required for optimal activity (Leader *et al.*, 1991). The viral enzymes are specific for substrates with arginyl residues, since they were unable to phosphorylate a peptide with replaced Arg by Lys in R_4ASVA. It was clear from these studies that the viral protein kinases have a similar substrate specificity which differs from the cellular protein kinases, protein kinase C and the cAMP-dependent protein kinase.

In contrast to its equal ability to infect cells in culture (Purves *et al.*, 1987b), the US3$^-$ mutant virus was much less virulent than the wild type when injected into mice, and had a lower ability to establish latency (Meignier *et al.*, 1988). On the basis of these studies, Leader and Purves (1988) suggested that a cellular homologue of the viral protein kinases might exist in cells in culture. If this is so, the cellular protein kinase would be expected to share unique structural features with the viral enzymes.

The eukaryotic cellular protein kinase family includes many individual members. Although the size and subunit structure of the enzymes vary widely, a striking common feature shared by these is a homologous catalytic domain with

several highly conserved amino acid motifs (for review, see Hanks *et al.*, 1988 and Taylor *et al.*, 1990). This was first pointed out by Barker and Dayhoff (1982), who found that the catalytic subunit of $pp60^{v\text{-}src}$, a tyrosyl protein kinase, and the cAMP-dependent protein kinase shared a number of amino acid identities, which were subsequently found to be a common feature of protein kinases (Fig. 1.4). The motifs include a glycine-rich loop (G-G--G) which resembles the nucleotide binding site in certain other proteins (Rossmann *et al.*, 1974; Pai *et al.*, 1989). A model for the ATP-binding site of the cAMP-dependent protein kinase, based on the crystal structure of lactate dehydrogenase, has been constructed in which the G-G--G sequence forms an elbow around ATP (Taylor *et al.*, 1990). In this model Lys^{72} of the cAMP-dependent protein kinase is close to the G-G--G loop and the γ-phosphate of ATP, which is consistent with the results of fluorescent displacement study using *lin*-benzo-ATP (Bhatnager *et al.*, 1984), and chemical modification with *p*-fluorosulfonylbenzoyl 5'-adenosine or acetic anhydride (Zooler *et al.*, 1981; Buechler *et al.*, 1989). It was suggested that residues Glu^{91} and Asp^{184} of the enzyme are also involved in ATP-binding, as these residues were protected by ATP from modification with dicyclohexylcarbodiimide (Buechler and Taylor, 1988). This was subsequently confirmed when it was shown that Asp^{184} can be cross-linked to Lys^{72} (Buechler and Taylor, 1989).

Some amino acid residues can be protected from chemical labelling by the substrate, or both substrate and MgATP, suggesting that these are involved in substrate binding or enzyme activity. It was observed that the residue Glu^{170} in the cAMP-dependent protein kinase, in subdomain VI of Hanks *et al.* (1988), was very sensitive to modification by 1-ethyl-3 (3-dimethyl-amino-propyl)-carbodiimide modification in the absence of MgATP and an inhibitor peptide, although this residue was fully protected in the presence of MgATP and the inhibitor (Buechler and Taylor, 1990). It is clear that Glu^{170} interacts with the substrate. This Glu is the key residue for the substrate specificity of protein kinases since all of the

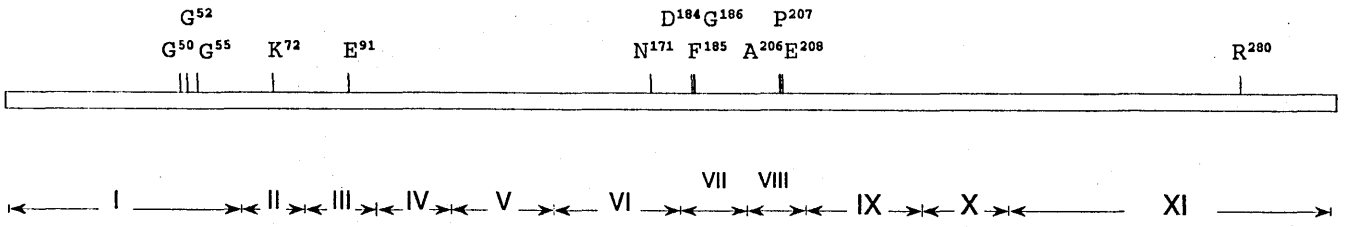


Fig. 1.4. Conserved amino acid residues in the catalytic domain of eukaryotic protein kinases. The numbering of the residues (one–letter code) is based on the cAMP-dependent protein kinase (Taylor *et al.*, 1990). Subdomains I-XI of Hanks *et al.* (1988) are indicated.

seryl/threonyl protein kinases with an acidic residue at the corresponding position have a preference for basic amino acid residues near phosphorylation site, *e.g.* cAMP-dependent protein kinase (Taylor *et al.*, 1990), protein kinase C (Knopf *et al.*, 1986), HSV-PK (McGeoch and Davison, 1986a; Purves *et al.*, 1986a). Instead of an acidic residue, casein kinase II has a His at the corresponding position, preferring peptides with acidic residues near the phosphorylation site (Kuenzel *et al.*, 1987). In tyrosyl protein kinases this Glu is replaced by either Arg or Ala (Hanks *et al.*, 1988), and these enzymes also prefer three to four acidic residues preceding the phosphorylation site (Hunter, 1982).

The conserved motif, APE, in subdomain VIII of Hanks *et al.* (1988), corresponding to $Ala^{206}Pro^{207}Glu^{208}$ of the cAMP-dependent protein kinase, has been shown to be a important region for the catalytic activity of protein kinases. Bryant and Parsons (1983) demonstrated that each of the residues in this motif was required for activity of $pp60^{v\text{-}src}$, although only the residue Glu in this motif is totally conserved in protein kinases (Hanks *et al.*, 1988). It should also be mentioned that an Arg near the C-terminus of the catalytic domain, corresponding to Arg^{280} in the cAMP-dependent protein kinase, is another conserved residue in the protein kinase family. Lörincz and Reed (1986) showed that in *Saccharomyces cerevisiae* CDC28 protein kinase replacement of the Arg with Glu caused the activity to become temperature-sensitive.

Of particular interest to this study is the homology of the viral protein kinases to cellular protein kinases. The predicted amino acid sequence of US3 of HSV-1 has most of these highly conserved motifs. However, HSV-PK is not closely related to any individual eukaryotic protein kinase for which the sequence has been determined, as can be seen from Fig. 3 of Hanks *et al.* (1988), which is reproduced in Fig. 1.5.

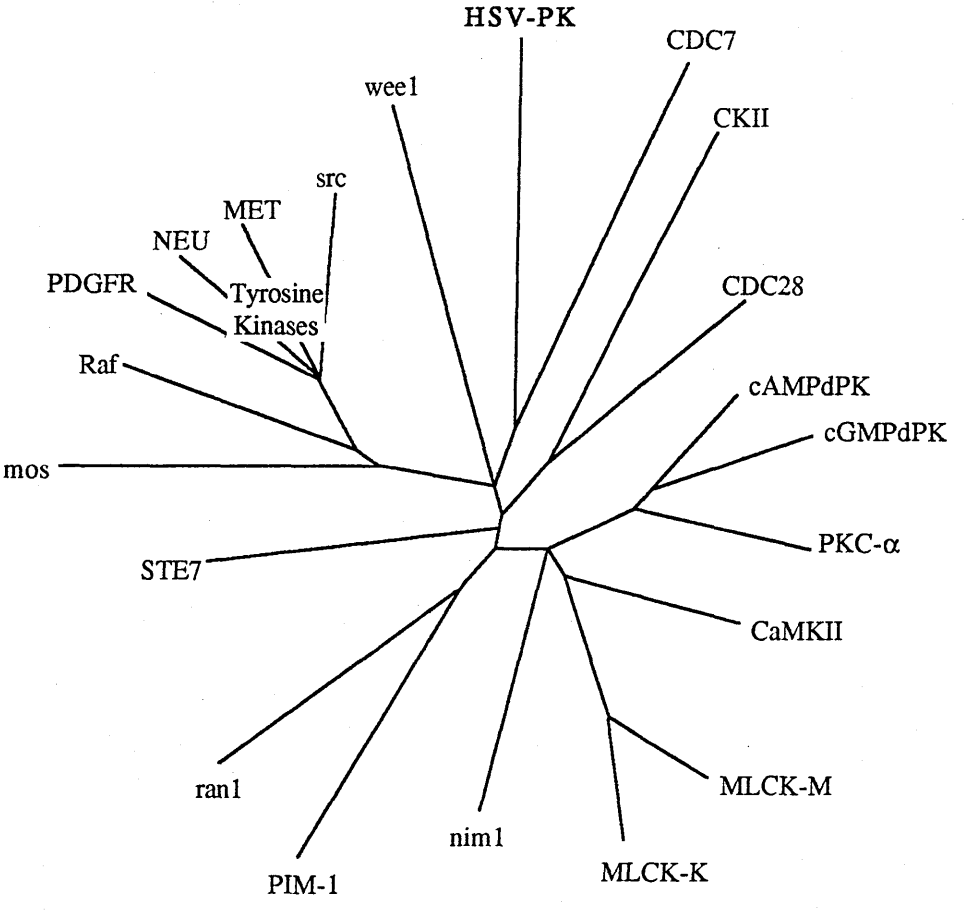


Fig. 1.5. Deduced phylogenetic relationship of the catalytic domain of protein kinases. This is a simplication of the tree derived by Hanks *et al.* (1988), which is shown 'unrooted'.

The abbreviations for the protein kinases are: CKII, casein kinase II; cAMPdPK, cyclic AMP-dependent protein kinase; cGMPdPK, cyclic GMP-dependent protein kinase; PKC-α, a variant of protein kinase C; CaMKII, calmodulin-dependent protein kinase II; MLCK-M and MLCK-K, variants of myosin light chain kinase; PDGFR, platlet-derived growth factor receptor. All other designations are of genetic loci, rather than abbreviations.

1.4. Objectives

When this work was initiated, PRV-PK had been completely purified (Purves *et al.*, 1987a), but the gene encoding this enzyme had not been determined. The initial object of this work was to identify the gene for PRV-PK in order to confirm that, like HSV-PK, it is encoded in the viral genome. This would also be useful for definition of common features of the herpesvirus protein kinase, which might help identify specific structure features of the viral protein kinases.

In order to define the gene encoding PRV-PK it became necessary in the course of this work to determine the nucleotide sequence of the adjacent region of the PRV genome. Part of this thesis deals with the identification of the Us/Rs junctions and the adjacent gene in the Rs region of the PRV genome.

CHAPTER 2
MATERIALS AND METHODS

2.1. Chemicals

Chemicals were AnalaR Grade purchased from BDH Chemicals Ltd or Fisons plc (FSA Laboratory Supplies), unless indicated in the text.

2.2. Media and antibiotics for bacterial culture

All of the solutions and media used for the growth of bacteria were sterilized by either autoclaving at 15 pounds square inch for 15 min, or (for heat-sensitive compounds such as antibiotics or vitamins) filtration through a 0.2 μm pore-size filter (Flow Research).

2.2.1. Liquid media

Luria broth (L-broth) was routinely used for the growth of bacteria except where indicated in the text, and contained 1% (w/v) bacto-tryptone (Difco 0123-01), 0.5% (w/v) yeast extract (Difco 1880-17), and 0.5% (w/v) NaCl. The pH was adjusted to 7.2 with 5 M NaOH. Antibiotics were added if required.

2x TY broth, a richer medium than L-broth for the culture of the bacteriophage M13, contained 1.6% (w/v) bacto-tryptone, 1% (w/v) yeast extract, and 0.5% (w/v) NaCl.

10x Hogness modified freezing medium (for 100 ml) contained 6.3 g

K_2HPO_4, 1.8 g KH_2PO_4, 0.45 g Na-citrate, 0.09 g $MgSO_4 \cdot 7H_2O$, 0.9 g $(NH_4)_2SO_4$, and 44 ml of glycerol. For use, one volume of this solution was mixed with nine volumes of a fresh overnight bacteria culture.

2.2.2. Agar plates

In general agar plates were prepared by dissolving 1.5% (w/v) bacto-agar (Difco 0140-01) in the particular growth medium, and pouring aseptically into sterile petri dishes.

Glucose/Minimal plates, used to store JM109 cells, were prepared from the following solutions, sterilized separately: 250 ml of 3% (w/v) bacto-agar, 0.5 ml of 1 M $MgSO_4$, 0.5 ml of 0.1 M $CaCl_2$, 0.5 ml of 1 M thiamine-HCl (Vit. B1), 2.5 ml of 40% (w/v) glucose, and 250 ml of M9 salts (3 g Na_2HPO_4, 1.5 g KH_2PO_4, 0.25 g NaCl, and 0.5 g NH_4Cl).

LB plates, routinely used for bacteria culture, were prepared by dissolving 1.5% (w/v) bacto-agar in L-broth with or without antibiotics, as required.

H plates, used for bacteriophage M13 culture, was prepared from 1% (w/v) bacto-tryptone, 0.8% (w/v) NaCl, and 1.2% (w/v) bacto-agar for the bottom layer, and 1% (w/v) bacto-tryptone, 0.8% (w/v) NaCl, and 0.8% (w/v) bacto-agar for the top layer.

Hammersmith stabs, used for storage of bacteria, contained 0.9% (w/v) nutrient broth (Difco 0003-02), 0.75% (w/v) bacto-agar, 0.5% (w/v) NaCl, and 0.1 mg/ml thymine (Sigma T 0376).

2.2.3. Antibiotics and inhibitors

Ampicillin (Sigma, A 9518) was a stock solution of 5 mg/ml in distilled water at -20°C. For routine bacterial growth it was used at a final concentration of 30 μg/ml.

Chloramphenicol (Sigma, C 0378) was stored as 25 mg/ml in 50% (v/v)

ethanol at 4°C, and for the amplification of plasmid pUC in bacterial cells, was used at a final concentration of 165 μg/ml.

Tetracycline (Sigma, T 3383) was stored at -20°C in a dark bottle at a concentration of 20 mg/ml in 50% (v/v) ethanol, and it was used at a final concentration of 20 μg/ml for liquid culture, or 13 μg/ml in agar plates.

Cycloheximide (Sigma, C 6255) was stored as a solution of 10 mg/ml in distilled water at -20°C, and used at a final concentration of 25 μg/ml.

2.3. Bacteria

2.3.1. Bacteria strains

The strains of *Escherichia coli* used in the work described in this thesis were as follows:

DH1 has the genotype of F^-, *recA*1, *endA*1, *gyrA*96, *thi*, *hsdR*17, *supE*44, *relA*1, and λ^- (Low, 1968; Meselson and Yuan, 1968).

JM109 has the genotype of *recA*1, *endA*1, *gyrA*96, *thi*, *hsdR*17, *supE*44, *relA*1, λ^-, *Δ(lac-proAB)*, and [F',*traD*36, *proAB*, *lacI*^{q}*ZΔ*M15] (Yanish-Perron *et al.*, 1985).

MC102 is a *lon*$^-$strain, originally obtained from the *E. coli* Genetic Stock Centre, Department of Human Genetics, Yale University, New Haven, Conn. the United States of America. Other details of the genotype are not cited (Bruner *et al.*, 1988).

2.3.2. Preparation of bacterial cells competent for transformation and transfection

A single colony was transferred to a 250 ml flask containing 50 ml of L-broth (section 2.2.1), the cells were grown at 37°C in an orbital shaker overnight. 2.5 ml of the overnight culture of bacterial cells were transferred into a

2 litre flask with 500 ml of fresh L-broth. This flask was incubated at 37°C with shaking. When the OD_{550} of the culture reached 0.80-0.89 (approx. 3 h), the cells were harvested by centrifugation in a Beckman JA–14 rotor at 5,000 rpm for 15 min at 4°C. The supernatant was immediately removed, and the cells were resuspended in 250 ml of pre-cooled 100 mM $CaCl_2$, which was equivalent to half the volume of the original medium. The suspension were left on ice for 20 min, re-centrifuged in the same rotor at 6,000 rpm for 10 min at 4°C, and then the cells were resuspended in 5 ml of pre-cooled 100 mM $CaCl_2$ at 4°C. The resulting bacterial suspension was mixed with glycerol to give 10% (v/v) final concentration, and frozen in liquid nitrogen. The frozen cells were stored at -80°C, where they retained activity for at last a year.

2.3.3. Storage of bacterial cells

Bacterial cells were stored in several ways:

For long-term storage (several years), bacteria were frozen in Hogness modified freezing medium, and stored at -80°C.

For medium-term storage (up to one year), bacteria were stored as Hammersmith stabs at room temperature.

For short-term storage (up to four weeks), bacteria were kept on an appropriated agar plate at 4°C.

2.4. Bacteriophages M13

Bacteriophage M13 strains mp18 and mp19 (Yanisch-Perron *et al.*, 1985) were used in this study for cloning DNA fragments for determination of nucleotide sequences, and grown on strain JM109.

2.5. Cell culture and virus infection

2.5.1. Virus and cell

Pseudorabies virus, used for this study was Ka strain, originally obtained from Kaplan and Vatter (1959), via Dr. W.S. Stevely. The virus was grown in BHK 21/C13 cells, an established cell line of baby hamster kidney fibroblasts (MacPherson and Stoker, 1962).

2.5.2. Cell culture

BHK 21/C13 cells were grown at 37°C in 2.5 litre roller bottles containing 180 ml of the Glasgow modification of Eagle's minimum essential medium, supplemented with 4 mM L-glutamine, 0.2% (w/v) Na_2CO_3, 100 units/ml penicillin, 100 µg/ml streptomycin, 10% (v/v) calf serum, and 2.5% (v/v) tryptose phosphate broth. As soon as the cells had been transferred into the growth medium (2x 10^7 cells in 180 ml of medium), the suspension was 'gassed' with CO_2 (95% air/5% CO_2 final concentration). The cells were grown at 37°C to 95% confluence before use. This usually took 3 days.

For subculture of the cells, the medium was poured off, and the cells were washed with freshly prepared trypsin/versene (1:4). The trypsin/versene wash was repeated, and then the cells were replaced at 37°C in 1-2 ml of the solution. As soon as the cells started to dislodge from the wall, the trypsin was diluted out with 20 ml of fresh growth medium. The cells were rinsed, and poured into a fresh bottle. This cell-suspension could be stored at 4°C for 24 h.

2.5.3. Virus infection

BHK cells at 95% confluence in roller bottles were used for infection. The medium was removed, and replaced by 20 ml of original medium to which PRV

had been added. After incubation at 37°C for 1 h at a multiplicity of infection of approximate 20 pfu per cell, this medium was removed and replaced by 90 ml of the original growth medium. The infection was allowed to proceed at 37°C for up to 6 h.

2.6. Vectors

2.6.1. Vectors for cloning and sequencing

The bacterial plasmid vectors used for cloning and sequencing in this study were pUC8 and pUC18. The two pUC vectors differ in restriction endonuclease sites in their multiple cloning region (Fig. 2.1.a and c).

The M13 vectors, M13mp18 and mp19, were used for the DNA sequencing. These differ in the orientation of their multiple cloning region (Fig. 2.1.b and c).

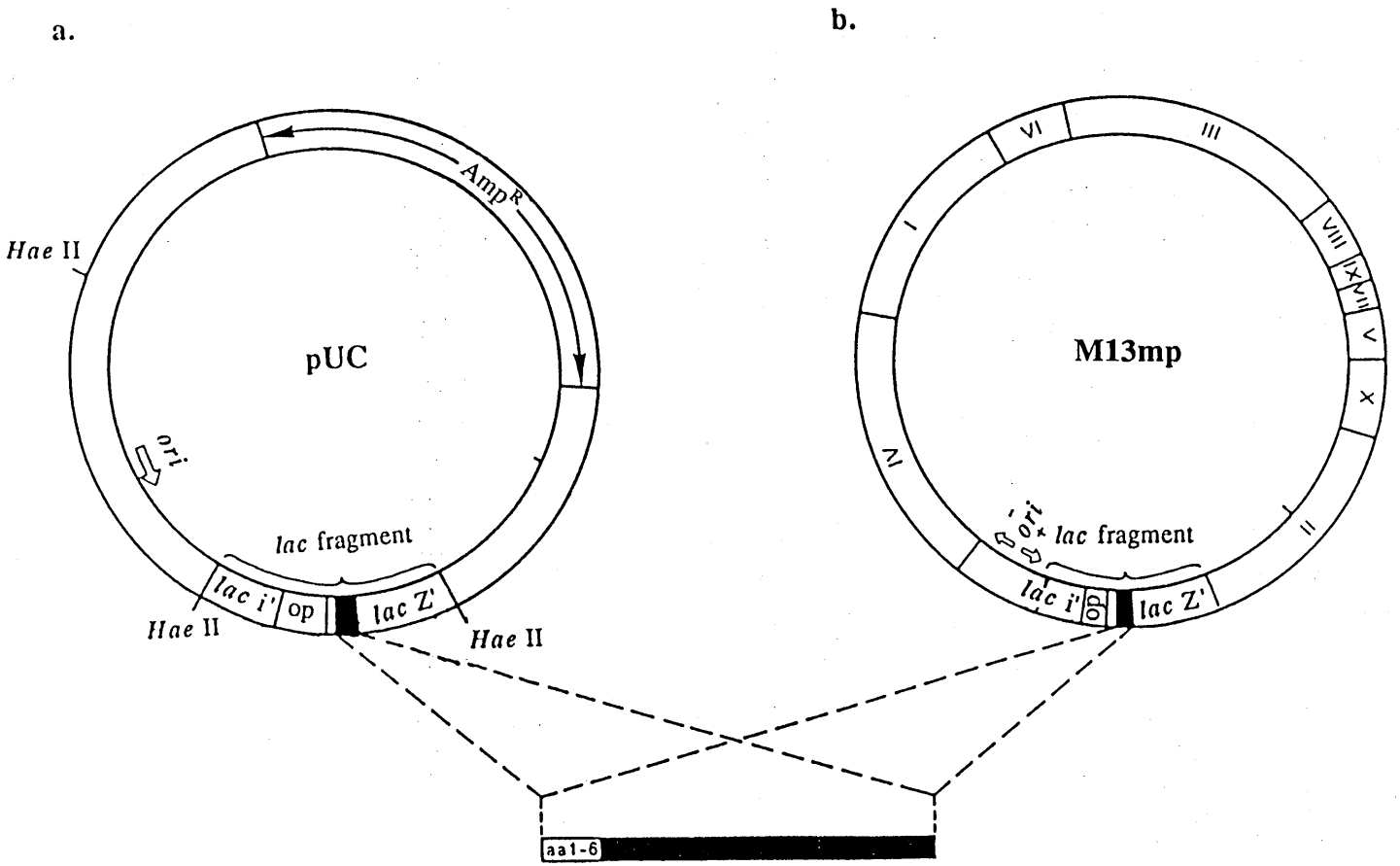
2.6.2. Expression vectors

Two prokaryotic expression vectors were used in attempt to express the PRV *pk* gene in *E. coli.*

The first was pKK233-2, which contains the *tac* promoter and an *Nco*I site overlapping the translational initiation codon which was preceded by a strong ribosome-binding site (Amann and Brosius, 1985). The structure of the vector is shown in Fig. 2.2.a.

The second was pJLA502 which contains the bacteriophage λ promoters P_R and P_L, the gene for the temperature-sensitive mutant of the cI repressor, the bacteriophage fd-transcription terminator, and a region of DNA containing an *Nco*I site at the translational initiation site between the promoter and terminator (Schauder *et al.*, 1987). The structure of this vector is presented in Fig. 2.2.b.

Fig. 2.1. Structures of cloning and sequencing vectors. (a) Map of pUC vectors. The multiple cloning region (which differs from different member of the series) lies in the *lacZ'* (truncated β-galactosidase) coding region, and the ampicillin resistant gene is marked as Amp^R. (b) Map of M13 vectors. M13mp18 and M13mp19 contain the same *lac* fragment as pUC vectors, and the bacteriophage genes I-X are indicated. The multiple cloning region of M13mp18 is the same as that of pUC18, but is in the opposite orientation in M13mp19. (c) Nucleotide sequences of the multiple cloning regions. The recognition sites of restriction endonucleases presented are those of pUC8, pUC18/M13mp18 and M13mp19, and the nucleotide sequences of the forward sequencing primer are shown. The reverse sequencing primers, however, are only shown in pUC8. Frames (a) and (b) are from Adams *et al.* (1986).



Multiple cloning region

c.

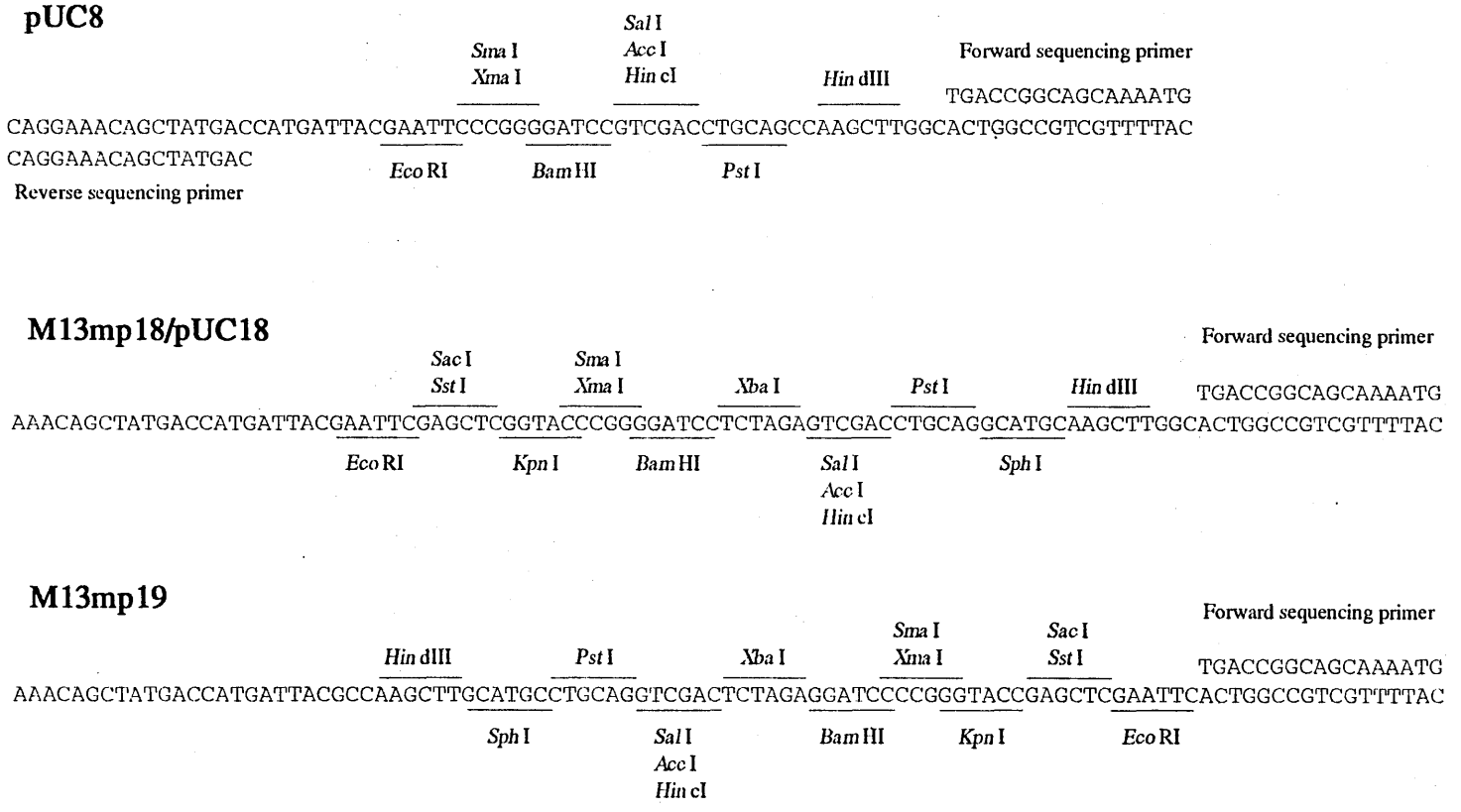
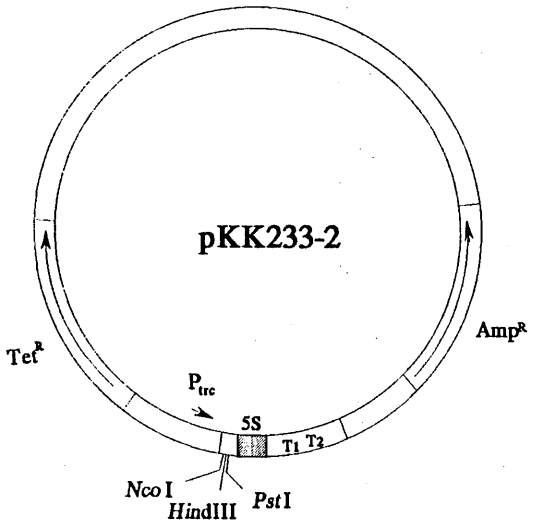
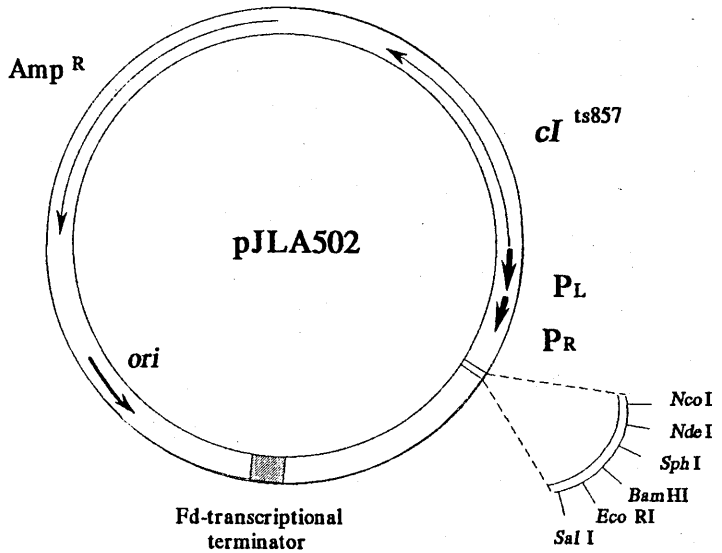

pUC8
Sma I
Xma I
Sal I
Acc I
Hin cI
Hin dIII
Forward sequencing primer
TGACCGGCAGCAAAATG
CAGGAAACAGCTATGACCATGATTACGAATTCCCGGGGATCCGTCGACCTGCAGCCAAGCTTGGCACTGGCCGTCGTTTTAC
CAGGAAACAGCTATGAC
Eco RI
Bam HI
Pst I
Reverse sequencing primer
M13mp18/pUC18
Forward sequencing primer
Sac I
Sst I
Sma I
Xma I
Xba I
Pst I
Hin dIII
TGACCGGCAGCAAAATG
AAACAGCTATGACCATGATTACGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTAC
Eco RI
Kpn I
Bam HI
Sal I
Acc I
Hin cI
Sph I
M13mp19
Forward sequencing primer
Hin dIII
Pst I
Xba I
Sma I
Xma I
Sac I
Sst I
TGACCGGCAGCAAAATG
AAACAGCTATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCACTGGCCGTCGTTTTAC
Sph I
Sal I
Acc I
Hin cI
Bam HI
Kpn I
Eco RI

Fig. 2.2. Structure of two prokaryotic expression vectors. (a) Expression vector pKK233-2 (Amann and Brosius, 1985). The vector is 4602 bp in length, and includes tetracycline and ampicillin resistant genes which are presented as Tet^R and Amp^R. The position of the *tac* promoter and the cloning region are shown. The two transcriptional terminators are also indicated. (b) Expression vector pJLA502. The size of the vector is 4.9 kb, and it contains bacteriophage lambda promoters P_R and P_L. The gene for the temperature-sensitive mutant of the cI repressor is indicated as cI^{ts857}, and the bacteriophage fd-transcription terminator is also presented. The multiple cloning region is shown as a open box, beside which the recognition sites of restriction endonucleases are indicated.

a.


pKK233-2
Tet^R
Amp^R
P_trc
5S
T1 T2
Nco I
HindIII
Pst I

b.


Amp R
cI ts857
pJLA502
P_L
P_R
ori
Fd-transcriptional terminator
Nco I
Nde I
Sph I
BamHI
Eco RI
Sal I

2.7. Agarose gel electrophoresis of DNA

Separation of DNA by agarose gel electrophoresis was performed in TBE buffer (90 mM Tris, 90 mM boric acid, and 2.5 mM EDTA; pH 8.3), to which ethidium bromide (a final concentration of 0.5 μg/ml) was added. The DNA samples were mixed and loaded with at least 0.1 volume of loading buffer (1% (v/v) bromophenol blue (10 mg/ml), 49% (v/v) TE buffer (10 mM Tris-HCl and 1 mM EDTA; pH 8.0), and 50% (v/v) glycerol), and the electrophoresis was performed at 40-100 V (for a 12x 11 cm gel) for an appropriated time, depending on the size of the DNA separated on the gel, and the percentage of agarose. The size markers routinely used were pUC18 digested with *Eco*RI and with *Rsa*I which gave fragments of 2686, 1739, 676 and 271 bp, and Lambda DNA digested with *Hin*dIII which gave fragments of 23.7, 9.46, 6.61, 4.26, 2.26, 1.98, and 0.58 kb.

2.8. Preparation of DNA

2.8.1. Preparation of plasmid DNA

(a). Small scale preparation (Mini-prep)

For rapid preparation of small amounts of DNA, a modification of the procedure of Holmes and Quigley (1981) was employed.

A single bacterial colony (transformed by a plasmid) was spread thoroughly onto an antibiotic plate, and allowed to grow at 37°C overnight. The bacteria were carefully scraped off the plate and resuspended in 1 ml of STET buffer (8% (w/v) sucrose, 50 mM Tris-HCl; pH 8.0, 50 mM EDTA, 5% (v/v) Triton X-100) in a 1.5 ml Eppendorf tube. The suspension was mixed with 10 μl of lysozyme (20 μg/ml in distilled water, stored at -20°C in aliquots), and then incubated in a boiling water both for 7 min, followed by centrifugation for 15 min.

A 0.6 ml portion of the supernatant was transferred into a fresh tube containing 2 µl of boiled RNase (1 mg/ml), and the tube was placed at 37°C for 15 min. After mixing with 1 µl of diethylpyrocarbonate (Sigma D-5758), the sample was incubated at 65°C for 15 min in order to denature the proteins. The DNA was precipitated by adding 0.24 ml of 5 M ammonium acetate and 0.54 ml of isopropanol, and then was kept at -70°C for 5 min. The DNA was collected by centrifugation for 15 min in an Eppendorf centrifuge, and washed with 0.3 M ammonium acetate/70% (v/v) isopropanol, and then with cold 100% ethanol. The DNA was briefly dried under vacuum, dissolved in 50 µl of TE buffer (section 2.7), and stored at -20°C.

(b). Preparation by alkaline lysis

This method used in the study was the modification of the methods of Birnboim and Doly (1979) and Ish-Horowicz and Burke (1981). The method was used for small scale preparation of plasmid DNA, and sometimes also for large scale preparation by increasing the volume of the solutions.

A single colony was transferred to 5 ml of appropriate medium with 30 mg/ml ampicillin in a universal flask, and bacteria were allowed to grow at 37°C overnight in an orbital shaker.

The bacteria from 1.5 ml of such an overnight culture were collected by centrifugation for 1 min in an Eppendorf centrifuge, and resuspended in 100 µl of ice-cold Solution I (50 mM glucose, 25 mM Tris-HCl; pH 8.0, and 10 mM EDTA; pH 8.0) by vortexing. The resulting suspension was mixed with 200 µl of freshly prepared Solution II (0.2 N NaOH and 1% (w/v) sodium dodecyl sulphate) by inverting the tube several times. After 5 min on ice, 150 µl of ice-cold 5 M sodium acetate, pH 5.0 was added, and the solutions mixed by vortexing gently. The sample was kept on ice for 5 min, followed by centrifugation for 5 min. The supernatant was transferred to a fresh tube and then vortexed with an equal volume

of a mixture of phenol:chloroform (1:1). The aqueous phase was separated by centrifugation and transferred to a fresh tube. The plasmid DNA was precipitated with 2 volumes of 100% ethanol for 2 min at room temperature. After centrifugation for 5 min in an Eppendorf centrifuge, the DNA was washed with 70% ethanol, dried briefly under vacuum, and resuspended in 100 µl of distilled water with 2 µl of DNase-free pancreatic RNase (1 mg/ml). The DNA was incubated at 37°C for 30 min, and re-precipitated with 60 µl of 2.5 M NaCl/20% (w/v) polyethylene glycol 6000 (Serva) at 0°C for 1 h. The DNA was sedimented by centrifugation in an Eppendorf centrifuge for 10 min, washed with 70% ethanol, and dried under vacuum. The DNA precipitate was re-dissolved in TE buffer (section 2.7) and stored at -20°C.

(c). Large scale preparation

The method for large scale preparation of DNA was that of Birnboim and Doly (1979).

A single colony was transferred to a 250 ml flask with appropriate medium and ampicillin at 30 µg/ml final concentration, and cells were allowed to grow at 37°C overnight in an orbital shaker. Next day 2.5 ml of fresh overnight culture were transferred into a 2 litre flask containing 500 ml of the medium, and cells were grown at 37°C until the OD_{600} of the bacterial culture reached 0.8. Chloramphenicol (165 µg/ml final concentration) (section 2.2.3) was then added, and the flask incubated at 37°C overnight in the shaker.

The bacteria were the harvested by centrifugation at 5,000 rpm for 20 min at 4°C. They were resuspended in 9.5 ml of Solution I (section 2.8.1.b), and transferred into a ice-cold Ti60 tube (Beckman). The suspension was mixed with 0.5 ml of lysozyme (40 µg/ml in Solution I), and left on ice for 30 min. The mixture was divided into two portions, each in a separate tube, and 10 ml of

Solution II (section 2.8.1.b) was added into each tube and mixed. After 5 min on ice, the suspension was mixed with 7.5 ml of 3 M sodium acetate (pH 5.0) in each tube by inverting several times, and then incubated on ice for 60 min. The bacterial debris was removed by centrifugation at 30,000 rpm in a Beckman Ti60 rotor at 4°C for 30 min. The supernatant was transferred to a Corex tube, and the nucleic acid was precipitated by adding 0.6 volume of isopropanol. After 10 min at room temperature, the nucleic acid was collected by centrifugation at 8,000 rpm for 15 min in a Sorvall SS-34 rotor at room temperature.

The DNA was briefly dried under vacuum and re-dissolved in 30 ml of TE buffer (section 2.7). Cesium chloride (28.9 g) and 1.8 ml of ethidium bromide (10 mg/ml) were added, and the mixture clarified by centrifugation in a Beckman 'Table Top' centrifuge at 1,500 rpm for 30 min. The solution was transferred to a Beckman VTi50 tube, and centrifugation was performed in a Beckman VTi50 rotor at 20°C at 50,000 rpm for 20 h. Two bands were usually seen; the upper one contained the nicked circular and linear form of the plasmid, and the lower one containing the closed circular plasmid DNA. The lower band was collected by piercing the tube with a 21 g needle, and transferred to a Beckman VTi65 tube. After centrifugation for 20 h at 65,000 rpm in a Beckman VTi65 rotor at 20°C, the plasmid DNA band was removed as previously.

Ethidium bromide was removed from the DNA solution by several extractions with an equal volume of isoamyl alcohol (3-methyl-1-butanol, Sigma). The solution was diluted with 4 volume of TE buffer (section 2.7), and the DNA was precipitated by 2.5 volumes of cold 100% ethanol. The suspension was placed at -70°C for 15 min (or left at -20°C overnight), and the DNA was obtained by centrifugation at 10,000 rpm in a Sorvall HB4 rotor for 20 min. The precipitation was re-dissolved in TE buffer (section 2.7), re-precipitated, washed with 70% ethanol, and briefly dried under vacuum. Finally, the DNA was re-dissolved in TE buffer, and stored at -20°C. The quantity of the DNA was measured by its

absorbance at 260 nm, using the relationship that double-stranded DNA at a concentration of 50 µg/ml has an A_{260} value of 1 in a cell with a 1 cm length path.

2.8.2. Preparation of M13 replicative form DNA

Bacteriophage M13mp18 or M13mp19 (10-500 pfu per cell) was added to 0.1 ml of fresh overnight culture of JM109, and then this suspension was mixed with 120 µl of 0.5% (w/v) IPTG/0.5% (w/v) X-gal and 3 ml of H-top agar (section 2.2.2) (kept at about 45°C). The mixture was briefly and quickly shaken and poured onto H plates (section 2.2.2). The plates were incubated at 37°C overnight.

A single plaque was removed into a 250 ml flask containing 50 ml of 2x TY broth (section 2.2.1) with 0.5 ml of fresh overnight culture. The culture was shaken on an orbital shaker at 37°C for at least 5 h, and the cells were harvested by centrifugation in a Sorvall SS-34 rotor at 5,000 rpm for 15 min. The supernatant was re-centrifuged, transferred into a sterilized 30 ml Corex tube, and stored at 4°C overnight. Next day, 5 ml of this phage stock were transferred in a 2 litre flask containing 500 ml of 2x TY broth (section 2.2.1) with 5 ml of fresh overnight culture. The cells were grown up as described above, and then collected in a Beckman JA-14 rotor at 5,000 rpm for 15 min for preparation of the M13 replicative form DNA.

Bacteriophage M13 replicative form DNA from bacteria infected with either strain mp18 or mp19 was prepared as described in section 2.8.1.c.

2.8.3. Preparation of single-stranded bacteriophage M13 DNA

A single bacteriophage plaque was transferred into a universal flask containing 1.5 ml of 2x TY broth (section 2.2.1) with 15 µl of fresh overnight culture of bacteria, and this culture was shaken on an orbital shaker at 37°C for at least 5 h. The cells were sedimented by centrifugation in an Eppendorf centrifuge for 5 min, and the supernatant was re-centrifuged.

The re-centrifuged supernatant (1 ml) was transferred into a fresh tube with 0.2 ml of 2.5 M NaCl/20% (w/v) polyethylene glycol 6,000 (Serva), mixed,and incubated at room temperature for 15 min. The phage particles were sedimented by centrifugation twice in an Eppendorf centrifuge for 5 min, and the supernatant was carefully and completely removed. The precipitate was resuspended in 0.1 ml of TE buffer (see section 2.7), and extracted with phenol:chloroform (1:1) by vortexing for at least 30 sec, incubating at room temperature for 15 min, and re-vortexing for 30 sec. The aqueous phase was carefully removed avoiding the interphase material, and the DNA was precipitated by adding 0.1 volume of 3 M sodium acetate; pH 6.0 and 2.5 volumes of 100% ethanol. After 15 min on dry ice, the DNA was collected by centrifugation in an Eppendorf centrifuge for 10 min, dried briefly under vacuum, and re-dissolved in 25 ml of TE buffer (section 2.7). This single-stranded DNA was stored at -20°C.

2.9. Isolation of DNA fragments from low melting point agarose gel

Low melting point agarose gel can be melted at 65°C which is below the Tm of most nucleic acids, which can thus be isolated from the gel.

Gel electrophoresis (section 2.7) of the DNA was performed at a voltage no greater than 75V. The part of the gel containing the DNA fragment of interest was excised under long wavelength ultraviolet light, transferred into an Eppendorf tube, and one volume of TE buffer (section 2.7) was added. The agarose was melted at 65°C for 10 min, and extracted three times with an equal volume of phenol (saturated with TE buffer). The DNA was precipitated as described in section 2.8.3, re-dissolved in TE buffer, and stored at -20°C.

2.10. DNA subcloning

2.10.1. Subcloning of restriction fragments

(a). Digestion with restriction endonucleases

Generally, 1.5-5 units of enzyme were used to digest 1 μg DNA (*e.g.* pUC18), and the volume of the enzyme added was not allowed to exceed one fifteenth of the total volume. The reaction was normally carried out for 1-1.5 h at the appropriate temperature, and if the reaction was at a high temperature (*e.g.* 65°C), repeated centrifugation was performed because of evaporation. The enzyme was inactivated by extracting three times with a mixture of phenol:chloroform (1:1), and the DNA was precipitated with ethanol (section 2.8.3).

(b). Treatment with alkaline phosphatase

After digestion, the vector was usually treated with alkaline phosphatase to reduce self-religation. For this purpose, the DNA (approx. 1-3 μg) was incubated with 5 units of alkaline phosphatase (Beohringer) in 50 mM Tris-HCl; pH 8.0 and 10 mM $MgCl_2$ at 37°C for 30 min. The phosphatase was inactivated by extraction with phenol:chloroform as mentioned above.

(c). Ligation

The vector (approx. 0.3 μg) and insert DNA were combined in approximately equal molar ratio, and incubated with 1 unit bacteriophage T4 DNA ligase (BRL) in the presence of 0.5 mM ATP, 100 mM Tris-HCl; pH 7.5 and 50 mM $MgSO_4$ at 15°C for 16-20 h.

(d). Transformation and transfection of bacterial cells

The competent cells (section 2.3.3) were allowed to thaw on ice for

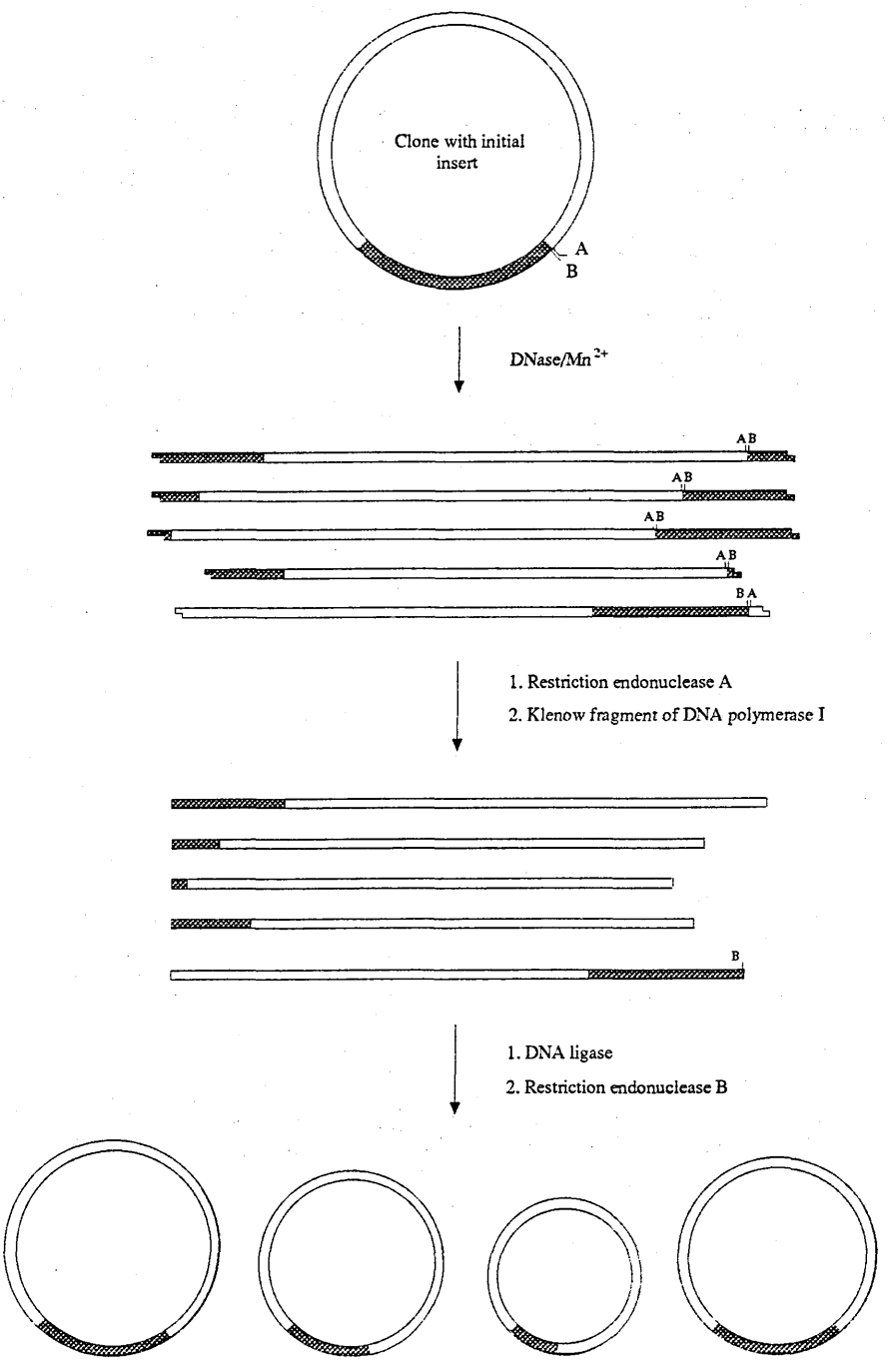
30 min, and 100 µl of the cells were mixed with a portion of the ligation reaction mixture (generally one tenth and one third). The suspension was kept on ice for 0.5-2.5 h, heated at 37°C for 3 min, and then poured onto an appropriate agar plate. For transformation of plasmid DNA, the suspension was spread on a plate containing antibiotic, which had been spread with 0.5% (w/v) IPTG and 0.5% (w/v) X-gal if appropriate. For transfection of bacteriophage M13, the suspension was quickly mixed with 60 µl of 1% (w/v) IPTG, 60 µl of 1% (w/v) X-gal, and 3-5 ml of Top agar (section 2.2.2) which had been cooled to approx. 45°C, and then immediately poured onto an H-plate (section 2.2.2). In both cases the plates were incubated at 37°C overnight, and the resulting colonies or plaques could be seen next day.

2.10.2. DNase/restriction deletion subcloning

DNase is capable of nicking double-stranded DNA in the presence of Mn^{2+} ions almost randomly, and can be used to generate a series of subclones with different deletions, providing different starting points for DNA sequencing. The method used here was a modification of that of Frischauf *et al.* (1980) by Labeit *et al.* (1987), and the principle of this method is presented in Fig. 2.3.

M13 RF DNA (approx. 10 µg) containing the cloned insert was dissolved in distilled water and mixed with 50 µl of 10^6-fold diluted DNase I (5 mg/ml stock solution in frozen aliquots) in dilution buffer containing 80 mM Tris-HCl; pH 7.4 and 2 mM $MnCl_2$ at room temperature. Samples (20 µl) of the reaction were withdrawn at two min time intervals, and immediately transferred into a fresh tube with 2 µl of 0.1 mM EDTA in order to stop the reaction. A portion (3 µl) of the samples from each fraction were subjected to electrophoresis on a 0.7% (w/v) agarose gel (section 2.7), and the remainder was kept at -20°C. The time point at which one third of the DNA was linearized was selected.

Fig. 2.3. Strategy of DNase deletion subcloning. M13 RF DNA containing the *Sst*I-*Sph*I fragment of *ca.* 2 kb (see Fig. 3.3.a) was digested with DNase in the presence of Mn^{2+}, followed by digestion with *Hin*dIII. The ends were filled in using the Klenow fragment of DNA polymerase I, and this DNA was self-ligated. The self-ligated DNA was digested with *Pst*I, and then the enzyme was inactivated by incubating at 70°C for 10 min. This was then used for transfection of JM109 cells.



Clones with deletion in the insert

Clone without deletion in the insert

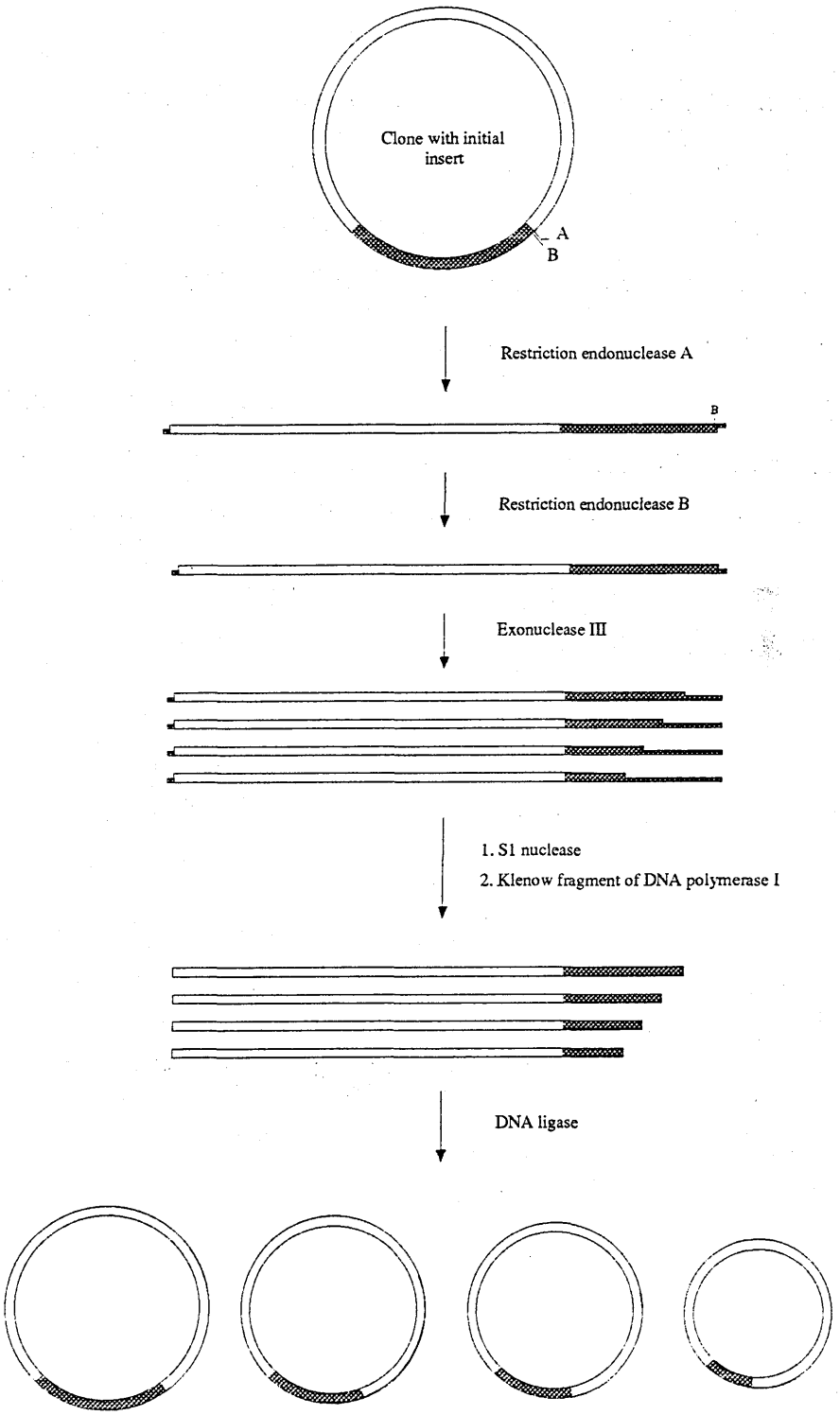
A suitable restriction enzyme (*Hin*dIII in this study) was used to cleave the DNA in the multiple cloning region beside the universal primer annealing site, and then the cleaved ends were filled in using the Klenow fragment of DNA polymerase I (in a buffer containing 100 mM Tris-HCl; pH 8.0, 50 mM $MgCl_2$, and 10 μM of each dNTP) or bacteriophage T4 DNA polymerase (in a buffer containing 15 mM Tris-HCl; pH 7.4, 75 mM NaCl, and 15 mM $MgSO_4$, and 10 μM of each dNTP) at 37°C for 30 min. The linearized DNA was religated (section 2.10.1.c), and digested with a second restriction enzyme (*Pst*I in this study) to cleave those religated DNA molecules that did not contain a deletion, so that they would not produce viable plaques. The enzyme was inactivated by incubation at 70°C for 10 min, and the reaction mixture used to transfect JM109 cells (section 2.10.1.d). The single-stranded DNA from each plaque was isolated (section 2.8.3), and its approximate size was determined by agarose gel electrophoresis (section 2.7) at 12 mA for 12-14 h. Clones containing different sized deletion in the insert were selected for sequencing (section 2.11.1).

2.10.3. Exonuclease III deletion subcloning

Exonuclease III is able to generate single-stranded DNA by releasing mono-nucleotides from the 3' end of double-stranded DNA containing a 3' hydroxyl group and 5'-protruding (or blunt) end. With subsequent S1 nuclease digestion, it is possible to create a deletion in the DNA. The principle of this subcloning method is presented in Fig. 2.4 (Henikoff, 1984).

Double-stranded target DNA (10-20 μg) cloned in pUC18, and prepared as in section 2.8.1.b was digested with a restriction endonuclease (*Sph*I or *Sst*I in this study) that gave a unique four-base 3'-protruding end in the clone at the both ends of the multiple cloning region of pUC18, or at the extreme ends of the insert. The DNA was then digested with a second restriction endonuclease which cut

Fig. 2.4. Strategy of exonuclease deletion subcloning. Target DNA cloned in pUC18 was digested with a suitable restriction endonulease A (see text),which generated a unique four-base 3'-protruding end. The linearized DNA was digested with a second restriction enzyme B (see text), which had a single recognition site beside the first cleavage point on the side of the insert, and provided a 5'-protruding end. This double-digested DNA was treated with exonuclease III, and the enzyme was inactivated by phenol. The single-stranded DNA generated by exonuclease III was deleted by using S1 nuclease, and the Klenow fragment of DNA polymerase I was used to fill in the ends of the DNA. This DNA was self-ligated and used to transform JM109 cells.



Clones with deletion in the insert

beside the first cleavage point (or less than 150 bp from it) on the insert side to give a 5'-protruding or blunt end. The DNA was precipitated with ethanol (section 2.8.3), resuspended in 100 µl of a buffer containing 50 mM Tris-HCl; pH 8.0 and 0.5 mM $MgCl_2$, and then incubated with 100 units exonuclease III (BRL) at 37°C. At 1 min intervals, 20 µl portion of the reaction were transferred to a fresh tube with 20 µl of a solution of 0.4 M NaCl and 10 mM EDTA; pH 8.0. The samples were heated at 70°C for 10 min followed by precipitation with ethanol (section 2.8.3). Each of the samples was resuspended in 50 µl of a buffer containing 33 mM sodium acetate; pH 4.5, 50 mM NaCl, and 0.03 mM $ZnSO_4$, and incubated with 5-10 Vogt units S1 nuclease (Boehringer) at 37 °C for 30 min. The extent of deletion was checked on a 0.7% (w/v) agarose gel (section 2.7). Next the Klenow fragment of DNA polymerase I was used to fill in the ends of the DNA (section 2.10.2), and the DNA was religated and used to transform JM109 cells (section 2.10.1.d). Double-stranded DNA from individual clones was isolated (section 2.8.1.b), and analysed by suitable restriction digestion in order to select clones with both the regions to which the sequencing primer anneals and the expected deletion in the insert.

2.11. Determination of nucleotide sequence

The dideoxy nucleotide chain-termination method was used to determine nucleotide sequences in this study (Sanger, 1977), and the procedure used was mainly based on the 'M13 Cloning and Sequencing Handbook" (Amersham International).

2.11.1. Determination of nucleotide sequence in single-stranded M13 DNA

(a). Annealing primer to template

The primer routinely used was the 17-mer universal sequencing primer

(BRL), which anneals immediate beyond the 3' end of the multiple cloning region (see Fig. 2.1.c) (the 15-mer primer was sometime also used). 1 µl of the primer (2 ng/µl) was mixed with 4 µl of single-stranded template DNA (approx. 1 µg), 1.5 µl of 10x Klenow buffer (section 2.10.2), and 3.5 µl of distilled water. This mixture was incubated at 55-60°C for 1-2 h, or alternatively heated at 70°C for 10 min, and then cooled down slowly to room temperature to allow the primer to anneal to the template.

(b). Nucleotide chain extension reaction using the Klenow Fragment

To the annealed template/primer mixture, 15 µCi [α-^{35}S]dCTP (Amersham, 1 mCi/0.1 ml) and 1 unit of the Klenow fragment of DNA polymerase I (diluted from the stock of 6 units/µl before use) were added, and mixed. The mixture was then divided into four portions of 2.5 µl each, and transferred to four fresh tubes labelled A, G, T, and C. The reaction was started by adding 2 µl of the appropriate chain termination solution (A°, G°, T°, and C°):

A°: 82 µM dGTP, 82 µM dTTP, 4 µM dATP, and 75 µM ddATP

G°: 82 µM dATP, 82 µM dTTP, 4 µM dGTP, and 25 µM ddGTP

T°: 82 µM dATP, 82 µM dGTP, 4 µM dTTP, and 250 µM ddTTP

C°: 82 µM of each dATP, dGTP,and dTTP, and 10 µM ddCTP

In some cases, 7-deaza-2'-dGTP was used in place of dGTP (at the same concentration) (Mizusawa *et al.*, 1986). After 20 min at 30°C, 2 µl of chase solution (0.125 µM of each dATP, dGTP, dTTP, and dCTP) was added to each tube (dGTP was used here even if the analogue had been used in the chain termination solutions), and the reaction continued for a further 15 min. The reaction was stopped by adding 5 µl of the electrophoresis loading solution (0.3% (w/v) xylene cyanol, 0.3% (w/v) bromophenol blue, and 20 mM EDTA in deionized formamide), and the samples stored at -20°C for up to two weeks.

(c). Nucleotide chain extension reaction using *Taq* polymerase

In some experiments the thermostable DNA polymerase from the thermophilic bacterium *Thermus aquaticus* (Chien *et al.*, 1976) was used instead of the Klenow fragment of DNA polymerase I in order to resolve the problem of 'false stops' (bands in all four lanes caused by the DNA polymerase pausing and allowing termination at regions of secondary structure during the sequencing reaction). The sequencing procedure described below was based on Brow (1990).

The annealed template/primer was mixed with 15 μCi [α-^{35}S]dCTP (Amersham, 1 mCi/0.1 ml), 1 μl of *Taq* Polymerase (BRL, diluted to 2 units/μl), 2 μl of the labelling solution (1.5 μM of each of 7-deaza-2'-dGTP, dATP, and dTTP), and 3 μl of distilled water. 7-deaza-2'-dGTP was used in place of dGTP in all of the reaction solution (Mizusawa *et al.*, 1986). After incubating at 42°C for 5 min, the tube was placed at room temperature, and the mixture was transferred into four tubes with 4 μl of the appropriate termination solution in each, labelled A, G, T, and C. The termination solutions were:

A: 20 mM of each dATP, 7-deaza-2'-dGTP, dTTP, and dCTP, and 800 mM ddATP
G: 20 mM of each dATP, 7-deaza-2'-dGTP, dTTP, and dCTP, and 60 mM ddGTP
T: 20 mM of each dATP, 7-deaza-2'-dGTP, dTTP, and dCTP, and 800 mM ddTTP
C: 20 mM of each dATP, 7-deaza-2'-dGTP, dTTP, and dCTP, and 400 mM ddCTP

The reaction was continued at 70°C for 5 min, and then cooled to room temperature, and 5 μl of the electrophoresis loading solution (0.3% (w/v) xylene cyanol, 0.3% (w/v) bromophenol blue, and 20 mM EDTA in deionized formamide) added to stop the reaction.

(d). Polyacrylamide buffer-gradient gel electrophoresis

The gel system used here was based on Biggin *et al.* (1983). It contained 9 M urea and 6% (w/v) deionized acrylamide/bisacrylamide (19:1), and two gel solutions were used. The top gel solution contained 0.5x TBE buffer (45 mM Tris, 45 mM boric acid, and 1.25 mM EDTA; pH 8.3), whereas the bottom gel solution contained 4x TBE buffer (0.36 M Tris, 0.36 M boric acid, and 10 mM EDTA; pH 8.3), 10% (w/v) sucrose, and 100 μg/ml bromophenol blue. For a single gel, 25 ml of top gel solution and 10 ml of bottom gel solution were used. Ammonium persulphate (0.05% (w/v) final concentration) and TEMED (NNN'N'-tetramethylethylenediamine, 0.2% (v/v) final concentration) were mixed quickly with each gel solution, and the top and the bottom gel solutions were drawn up into a 25 ml pipette subsequently. These solutions were gently mixed by drawing several air bubbles into the pipette to create a buffer gradient. This gel solution was then poured into a pre-prepared gel plate (40 cm x 20 cm),which had been siliconized with Repelcote (2% dimethyldichlorosilane in 1,1,1,-trichloroethane) and assembled with 0.4 mm spacers. Either a normal 32 tooth or sharks-tooth comb was inserted to make the wells. The gel was usually left at room temperature overnight in order to allow complete polymerization. The samples in the loading solution (0.3% (w/v) xylene cyanol, 0.3% (w/v) bromophenol blue, and 20 mM EDTA in deionized formamide) were denatured at 100°C for 3 min before applying to the wells on the gel. Electrophoresis was carried out at 25 mA and stopped when the first dye reached the bottom of the gel.

(e). Polyacrylamide buffer-linear gel electrophoresis

In this gel system, the top gel solution contained 0.5x TBE buffer (45 mM Tris, 45 mM boric acid, and 1.25 mM EDTA; pH 8.3), 9 M urea, and 6% (w/v) deionized acrylamide/bisacrylamide (19:1) was used alone, and electrophoresis was performed for a longer time than for a buffer gradient gel; This was normally

2.5-4.5 h longer, and allowed 50-100 more bases of the nucleotide sequence to be determined.

(f). Autoradiography

After the electrophoresis, the gel was soaked in a fixing solution containing 10% (v/v) acetic acid and 10% (v/v) methanol for 20 min. The gel was then dried under vacuum at 80°C for 20 min, and exposed to a Kodak X-Omat H film for autoradiography at room temperature.

2.11.2. Determination of nucleotide sequence in double-stranded plasmid DNA

This was performed according to Hattori and Sakaki (1986).

Double-stranded plasmid DNA (approx. 1.5-2.5 μg) was isolated (section 2.8.1.b), denatured in 20 μl of 0.5 N sodium hydroxide at room temperature for 10 min, and then neutralized by mixing with 20 μl of 1 M sodium acetate; pH 4.5. The DNA was precipitated with 120 μl of cold 100% ethanol, and kept in dry ice for 15 min. After centrifugation in an Eppendorf centrifuge for 10 min, the DNA was dried briefly under vacuum, and dissolved in 9 μl of the Klenow buffer containing 100 mM Tris-HCl; pH 8.0, 50 mM $MgCl_2$, and 10 μM of each of the dNTPs.

The denatured DNA was mixed with 1 μl of sequencing primer (approx. 0.5 pmol of the 17-mer, BRL), and incubated at 37°C for 20 min, and then left at room temperature at least for 10 min. The chain extension reaction was performed as in section 2.11.1.b and c.

2.12. Preparation of RNA from BHK cells

2.12.1. Preparation of total RNA

The acid guanidinium thiocyanate method was used for preparation of total RNA (Chomczynski and Sacchi, 1987).

The medium was poured off, and the cells were washed once with PBS buffer (130 mM NaCl, 5 mM Na_2HPO_4, and 5 mM NaH_2PO_4; pH 7.2), scraped off the wall of the bottle into PBS buffer, and sedimented in a autoclaved 30 ml Corex tube by centrifugation at 3,000 rpm, for 10 min at 20°C. The cells were then resuspended, by vortexing, in 5 ml of Guanidinium lysis buffer consisting of 50% (w/v) guanidinium thiocyanate, 0.5% (w/v) N-lauryl sarcosine, 50 mM Tris-HCl; pH 7.5, 2 mM EDTA; pH 8.0, and 0.7% (v/v) 2-mercaptoethanol. After adding 0.5 ml of sodium acetate (2 M; pH 4), 5 ml of phenol (saturated with distilled water) and 1 ml of chloroform:isoamyl alcohol (49:1), the mixture was vortexed for 2 min, and left on ice for 15 min followed by centrifugation at 9,000 rpm, for 20 min at 4°C. The upper aqueous phase was carefully transferred to a fresh tube, total RNA was precipitated with an equal volume of isopropanol and left at -20°C for at least 1 h. The precipitated RNA was sedimented and briefly dried under vacuum. The RNA was resuspended in 0.7 ml of the guanidinium lysis buffer (see above), and precipitated again with an equal volume of isopropanol in an Eppendorf tube. The sample was left at -20°C for 1 h. The RNA was collected by centrifugation in an Eppendorf centrifuge at 20°C for 10 min, dried under vacuum,and dissolved in 300 µl of distilled water. It was then mixed with an equal volume of chloroform/butan-1-ol (4:1) by vortexing, and subjected to centrifugation for 4 min to separate the two phases. The upper aqueous phase was carefully transferred to a fresh tube, and the organic phase was re-extracted once with 100 µl of water, and the two aqueous extracts were pooled. The RNA was re-precipitated with ethanol (section 2.8.3) at -20°C for at least 2 h, centrifuged in

an Eppendorf centrifuge, dissolved in a suitable volume of water, and stored at -80°C.

The quality of the RNA was determined by agarose gel electrophoresis, visualizing with ethidium bromide (section 2.7) and by the ratio of A_{260}/A_{280}, which was greater or equal to 1.8. The quantity of RNA was established from its A_{260}, using the relationship that single-stranded RNA at a concentration of 40 µg/ml has an A_{260} value of 1 in a cell with a 1 cm length path.

2.12.2. Preparation of polyadenylated RNA

The method for isolation of polyadenylated RNA was that of Aviv and Leder (1972).

A column of 100 mg of oligo(dT) cellulose (type 2, BRL) in distilled water containing diethylpyrocarbonate ($1/10^4$) was prepared. The column was washed with approx. 2 ml of 0.2 N NaOH, and then with water containing diethylpyrocarbonate ($1/10^4$) until the pH of elution was lower than 7. Before use the column was equilibrated with about 5 ml of loading buffer containing 10 mM Tris-HCl; pH 7.5, 500 mM NaCl, and 1 mM EDTA. 0.5 ml of the total RNA (*ca.* 0.5 mg/1 ml) in the loading buffer was heated at 65°C for 5 min, cooled on ice, and then applied to the column. The eluate was collected, re-heated, and applied to the column again. The column was washed with loading buffer containing 0.5% (w/v) sodium dodecyl sulphate until the A_{260} of the eluate fractions (1 ml fractions) fell below 0.05, and the polyadenylated RNA was then eluted from the column using elution buffer (10 mM Tris-HCl; pH 7.5, 1 mM EDTA, and 0.5% (w/v) sodium dodecyl sulphate). The A_{260} of each fraction (0.2 ml for each) was determined, and the fractions containing the peak of the polyadenylated RNA was pooled.

The polyadenylated RNA was precipitated by ethanol (section 2.8.3) and stored at -80°C.

2.13. Labelling DNA

2.13.1. End-labelling with [γ-^{32}P]ATP

This was used to label restriction endonuclease digested DNA for use as a size marker.

The DNA fragments with 5' phosphorylated ends (1-50 pmole ends) were treated with alkaline phosphatase as described in section 2.10.1.b in order to produce free 5' hydroxyl groups. This DNA was re-phosphorylated with 60 µCi [γ-^{32}P]ATP (1 mCi/0.1 ml, Amersham) using 0.2 units of polynucleotide kinase in a buffer containing 50 mM Tris-HCl; pH 8.0; 10 mM $MgCl_2$, and 5 mM dithiothreitol at 37°C for 30 min. The solution was then mixed with 40 µl of 2.5 M ammonium acetate and 160 µl of 100% ethanol, the tube was placed on dry ice for 15 min, and centrifuged in an Eppendorf centrifuge for 5 min. The DNA was resuspended in 0.1 ml of sodium acetate (0.3 M; pH 6), and re-precipitated in 0.3 ml of 100% ethanol. After 15 min on dry ice the DNA was collected by centrifugation in an Eppendorf centrifuge for 5 min and dissolved in 25 µl of TE buffer (section 2.7).

2.13.2. Nick-translation

The method used here was based on that described by Rigby *et al.* (1977). 0.3-1 µg DNA was mixed with dATP, dGTP,and dTTP (each at 5 µM), 50 µCi [α-^{32}P]dCTP (1 mCi/0.1 ml, Amersham), DNase I (0.01 ng/µl), 5 units of DNA polymerase I (BRL) in a buffer containing 0.15 M Tris-HCl; pH 7.4, 0.15 M $MgSO_4$, 15 mM dithiothreitol, and 0.75 M NaCl, and total volume of the reaction was 12 µl. The reaction was performed at 15°C for 4 h.

After nick-translation, the sample was applied on a Biogel P60 column which had been pre-equilibrated with NE (100 mM NaCl and 0.5 mM EDTA: pH 7.0). The column was eluted with 1 ml of NE, and 100 µl fractions collected.

The fractions which contained the first peak of radiation of ^{32}P were pooled, and then the radioactivity was determined by measuring Cherenkov radiation in a scintillation spectrometer. The labelled DNA was stored at -20°C.

2.14. Hybridization analysis of nucleic acids

2.14.1. Southern hybridization (DNA-DNA hybridization)

This is the method of Southern (1975). The digested DNA was separated on an agarose gel (section 2.7). The gel was then soaked in approx. 200 ml of 0.5 N NaOH and 1.5 M NaCl for 30 min, rinsed twice with distilled water, and then neutralized by rinsing in 250 ml of 0.5 M Tris-HCl; pH 7.6 and 1.5 M NaCl for 30 min each time. The gel was then placed on 3MM paper on the lid of a transfer tank, in contact with 20x SSC (3 M NaCl and 0.3 M sodium citrate). A nitrocellulose membrane (approx. 12x 12 cm), wetted with 20x SSC, was placed on top of the gel, and six sheets of 3MM paper and a *ca.* 6 cm stack of napkin tissues placed on top, under pressure. The transfer was allowed to proceed overnight at room temperature.

Next day, the membrane was washed in 2x SSC for 5 min, allowed to dry at room temperature, and baked in a vacuum oven for 2 h at 80°C in order to immobilize the DNA on the membrane. Pre-hybridization was set up at 45°C for 3 h in hybridization buffer containing 5x SSPE (0.9 M NaCl, 50 mM NaH_2PO_4, and 5 mM EDTA, adjusted pH to 7.2), 50% (v/v) deionized formamide, 0.1 % (w/v) sodium dodecyl sulphate, and 10x Denhardt's solution (diluted from filtered (Whatman No 1) 50x Denhardt's composed of 1% (w/v) Ficoll, 1% (w/v) polyvinylpyrolidine, and 1% (w/v) bovine serum albumin; all from Sigma). The labelled DNA probe (see section 2.13.2) was denatured in 0.1 volume of 0.1 N NaOH, and neutralized in 0.1 volume of 1 M Tris-HCl; pH 7.5 and 0.1 volume of 1 N HCl. When pre-hybridized already, the membrane was incubated with the

denatured DNA probe at 45°C for 20 h in the fresh hybridization buffer. The membrane was then washed three times at room temperature with 2x SSC (diluted from 20x SSC) and 0.1% (w/v) sodium dodecyl sulphate for 10 min each time, and twice at 65°C in 0.1x SSC and 0.1% (w/v) sodium dodecyl sulphate for 30 min each time.

The membrane was then exposed to Fuji *RX* film using a Lightening-Plus intensifying screen at –80°C for autoradiography.

2.14.2. Northern hybridization (RNA–DNA hybridization)

This DNA–RNA hybridization procedure was based on the method of Goldberg (1980).

(a). Agarose gel electrophoresis of RNA

In this system, a buffer containing 20 mM MOPS (Sigma), 5 mM sodium acetate, and 1 mM EDTA pH 7.0, was employed, instead of TBE buffer, and formaldehyde included as a denaturing reagent. To make a 1% (w/v) agarose gel, 1 g agarose (BRL 5510UB) was melted in 90 ml MOPS buffer, and the solution was cooled to about 50°C. This was transferred to a fume cupboard (formaldehyde is toxic). The solution was quickly mixed with 18 ml of 35-37% formaldehyde (2.2 M final concentration) and poured into a gel mould (12x 11 cm). RNA samples were added in 15 µl of loading buffer containing 13% (v/v) 10x MOPS buffer, 20% (v/v) 35-37% formaldehyde, 66% (v/v) deionized formamide (Fluka), 0.5% (v/v) bromophenol blue (10 mg/ml), heated at 65°C for 10 min, and then quickly transferred to ice. The size marker was denatured DNA makers mentioned in section 2.7, either radioactively labeled, or stained with ethidium bromide. These samples were subjected to the gel, and electrophoresis was stopped when the dye had migrated to the other end of the gel. (This was performed at 25-50 V for 4-6 h.)

(b). Hybridization

After the electrophoresis (see above), the gel was rinsed briefly with 20x SSC (3 M NaCl and 0.3 M sodium citrate), and RNA was directly transferred onto a nitrocellulose membrane as described in section 2.14.1. The nitrocellulose membrane was removed and baked in a 80°C vacuum oven for 2 h without premier washing. Pre-hybridization, hybridization, stringent washing and the autoradiography was performed as described in section 2.14.1.

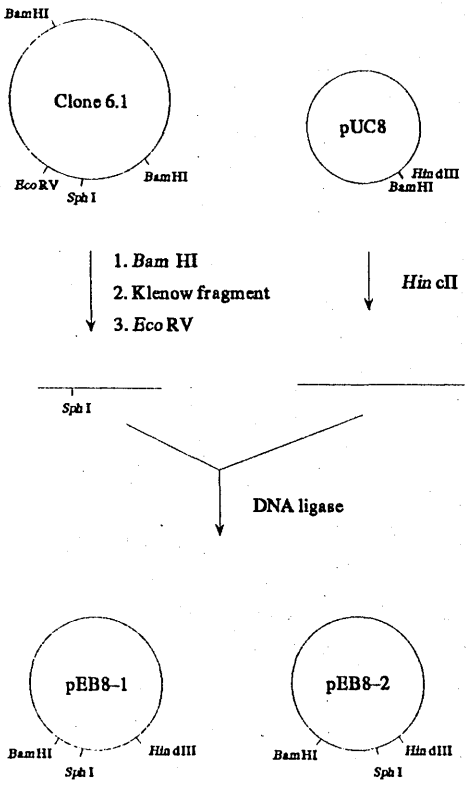
2.15. Site-directed mutagenesis using the polymerase chain reaction

The polymerase chain reaction amplification method described here was based on that developed by Saiki *et al.* (1988) using reagents from Cetus.

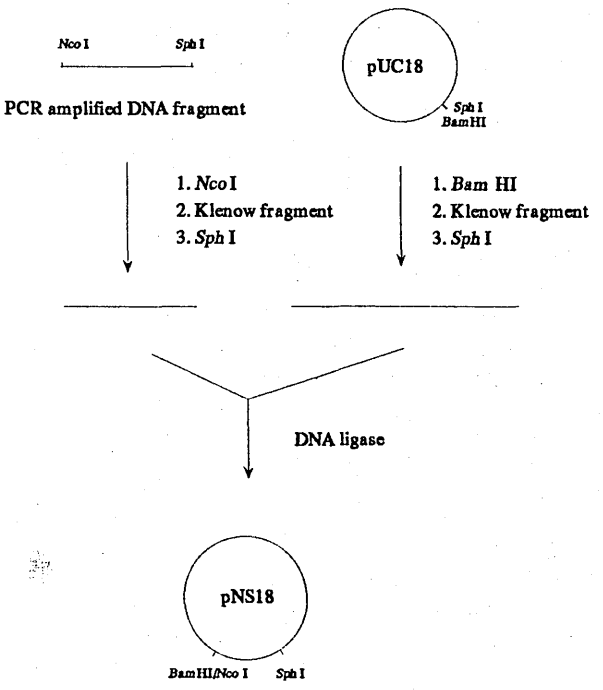
In order to allow the gene *pk* to be cloned into the prokaryotic expression vectors (section 2.6.2), the polymerase chain reaction was used to introduce an *Nco*I site at the translational initiation site of the *pk* gene. For performing this site-directed mutagenesis, an oligonucleotide, primer 1 was designed with a single mismatch (bold) against the anti-sense strand of the *pk* gene, so as to produce an *Nco*I side: $^{5'}$ACGCAGCAACCATGGCCGAC$^{3'}$. The forward universal sequencing primer ($^{5'}$GTAAAACGACGGCCAGT $^{3'}$) was used as the other primer for the sense strand. The target DNA was the *Eco*RV-*Sph*I fragment of the *Bam*HI 10 fragment of PRV cloned in M13mp18 (see Fig. 3.3.a). 0.1 µg target DNA was mixed with 2.5 units of *Taq* polymerase, 25 µg of the primer 1, 20 µg of the universal forward primer, and dNTPs at a concentration of 50 µM within 1x *Taq* buffer in a total volume of 50 µl, and covered by light paraffin. The reaction cycle was 94°C for 1.5 min; 50°C for 2 min; 72°C for 1 minute, and 30 cycles were performed. Then, the products were analysed by electrophoresis on a 1% (w/v) agarose gel (section 2.7), and the sample was stored at -20°C.

Fig. 2.5. Strategy of reconstruction of the *pk* gene of PRV. (a) DNA from clone 6.1 (see section 3.1.1), which contains the *Bam*HI 10 fragment of PRV, was digested with *Bam*HI, and the ends were filled in by the Klenow fragment of DNA polymerase I. This DNA was then digested with *Eco*RV, and the released DNA fragment containing the *pk* gene was subcloned into *Hinc*II site of pUC8. Clones with the two alternative orientations were designated pEB8-1 and pEB8-2. (b) Cloning of DNA fragments generated by site-directed mutagenesis using the polymerase chain reaction. DNA fragments generated by the polymerase chain reaction were digested with *Nco*I, filled in by the Klenow fragment of DNA polymerase I, and digested with *Sph*I. This DNA was cloned between the *Bam*HI (filled in by the Klenow fragment of DNA polymerase I) and *Sph*I sites of pUC18, and the resulting plasmid was designated pNS18. (c) Reconstruction of the *pk* gene. (i). The DNA of pNS18 was digested with *Bam*HI and *Sph*I, and the released fragment containing the mutation was subcloned between *Bam*HI and *Sph*I sites of pEB8-1. This new plasmid, pMK8-1 was used to place the *pk* gene into pKK233-2. (ii). The pNS18 DNA was also digested with *Bam*HI, and the ends were filled by the Klenow fragment of DNA polymerase I. The linearized DNA was digested with *Sph*I, and the released fragment containing the mutation was subcloned between the *Sph*I and the *Hin*dIII (filled by the Klenow fragment of DNA polymerase I) of pEB8-2. This new plasmid, pMK8-2 was used to clone the *pk* gene into pJLA502.

a.

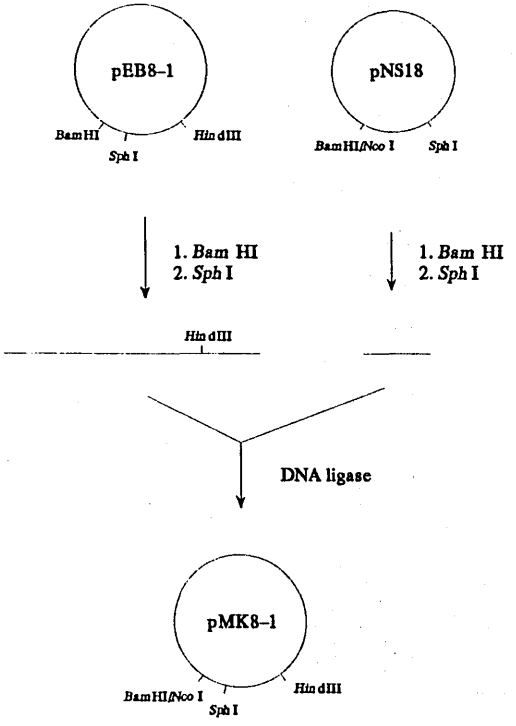

BamHI
Clone 6.1
EcoRV
BamHI
Sph I
pUC8
HindIII
BamHI
1. Bam HI
2. Klenow fragment
3. EcoRV
Hin cII
Sph I
DNA ligase
pEB8-1
BamHI
HindIII
Sph I
pEB8-2
BamHI
HindIII
Sph I

b.

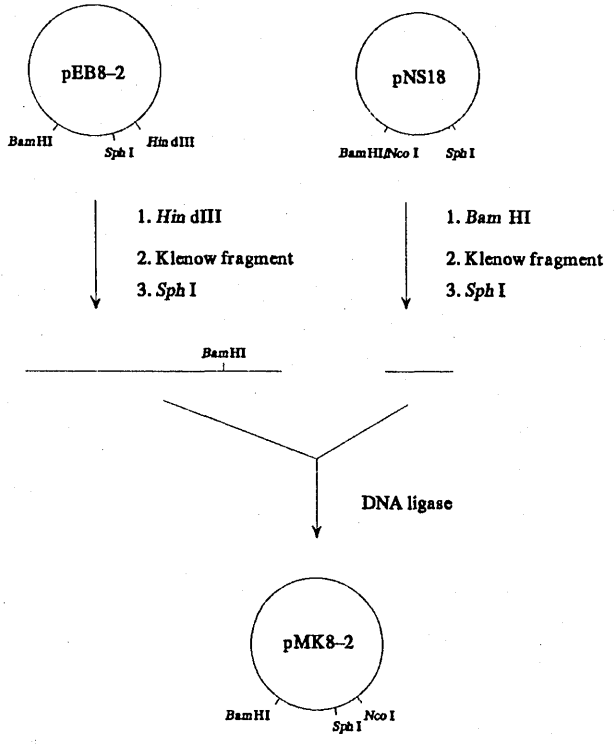

Nco I
Sph I
PCR amplified DNA fragment
pUC18
Sph I
BamHI
1. Nco I
2. Klenow fragment
3. Sph I
1. Bam HI
2. Klenow fragment
3. Sph I
DNA ligase
pNS18
BamHI/Nco I
Sph I

c.

i.


pEB8–1
BamHI
Hin dIII
Sph I
pNS18
BamHI/Nco I
Sph I
1. Bam HI
2. Sph I
1. Bam HI
2. Sph I
Hin dIII
DNA ligase
pMK8–1
BamHI/Nco I
Sph I
Hin dIII

ii.


pEB8–2
BamHI
Sph I
Hin dIII
pNS18
BamHI/Nco I
Sph I
1. Hin dIII
2. Klenow fragment
3. Sph I
1. Bam HI
2. Klenow fragment
3. Sph I
BamHI
DNA ligase
pMK8–2
BamHI
Sph I
Nco I

The gene *pk* was re-constructed in pUC8 (Fig. 2.5) before subcloned into the expression vectors (section 2.6.2).

2.16. Gel electrophoresis for protein analysis

The electrophoresis system was according to Laemmli (1970). The gels contained 10% (w/v) acrylamide/bisacrylamide (19:1), 0.1% (w/v) sodium dodecyl sulphate, 0.05% (w/v) ammonium persulphate, 0.002% (v/v) TEMED, and 0.4 M Tris-HCl; pH 8.7. The stacking gel was made as 6 % (w/v) acrylamide/bisacrylamide (19:1), 0.1% (w/v) sodium dodecyl sulphate, 0.05% (w/v) ammonium persulphate, 0.002% (v/v) TEMED, and 0.125 M Tris-HCl; pH 6.8. The buffer used for this gel electrophoresis contained 50 mM Tris-HCl, 192 mM glycine, and 0.1% (w/v) sodium dodecyl sulphate; pH 8.5.

The samples in 150 µl of the bacteria lysis buffer containing 50 mM Tris-HCl; pH 6.8, 10% (v/v) 2-mercaptoethanol,4% (w/v) sodium dodecyl sulphate, 30% (v/v) glycerol, and 0.002% (w/v) bromophenol blue, were heated at 100°C for 10 min, and 10 µl of the sample was applied. The electrophoresis was performed in the electrophoresis buffer (see above) at 5-6 mA overnight (or in some cases 20-30 mA for 4-5 h) until the blue dye approached the end of the separation gel. The gel was stained with Coomassie blue solution containing 20% (v/v) trichloroacetic acid, 10%(v/v) methanol, and 0.1% (w/v) Coomassie brilliant blue R 250 (Sigma) for 15 min, and then, destained in 10% (v/v) acetic acid and 10% (v/v) methanol.

The destained gel was photographed.

2.17. Computer programs for DNA sequence analysis

Computer analysis of DNA sequence was made using programs (version 6.1) of the GCG (Genetic Computer Group, University of Wisconsin) package

(Devereux *et al.*, 1984) mounted on a Micro Vax 3600 in the Glasgow University local area VAX cluster running version 5 of the VMS operating system. The databases used were GenBank (release 64.0), EMBL (release 23), NBRF-protein (release 25.0), and SwissProt (release 14), and the programs most frequently used were:

SeqEd: for creating and editing files of DNA and protein sequences.

Map: for displaying a restriction map of a DNA sequence and its possible protein translations.

MapSort: for displaying the coordinates of the restriction cuts in a DNA sequence and sorting the fragments of the resulting digest by size.

Compare: for comparing two proteins or nucleotide sequences and creating a file of the points of similarity between them for plotting with DotPlot.

DotPlot: for making a dot matrix comparison with the output file from Compare.

BestFit: for making an optimal alignment of the best segment of similarity between two sequences.

Gap: for making an optimal alignment between two complete sequences by inserting gaps to maximize the number of matches.

CodonPreference: a frame-specific gene finder that tries to recognize protein coding sequence by virtue of their similarity to a coding frequency table.

TestCode: an aid to identifying protein coding sequences by plotting a measure of the non-randomness of the composition at every third base.

Translate: for translating nucleotide sequences into peptide sequences.

PepPlot: for measuring protein secondary structure and hydrophobicity.

Frames: for showing open reading frames for the six possible translation frames of a DNA sequence.

WordSearch: for identifying sequences similar to any known sequence in the database.

Segments: for aligning and displaying the segments of similarity found by

WordSearch.

FastA: similar to WordSearch, but arising a different alignment.

TFastA: for searching predicted peptide sequences encoded in a nucleotide sequence similar to that in any known nucleotide sequence in the database. It translates the nucleotide sequence in six frames before performing the comparison.

Profile: for calculating a table (profile) that quantitively represents a group of aligned sequences.

ProfileSearch: for using Profile to search the database for new sequences with similarity to the group.

PeptideSort: for showing molecular weight, isoelectric point and amino acid analysis. It can also list peptides cleaved by various enzymes, sorted by position in the sequence, weight, and retention on HPLC at pH 2.1.

CHAPTER 3
RESULTS

3.1. Cloning and sequencing of the *Bam*HI 10 fragment of PRV genomic DNA

Because the HSV-1 protein kinase gene, US3, lies upstream of the gene (US4) for glycoprotein G, it seemed possible that the PRV protein kinase gene might be located upstream of the gX gene, homologous to US4 (see Fig. 1.2 and Table 1.6). From the restriction map of PRV genomic DNA (Fig. 1.3) and the nucleotide sequence of the gX gene (Rea *et al.*, 1985) , it transpired that the 5'-end of gX is located just 5' of the *Bam*HI site between the *Bam*HI 10 fragment and the *Bam*HI 16 fragment. As the molecular weight of the subunit of the PRV protein kinase is 38,000 (Purves *et al.*, 1987a), it would be expected that the protein kinase gene would occupy approximately 1 kb in the virus genomic DNA. Thus, the 4 kb *Bam*HI 10 fragment should have ample room to contain the complete protein kinase gene. For this reason, the *Bam*HI 10 fragment was chosen as the initial target in the search for the protein kinase gene.

3.1.1. Cloning and identification of *Bam*HI fragments in pUC18

PRV genomic DNA was digested with *Bam*HI, ligated into the *Bam*HI site of pUC18, and the product used to transform *E. coli*. Plasmid DNA isolated from 46 different colonies was digested with *Bam*HI, and the sizes of the inserts

Fig. 3.1. Selection of cloned *Bam*HI fragments of PRV DNA. DNA from clones 2.3 (lane 3), 4.8 (lane 4), 4.10 (lane 5), 1.6 (lane 6), 6.3 (lane 7), 1.1 (lane 8), 6.1 (lane 9), 1.0 (lane 10), 6.2 (lane 11), 3.1 (lane 12) and 1.11 (lane 13) was digested with *Bam*HI, and subjected to electrophoresis in 0.7% agarose. The linearized pUC18 is shown in lane 1, and the size markers were λ DNA digested with *Hin*dIII (lanes 2 and 14), as well as pUC18 digested with *Eco*RI and with *Rsa*I (lane 15).

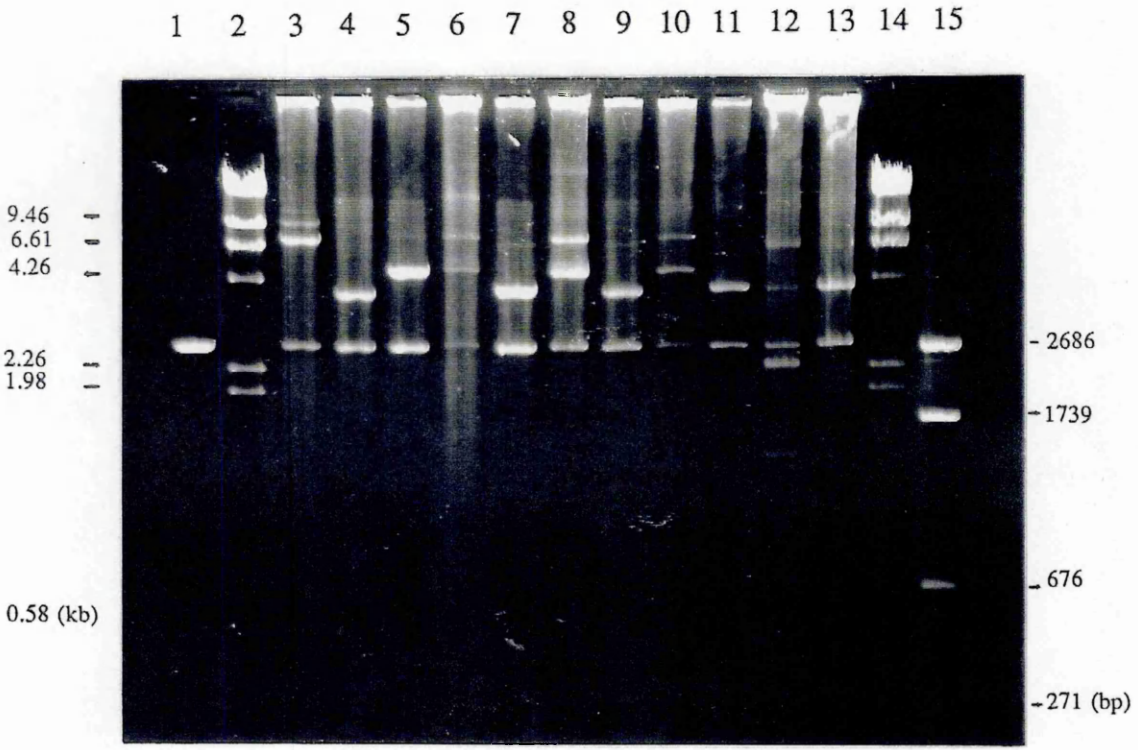

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
9.46
6.61
4.26
2.26
1.98
0.58 (kb)
2686
1739
676
271 (bp)

Table 3. *Bam*HI fragments of PRV DNA cloned into pUC18

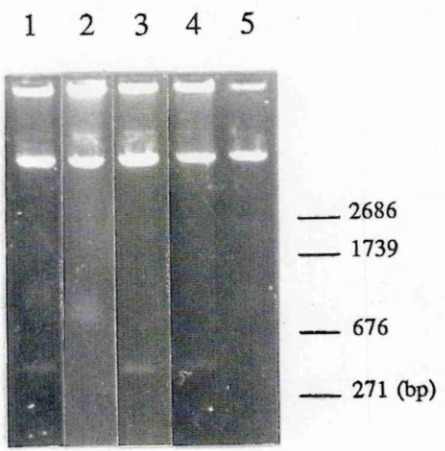
Clone	*Bam*HI fragment	Designation	Insert size (kb)	Evidence
11.3	*Bam* 7	pPRVB7	*c.* 7.0	Sequence at one end corresponds to *gX* gene [Rea *et al.*, 1985].
4.10	*Bam* 8'	pPRVB8'	*c.* 5.2	Size. Sequence at ends show it is not *Bam*8 fragment [Lomniczi *et al.*,1984].
6.3	*Bam* 9	pPRVB9	*c.* 4.0	Size. Although *Bam*10 fragment is also of this size, restriction map and terminal sequences are different.
6.1	*Bam* 10	pPRVB10	4.0	Sequence at one end corresponds to 5'end of *gX* gene [Rea *et al.*, 1985].
6.10	*Bam* 11	pPRVB11	3.1	Size and single *Kpn*I site.
9.8	*Bam* 12	pPRVB12	2.7	Size and restriction sites in common with *Bam*10 fragment. (Confirmed by sequencing: see Section 3.4)
2.9	*Bam* 14	pPRVB14	1.5	Size only.
3.8	*Bam* 15	pPRVB15	0.8	Size only.
4.1	?		*c.* 0.4	
1.10	?		*c.* 0.3	

were estimated by comparison with DNA size markers (see section 2.7.1). Ten different sizes of *Bam*HI fragment were distinguished (a selection of these are shown in Fig. 3.1), and eight of these were tentatively identified from published values of the sizes of the *Bam*HI fragments of PRV (Lomniczi *et al.*, 1984) (Table 3). Those of a size (*ca.* 3.7 kb) corresponding approximately to that reported for *Bam*HI 10 were picked as candidates for further identification.

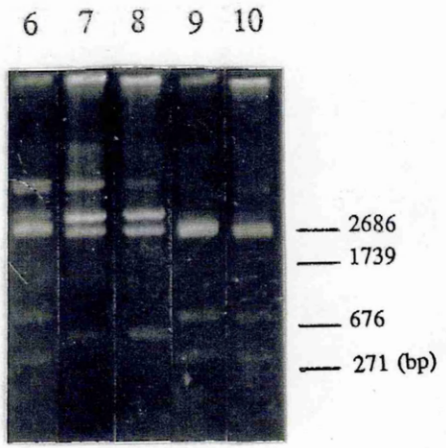
The candidate clones, 1.11, 4.8, 6.1, 6.2 and 6.3, were digested with restriction endonucleases. Of particular interest were *Xho*I and *Pst*I, as Thomsen *et al.* (1987) had shown that an *Xho*I site is located approximately 1 kb from the 'right hand' *Bam*HI site of *Bam*HI 10, and Ben-Porat and Kaplan (1985) had shown there to be a *Pst*I site at a similar position. The results of digestion with *Xho*I showed that these five clones, although of very similar size, represented two distinct species, as 1.11, 6.2, and 6.3 generated two fragments with this enzyme, whereas clones 6.1 and 4.8 were only linearized (Fig. 3.2.a). As there is no *Xho*I site in pUC18, this indicated two and one sites for the enzyme in the respective inserts. Double digestion with *Xho*I and *Bam*HI showed that the *Xho*I site in the clones 6.1 and 4.8 was at about 0.6 kb from one end of the insert, and the *Xho*I sites in the clones 1.11, 6.2, and 6.3 were approximately 0.4 and 1.2 kb from one end (Fig. 3.2.b). The digestion with *Pst*I showed that clone 6.1 contained a single *Pst*I site at about 0.8 kb from the *Pst*I site of the vector, whereas the inserts in 6.2 and 6.3 contained two *Pst*I sites. Hence both types of clone remained candidates for *Bam*HI 10, and clones 6.1 and 6.3 were chosen for sequencing to resolve this uncertainty. Partial nucleotide sequences were determined from the *Bam*HI sites of both clones. It was found that the sequence from the end of the *Bam* fragment subsequently shown to be nearest to the *Xho*I site in clone 6.1 corresponded to that of the 5'-terminus of the gX gene (Rea *et al.*, 1985), indicating that clone 6.1 contained the *Bam*HI 10 fragment. This clone was therefore renamed pPRVB10.

Fig. 3.2. Identification of the *Bam*HI 10 fragment of PRV DNA. DNA from clones, 1.11 (lanes 1, 6 and 11), 4.8 (lanes 2, 7 and 12), 6.1 (lanes 5, 8 and 13), 6.2 (lanes 4, 9 and 14), and 6.3 (lanes 3, 10 and 15), was digested with *Xho*I (a), *Xho*I/*Bam*HI (b), and *Pst*I (c), and subjected to electrophoresis of 1% agarose. The size marker was pUC18 digested with *Eco*RI, and with *Rsa*I.

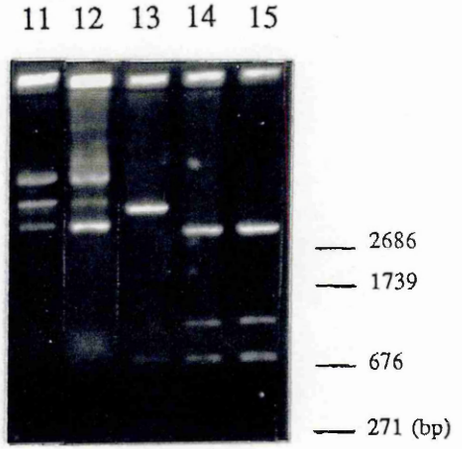
a.


1 2 3 4 5
2686
1739
676
271 (bp)

b.


6 7 8 9 10
2686
1739
676
271 (bp)

c.


11 12 13 14 15
2686
1739
676
271 (bp)

The partial nucleotide sequence of clone 6.3 did not correspond to any previously published sequence of PRV. However, as it contained an insert slightly larger than that in pRVB10, it might contain the *Bam*HI 9 fragment. It should, however, be mentioned that in a previous study the *Bam*HI 9 fragment of the Ka strain of PRV was clearly different in size from the *Bam*HI 10 fragment on agarose gel electrophoresis (Lomniczi *et al.*, 1984), in contrast to the results of the present study. The reason for this difference is unclear.

3.1.2. The nucleotide sequence of the *Bam*HI 10 fragment

Before starting to sequence the *Bam*HI 10 fragment of PRV DNA, a restriction map was made using a range of restriction endonucleases which have six base-pair recognition sites. The restriction map is presented in Fig. 3.3.a, from which it can be seen that the restriction endonuclease sites are mainly located in the regions near the ends of the *Bam*HI 10 fragment, but not in the middle. It was decided to divide the determination of the nucleotide sequence of the *Bam*HI 10 fragment into three stages, involving the regions *Bam*HI-*Sst*I, *Sst*I-*Eco*RV, and *Eco*RV-*Bam*HI. These fragments were subcloned into pUC18, and named pB10-BS1, pB10-SE2, and pB10-EB1, respectively. The sequencing strategy is shown in Fig. 3.3.b, and the complete nucleotide sequence of *Bam*HI 10 fragment is presented in Fig. 3.4. The particular details of the sequence determination for these regions are described, in turn, below.

The nucleotide sequence in the region between the *Eco*RV site and the 'right hand' side of the *Bam*HI site was initially determined on M13 clones of fragments generated using the restriction endonuclease sites identified by restriction mapping (see Fig. 3.3.a). Then, random clones in M13 were prepared from fragments generated by restriction endonucleases (*Hin*PI, *Hpa*II, and *Taq*I) which have four base-pair recognition sites, these being chosen on the basis of the

Fig. 3.3. Strategy for sequencing the *Bam*HI 10 fragment of PRV. (a) Restriction map of the *Bam*HI 10 fragment, showing restriction endonuclease recognition sites used for subcloning and sequencing. (b) Strategy for determination of the nucleotide sequence of the *Bam*HI 10 fragment. The horizontal arrows represent the regions and the directions sequenced in individual clones which were obtained using the methods described in section 2.10. The nucleotide sequences in each clone were determined in at least two independent experiments. (c) Fragments used as probes for hybridization analyses. The fragment *a*, from nucleotides 3300-3585 (see text), was used as a probe for hybridization of the transcript of the *pk* gene, and the fragment *s*, the *SalI* fragment between nucleotides *ca.* 645 and 1140, was used for hybridization of RSp40 mRNA. The fragments r_1 and r_2, (approx. nucleotides 1805-2303 and 2303-2553), together with the fragments *a* and *s*, were used as probes for determining the region of homology between the *Bam*HI 10 and the *Bam*HI 12 fragments.

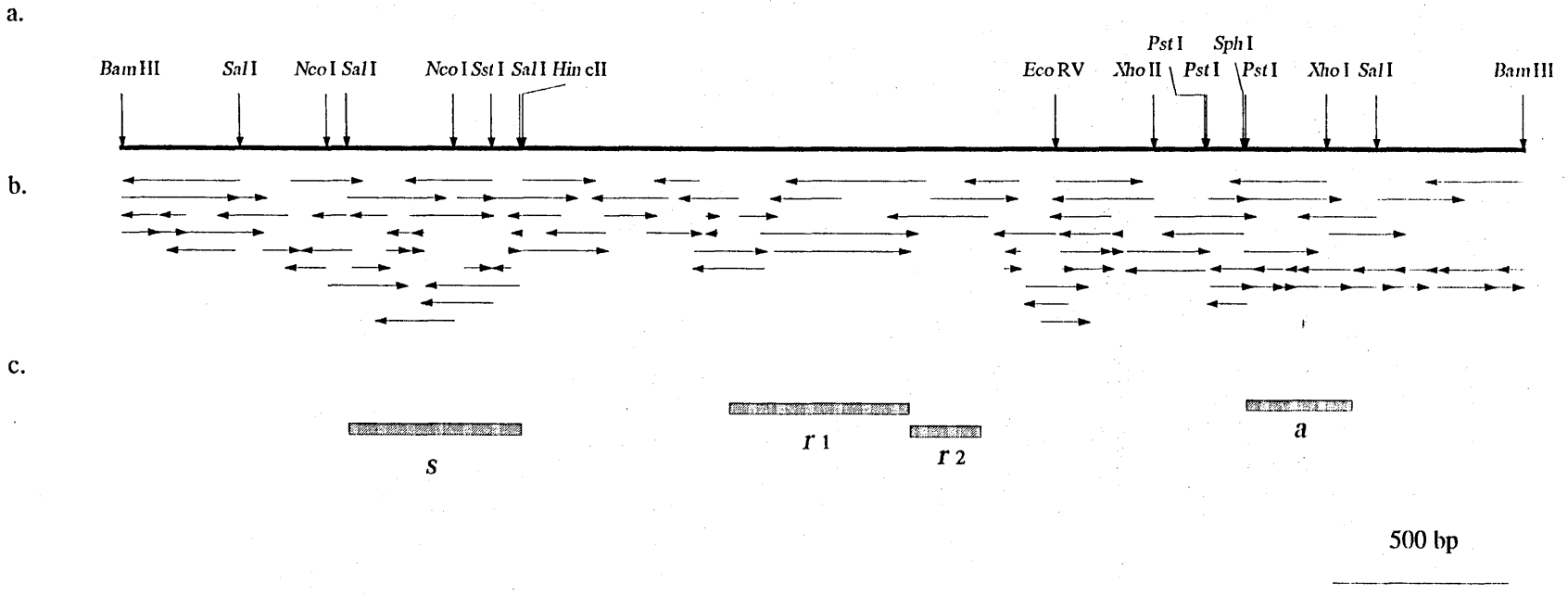

a.
b.
c.
BamHI
SalI
NcoI SalI
NcoI SstI SalI HincII
EcoRV
XhoII
PstI
PstI
SphI
XhoI SalI
BamHI
s
r1
r2
a
500 bp

Fig. 3.4. Nucleotide sequence of the *Bam*HI 10 fragment of PRV DNA. Each identical copy of the two tandemly reiterated sequences, R_1 and R_2, is numbered, and the position of the Us/IRs junction is indicated. The termini of fragments generated by digestion of the pB10-BS1 DNA with S1 nuclease (see text) are also marked, and the 25 nucleotides sequenced on one strand in three individual clones are underlined.

GGATCCCCCGGTCGGGGTGGGCCGGGAGTGGGCGTGGCAACTCCTCCCCGTCCCCCCTCC 60
CCTAGGGGGCCAGCCCCACCCGGCCCTCACCCGCACCGTTGAGGAGGGGCAGGGGGGAGG

CCGGAAGAAACGCCTCGGGGGTCGCGCGTCCGTCGCCTAGGTGAGTCGGGGGGCTTCCCG 180
GGCCTTCTTTGCGGAGCCCCCAGCGCGCAGGCAGCGGATCCACTCAGCCCCCCGAAGGGC

GGTGCGGCTCCTCCCCCCTCCCCTCCCCCGGGTTGGGTTGGGCGGGTGGGCGGGTGGGCG 240
CCACGCCGAGGAGGGGGGAGGGGAGGGGGCCCAACCCAACCCGCCCACCCGCCCACCCGC

GGGGTCGGGAGGGATCGGGGGAAGGGGGGGGGTCGTGGCTCACGCTGACGTCATGCGTTT 300
CCCCAGCCCTCCCTAGCCCCCTTCCCCCCCCCAGCACCGAGTGCGACTGCAGTACGCAAA

CCCCCCCCCCCACAGGGACGCGAGCGGGACCGGCTGGTCGACCACCCATTCAATCCATCC 360
GGGGGGGGGGGTGTCCCTGCGCTCGCCCTGGCCGACCAGCTGGTGGGTAAGTTAGGTAGG

AATCCATCCAGCCACCCCCACCCGTCGGGGCGGCCGGGCTCGGGTCCGGTTCCGGGTCGG 420
TTAGGTAGGTCGGTGGGGGTGGGCAGCCCCGCCGGCCCGAGCCCAGGCCAAGGCCCAGCC

CGTCGGGGAGCCTCGCCCACCCACCCGGCCCCCGTCCCAGCCCAGCCCCCCCCCCCCCCC 480
GCAGCCCCTCGGAGCGGGTGGGTGGGCCGGGGGCAGGGTCGGGTCGGGGGGGGGGGGGGG
← S1 S1 →

CCCCCCCCCCCCACGCCCCCATCGCACCCCCTCCCACCCCGTTCCCACCACCCCGCCGACC 540
GGGGGGGGGGGTGCGGGGGTAGCGTGGGGGAGGGTGGGGCAAGGGTGGTGGGGCGGCTGG

CCCCGCTCGACCGCCCGCCCCCGACCCACCGCGTCCGCCGCGCCCATGGACCGGGTCTGG 600
GGGGCGAGCTGGCGGGCGGGGGCTGGGTGGCGCAGGCGGCGCGGGTACCTGGCCCAGACC

GCCGACTGGTACGAGCCCGTGCCCTCCCCGCCGTTCTCGCCCGTCGACCCGCCCGGGCCC 660
CGGCTGACCATGCTCGGGCACGGGAGGGGCGGCAAGAGCGGGCAGCTGGGCGGGCCCGGG

CGGCCCACGACCCCGGTCCCGGGGAGCAGCCCCCCGTCCCCCGCCTCGACCCCCACGCCC 720
GCCGGGTGCTGGGGCCAGGGCCCCTCGTCGGGGGGCAGGGGGCGGAGCTGGGGGTGCGGG

CCCAAGCGCGGGCGCTACGTCGTCGAGCACCCCGAGTACGGGCCCCCGCCCGACCCCGAG 780
GGGTTCGCGCCCGCGATGCAGCAGCTCGTGGGGCTCATGCCCGGGGGCGGGCTGGGGCTC

GAGGTGCGCGTCCACGGCGCGCGGGGCCCCGGCGCCTTCTGCGCGGCCCCCTGGCGCCCC 840
CTCCACGCGCAGGTGCCGCGCGCCCCGGGGCCGCGGAAGACGCGCCGGGGGACCGCGGGG

GACACGCGGCGCCTCGGGGCCGACGTGAACCGCCTCTTTCGCGGCATCGCCGTCTCGGCC 900
CTGTGCGCCGCGGAGCCCCGGCTGCACTTGGCGGAGAAAGCGCCGTAGCGGCAGAGCCGG

GCCGACGTGACGGGCGACACGCGCGCCCTGCGCCGCGCCCTCTTTGACTTTTACGCCATG 960
CGGCTGCACTGCCCGCTGTGCGCGCGGGACGCGGCGCGGGAGAAACTGAAAATGCGGTAC

GGCTACACGCGCCAGCGCCCCTCGGCCCCCTGCTGGCAGGCCCTGCTCCAGCTCTCGCCC 1020
CCGATGTGCGCGGTCGCGGGGAGCCGGGGGACGACCGTCCGGGACGAGGTCGAGAGCGGG

GAGCAGAGCGCCCCGCTGCGCAGCGCGCTGCGCGAGCTCAACGAGCGCGACGTCTACGAC 1080
CTCGTCTCGCGGGGCGACGCGTCGCGCGACGCGCTCGAGTTGCTCGCGCTGCAGATGCTG

CCGCGCGTCCTCTCCCCGCCGGTCATCGAGGGCCCGCTCTTTGGGGAGGAGTGCGACGTC 1140
GGCGCGCAGGAGAGGGGCGGCCAGTAGCTCCCGGGCGAGAAACCCCTCCTCACGCTGCAG

GACGAGGACGACGCCGGCTCGGACACCACCGTCGCGTCCGAGTTTAGCTTCCGCGGCTCG 1200
CTGCTCCTGCTGCGGCCGAGCCTGTGGTGGCAGCGCAGGCTCAAATCGAAGGCGCCGAGC

GTGTGCGAGGACGACGGGGAGGACGAGGACGAAGAGGAAGACGGGGAGGAGGAAGACGAG 1260
CACACGCTCCTGCTGCCCCTCCTGCTCCTGCTTCTCCTTCTGCCCCTCCTCCTTCTGCTC

GACGAGGAGGGGGAAGAGGAAGAAGACGAGGAGGAAGAGGAAGGGGACGAGGACGGGGAG 1320
CTGCTCCTCCCCCTTCTCCTTCTTCTGCTCCTCCTTCTCCTTCCCCTGCTCCTGCCCCTC

ACGGACGTGTACGAGGAGGACGACGAGGCCGAGGACGAGGAGGACGAGGAGGACGGGGAC 1380
TGCCTGCACATGCTCCTCCTGCTGCTCCGGCTCCTGCTCCTCCTGCTCCTCCTGCCCCTG

GACTTTGACGGGGCCAGCGTGGGCGACGACGACGTGTTTGAGCCCCCCGAGGACGGCTCG 1440
CTGAAACTGCCCCGGTCGCACCCGCTGCTGCTGCACAAACTCGGGGGGCTCCTGCCGAGC

GACGGAGAGGGCTCGGGCTCGGACGACGGCGGGGACGGGGAAGATGAAGACGAGGACGAA 1500
CTGCCTCTCCCGAGCCCGAGCCTGCTGCCGCCCCTGCCCCTTCTACTTCTGCTCCTGCTT

GATGAAGACGAGGACGAGGACGATGGAGAGGACGAGGAGGACGAGGAAGGGGAGGACGGG 1560
CTACTTCTGCTCCTGCTCCTGCTACCTCTCCTGCTCCTCCTGCTCCTTCCCCTCCTGCCC

GGGGAAGACGGCGAAGACGGTGAAGAAGACGAAGACGAAGACGGAGAGGGCGAGGAGGGC 1620
CCCCTTCTGCCGCTTCTGCCACTTCTTCTGCTTCTGCTTCTGCCTCTCCCGCTCCTCCCG

GGGAAGGACGCCGCCCGCCGGGGGACGCGCGCCCCGACGCGGCCCGCCGCCGCCCCGTGA 1680
CCCTTCCTGCGGCGGGCGGCCCCCTGCGCGCGGGGCTGCGCCGGGCGGCGGCGGGGCACT

GGCGGGCCTCGCCCCGCGCTGTTCTTTCCTCCCCCCACCCCCCCCACCCCACCGTCCGCC 1740
CCGCCCGGAGCGGGGCGCGACAAGAAAGGAGGGGGGTGGGGGGGGTGGGGTGGCAGGCGG

CCATCGTTTGCCCCTCTCCCCCTCCCCTCTGTTGTGCCCTCAATAAACACGGCGGCCCGC 1800
GGTAGCAAACGGGGAGAGGGGGAGGGGAGACAACACGGGAGTTATTTGTGCCGCCGGGCG

CGCTCGAACCTCAACTCTCTCGTCTCTCGGGCGTTTTTTCCCTCCCGGCCCTCGTGGGGA 1860
GCGAGCTTGGAGTTGAGAGAGCAGAGAGCCCGCAAAAAAGGGAGGGCCGGGAGCACCCCT

R1: ←——— 1 ———→←———

AGGGAGATGGGGGTGGAGACGGTGGAGGGAGAGGGGAGTGGGATGGGGTGGAGACGGTGG 1920
TCCCTCTACCCCCACCTCTGCCACCTCCCTCTCCCCTCACCCTACCCCACCTCTGCCACC

2 ———→←——— 3 ———→←———

AGGGAGAGGGGAGTGGGATGGGGTGGAGACGGTGGAGGGAGAGGGGAGTGGGATGGGGTG 1980
TCCCTCTCCCCTCACCCTACCCCACCTCTGCCACCTCCCTCTCCCCTCACCCTACCCCAC

——— 4 ———→←——— 5 ———

GAGACGGTGGAGGGAGAGGGGAGTGGGATGGGGTGGAGACGGTGGAGGGAGAGGGGAGTG 2040
CTCTGCCACCTCCCTCTCCCCTCACCCTACCCCACCTCTGCCACCTCCCTCTCCCCTCAC

—→←——— 6 ———→←——— 7 –

GGATGGGGTGGAGACGGTGGAGGGAGAGGGGAGTGGGATGGGGTGGAGACGGTGGAGGGA 2100
CCTACCCCACCTCTGCCACCTCCCTCTCCCCTCACCCTACCCCACCTCTGCCACCTCCCT

———→←——— 8 ———→←———

GAGGGGAGTGGGATGGGGTGGAGACGGTGGAGGGAGAGGGGAGTGGGATGGGGTGGAGAC 2160
CTCCCCTCACCCTACCCCACCTCTGCCACCTCCCTCTCCCCTCACCCTACCCCACCTCTG

— 9 ———→←——— 10 ———→←

GGTGGAGGGAGAGGGGAGTGGGATGGGGTGGAGACGGTGGAGGGAGAGGGGAGTGGGATG 2220
CCACCTCCCTCTCCCCTCACCCTACCCCACCTCTGCCACCTCCCTCTCCCCTCACCCTAC

11 12

GGGTGGAGACGGTGGAGGGAGAGGGGAGTGGGATGGGGTGGAGACGGTGGAGGGAGAGAA 2280
CCCACCTCTGCCACCTCCCTCTCCCCTCACCCTACCCCACCTCTGCCACCTCCCTCTCTT

R2:

AGGACGGAGGCGTGTCATCGGGTGCCAGAGCGAGAGCGTTTATTGTTAAAGTTGTTGGTG 2340
TCCTGCCTCCGCACAGTAGCCCACGGTCTCGCTCTCGCAAATAACAATTTCAACAACCAC

1 2 3 4 5 6

GGGGGGGAGTCCGGGGGAGTCCGGGGGAGTCCGGGGGAGTCCGGGGGAGTCCGGGGGAGT 2400
CCCCCCCTCAGGCCCCCTCAGGCCCCCTCAGGCCCCCTCAGGCCCCCTCAGGCCCCCTCA

7 8 9

CCGGGGGAGTCCGGGGGAGTCCGGGGGAGTCCTCGGCGGCTATTGGTGGGAAGGGGGAGT 2460
GGCCCCCTCAGGCCCCCTCAGGCCCCCTCAGGAGCCGCCGATAACCACCCTTCCCCCTCA

GACGTCAGGGCAGCGGGGGGGGGGGAAGGGGGGGAAGGGGGGAGATGGGGGAGGGAGAGA 2520
CTGCAGTCCCGTCGCCCCCCCCCCCTTCCCCCCCTTCCCCCCTCTACCCCCTCCCTCTCT

Rs Us

GAGGACGGAGGCGGCGCGAGTCTGTCGGATGGCGCCAGGGGGGGGCGGGCGGCCGCGGGG 2580
CTCCTGCCTCCGCCGCGCTCAGACAGCCTACCGCGGTCCCCCCCCGCCCGCCGGCGCCCC

AGGGATGACGGGCGCGCGTGGGGCGGCGGCGCCCCCGTCGCGGTCGAGAACCACCGCCGC 2640
TCCCTACTGCCCGCGCGCACCCCGCCGCCGCGGGGGCAGCGCCAGCTCTTGGTGGCGGCG

CGTCACCGCCGCCTCCCATCCGATGTGATATCGCGGCACGCCGGCCGTCCCGGCGCTCAT 2700
GCAGTGGCGGCGGAGGGTAGGCTACACTATAGCGCCGTGCGGCCGGCAGGGCCGCGAGTA

TCACACCGCACCCGTTCGCCCACGTCCCCGCGGCCAAGCACACACACCAACTCGCGCACC 2760
AGTGTGGCGTGGGCAAGCGGGTGCAGGGGCGCCGGTTCGTGTGTGTGGTTGAGCGCGTGG

ATGTTGGCGATGTGGAGATGGGTCACCAAGAGGTCGCGGCTCCGCCGAGGCCACGCCCAT 2820
TACAACCGCTACACCTCTACCCAGTGGTTCTCCAGCGCCGAGGCGGCTCCGGTGCGGGTA

CTTGGGGGAAATAAAGGAGTCCGGGGAATTTGTTCCTTATACCTTGCCGGGCTCAGCAGG 2880
GAACCCCCTTTATTTCCTCAGGCCCCTTAAACAAGGAATATGGAACGGCCCGAGTCGTCC

GGGTTGTCGCGCGTCCACGCCCAGCGCTCGCACGCAGCAACAATGGCCGACGCCGGAATC 2940
CCCAACAGCGCGCAGGTGCGGGTCGCGAGCGTGCGTCGTTGTTACCGGCTGCGGCCTTAG

CCCGACGAGATCCTGTACTCGGACATCAGCGACGACGAGATCATCATCGACGGCGACGGC 3000
GGGCTGCTCTAGGACATGAGCCTGTAGTCGCTGCTGCTCTAGTAGTAGCTGCCGCTGCCG

GACAGCAGCGGGGACGAGGACGACGATGACGGGGGGCTGACGCGGCAGGCCGCGGCGCGC 3060
CTGTCGTCGCCCCTGCTCCTGCTGCTACTGCCCCCCGACTGCGCCGTCCGGCGCCGCGCG

ATCGTCACGGACCTGGGCTTCGAGGTGCTGCAGCCCCTGCAGTCGGGCTCGGAGGGCCGC 3120
TAGCAGTGCCTGGACCCGAAGCTCCACGACGTCGGGGACGTCAGCCCGAGCCTCCCGGCG

GTCTTCGTGGCCCGCCGGCCGGGCGAGGCGGACACGGTGGTGCTGAAGGTGGGCCAGAAG 3180
CAGAAGCACCGGGCGGCCGGCCCGCTCCGCCTGTGCCACCACGACTTCCACCCGGTCTTC

CCCTCGACGCTGATGGAGGGCATGCTGCTGCAGCGCCTGTCCCACGATAACGTCATGCGC 3240
GGGAGCTGCGACTACCTCCCGTACGACGACGTCGCGGACAGGGTGCTATTGCAGTACGCG

ATGAAACAGATGCTCGCCCGGGGCCCGGCGACGTGCCTGGTCCTGCCGCACTTTCGGTGC 3300
TACTTTGTCTACGAGCGGGCCCCGGGCCGCTGCACGGACCAGGACGGCGTGAAAGCCACG

GATCTGTACAGCTACCTGACCATGCGGGACGGGCCGCTGGACATGCGCGACGCCGGGTGC 3360
CTAGACATGTCGATGGACTGGTACGCCCTGCCCGGCGACCTGTACGCGCTGCGGCCCACG

GTGATCCGGGCCGTGCTCCGCGGGCTCGCCTACCTGCACGGGATGCGCATCATGCATCGC 3420
CACTAGGCCCGGCACGAGGCGCCCGAGCGGATGGACGTGCCCTACGCGTAGTACGTAGCG

GACGTCAAGGCGGAGAACATCTTCCTCGAGGACGTGGACACGGTGTGCCTAGGGGACCTC 3480
CTGCAGTTCCGCCTCTTGTAGAAGGAGCTCCTGCACCTGTGCCACACGGATCCCCTGGAG

GGGGCCGCGCGCTGCAACGTGGCGGCGCCCAACTTTTACGGGCTCGCCGGGACCATCGAG 3540
CCCCGGCGCGCGACGTTGCACCGCCGCGGGTTGAAAATGCCCGAGCGGCCCTGGTAGCTC

ACCAACGCCCCCGAGGTGCTCGCGCGCGACCGCTACGACACCAAGGTCGACGTCTGGGGC 3600
TGGTTGCGGGGGCTCCACGAGCGCGCGCTGGCGATGCTGTGGTTCCAGCTGCAGACCCCG

GCGGGGGTGGTGCTCTTCGAGACGCTGGCCTACCCCAAGACGATCACCGGCGGGGACGAG 3660
CGCCCCCACCACGAGAAGCTCTGCGACCGGATGGGGTTCTGCTAGTGGCCGCCCCTGCTC

CCCGCGATCAACGGGGAGATGCACCTGATCGACCTCATCCGCGCCCTCGGGGTGCACCCC 3720
GGGCGCTAGTTGCCCCTCTACGTGGACTAGCTGGAGTAGGCGCGGGAGCCCCACGTGGGG

GAGGAGTTCCCGCCCGACACGCGCCTCCGGAGCGAGTTCGTCCGGTACGCCGGGACCCAC 3780
CTCCTCAAGGGCGGGCTGTGCGCGGAGGCCTCGCTCAAGCAGGCCATGCGGCCCTGGGTG

CGCCAGCCGTACACGCAGTACGCGCGCGTGGCTCGCCTCGGGCTGCCCGAGACGGGGGCT 3840
GCGGTCGGCATGTGCGTCATGCGCGCGCACCGAGCGGAGCCCGACGGGCTCTGCCCCCGA

TTCCTGATTTACAAGATGTTGACGTTTGATCCCGTCCGCCGCCCTTCCGCTGATGAGATA 3900
AAGGACTAAATGTTCTACAACTGCAAACTAGGGCAGGCGGCGGGAAGGCGACTACTCTAT

CTCAACTTTGGAATGTGGACCGTATAAAACGGGCCGGCTCCGAGCGGTAGGACACACACA 3960
GAGTTGAAACCTTACACCTGGCATATTTTGCCCGGCCGAGGCTCGCCATCCTGTGTGTGT

CCTTTGCGCATCTCCACAGCTCAACAATGAAGTGGGCAACGTGGATCC 4008
GGAAACGCGTAGAGGTGTCGAGTTGTTACTTCACCCGTTGCACCTAGG

sequence already determined. This allowed the determination of the nucleotide sequence on both strands with overlap of all the restriction sites. As mentioned in the Introduction, the DNA of PRV is relatively GC-rich (73%). The major problems encountered in sequencing such GC-rich DNA are band 'compressions' and 'false stops'. The band 'compressions' (closely spaced bands or even wrong migration on the sequencing gels) arise from secondary structure being retained in the gel despite the presence of 8 M urea, and this causes anomalous migration. The 'false stops' (bands in all four lanes) result from termination of the polymerization of the chain at regions where secondary structure is present during the sequencing reaction. As a matter of fact, no strong 'false stops' was observed in this region, and, although there were several regions with 'compressions' found on one strand, these were always resolved by determination of the nucleotide sequence on the opposite strand, where any corresponding compression was at a different position.

Much greater difficulty was encountered in sequencing the *Bam*HI 10 fragment in the region between *Sst*I and *Eco*RV. The first problem was that the restriction map revealed no suitable restriction endonuclease sites in this 1.7 kb fragment for further subcloning. For this reason, this fragment was cloned into M13mp18, and a series of clones with different sized deletions in the insert were generated from the double-stranded RF DNA using the DNase I deletion subcloning method (see section 2.10.2). These clones allowed the nucleotide sequence in this region to be determined on overlapping fragments from different starting points, except for one small gap. On the basis of this partial nucleotide sequence, restriction endonucleases (*Hin*PI and *Taq*I) with four base-pair recognition sites were used to generate random clones which were sequenced in single-stranded M13 templates. The sequencing data obtained from the deletion clones and random clones covered this fragment on both strands, except for some small regions where the determination was only performed on one strand. In order to determine the

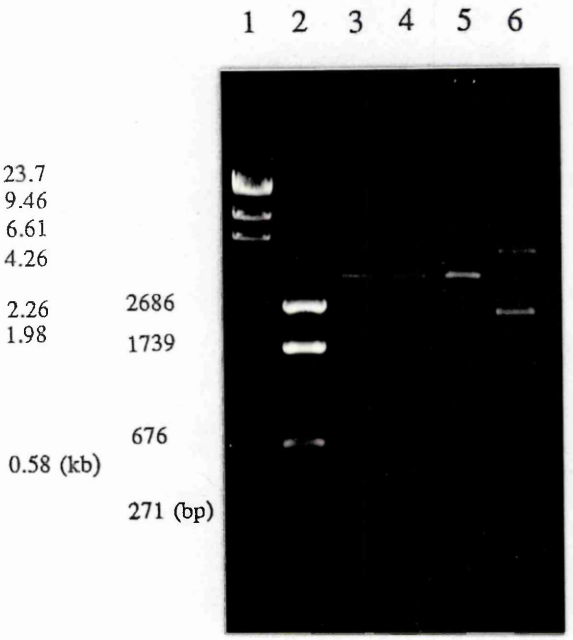
nucleotide sequence on both strands, the fragments containing those regions were further cloned into pUC18, and parts of the inserts were deleted using the exonuclease III deletion subcloning method (see section 2.10.3). Then, the clones with deletions were directly sequenced from the plasmid DNA. From these combined results the nucleotide sequence of this 1.7 kb fragment was obtained on both strands except for one small stretch of 25 nt (2276-2300), although this was determined from three separate clones.

The 1.7 kb fragment also suffered quite badly from the two problems mentioned above: 'false stops' and 'compressions'. The sequences in the regions of 'false stops' were confidently read on the sequencing gels from the opposite strands, as explained above, and 'compressions' were resolved in the same way. In addition, the compressions were eliminated by replacing dGTP with its analogue, 7-deaza-2'-dGTP, which forms base pairs more weakly than dGTP. It should be stressed that this generally had the side-effect of producing 'false stops' elsewhere in the sequence, so that gels both with and without the analogue were required.

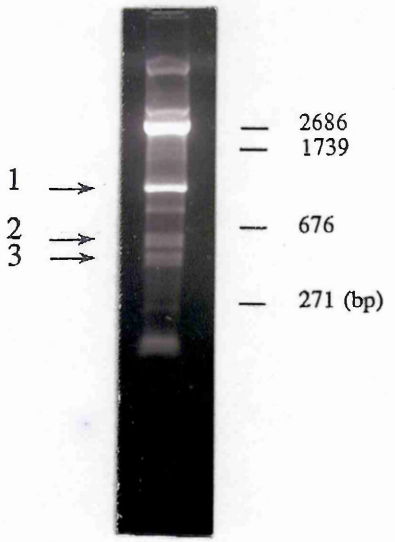
For the 1 kb fragment between the *Bam*HI and *Sst*I sites, the nucleotide sequence was initially determined on single-stranded M13 templates using the restriction sites shown in Fig. 3.3.a, but only part of the sequence was obtained in this way. A major problem in sequencing this fragment was the fact that the *Sal*I fragment (nt 337-648) could not be cloned into M13 without suffering large deletions of varying sizes. This fragment did appear to be stable in pUC18 because no variation was observed in the size of the insert in the plasmid. It was not, however, possible to determine accurately more than 60 bp from each end of the fragment in the plasmid DNA because of extremely strong 'false stops'. It seemed that both this problem and the deletions might have arisen from a palindromic structure, and experiments with S1 nuclease was performed to determined whether this was so. The rationale for this was the fact that the single-stranded portion of the stem of looped-out DNA from such palindromes is susceptible to digestion by S1

Fig. 3.5. Determination of S1-sensitive sites in pB10-BS1 DNA. (a) Effect of S1 nuclease digestion on pB10-BS1 DNA. The pB10-BS1 DNA was incubated with S1 nuclease, and subjected to electrophoresis in 1% agarose (lane 3 and 4). Undigested DNA and pB10-BS1 DNA linearized with *Hin*dIII are shown in lanes 6 and 5, respectively, and the two size markers used in Fig. 3.1 are presented in lanes 1 and 2. (b) Determination of sites susceptible to S1 nuclease in pB10-BS1. DNA of pB10-BS1 treated with S1 nuclease was digested with *Bam*HI, and subjected to electrophoresis in 1% agarose. The arrow (1) shows the released insert, the *Bam*HI-*Sst*I 1 kb fragment (see text). The two sets of small fragment (*ca.* 0.45 kb and 0.55 kb) generated by this double digestion are marked with arrows 2 and 3. The size marker was pUC18 DNA digested with *Eco*RI and with *Rsa*I (see Fig. 3.1).

a.

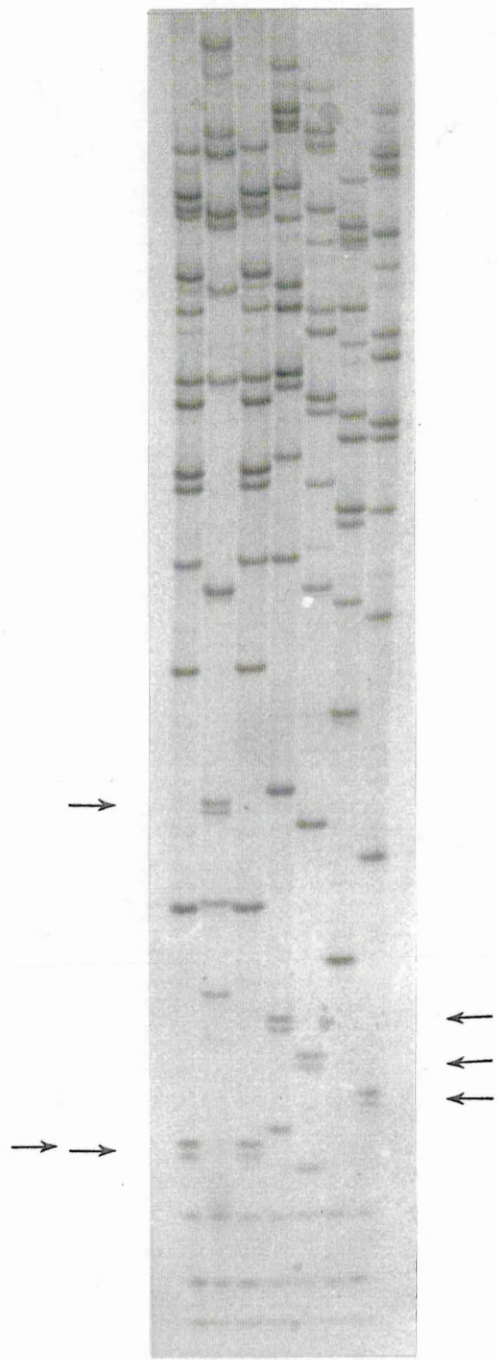

1 2 3 4 5 6
23.7
9.46
6.61
4.26
2.26
1.98
0.58 (kb)
2686
1739
676
271 (bp)

b.


1
2
3
2686
1739
676
271 (bp)

nuclease (*cf.* Stow, 1985). As can be seen in Fig. 3.5.a, the DNA of plasmid pB10-BS1 contains supercoiled and relaxed circular DNA. After incubation with S1 nuclease, the supercoiled DNA band was replaced by a band corresponding to linearized plasmid DNA, although the relaxed form was not affected. These results suggested that there was a position of susceptibility to S1 nuclease in the supercoiled circular plasmid DNA of pB10-BS1, but not in the relaxed circular form. To determine whether the linearization by S1 nuclease occurred at a single cleavage point, the pB10-BS1 DNA was digested with S1 nuclease followed by *Bam*HI, for which there were flanking sites on each side of the insert. The result in Fig. 3.5.b. shows that this double digestion generated not only the free insert, but also two new sets of fragments with average sizes of approximately 0.55 kb and 0.45 kb. These new fragments were cloned after ligating between the *Hinc*II-*Bam*HI sites of M13mp18, and the nucleotide sequence was determined. The sequencing results showed a variety of cleavage points by S1 nuclease. Fig. 3.6 illustrates this for clones on the 5' site of the cleavage point. The largest fragments extended to near each end of the region of 26 Cs (at approx. nucleotide 480 in Fig. 3.4). Surprisingly, no potentially palindromic structure could be found by inspection of the sequence (other than those that might be formed by long-range interaction of these Cs with G-rich regions; *e.g.* approximately nucleotide 390 and 420). There are three reasons for regarding that this result dose not reflect a sequencing error. First, the size of the *Sal*I fragment (306 nt) obtained from sequencing results agrees very well with the value obtained from restriction analysis (305 ± 5 nt, see Fig. 3.17). Second, the nucleotide sequence determined on one M13 clone, starting from the *Nco*I site (approx. nucleotide 585) in the direction of the *Bam*HI site (nucleotide 1), overlapped the gap generated by S1 nuclease cleavage, before a deletion was encountered at a position near the 5'-end of the 26 Cs. Third, the sequence in this region could also be generally confirmed in a pUC18 clone from the *Nla*IV site

Fig. 3.6. Identification of S1 nuclease cleavage points by nucleotide sequencing. The two sets of the fragments (*ca.* 0.45 kb and 0.55 kb, see Fig. 3.5.b) generated by S1 nuclease and *Bam*HI double digestion were sequenced in M13mp18 (see section 2.11). An autoradiograph of a sequencing gel shows the T reactions of seven clones obtained from one (0.55 kb) set of the fragments. The six sets of double bands indicated by arrows correspond to nucleotides 521 and 522 in Fig. 3.4.



near the 5'-end of the 26 Cs in the direction of the *Sst*I site. Thus, the reason for the susceptibility to S1 nuclease in this 1 kb fragment remains unclear.

The nucleotide sequence in this *Bam*HI-*Sst*I fragment was also determined in pUC18 using randomly cloned fragments generated by *Nla*IV. This allowed all of the cleavage points to be fully overlapped, and the fragment to be completely sequenced on both strands.

Although the determination of the nucleotide sequence was performed in three sections, each of these sections was completely overlapped by different clones (Fig. 3.3.b). As can be seen in Fig. 3.4, the *Bam*HI 10 fragment contains a total of 4,008 bp with 71% GC, and can be regarded as being divided into two parts, 1-1869 nt and 2433-4008 nt, by two regions of reiterated DNA. The first reiterated region (R_1) contains 11 full and apparently identical copies of a 35 nt repeat and 24 nt of a twelfth copy, and the second (R_2) has 9 full and identical copies of a 10 nt repeat. It must be emphased that the nucleotide sequence of the region R_1 was too long for each component repeat to be read in detail. However, the nucleotide sequence was determined starting from both 5' and 3'-ends of the R_1 region, and the sequencing gels were read with complete confidence for the first five repeats and for 19 nt of the sixth; and for the last two repeats (10 and 11) and for the 24 nt of the twelfth. The total number of copies was counted from the specific pattern of the reiterated sequence on the gels, and the number of the copies counted from each end of R_1 on both normal gels and one run for a longer time was exactly same. It is unlikely, but cannot be excluded, that some of the repeats deviate from those read with certainty.

Fig. 3.7 shows all of the possible open reading frames in this 4 kb fragment, of which the longest two, designated A and B, will be considered in section 3.3 and 3.2, respectively.

Fig. 3.7. Alignment of possible open reading frames in the *Bam*HI 10 fragment of PRV. The possible open reading frames on each frame of the *Bam*HI 10 fragment of PRV are presented on the basis of analysis using the program Frames (see section 2.17). Those shown in frame a, b, and c lie in the forward strand in Fig. 3.4, starting from the first (frame a), second (frame b), and third (frame c) nucleotide. Those shown in frame d, e, and f lie in the reverse strand in Fig. 3.4, starting from nucleotides 4008 (frame d), 4007 (frame e), and 4006 (frame f). The translational initiation codon (..⊥..) and the stop codon (··⊣··) are marked, and the longest two possible open reading frame are also marked as A and B.


a
A
B
Forward translation
b
5'
3'
c
0
1
2
3
4
kb
e
Reverse translation
f
3'
5'
g

3.2. The *pk* gene

3.2.1. The open reading frame

As mentioned in the Introduction and Section 3.1, the PRV protein kinase gene was believed to be located in the 'right hand' side of the *Bam*HI 10 fragment, and Fig. 3.7 shows that there is a potential open reading frame (B) in this region. Fig. 3.8 shows the predicted amino acid sequence of the gene lying in this open reading frame, which is thought to occupy only the latter part of this, as discussed further below. The open reading frame indicated in the region between nucleotide 2923 and 3924 encodes a predicted protein of 334 amino acid residues. A Dot-matrix comparison using the program DotPlot (see section 2.17) was made between this amino acid sequence and those of protein kinases predicted to be encoded by HSV-1 gene US3 and VZV gene 66. The results of this analysis show that this amino acid sequence has extensive homology to that of the products of HSV-1 US3 and VZV 66 genes (Fig. 3.9), although the N-terminal regions were distinct. It is clear that this region of the gene of PRV contains the homologue of HSV-1 US3 and VZV 66, and the gene was therefore designated *pk*. A more detailed comparison of these amino acid sequences is presented in section 4.

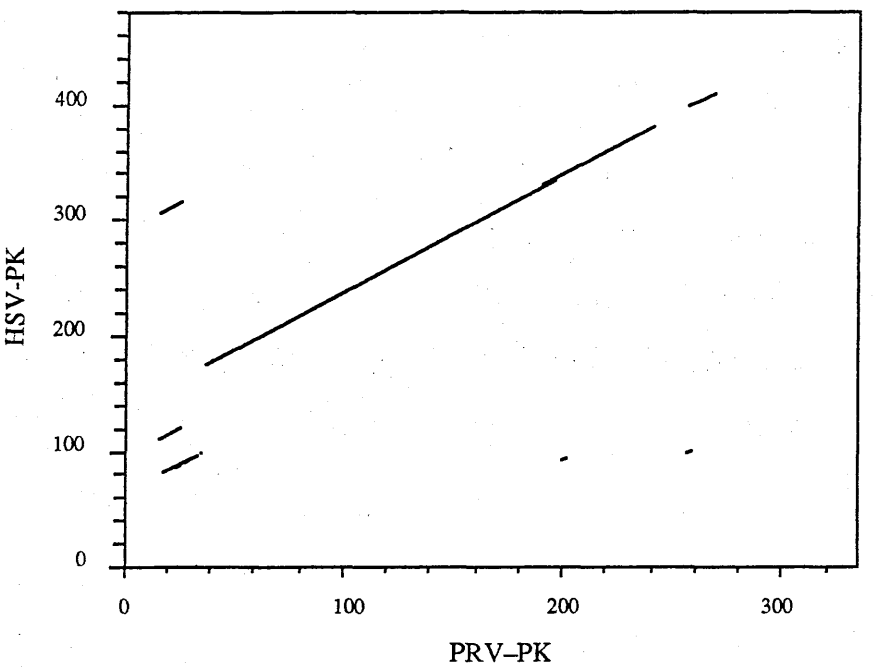
In Fig. 3.8 it can be seen that there are several in-frame Met codons 5' of this open reading frame, which could have been chosen as alternatives for the translational initiation site. The basis for the choice made rests on analysis with two computer programs, CodonPreference and TestCode (see section 2.17), which make independent assessments of the likelihood that an open reading frame encodes protein. The results from both analyses (Fig. 3.10) supported the choice made, which is in fact the first Met codon 5' of the coding region of the catalytic domain (section 4.1). It can also be seen in Fig. 3.8 that there are two AT-rich regions (underlined in Fig. 3.8) 5' of the predicted translational initiation site, which might

Fig. 3.8. Nucleotide sequence of the *Eco*RV-*Bam*HI region of the *Bam*HI 10 fragment of PRV DNA. The conceptual translation of the region in the possible open reading frame B is shown, which is predicted to be the homologue of the product of US3 of HSV-1. The N-terminal portion of the translation at the 3' region is that of the gX gene, and the one-letter code is used for presenting amino acid sequences. The two in-frame Met codons upstream of the predicted translational initiation codon are underlined, and the broken overlines indicate the AT-rich regions that might contain a TATA box.

```
GATATCGCGGCACGCCGGCCGTCCCGGCGCTCATTCACACCGCACCCGTTCGCCCACGTCCCCGCGGCCAAGCACACACACCAACTCGCGCACCATGTTGGCGATGTGGAGATGGGTCAC   120

CAAGAGGTCGCGGCTCCGCCGAGGCCACGCCCATCTTGGGGGAAATAAAGGAGTCCGGGGAATTTGTTCCTTATACCTTGCCGGGCTCAGCAGGGGGTTGTCGCGCGTCCACGCCCAGCG   240

                  M  A  D  A  G  I  P  D  E  I  L  Y  S  D  I  S  D  D  E  I  I  I  D  G  D  G  D  S  S  G  D  E  D  D  D    35
CTCGCACGCAGCAACAATGGCCGACGCCGGAATCCCCGACGAGATCCTGTACTCGGACATCAGCGACGACGAGATCATCATCGACGGCGACGGCGACAGCAGCGGGGACGAGGACGACGA   360

  D  G  G  L  T  R  Q  A  A  A  R  I  V  T  D  L  G  F  E  V  L  Q  P  L  Q  S  G  S  E  G  R  V  F  V  A  R  R  P  G  E    75
TGACGGGGGGCTGACGCGGCAGGCCGCGGCGCGCATCGTCACGGACCTGGGCTTCGAGGTGCTGCAGCCCCTGCAGTCGGGCTCGGAGGGCCGCGTCTTCGTGGCCCGCCGGCCGGGCGA   480

  A  D  T  V  V  L  K  V  G  Q  K  P  S  T  L  M  E  G  M  L  L  Q  R  L  S  H  D  N  V  M  R  M  K  Q  M  L  A  R  G  P   115
GGCGGACACGGTGGTGCTGAAGGTGGGCCAGAAGCCCTCGACGCTGATGGAGGGCATGCTGCTGCAGCGCCTGTCCCACGATAACGTCATGCGCATGAAACAGATGCTCGCCCGGGGCCC   600

  A  T  C  L  V  L  P  H  F  R  C  D  L  Y  S  Y  L  T  M  R  D  G  P  L  D  M  R  D  A  G  C  V  I  R  A  V  L  R  G  L   155
GGCGACGTGCCTGGTCCTGCCGCACTTTCGGTGCGATCTGTACAGCTACCTGACCATGCGGGACGGGCCGCTGGACATGCGCGACGCCGGGTGCGTGATCCGGGCCGTGCTCCGCGGGCT   720

  A  Y  L  H  G  M  R  I  M  H  R  D  V  K  A  E  N  I  F  L  E  D  V  D  T  V  C  L  G  D  L  G  A  A  R  C  N  V  A  A   195
CGCCTACCTGCACGGGATGCGCATCATGCATCGCGACGTCAAGGCGGAGAACATCTTCCTCGAGGACGTGGACACGGTGTGCCTAGGGGACCTCGGGGCCGCGCGCTGCAACGTGGCGGC   840

  P  N  F  Y  G  L  A  G  T  I  E  T  N  A  P  E  V  L  A  R  D  R  Y  D  T  K  V  D  V  W  G  A  G  V  V  L  F  E  T  L   235
GCCCAACTTTTACGGGCTCGCCGGGACCATCGAGACCAACGCCCCCGAGGTGCTCGCGCGCGACCGCTACGACACCAAGGTCGACGTCTGGGGCGCGGGGGTGGTGCTCTTCGAGACGCT   960

  A  Y  P  K  T  I  T  G  G  D  E  P  A  I  N  G  E  M  H  L  I  D  L  I  R  A  L  G  V  H  P  E  E  F  P  P  D  T  R  L   275
GGCCTACCCCAAGACGATCACCGGCGGGGACGAGCCCGCGATCAACGGGGAGATGCACCTGATCGACCTCATCCGCGCCCTCGGGGTGCACCCCGAGGAGTTCCCGCCCGACACGCGCCT  1080

  R  S  E  F  V  R  Y  A  G  T  H  R  Q  P  Y  T  Q  Y  A  R  V  A  R  L  G  L  P  E  T  G  A  F  L  I  Y  K  M  L  T  F   315
CCGGAGCGAGTTCGTCCGGTACGCCGGGACCCACCGCCAGCCGTACACGCAGTACGCGCGCGTGGCTCGCCTCGGGCTGCCCGAGACGGGGGCTTTCCTGATTTACAAGATGTTGACGTT  1200

  D  P  V  R  R  P  S  A  D  E  I  L  N  F  G  M  W  T  V                                                                   334
TGATCCCGTCCGCCGCCCTTCCGCTGATGAGATACTCAACTTTGGAATGTGGACCGTATAAAACGGGCCGGCTCCGAGCGGTAGGACACACACACCTTTGCGCATCTCCACAGCTCAACA  1320

 M  K  W  A  T  W  I
ATGAAGTGGGCAACGTGGATCC  1342
```

Fig. 3.9. Dot-Matrix comparison of the similarity of the predicted amino acid sequence encoded by the open reading frame B with those of other alphaherpesvirus protein kinases. The comparison was made using Compare, and aligned using DotPlot (see section 2.17). The predicted amino acid sequence encoded by the open reading frame B (see Fig. 3.8) is presented on ordinate, and those of US3 of HSV-1 (a) or gene 66 of VZV (b) is arranged on the abscissa. The amino acid sequence of the product of US3 of HSV-1 is from McGeoch *et al.* (1985), and that of gene 66 of VZV is from Davison and Scott (1986).

a.


400
300
200
100
0
HSV-PK
0
100
200
300
PRV–PK

b.

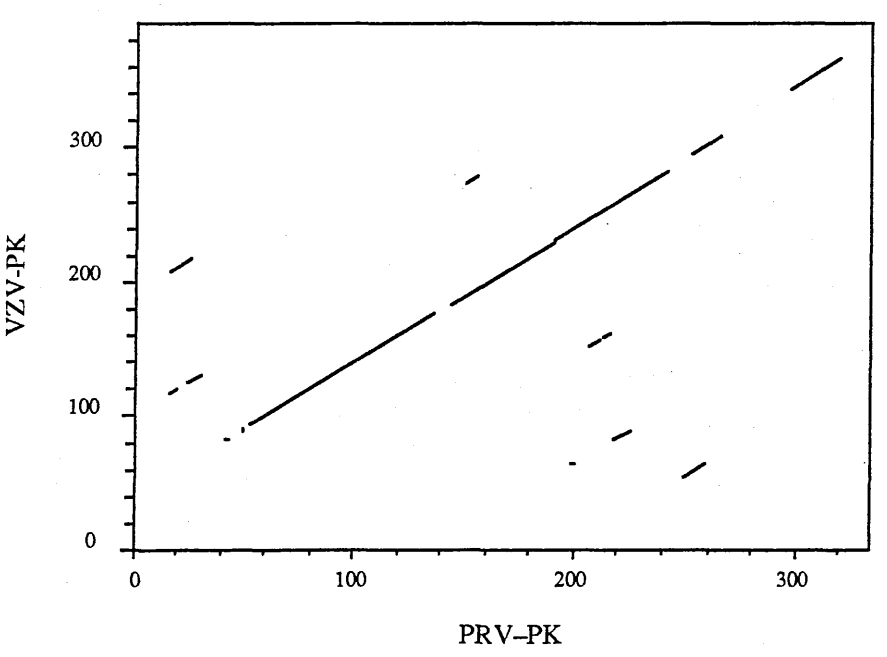
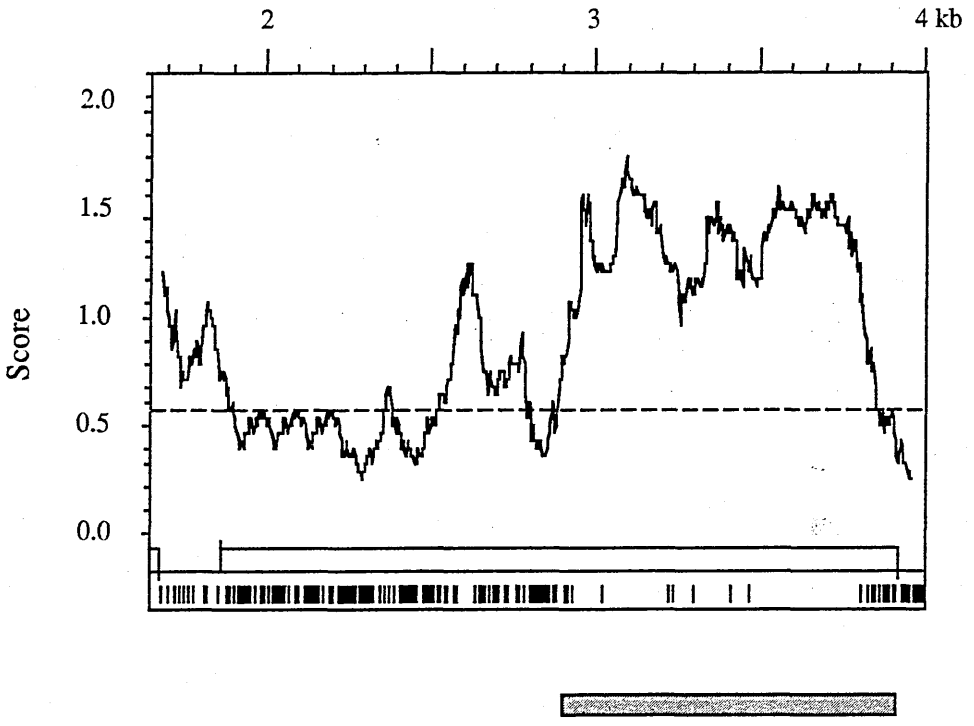
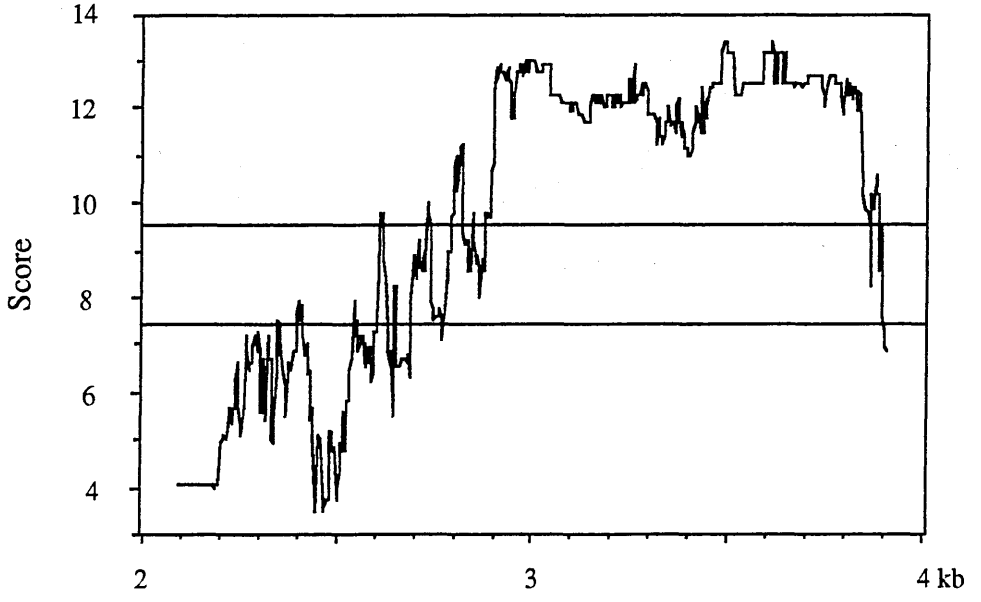

300
200
100
0
VZV-PK
0
100
200
300
PRV–PK

Fig. 3.10. Protein-encoding potential of the open reading frame B (from nucleotides 1800 to 4008). (a) Analysis by CodonPreference. The GCG program, CodonPreference (see section 2.17), was used for analysing the protein-encoding potential of the open reading frame. The regions where the curve rises above the broken line are predicted to have protein-encoding potential, and the grey bar shows the predicted coding region. (b) Analysis by TestCode. The open reading frame B was analyzed by TestCode (see section 2.17). The portions of curve in the upper window are predicted to lie in protein-encoding regions, whereas the portions in the lower window are predicted to lie in non-encoding regions. The middle window is an area for which the program makes no prediction.

a.



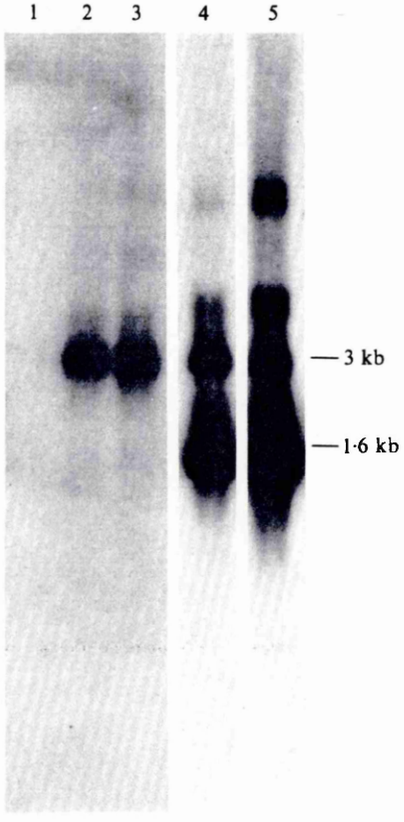
b.



contain (especially in the the case of the more 3' one) a TATA box.

In Fig. 3.8, downstream from the translational termination codon of the gene *pk*, there is no putative polyadenylation/processing signal (AATAAA or ATTAAA) before the initiation codon of gX, or even within the coding sequence of gX (Rea *et al.*, 1985). A similar situation was found regarding US3 and US4 in HSV-1 and HSV-2 (Rixon & McGeoch, 1985; McGeoch *et al.*, 1987), in which case the transcript of US3 was shown to be co-terminal with that of US4. To determine if a similar situation existed in PRV, polyadenylated mRNAs were prepared from cells infected for different times with PRV, separated by electrophoresis in agarose gels containing formaldehyde, transferred to nitrocellulose membrane, and hybridized with a radioactively labelled DNA probe *a*, from nucleotides 3300-3585 of the coding region of the gene *pk* (produced using the polymerase chain reaction, and kindly provided by Dr. B. Favre), and probe *b*, a *Bam*HI-*Nde*I fragment from the coding region of the gene gX (Rea *et al.*, 1985). The results are presented in Fig. 3.11, and, as can be seen from the autoradiograph, the probe from the coding region of gX hybridized most strongly to two mRNAs; approximately 1.6 and 3 kb long, whereas the probe from the coding region of *pk* only hybridized to a species of 3 kb. The 1.6 kb mRNA is the expected size of the transcript of the gene gX alone, of which the TATA box overlaps the termination codon of the gene *pk* (see Fig. 3.8); and the 3 kb mRNA is of the size predicted for a transcript initiated from a promoter preceding the gene *pk* and terminated at the polyadenylation/processing signal following the gene gX. These results suggest that the transcript of the gene *pk* is probably co-terminal with that of the gene gX, with these using a single polyadenylation/processing site.

Fig. 3.11. Hybridization analysis of the transcripts of the *pk* gene. Polyadenylated mRNA was isolated from cells infected for 0 h (lane 1), 3 h (lanes 2 and 4) or 6 h (lanes 3 and 5) with PRV, separated by electrophoresis in 1% agarose under denaturing condition (see section 2.14.2.a), transferred to nitrocellulose and hybridized with radioactively labelled probes *a*, from the *pk* gene (lanes 1 to 3), or *b*, from the gX gene (lane 4 and 5), as described in section 3.2.1 (also see Fig. 3.3.c). The figure shows an autoradiograph with the sizes of the major bands, estimated from comparison with *Hin*dIII digested λ DNA (see section 2.7).

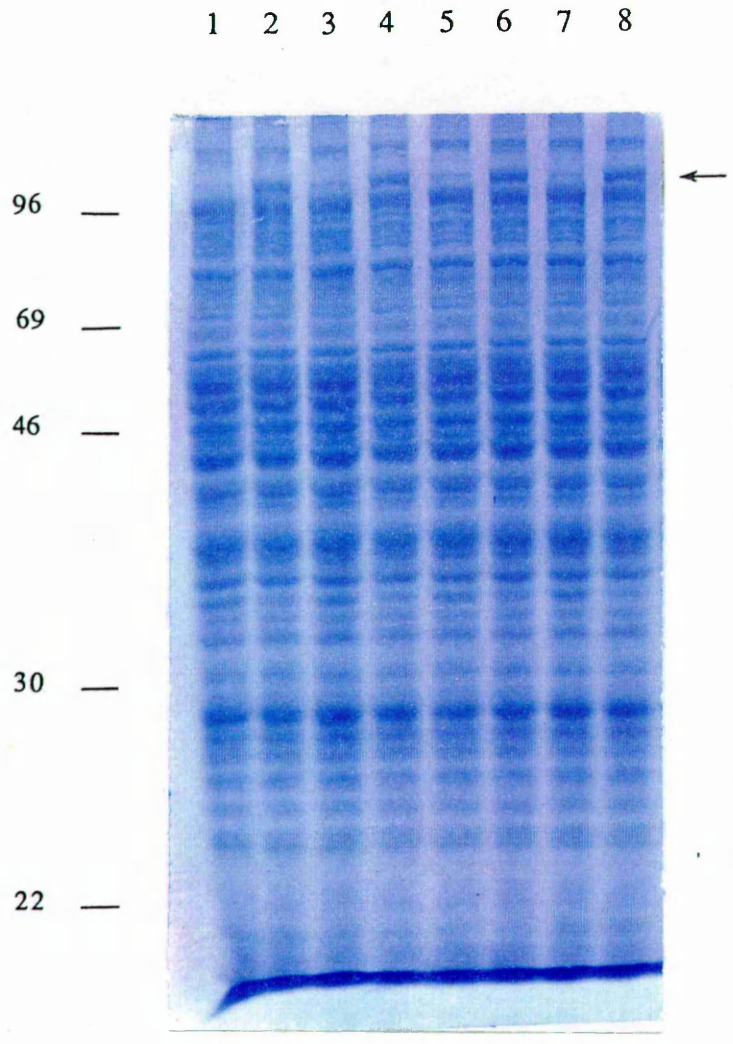

1
2
3
4
5
3 kb
1·6 kb

3.2.2. Expression of the gene *pk* in *E. coli*

In an attempt to express the gene *pk* in *E. coli*, two vectors, pKK233-2 and pJLA502, were chosen (see section 2.6.2), and an *Nco*I site was introduced over the translational initiation site of the *pk* gene by site-directed mutagenesis, as described in section 2.15. The *pk* gene was then subcloned between the *Nco*I-*Hin*dIII sites of pKK233-2, and expression was initially performed in strain JM109 (see section 2.3.1). The time for induction was varied from 2 h to 6 h at 5 mM IPTG, and an induction was also performed for 16 h at 50 μM IPTG, and the total cell extract was analyzed by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulphate (see section 2.16). From the results presented in Fig. 3.12, it can be seen that induction causes some differences in the stained protein bands, but there is no newly expressed protein visible in the region of approximately 40kDa. Next, a *lon*$^-$ strain, MC102 (see section 2.3.1), was used as it was thought that the failure to observe the expression of *pk* could be because of instability of the protein in strain JM109 which contains the *lon* protease (Bruner *et al.*, 1988). The induction was therefore performed in MC102 in a similar manner to that in JM109, but no new protein of 40kDa was observed. The *pk* gene was also subcloned between the *Nco*I-*Bam*HI sites of pJLA502, and induction in strain DH1 (see section 2.3.1) and MC102 was performed from 2 h to 16 h at 42°C. However, expression of the *pk* gene was still not achieved.

It is known that GC-rich DNA is not easy to express in *E coli*. However, this is unlikely to be the reason for the lack of expression in this case, since a part of the *pk* gene has been expressed as a fusion protein in *E coli* (Stevens, 1989). It cannot be excluded that the expression was at a very low level, producing insufficient protein to be seen by staining with Coomassie Blue. In any case, the reason for the poor expression is unclear.

Fig. 3.12. Analysis by polyacrylamide gel electrophoresis of expression of the *pk* gene in *E. coli*. JM109 cells containing pMK-1 and pKK233-2 were lysed in the bacteria lysis buffer, heated at 100°C for 10 min, and subjected to electrophoresis in 10% polyacrylamide gel under denaturing condition (see section 2.16). The gel then was stained with Coomassie brilliant blue R, and destained. Culture of JM109 cells containing pKK233-2 were also analysed after 2 h induction with IPTG (lane 2) or without IPTG (lane 1). Expression in JM109 cells containing pMK-1 was induced for 2 h (lane 4) and 6 h (lane 6) in the presence of 5 mM IPTG, and bacterial cells grown under the same condition, but in the absence of IPTG were also analysed at the same time points (lane 3 and lane 5). The expression was also performed for 16 h in the presence of 50 μM IPTG (lane 8), or in the absence of IPTG (lane 7). The arrow shows the migration of β-galactosidase (*ca.*120 kDa), and the size markers were of 96, 69, 46, 30 and 21.5 kDa.


1 2 3 4 5 6 7 8
96
69
46
30
22

3.3. The gene RSp40

3.3.1. The open reading frame

The translation of the open reading frame A (Fig. 3.7) shown in Fig. 3.13. represents a predicted protein of 364 amino acids (molecular weight 40,000). The translational initiation codon chosen is preceded by an in-frame stop codon (see Fig. 3.7), and the termination codon is followed by a putative polyadenylation/processing signal (AATAAA). Since the open reading frame falls in the Rs region of the genome of PRV (see section 3.4), it was designated RSp40. Comparison of this amino acid sequence with those predicted for the products of HSV-1 genes using the program BestFit (see section 2.17) showed that the predicted product of the gene RSp40 was related to that of the gene US1 (McGeoch *et al.*, 1985). A Dot-matrix comparison using the program DotPlot (see section 2.17) is presented in Fig. 3.14 to show the overall similarity between the predicted products of the gene RSp40 and these of genes 63 of VZV (Davison & Scott, 1986), ORF4 in the Us region of equine herpes virus 4 (EHV-4) (Cullinane *et al.*, 1988), and US1 of HSV-1 (McGeoch *et al.*, 1985). It should be mentioned that the next Met codon (amino acid 125), possible to chose as the translational initiation codon of the gene, falls within the region of similarity (Fig. 3.14). The programs CodonPreference and Testcode (see section 2.17) were again used to analyse the reading frame chosen (Fig. 3.15). It can be seen in Fig. 3.15 that the translational initiation and termination codons are consistent with the results of the computer analysis. However, two internal sections in RSp40 have poor protein-coding potential, although the analysis by TestCode did not exclude protein-encoding possibility in these regions. These sections correspond to an unusually long acidic domain near the C-terminus of the predicted product of RSp40, encoded by a region extraordinarily rich in G and A, and will be considered in section 4.2.

Fig. 3.13. Nucleotide sequence of the region between nucleotides 1 and 1807 of the *Bam*HI 10 fragment of PRV DNA. The conceptual translation of open reading frame A is shown using the one-letter code. The putative polyadenylation/processing site of the RSp40 gene is indicated by underlining, a possible CCAAT box by a solid overline and possible Sp1 binding sites by broken overlines.

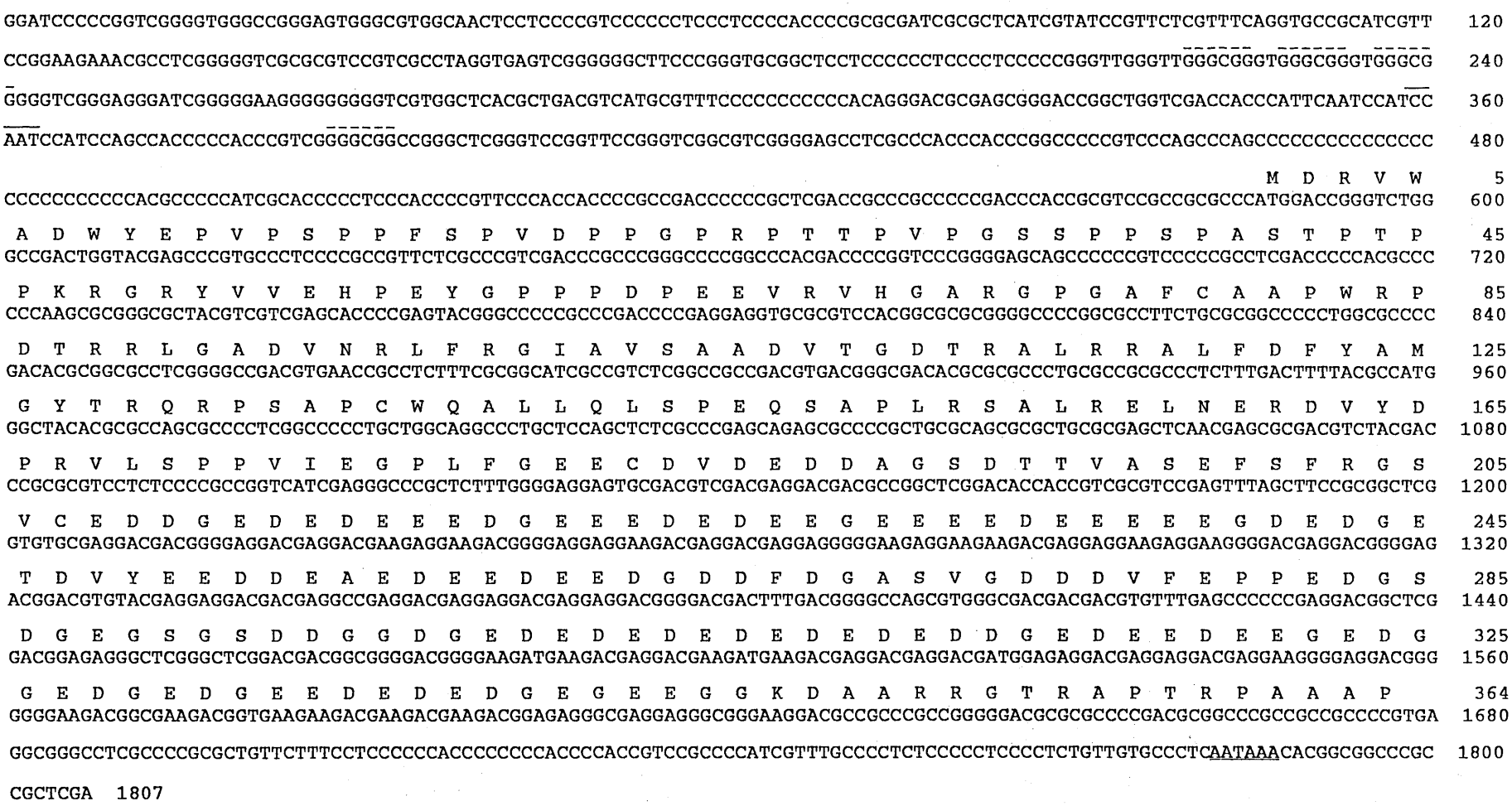
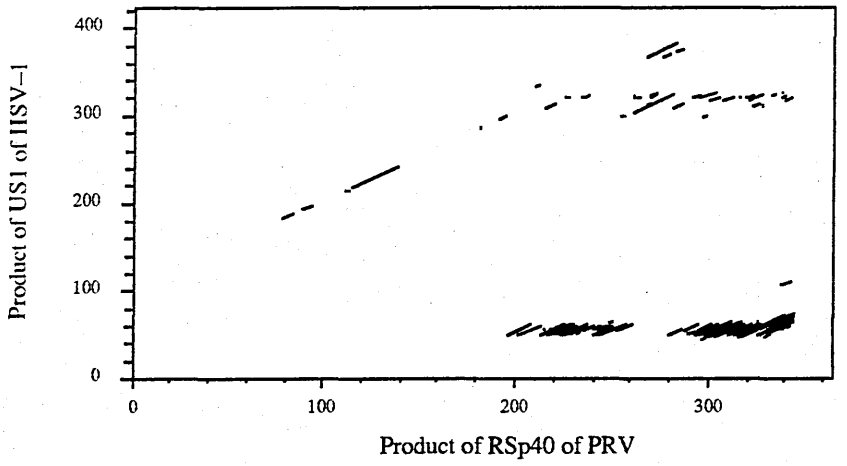
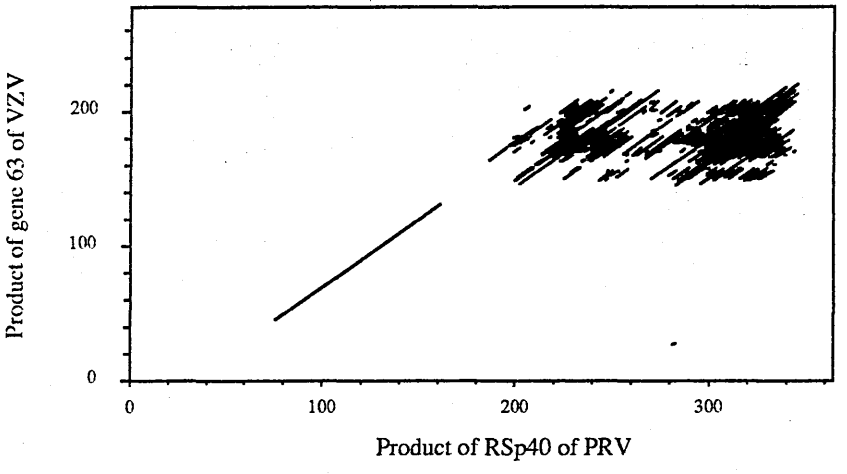


Fig. 3.14. Dot-Matrix comparison of the amino acid sequence encoded by the open reading frame A with that of US1 of HSV-1 and homologues in other alphaherpesviruses. The comparison was made using Compare, and aligned using DotPlot (see section 2.17). The predicted amino acid sequence encoded by the open reading frame A of the *Bam*HI 10 fragment (see Fig. 3.13) is presented on the ordinate, and those of US1 of HSV-1 (a), gene 63 of VZV (b) or ORF4 of EHV-4 on the abscissa. The amino acid sequence of the product of US1 of HSV-1 is from McGeoch *et al.*(1985), those of gene 63 of VZV and ORF4 of EHV-4 are from Davison and Scott (1986) and Cullinane *et al.* (1988).

a.


Product of US1 of HSV-1
400
300
200
100
0
0
100
200
300
Product of RSp40 of PRV

b.


Product of gene 63 of VZV
200
100
0
0
100
200
300
Product of RSp40 of PRV

c.

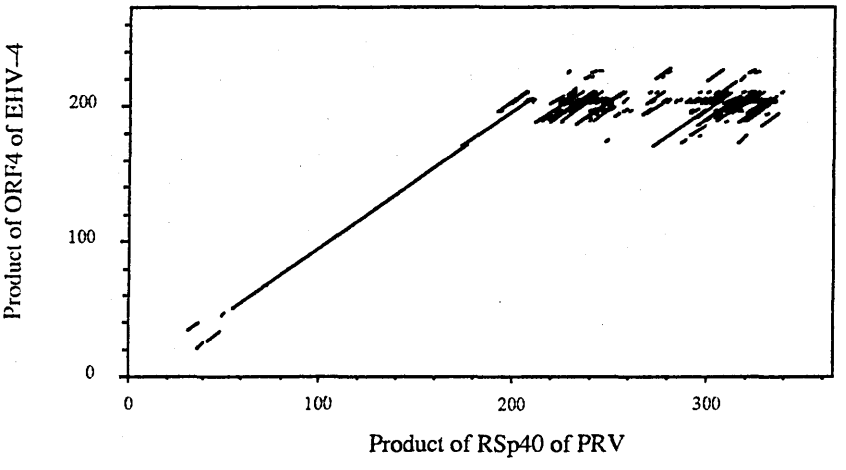
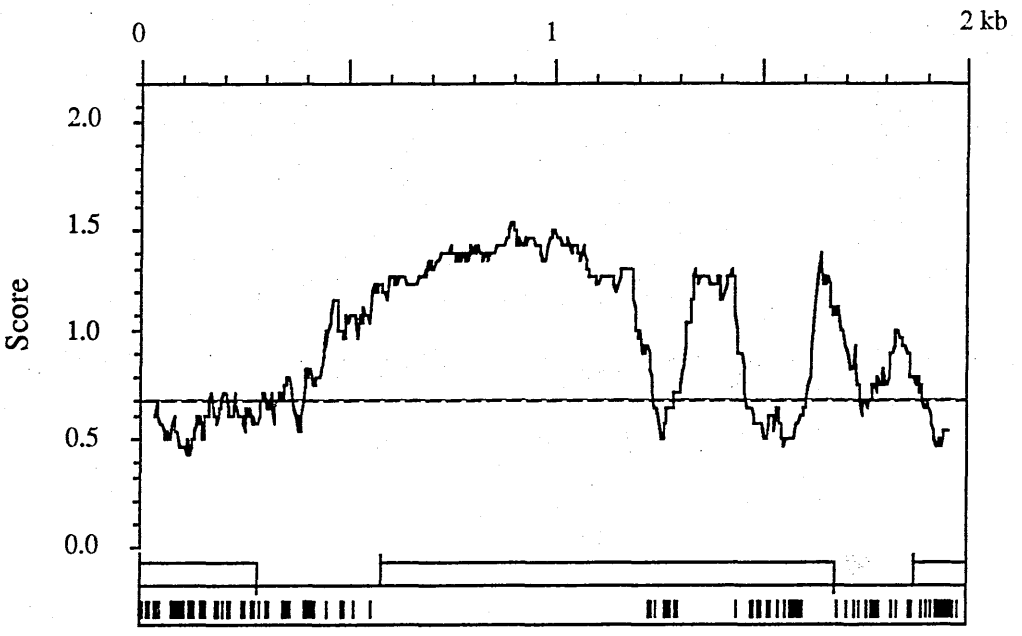
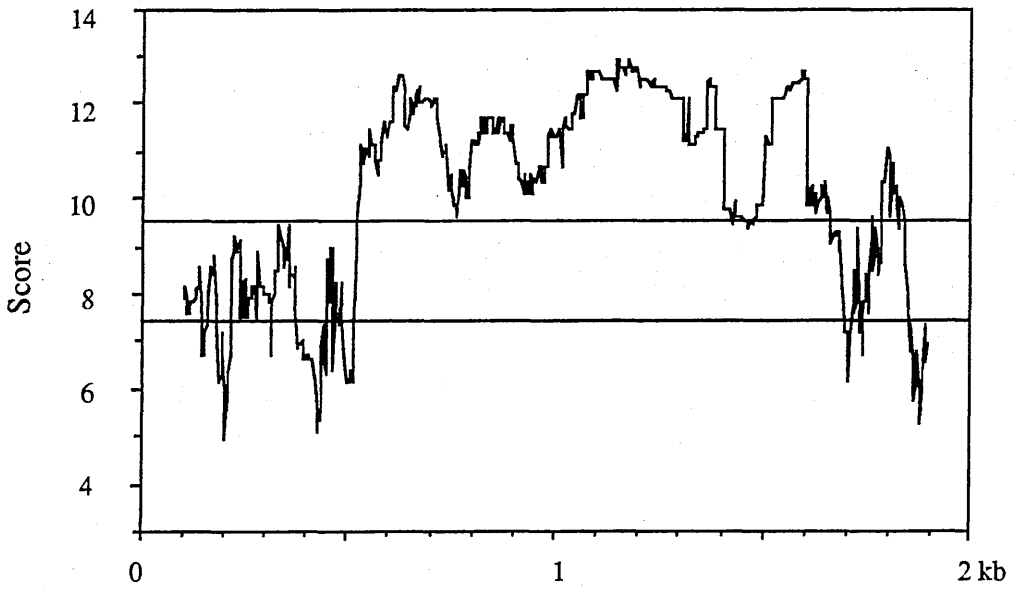

Product of ORF4 of EHV-4
200
100
0
0
100
200
300
Product of RSp40 of PRV

Fig. 3.15. Protein-encoding potential of the open reading frame A (from nucleotide 1 to 1990). (a) Analysis by CodonPreference. The GCG program, CodonPreference (see section 2.17), was used for analysing the protein-encoding potential of this open reading frame. The regions where the curve rises above the broken line are predicted to have protein-encoding potential, and the grey bar shows the predicted coding region. (b) Analysis by TestCode. The open reading frame A was also analyzed using the GCG program TestCode (see section 2.17). The portions of curve in the upper window are predicted to lie in protein-encoding regions, whereas the portions in the lower window are predicted to lie in non-encoding regions. The middle window is an area for which the program makes no prediction.

a.



b.



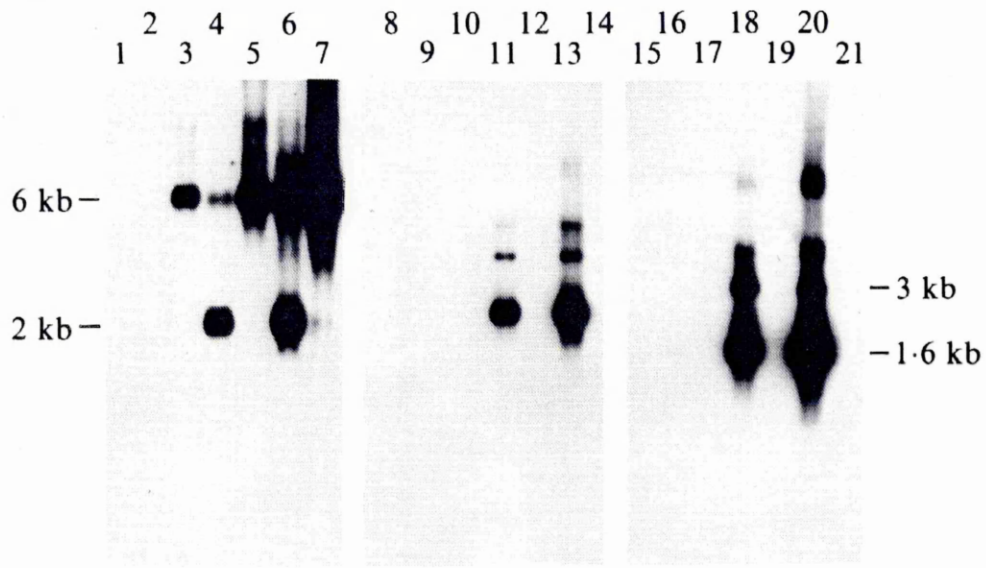
Upstream of the gene RSp40, it is difficult to identify a putative TATA box (see Fig. 3.13), although not all genes possess such a signal. In this region, there is a potential CAT box and several GC boxes (GGGCGG) which may possibly control the transcript of this gene, although the sequence GGGCGG occurs frequently in the GC-rich PRV DNA.

3.3.2. Characterization of the kinetic class of RSp40 mRNA

As mentioned in the Introduction, US1 (McGeoch *et al.*, 1985) is one of the immediate-early genes of HSV-1 (Honess & Roizman, 1974). However, in PRV only a single immediate-early gene has been identified (Ihara *et al.*, 1983): IE180, the homologue of the major trans-activating protein of HSV-1 (Watson & Clements, 1980), known variously as α4, $V_{mw}175$ and IE3. As the predicted amino acid sequence of RSp40 is homologous to that of the product of US1 of HSV-1, it was of interest to determine whether the mRNA of RSp40 in PRV was of the same kinetic class as that of US1 in HSV-1.

As explained in the Introduction (see section 1.2.1), immediate-early mRNA, rather than those of the other kinetic classes, can be synthesized in the presence of cycloheximide. Under these conditions, the amount of transcription is increased, presumably because there is no down-regulation by the products of delayed-early genes. In order to determine if the transcript of RSp40 was an immediate-early mRNA, cycloheximide was added 1 h before infection with PRV. Polyadenylated mRNA was isolated from the cells infected for 1 h, 3 h, and 6 h in the presence or absence of cycloheximide, separated on a denaturing agarose gel, and transferred onto nitrocellulose membranes. These membranes were separately hybridized with the different probes, probe *s*, a *SalI* fragment (approx. nucleotides 645-1140) containing a region of the RSp40 gene (see Fig. 3.3.c); probe *b*, from the PRV gX gene (thought not to be an immediate-early gene) as used before

Fig. 3.16. Effect of cycloheximide on the transcription of gene RSp40 of PRV. Polyadenylated mRNA was isolated from BHK cells infected for 0 h (lanes 1, 8 and 15), 1 h (lanes 2, 3, 9, 10, 16 and 17), 3 h (lanes 4, 5, 11, 12, 18 and 19) or 6 h (lanes 6, 7, 13, 14, 20 and 21) with PRV, separated by electrophoresis in 1% agarose under denaturing condition (see section 2.14.2.a), and transferred to nitrocellulose membranes. The membranes were hybridized with the radioactively labelled probes (see text) from the IE180 gene, *i* (lanes 1 to 7), from gene RSp40, *s* (lanes 8 to 14), and from the gene gX, *b* (lanes 15 to 21). Cycloheximide was present in the cells from which the mRNA in lanes 3, 5, 7, 10, 12, 14, 17, 19 and 21 was isolated. The figure shows an autoradiograph, and the sizes of some bands presented were deduced by comparing with λ DNA digested with *Hin*dIII, and pUC18 DNA digested with *Eco*RI and with *Rsa*I (see section 2.7). The size of the major band in lanes 11 and 13 (not indicated) was 1.9 kb.


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
6 kb
2 kb
3 kb
1·6 kb

(see section 3.2.2), serving as control; and probe *i*, the PRV *Bam*HI 8 fragment containing the gene for IE180 (Ihara *et al.*, 1983; Cheung, 1989a). From the result presented in Fig. 3.16, it can be seen that the 6 kb mRNA corresponding to IE180 is synthesized in infected cells in the presence of cycloheximide; (indeed, the amount of this mRNA was increased under these conditions) whereas the transcripts hybridized to the probe from gene gX were not present under these conditions. Thus, as anticipated, gX behaves as an early gene. The mRNA for RSp40 contained an abundant major species of 1.9 kb, together with some minor species of higher molecular weight, but none of these were detectable in the presence of cycloheximide. These results suggest that RSp40 is not an immediate-early gene.

In Fig. 3.16, the *Bam*HI 8 probe also hybridized to a second, 2 kb, transcript of a different kinetic class to IE180, which, incidentally, serves as an internal control. This would correspond to the open reading frame which overlaps the IE180 gene, and lies on the opposite strand (Cheung, 1989b).

3.4. The Us/Rs junction

The restriction map of PRV DNA indicated that the *Bam*HI 10 and the *Bam*HI 12 fragments contain the junctions of the Rs and Us regions of the genome (Rixon & Ben-Porat, 1979). As the work described in section 3.1-3.3 had established the sequence of the *Bam*HI 10 fragment, it was decided to identify the precise position of the junction in the PRV genome by comparison of this nucleotide sequences with that of the *Bam*HI 12 fragment.

Initially, these two fragments were compared by double digestion of the pPRVB10 and the pPRVB12 DNA with *Sal*I and *Bam*HI, followed by electrophoresis, transfer onto nitrocellulose, and hybridization to radioactively labelled DNA probes corresponding to different regions in the *Bam*HI 10 fragment. Probes *a* and *s* were those used before (see sections 3.2.1 and 3.3.2), and r_1 and r_2 were a *Taq*I-*Nla*VI and an *Nla*VI fragment (approx. nucleotides 1805-2303 and 2303-2553) containing the R_1 and the R_2 regions (see Fig. 3.3.c and Fig. 3.4). The results show that only probe *a* did not hybridize to the *Bam*HI 12 fragment (Fig. 3.17), and demonstrate that the *Bam*HI 12 fragment must therefore contain regions corresponding to the gene RSp40 and the two reiterated sequences.

Part of the nucleotide sequence of the *Bam*HI 12 fragment was determined using the restriction sites identified by restriction analysis and partial sequencing. The 'compressions' and 'false stops' mentioned in section 3.1.2 occurred frequently in this region. 7-deaza-2'-dGTP was used to resolve the 'compressions', and in some cases, the *Taq* polymerase substituted for the Klenow fragment of *E. coli* polymerase I for clearing up the 'false stops'. In fact, the *Taq* polymerase was not able to eliminate all the 'false stops', and sequencing from the opposite direction was the only way to resolve them. The sequencing strategy is presented in Fig. 3.18, and

Fig. 3.17. Comparison of the *Bam*HI 10 and the *Bam*HI 12 fragments of PRV DNA. DNA from the cloned *Bam*HI 12 fragment (lanes 1, 3, 5, 7 and 9) or the *Bam*HI 10 fragment (clone 6.1) (lanes 2, 4, 6, 8 and 10) was digested with *Bam*HI and *Sal*I, subjected to electrophoresis in 0.7% agarose and transferred to nitrocellulose membranes. The membranes were hybridized, as described in section 2.14.1, with the radioactively labelled probes *s*, r_1, r_2 and *a* (see Fig. 3.3.c): Lanes 1 and 2 with *s*, lanes 3 and 4 with r_1, lanes 5 and 6 with r_2 and lanes 7 and 8 with *a*. The hybridized membranes were used for autoradiography. Lanes 1 to 8 show an autoradiograph, and lanes 9 and 10 are from an agarose gel stained with ethidium bromide. The sizes of the DNA fragments (kb) were determined from the DNA fragments of pUC18 digested by *Eco*RI and by *Rsa*I (see section 2.7), and those from 0.300 to 0.490 kb were estimated form a separate 2% agarose gel with size markers generated by digestion of pBR322 DNA with *Taq*I and *Hin*fI.

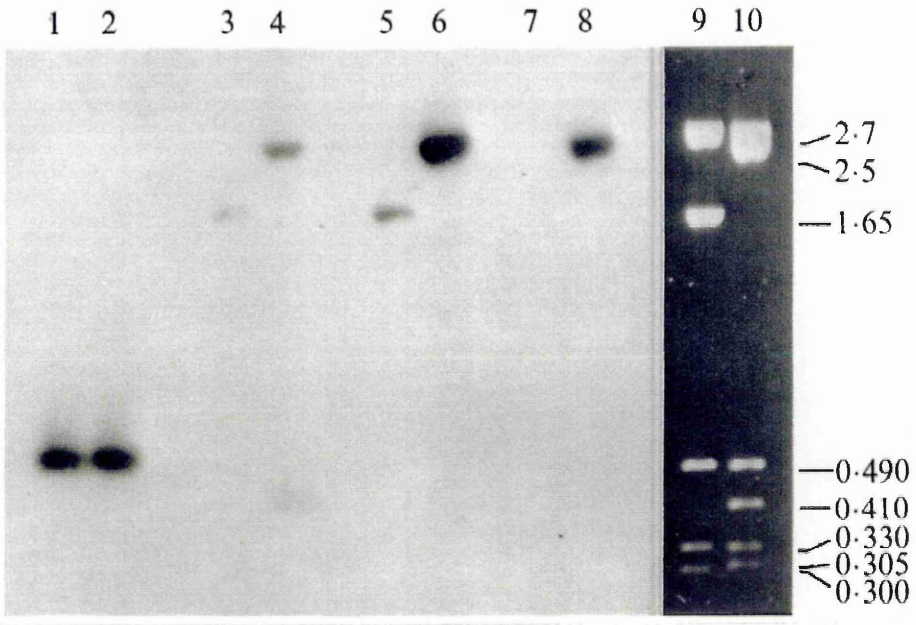
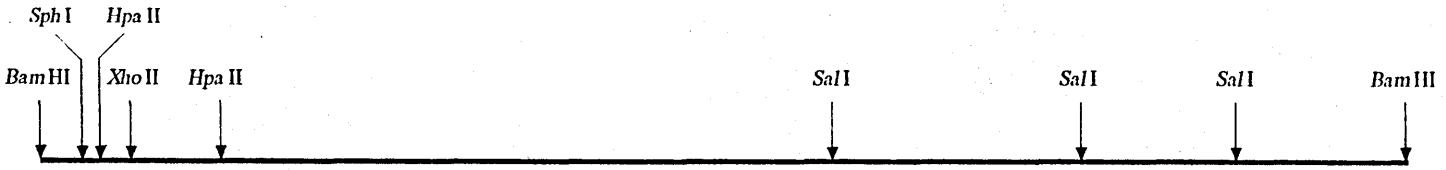

1 2 3 4 5 6 7 8 9 10
2·7
2·5
1·65
0·490
0·410
0·330
0·305
0·300

Fig. 3.18. Strategy for partial sequencing the of the *Bam*HI 12 fragment of PRV. (a) Restriction map of the *Bam*HI 12 fragment, showing restriction endonuclease recognition sites used for subcloning and sequencing of *Bam*HI 12 fragment, or for identifying regions of similarity in the *Bam*HI 10 and the *Bam*HI 12 fragment by Southern blotting. (b) Strategy for determination of the nucleotide sequence of the part of the *Bam*HI 12 fragment. The horizontal arrows represent the regions and the directions sequenced in individual clones, and the dashed lines show the R_2 region in the *Bam*HI 12 fragment, where it was possible to count the number of repeat units. The nucleotide sequences in each clone were determined in at least two independent experiments.

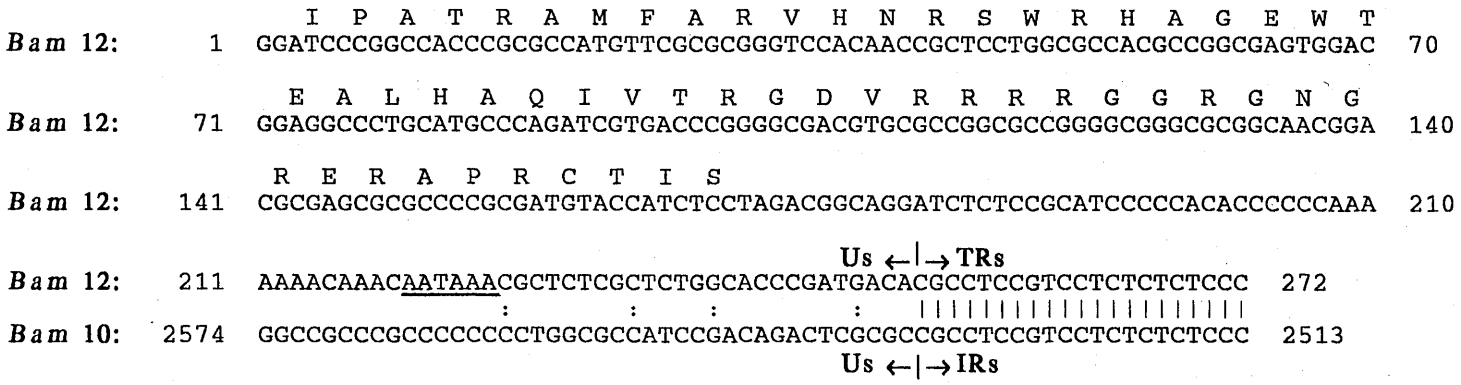
a.



b.

500 bp

Fig. 3.19. Nucleotide sequence of the region of the *Bam*HI 12 fragment of PRV DNA at the Us/TRs junction. The first 272 nucleotides of the *Bam*HI 12 fragment from the end in the Us region are shown, and the comparison with the nucleotide sequence between nucleotides 2513 and 2574 of the *Bam*HI 10 fragment is also presented. The short vertical lines show the identities thought to be indicative of the Rs sequences in the two fragments, whereas colons indicate what are thought to be random identities in the regions compared. The Us/Rs junctions are marked, and amino acid residues of the C-terminal portion of an open reading frame are shown, followed by a putative polyadenylation/processing site underlined.


I P A T R A M F A R V H N R S W R H A G E W T
Bam 12: 1 GGATCCCGGCCACCCGCGCCATGTTCGCGCGGGTCCACAACCGCTCCTGGCGCCACGCCGGCGAGTGGAC 70
E A L H A Q I V T R G D V R R R R G G R G N G
Bam 12: 71 GGAGGCCCTGCATGCCCAGATCGTGACCCGGGGCGACGTGCGCCGGCGCCGGGGCGGGCGCGGCAACGGA 140
R E R A P R C T I S
Bam 12: 141 CGCGAGCGCGCCCCGCGATGTACCATCTCCTAGACGGCAGGATCTCTCCGCATCCCCCACACCCCCCAAA 210
Us ←|→ TRs
Bam 12: 211 AAAACAAACAATAAACGCTCTCGCTCTGGCACCCGATGACACGCCTCCGTCCTCTCTCTCCC 272
Bam 10: 2574 GGCCGCCCGCCCCCCCCTGGCGCCATCCGACAGACTCGCGCCGCCTCCGTCCTCTCTCTCCC 2513
Us ←|→ IRs

the nucleotide sequence of the first 272 bp on the 'left hand' side of the *Bam*HI 12 fragment (sequenced on both strands) in Fig. 3.19. It was found that nucleotides 1-251 in the *Bam*HI 12 fragment were unique to this fragment, but nucleotides 252-272 corresponded to the complement of nucleotide 2513-2533 of the *Bam*HI 10 fragment (Fig. 3.4). Thus, the Us/Rs junctions are located at positions 251/252 in the *Bam*HI 12 fragment, and 2533/2534 in the *Bam*HI 10 fragment.

It should be mentioned that some differences were observed between the nucleotide sequences of the *Bam*HI 10 and the *Bam*HI 12 fragments in the Rs regions. Although the sequence determined from the *Sal*I site to the Us/TRs junction (Fig. 3.18) shows no difference from the complement of corresponding region of the *Bam*HI 10 fragment, ten, rather than nine, copies of the repeat unit of R_2 of the *Bam*HI 12 fragment could be counted on the sequencing gels performed from both of the *Xho*II and *Sph*I sites to the junction. Another difference can be seen in Fig. 3.17: the smallest *Sal*I fragment in the cloned *Bam*HI 12 fragment was approximately 5 nt shorter than the corresponding region in the *Bam*HI 10 fragment. Since the *Sal*I fragment of the *Bam*HI 12 was not sequenced, the exact reason for the difference is unknown. These features of the PRV DNA will be discussed in section 4.3.

From the nucleotide sequence shown in Fig. 3.19, it can also be seen that a potential polyadenylation/processing site is located approximately 30 bp 5' of the Us/TRs junction in *Bam*HI 12 fragment, and 5' of this, the C-terminal portion of a possible open reading frame is indicated. This will be discussed in section 4.3.

CHAPTER 4
DISCUSSION

4.1. The gene *pk*

The initial objective of this work was to identify the PRV protein kinase gene analogous to US3 of HSV-1. This has clearly been achieved (see section 3.2.1). Furthermore, it has been shown by another worker from this laboratory that the *pk* gene of PRV encodes the protein kinase previously purified from cells infected with the virus (Purves *et al.*, 1987a). This was done using a DNA fragment containing most of the *pk* gene described in this work (approx. nucleotides 3021-4008, see Fig. 3.8) to generate a fusion protein antibody which reacts with the purified enzyme (Stevens, 1989).

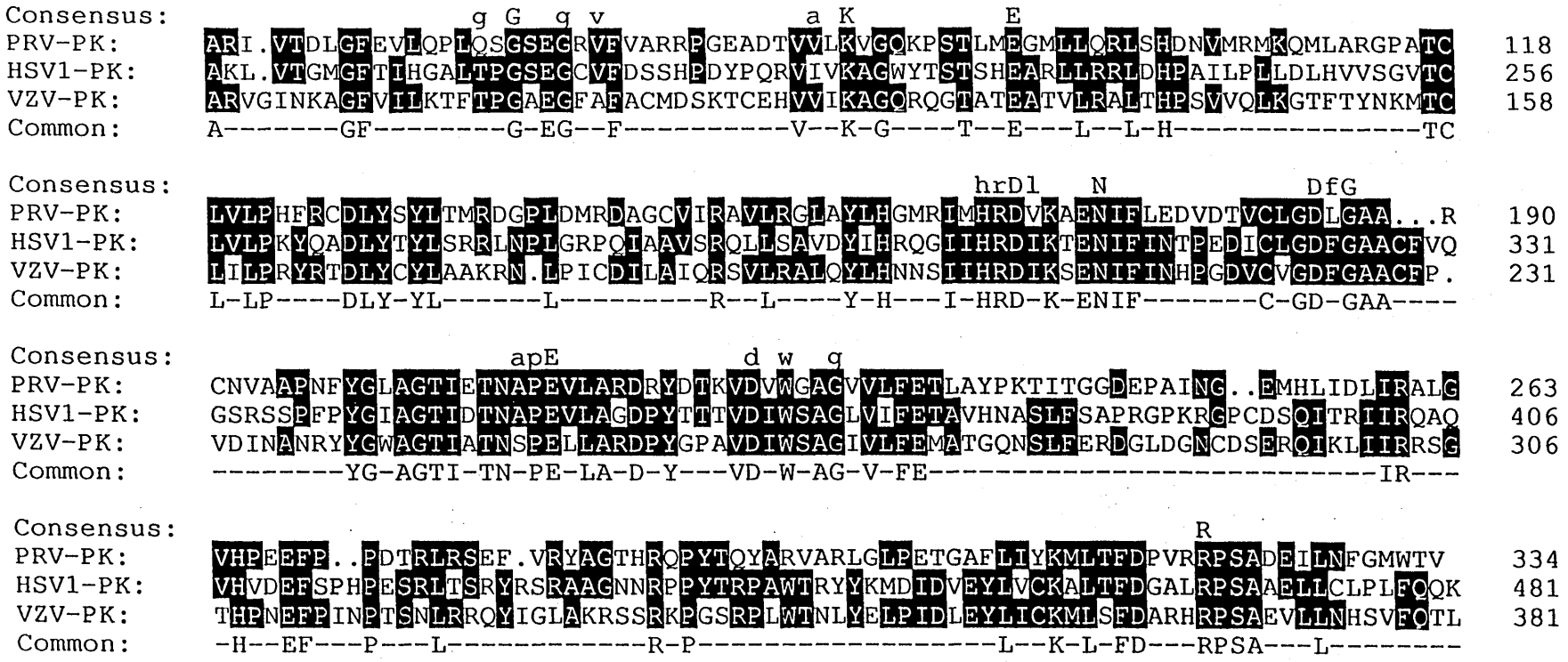
As mentioned in section 3.2, there was some difficulty in making a choice of translational initiation codon for the *pk* gene. The product of the gene (Fig. 3.8) is predicted to have a molecular weight of 37,000 and an isoelectric point of 4.7 (from analysis by the program PeptideSort, see section 2.17) which correspond well to the values obtained for the purified enzyme (Purves *et al.*, 1987a). This, therefore, is consistent with the choice of initiation codon based on the results of the computing analyses, described in section 3.2.1. While this work was in progress, another group were determining the nucleotide sequence of the protein kinase gene in the NIA-3 strain of PRV (van Zijl *et al.*, 1990), and the two sequences were published in the same issue of *Journal of General Virology* (Zhang *et al.*, 1990; van Zijl *et al.*,

1990). Van Zijl *et al.* (1990) determined the transcriptional initiation site using primer extension and found that the major transcript of the *pk* gene starts approximately 70 nt upstream of the translational initiation codon predicted for the Ka strain (corresponding to nucleotides 2856-2859 in Fig. 3.4). The translational initiation codon predicted is the first Met following this transcriptional start, consistent with Kozak's rules (see, *e.g.* Kozak, 1989). Van Zijl *et al.* (1990) also found a minor transcriptional initiation site (5% of the total transcript) at a position corresponding to nucleotide 2699 in Fig. 3.8. The northern analysis in this work (Fig. 3.11) gave no indication of two transcripts in the Ka strain, and only one translation product was detected immunologically (Stevens, 1989).

Since three distinct alphaherpesvirus protein kinase genes have been sequenced, it is possible to look for common structural features among the predicted amino acid sequence of these genes. A detailed comparison of the predicted amino acid sequences in the catalytic domain of the alphaherpesvirus protein kinases is presented in Fig. 4.1. In this domain PRV-PK shares 39% and 37% identity with HSV-PK and VZV-PK, respectively, whereas there is 44% identity between HSV-PK and VZV-PK. There are 27% positions of perfect identity among these three protein kinases. Some of these identities correspond to those motifs highly conserved in eukaryotic cellular protein kinases (Hanks *et al.*, 1988), and are indicated on the 'consensus' line in Fig. 4.1. It is clear that, as for HSV-PK and VZV-PK, PRV-PK is related to eukaryotic cellular protein kinases. Although some features of the motifs in eukaryotic cellular protein kinases are missing in PRV-PK (*e.g.* the first G in motif G-G--G, the A in motif A-K, and the F in motif DFG), similar minor deviations from the consensuses have also been observed in individual cellular protein kinases (Montell and Rubin, 1988; Saxena *et al.*, 1987; Levin *et al.*, 1987; Osmani *et al.*, 1988).

Of more particular interest are some regions highly conserved in the viral enzymes, but not found to be conserved in other protein kinases. The most striking

Fig. 4.1. Comparison of the catalytic domains of alphaherpesvirus protein kinases. The protein kinases compared are PRV-PK (see Fig. 3.8), HSV-PK (McGeoch *et al.*, 1985) and VZV-PK (Davison and Scott, 1986). The amino acid residues that are identical for at least two of the three protein kinases are presented in white on black, and those identical in all three protein kinases are shown on the line designated Common. The motifs highly conserved in eukaryotic protein kinases are shown on the line designated Consensus, and gaps are indicated by dots. Positions absolutely conserved in eukaryotic protein kinases currently known are in upper case, whereas those not completely conserved are in lower case. The last residues presented in the case of PRV-PK and HSV-PK are the C-termini, but there are 12 amino acids extending further downstream in VZV-PK, ending at 393.


Consensus:
PRV-PK:
HSV1-PK:
VZV-PK:
Common:
Consensus:
PRV-PK:
HSV1-PK:
VZV-PK:
Common:
Consensus:
PRV-PK:
HSV1-PK:
VZV-PK:
Common:
Consensus:
PRV-PK:
HSV1-PK:
VZV-PK:
Common:

examples of these (numbered for PRV-PK) are: Glu^{64} (preceding the third G of the G-G--G motif), Thr^{117} and Cys^{118}, Cys^{189} (preceding the motif DFG), and Thr^{207} and Asn^{208} (preceding the motif APE). As described in the Introduction (see section 1.3), the G-G--G motif is thought to reside at the ATP binding site of protein kinases, and the Asp in the DFG motif is also implicated in ATP binding in the cAMP-dependent protein kinase (Buechler and Taylor, 1988, 1989). The APE motif is required for activity of $pp60^{v\text{-}src}$ (Bryant and Parsons, 1984), and has also been found near the catalytic site of cAMP-dependent protein kinase (Bramson *et al.*, 1982). Since the Glu^{64}, Cys^{189} and $Thr^{207}Asn^{208}$ are very close to these motifs, they should also be near the catalytic site. As they are not conserved in eukaryotic protein kinases, they are unlikely to be involved in the reaction mechanism, and it is therefore more likely that they interact with the specific substrate of these viral protein kinases. In contrast to the residues just mentioned, the Thr^{117} and Cys^{118} lie in subdomain V of Hanks *et al.* (1988), which is not conserved in eukaryotic protein kinases, and the function of which is unclear. Nevertheless, it is the second most highly conserved region in the alphaherpesvirus protein kinases, having 42% amino acid residues in common. (Subdomain VIII is the most conserved, with 73% identity)

As the protein kinase gene has also been determined in strain NIA-3 of PRV (van Zijl *et al.*, 1990), it is of interest to compare the *pk* gene in the two different PRV strains. Table 4.1 shows the variant amino acid residues of the predicted products of *pk* in the Ka and NIA-3 strains. Following amino acid residue 27, two residues, GlyGlu, are absent from the Ka strain, but present in the NIA-3 strain. Ten other amino acid residues differ between these two strains: Ala^{35}, Val^{38}, Gln^{97}, Ser^{100}, Arg^{106}, Met^{107}, Ala^{116}, Cys^{146}, Ala^{150}, and Thr^{242} (Ka strain numbering). Eight of these amino acid residues are found in the catalytic domain of the protein kinase, but none of these is part of a conserved motif of cellular protein kinases or a

Table 4.1. Amino acid differences between the protein kinases of the Ka and the NIA-3 strains of PRV

Position*	Amino acid residues in the Ka strain	Amino acid residues in the NIA-3 strain
after 27	–	GlyGlu
35	Ala	Ser
39	Val	Ala
97	Gln	Lys
100	Ser	Ala
106	Arg	Ser
107	Met	Leu
116	Ala	Val
146	Cys	Arg
150	Ala	Ser
242	Thr	Ala

* The positions of the amino acid residues are those for the Ka strain of PRV.

region specifically conserved among the viral enzymes.

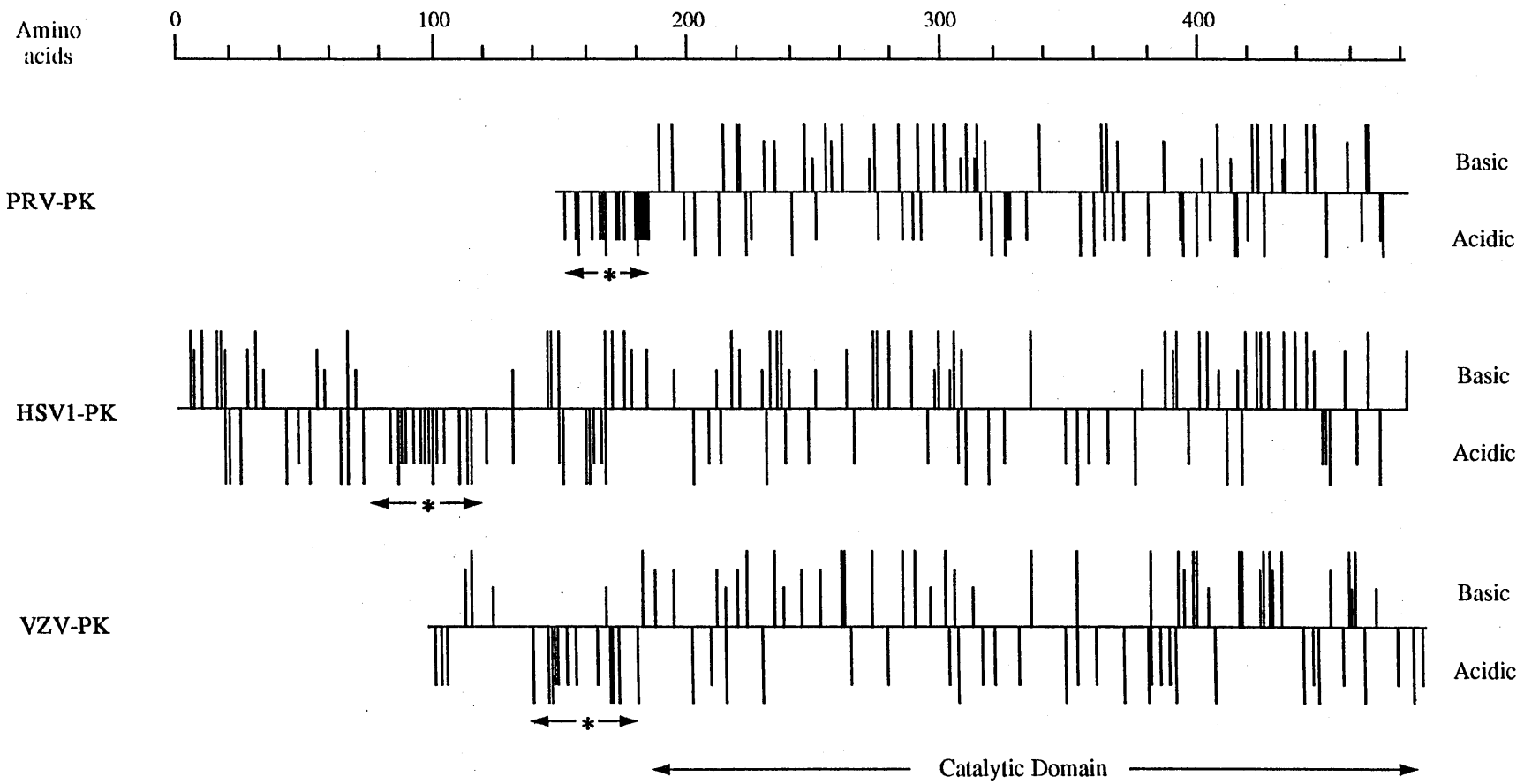
As described in the Introduction, HSV-PK is not closely related to any individual cellular protein kinase (Hanks *et al.*, 1988). Leader and Purves (1988) suggested that there might be a cellular protein kinase homologous to these viral enzymes. If this were true, this cellular enzyme might be expected to have those amino acid sequences that are conserved between the viral protein kinases, as discussed above. Comparison of the amino acid sequence of the viral enzymes with the sequences in the most recent versions of the GenBank, EMBL, NBRF, and SwissProt databases (section 2.17) shows that none of the known cellular protein kinases is closely related to PRV-PK and the other alphaherpesvirus protein kinases. PRV-PK does not share more than 27% identity with any known cellular protein kinases in its catalytic domain, and no cellular protein kinase has even 50% identity with the the region specifically common to the alphaherpesvirus protein kinases. The highest identity was obtained for the products of *Saccharomyces cerevisiae* gene CDC 28 (Lörincz and Reed, 1986), the two domains of *Xenopus* ribosomal protein S6 kinase II (Jones *et al.*, 1988), human *A-raf-1* (Beck *et al.*, 1987), and the γ-subunit of rat calmodulin kinase II (Tobimatsu *et al.*, 1988). However, these cellular protein kinases are not closely related to each other.

Recently, the nucleotide sequence of a putative protein kinase gene, B1R, in vaccinia virus (a pox virus) has been reported (Howard and Smith, 1989). The predicted amino acid sequence of the gene B1R has even less identity to the alphaherpesvirus protein kinases than do the cellular protein kinases mentioned above, and it is clear that these different viral protein kinases are not closely related to each other.

In contrast to the 37-44% identity of the catalytic domain of the alphaherpesvirus protein kinases, in the non-catalytic domain of these enzymes, not only do the amino acid sequences differ completely from each other, but also the

number of amino acid residues varies from 182 in HSV-PK, 83 in VZV-PK, to 44 in PRV-PK of the Ka strain (the NIA-3 strain has 46 amino acid residues). The only common feature that can be suggested in this region is a stretch of acidic amino acid residues (Fig. 4.2), although this is less marked in VZV-PK than in HSV-PK and PRV-PK. This acidic portion of the non-catalytic domain might be necessary for recognition of basic substrates, or alternatively act to inhibit the catalytic domain until it is activated in some way.

Fig. 4.2. Comparison of the distribution of acidic and basic amino acid residues in alphaherpesvirus protein kinases. The amino acid sequences cited in the legend to Fig. 4.1 were analysed using the program PepPlot (see section 2.17), and the catalytic domains are indicated. Indications of hydrophobic and hydrophilic residues were edited out, and individual amino acid residues are distinguished by the lengths of the lines (Glu>Asp, and Arg>Lys>His). The regions rich in acidic amino acid residues in the non-catalytic domains are indicated by asterisks.


Amino acids
0
100
200
300
400
PRV-PK
Basic
Acidic
HSV1-PK
Basic
Acidic
VZV-PK
Basic
Acidic
Catalytic Domain

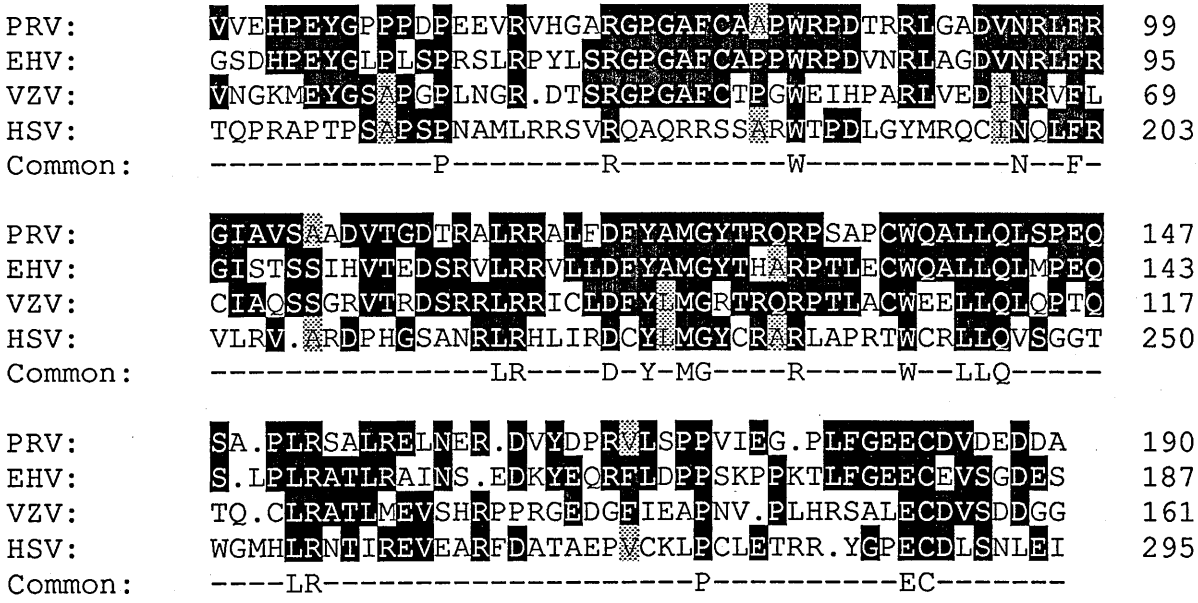
4.2. The gene RSp40

The gene RSp40 was identified in the region between nucleotides 586 and 1680 in the *Bam*HI 10 fragment of PRV DNA (see section 3.3). A striking feature of this open reading frame is the very AG-rich region already mentioned in section 3.3.1, which is predicted to encode a long and extremely acidic stretch of over 120 amino acid residues near the C-terminus of the protein. The results from computing analysis showed two sections in this region with poor protein-encoding potential (see Fig. 3.15). Thus, one must consider the possibility that a part of this AG-rich region may be spliced out of the initial transcript. Spliced genes are very rare in alphaherpesviruses: in HSV-1 only IE110, UL15, US1, and US12 are spliced (Watson *et al.*, 1981; Rixon and Clement, 1982; Costa *et al.*, 1985; Perry *et al.*, 1986). Although the mRNA of US1, the homologue of RSp40 in HSV-1, is spliced out of the initial transcript of the gene (Watson *et al.*, 1981; Rixon and Clements, 1982), the intron is in the 5' non-coding region rather than in the coding region. It is difficult to identify splice donor [$^{C}_{T}$AG ↓ GT$^{A}_{G}$AGT] and acceptor [$(^{C}_{T})_{11}N_7{}^{C}_{T}$AG ↓ G] consensus sequences (Mount, 1982) in the AT-rich region of RSp40: in particular the polypyrimidine portion of the acceptor site is lacking. The results of northern blotting analysis showed that the size of the transcript of RSp40 is about 1.9 kb (see Fig. 3.16). As the distance from the translational initiation site to the putative polyadenylation/processing signal is only 1.2 kb (see Fig. 3.13), the transcript is certainly long enough to contain the AG-rich region. Thus, it seems likely that the AG-rich region is present in the mRNA. Regions of proteins unusually rich in particular amino acid residues have also been found in other alphaherpesviruses. For example, the gene 11 of VZV encodes a highly acidic region, lying near the N-terminus of its predicted product (Davison and Scott, 1986).

A comparison between the predicted amino acid sequences of the gene RSp40 and the analogous genes of other alphaherpesviruses is presented in Fig. 4.3. It is clear that the predicted primary structure of the product of gene RSp40 of PRV shares more identities with those of the US1 homologues, gene 63 of VZV and ORF-4 of EHV-4, than with that of US1 of HSV itself. The transcript of US1 has been identified as an immediate-early mRNA (Honess and Roizman, 1974), and the product of US1 is involved in the expression of certain late genes (Sears *et al.*, 1985). However, the precise function of the gene is unclear. In some cell lines, but not others, US1 is not essential for virus growth, implying that a host cellular component might be able to subserve the function of the viral protein (Sears *et al.*, 1985; Ackerman *et al.*, 1985). It was therefore decided to perform a computing search for a possible cellular protein containing the conserved regions shown in Fig. 4.3. The program Profile (section 2.17) was used, but revealed no related cellular protein. Another program, Find (section 2.17), showed that no cellular protein even contains both of the conserved motifs D-Y-MG and W--LLQ.

The result of northern blotting analysis indicated that RSp40 does not specify an immediate-early mRNA (section 3.3.2), in contrast to US1 in HSV. Since the predicted amino acid sequence of the product of gene RSp40 is more closely related to that of gene 66 of VZV and ORF4 of EHV-4 (Fig. 3.16 and Fig. 4.3), it would be of interest to know the kinetic class of the transcripts of those genes. It is possible that the PRV, EHV-4 and VZV homologues of HSV US1 may have a similar function which differs in some respects from that of the HSV protein.

Fig. 4.3. Comparison of the conserved region in the predicted product of gene RSp40 with that of the homologues of alphaherpesviruses. The proteins compared are: the predicted product of RSp40 of PRV (see Fig. 3.13), the product of US1 of HSV-1 (McGeoch *et al.*, 1985), the product of gene 63 of VZV (Davison and Scott, 1986) and the product of ORF4 of EHV-4 (Cullinane *et al.*, 1988). Gaps introduced to optimize alignment are indicated by dots. The amino acid residues which are identical for at least two of the four sequences are presented with either a black or grey background, those identical in all four sequences are shown on the line designated Common.


PRV: VVEHPEYGPPPDPEEVRVHGARGPGAFCAAPWRPDTRRLGADVNRLFR 99
EHV: GSDHPEYGLPLSPRSLRPYLSRGPGAFCAPPWRPDVNRLAGDVNRLFR 95
VZV: VNGKMEYGSAPGPLNGR.DTSRGPGAFCTPGWEIHPARLVEDINRVFL 69
HSV: TQPRAPTPSAPSPNAMLRRSVRQAQRRSSARWTPDLGYMRQCINQLFR 203
Common: ----------P--------R---------W----------N--F-
PRV: GIAVSAADVTGDTRALRRALFDFYAMGYTRQRPSAPCWQALLQLSPEQ 147
EHV: GISTSSIHVTEDSRVLRRVLLDFYAMGYTHARPTLECWQALLQLMPEQ 143
VZV: CIAQSSGRVTRDSRRLRRICLDFYIMGRTRQRPTLACWEELLQLQPTQ 117
HSV: VLRV.ARDPHGSANRLRHLIRDCYIMGYCRARLAPRTWCRLLQVSGGT 250
Common: ---------------LR----D-Y-MG----R-----W--LLQ-----
PRV: SA.PLRSALRELNER.DVYDPRVLSPPVIEG.PLFGEECDVDEDDA 190
EHV: S.LPLRATLRAINS.EDKYEQRFLDPPSKPPKTLFGEECEVSGDES 187
VZV: TQ.CLRATLMEVSHRPPRGEDGFIEAPNV.PLHRSALECDVSDDGG 161
HSV: WGMHLRNTIREVEARFDATAEPVCKLPCLETRR.YGPECDLSNLEI 295
Common: ----LR--------------------P----------EC-------

4.3. The genetic organization of the Us region of PRV

In this work two genes in the *Bam*HI 10 fragment have been defined, *pk* and RSp40, homologous to US3 and US1 of HSV-1. In a study of the *Bam*HI 7 and the *Bam*HI 12 fragments, van Zijl *et al.* (1990) identified another gene, 28K, of which the 3' terminal portion corresponds to the potential open reading frame indicated in section 3.4 (see also Fig. 3.19). 28K is homologous to US2 of HSV-1 (van Zijl *et al.*, 1990). These results, together with previous studies on the *Bam*HI 7 fragment (Rea *et al.*, 1985; Petrovskis *et al.*, 1986a, 1986b; Petrovskis and Post 1987), have established the complete genomic organization of the Us region of PRV (see Fig. 4.5). The length of this region is approximately 8.7 kb, although an exact figure cannot be stated as the data are from three different strains of PRV; Ka, NIA-3, and Rice, which have been shown to have differences in their restriction maps (Paul *et al.*, 1982; Gielkens and Berns, 1982; Ludwig *et al.*, 1982).

Table 4.2 lists the genes located in the S segment of PRV, HSV-1, and VZV. Of the thirteen genes of HSV-1, six genes have counterparts in both PRV and VZV: IE175, US1, US3, US7, US8, and US9 (Rea *et al.*, 1985; Petrovskis *et al.*, 1986a, 1986b; Petrovskis and Post 1987; Cheung, 1989a; van Zijl *et al.*, 1990; Davison and McGeoch, 1986). Four genes — US5, US10, US11, and US12 — lack counterparts in PRV; and six — US2, US4, US5, US6, US11, and US12 — lack counterparts in VZV. The only gene of HSV-1 having a counterpart in VZV, but apparently not in PRV, is US10, which also has a counterpart in EHV-4 (Cullinane *et al.*, 1988). How certain is it that US10 has no counterpart in PRV? No US10 homologue was found in a careful analysis of the *Bam*HI 10 fragment (Fig. 4.4, which also shows that there is no additional open reading frame predicted to encode protein), nor in the *Bam*HI 12 fragment (van Zijl *et al.*, 1990). It is

Table 4.2. Homology of the genes in the S segments between three alphaherpesviruses*

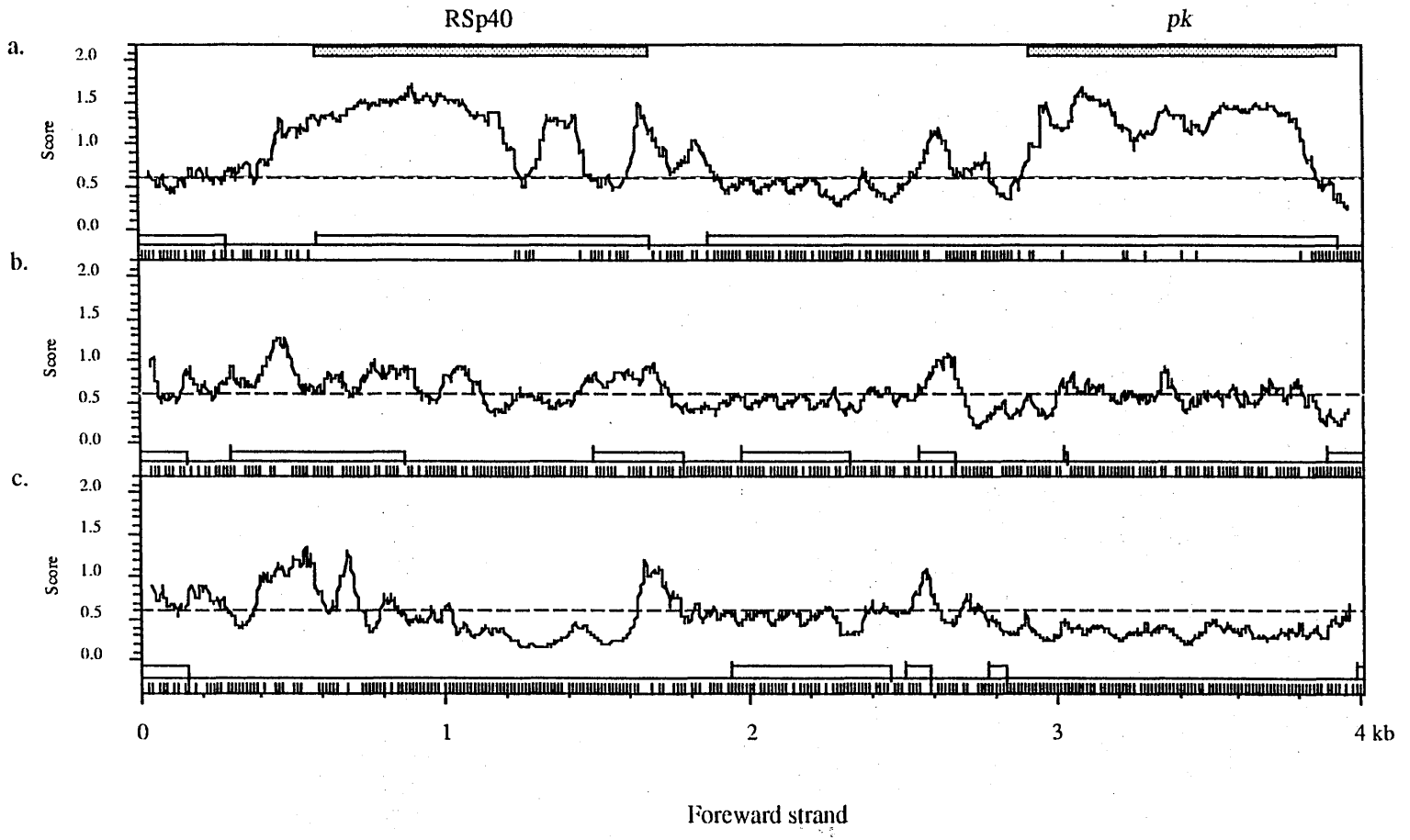
HSV–1		VZV		PRV	
Gene	Location	Gene	Location	Gene	Location
IE175	Rs	62	Rs	IE180	Rs
US1	Us	63	Rs	RSp40	Rs
US2	Us	—	—	28K	Us
US3	Us	66	Us	*pk*	Us
US4	Us	—	—	gX	Us
US5	Us	—	—	?	—
US6	Us	—	—	g50	Us
US7	Us	67	Us	g63	Us
US8	Us	68	Us	gI	Us
US9	Us	65	Us	11K	Us
US10	Us	64	Rs	?	—
US11	Us	—	—	?	—
US12	Us	—	—	?	—

– Absent from the S segment of VZV or from the Us region of PRV.

? Not detected in PRV: Possibly in unsequenced portion of the Rs region.

* For references, see text.

Fig. 4.4. Protein-encoding potential in the *Bam*HI 10 fragment of PRV. Protein-encoding potential in the six frames of the *Bam*HI 10 fragment was analysed using the program CodonPreference (see section 2.17). Open reading frames in the *Bam*HI 10 fragment shown in Fig. 3.7 are indicated as in Fig. 3.7. Analysis was started from the first (a), the second (b) and the third (c) nucleotide of the forward strand as shown in Fig. 3.4, and from the 4008th (d), the 4007th (e) and the 4006th (f) nucleotide of the reverse strand. The parts of the curve above the broken lines are the regions predicted to have protein-encoding potential. The two open reading frames identified in this work are indicated as grey bars above window a.


RSp40
pk
a.
b.
c.
Score
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4 kb
Foreward strand

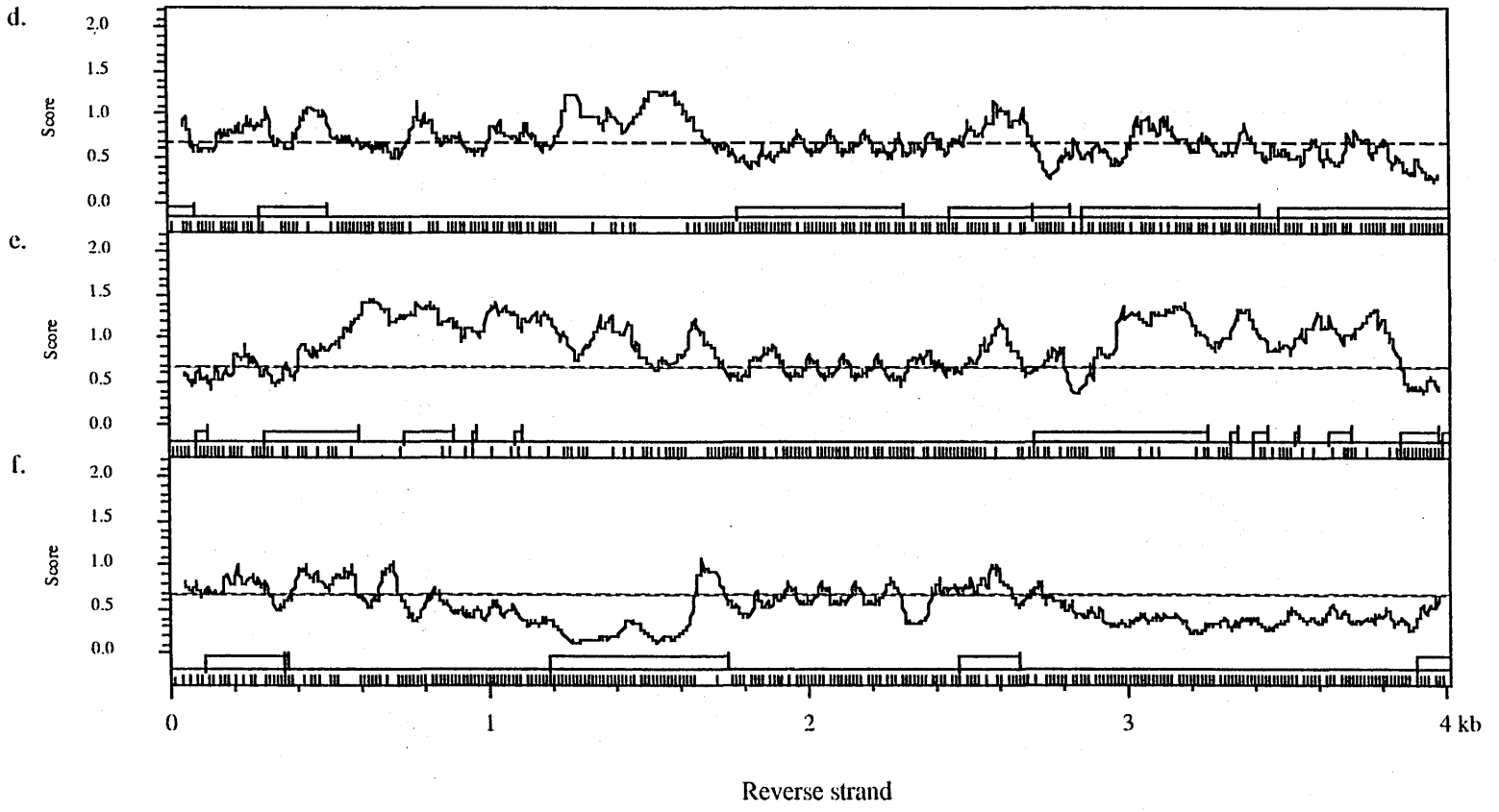
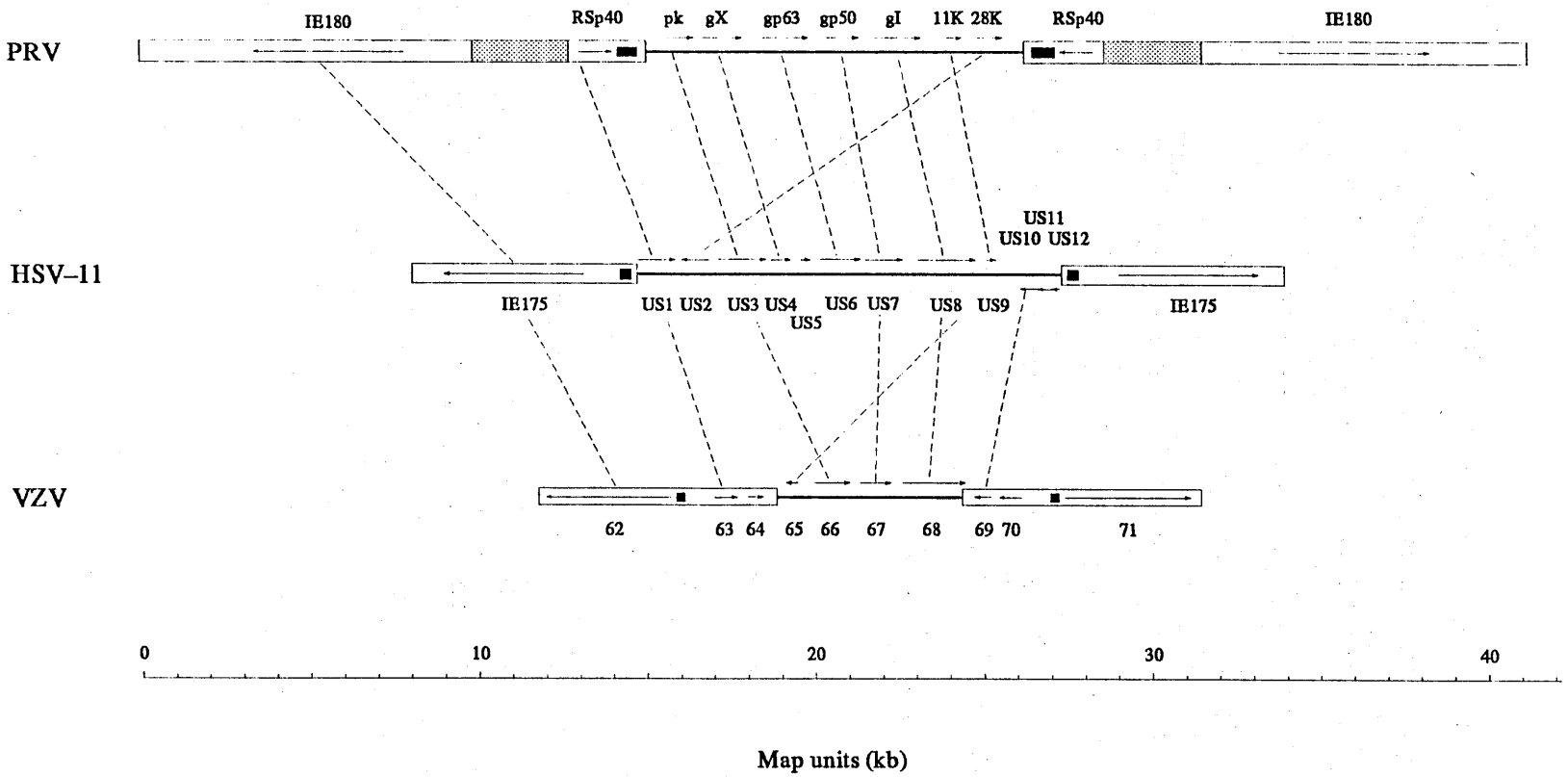

d.
e.
f.
Score
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4 kb
Reverse strand

Fig. 4.5. Comparison of the gene organization in the S segment of some alphaherpesviruses. The arrows show the coding regions and the orientation of genes, the reiterated sequences are represented as grey boxes and sequences of R_1 and R_2 of PRV are shown as a single unit. The position of the junctions are indicated by solid vertical lines, and the unsequenced region of PRV is represented by grey shading. Genes in PRV and VZV are joined to their counterparts in HSV-1 by brokern lines. The data used are after McGeoch *et al.* (1985, 1986b), Davison & Scott (1986), Rea *et al.* (1985), Petrovskis *et al.* (1986a, 1986b), Petrovskis and Post (1987), Cheung (1989a), van Zijl *et al.* (1990) and Vlcek *et al.* (1990).

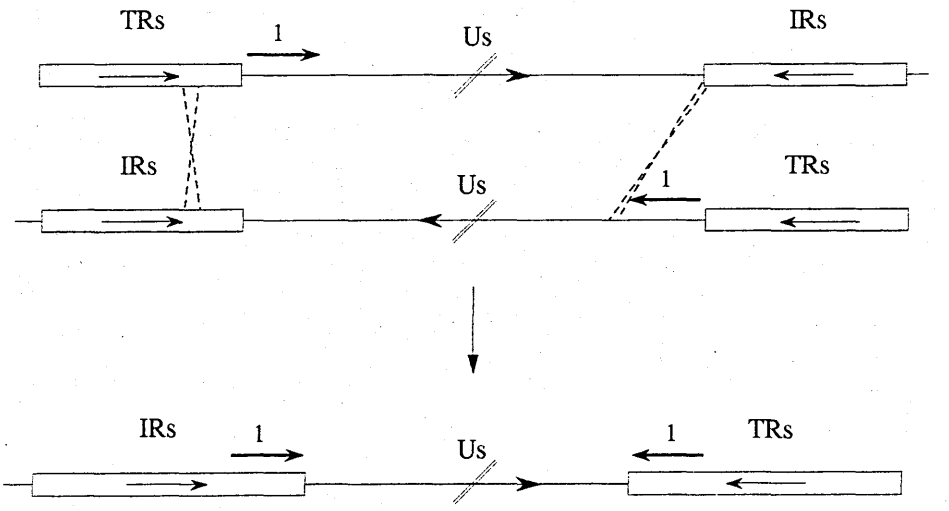

PRV
IE180
RSp40
pk
gX
gp63
gp50
gI
11K 28K
RSp40
IE180
HSV–11
US11
US10 US12
IE175
US1 US2
US3 US4
US5
US6 US7
US8
US9
IE175
VZV
62
63 64
65
66
67
68
69 70
71
0
10
20
30
40
Map units (kb)

nevertheless possible that a homologue of US10 of HSV-1 might lie in a part of the Rs region of PRV which has not yet been sequenced. Recently the nucleotide sequence of the *SalI* 1 fragment containing most of the Rs region has been determined, and no US10 homologue was identified in this (Vlcek *et al.*, 1990). There remains only *ca.* 3 kb unsequenced DNA — between the *SalI* 1 and *BamHI* 10 fragment (see Fig. 1.3) — in the Rs region, which might contain a US10 homologue.

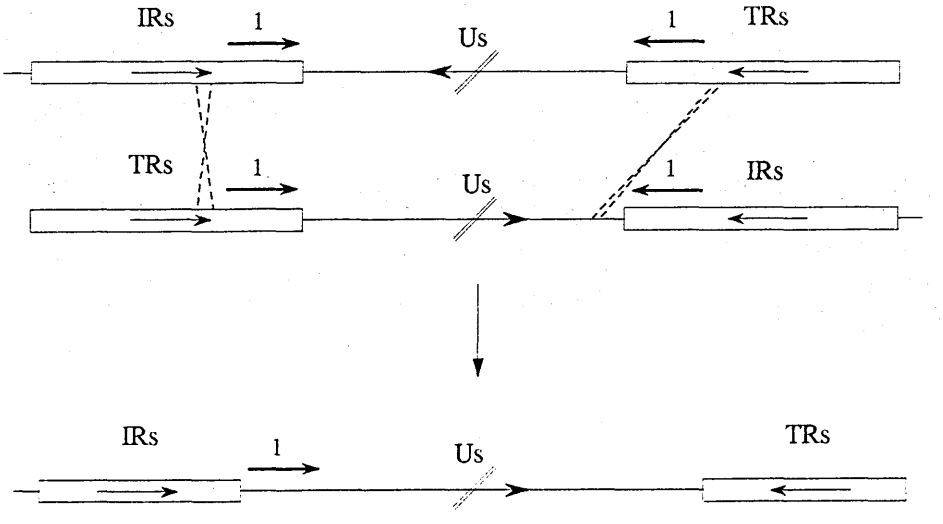
The organization of the genes in the S segment of PRV is congruent with that of HSV-1 (McGeoch *et al.*, 1985), except for the case of 28K (Fig. 4.5). 28K does not lie in the same relative position or in the same orientation to its counterpart in HSV-1, US2 (van Zijl *et al.*, 1990). Although RSp40 of PRV is in the same relative position and has the same orientation as its counterpart in HSV-1, US1, it lies in the Rs region rather than the Us region. A similar phenomenon was found in the relative gene organization of the S segments of HSV-1 and VZV (Davison and McGeoch, 1986). VZV gene 65 does not occupy the same relative position, nor have the same orientation as US9, its counterpart in HSV-1. Moreover, not only is gene 63, the counterpart of US1 of HSV-1 and RSp40 of PRV, located in the Rs region, but so is gene 64, the counterpart of US10. It has been suggested that the different gene organizations of the S segments of HSV-1 and VZV have been generated primarily by genetic recombination (Davison and McGeoch, 1986). Homologous recombination between the inverted repeats of the alphaherpesviruses is probably the major mechanism by which these retain their sequence identity, and by which mutations in one repeat are transmitted to the other (Wilkie *et al.*, 1977). However, non-homologous recombination may also occur between dissimilar sequences with limited homology (Davison and McGeoch, 1986). A possible model for non-homologous recombination is shown in Fig. 4.6 to illustrate how the different gene organization in the S segments of alphaherpesviruses might have

Fig. 4.6. A simple model to illustrate how different gene organization in the S segment of alphaherpesviruses could have arisen by non-homologous recombination. The Rs and Us regions are indicated, and the broken lines show the positions of recombination. Horizontal arrows show the orientations of the Us and the Rs regions, and the arrow labelled 1 represents a gene and indicates its orientations. Both expansion (a) and contraction (b) of the Rs regions by non-homologous recombination are presented.

a.


TRs
1
Us
IRs
IRs
Us
1
TRs
IRs
1
Us
1
TRs

b.


IRs
1
Us
1
TRs
TRs
1
Us
1
IRs
IRs
1
Us
TRs

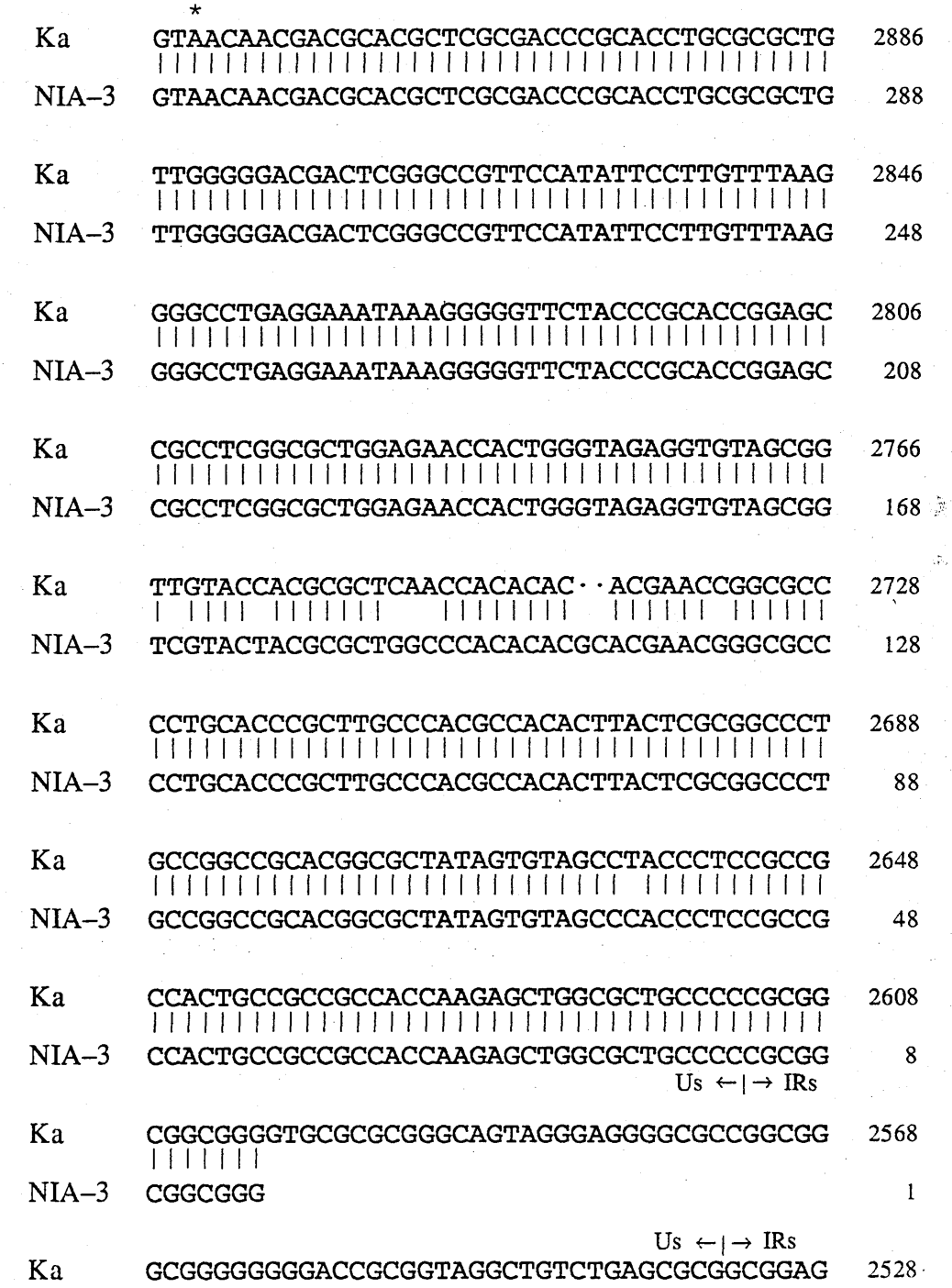
arisen. One recombination event occurs between completely homologous sequences, and the other takes place between sequences of limited homology in the Us and the Rs regions. Thus, the gene 1, which was in the Us region, becomes part of the Rs region (Fig. 4.6.a). Alternatively it is possible that gene 1 originally in the Rs region has been transferred to the Us region by a similar mechanism (Fig. 4.6.b). It is also possible that gene 1 could be transferred from one end of the Us region to the other, with a change in orientation, by mechanisms similar to those shown in Fig. 4.6. This could account for the different relative positions of 28K of PRV and US2 of HSV-1.

A striking feature of the Rs regions of alphaherpesviruses is the existence of short reiterated sequences near the Us/Rs junctions, although these sequences are different in individual viruses (McGeoch *et al.*, 1986b; Davison and Scott, 1986; Cullinane *et al.*, 1988; van Zijl *et al.*, 1990). The reiterated sequences found in the Ka strain of PRV are also distinct from those of the NIA-3 strain (van Zijl *et al.*, 1990). In the region corresponding to R_2 of the Ka strain, the NIA-3 strain has only two copies of the repeat unit, rather than the ten copies in the Ka strain; and R_1 of the Ka strain is very different from the reiterated sequence designated DR in the NIA-3 strain. Different reiterated sequences were previously observed in the two serotypes of HSV, HSV-1 and HSV-2 (McGeoch *et al.*, 1986b; Whitton and Clements, 1984). These short reiterated sequences at the end of the Rs region clearly undergo rapid changes in length and in nucleotide sequence, which may cause misalignment of DNA molecules, and hence encourage non-homologous recombination.

Evidence of non-homologous recombination of the type proposed above is found not only in the gene organization of the S segment of alphaherpesviruses (Davison and McGeoch, 1986), but also in the fact that the position of the Us/Rs junction is different in the Ka and the NIA-3 strains of PRV (this work and van Zijl *et al.*, 1990). The distance from the translational initiation site of the *pk* gene to the

Fig. 4.7. Comparison of the positions of the Us and Rs junctions in the Ka and the NIA-3 strains of PRV. (a) The nucleotide sequences between the translational initiation codons of the *pk* gene (marked with an asterisk) and the Us/IRs junctions of the Ka and the NIA-3 strains of PRV are presented in reverse. Identities between the sequences are shown with vertical bars. Gaps introduced for alignment purposes are represented by dots, and the Us/IRs junctions are also indicated. (b) The nucleotide sequences between the translational stop codons of the gene 28K (marked by •) and the Us/TRs junctions of the Ka and the NIA-3 strains of PRV are shown, and identities between the sequences are presented with vertical bars. The Us/IRs junctions are also indicated. The nucleotide sequences presented in (a) and (b) are from this work (see Fig. 3.4 and 3.19) and van Zijl *et al.* (1990).

a.


*
Ka GTAACAACGACGCACGCTCGCGACCCGCACCTGCGCGCTG 2886
NIA–3 GTAACAACGACGCACGCTCGCGACCCGCACCTGCGCGCTG 288
Ka TTGGGGGACGACTCGGGCCGTTCCATATTCCTTGTTTAAG 2846
NIA–3 TTGGGGGACGACTCGGGCCGTTCCATATTCCTTGTTTAAG 248
Ka GGGCCTGAGGAAATAAAGGGGGTTCTACCCGCACCGGAGC 2806
NIA–3 GGGCCTGAGGAAATAAAGGGGGTTCTACCCGCACCGGAGC 208
Ka CGCCTCGGCGCTGGAGAACCACTGGGTAGAGGTGTAGCGG 2766
NIA–3 CGCCTCGGCGCTGGAGAACCACTGGGTAGAGGTGTAGCGG 168
Ka TTGTACCACGCGCTCAACCACACAC··ACGAACCGGCGCC 2728
NIA–3 TCGTACTACGCGCTGGCCCACACACGCACGAACGGGCGCC 128
Ka CCTGCACCCGCTTGCCCACGCCACACTTACTCGCGGCCCT 2688
NIA–3 CCTGCACCCGCTTGCCCACGCCACACTTACTCGCGGCCCT 88
Ka GCCGGCCGCACGGCGCTATAGTGTAGCCTACCCTCCGCCG 2648
NIA–3 GCCGGCCGCACGGCGCTATAGTGTAGCCCACCCTCCGCCG 48
Ka CCACTGCCGCCGCCACCAAGAGCTGGCGCTGCCCCCGCGG 2608
NIA–3 CCACTGCCGCCGCCACCAAGAGCTGGCGCTGCCCCCGCGG 8
Us ←|→ IRs
Ka CGGCGGGGTGCGCGCGGGCAGTAGGGAGGGGCGCCGGCGG 2568
NIA–3 CGGCGGG 1
Us ←|→ IRs
Ka GCGGGGGGGGACCGCGGTAGGCTGTCTGAGCGCGGCGGAG 2528

b.

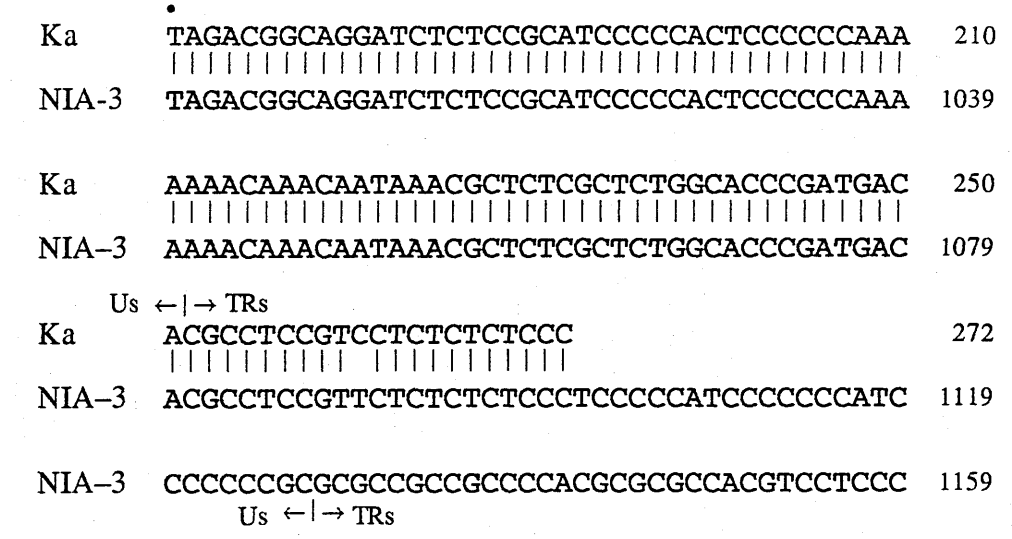
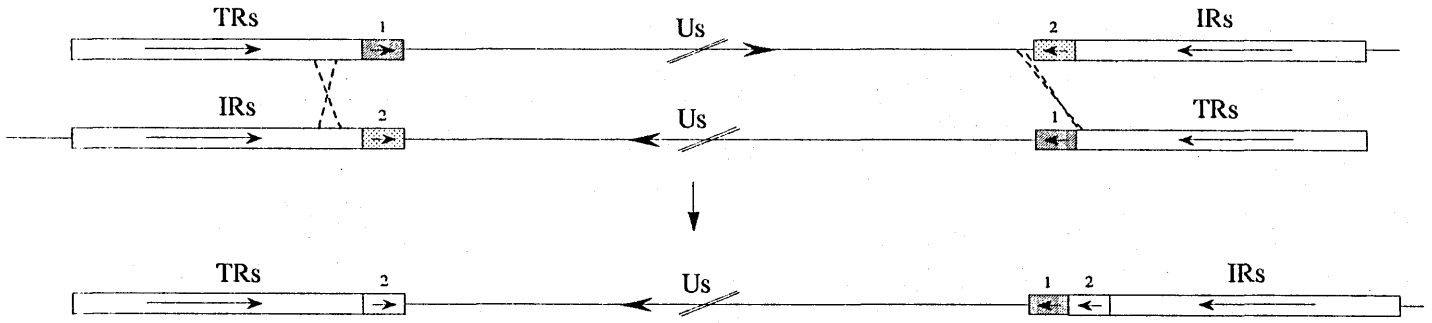
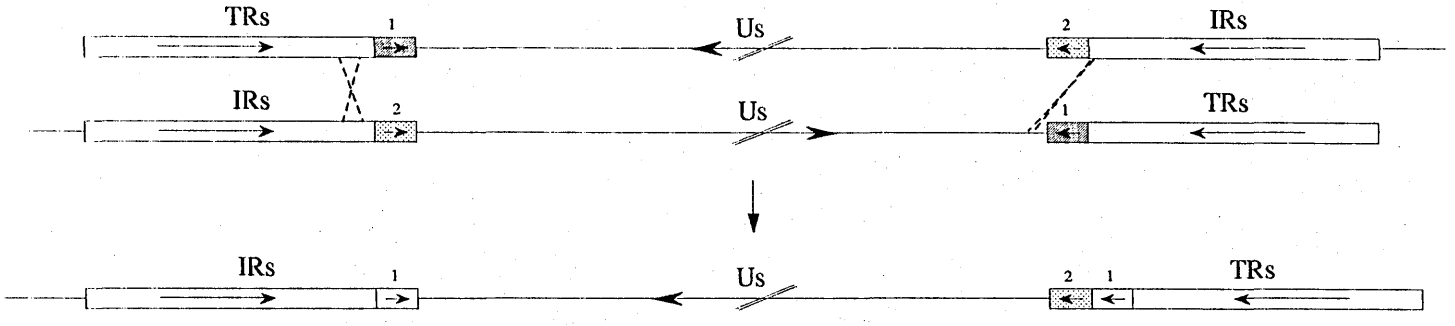

Ka TAGACGGCAGGATCTCTCCGCATCCCCCACTCCCCCCAAA 210
NIA-3 TAGACGGCAGGATCTCTCCGCATCCCCCACTCCCCCCAAA 1039
Ka AAAACAAACAATAAACGCTCTCGCTCTGGCACCCGATGAC 250
NIA–3 AAAACAAACAATAAACGCTCTCGCTCTGGCACCCGATGAC 1079
Us ←|→ TRs
Ka ACGCCTCCGTCCTCTCTCTCCC 272
NIA–3 ACGCCTCCGTTCTCTCTCTCCCTCCCCCATCCCCCCCCATC 1119
NIA–3 CCCCCCGCGCGCCGCCGCCCCACGCGCGCCACGTCCTCCC 1159
Us ←|→ TRs

Fig. 4.8. Model to account for the different positions of the Us and Rs junctions in individual strains of PRV. The Us regions are shown as long horizontal lines and grey boxes, whereas the Rs regions are presented as open boxes, with their orientations shown by horizontal arrows. The regions indicated 1 or 2 represent those flanking the Us/IRs junction of the Ka strain (*i.e.* 80 bp in the Us region and 50 bp in the IRs region). The broken lines show the positions of recombination, and two possibilities of non-homologous recombination are presented, to account for the formation of the Ka (a) and the NIA-3 (b) strains of PRV from the same ancient parent.

a.


TRs
1
Us
2
IRs
IRs
2
Us
1
TRs
TRs
2
Us
1
2
IRs

b.


TRs
1
Us
2
IRs
IRs
2
Us
1
TRs
IRs
1
Us
2
1
TRs

Us/Rs junction is about 80 bp shorter in the NIA-3 strain than in the Ka strain (Fig. 4.7.a), whereas that from the translational stop codon of the gene 28K to the Us/TRs juction is about 50 bp longer (Fig. 4.7.b). The DNA absent from the Rs region in each strain is found in the respective Us region. Fig. 4.8 shows a model of how the different positions of the Us/Rs junction in the Ka and the NIA-3 strains of PRV could have arisen by non-homologous recombination. This assumes that both PRV strains have evolved from the same ancient parent, which contained both the DNA segments 1 and 2 flanking the Us/IRs junction of the Ka strain (*i.e.* 80 bp in the Us region and 50 bp in the IRs region) at each end of the Us region, and in the same orientation. Fig. 4.8.a shows one recombination event occurring between completely homologous sequences, and the other between the Rs region and the 5' end of segment 2 in the Us region. Fig. 4.8.b presents a similar scheme to that in Fig. 4.8.a, but with the site of recombination in the Us region at the 3' end of segment 1. Different positions for the Us/Rs junctions have also been observed in different HSV-1 strains (McGeoch *et al.*, 1986b; Umene, 1989), and in the two serotypes of HSV, HSV-1 and HSV-2 (McGeoch *et al.*, 1986; Whitton and Clements, 1984). These expansions and contractions of the Us and Rs regions between different strains of the same virus are relatively minor, but they provided some support for the mechanisms proposed to explain the more extreme differences observed between different viruses.

Additional reference:

The reference listed here is missing in the References.

Campbell, M. E. M. & Preston, C. M. (1987). DNA sequences which regulate the expression of the pseudorabies virus major immediate early gene. *Virology* **157**, 307–316.

References:

Ackerman, M., Sarmiento, M., Lerner, R. A. & Roizman, B. (1985). Application of antibody to synthetic peptide of a protein specified for characterization of the intact and truncated α22 protein specified by herpes simplex virus 1 and the R325 α22- deletion mutant. *J. Virol.* **56**, 207–215.

Adams, R. L. P., Knowler, J. T. & Leader, D. P. (1986). *The biochemistry of the nucleic acids.* Chapman and Hall, London, New York, tenth edition.

Ahler, S. & Feldman, L. T. (1987). Effects of a temperature–sensitive mutation in the immediate–early gene of pseudorabies virus on class II and class III gene transcription. *J. Virol.* **61**, 1103–1107.

Allet, B., Ktagiri, K. J., & Gesteland, R. F. (1973). Characterization of polypeptides made in vitro from bacteriophage lambda DNA. *J. Mol. Biol.* **78**, 589–600.

Amann, E. & Brosius, J. (1985). 'ATG vectors' for regulated expression of cloned gene in Escherichia coli. *Gene* **40**, 183–190.

Aviv, H. & Leder, P (1972). Purification of biologically active globin mRNA by chromatography on oligothymidylic acid–cellulose. *Proc. Nat. Acad. Sci., U.S.A.* **69**, 1408–1412.

Barker, W. C. & Dayhoff, M. O. (1982). Viral *src* gene products are related to the catalytic chain of mammalian cAMP–dependent protein kinase. *Proc. Nat. Acad. Sci., U.S.A.* **79**, 2836–2839.

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

Beck, T. W., Huleihel, M., Gunnell, M., Bonner, T. I. & Rapp, U. R. (1987). The

complete coding sequence of the human *A-raf-1* oncogene and transforming activity of a human *A-raf-1* carring retrovirus. *Nuc. Acids Res.* **15**, 595–609.

Ben–Porat, T. & Kaplan, A. S. (1985). Molecular biology of pseudorabies virus. In *Herpesviruses* vol. 3, **chapter 3**, pp. 105–173. Ed. B. Roizman. Plenum Press, New York and London.

Ben–Porat, T. & Veach, R. A. (1980). Origin of replication of the DNA of a herpesvirus (peudorabies). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 172–175.

Ben–Porat, T., Rixon, F. J. & Blankenship, M. L. (1979). Analysis of the structure of the genome of peudorabies virus. *Virology* **95**, 285–294.

Ben–Porat, T., Veach, R. A. & Latin, B. F. (1980). Replication of herpesvirus DNA. VI. Virions containing either isomer of peudorabies virus DNA are infectious. *Virology* **102**, 370–380.

Ben–Zéev, A. & Becker, Y. (1977). Requirement of host cell RNA polymerase II in the replication of herpes simplex virus in alpha–amanitin–sensitive and –resistant cell lines. *Virology* **76**, 246–253.

Ben–Zéev, A., Abulafia, R. & Bratosin, S. (1983). Herpes simplex virus and protein transport are associated with the cytoskeletal framework and the nuclear matrix in infected BSC–1 cells. *Virology* **129**, 501–507.

Bhatnager, D., Hartl, F. T., Roskoski, R., Jr., Lessor, R. A. & Leonard, N. J. (1984). Adenosine cyclic 3', 5'–monophosphate dependent protein kinase: nucleotide binding to the chemically modified catalytic subunit. *Biochem.* **23**, 4350–4357.

Bibor–Hardy, V., Suh, M., Pouchelet, M. & Simard, R. (1982). Modification of the nuclear envelope of BHK cells after infection with herpes simplex virus type 1. *J. Gen. Virol.* **63**, 81–94.

Bibor–Hardy, V., Bernard, M. & Simard, R. (1985). Nuclear matrix modification at different stages of infection by herpes simplex virus type 1. *J. Gen. Virol.* **66**, 1095–1103.

Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983). Buffer gradient gels and 35S label as an aid to rapid DNA sequence determination. *Proc. Nat. Acad. Sci., U.S.A.* **80**, 3963–3965.

Birnboim, H. C. & Doly, J. (1979). A rapid extraction procedure for screenig recombinant plasmid DNA. *Nucl. Acids Res.* **7**, 1513–1523.

Bramson, H. N., Thomas, N., Matsueda, R., Nelson, N. C., Taylor, S. S. & Kaiser, E. T. (1982). Modification of the catalytic subunit of bovine heart cAMP–dependent protein kinase with affinity labels related to peptide substrate. *J. Biol. Chem.* **257**, 10575–10581.

Brow, M. A. D. (1990). Sequening with Taq DNA polymerase. In *PCR Protocols:A Guide to Methods and Applications* **chapter 24**, pp. 189–196. Ed. M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White. Academic Press, Inc.

Bruner, M., Dalie, B., Spangler, R. & Harter, M. L. (1988). Purification and biological characterization of an adenovirus type 2 E1A protein expressed in Escherichia coli. *J. Biol. Chem.* **263**, 3984–3989.

Bryant, D. & Parsons, J. T. (1983). Site–directed point mutation in the *src* gene of rous sarcoma virus results in an inactive *src* gene product. *J. Virol.* **45**, 1211–1216.

Bryant, D. & Parsons, J. T. (1984). Amino acid alteration with a highly conserved region of the rous sarcoma virus src gene product pp60^{src} inactivate tyrosine protein kinase activity. *Mol. Cell. Biol.* **4**, 862–866.

Buechler, J. A. & Taylor, S. S. (1988). Identification of aspartate–184 as an essential residue in the catalytic subunit of cAMP–dependent protein kinase. *Biochem.* **27**, 7356–7361.

Buechler J. A. & Taylor, S. S. (1989). Dicyclohexylcarbodiimide cross–links two conserved residues, Asp–184 and Lys–72, at the active site of the catalytic

subunit of cAMP–dependent protein kinase. *Biochem.* **28**, 2065–2070.

Buechler, J. A. & Taylor, S. S. (1990). Differential labeling of the catalytic subunit of cAMP–dependent protein kinase with a water–soluble carbodiimide: Identification of carboxyl groups protected by MgATP and inhibitor peptides. *Biochem.* **29**, 1937–1943.

Buechler, J. A., Vedvick, T. A. & Taylor, S. S. (1989). Differetial labeling of the catalytic subunit of cAMP–dependent protein kinase with Acetic anhydride: Substrate–induced conformational changes. *Biochem.* **28**, 3018–3024.

Bzik, D. J., Fox, B. A., DeLuca, N. A. & Person, S. (1984). Nucleotide sequence specifying the glycoprotein gene, gB, of herpes simplex virus type 1. *Virology* **133**, 301–314.

Cai, W., Gu, B. & Person, S. (1988). Role of glycoprotein B of herpes simplex virus type 1 in viral entry and cell fusion. *J. Virol.* **62**, 2596–2604.

Calder, J. M. & Stow, N. D. (1990). Herpes simplex virus helicase–primase: the UL8 protein is not required for DNA–dependent ATPase and DNA helicase activity. *Nuc. Acids Res.* **18**, 3573–3578.

Campbell, M. E. M., Palfreyman, J. W. & Preston, C. M. (1984). Identification of herpes simplex virus DNA sequences which encode a *trans*–acting polypeptide resposible for stimulation of immediate early transcription. *J. Mol. Biol.* **180**, 1–19.

Carmichel, E. P. & Weller, S. K. (1989). Herpes simplex virus type 1 DNA synthesis requires the product of the UL8 gene: Isolation and characterization of an ICP6::*lacZ* insertion mutation. *J. Virol.* **63**, 591–599.

Cheung, A. (1989a). DNA nucleotide sequence analysis of the immediate–early gene of pseudorabies virus. *Nucl. Acids Res.* **17**, 4637–4646.

Cheung, A. (1989b). Detection of pseudorabies virus transcripts in trigeminal ganglia of latently infected swine. *J. Virol.* **63**, 2908–2913.

Chien, A., Edgar, D. B. & Trela, J. M. (1976). Deoxyribonucleic acid polymerase

from the extreme thermopile *Thermus aquaticus*. *J. Bacteriol.* **127**, 1550–1557.

Chomczynski, P. & Sacchi, N. (1987). Single–step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162**, 156–159.

Chou, J. & Roizman, R. (1989). Characterization of DNA sequence–common and sequence–sepecific protein binding to *cis*–acting sites for cleavage of the terminal *a* sequence of the herpes simplex virus 1 genome. *J. Virol.* **63**, 1059–1068.

Conley, A. J., Knipe, D. M., Jones, P. C. & Roizman, B. (1981). Molecular genetics of herpes simplex virus. **VII**. Characterization of a temperature–sensitive mutant produced by *in vitro* mutagenesis and defective in DNA synthesis and accumulation of γ polypeptide. *J. Virol.* **37**, 191–206.

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

Cook, M. L., Bastone, V. B. & Stevens, J. G. (1974). Evidence that neurons harbor latent herpes simplex virus. *Infect. Immun.* **9**, 946–951.

Costa, R. H., Cohen, G., Eisenberg, R., Long, D. & Wagner, E. K. (1984). Direct demonstration that the abundant 6–kilobase herpes simplex virus type 1 mRNA mapping between 0.23 and 0.27 map units encodes the major capsid protein VP5. *J. Virol.* **49**, 287–292.

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

Crute, J. J., Tsurumi, T., Zhu, L., Weller, S. K., Olivo, P. D., Challberg, M. D., Mocarski, E. S. & Lehman, I. R. (1989). Herpes simplex virus 1 helicase–primase: A complex of three herpes–encoded gene products. *Proc.*

Nat. Acad. Sci., U.S.A. **86**, 2186–2189.

Cullinane, A. A., Rixon, F. J. & Davison, A. J. (1988). Characterization of the genome of equine herpesvirus 1 subtype 2. *J. Gen. Virol.* **69**, 156–159.

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

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

Davison, A. J. & McGeoch, D. J. (1986). Evolutionary comparisons of the S segment in the genomes of herpes simplex virus type 1 and varicella–zoster virus. *J. Gen. Virol.* **67**, 597–611.

Davison, A. J. & Scott, J. E. (1986a). The complete DNA sequence of varicella–zoster virus. *J. Gen. Virol.* **67**, 1759–1816.

Davison, A. J. & Scott, J. E. (1986b). DNA sequence of the major capsid protein gene of herpes simplex virus type 1. *J. Gen. Virol.* **67**, 2279–2286.

DeLuca, N. A., Bzik, D. J., Bond, V. C., Person, S. & Snipes, W. (1982). Nucleotide sequences of herpes simplex virus type 1 (HSV–1) affecting virus entry, cell fusion and production of glycoprotein gB (VP7). *Virology* **122**, 411–423.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 383–395.

Dodson, M. S., Curte, J. J., Bruckner, R. C. & Lehman, I. R. (1989). Overexpression and assembly of the herpes simplex virus type 1 helicase–primase in insect cells. *J. Biol. Chem.* **264**, 20835–20838.

Draper, K. G., Costa, R. H., Lee, G. T. Y., Spear, P. G. & Wagner, E. K. (1984). Molecular basis of the glycoprotein–C–negative phenotype of herpes simplex

virus type 1 macroplaque strain. *J. Virol.* **51**, 578–585.

Draper, K. G., Devi–Rao, G., Costa, R. H., Blair, E. D., Thompson, R. L. & Wagner, E. K. (1986). Characterization of the genes encoding herpes simplex virus type 1 and type 2 alkaline exonucleases and overlapping proteins. *J. Virol.* **57**, 1023–1036.

Draper, K. G., Frink, R. J. & Wagner, E. K. (1982). Detailed characterization of an apparently unspliced β herpes simplex virus type 1 gene mapping in the interior of another. *J. Virol.* **43**, 1123–1128.

Dumas, A. M., Geelen, J. L. M., Westrate, M. W., Wertheim, P. & Van der Noordaa, J. (1981). *Xba*I, *Pst*I, and *Bgl*II restriction enzyme maps of the two orientations of the varicella–zoster virus genome. *J. Virol.* **39**, 390–400.

Elkareh, A. A., Murphy, A. J. M., Fichter, T., Efstradiatis, A & Silverstein, S. (1985). " Transactivation" control signals in the promoter of the herpesvirus thymidine kinase gene. *Proc. Nat. Acad. Sci., U.S.A.* **82**, 1002–1006.

Fisher, F. B. & Preston, V. G. (1986). Isolation and characterization of herpes simplex virus type 1 mutant which fail to induce dUTPase activity. *Virology* **148**, 190–197.

Fleckenstein, B., Bornkamm, G. W. & Ludwig, H. (1975). Repetitive sequence in complete and defective genomes of *herpesvirus saimiri*. *J. Virol.* **15**, 389–406.

Fleckenstein, B., Bornkamm, G. W., Mulder, C., Werner, F. –J., Daniel, M. D., Falk, L. A. & Delius, H (1978). *Herpesvirus ateles* DNA and its homology with *herpesvirus saimiri* nucleic acid. *J. Virol.* **25**, 361–373.

Frame. M. C., Marsden, H. S. & McGeoch, D. J. (1986). Novel herpes simplex virus type 1 glycoproteins identified by antiserum against a synthetic oligopeptide from the predicted product of gene US4. *J. Gen. Virol.* **67**, 745–751.

Frame. M. C., Purves, F. C., McGeoch, D. J. & Leader, D. P. (1987). Identification

of the herpes simplex virus protein kinase as the product of viral gene US3. *J. Gen. Virol.* **68**, 2699–2704.

Frink, R. J., Eisenberg, R., Cohen, G. & Wagner, E. K. (1983). Detailed analysis of the portion of the herpes simplex virus type 1 genome encoding glycoprotein C. *J. Virol.* **45**, 634–647.

Frischauf, A. M., Garoff, H. & Lehrach, H. (1980). A subcloning strategy for DNA sequencing analysis. *Nucl. Acids Res.* **8**, 5541–5549.

Fuller, A. O. & Spear, P. G. (1987). Anti–glycoprotein D antibodies that permit adsorption but block infection by herpes simplex virus 1 prevent virion–cell fusion at the cell surface. *Proc. Nat. Acad. Sci., U.S.A.* **84**, 5454–5458.

Gibbs, J. S., Chiou, H. C. Hall, J. D., Mount, D. W., Retondo, M. J., Weller, S. K. &Coen, D. M. (1985). Sequence and mapping of the herpes simplex virus DNA polymerase gene predict a C–terminal substrate binding domain. *Proc. Nat. Acad. Sci., U.S.A.* **82**, 7969–7973.

Gielkens, A. L. J. & Berns, A. J. M. (1982). Differentiation between Aujeszky's disease virus strains by restriction endonuclease analysis of the viral DNA's. In *Aujeszky's Disease* pp. 3–13. Ed. G. Wittmann & S. A. Hall. Nijhoff, The Hague.

Given, D. & Kieff, E. (1979). DNA of Epstein–Barr virus. VI. Mapping of the internal tandem reiteration. *J. Virol.* **31**, 315–324.

Goding, C. R. & O'Hare, P. O. (1989). Herpes simplex virus Vmw65–octamer binding protein interaction: A paradigm for combinatorial control of transcription. *Virology* **173**, 363–367.

Goldberg, D. A. (1980). Isolation and partial characterization of the Drosophila alcohol dehydrogenase gene. *Proc. Nat. Acad. Sci., U.S.A.* **77**, 5794–5798.

Goldstein, D. J. & Weller, S. K. (1988). Herpes simplex virus type 1–induced ribonucleotide reductase activity is dispensable for virus growth and DNA synthesis: isolation and characterization of an ICP6 *lacZ* insertion mutant. *J.*

Virol. **62**, 196–205.

Gompels, U. & Minson, A. (1986). The properties and sequence of glycoprotein H of herpes simplex virus type 1. *Virology* **153**, 230–247.

Hanks, S. K., Qinn, A. M. & Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52.

Hattori, M. & Sakaki, Y. (1986). Dideoxy sequencing method using denatured plasmid templates. *Anal. Biochem.* **152**, 232–238.

Hayward, S. D., Nogee, L. & Hayward, G. S. (1980). Organization of repeated regions within Epstein–Barr virus DNA molecule. *J. Virol.* **33**, 507–521.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.

Herold, B. C., Wudunn, D., Soltys, N. & Spear, P. G. (1991). Glycoprotein C of herpes simplex virus type 1 plays a principal role in the adsorption of virus to cells and in infectivity. *J. Virol.* **65**, 1090–1098.

Hill, T. J. (1985). Herpes simplex virus latency. In *Herpesviruses* vol. 3, **chapter 4**, pp. 175–240. Ed. B. Roizman. Plenum Press, New York and London.

Holland, T. C., Homa, F., Marlin, S. D., Levin, M. & Glorioso, J. (1984). Herpes simplex virus type 1 glycoprotein C–negative mutant exhibit multiple phenotypes, including secretion of truncated glycoproteins. *J. Virol.* **52**, 566–574.

Holmes, D. W. & Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114**, 193–197.

Honess, R. D. & Roizman, B. (1974). Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral protein. *J. Virol.* **14**, 8–19.

Howard, S. T. & Smith, G. L. (1989). Two early vaccinia virus genes encode

polypeptides related to protein kinases. *J. Gen. Virol.* **70**, 3187–3201.

Hunter, T. (1982). Synthetic peptide substrates for a tyrosine protein kinasse. *J. Biol. Chem.* **257**, 4843–4848.

Ihara, S., Feldman, L., Watanabe, S. & Ben–Porat, T. (1983). Characterization of the immediate–early functions of pseudorabies virus. *Virology* **131**, 437–454.

Ish–Horowicz, D. & Burke, J. F. (1981). Rapid and efficient cosmid cloning. *Nucl. Acids Res.* **9**, 2989–2998.

Jacob, R. J. & Roizman, B. (1977). Anatomy of herpes simplex virus DNA. VIII. Properties of the replicating DNA. *J. Virol.* **23**, 394–411.

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

Jakubowicz, T. & Leader, D. P. (1987a). Induction, partial purification and characterization of a hamster fibroblast protein kinase activity that phosphorylates ribosomal protein S6. *Eur. J. Biochem.* **164**, 83–88.

Jakubowicz, T. & Leader, D. P. (1987b). Activation of a ribosomal protein S6 kinase in mouse fibroblast during infection with herpesvirus. *Eur. J. Biochem.* **168**, 371–376.

Johnson, R. M. & Spear, P. G. (1989). Herpes simplex virus glycoprotein D mediates interference with herpes simplex virus infection. *J. Virol.* **63**, 819–827.

Jones, S. W., Erikson, E., Blenis, J., Maller, J. L. & Erikson, R. L. (1988). A *Xenopus* ribosomal protein S6 kinase has two apparent kinase domains that are each similar to distinct protein kinases. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 3377–3381.

Kaner, R. J., Baird, A., Mansukhani, A., Basilico, C., Summers, B. D., Florkiewicz, R. Z. & Hajjar, D. P. (1990). Fibroblast growth factor receptor is a

portal of cellular entry for herpes simplex virus type 1. *Science* **248**, 1410–1413.

Kaplan, A. S. & Vatter, A. E. (1959). A comparison of herpes simplex and pseudorabies virus. *Virology* **7**, 394–407.

Katan, M., Stevely, W. S. & Leader, D. P. (1985). Partial purification and characterization of a new phosphoprotein kinase from cells infected with pseudorabies virus. *Eur. J. Biochem.* **152**, 57–65.

Katan, M., McGarvey, M. J., Stevely, W. S. & Leader, D. P. (1986). The phosphorylation of ribosomal protein S6 by protein kinases from cells infected with pseudorabies virus. *Biochem. J.* **239**, 205–211.

Kennedy, I. M., Stevely, W. S. & Leader, D. P. (1981). Phosphorylation of ribosomal proteins in hamster fibroblasts infected with pseudorabies virus or herpessimplex virus. *J. Virol.* **39**, 359–366.

Kit, S., Kit, M. & Otsuka, H. (1984). Structure of viral thymidine kinases. In *Herpesvirus* pp. 311–327. Ed. F. Rapp. Alan R. Liss, Inc.

Knopf, J. L., Lee, M. –H., Sultzman, L. A., Kriz, R. W., Loomis, C. R., Hewick, R. M. & Bell, R. M. (1986). Cloning and expression of mutiple protein kinase C cDNAs. *Cell* **46**, 491–502.

Kops, A. B. & Knipe, D. M. (1988). Formation of DNA replication structures in herpes virus–infected cells requires a viral DNA binding protein. *Cell* **55**, 857–868.

Kozak, M. (1989). The scanning model for translation: an update. *J. Cell Biol.* **108**, 229–241.

Kristie, T. & Roizman, B. (1986a). α4, the major regulatory protein of herpes simplex virus type 1, is stably and specifically associated with promoter–regulatory domains of α genes and of selected other viral genes. *Proc. Nat. Acad. Sci., U.S.A.* **83**, 3218–3222.

Kristie, T. & Roizman, B. (1986b). The binding sits of the major regulatory protein α4 specially associated with the promoter–regulatory domain of α genes of herpes simplex virus type 1. *Proc. Nat. Acad. Sci., U.S.A.* **83**, 4700–4704.

Kuenzel, E. A., Mulligan, J. A., Sommercorn, J. & Krebs, E. G. (1987). Substrate specificity determinants for casein kinase II as deduced from studies with synthetic peptides. *J. Biol. Chem.* **262**, 9136–9140.

Kwong, A. D. & Frenkel, N. (1989). The herpes simplex virus virion host shutoff function. *J. Virol.* **63**, 4834–4839.

Labeit, S., Lehrach, H. & Goody, R. S. (1987). DNA sequencing using α–thiodeoxynucleotides. *Methods Enzymology* **155**, 166–177.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, London* 227, 680–685.

Leader, D. P. & Katan, M. (1988). Viral aspects of protein phosphorylation. *J. Gen. Virol.* **69**, 1441–1464.

Leader, D. P. & Purves, F. C. (1988). The herpesvirus protein kinase. *Trends in Biochemical Science* **13**, 244–246.

Leader, D. P., Deana, A. D., Marchiori, F., Purves, F. C. & Pinna, L. A. (1991). Further definition of the substrate specificity of the alphaherpesvirus protein kinase and comparison with protein kinases A and C. *Biochim. Biophys. Acta* **1091**, 426-431.

Levin, D. E., Hammond, C. I., Ralston, R. O. & Bishop, J. M., (1987). Two yeast genes that encode unusual protein kinases. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 6035–6039.

Lomniczi, B., Watanabe, S., Ben–Porat, T. & Kaplan, A. S. (1984). Genetic basis of the neurovirulence of pseudorabies virus. *J. Virol.* **52**, 198–205.

Longnecker, R. & Roizman, B. (1986). Generation of an inverting herpes simplex virus 1 mutant lacking the L–S junction *a* sequences, an origin of DNA synthesis, and several genes including those specifying glycoprotein E and the

α47 gene. *J. Virol.* **58**, 583–591.

Longnecker, R. & Roizman, B. (1987). Clustering of genes dispensable for growth in culture in the S component of the HSV–1 genome. *Science* **236**, 573–576.

Longnecker, R., Chatterjee, S., Whitley, R. J. & Roizman, B. (1987). Idetification of a herpes simplex virus 1 glycoprotein gene with a gene cluster dispensable for growth in cell culture. *Proc. Nat. Acad. Sci., U.S.A.* **84**, 4303–4307.

Lörincz, A. T. & Reed, S. I. (1986). Sequence analysis of temperature–sensitive mutation in the *Saccharomyces cerevisiae* gene *CDC28. Mol. Cell. Biol.* **6**, 4099–4103.

Low, B. (1968). Formationof merodiploids in mating with a class of *Rec*⁻ recipient strains of Escherichia coli K12. *Proc. Nat. Acad. Sci., U.S.A.* **60**, 160–167.

Ludwig, H., Heppner, B. & Herrman, S. (1982). The genomes of different field isolates of Aujeszky's disease virus. In *Aujeszky's Disease* pp. 15–22. Ed. G. Wittmann & S. A. Hall. Nijhoff, The Hague.

Lukacs, N., Thiel, H. –J., Mettenleiter, T. C. & Rziha, H. –J. (1985). Demonstration of three major species of pseudorabies virus glycoproteins and identification of a disulfide–linked glycoprotein complex. *J. Virol.* **53**, 166–173.

MacPherson, I. & Stoker, M. (1962). Polyoma transformation of hamster cell clones — an investigation of genetic factors affecting cell competence. *Virology* **16**, 147–151.

Mavromara–Nazos, P. & Roizman, B. (1986). Construction and properties of a viable herpes simplex virus 1 recommbinant lacking the coding sequences of the α47 gene. *J. Virol.* **60**, 807–812.

McGeoch, D. J. & Davison, A. J. (1986a). DNA sequence of the herpes simplex virus type 1 gene encoding glycoprotein gH, and identification of homologous in the genome of varicella–zoster virus and Epstein–Barr virus. *Nuc. Acids Res.* **14**, 111–114.

McGeoch, D. J. & Davison, A. J. (1986b). Alphaherpesviruses possess a gene

homologous to the protein kinase gene family of eukaryotes and retroviruses. *Nuc. Acids Res.* **14**, 1765–1777.

McGeoch, D. J., Dolan, A., Donald, S. & Rixon, F. J. (1985). Sequence determination and genetic content of the short unique region of the genome of herpes simplex virus type 1. *J. Mol. Biol.* **181**, 1–13.

McGeoch, D. J., Dolan, A. & Frame, M. C. (1986a). DNA sequence of the region in the genome of herpes simplex virus type 1 containing the exonluclease gene and neighbouring genes. *Nuc. Acids Res.* **14**, 3435–3448.

McGeoch, D. J., Dolan, A., Donald, S. & Brauer, D. H. K. (1986b). Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type 1. *Nuc. Acids Res.* **14**, 1727–1745.

McGeoch, D. J., Moss, H. W., McNab, D. & Frame, M. C. (1987). DNA sequence and genetic content of the *Hin*dIII l region in the short unique component of the herpes simplex virus type 2 genome: identification of the gene encoding glycoprotein G, and evolutionary comparisons. *J. Gen. Virol.* **68**, 19–38.

McGeoch, D. J., Dalrymple, M. A., Davison, A. J., Dolan, A., Frame, M. C., McNab, D., Perry, L.J., Scott, J. E. & Taylor, P. (1988a). The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *J. Gen. Virol.* **69**, 1531–1574.

McGeoch, D. J., Dalrymple, M. A., Dolan, A., McNab, D., Perry, L.J., Taylor, P. & Challberg, M. D. (1988b). Structure of herpes simplex virus type 1 genes requried for replicatin of virus DNA. *J. Virol.* **62**, 444–453.

McKnight, S. L. (1980). The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. *Nuc. Acids Res.* **8**, 5949–5963.

McLauchlan, J & Clements, J. B. (1983). Organization of the herpes simplex virus type 1 transcription unit encoding two early proteins with molecular weight of 140000 and 40000. *J. Gen. Virol.* **64**, 997–1006.

Meignier, B., Longnecker, R., Mavromara–Nazos, P., Sears, A. & Roizman, B. (1988). Virulence of and establishment of latency by generally engineeered deletion mutants of herpes simplex virus 1. *Virology* **162**, 251–254.

Meselson, M. & Yuan, R (1968). DNA restriction enzyme from *E. coli*. *Nature, London* **217**, 1110–1114.

Mettenleiter, T. C., Lukacs, N. & Rziha, H. –J. (1985). Pseudorabies virus avirulent strains fail to express a major glycoprotein. *J. Virol.* **56**, 307–311.

Mettenleiter, T. C., Schreurs, C., Zuckerman, F., Ben–Porat, T. & Kaplan, A. (1988). Role of glycoprotein gIII of pseudorabies virus in virulence. *J. Virol.* **62**, 2712–2717.

Mettenleiter, T. C., Zsak, L., Zuckermann, F., Sugg, N., Kern, H. & Ben–Porat. (1990). Interaction of glycoprotein gIII with a cellular heparinlike substance mediates adsorption of pseudorabies virus. *J. Virol.* **64**, 278–286.

Mizusawa, S., Nshimura, S. & Seela, F. (1986). Improvment of the dideoxy chain termination method of DNA sequencing by use of deoxy–7–deazaguanosine triphosphate in place of dGTP. *Nucl. Acids Res.* **14**, 1319–1324.

Montell, C. & Rubin, G. M. (1988). The Drosophila *ninaC* locus encodes two photoreceptor cell specific proteins with domains homologous to protein kinases and the myosin heavy chain head. *Cell* **52**, 757–772.

Mount, S. M. (1982). A catalogue of splice junction sequences. *Nucl. Acids Res.* **10**, 459–472.

Mullaney, J., Moss, H. W. M. & McGeoch, D. J. (1989). Gene UL2 of herpes simplex virus type 1 encodes a uracil–DNA glycosylase. *J. Gen. Virol.* **70**, 449–454.

Murchie, M. –J. & McGeoch, D. J. (1982). DNA sequence analysis of an immediate–early gene region of the herpes simplex virus type 1 genome (map coordinates 0.950 to 0.978). *J. Gen. Virol.* **62**, 1–15.

Nikas, I., McLauchlan, J., Davison, A. J., Taylor, W. R. & Clemnets, J. B. (1986).

Structural features of ribonucleotide reductase. *Proteins* **1**, 376–384.

O'Hare, P. & Hayward, G. S. (1984). Expression of recombinant genes containing herpes simplex virus delayed early and immediate early regulatory regions and *trans* activation by herpesvirus infection. *J. Virol.* **52**, 522–531.

O'Hare, P. & Hayward, G. S. (1985). Evidence for a direct role for both the 175,000 and 110,000 molecular weight immediate–early proteins of herpes simplex virus in the transactivation of delayed–early promoters. *J. Virol.* **53**, 751–760.

Openshaw, H. (1983). Latency of herpes simplex virus in ocular tissue of mice. *Infect. Immun.* **39**, 960–962.

Osmani, S. A., Pu, R. T. & Morris, N. R. (1988). Mitotic induction and maintenance by overexpression of a G2–sepcific gene that encodes a potential protein kinase. *Cell* **53**, 237–244.

Pai, E. F., Kabsch, W., Krengel, U., Holmes, K. C., John, J. & Wittinghofer, A. (1989). Structure of the guanine–nucleotide–binding domain of the Ha–*ras* oncogene product p21 in the triphosphate conformation. *Nature, London* **341**, 209–214.

Parris, D. S., Cross, A., Haarr, L., Orr, A., Frame, M. C., Murphy, M., McGeoch, D. J. & Marsdon, H. S. (1988). Identification of the gene encoding the 65 kilodalton DNA binding protein of herpes simplex virus type 1. *J. Virol.* **62**, 818–825.

Paul, P. S., Mengeling, W. L. & Pirtle, E. C. (1982). Differentiation of peudorabies (Aujeszky's disease) virus strains by restriction endonuclease analysis. *Arch. Virol.* **73**, 193–198.

Pederson, N. E. & Enquist, L. W. (1988). The nucleotide sequence of a pseudorabies virus gene similar to ICP18.5 of herpes simplex virus type 1. *Nuc. Acids Res.* **17**, 3597.

Pellett. P. E., Kousoulas, K.G., Pereira. L. & Roizman, B. (1985a). The anatomy of

the herpes simplex virus 1 strain F glycoprotein B gene: primary sequence and predicted protein structure of the wild type and of monoclonal antibody–resistant mutant. *J. Virol.* **53**, 243–253.

Pellett. P. E., McKnight, J.L. C., Jenkins, F. J. & Roizman, B. (1985b). Nucleotide sequence and predicted amino acid sequence of a protein encoded in a small herpes simplex virus DNA fragment capable of *trans*–inducing α genes. *Proc. Nat. Acad. Sci., U.S.A.* **82**, 5870–5874.

Perry, L. J. & McGeoch, D. J. (1988). The DNA sequences of the long repeat region and adjoining parts of the long unique region in the genome of herpes simplex virus type 1. *J. Gen. Virol.* **69**, 2831–2846.

Perry, L. J., Rixon, F. J., Everett, R. D., Frame, M. C. & McGeoch, D. J. (1986). Characterization of the IE110 gene of herpes simplex virus type 1. *J. Gen. Virol.* **67**, 2365–2380.

Pertuiset, B., Boccara, M., Cebrian, J., Berthelot, N., Chousterman, S., Puvion–Dutilleul, F., Sisman, J. & Sheldrick, P. (1989). Physical mapping and nucleotide sequence of a herpes simples virus type 1 gene required for capsid assembly. *J. Virol.* **63**, 2169–2179.

Petrovskis, E. A. & Post, L. E. (1987). A small open reading frame in pseudorabies virus and implications for evolutionary relationships between herpesviruses. *Virology* **159**, 193–195.

Petrovskis, E. A., Timmins, J. G. & Post, L. E. (1986a). Use of λgt11 to isolate genes for two peudorabies virus glycoproteins with homology to herpes simplex virus and varicella–zoster virus glycoproteins. *J. Virol.* **60**, 185–193.

Petrovskis, E. A., Timmins, J. G., Armentrout, M. A., Marchioli, C. C., Yancey, R. J. & Post, L. E. (1986b). DNA sequence of the gene for pseudorabies virus gp50, a glycoprotein without N–linked glycosylation. *J. Virol.* **59**, 216–223.

Petrovskis, E. A., Timmins, J. G., Gierman, T. M. & Post, L. E. (1986c). Deletions

in vaccine strains of pseudorabies virus and their effect on synthesis of glycoprotein gp63. *J. Virol.* **60**, 1166–1169.

Petrovskis, E. A., Meyer, A. & Post, L. E. (1988). Reduced yield of infectious pseudorabies virus and herpes simplex virus from cell lines producing viral glycoprotein gp50. *J. Virol.* **62**, 2196–2199.

Poffenberger, K. L. & Roizman, B. (1985). Studies on a non–inverting genome of a viable herpes simplex virus 1. Presence of head–to–tail linkage in packaged genomes and requirements for circularization after infection. *J. Virol.* **53**, 589–595.

Post, L. E. & Roizman, B. (1981). A generalized technique for deletion of sepecific genes in large genomes: a gene 22 of herpes simplex virus 1 is not essential for growth. *Cell* **25**, 227–232.

Preston, C. M. & Cordingley, M. J. (1982). mRNA– and DNA–directed synthesis of herpes simplex virus–coded exonuclease in *Xenopus laevis* oocytes. *J. Virol.* **43**, 386–394.

Preston, V. G. & Fisher, F. B. (1984). Identification of the herpes simplex virus type 1 gene encoding the dUTPase. *Virology* **138**, 58–68.

Preston, V. G., Coates, J. A. V. & Rixon, F. J. (1983). Identification and charaterization of a herpes simplex virus gene product required for encapsidation of virus DNA. *J. Virol.* **45**, 1056–1064.

Preston, V. G., Palfreyman, J. W. & Dutia, B. M. (1984). Identification of a herpessimplex virus type 1 polypeptide which is a component of the virus–induced ribonucleotide reductase. *J. Gen. Virol.* **65**, 1457–1466.

Purves , F. C. (1987). *The protein kinases induced in cells infected with herpesviruses.* Ph.D. thesis, University of Glasgow.

Purves, F. C., Donella Deana, A., Marchiori, F., Leader, D. P. & Pinna, L. A. (1986a). The substrate specificity of the protein kinase induced in cells infected with herpesviruses: studies with synthetic substrates indicate structural

requirements distinct from other protein kinases. *Biochim. Biophys. Acta* **889**, 208–215.

Purves, F. C., Katan, M., Stevely, W. S. & Leader, D. P. (1986b). Characteristics of the induction of a new protein kinase in cells infected with herpesviruses. *J. Gen. Virol.* **67**, 1049–1057.

Purves, F. C., Katan, M. & Leader, D. P. (1987a). Complete purification of the pseudorabies virus protein kinase. *Eur. J. Biochem.* **167**, 507–512.

Purves, F. C., Longnecker, R. M., Leader, D. P. & Roizman, B. (1987b). Herpes simplex virus 1 protein kinase is encoded by open reading frame US3 which is not essential for virus growth in cell culture. *J. Virol.* **61**, 2896–2901.

Quinn, J. P. & McGeoch, D. J. (1985). DNA sequence of the region in the genome of herpes simplex virus type 1 containing the genes for DNA polymerase and the major DNA binding protein. *Nuc. Acids Res.* **13**, 8143–8163.

Rea, J., Timmins, J. G., Long, G. W. & Post, L. E. (1985). Mapping and sequence of the gene for the pseudorabies glycoprotein which accumulates in the medium of infected cells. *J. Virol.* **54**, 21–29.

Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**, 237–251.

Rixon, F. J. & Ben–Porat, T. (1979). Structural evolution of the DNA of pseudorabies–defective virus particles. *Virology* **97**, 151–163.

Rixon, F. J. & Clement, J. B. (1982). Detailed structure analysis of two spliced HSV–1 immediate–early mRNAs. *Nuc. Acids Res.* **12**, 2473–2487.

Rixon, F. J. & McGeoch, D. J. (1985). Detailed analysis of the mRNAs mapping in the short unique region of herpes simplex virus type 1. *Nucl. Acids Res.* **13**, 953–973.

Rixon, F. J., Davison, M. D. & Davison A. J. (1990). Identification of the genes

encoding two capsid proteins of herpes simplex type 1 by direct amino acid sequencing. *J. Gen. Virol.* **71**, 1211–1214.

Robbins, A. K., Dorney, D. J., Wathen, M. W., Whealy, M. E., Gold, C., Watson, R. J., Holland, L. E., Weed, S. D., Levine, M., Glorioso, J. C. & Enquist, L. W. (1987). The pseudorabies virus gII gene is closely related to the gB glycoprotein gene of herpes simplex virus. *J. Virol.* **61**, 2691–2701.

Robbins, A. K., Watson, R. J., Whealy, M. E., Hays, W. W. & Enquist, L. W. (1986a). Characterization of a pseudorabies virus glycoprotein gene with homology to herpes simplex virus type 1 and type 2 glycoprotein C. *J. Virol.* **58**, 339–347.

Robbins, A. K., Whealy, M. E., Watson, R. J. & Enquist, L. W. (1986b). Pseudorabies virus gene encoding glycoprotein gIII is not essential for growth in tissue culture. *J. Virol.* **59**, 635–645.

Roizman, B. (1982). The family Herpesviridae: general description, taxonomy, and classification. In *Herpesviruses* vol. 1, **chapter 1**, pp. 1–23. Ed. B. Roizman. Plenum Press, New York and London.

Roizman, B. & Batterman, W. (1984). The replication of herpesviruses. In *General Virology* pp.497–526. Ed. B.Fields. New York: Raven.

Roizman, B. & Furlong, F. (1974). The replication of herpesviruses. In *Comprehensive Virology* vol. 3, pp. 229–403. Ed. H. Fraenkel–Conrad & R. R. Wagner. Plenum Press, New York.

Roizman, B. & Sears, A. E. (1987). An inquiry into the mechanisms of herpes simplex virus latency. *Ann. Rev. Microbiol.* **41**, 543–571.

Roller, R. J. & Roizman, B. (1990). The herpes simplex virus Us11 open reading frame encodes a sequence–sepecific RNA–binding protein. *J. Virol.* **64**, 3463–3470.

Rossmann, M. G., Moras, D. & Olsen, K. W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *Nature, London* **250**, 194–199.

Sacks, W. R., Greene, C. C., Aschman, D. P. & Schaffer, P. A. (1985). Herpes simplex virus type 1 ICP27 is an enssetial regulatory protein. *J. Virol.* **55**, 796–805.

Saiki, R. K., Gelfand, D. H., Stoffel, S. J., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988). Primer–directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.

Sanders, P. G., Wilkie, N. M. & Davison, A. J. (1982). Thymidine kinase deletion mutants of herpes simplex virus type 1. *J. Gen. Virol.* **63**, 277–295.

Sanger, F., Nicklin, S. & Coulson, A. R. (1977). DNA sequencing with chain–terminating inhibitors. *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.

Saxena, A., Padmanabha, R. & Glover, C. V. C. (1987). Isolation and sequencing of cDNA clones encoding alpha and beta subunits *Drosophila melanogaster* casein kinase II. *Mol. Cell. Biol.* **7**, 3409–3417.

Schauder, B., Blöcker, H., Frank, R. & McCarthy, J. E. G. (1987). Inducible expression vectors incorporating the *Escherichia coli* atpE translational initiation region. *Gene* **52**, 279–283.

Sears, A. E., Halliberton, I. W., Meignier, B., Silver, S. & Roizman, B. (1985). Herpes simplex virus mutant deleted in the α22 gene: growth and gene expression in permissive and restrictive cells, and establishment of latency in mice. *J. Virol.* **55**, 338–346.

Sheldrick, P. & Berthelot, N. (1974). Inverted repetitions in the chromosome of herpes simplex virus. *Cold Spring Harbor Symposia on Quantitative Biology* **39**, 667–678.

Smith, R. F. & Smith, T. F. (1989). Identification of new protein kinase–related genes in three herpes viruses, herpes simplex viruse, varicella–zoster virus, and Epstein–barr virus. *J. Virol.* **63**, 450–455.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–517.

Stevely, W. S. (1977). Inverted repetition in the chromosome of pseudorabies virus. *J. Virol.* **22**, 232–234.

Stevens, R. (1989). *Studies of the protein kinases of α–herpes viruses*. Ph. D. thesis, Glasgow University.

Stow, N. D. (1985). Mutagenesis of a herpes simplex virus origin of DNA replication and its effect on viral interference. *J. Gen. Virol.* **66**, 31–42.

Stow, N. D. & Davison, A. J. (1986). Identification of a varicella–zoster virus origin of DNA replication and its activation by herpes simplex virus type 1 gene products. *J. Gen. Virol.* **67**, 1613–1623.

Stow, N. D. & McMonagle, E. C. (1983). Characterization of the TRs/IRs origin of DNA replication of herpes simplex virus type 1. *Virology* **130**, 427–438.

Taylor, S. S., Buechler, J. A. & Yonemoto, W. (1990). cAMP–dependent protein kinase: Framework for a diverse family of regulatory enzyme. *Annu. Rev. Biochem.* **59**, 971–1005.

Thompson, R., Honess, R. W., Taylor, L., Morran, J. & Davison, A. J. (1987). Varicella–zoster virus specifies a thymidylate synthetase. *J. Gen. Virol.* **68**, 1449–1455.

Thomsen, D. R., Marchioli, C. C., Yancey, JR., R. J. & Post, L. E. (1987). Replication and virulence of pseudorabies virus mutants lacking glycoprotein gX. *J. Virol.* **61**, 229–232.

Tobimatsu, T., Kameshita, I. & Fujisawa, H. (1988). Molecular cloning of the cDNA encoding the third polypeptide (γ) of brain calmodulin–dependent protein kinase II. *J. Biol. Chem.* **263**, 16082–16086.

Umene, K. (1989). Short duplicated sequence indicative of the recombinogenicity of the junction between a unique and an inverted repeat sequence in the S component of the herpes simplex virus type 1 genome. *J. Virol.* **63**, 1877–1883.

Van Zijl, M., Van der Gulden, H., De Wind, N., Gielkens, A. & Berns, A. (1990). Identification of two genes in the unique short region of peudorabies virus; comparison with herpes simplex virus and varicella–zoster virus. *J. Gen. Virol.* **71**, 1747–1755.

Vlcek, C., Kozmík, Z., Paces, V., Schirm, S. & Schwyzer, M (1990). Pseudorabies virus immediate–early gene overlaps with an oppositely orientated openreading frame: characterization of their promoter and enhancer regions. *Virology* **179**, 365–377

Wadswoth, S., Jacob, R. J & Roizman, B. (1975). Anatomy of herpes simplex virus DNA. II.Size, composition, and arrangement of inverted repeat terminal repetitions. *J. Virol.* **15**, 1487–1497.

Wagner, M. J., Sharp, J. A. & Summers, W. C. (1981). Nucleotide sequence of the thymidine kinase gene of herpes simplex type 1. *Proc. Nat. Acad. Sci., U.S.A.* **78**, 1441–1445.

Wathen, M. W. & Wathen, L. M. (1986). Characterization and mapping of a nonessential pseudorabies virus glycoprotein. *J. Virol.* **58**, 173–178.

Watson, R. J. & Clements, J. B. (1980). A herpes simplex virus function continuously required for early and late virus RNA synthesis. *Nature, London* **285**, 329–330.

Watson, R. J., Sullivan, M. & Van de Woude, G. F. (1981). Structure of two spliced herpes siplex virus type 1 immediate–early mRNA's which map at the junctions of the unique and reiterated regions of the virus DNA S component. *J. Virol.* **37**, 431–444.

Watson, R. J., Weis, J. H., Salstrom, J. S. & Enquist, L. W. (1982). Herpes simplex virus type–1 glycoprotein D gene: nucleotide sequence and expression in *Escherichia coli. Science* **218**, 381–383.

Weir, H. M., Calder, J. M. & Stow, N. D. (1989). Binding of the herpes simplex virus type 1 UL9 gene product to an origin of viral DNA replication. *Nuc.*

Acids Res. **17**, 1409–1425.

Weller, S. K., Spadaro, A., Schaffer, J. E., Murray, A. W., Maxam, A. M. & Schaffer, P. A. (1985). Cloning, sequencing, and functional analysis of ori_L, a herpes simplex virus type 1 origin of DNA synthesis. *Mol. Cell. Biol.* **5**, 930–942.

Westrate, M. W., Geelen, J. L. M. C. & Van der Noordaa, J. (1980). Human cytomegalovirus DNA: Physical maps for the restriction endonucleases *Bgl*II, *Hin*dIII and *Xba*I. *J. Gen. Virol.* **49**, 1–21.

Whalley, J. M., Robertson, G. R. & Davison, A. J. (1981). Analysis of the genome of equine herpesvirus type 1: Arrangement of cleavage sites for restriction endonuleases *Eco*RI, *Bgl*II and *Bam*HI. *J. Gen. Virol.* **57**, 307–323.

Whealy, M. E., Robbins, A. K. & Enquist, L. W. (1988). Pseudorabies virus glycoprotein gIII is required for efficient virus growth in tissue culture. *J. Virol.* **62**, 2512–2515.

Whealy, M. E., Robbins, A. K. & Enquist, L. W. (1990). The export pathway of the pseudorabies virus gB homolog gII involves oligomer formatiom in the endoplasmic reticulum and protease processing in the golgi apparatus. *J. Virol.* **64**, 1946–1955.

Whitton, J. L & Clements, J. B. (1984). The junctions between the repetitive and the short unique sequences of the herpes simplex virus genome are determined by the polypeptide–coding regions of two spliced immediate–early mRNA. *J. Gen. Virol.* **65**, 451–466.

Widly, P., Russel, W. C. & Horne, R. W. (1960). The morphology of herpes viruses. *Virology* **12**, 204–222.

Wilcox, K. W., Kohn, A., Skylanskaya, E. & Roizman, B. (1980). Herpes simplex virus phosphoproteins. I. Phosphate cycles on and off some viral polypeptides and can alter their affinity for DNA. *J. Virol.* **33**, 167–182.

Wilkie, N. M., Stow, N. D., Marsden, H. S., Preston, V., Cortini, R., Timbury, M.

C. & Subak–Sharpe, J. H. (1977). Physical mapping of herpes simplex virus coded functions and polypeptides by marker rescue and analysis of HSV–1/HSV–2 intertypic recombinants. In *Oncogenesis and Herpesviruses III.* vol. I, pp. 11–31. Ed. G. de The, W. Henle & F. Rapp. Lynon: IARC Scientific Publications, no. 24.

Worrad, D. M. & Caradonna, S. (1988). Identification of the coding sequence for herpes simplex virus uracil–DNA glycosylase. *J. Virol.* **62**, 4774–4777.

Wölfer, U., Kruft, V., Sawitzky, D., Hampl, H., Wittmann–Liebold, B. & Habermehl, K. –O. (1990). Processing of pseudorabies virus glycoprotein gII. *J. Virol.* **64**, 3122–3125.

Wu, C. A., Nelson, N. J., McGeoch, D. J. & Challberg, M. D. (1988). Idetification of herpes simplex virus type 1 genes required for origin–dependent DNA sythesis. *J. Virol.* **62**, 435–443.

Wu, C. –L. & Wilcox, K. W. (1991). The conserved DNA–biding domains encoded by the herpes simplex virus type 1 ICP4, pseudorabies virus IE180, and varicella–zoster virus ORF62 genes recognize similar sites in the corresponding promoters. *J. Virol.* **65**, 1149–1159.

Wudunn, D. & Spear, P. G. (1989). Initial interaction of herpes simples virus with cells is binding to heparan sulfate. *J. Virol.* **63**, 52–58.

Yanish–Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vector. *Gene* **33**, 103–119.

Zhang, G., Stevens, R. & Leader, D. P. (1990). The protein kinase encoded in the short unique region of peudorabies virus: description of the gene and identification of its product in virions and in infected cells. *J. Gen. Virol.* **71**, 1757–1765.

Zooler, M. J., Nelson, N. C. & Taylor, S. S. (1981). Affinity labeling of

cAMP–dependent protein kinase with *p*–fluorosulfonylbenzoyl adenosine. *J. Biol. Chem.* **256**, 10837–10842.

Zsak, L., Mettenleiter, T. C., Sugg, N. & Ben–Porat, T. (1989). Release of peudorabies virus from infected cells is controlled by several viral functions and is modulated by cellular components. *J. Virol.* **63**, 5475–5477.

Zsak, L., Mettenleiter, T. C. Sugg, N. & Ben–Porat, T. (1990). Effect of polylysine on the early stages of infection of wild type pseudorabies virus and of mutants defective in gIII. *Virology.* **179**, 330–338.

Zuckermann, F. A., Zsak, L., Mettenleiter, T. C. & Ben–Porat, T. (1990). Pseudorabies virus glycoprotein gIII is a major target antigen for murine and swine virus–specific cytotoxic T lymphocytes. *J. Virol.* **64**, 802–812.