RECOMBINATIONAL AND PACKAGING SIGNALS

IN

HERPES SIMPLEX VIRUS

DEOXYRIBONUCLEIC ACID

by

SUSANNAH LOUISE VARMUZA, BSc., MSc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

June, 1985

RECOMBINATION AND PACKAGING OF HERPES SIMPLEX VIRUS DNA

DOCTOR OF PHILOSOPHY (1985) McMASTER UNIVERSITY

(Medical Sciences) Hamilton, Ontario

TITLE: Recombinational and Packaging Signals in Herpes Simplex Virus Deoxyribonucleic Acid

AUTHOR: Susannah Louise Varmuza B.Sc. (McMaster University)

M.Sc. (McMaster University)

SUPERVISOR : Dr. J.R. Smiley

NUMBER OF PAGES: xiv, 267

ABSTRACT

Herpes Simplex Virus DNA displays a number of unusual features which have been the subject of intense scrutiny in a number of laboratories. The genome is composed of two segments, each of which is flanked by inverted repeats. These segments invert freely with respect to each other generating equimolar quantities of four different isomers. This phenomenon, called segment inversion, was reputed to be the result of a site specific recombination mechanism operating on the terminal repeat, the "a" sequence, which is part of the inverted repeats flanking each segment. The "a" sequence was also implicated as the cleavage/packaging signal utilized by the virus to process viral DNA concatemers. The underlying mechanism of this process was believed to be a double strand break at a specific site between two "a" sequences. The models of HSV maturation were deficient, however, in explaining several phenomena, namely the tendency of the "a" sequence to accumulate tandem iterations of itself, the asymmetric distribution of these tandem iterations to one end of the genome, but not to the other, and the ability of defective genomes, which do not have tandemly iterated "a" sequences, at least initially, to be efficiently packaged. I have shown that the "a" sequence actually contains two signals for cleavage/packaging, not one, that the cleavage occurs at specific distances from these signals, not in specific sequences, and that the cleavage mechanism results in a duplication of the cleavage signal and flanking DNA. Furthermore, I have determined that the "a" sequence is not a target for site specific recombination, and that there is better evidence to support the idea

that segment inversion is accomplished by a number of related, but independent mechanisms, including generalized recombination.

Acknowledgements

This work is the distillation of many hours of stimulating conversation, shared ideas, disappointing failures, and soaring triumphs. For much of this I owe a debt of gratitude to my supervisor, Jim Smiley. His critical appraisal has been invaluable. Thanks are also due to Silvia Bacchetti for lending me her excellent editorial skills. To my colleagues whose unqualified moral support helped me over the rough spots - Marnie Halpern and Eddy Littler in particular - I offer thanks with gratitude and affection. Finally, words are inadequate to express my debt to my husband, Petr Varmuza, whose boundless patience and sense of humour have been the staying forces in my life.

This Thesis is Dedicated

to the Memory of

my stepfather

Samuel Burke

TABLE OF CONTENTS

Table of Contents contd.

LIST OF ABBREVIATIONS

HSV	Herpes Simplex Virus
HSV-1	Herpes Simplex Virus Type 1
HSV-1(KOS).	Herpes Simplex Virus Type 1 Strain KOS
KOS	Herpes Simplex Virus Type 1 Strain KOS
F	Herpes Simplex Virus Type 1 Strain F
17	Herpes Simplex Virus Type 1 Strain 17
USA-8	Herpes Simplex Virus Type 1 Strain USA-8
Justin	Herpes Simplex Virus Type 1 Strain Justin
HSV-2	Herpes Simplex Virus Type 2
CMV	Human Cytomegalovirus
EBV	Epstein-Barr Virus
VZV	Varicella Zoster Virus
SV40	Simian Virus 40
TK	Thymidine Kinase
TK^-	Thymidine Kinase Deficient
ATP	Adenosine triphosphate
G	Deoxyguanosine monophosphate
A	Deoxyadenosine monophosphate
T	Deoxythymidine monophosphate
C	Deoxycytosine monophosphate
bp	base pair
kb	kilobase pair (bp x 10^3)
kbp	kilobase pair
kdalton	kilodalton

Abbreviations contd.

MOI	Multiplicity of Infection
pfu	plaque forming unit
^{32}P	radioisotopic phosphorus
DNA	Deoxyribonucleic Acid

LIST OF FIGURES AND TABLES

List of Figures and Tables contd.

List of Figures and Tables contd.

INTRODUCTION

When Gregor Mendel's brilliant insights emerged from obscurity at the beginning of this century, biology moved firmly into the scientific realm. The mechanism of heredity could be seen as a process that obeyed Newtonian-like laws, and could therefore be studied with scientific approaches based on experimentation of cause and effect.

The rapid expansion of knowledge in the field of molecular biology has produced a set of fairly flexible rules by which most organisms seem to conduct their inner affairs. I say flexible because for every "rule" there are usually exceptions of one form or another (Nature is something of an opportunist). Indeed, even Mendel, whose experiments have been branded too perfect by some (Whitehouse, 1977), apparently failed to make observation of one basic feature of heredity which may eventually hold the key to an understanding of larger mysteries such as evolution. That feature is the ordered linkage of genes within a chromosome. It was discovered by early geneticists attempting to repeat Mendel's experiments with other organisms and expanded libraries of "characters" or genes. Even after linkage was established as a general feature of heredity, however, exceptions again were found to the new "rule". These were discovered to be the result of a fairly ubiquitous process, known as recombination, in which alleles of different genes appeared to move from one chromosome to another. This ability of genetic material to move around and still maintain

order has fascinated geneticists for decades. In the process of unravelling the mysteries of genetic recombination, investigators have discovered some amazing tricks devised by nature to preserve and expand genetic information.

Recombination

Recombination during Meiosis and Mitosis in Eucaryotes

Mendel's experiments measured the random assortment of unlinked genes during sexual reproduction. Later work showed that the frequency of co-segregation of two linked markers reflected (inversely) the distance between them; close markers frequently segregated together, while distant markers segregated separately. This property allowed geneticists to order, or map, different genes along a chromosome. Thus, recombination became a tool as well as a phenomenon.

Sexually reproducing organisms recombine their chromosomes during meiosis. This allows even greater mixing of genetic material than that afforded by random assortment of chromosomes, presumably to ensure variation within the population as a hedge against genetic catastrophe. Recombination is believed to occur at an early stage of meiosis, after DNA replication. The chromosomes form a structure called a "synaptonemal complex" characterized by the synapsis of homologous chromosomes along a polymeric protein backbone. Shortly after it is formed, chiasmata can be observed at various sites along the paired chromosomes. Chiasmata have been positively correlated with chromosome cross-overs and with recombination frequency. For example, male Drosophila and the anthers of Fritillaria japonica do not exhibit

chiasmata during meiosis and do not recombine their genes (Catcheside, 1977; Henderson, 1970).

Nature has kindly supplied geneticists with wonderful tools to study recombination in the form of fungi and yeasts. These organisms retain the property of independent growth of both haploid and diploid forms, although in some fungi the diploid stage is short and highly specialized. Moreover, in some fungi, the meiotic products (the spores) are conveniently ordered in the ascus according to the planes of division of the original precursor cell. Many fungi undergo a post-meiotic mitosis, yielding 8 spores from a single diploid cell. The eight spores are thus analogous to the eight strands of DNA within the interacting chromosomes at the diplotene stage of meiosis during which chiasmata form, and, by deductive reasoning, during which recombination takes place. This allows the investigator to follow the fate of all the participating chromatids during meiosis (reviewed in Catcheside, 1977; Stahl, 1979). This is an important concept in the study of recombination because it ensures an "unbiased" sampling of the population. Analysis of ascospores, referred to in the trade as tetrad analysis, has provided biology with some basic ground rules of the behaviour of recombining chromosomes. First, distant markers almost always recombine reciprocally, ie. there is no net loss of genetic material during recombination, only rearrangement. Second, this rule does not apply to very close markers. Tetrad analysis of ordered asci from Neurospora crassa and Sordaria fimicola occasionally yields ascospores with asymmetrically distributed genes. These are referred to as "6:2", "5:3", and "aberrant 4:4" tetrads, and were originally

discovered during tetrad analysis of spores produced by a cross of two different strains of Neurospora which differed at a locus defining spore colour. Most ascospores contained 4 coloured (wild type - wt):4 colourless (mutant - m) spores, indicating that no rearrangement of the spore colour gene with respect to the centromere occurred. Some contained 2 wt:2 m:2 wt:2m spores, indicating that the spore colour gene had been exchanged between the two internal chromatids. The beauty of ordered asci is that the products reflect exchanges between individual chromatids within the division plane. Thus, a 2:4:2 arrangement reflects recombination between non-adjacent chromatids within the division plane. Rare ascopores with 6:2 (or 2:6) and 5:3 (or 3:5) arrangements were also detected. These asci were interpreted as the result of nonreciprocal exchange of genetic information within the spore colour locus. The best mechanistic explanation of the phenomenon so far is the repair of mismatched heteroduplexes formed during recombination, or, gene conversion.

The observation of gene conversion during meiosis, and the patterns exhibited by different organisms, allowed investigators to infer mechanisms of recombination which encompassed this unusual trait. All of these models begin with the assumption that gene conversion is a consequence of the mechanism employed by the cell to recombine its chromosomes. This interpretation is based on the observation that approximately 50% of the time conversion is associated with recombination of flanking markers.

Models of recombination are themselves subject to evolution as data from different organisms becomes available. The first widely known

model was proposed by Holliday (1964). He envisioned two nicks in the equivalent DNA strands of homologous chromatids followed by strand exchange and coincident migration of the two branches of heteroduplex DNA. Such cross strand structures, referred to as chi structures, have been observed in electron micrographs of recombination intermediates obtained from several bacterial systems (reviewed in Potter and Dressler, 1982). Repair synthesis of mismatched bases would account for gene conversion if the repair was asymmetric. For example, if both heteroduplexes were repaired in the same direction with respect to the chosen alleles, 6:2 tetrads would be generated, if one heteroduplex was repaired while the other was not, 5:3 tetrads would be generated, and if neither heteroduplex was repaired, aberrant 4:4 tetrads would be generated. This did not fit well with the observation that some alleles which exhibited certain behaviours failed to fulfill other predictions of the model; for example, a preference for 5:3 segregation of conversion asci should be accompanied by a similar preference for aberrant 4:4 asci, but is not in some organisms. Moreover, analysis of 5:3 conversion asci revealed the absence of asci with certain predicted genotypes if the original exchange was reciprocal. For example, consider the cross (AbCD) x (aBcd), where C is the marker being scored, and where B is a very close marker, while A and D are more distant markers. Reciprocal exchanges across the C locus in which one heteroduplex is repaired and the other is not will result in 5:3 asci. Ignoring recombination of flanking markers, the genotypes of the asci should be : (AbCD) (AbCD) (AbCD) <u>(AbcD)</u> <u>(aBCd)</u> <u>(aBCd)</u> (aBcd) (aBcd), where the underlined combinations represent the reciprocal exchanges.

Analysis of 5:3 asci revealed the absence of reciprocal products, and typically yielded genotypes : (AbCD) (AbCD) (AbCD) (AbCD) (aBCd) (aBcd) (aBcd) (aBcd). This suggested that, instead of asymmetric repair of reciprocal exchanges, the underlying reason for nonreciprocality of conversion was the presence of a more extensive heteroduplex on one chromatid than on the other (reviewed in Catcheside, 1977; Whitehouse, 1977; Stahl, 1979; Dressler and Potter, 1982). This led to a modification of Holliday's model, proposed by Meselson and Radding (1975), which is sometimes referred to as the Aviemore model.

Meselson and Radding (1975) proposed that recombination proceeds via a single strand invasion of one DNA duplex following a single nick in the homologous duplex. The displaced strand is degraded until it decides to invade its neighbour further down the chromosome. The space left by the original invading strand is filled in by repair synthesis until it reaches the roadblock set up by the tardy invading strand. In addition to accounting for observed results from genetic studies of fungi and yeast, this model is singular in being supported by biochemical studies of recombining molecules in bacterial systems. Elegant experiments, mainly from Radding's lab, performed in vitro with various DNA substrates and RecA protein from E. coli have indicated that exchange appears to be directional, with the invading strand proceeding in a 5'to 3' direction (reviewed in Radding, 1982). This would leave a 3' end behind at the original nick, if one assumes that the same directionality is exhibited in eucaryotes. DNA synthesis also proceeds in a 5' to 3' direction; thus the 3' end left behind would serve as a primer for repair synthesis.

The double-strand-break model proposed by Szostak et al (1982) was a response to results they obtained with yeast transformants. Yeast investigators enjoy the happy fact that exogenous DNA can be introduced into the cells and will recombine at the homologous site in the yeast genome, thus facilitating the construction of mutants. Szostak and co-workers found that transformation frequency could be increased by several thousand fold if the exogenous DNA was first cleaved within the region of homology. Moreover, if the input DNA contained a gap, the gap was repaired during insertion of the exogenous DNA. The double-strand-break repair model explains this serendipitous fact by proposing that recombination in yeast is initiated by double strand breaks which are expanded by exonuclease to form gaps. One of the 3' ends thus created invades the homologue and begins repair synthesis, creating a D-loop, which will hybridize to the other free end as soon as the gap is traversed. This creates two Meselson-Radding exchanges on either side of the gap.

Recombination is not confined to meiotically dividing cells, and may occur at reduced frequency in mitotic cells, probably in response to DNA or chromosome damage (reviewed in Catcheside, 1977; Kunz and Haynes, 1981). Mitotic recombination was first described in Drosophila in the 1930's as the underlying mechanism by which "twin spots" of homozygous tissue arose in heterozygous females (remember that male Drosophila do not recombine their chromosomes) (Stern, 1936). The frequency of this phenomenon could be increased by irradiating the insect, and in fact embryologists have employed this technique to prepare fate maps of developing insect embryos (reviewed in Crick and

Lawrence, 1975). Analysis of mitotic recombination is aggravated by the inability to isolate individual products of a recombination event. Nevertheless, by using specially constructed strains of fungi and yeast, investigators have established that mitotic recombination proceeds most frequently by non reciprocal conversion without associated exchange of flanking markers, although the latter is not strictly excluded (Catcheside, 1977; Jackson and Fink, 1984; Klein, 1984; Klar and Strathern, 1984). This has led to the suggestion that mitotic recombination may not utilize the same mechanisms as its meiotic counterpart (Fink and Petes, 1984).

Recombination in Procaryotes

Bacterial organisms have been a continual source of delight to molecular biologists because of their rapid growth, relatively simple organization, and especially because of the ease with which mutants in various genes can be isolated and analysed. So it is not surprising that recombinational pathways in procaryotes are much better understood than their counterparts in eucaryotes.

The first recombinational pathway to be characterized was that employed by bacteria during conjugation (reviewed in Lewin, 1977). Bacterial cells can exchange genetic material by means of a unidirectional process initiated in some cells harbouring a piece of DNA called an F episome. The F episome may exist separately (F^+), may be integrated into the bacterial genome (Hfr), or may be separate but contain bacterial sequences, presumably picked up by successive integration into and excision from the bacterial genome (F'). The F DNA encodes the

machinery required to make structures called F or sex pili, proteinaceous tubes which connect the donor (F containing) cell with the recipient (F-) cell. F DNA moves through the tube from the donor to the recipient. If bacterial sequences are attached, as they are in Hfr and F' cells, then they too are transferred. Transfer starts at a specific site within F DNA. The proximity of the linked genes to this site has been shown to reflect the frequency and time of transfer, indicating that it is a temporally limited event. This latter property allowed investigators to use Hfr mediated transfer of marked genes to map the E.coli chromosome.

The DNA being transferred through the F pilus was shown to be single stranded, with a 5' leading end. This DNA integrated at its homologous site in the recipient chromosome by displacing its equivalent strand. The integration did not require DNA synthesis, but it did require a number of gene products defined by this property as *rec* genes. Most prominent among these were *recA*, *recB* and *recC*. It is now known that *recB* and *recC* encode the two subunits of the *recBC* nuclease, exonuclease V, while *recA* encodes a protein which exerts pleiotropic effects through two different activities, a stoichiometric DNA binding activity required by all of the generalized recombination pathways in *E. coli*, and a protease activity specific for certain repressors of the SOS DNA repair pathway (Radding, 1982; Dressler and Potter, 1982; Little and Mount, 1982). These are by no means the only bacterial genes, nor is *recBC* the only pathway, involved in general recombination; however, it is the major pathway. The essential features of general recombination in bacterial cells, determined from in vitro

studies with purified recA protein and other components (eg. single strand binding protein), is the invasion of a duplex molecule by a single strand in a directed fashion, ie from 5' to 3'. This fits well with the mechanism of genetic transfer mediated by F factor DNA in which single stranded DNA from the donor is recombined with the duplex DNA of the recipient. The role played by recBC nuclease is not clear although the suggestion has been made that it promotes the formation of single stranded DNA, a requisite for recombination in vitro. An unwinding activity in the presence of ATP, double stranded DNA, and single stranded DNA binding protein has been assigned to recBC nuclease in vitro (Dressler and Potter, 1982).

Site Specific Recombination

The other bacterial recombination systems are specialized and fall into two main classes:

1) integration of transposable elements, and

2) site specific recombination.

I will not deal with transposable elements here in the interests of brevity, but refer the reader to an excellent collection of reviews in a book edited by J. Shapiro (1983). Site specific recombination has best been studied for the integration of bacteriophage lambda into the E. coli chromosome. However, a number of other site specific recombination systems have been found which share certain similarities with the lambda mechanism. These are the inversion of the G loop in bacteriophage Mu (gin) (and its equivalent in bacteriophage P1 - cin), the inversion of the H2 gene in Salmonella typhimurium (hin), and the cointegrate resolution mechanism of transposon Tn3, TnpR (Campbell,

1983; Silverman and Simon, 1983). An analogous site specific recombination system has also been described for the yeast plasmid 2 micron circle, in which a region of the circular plasmid DNA which is bounded by inverted repeats is "flipped" by the plasmid encoded FLP protein (Broach et al, 1982).

Bacteriophage lambda is a temperate phage which is capable of either lytic infection or lysogenic integration into the bacterial genome. The latter is accomplished by a site specific mechanism involving specific sites (frequently referred to as att sites) on both the bacterial and phage genomes. These are designated BOB' and POP', respectively. B, B',P and P' are all different from each other. The O sequence is a 15 bp common core region. Integration is independent of the bacterial recombination pathways which utilize recA, but is dependent on a phage protein encoded by the int gene. Int catalyzes two staggered double strand breaks and religations within the 15 bp core O region of BOB' and POP', generating reciprocal recombinant sequences BOP' (att L) and POB' (att R) (Nash, 1981). Little if any branch migration of a transitory Holliday structure occurs. This was shown to be the result of the concerted nature of int mediated recombination and not due to the lack of sequence homology between B, B', P, and P' (Echols and Green, 1979; Enquist et al, 1979).

The other bacterial site specific recombination systems share features in common with lambda int mediated recombination, mainly requirements for cis acting signals (the site of action) and trans acting components which interact with the target sites in a concerted fashion (Silverman and Simon, 1983). The trans-acting components from

the Salmonella hin and bacteriophage Mu gin systems were found to cross react with the respective target sequences, which share some sequence homology, indicating a possible evolutionary relationship. None of the other site specific systems showed any functional homology; however hin, gin, and InpR proteins share 36% amino acid identity, suggesting a common ancestry (Simon et al, 1980). All of these site specific recombination systems are independent of recA.

Bacteriophage Recombination Systems

A number of bacteriophage families utilize recombination mechanisms during their lytic growth cycles which are part of the phage genetic heritage. I will discuss briefly the mechanisms employed by phages lambda, T4, and T7.

All three of these phages form concatemers during the final stages of DNA replication from which unit length phage genomes are cleaved and packaged. The cleavage/packaging mechanisms will be discussed later; at present it is sufficient to say that phages lambda and T7 contain specific termini, while phage T4 contains non-specific, but redundant, termini, indicating some divergence in their mechanistic strategies. However, the recombination pathways in all three phages are involved in the formation of the packaging intermediates, the concatemers.

Phage lambda can utilize any one of three recombination pathways to form the requisite packaging intermediates. These are the recA pathways of the host cell, the phage site specific int mediated pathway, and a third phage system designated red. If all three pathways are

blocked, the phage is unable to package its DNA, and "dies".

The red system, in conjunction with other phage gene products, notably the gam gene product, is involved in the transition of DNA replication from the theta to the sigma, or rolling circle, mode. Lambda DNA is a linear duplex with single stranded complementary ends which ligate together after infection to form a circular duplex. The circular lambda genome replicates first in the theta mode, and then switches to the sigma mode. Lambda red components include the products of two genes - exo and beta (Lewin, 1977; Kornberg, 1974). The beta protein has an unknown function, but the exo protein is a 5'-3' exonuclease. Stahl (1979) envisions the theta to sigma switch as being the result of branch migration of a Holliday structure between two lambda circles across a nick in one strand, thus creating a replication fork. Chain extension is facilitated by continual exposure, mediated by the red exonuclease, of single stranded template at the opposite end which is subsequently hybridized to the displaced strand. The latter rolls around as it is displaced by DNA synthesis on one side and hybridized to exposed template on the other. Red exonuclease falls off as soon as a nick (or break) in the template strand is encountered. The end result is a sigma structure. Still to be explained by this mechanism is the formation of the initial Holliday structure in terms of recombination, although as a replication intermediate, incompletely replicated theta structures may be the source.

Bacteriophage T7 possesses a linear duplex genome which is bounded by terminal repeats of approximately 160 bp. Replication is initiated from a preferred origin at approximately 17% from the left

end of the genome and proceeds in a bidirectional manner. Linear unit length molecules then become incorporated into branched concatemers by one or more recombination mechanisms which require the activities of four phage gene products - gene 3 (endonuclease), gene 4 (an RNA polymerase associated protein), gene 5 (DNA polymerase), and gene 6 (exonuclease) (reviewed in Kornberg, 1974; Hausmann, 1976; Powling and Knippers, 1976; Lewin, 1977; Roeder and Sadowski, 1978; Kruger and Schroeder, 1981). Roeder and Sadowski (1978) developed an in vitro recombination system to test the activities of various combinations of the four gene products and concluded that gene 3 and gene 6 recombine T7 DNA by separate, but not mutually exclusive pathways which they termed the "endonuclease" and "exonuclease" pathways. The molecular mechanisms by which these pathways operate is not yet clear. However, the product of the reactions, the concatemer of phage DNA, is known to be an obligatory precursor for packaging of phage DNA into viral capsids, and the maturation of unit length phage DNA from concatemers appears to be coupled with packaging. Langman et al (1978) demonstrated that concatemers contain unit length genomes arranged head-to-tail and separated by a single copy of the terminal repeat. Watson (1972) had proposed that incomplete replication of the termini, coupled with a 5' exonuclease activity would expose complementary 3'ends which could ligate in a head-to-tail fashion to produce concatemers from which unit length genomes could be cleaved. He saw the necessity for such a mechanism in order to account for the preservation of T7 termini, which would gradually be lost if the phage replicated its unit length genomes by a semiconservative scheme. He proposed further that the terminal

repeat, lost during formation of the concatemer, would be regenerated during cleavage of the precursor by staggered nicks on either side of the repeat followed by repair synthesis in both directions across the repeat. The first part of his proposal now seems unlikely for several reasons (no circles have ever been found, for example); however, while there is no evidence to support or deny the second part of the proposal, it remains the only explanation to be offered so far.

Bacteriophage T4 also possesses a linear genome with a terminal redundancy. However, the latter can be derived from any sequence in the genetic map and is substantially longer than that in T7. T4 thus has a circularly permuted genetic map and is diploid for the two or three genes encoded by the terminal repeat (reviewed in Kornberg, 1974; Broker and Doermann, 1975; Lewin, 1977; Stahl, 1979; Luder and Mosig, 1982). This is accomplished by head full packaging of a greater than unit length genome into the phage capsid from a head-to-tail concatemer. Packaging may begin at any point in the genome, hence the circular permutation of the genetic map. Since concatemers are an essential precursor for packaging, the mechanism which generates the concatemer, recombination, is also essential. The phage encodes its own recombination machinery, which includes several nucleases (among which is endonuclease VII, an enzyme which cleaves Holliday structures in vitro - Mizuuchi et al, 1982), DNA binding protein, DNA polymerase, and DNA ligase. Luder and Mosig (1982) demonstrated that T4 DNA synthesis goes through two phases. The first phase requires host RNA polymerase and coincides with replication of unit length linear genomes, which, like T7, employs one or more preferred sites or origins located close

to or at strong promoters of RNA synthesis. The second phase, coincident with the appearance of concatemers, is independent of RNA polymerase but dependent on recombination. These authors proposed that recombination between phage termini, which might by this time be single stranded from incomplete replication of 5' ends or nuclease digestion or both, and the homologous sequence in another genome would create a replication fork with a 3'-OH DNA strand as a primer. The genetic data are entirely consistent with this proposal, as are the highly branched concatemers of intracellular T4 DNA. Note that this proposal requires that the invading strand be a 3' end. Rec A mediated recombination depends on the invasion of a 5' end followed by 5' to 3' branch migration of the heteroduplex. T4 recombination appears to utilize similar principles but different operational mechanisms.

Special Recombination Mechanisms

Recombinational systems which are not fully understood as yet, or which display features different from the classical homologous or site specific mechanisms described above, are frequently classed as "special". Falling into the former category are the phenomena of immunoglobulin gene rearrangement (Alt, et al, 1984; Alt and Baltimore, 1982), trypanosome antigenic variation (Borst and Cross, 1982), and transposable element movement (Toussaint and Resibois, 1983; Iida et al, 1983; Heffron, 1983; Kleckner, 1983; Roeder and Fink, 1983; Varmus, 1983). Immense volumes have been written on all three topics, and I do not propose to add to the list here. The best (and only?) example of the latter category is mating type conversion in the yeast

Saccharomyces cerevisiae.

Yeast can grow vegetatively in either haploid or diploid states. The latter is possible as a result, usually, of perturbations in the mating and sporulation mechanisms. Mating of haploid cells can only occur between two cells of opposite mating type. These have been designated a and α, and two different alleles of a locus called MAT determine the a or α phenotype. The exact phenotypic consequences of the a and α genotypes are not clearly understood at present, although there is some evidence that the two alleles encode pleiotropic regulatory proteins that turn on and off batteries of genes involved in mating and sporulation (Strathern et al, 1981; Haber, 1983).

Heterothallic strains of yeast stably retain their mating type phenotype, and their haploidy, until they are mated with cells of the opposite mating type. At this point, the a/α diploid progeny become incapable of further mating, but may be induced to sporulate under certain defined conditions (usually starvation). Individual spores from a mating between two heterothallic strains will themselves be heterothallic. The stability of heterothallic strains was found to result from a recessive trait at an unlinked locus, ho. Homothallic strains, which contain the dominant allele, HO, at that locus are frequently diploid as a result of constant mating between haploid cells of opposite mating type, which arise at equal frequencies within a single population. This was found to be the result of mating type switching at the MAT locus from a to α and vice versa (reviewed in Haber, 1983).

Intense analysis of the molecular underpinnings of this phenomenon revealed that the genetic information at MAT is duplicated

at two other "silent" loci on the same chromosome (III) - HML and HMR. Usually, a information resides at HMR and α information at HML. The switch involves transposition of the information at HMR or HML into MAT, replacing whatever sequences resided there before the switch.

All three loci were found to share homologous sequences at the borders of the a or α coding sequences, which are different. Thus, in a MATa haploid, HMLα contains sequences W-X-Yα-Z1-Z2, MATa contains W-X-Ya-Z1-Z2, and HMRa contains X-Ya-Z1. The product of the HO gene is a specific endonuclease which makes a double stranded staggered cleavage in Z1 at MAT, but not at HMR or HML.

The current model of mating type switching proposes that the double-stranded cleavage at MAT initiates the switch by creating ends composed of Z1 sequences which are capable of invading their homologues at either HMR or HML. Repair synthesis across HM will displace a non homologous strand until X sequences are reached. The newly displaced X sequences may then invade their homologues at MAT and serve as templates for additional repair synthesis in the opposite direction. The Holliday structures in Z1 and X would then be resolved in such a way that recombination of flanking sequences is avoided. This constraint must be part of the concerted nature of the mating type switch since Klar and Strathern (1984) showed that switching between unlinked HM and MAT loci rarely results in recombination of flanking markers. Another constraint on the model is the unidirectional flow of information from HM to MAT. Obviously there is some mechanism which distinguishes the displaced strand at HM from the replaced strands at MAT. Perhaps both strands at MAT are degraded while the replacement strand from HM is

being displaced.

Herpes Simplex Virus

General

The rather long preamble above on recombination mechanisms sets the stage for understanding some of the phenomena associated with the behaviour of Herpes Simplex Virus DNA, the subject of this thesis.

The Herpesviruses comprise a large family of DNA viruses which infect eukaryotes and cause a wide variety of pathogenic diseases (reviewed in Honess, 1984). These have been classified under three broad categories determined mainly by their tissue specific pathogenicity. Thus, alphaherpesviruses (eg. Herpes Simplex Virus) infect epithelial cells acutely, and are capable of latent infection of underlying nervous tissue, allowing periodic reactivation of lytic infections in the innervated epithelial tissue. Betaherpesviruses (eg. Cytomegalovirus) cause persistent infections of a wide variety of tissues, including salivary glands and kidneys, and while they are usually asymptomatic in adults, they may cause generalized viremia in newborns. Alphaherpesviruses may also cause persistent infections in adults and viremia in newborns. Gammaherpesviruses (eg. Epstein-Barr Virus) infect lymphatic tissues, and have been associated with lymphoproliferative diseases such as Burkitt's Lymphoma.

The most intensively studied of the Herpesviruses is Herpes Simplex Virus, whose natural host is humans. The virion consists of a dense nucleoprotein core contained within an icosahedral nucleocapsid

of 100nm diameter, which is itself surrounded by a protein "tegument" and a membrane envelope. The nucleocapsids are assembled in the nucleus, and enveloped on their way out of the nucleus.

The viral replication cycle is similar to those of many phage and eukaryotic viruses in that the viral genes, of which there are more than 50, are classed according to their temporally regulated modes of expression. As soon as the viral genome is released in the nucleus, five genes classed "immediate early" are transcribed and translated. One of these, ICP4, has been shown to be essential for the induction of several of the second class of viral genes, the "delayed early" genes. Among this latter group is the gene encoding the viral thymidine kinase gene, one of several delayed early viral genes involved in DNA metabolism and synthesis. The third class of viral genes are the "late" genes, many of which encode structural components of the nucleocapsid and membrane glycoproteins. DNA synthesis is required for the enhanced expression of these genes, suggesting that the large quantities produced may be a function of gene amplification. Details of the control mechanisms involved in the switches from immediate early to delayed early to late gene expression are currently the subjects of exciting research in many labs.

Structure of HSV DNA

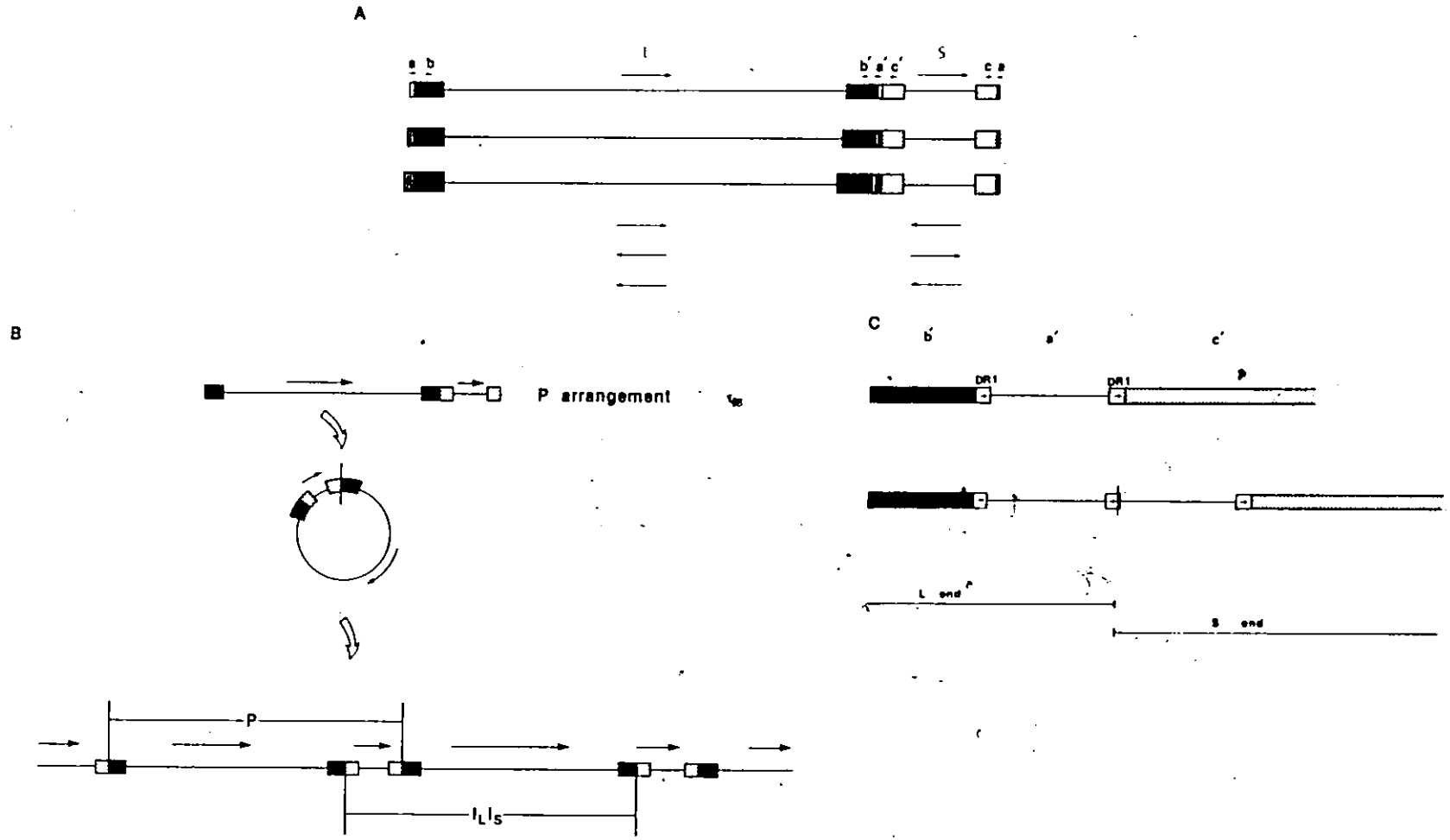
The HSV genome is a linear duplex of approximately 100×10^6 daltons, or 150 -160 kbp, arranged in an unusual manner (Fig. 1). Two segments of unique sequences, designated U_L and U_S (for Long and Short), are each bracketed by substantial inverted repeats. The repeats

FIGURE 1

(A) Structure of HSV DNA. The figure shows the locations and orientations of repeated sequences in HSV DNA, and diagrams the heterogeneity arising from variable numbers of tandemly reiterated copies of the "a" sequence at L-S junctions and L termini. The long arrows below the figure represent the three additional relative orientations of the L and S components of viral DNA. Solid bars represent "b" repeats and stippled bars represent "c" repeats.

(B) HSV-1 replicates as a concatemer. The figure outlines the current model for the replication of HSV DNA. Following infection, end-to-end ligation of the L and S termini gives rise to a circle, which then generates head-to-tail concatemers, perhaps by rolling circle replication. Unit length genomes are then cleaved out of the concatemer by cleavage at L-S junctions. Note that alternative packaging phases can produce two of the four isomers, in this case P and I_{LS}. A concatemer of either I_L or I_S genomes could produce the other two isomers in this same fashion.

(C) Tandemly reiterated "a" sequences share a single copy of DR1. The single "a" sequence present on some L-S junctions is flanked by a 20 bp repeat, DR1, whereas the tandemly reiterated "a" sequences present on other L-S junctions share a single copy of DR1. The lower portion of the figure compares the structure of junctions and termini, showing that the termini appear to arise by a single cleavage event within a copy of DR1 separating two adjacent "a" sequences.


A
a b
L
b' a' c'
S
c a
B
P arrangement
P
$I_L I_S$
C
b'
a'
c'
DR1
DR1
P
L end
S end

flanking U_L, designated "b" and "b'", are approximately 8,000 bp long, while the repeats flanking U_S, designated "c" and "c'", are approximately 6,500 bp long. The gene for ICP4 is wholly encoded within the "c" sequences. The termini of HSV DNA contain a short (250-550 bp) direct repeat designated "a". This same sequence also flanks each segment as an inverted repeat. The genome may thus be represented as follows: ab-U_L-b'a'c'-U_S-ca.

The structure of HSV DNA was first intimated by Sheldrick and Berthelot (1974), who found both circularized and bar-bell shaped single stranded DNA molecules in electron micrographs of denatured HSV DNA. One of the loops in the latter structure was much larger than the other. These authors concluded that the genome was composed of two unique sequences, L (long) and S (short), each of which was bracketed by its own set of inverted repeats, and that the genome was terminally redundant. They then proposed that inversion of the gene sequences in the two unique segments, L and S, could result from an odd number of recombinational events between the terminal and internal inverted repeats.

Restriction endonuclease digestion of HSV DNA revealed the presence of several submolar fragments (Wilkie et al, 1974; Skare et al, 1975; Hayward et al, 1975). Subsequent restriction mapping of the viral genome indicated that the submolar fragments could be accounted for by heterogeneous arrangements of HSV DNA similar to those proposed by Sheldrick and Berthelot (1974) (Wilkie, 1976; Skare and Summers, 1977). These results, and those obtained from partial denaturation maps, supported the idea that the genome existed in four equimolar

isomeric arrangements (Wadsworth et al, 1975,1976; Delius and Clements, 1976). It is now well established that the genome has the structure represented in Fig. 1, and that the two segments, L and S, seem to invert freely with respect to each other to generate the four arrangements - P, I_L (inverted L), I_S (inverted S), and I_{LS} (inverted L and S).

The terminal redundancy (referred to as the "a" sequence) has been the subject of close scrutiny for several years. Exonuclease digestion of the termini exposed an approximately 800 bp sequence capable of forming circular viral DNA upon reannealing (Grafstrom et al, 1974; Wadsworth et al, 1976). Since many of Sheldrick and Berthelot's bar-bells contained no single stranded tails (some did), the best interpretation was that the "a" sequence is repeated at the junction between L and S in inverted form. This has indeed turned out to be the case. Fine structure mapping of the junction between L and S revealed that the termini overlap completely with the junction (Wagner and Summers, 1978). These authors also found that the most abundant L-S junctions contained only one copy of "a" (which is 280 bp long in the strain they analysed). However, they also discovered that fragments from the L termini and the L-S junction consisted of families of fragments which differed in size by multiples of 280 bp, the size of the "a" sequence. The pattern displayed by restriction fragments bearing these iterations was highly stereotyped, such that each succeeding fragment within a family or stepladder, in ascending size, was less abundant than the preceding fragment. These were interpreted as tandem iterations of "a" which accumulated at the L-S junction and

at the L terminus, but not at the S terminus. Detailed sequence analysis of L-S junctions has since confirmed this idea. Moreover, it has revealed that "a" itself is flanked by a 20 bp direct repeat which is present only once between tandem copies (see Fig. 1) (Davison and Wilkie, 1981; Mocarski and Roizman, 1982a).

The three mysterious features of HSV DNA, namely segment inversion, tandem iteration of "a", and the asymmetric distribution of "a" to one terminus and the junction, may result from a concerted mechanism involved with HSV DNA replication rather than stochastic fluctuations in the population. This is evident from 1) the non random pattern of "a" sequence amplification, 2) the asymmetric distribution of iterated "a" sequences to the L terminus and the L-S junction, but not the S terminus, and 3) the invariance of these patterns from one plaque isolate to another. That is, the four isomers will be equimolar in amount, and the distribution of tandem iterations of the "a" sequence will be identical from stock to stock. Thus, any explanation of HSV DNA replication and processing must account for these observations. That they are consequences of a replication mechanism is given somewhat oblique support by the observation that many Herpesviruses, even distantly related ones, share some of these features (reviewed in Roizman, 1979; Honess, 1984).

The mechanism of HSV DNA replication is not fully understood as yet, but there are some tantalizing clues. Large "tangles" of viral DNA were observed in cells infected with HSV (Hirsch et al, 1977; Jacob and Roizman, 1977), and Pseudorabies virus (Ben-Porat, et al, 1976). These were believed to be concatemeric intermediates of replication. This

idea was reinforced by restriction endonuclease analysis of replicating viral DNA; both pulse labelled and unlabelled viral DNA from the nuclei of infected cells contained a much larger proportion of L-S junction fragments than L or S terminal fragments (Hirsch et al, 1977; Roizman et al, 1978). This was interpreted as evidence that the viral DNA was probably arranged in head-to-tail concatemers. Further evidence confirming head-to-tail concatemers as the replicative intermediates came from studies of defective viral genomes. Undiluted passage of HSV frequently results in the accumulation of defective interfering particles in the population. These virions are incapable of independent growth, and require the continued presence of "helper", wild type virus (Frenkel et al, 1975). The DNA in defective viruses was found to be unit length tandem iterations of short, subgenomic regions of the wild type genome (Frenkel et al, 1976). These have been found to fall into two classes depending on the sequences from the wild type genome which are amplified. Type I defectives are derived completely from the S terminus, and contain most or all of the "ca" inverted repeat, and may contain some unique sequences from the S segment. Type II defectives contain a small portion of the S terminus covalently linked to unique sequences from L at map coordinates 0.35-0.42 (Locker and Frenkel, 1979; Kaerner et al, 1979). Transfection of cells with monomer DNA from naturally occurring defectives, together with helper viral DNA, has been shown to result in production of defective genomes consisting of head-to-tail concatemers of the input monomer units (Spaete and Frenkel, 1982, Barnett, et al, 1983).

The problem associated with replicating linear molecules is the

faithful generation of sequences at the ends within the confines of the known biochemistry of DNA synthesis, specifically, the need for a 3'-OH primer to which nucleotides may be polymerized. Phages T7 and T4 have overcome this problem by utilizing intermolecular recombination between terminally redundant genomes, while bacteriophage lambda circularizes its genome upon entry into the cell and replicates it as a circle, either in the theta mode or as a rolling circle. In all three cases, the end result is a concatemer from which unit length genomes are cleaved and packaged. HSV has adopted a similar strategy to replicate its ends. Not all viruses with linear genomes utilize this sort of pathway. Adenovirus has evolved a very different scheme in which a viral protein provides the missing primer function at the 5' ends (Tooze, 1980). The major question still to be resolved is how HSV generates the concatemer. If the genome ligates intramolecularly to create a circle, the large tangles must be generated by an amplification mechanism similar to the rolling circle of lambda phage. Intermolecular ligation or recombination of linear molecules, or intermolecular recombination of circular molecules, would be sufficient to produce the tangles observed. However, there is some genetic evidence which refutes the idea that linear molecules ligate together, or undergo intermolecular recombination between termini. Honess et al (1980) found that some markers which map in the U_S or "c" regions behave as if they are linked to markers in U_L. Frequent intermolecular interactions would be expected to generate such recombinants at a frequency close to 50%, since the markers would be essentially unlinked. Roizman et al (1978) also found evidence of linkage between

markers in the L and S repeats, although the frequency of (genetic) recombination was higher than would be expected from the molecular distance. The genetic data described in these studies is not inconsistent with a lower frequency of intermolecular ligation, or compartmentalization of input DNA, or both, operating on HSV DNA.

Whatever mechanism HSV utilizes to generate the concatemeric precursor of viral DNA, it is evident from the fact that the termini are specific that a specific cleavage event cuts the unit length genomes from the concatemer. Analysis of defective genomes provides the best evidence that the cleavage signal must reside within terminal sequences. Both type I and type II defectives share the last (approximately) 500 bp of the S terminus, comprising part or all of the "a" sequence (Locker and Frenkel, 1979; Kaerner et al, 1979). Moreover, one terminus from the defective genome ends in this sequence, ie the integrity of the S terminal sequences is maintained. Recent experiments have indicated that plasmids containing HSV sequences from either the S or the L termini can be propogated as defective viral DNA, confirming the idea that the "a" sequence contains the cleavage signal (Stow et al, 1983). Engineered defectives also require the presence of HSV sequences containing a viral origin of DNA replication. These can be derived from the "c" region (Stow and McMonagle, 1983) or the U_L region (Vlazny and Frenkel, 1982), suggesting that natural defectives arise by illegitimate amplification of HSV sequences containing the two signals - type I defectives contain the origin and cleavage signals from the "ca" sequences, and type II defectives contain the origin in U_L covalently attached to the cleavage signal from the S terminus.

Since defective genomes contain iterations of the monomer with the two cis requirements, the virus must have a mechanism which distinguishes the internal cleavage signals from the ones that are utilized in the generation of unit length genomes. The simplest explanation is a spatial recognition of a specific length of DNA coupled with a chemical recognition of the specific cleavage signal. The same arguments apply to wild type viral DNA. Concatemers of full length viral DNA consist of consecutive L and S segments separated by inverted junctions which contain the cleavage signal (Fig. 1). The inverted orientation of the junctions in the concatemer introduce a biphasic organization of unit length genomes within the concatemer. If both phases are used, then concatemers of one isomer can produce equal proportions of genomes of two isomers; concatemers of P isomers will produce both P and I_{LS} isomers, and concatemers of I_L or I_S isomers will produce both of the latter (see Fig. 1). Some other mechanism, such as intramolecular recombination, must be invoked to derive I_L or I_S concatemers from P or I_{LS} concatemers. Thus, a simple maturation scheme would have one cleavage event at any junction followed by a measuring mechanism such as head full packaging, ending with a second cleavage at the next closest junction. While this strategy accounts for the defective data, without invoking additional events, it fails to account for the three invariant features of HSV DNA discussed above, namely production of four isomers from a single pfu, "a" sequence amplification, and asymmetric distribution of amplified "a" sequences.

In spite of the obvious intellectual challenges, few

investigators have proposed any models of HSV DNA replication and cleavage. Only two models have been formulated, both by Roizman and co-workers. In the first model, Roizman et al (1978) proposed that HSV DNA circularized by annealing of complementary strands of the terminal "a" sequence which were exposed by exonuclease digestion. The junction thus formed, bearing a single "a" sequence, was referred to as a "modified" junction, and was postulated to serve as the cleavage site for unit length genomes within the concatemer. Thus, the concatemer had two kinds of junctions, those with two tandem "a" sequences and those with one "a" sequence, arranged consecutively, and inverted with respect to one another. The cleavage signals at the junctions were distinguishable such that subgenomic L or S segments would not be cleaved out of the concatemer. This was a necessary constraint since the proposal demanded that cleavage occur prior to, and separately from, encapsidation. In the original proposal, the junction with the single "a" sequence was chosen as the cleavage site, and cleavage was deemed to occur on one side or the other of the "a" sequence in a random fashion. This generated a terminus, either L or S, lacking an "a" sequence. To account for the observation that both termini contained "a" sequences, the authors proposed that the "a"-less terminus underwent a strand exchange with an inverted repeat from the L-S junction which had been nicked between the adjacent "a" sequences. The 3' end of the "a"-less terminus used the nicked strand as a template for repair synthesis, thus displacing it from the junction. The 3' end of the nicked junction strand primed repair synthesis back across the gap left by the displaced strand. In the process, the Holliday

structure thus formed could be resolved to generate inverted segments, and it could also migrate into flanking sequences to initiate repair of mismatched heteroduplexes.

The model attempted to explain HSV DNA cleavage within the constraints of 1) a single isomer only being able to replicate, 2) obligatory identity of the inverted repeats, and 3) the observed concatemers of both viral DNA in infected cells and the concatemeric genomes of defective interfering particles. (Roizman et al, 1978 proposed that the concatemer arose by a rolling circle replication mechanism). However, it is now clear that the first two constraints were the result of faulty interpretation of the following experiments.

HSV types 1 and 2 differ from each other by approximately 50%. in terms of DNA sequence. The arrangement of genes, however, is colinear, allowing the ready formation of viable intertypic recombinants (Esparaza et al, 1976). Morse et al (1977) prepared a series of such recombinant viruses and examined their DNAs by restriction analysis in order to map the cross-over sites. Most (21/28) exhibited multiple crossovers. Morse et al (1977) reasoned that single crossover events between three of the four isomers would appear as two crossovers when the recombinant is displayed in the fourth isomeric arrangement, and should therefore seem to be rare. The same reasoning applied to any recombinant which displayed an odd number of crossovers. (In fact, this logic only applies to two out of four arrangements, in which the segment containing the crossover remains fixed.) Since they were apparently able to generate 7 recombinant maps with single crossovers, 1 with three crossovers, and 2 with five crossovers, they concluded that the

arrangement in which they chose to display their maps must reflect positive selection by the virus for that particular arrangement. This was interpreted to mean that only the P arrangement was capable of replication.

There are several defects with their logic. First, the fact that multiple recombination events were the norm removed the necessity to minimize the number of crossovers experienced by interacting molecules. Second, recombination within the inverted repeats does not distinguish any isomer as the starting material. Two of the molecules with single crossovers, the only one with three crossovers, and one of the two molecules with five crossovers all had undergone recombination within the inverted repeats. In a similar series of experiments Wilkie et al (1978) showed that intertypic recombinants which had apparently resulted from single crossovers between P isomers in fact had undergone multiple crossovers, at least one of which occurred in the inverted repeats. Roizman et al (1978) realised that their model was incapable of explaining the generation of more than three of the four isomers unless they also allowed genomes in the I_S arrangement to replicate in subsequent rounds. This was deemed acceptable since five of their single crossover recombinants could have been derived from genomes in either P or I_S.

The second source of confusion with regard to the proposed model was the assertion that the inverted repeats flanking L and S were "obligatorily" identical. This conclusion is based on marker rescue experiments performed with temperature sensitive mutants of ICP4, the gene encoding an essential immediate early protein. ICP4 is wholly

encoded by the inverted repeats flanking the S segment, and is therefore diploid. Nevertheless, mutants of ICP4, which of necessity contain two mutant copies of the gene, are relatively easy to obtain (Schaffer et al, 1978). Knipe et al (1979) used restriction fragments from HSV-2 to "rescue" HSV-1 mutants by their ability to grow at the restrictive temperature. The ICP4 proteins from HSV-1 and HSV-2 are distinguishable on SDS-gels, as are the restriction fragments encoding their respective genes. Thus, rescued viruses were analysed both for the restriction patterns of their DNAs and for the migration patterns of their ICP4 proteins. Of ten such viruses analysed, 8 displayed patterns indicative of heterodiploids, while 2 were homodiploids. The latter where found to be insertions of HSV-2 DNA close to the terminal "a" sequence, while the others were insertions mapping further away from the "a" sequence. These data were interpreted to mean that the "a" sequence and part of the "c" sequence, and by analogy, the "a" sequence and part of the "b" sequence from the other end (see Fig. 1), were obligatorily identical, as a consequence of repair synthesis to generate the missing "a" sequence coupled with branch migration into the inverted repeats and repair of heteroduplexes thus formed. This assertion was difficult to reconcile with the observation from the same lab (Morse et al, 1977) that intertypic recombinants with heterologous termini (and therefore different "a" sequences) were perfectly viable.

Apart from the above criticisms, the major drawback of this model is its inability to explain cleavage of unit length defective genomes. All of the "junctions" in defectives are identical, at least at a gross level. Uncoupled cleavage and packaging would be expected to

generate monomer units in abundance, and few if any unit length genomes. Furthermore, no attempt was made to explain the tandem accumulation of "a" sequences or their asymmetric distribution to the L terminus and the joint, but not the S terminus. This feature is as important as segment inversion in terms of the dynamic fashion with which it is reproduced.

Some recent advancements of our knowledge have led to virtual abandonment of the model described above. Two groups demonstrated that segment inversion could be induced to occur in engineered viruses containing junction spanning fragments inserted at the viral thymidine kinase locus (Mocarski et al, 1980; Smiley et al, 1981). The inserted sequences served as focal points for inversion of segments between the insert and legitimate junctions when the two sequences were in inverted orientation with respect to each other. Smiley et al (1981) also noted that intervening segments could be deleted when the two sequences were in direct orientation. Mocarski et al (1980) tested repeats of non junction sequences as focal points for segment inversion and found them to be inert, although Pogue-Geile et al (1985) later found that a different non-junction fragment was quite active. Subsequent experiments showed that sequences from either terminus were suitable targets for segment inversion in the same assay (Mocarski and Roizman, 1982a). Moreover, these same authors claimed that a thymidine kinase gene flanked by two inverted junction fragments, when integrated into cellular DNA, could be induced to undergo segment inversion between the two junction fragments after infection of the cells with HSV (Mocarski and Roizman, 1982b). This was interpreted as evidence of a trans acting

substance induced by the viral infection interacting with a cis acting target sequence in the junction fragment. Since 1) both termini served as targets in the segment inversion of viral DNA, 2) the termini shared a common sequence, "a", and 3) non "a" containing sequences were inert, the "a" sequence was deemed to be the target of a site specific recombination mechanism.

Before looking at the new model, it would be expedient to bury the old one. How do the new data fit into the old model? The answer is "not at all" by several criteria:

1) Mocarski et al (1980) inserted a junction containing two adjacent "a" sequences into the viral TK gene, but Smiley et al (1981) inserted a junction with one "a" sequence. Unless the circularization process tags the "modified" junction in a cryptic fashion, Smiley et al's insert would be expected to serve as an internal cleavage site. This would generate concatemers with cleavage sites at less than genome length intervals, and would be expected to be lethal. If, on the other hand, the "modified" junction is distinguishable from the insert, then, if the inversions are a consequence of repair of "a"-less termini from junctions with two adjacent "a" sequences, as demanded by Roizman et al's (1978) model, Smiley et al's (1981) construction should have been an unsuitable substrate and should have remained inert, which it clearly did not.

2) The assertion that only genomes in the P or I_S orientation are capable of replication predicts that of the two constructions made by Mocarski et al (1980), one would produce 8 isomers, the four original ones and four new ones, and the other would only produce the four

original isomers and no new ones. These are the isomers produced by genomes in which the insert, in the P isomer, is inverted with respect to the termini. In fact, both constructions produced 12 isomers, some of which must have arisen by inversion of DNA in the I_L orientation, indicating no restrictions on replication of any isomers of HSV DNA.

3) Genome isomerization was deemed to be the result of post cleavage processing of HSV DNA. Yet, Smiley et al (1981) found that direct repeats of junction sequences promoted the production of deletions of intervening sequences. These authors found that the deleted segments were packaged as unit length dimers, indicating that the deletion was part of a concatemer. The only way to reconcile this result with Roizman et al's (1978) model is to demand repeated rounds of cleavage and replication. However, there is good genetic evidence from studies on another Herpesvirus, Pseudorabies virus, that cleavage is intimately coupled with packaging (Ladin et al, 1980). These authors found mutations which fell into 9 different complementation groups in which both cleavage of concatemers and packaging of viral DNA were deficient.

4) Mocarski et al (1982a) had determined that both termini were capable of generating the full complement of new inversion products. Since, for example, L terminal sequences lack "c" sequences, S termini without an "a" sequence, generated by cleavage between "c" and "a" at a modified junction, would be incapable of forming a heteroduplex with the L terminal insert. Thus, each of the terminal inserts would be expected to generate only a subset of the inversion fragments. This was not the case. Both termini were equally adept at generating the full complement of inversion products.

Mocarski et al (1982a) undertook to sequence the termini from both cloned and viral DNA. These results were illuminating in that the termini appeared to be the products of a single cleavage event within a junction containing two "a" sequences. Other investigators had determined the sequence of several strains of HSV types 1 and 2 (Davison and Wilkie, 1981) and had found that "a" was itself flanked by direct repeats of a 20 bp sequence, now generally referred to as DR1. Mocarski et al (1982a) found that:

1) tandem copies of "a" were separated by a single copy of DR1, and

2) termini appeared to be generated by a specific cleavage within DR1 such that direct apposition of L and S termini regenerated a complete copy of DR1 and a junction with two tandem "a" sequences.

The revised model incorporates these new observations. Cleavage of concatemers is still believed to occur by preferential selection of some junctions over others, except that the cleavage junctions must now contain two adjacent "a" sequences, which are cleaved by a double strand break between the two "a"s in the DR1 separating them. The cleaved DNA is rolled into the viral capsid by a headfull mechanism until another junction containing two "a" sequences is encountered, at which time cleavage between "a"s is repeated. Segment inversion is accomplished by a site specific recombination mechanism at an earlier time during DNA replication, and genome circularization is accomplished by ligation of the termini rather than the original exonuclease exposure of single strands followed by annealing.

This model is clearly superior to its predecesor. Nevertheless, there are serious flaws. The most glaring is its inability to explain

the stereotypic amplification of "a" sequences. Mocarski and Roizman (1982a) proposed that unequal intermolecular recombination between the flanking DRIs might be responsible; however, this would not be expected to produce the stereotypic pattern, and might be expected to produce termini or joints lacking "a"s if recombination were intramolecular. Another is the unproven prediction that the most common junction should be one with two tandem "a" sequences. In fact, the most common (50%) junction has a single "a" sequence. Third, while site specific recombination explains segment inversion admirably, it is not the only possible interpretation of the data, and is flawed, at least in a semantic sense, in that site specific recombination has generally not been associated with heteroduplex exchange. Yet clearly, heteroduplexes do occur at least some of the time as evidenced by the appearance of homozygous recombinants in the diploid regions of the HSV genome. Even special recombination, similar to mating-type switching in yeast, seems inadequate in that it is a concerted mechanism, while the exchange events in HSV seem too random. Finally, the proposed mechanism fails to explain how defective genomes are packaged. These have been shown to contain tandem iterations of S termini, which only contain a single "a" sequence. So far, no alteration in the number of "a" sequences within a single repeat unit have been demonstrated. How then does it get packaged if two tandem "a" sequences (at a minimum) are required?

Several issues remained unresolved at the time that this thesis was begun, namely:

1) Obligatory identity of the inverted repeats. While the genetic data upon which the original contention was based was itself insupportive, a

clear demonstration of non-obligatory identity was deemed necessary.

2) Amplification and asymmetric distribution of the "a" sequence needed a rational explanation. Insight into the mechanism which generated this unusual feature would probably provide important clues to the cleavage and/or replication strategy employed by HSV.

3) What constitutes a viable cleavage signal? Does it indeed require the presence of two adjacent "a" sequences? Can the signal sequence be confined to a smaller segment?

4) What is the relationship between cleavage and site specific recombination? What is the relationship between site specific recombination and the formation of HSV DNA concatemers?

The object of this thesis was to examine some of these issues in greater detail. More specifically, the necessity for sequence identity of the inverted repeats, including the "a" sequence, was tested. The junction region from HSV-1 (KOS) was examined closely for biological activities which could be localized to specific sub-fragments, and correlated with each other. These experiments provided helpful insights into the amplification of the "a" sequence, and into the mechanism by which HSV packages its DNA, and allowed me to formulate a model for HSV cleavage/packaging which more closely agrees with all of the data described above.

MATERIALS AND METHODS

Bacterial Cells

Suspension cultures of bacteria were grown by gentle agitation at 37°C in Luria Broth, a medium consisting of 1% (W/v) bactotryptone (Difco), 0.5% (W/v) yeast extract (Difco), 85.5 mM NaCl, 10 mM Tris HCl pH 7.6, and 0.4% (W/v) glucose. Where appropriate, ampicillin (20 μg/ml) (Sigma) or tetracycline (10 μg/ml) (Sigma) were added.

Bacterial colonies were grown at 37°C on solid media consisting of Luria Broth supplemented with 1.5% (W/v) bacto-agar (Difco) and, if appropriate, ampicillin or tetracycline. Some strains of bacteria (eg JM109) were grown on 2YT plates which contained 1.6% (W/v) bacto-tryptone, 1% (W/v) yeast extract, 85.5 mM NaCl, 1.5% (W/v) bacto-agar, 0.004% (W/v) x-gal (5-Bromo-4-chloro-3-indolyl β -D-galacto-pyranoside) (Sigma), and 0.04mM IPTG (isopropyl -D-arabinofuranoside) (Sigma).

Amplification of plasmid DNA (see below) was accomplished by growing bacteria in M9 medium consisting of 42.26 mM Na_2HPO_4, 22.14 mM KH_2PO_4, 8.55 mM NaCl, 20 mM NH_4Cl, 0.01% (W/v) casamino acids (Difco), 1mM $MgSO_4$, .0.1 mM $CaCl_2$, 0.2% (W/v) glucose, 1 g/ml thiamine, and ampicillin or tetracycline where appropriate.

Plasmid Preparation

Saturated cultures of bacteria were diluted in M9 medium 1:50 and grown until the O.D. at 600 nm reached (approx.) 0.6, at which

time chloramphenicol (200 μg/ml) was added and the culture was further incubated overnight. Cells were pelleted by centrifugation at 5000 rpm in a Sorvall GSA rotor, and the pellets were resuspended in a solution containing 50mM TrisHCl pH 8 and 25% (W/v) sucrose. At 5 min. intervals, solutions containing lysozyme (1.5 mg/ml final conc.) (Sigma), EDTA (ethylenediaminetetraacetic acid) (50 mM final conc.), and Triton-X (0.04% V/v final conc.) were added, with gentle mixing. The lysate was centrifuged at 29,000 rpm in a Beckman 50Ti fixed angle rotor for 1 hour, and 0.9 g/ml CsCl were dissolved in the supernatant. Ethidium bromide (Sigma) was added (0.3% W/v) to the solution, which was placed in either 16x76 mm Beckman polyallomar tubes (approximately 7 ml per tube) or in 25x89 mm Beckman Quick seal tubes (approximately 25 ml per tube). Mineral oil was layered on top of the solution and the tubes were sealed. The small tubes were centrifuged at 35,000 rpm in a Beckman 50Ti fixed angle rotor for approximately 60 hours, while the quick seal tubes were centrifuged at 43,000 rpm for 20 hours in a Beckman VTi50 vertical rotor. Plasmid DNA was visualized with a hand held UV lamp (Mineralight) and removed by aspiration with a 20 gauge needle and syringe. The plasmid solution was extracted three times with equal volumes of 1-butanol and dialysed against 10mM TrisHCl, 1 mM EDTA, pH 7.4. Plasmid DNA was precipitated in 0.15M NaCl and 70% ethanol (V/v) at -70° C. The precipitate was collected by centrifugation in a Sorvall SS-34 rotor at 10,000 rpm for 30 minutes. The plasmid precipitate was air dried and resuspended in a small volume of 10mM TrisHCl, 1 mM EDTA, pH 7.4.

Rapid Extraction of Plasmid DNA

This procedure was first described by Birnboim and Doly (1979). Small (5 ml) cultures of bacteria were prepared as described above, and the cells were pelleted by centrifugation at 2,500 rpm in a Chilspin bench top centrifuge. The pellets were resuspended in 0.1 ml of a solution containing 2 mg/ml lysozyme, 50 mM glucose, 10 mM EDTA, and 25 mM TrisHCl pH 8, and placed on ice for 30 min. The suspension was lysed by the addition of 0.2 ml of a solution containing 0.2 N NaOH and 1% (W/v) SDS (sodium dodecyl sulphate), and then precipitated with 0.15 ml of 3 M sodium acetate pH 4.8 for 1 hour at 0° C. Samples were centrifuged at 10,000 rpm in an Eppendorf centriguge for 5 minutes and the supernatant was precipitated with the addition of 70% (V/v) ethanol at -20° C for 30 minutes. The plasmid DNA precipitate was collected by centrifugation in an Eppendorf centrifuge and washed twice by redissolving it in 0.1 M sodium acetate, 0.05 mM Tris HCl pH 8, and precipitating it with 70% ethanol. The washed precipitate was either air dried or centrifuged in a Speed-Vac lyophilizer, and redissolved in a small volume of 10 mM TrisHCl, 1 mM EDTA pH 7.4.

Transformation of Bacteria With Plasmid DNA

A saturated culture of bacteria was diluted 1:50 in Luria Broth and incubated until the O.D. at 600 nm was approximately 0.6. The cells were then pelleted by centrifugation at 2,500 rpm for 15 minutes in a Chilspin bench top centrifuge. The pellet was resuspended in ice cold 50 mM $CaCl_2$ (half the volume of the culture) and placed in an ice bath for 20 minutes. After centrifugation, the cell pellet was resuspended in 1/5 the volume of 50 mM $CaCl_2$. Small aliquots (0.2 - 0.3

ml) were incubated at 0°C for 45 minutes with 50 ng - 1 μg of plasmid DNA, and then gently agitated for two minutes at 42°C. The cells were spread evenly over a 100 mm plate (Canlab) containing solid medium and incubated overnight at 37°C.

Mammalian Cells

Monolayer cultures of African green monkey kidney cells (Vero cells) and mouse Ltk⁻ cells were maintained in α-modified minimal essential medium (α-MEM) (GIBCO) at 37° C. All cells were grown in Corning plastic tissue culture bottles or plates. Media were supplemented with either of two sets of additives : 1) 10% (v/v) fetal bovine serum (GIBCO), 0.03% (w/v) l-glutamine (GIBCO), 100 units/ml penicillin (GIBCO), 100 μg/ml streptomycin (GIBCO), 10 mM HEPES (N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid),and 0.075% (w/v) $NaHCO_3$; or 2) 5% (v/v) fetal bovine serum, 0.03% (w/v) l-glutamine, 100 units/ml penicillin, 100 μg/ml streptomycin, and 0.15% (w/v) $NaHCO_3$. Cells grown in media without HEPES, and cells grown in unsealed vessels required a humidified atmosphere enriched with 5% CO_2.

Virus Growth and Titration

Herpes Simplex Virus type 1, strain KOS, and various mutants of this strain, were used throughout this study. Occasionally, strain F and its derivative mutants were also used. Strain F was kindly supplied by Dr. B. Roizman.

Virus was passaged in Vero cells. Confluent monolayers were infected with varying multiplicities of virus in a small volume of α-MEM lacking fetal bovine serum for 2 hours, after which overlay

medium containing 5% fetal bovine serum was added and the infection was allowed to proceed until all cells displayed the characteristic cytopathic effect. Infected cells were dislodged by a sharp blow to the vessel and the cell suspension was centrifuged at 2,500 rpm for 5 minutes in a Chilspin benchtop centrifuge. The cell pellet was resuspended in a small volume of α-MEM lacking serum and the cells were disrupted by sonication in a water bath sonicator (Bransonic). Cell debris was removed by centrifugation and the virus suspension was stored in 1 ml aliquots at -70°C.

Virus was titrated on confluent monolayers of Vero cells laid down in 24 well Linbro plates. Serial $^{1}/_{10}$ dilutions of the virus stock were made in α-MEM lacking serum, and 0.2 ml of each dilution was added to each well. After a two hour infection, the monolayers were overlaid with α-MEM containing 5% fetal bovine serum and 0.05% ($^{v}/_{v}$) human immune serum (Connaught). Occasionally, 100 μg/ml of Ara-T (thymine-1-β-D-arabinofuranoside) (Sigma) would also be added to the overlay medium. Infection would be allowed to proceed for 2 to 3 days, at which time viral plaques would be visible.

Selection of Thymidine Kinase Deficient HSV Mutants

Thymidine Kinase deficient (TK^-) mutants of HSV-1 were selected on the basis of their survival in the presence of Ara-T, a thymidine analogue which causes chain termination during DNA synthesis. HSV TK efficiently phosphorylates Ara-T. 5000 plaque forming units (pfu) of virus were used to infect a monolayer of 10^7 cells in a 100 mm plate. After adsorption, overlay medium containing human immune serum and Ara-T was added to the plate and infection was allowed to proceed until

plaques became visible (approximately 2 days). The medium was aspirated and plaques were gently touched with a sterile wooden stick, which was then swished in the medium covering a confluent monolayer of Vero cells laid down in the well of a Linbro plate. Infection was allowed to proceed until all of the cells displayed the cytopathic effect. The medium from the well was harvested and stored at -70°C.

Extraction of Viral DNA

Confluent monolayers of Vero cells were infected at a mutiplicity of 2 to 10 pfu per cell. Infected cells were pelleted as described above, and the pellet was resupended in 0.1 M TrisHCl, 0.01 M EDTA, and 0.1% Triton-X, pH 7.8. The cell suspension was gently dounced five times in a loose fitting dounce homogenizer, and centrifuged at 5000 rpm in a Sorvall SS-34 rotor for 5 minutes, twice, to remove nuclei. The supernatant was then centrifuged at 12,000 rpm for 30 minutes in a Sorvall SS-34 rotor, and the viral capsid pellet was resuspended in 0.15 M NaCl, 0.01 M TrisHCl, and 0.01 M EDTA pH 7.5. Pancreatic ribonuclease (Sigma) was added to a final concentration of 0.05 mg/ml, and incubated at room temperature for 10 minutes. The capsids were lysed by the addition of SDS to a final concentration of 0.5%, and the solution was extracted gently, twice, with H_2O saturated phenol. The aqueous phase was dialysed against 10 mM TrisHCl, 1 mM EDTA, pH 7.5, and stored in small aliquots at -20°C.

Rapid Extraction of Viral DNA

This method was first described by Hirt (1967) for the extraction of polyoma DNA. Infected cell pellets were resuspended in 10 mM TrisHCl, 10 mM EDTA, pH 8 and lysed by the addition of SDS to a final

concentration of 0.006%. High molecular weight DNA was precipitated by incubation of the solution with 1 M NaCl at 4°C overnight. The precipitate was removed by centrifugation at 10,000 rpm in an Eppendorf centrifuge. The supernatant was extracted three times with H_2O saturated phenol and twice with ether before being precipitated with 70% ethanol. The DNA precipitate was dried and resuspended in a small volume of 10 mM TrisHCl, 1 mM EDTA pH 7.5.

Transfection of Cells With Viral DNA

Vero cells were transfected with HSV DNA using a modification of the method described by Graham and Van der Eb (1973). Half confluent monolayers of Vero cells were incubated for 3 to 5 hours with a $CaCl_2$ precipitate of viral DNA, prepared as described below. The medium containing the $CaCl_2$ precipitate was aspirated off the cells, and the monolayer was washed twice with sterile phosphate buffered saline (PBS: 0.137 M NaCl, 2.68 mM KCl, 0.88 mM KH_2PO_4, 6.41 mM Na_2HPO_4, pH 7.3) before α-MEM containing 5% fetal bovine serum was added to the cells. After several days, virus from infected cells was harvested as described above.

$CaCl_2$ precipitates were prepared by mixing 0.01 to 0.05 ml of viral DNA, 10 μg salmon sperm carrier DNA, 0.25 ml. of HEPES buffered saline (42 mM HEPES, 27 mM NaCl, 10 mM KCl, 1.4 mM Na_2HPO_4, 11 mM dextrose, pH 7.05-7.1), and H_2O to a final volume of 0.475 ml. 0.025 ml of 2.5 M $CaCl_2$ were added to this mixture and gently mixed in. Occasionally, 1 μg of plasmid DNA was included in the precipitate. Salmon sperm DNA (Sigma) was dissolved in H_2O at a concentration of 10 mg/ml and phenol extracted to sterilize it. It was dialysed against

H_2O to remove the phenol and stored at -20°C.

Nick Translation of DNA

DNA was labelled in vitro with ^{32}P by the nick translation procedure. 1 µg of DNA was mixed with the following reagents in a final volume of 0.045 ml: 55 mM TrisHCl pH 7.4, 5.5 mM $MgCl_2$, 5.5 mM 2-mercaptoethanol, 0.17 mM dATP, 0.17 mM dTTP, 0.17 mM dGTP, and 100 µCi ^{32}P-labelled dCTP, sp. act. > 2000 Ci/mmole (New England Nuclear). Five pg of activated DNase I (Boeringer Mannheim) was added to the reaction in a volume of 2.5 µl, and the mixture was incubated for 1 minute at room temperature. 20 units of E. coli DNA Polymerase I were added and the reaction was incubated at 14°C for 2 hours. The volume was increased to 0.5 ml with H_2O and the mixture was extracted once with phenol and once with ether. 100 µg of E. coli tRNA were added and the reaction mix was precipitated with 10 ml of ice cold 10% trichloroacetic acid (TCA) for 15 minutes at 0°C. The precipitate was pelleted by centrifugation at 15,000 rpm in a Sorvall SS-34 rotor for 10 minutes. The pellet was washed twice with 70% ethanol plus 10 mM Tris HCl, 1 mM EDTA, pH 7.5 and drained on a paper tissue. The pellet was dissolved in 1 ml of 100 mM TrisHCl pH 8 and 1 µl was counted in the ^{3}H channel on a Beckman scintillation counter to determine the Cerenkov radiation produced by the ^{32}P incorporated into the DNA. The sample was then boiled for 10 minutes and added to the hybridization mix.

DNase I was activated by incubating 10 µg with 0.01 M TrisHCl pH 7.5, 5 mM $MgCl_2$, 0.15 M NaCl, and 1 mg/ml bovine serum albumin (Sigma) at 0°C for 2 hours. Before use in the reaction it was diluted

1:50 in the same buffer.

Southern Blots

This procedure was first described by Southern (1975). It works best on agarose gels, although it can be used with limited success on acrylamide gels.

The gel was soaked successively in 0.5 N NaOH, 1.5 M NaCl plus 1 M TrisHCl pH 7.0, and 20xSSC (3 M NaCl, 0.3 M NaCitrate, pH 7) for 30 minutes each. It was then laid on top of 5 to 6 layers of Whatman paper which had been prewetted with 20xSSC. Saran wrap was used to create a window on the gel and prewetted nitrocellulose paper (Mandel) was carefully layered on top of the window. Two layers of dry Whatman paper were placed on top of the nitrocellulose and a wad of paper towels was layered on top and weighed down. After allowing the gel to blot to it overnight, the nitrocellulose was peeled off the gel and soaked for 10 minutes in 2x SSC and then baked for 2 hours at 80°C. It was then ready for hybridization.

The baked nitrocellulose was prewetted in H_2O and then soaked for 10 minutes in 2xSSC. It was then sealed in a plastic bag, and 20 ml of a prehybridization solution, consisting of 2xSSC, 0.4 mg/ml denatured calf thymus DNA, 0.4% bovine serum albumin, 0.4% Ficoll, and 0.4% polyvinylpyrolidone (Sigma) was added to the bag. Prehybridization was performed at 60° C for 5 hours, at which time the prehybridization solution was replaced with the hybridization solution and hybridization was allowed to proceed for 16 hours. The hybridization solution contained 2xSSC, 0.4 mg/ml denatured calf thymus DNA, 0.02% bovine serum albumin, 0.02% Ficoll, 0.02% poyvinylpyrolidone, 10% dextran

sulfate, and the nick translated DNA being used as the probe. When hybridization was complete, the probe solution was retrieved and stored at -20°C for future use.

The blot was washed with three changes at 30 minute intervals of 2xSSC plus 0.1% SDS at 67°C, followed by one wash with 0.1xSSC plus 0.1% SDS at 50°C. This last wash could be extended to 16 hours to get rid of excess background binding. The blot was then exposed to X-ray film, Kodak X-OMAT, with or without an intensifier screen.

Colony Hybridization

Individual colonies to be screened were replica plated, with the aid of a transparent grid, by touching the colonies with sterile toothpicks, and then applying the toothpicks to the same grid locations on two separate plates containing solid medium. The replicas were incubated overnight at 37°C to allow the colonies to grow. One replica was selected from each pair for colony transfer.

Circular discs of Gene Screen (New England Nuclear) were placed on top of the colonies and carefully smoothed in place to ensure contact between the disc and all of the colonies. The disc was then carefully peeled off the solid medium support with the colonies attached. The disc was placed, with the colonies on the upper surface, on top of several sheets of Whatman paper prewetted in 0.5 M NaOH. The disc was allowed to soak for 1 hour, and then was "washed" by laying it on top of Whatman paper soaked with various solutions. In between each "wash" any excess liquid was drained from the disc. The "washes", each 10 minutes long, consisted of once with 1.5 M NaCl plus 1 M TrisHCl, pH 7, and twice with 2xSSC. The disc was the placed in a small dish and

1 mg Proteinase K in 2xSSC (variable volumes) was layered on top of the colonies. The disc was incubated for 30 minutes, air dried, and washed with 5 changes of chloroform. After allowing the chloroform to evaporate, the disc was baked at 80°C for 2 hours. It was then hybridized with a radio-labelled probe as described above except that both the prehybridization and hybridization solutions contained 0.05 mg denatured E. coli DNA plus pBR325 DNA.

Gel Electrophoresis

Two kinds of gel matrices were employed throughout this study - agarose gels and polyacrylamide gels. Except for the specialized gels used in the work up of material for sequence analysis, both agarose and acrylamide gels were prepared and run with the same buffer system - 40 mM Tris acetate, 2 mM EDTA, pH 8.3. Agarose (Sigma) was melted in the electrophoresis buffer and either poured into a vertical gel apparatus, or into a horizontal gel apparatus. The % used refers to a W/v value. Polyacrylamide gels contained 40 parts acrylamide (Bio-Rad) to 1 part bis-acrylamide (Bio-Rad); the % used refers to the W/v amount of acrylamide. The 40% stock acrylamide solution was deionized by mixing with mixed bead resin (BioRad) and filtered through Whatman paper before storage.

Cloning Procedures

Many of the cloning procedures and reaction conditions used in this study are written up in exhaustive detail in Maniatis et al,(1982). Restriction enzymes and other enzymes used in molecular biology were obtained from a number of different companies, depending on availability and other exigencies of the market.

The basic protocol used to clone DNA fragments into plasmid vectors is as follows. The DNA to be cloned was digested with the appropriate restriction enzyme, electrophoresed through an agarose or an acrylamide gel, and eluted from the gel. Fragments were eluted from agarose gel slices by electroelution inside a dialysis sac. DNA was eluted from acrylamide by the method described by Maxam and Gilbert (1980), except that the buffer contained Na acetate instead of NH_4 acetate (see below). The vector DNA was digested with the appropriate restriction enzyme and then treated for 30 minutes with calf intestinal alkaline phosphatase (CIAP). The CIAP was inactivated by heating at 67°C in the presence of 0.5% SDS. The eluted DNA fragment was then mixed with the treated vector DNA, and the mixture was phenol extracted several times, extracted once with ether, and precipitated in the presence of 0.15 M NaCl with 70% ethanol. The precipitate was pelleted, the pellet was dried and redissolved in 20 µl of buffer containing 50 mM Tris (pH 7.4), 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM spermidine, 1 mM ATP, 0.1 mg/ml BSA, and 20 units of T4 DNA ligase. Ligation was performed at room temperature overnight, and bacteria were transformed with the ligation mix.

Sequencing DNA

The method used to sequence the various fragments in this study was that described by Maxam and Gilbert (1980). Except for one minor adjustment, no changes were made in their protocol. The one small change was to substitute Na acetate for NH_4 acetate in the elution buffer. This was to avoid the problem of NH_4 ion inhibition of T4 kinase, since many fragments had to be end labelled after elution.

RESULTS

Gene Conversion and/or Recombination in the "c" Inverted Repeats

At the time this thesis was begun (1980), the notion of "obligatory identity" of the inverted repeats in HSV DNA was still widely held to be true, in spite of the weak data supporting it. Moreover, Roizman et al's (1978) first model of HSV DNA replication and cleavage was also still in vogue. Since a major foundation of that model was "obligatory identity", which I and others (Smiley, pers. comm.) believed to be untrue, it seemed necessary to demonstrate its fallacy.

The diploid inverted "c" sequences flanking S contain the gene for ICP4, an immediate early polypeptide believed to be essential for the expression of delayed early and late viral genes (Clements et al, 1979; Mackem and Roizman, 1980; Watson et al, 1979). Temperature sensitive mutations affecting this diploid gene are recovered at a fairly high frequency (Schaffer et al, 1978), indicating that both copies of the gene are readily mutated. By transfecting cells with mutant viral DNA and subfragments of wild type HSV-1 or HSV-2 DNA, it is possible to "rescue" wild type "revertants" of ICP4 ts mutants. When Knipe et al (1978; 1979) performed this experiment, they found that, of 10 rescue products recovered, 8 were heterozygous and 2 were homozygous for the rescued gene. Restriction analysis of DNA from one of the heterozygous rescue products indicated that the viral terminal 1.6 kb

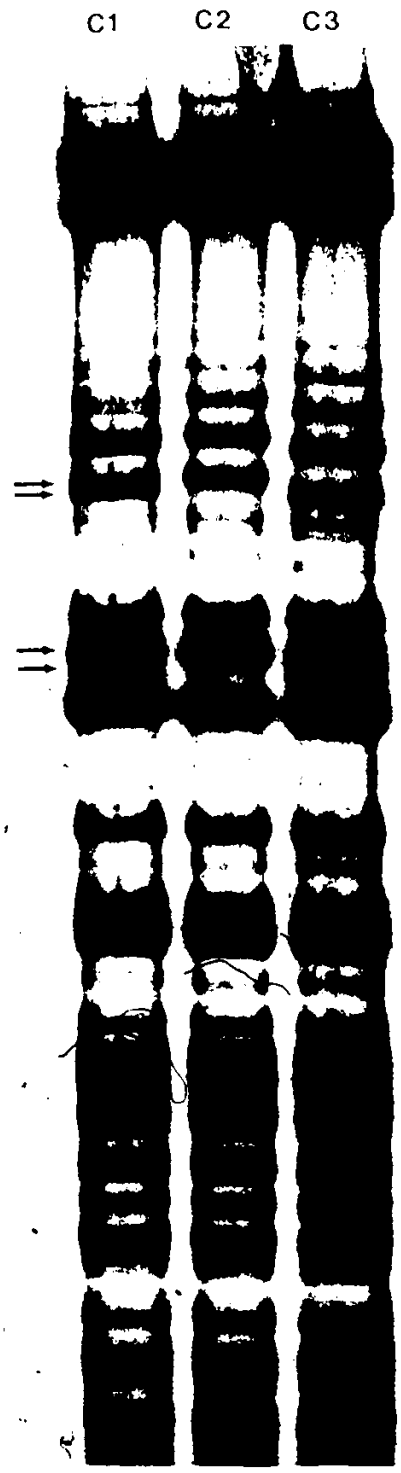
were identical in both copies even though the (internal) gene sequences were not.

This led to the notion that sequence identity of the terminal "ac" region was obligatory, and possibly reflected a vital copying mechanism for preserving essential information at the ends of the genome. Indeed, Jacob et al (1979) proposed that segment inversion resulted from the obligatory regeneration of "a" sequences by a mechanism involving the use of internal "a" sequences as templates to "repair" the loss suffered during cleavage/packaging of concatemeric precursor viral DNA (see Introduction). The regular occurrence of homozygous rescue products would be the result of occasional extension of this copying mechanism into the distal "c" and (by analogy) "b" sequences.

In our hands at that time was a subclone of HSV-1 (KOS) which contained heterozygous "c" sequences (KOSc). By itself the existence of this virus was interesting as it indicated that a clonal isolate could contain non identical sequences within the inverted repeats, although whether the non identity extended to individual molecules remained to be demonstrated. The heterozygosity was evident on restriction patterns of viral DNA as a size difference of approximately 60 -100 bp. Fragments containing the "c" sequences (ie joint and S terminal fragments) appeared as doublets. The lane labelled c1 in Fig. 2 shows an AluI digest of ^{32}P-labelled viral DNA from a plaque purified stock of KOSc which displays the [illegible]estriction pattern as its parent. Lanes c2 and c3 contain p[illegible]cified progeny of KOSc which have become homozygous for the larger (c2) or the smaller (c3) form of the

FIGURE 2

AluI digestion pattern of DNA from HSV-1 (KOSc) progeny. The heterozygote KOSc was plaque purified by end point dilution, and three progeny were expanded. DNA was labelled in vivo, and AluI restriction patterns were compared. Arrows indicate the two forms of the joint fragment (upper set) and S terminus (lower set). The three stocks shown here were designated KOSc1, KOSc2, and KOSc3. This experiment was performed by Jim Smiley.


C1
C2
C3

"c" sequence. Several observations can be made from this figure :

1) Joint and S terminal fragments of c1 both display the heterozygosity. The upper arrows indicate the joints, while the lower arrows indicate the S termini. In lane c1, the larger form of the S terminus comigrates with another band.

2) AluI cleaves the viral DNA within the inverted repeats, such that joint and terminal fragments are present in molar quantities, in contrast with joint spanning and terminal restriction fragments derived by cleavage outside of the repeat regions, which are in submolar quantities. The S terminus in lane c3 (lowest arrow) is clearly present in molar quantities, while the band indicated by the second arrow from the bottom is present in greater than molar quantities in lane c2 (compare it with the band above). The lower S terminal band in lane c1, however, is half molar in quantity. This means that half of the S termini in the c1 population have one type of S terminus and the other half have the other type. Similarly, the joint fragments in c1 are submolar in intensity such that the two together are roughly equivalent to one from either c2 or c3. Thus, the S terminal and joint fragments of c1 are divided into two groups which, together, add up to the full complement.

3) All of the forms of the joint fragment within the "a" sequence stepladder display the heterozygosity. This means that the heterozygosity is present in all forms of viral DNA. Remember that the terminal redundancy, "a", is accumulated tandemly on L termini and L-S joints in the population of viral DNA (Fig. 1).

Does this mean that all viral genomes within the KOSc stock contain two different "c" sequences? Attempts to determine whether the

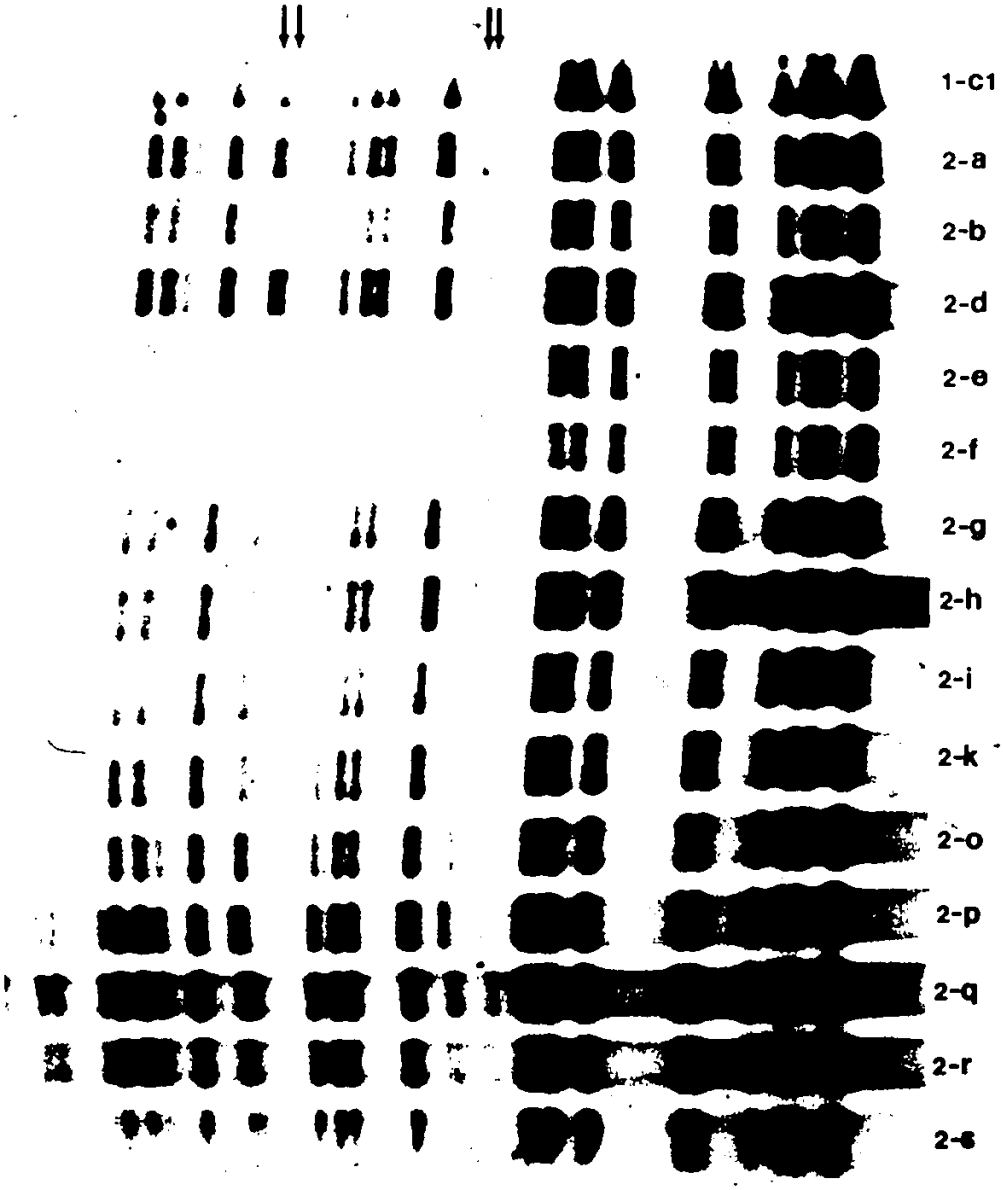
heterozygosity was fixed with respect to flanking unique sequences were unsuccessful. HindIII, which does not cleave within the repeats, also produced heterozygous joint and S terminal fragments (Jim Smiley, pers. comm.) of all forms. This indicated that either the heterozygosity was not fixed in either repeat and could move back and forth, or the population was a mixture of homozygous forms.

As indicated above, plaque purification of KOSc had yielded three different progeny: 2 homozygotes, each containing either the larger or the smaller "c" sequence, and a "heterozygote" in which both forms of the "c" sequence appeared as doublet bands (Fig. 2). The latter could have been a true heterozygote or the result of a mixed infection with two homozygotes. Since the heterozygosity was known to lie within the terminal 2.5 kb, a region of supposed obligatory identity, it was deemed necessary to determine 1) the location of the heterozygosity, and 2) the true nature of heterozygous stocks, ie were they true heterozygotes or mixed populations.

These two questions were addressed by applying more rigourous plaque purification conditions. The heterozygous progeny of KOSc, KOSc1, was plaque purified by endpoint dilution. This was accomplished by diluting KOSc1 and plating 0.1 pfu (plaque forming units) per well into each well of a Linbro plate containing a confluent monolayer of Vero cells. Fifteen wells out of 120 infected produced virus. Using a variation of Poisson distribution designed for application to Random Particle Analysis (Stahl, 1979), the probability of infecting a cell with one or more pfu is $(1-e^{-MOI})$, and the probability of infecting a cell with at least two different types of pfu is $(1-e^{-MOI})^2$, or

FIGURE 3

SacI digestion pattern of plaque purified progeny. KOSc1 was plaque purified by end point dilution, and labelled DNA from the progeny viruses was examined as described in the legend to Fig. 2, except that SacI was used instead of AluI. Upper arrows indicate the two forms of the joint fragment; lower arrows indicate the two forms of the S terminal fragment. The first number in the code above each track indicates the "generation", ie. 1-c1 is first generation (KOSc1 - the parent), 2-a is second generation, and in Fig. 5, 3-19 is third generation. KOSc1 was the parental stock from which all of the second generation progeny were derived, and KOSc1r (lane 2-r) was subjected to TK^- selection to generate all of the third generation progeny.

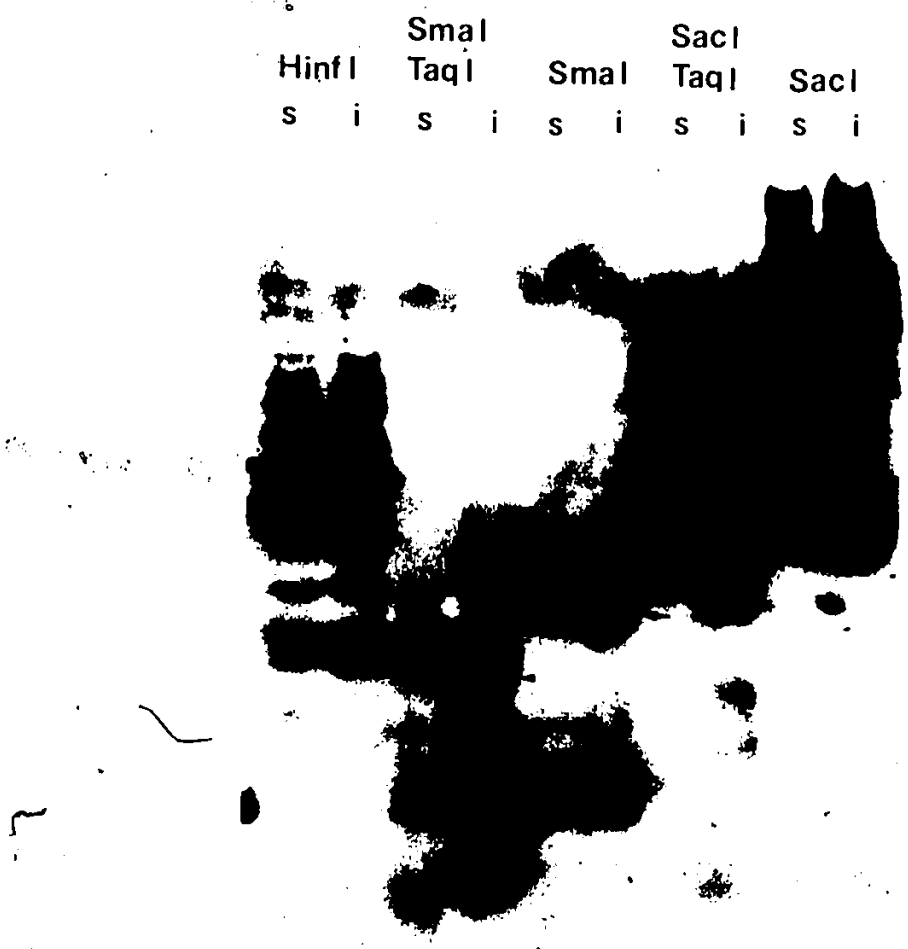

1-c1
2-a
2-b
2-d
2-e
2-f
2-g
2-h
2-i
2-k
2-o
2-p
2-q
2-r
2-s

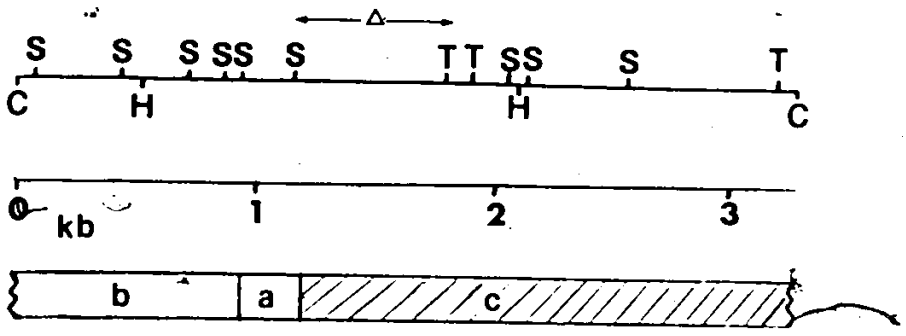
0.009 for a MOI (multiplicity of infection) of 0.1. These fifteen different isolates were examined for the restriction pattern of ^{32}P-labelled viral DNA (Fig. 3). The SacI joint fragment, indicated by the upper set of arrows, served as the diagnostic determinant. Where possible, the S terminus was also identified, although low incorporation of ^{32}P into viral DNA precluded this on several occasions. Each isolate was examined several times to ensure correct identification. Although the sample size was small, the proportion of heterozygotes in this population was fairly high - 4 of 15 (only 3 are on this particular gel - see Fig. 6 for designations). This is a much higher frequency than that predicted for random double infections (0.009).

Two of the homozygous progeny, KOScli and KOScls, were digested with various combinations of restriction enzymes, blotted to nitrocellulose, and hybridized with ^{32}P-labelled DNA from a cloned SacI joint, pINJ2 (Fig. 4). The map of the KOS joint (Wagner and Summers, 1978) displayed below the blot indicates the position of the relevant restriction sites within the joint region. The SacI joints (dot indicates the bottom step of a SacI joint stepladder) and S termini (asterisk) from KOScls (lanes marked s) and from KOScli (lanes marked i) both display one of the two forms of the "c" sequence, while the L terminal stepladders, clearly visible in the blot below the S termini, comigrate. This confirms the location of the heterozygosity within the "ca" region of the genome. The HinfI digests on the left indicate that the heterozygosity is contained within the HinfI joint and S terminus. This is confirmed by the SacI/TaqI digests. SmaI digestion of the two DNAs indicated that the SmaI A fragment (arrows) contained the hetero-

FIGURE 4

Southern (1975) blot map of deletion. The DNA from two homozygous progeny of KOSc1 - KOSc1i (smaller allele) and KOSc1s (larger allele) - were digested with various restriction enzymes, electrophoresed on a 1.4% agarose gel, blotted to nitrocellulose, and hybridized with a cloned joint spanning fragment (Smiley et al, 1981). Dots on the left of lanes marked "s" and on the right of lanes marked "i" lie beside the bottom step of junction stepladder fragments. Note the allelic differences between KOSc1s and KOSc1i; all fragments bearing the allelic region display the "phenotype", ie. they are either larger (KOSc1s) or smaller (KOSc1i). Asterisks indicate the S terminal fragments. Arrows indicate the heterozygous SmaI A fragment (refer to Fig. 14 for letter designations), and arrowheads indicate the heterozygous SmaI/TaqI fragments. The map below the blot is derived from data published by Wagner and Summers (1978) and shows the SmaI/TaqI fragment which contains the heterozygosity. The cartoon of the junction underneath the map orients the restriction sites within that region. S=SmaI, T=TaqI, H=HinfI, and C=SacI.


HinfI
SmaI
TaqI
SmaI
SacI
TaqI
SacI
s i s i s i s i s i

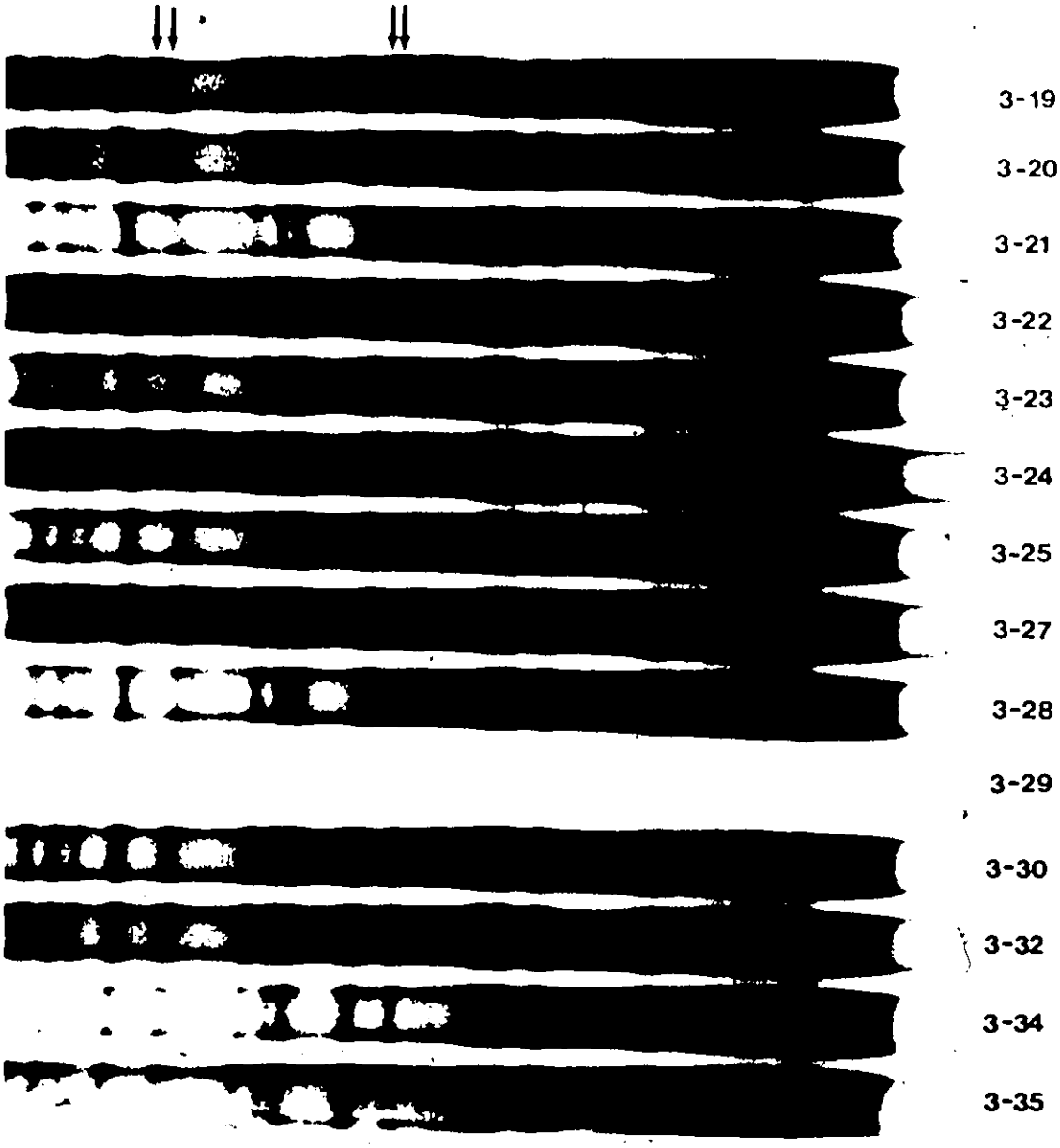

S S S SS S TT SS S T
C H H C
0 kb 1 2 3
b a c

zygosity (see Fig. 14 for letter designations of SmaI fragments of the joint). The region containing the heterozygosizy was finally mapped to the SmaI/TaqI fragment directly abutting "a". A dearth of appropriate restriction sites in this sequence prevented further delimitation of the heterozygosity site. Within the "ca" sequences of the S terminus present in the joint, the maximum distance from the end of "a" (see diagram below map) to the TaqI site defining the outer limits of the heterozygous region is 880 bp, well within the region of so-called obligatory identity.

It was still possible that the heterozygous stocks arising from the last plaque purification arose by mixed infection with two homozygotes. To reduce the probability of this event, one of the heterozygous progeny was plaque purified by selecting TK^- progeny. Mutants of this sort arise spontaneously in the population at a frequency of approximately 10^{-3}. Since selection is strictly dependent on the absence of thymidine kinase activity (see Materials and Methods), mixed TK^- infections would be expected to occur at a frequency of approximately 10^{-6} of mixed infections. Therefore this procedure selects against plaques arising from doubly infected cells. TK^- progeny of KOSclr were isolated by infecting $8x10^7$ Vero cells with $4x10^4$ pfu in the presence of Ara-T. TK^- plaques were picked into monolayers of Vero cells laid down in Linbro plates. The resulting stocks were titred in the presence and absence of Ara-T to confirm their TK^- phenotype, and then used to make ^{32}P-labelled viral DNA. SacI digests were screened for the type of joint and S terminal fragments present, ie whether large, small, or both. Several of the

FIGURE 5

SacI digests of TK⁻ progeny of KOSc1r. A heterozygous stock derived from the plaque purification of KOSc1, KOSc1r, was plaque purified by selecting naturally occurring TK⁻ progeny. The DNA from 31 different viruses was labelled *in vivo* and digested with SacI as described. Those shown here are therefore only a subset of the total sample. The arrows indicate heterozygous junction and S terminal fragments as described in the legend to Fig. 3. Heterozygous progeny include 3-19, 3-20, 3-21, 3-23, 3-25, 3-28, and 3-35; homozygous progeny with the smaller allele include 3-22, 3-29, 3-30, 3-32, and 3-34; homozygous progeny with the larger allele include 3-24 and 3-27.


3-19
3-20
3-21
3-22
3-23
3-24
3-25
3-27
3-28
3-29
3-30
3-32
3-34
3-35

digests are displayed in Fig. 5. The different isolates fell into three classes of approximately equal proportions : 7 homozygotes with the larger "c" sequence, 14 homozygotes with the smaller "c" sequence, and 10 heterozygotes (Fig. 6). The frequency of heterozygotes is far too high to be accounted for by mixed infection and must reflect the level of true heterozygotes - HSV genomes with two different "c" sequences - in the population. This result strongly supports the idea that sequence identity is not obligatory, at least within the 600 bp sequence abutting the "a" sequence.

The three "generations" of KOSc progeny are displayed together in Figures 7 and 8 as whole viral DNA or blots of selected sequences. Since AluI cleaves the same sequence as SacI (although not vice versa), the two figures can be directly compared, although different third generation progeny were used. Note again that the heterozygosity persists throughout the population of fragments containing "c" sequences in the heterozygous stocks, and disappears from all of the affected fragments within the homozygous stocks, indicating that the different forms of viral DNA have a common lineage. The faint band below the S terminus in the BamHI digest of 2-s in Figures 7 and 8 is probably the second step of the L terminus stepladder.

At about the same time that this work was published, Davison and Wilkie (1983) demonstrated that intertypic recombinants harbouring different "a" sequences were viable, suggesting that the region of non-identity could be extended to the ends of the genome. I will show below that different (although similar) "a" sequences may be inserted into the genome without affecting the "activity" of either the resident

FIGURE 6

Genealogy of KOSc progeny. The "phenotype" is indicated by wt/Δ (heterozygous), wt/wt (homozygous for the larger allele of "c"), or Δ/Δ (homozygous for the smaller allele of "c"). The letters in the KOSc1 progeny group are individual codes (see Fig. 3), and the numbers in the last group indicate the number of progeny with a particular "phenotype".

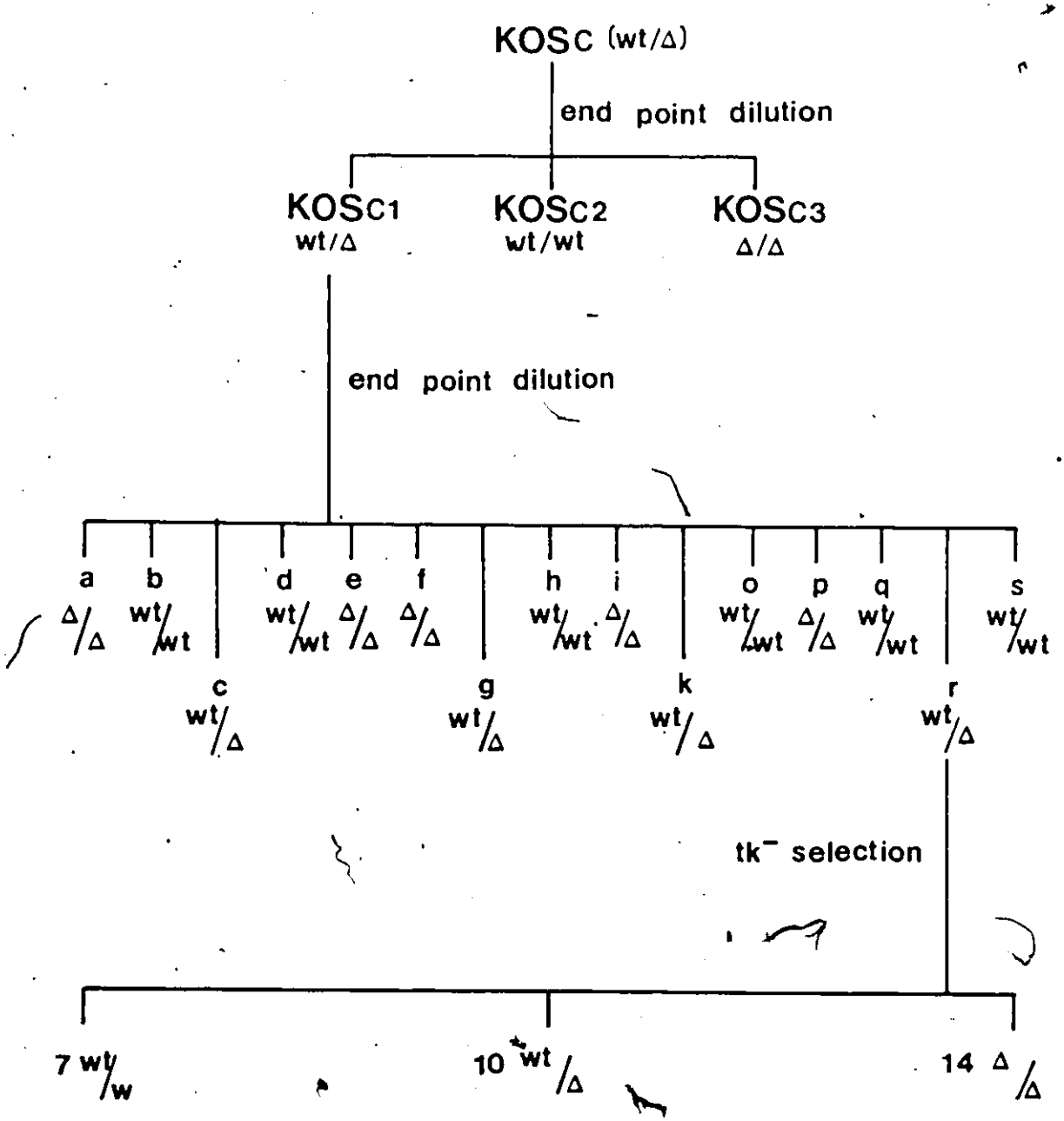

KOSC (wt/Δ)
end point dilution
KOSC1 wt/Δ
KOSC2 wt/wt
KOSC3 Δ/Δ
end point dilution
a Δ/Δ
b wt/wt
c wt/Δ
d wt/wt
e Δ/Δ
f Δ/Δ
g wt/Δ
h wt/wt
i Δ/Δ
k wt/Δ
o wt/wt
p Δ/Δ
q wt/wt
r wt/Δ
s wt/wt
tk⁻ selection
7 wt/w
10 wt/Δ
14 Δ/Δ

FIGURE 7

Restriction profile of parental and plaque purified progeny. DNA from representative stocks of plaque purified progeny of the three different phenotypes was labelled *in vivo* with ^{32}P, and digested with either BamHI or AluI. The three generations are represented among the digests; the first number in the code above each lane designates the generation. Arrows indicate joint (upper set) and S terminal (lower set) fragments. Heterozygous stocks are in lanes marked 1-c1, 2-r, and 3-10. Stocks homozygous for the small allele are in lanes marked 2-i and 3-5; stocks homozygous for the large allele are in lanes marked 2-s and 3-8.

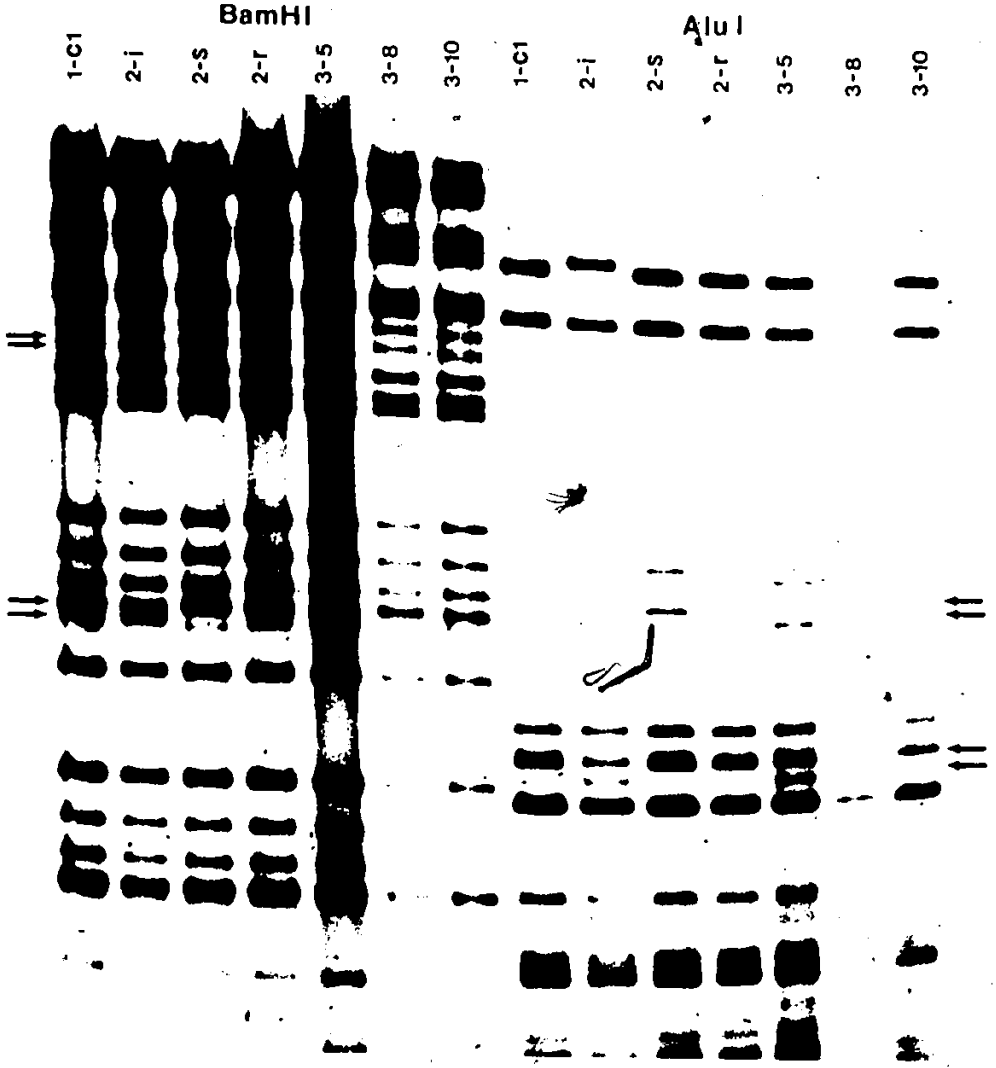
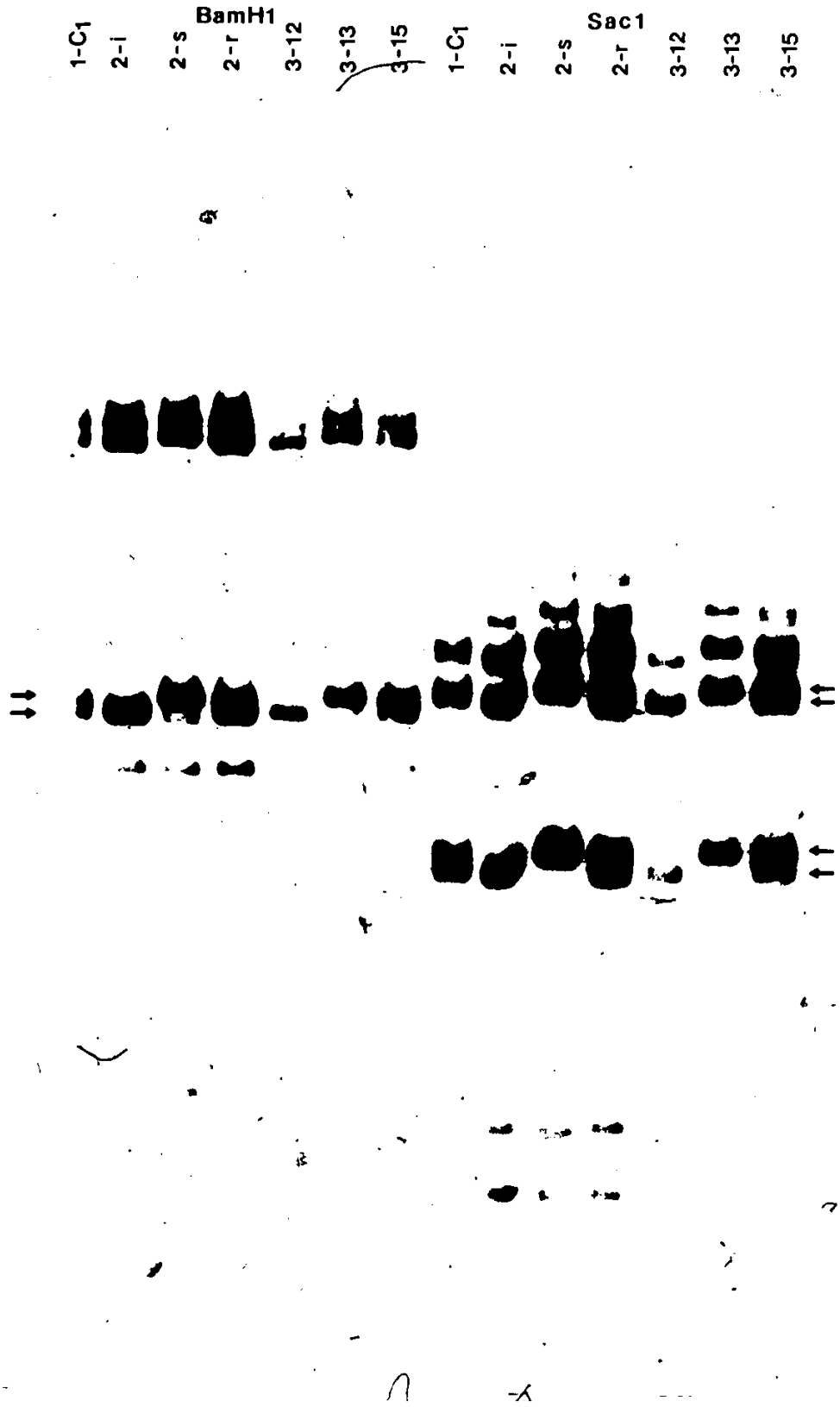

BamHI
1-C1
2-i
2-s
2-r
3-5
3-8
3-10
Alu I
1-C1
2-i
2-S
2-r
3-5
3-8
3-10

FIGURE 8

Restriction profile of parental and plaque purified progeny virus. DNA from representative stocks of plaque purified progeny of the three different "phenotypes" was digested with either BamHI or SacI, electrophoresed through a 1% agarose gel, and blotted to nitrocellulose. The blot was hybridized with ^{32}P-labelled DNA from a cloned junction (Smiley et al, 1981). Arrows on the left indicate S terminal fragments. The different alleles of the junction fragments, grouped above the S terminal fragments, are difficult to distinguish in the blot, particularly in the heterozygous stocks, although the difference between the two forms can be discerned by comparing lanes 2-i and 2-s. Arrows on the right indicate junction (upper set) and S terminal (lower set) fragments in the SacI digests. The SacI digests in this figure can be directly compared with the AluI digests in Fig. 7, although different representative stocks were used. AluI cleaves at all SacI sites, although the reverse is not the case. However, the junction and terminal fragments do not contain additional AluI sites. Heterozygous stocks are in lanes marked 1-c1, 2-r, and 3-15; stocks homozygous for the small allele are in lanes marked 2-i and 3-12; stocks homozygous for the large allele are in lanes marked 2-s and 3-13.


BamH1
Sac1
1-C1
2-i
2-s
2-r
3-12
3-13
3-15
1-C1
2-i
2-s
2-r
3-12
3-13
3-15

"a" sequences or the inserted "a" sequence, supporting the idea that sequence nonidentity is perfectly acceptable.

Apart from refuting the idea of obligatory identity, the most interesting observation to come out of these experiments was the pattern of segregation of the different classes of progeny. At each step, all three classes arose at approximately equal frequency. If the parent genome was a true heterozygote, as indicated by these experiments, then it quickly gave rise to homozygous progeny. This could occur by a copying mechanism, such as gene conversion, during segment inversion, or by intermolecular exchange, possibly mediated by activities relating to the "a" sequence (see below) (Varmuza and Smiley, 1984). Whatever the mechanism is, it is very efficient.

Sequence Identity of "a" Is Not Obligatory

The experiment described above clearly indicated that the flanking "c" sequences were not obligatorily identical, although they quickly became so. The question still remained as to whether the "a" sequences were as flexible. At about this time, Mocarski and Roizman 1982a, 1982b) had proposed that the "a" sequence was a cis acting site engaged in site specific recombination, a mechanism purported to be responsible for segment inversion. Smiley et al (1981) and Mocarski et al (1980) had both shown that insertion of a joint spanning fragment into the L unique sequences resulted in a virus with three freely inverting segments delineated by the position and orientation of the repeats - inverted repeats promoted inversions while direct repeats promoted deletions. Mocarski and Roizman (1982a) further showed that

either terminus was capable of producing the same results, while non junction fragments were inert. These data were interpreted as support for the idea that HSV DNA undergoes segment inversion by means of site specific recombination mediated by cis acting signals in the "a" sequence.

If the "a" sequence is indeed a cis acting signal for site specific recombination, then sequence identity of this segment is not an unreasonable hypothesis, since other site specific mechanisms have been shown to be strictly dependent on the integrity of the target sequence (see Introduction). This latter aspect could be tested directly by making a recombinant containing two different "a" sequences and asking whether or not the heterogeneity persisted and also whether it affected the "recombinogenic" nature of either the insert or the resident "a" sequences.

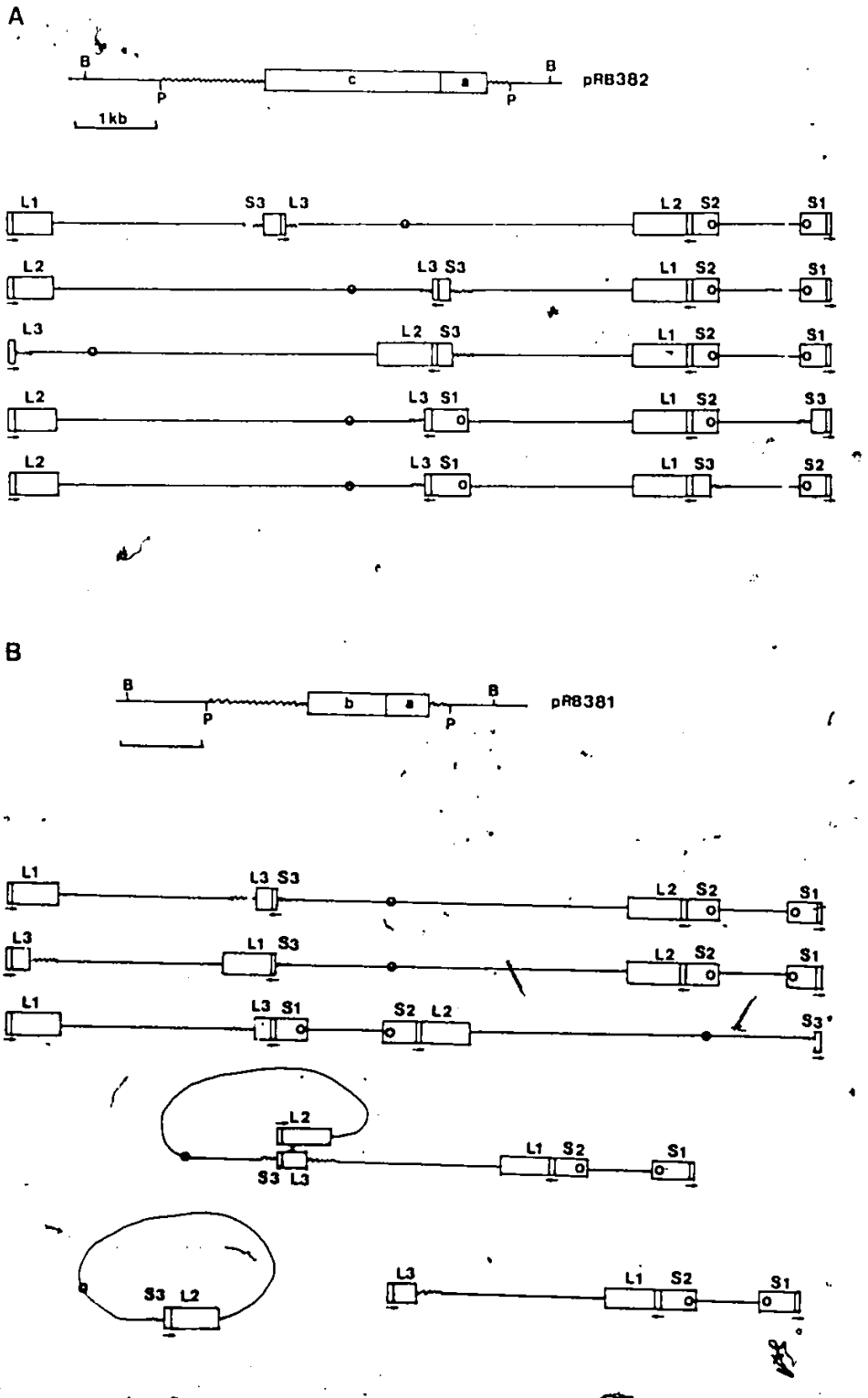
Strains F and KOS differ from each other in the size of their "a" sequences. Strain F "a" sequence is 550 bp long, while the KOS "a" sequence is 310 bp long. Mocarski and Roizman (1982a) have published the sequence of the strain F "a" sequence, and I have determined the sequence of the KOS "a" sequence (see below and Varmuza and Smiley, 1985). Apart from a few mismatched bases, the difference in size can be accounted for by the number of iterations of a 12 bp direct repeat (DR2) which is iterated 9 times in KOS and 18-20 times in strain F, and in the number of iterations of a 36 bp repeat (DR4) present once in KOS as part of the unique sequence and three times in F (Fig. 13).

Dr. Roizman kindly supplied me with two plasmids containing the viral termini from strain F cloned into the strain F TK gene. These

FIGURE 9

(A) Map of recombinant site in pRB382, and diagram of segment inversion products in KOSxF recombinant viruses containing the insert. pRB382, kindly supplied by Dr. B. Roizman, contains the SacI S terminal fragment cloned into the SacI and BglII sites in a BamHI TK fragment from strain F (Mocarski and Roizman, 1982a). This clone was used to make KOSxF and FxF recombinant viruses; the letter c in coded names of these viruses indicates that they contain this construction, ie. that they contain the "ca" sequence at the TK locus. The wavy line bounded by Ps indicates the extent of homology between the new TK locus in the recombinant viruses and the probe used to identify and analyse them. The viral genomes represented in the figure have been drawn in several, but not all, of the predicted orientations of the various segments in order to show the structures of recombinant fragments. The arrows underneath each genome indicate the orientation of the designated "a" sequences. The letter o represents viral origins of DNA replication. P=PvuII, B=BamHI. PvuII cuts viral DNA within the "c" sequence to generate a 4700 bp S terminal fragment (S1 and S2), and within the U_L sequences to produce 8000 bp and 11800 bp L terminal fragments (L1 and L2). The positions of these PvuII sites within the P isomer is not known, and the designations L1 and L2 are arbitrary. BamHI cuts viral DNA in the "c" sequence to produce 3200 bp S terminal fragments (S1 and S2), and within the "b" sequence to produce 2100 bp L terminal fragments (L1 and L2). The new arrangements can be derived by inverting segments flanked by inverted repeats of the "a" sequence.

(B) Map of insert site in pRB381. pRB381, kindly supplied by Dr. B. Roizman, contains the SacI L terminus cloned into the SacI and BglII sites in TK (Mocarski and Roizman, 1982a). The bar below the map indicates 1 kb. As described above, only a subset of new arrangements are shown here to indicate possible mechanisms for generating new fragments. The bottom two panels show the derivation of S3L2 recombinant fragments by deletion of a segment containing an origin of replication and flanked by "a" sequences in direct orientation. KOSxF and FxF recombinant viruses which have the letter b in their code names contain this construction.


A
B
P
c
a
pRB382
1kb
L1
S3
L3
L2
S2
S1
L2
L3 S3
L1
S2
S1
L3
L2
S3
L1
S2
S1
L2
L3 S1
L1
S2
S3
L2
L3 S1
L1
S3
S2
B
P
b
a
pRB381
L1
L3 S3
L2
S2
S1
L3
L1 S3
L2
S2
S1
L1
L3 S1
S2 L2
S3
L2
S3 L3
L1
S2
S1
S3 L2
L3
L1
S2
S1

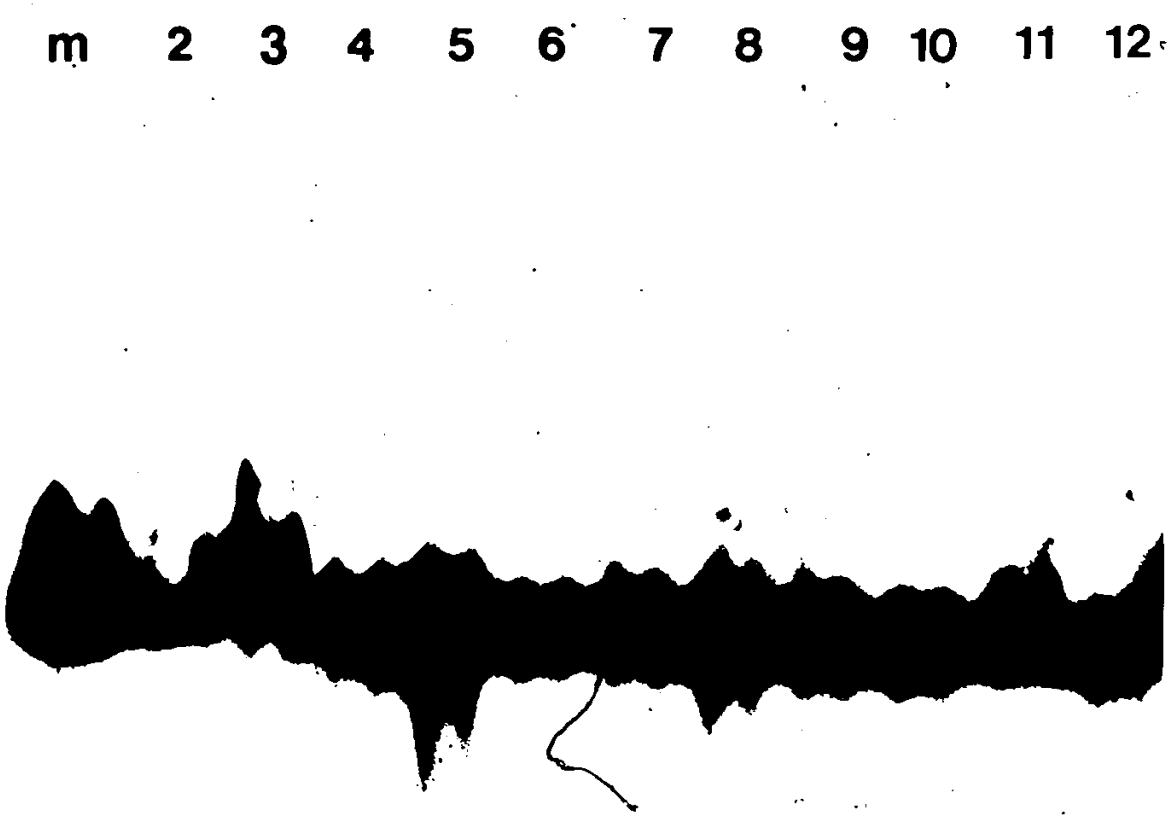
were pRB381 (L terminus in TK) and pRB382 (S terminus in TK). The stucture of these hybrid TK genes is shown in Fig. 9A and 9B.

These hybrid genes were rescued into KOS by cotransfection of Vero cells with viral and plasmid DNA. TK$^-$ viruses were isolated from the resulting stocks of virus as described above. Because the spontaneous mutation rate in HSV is quite high, the TK$^-$ virus stocks had to be screened for the presence of the desired recombinant. This was done by blotting BamHI digests of Hirt (1967) extracts and hybridizing them with ^{32}P-labelled pTK173 DNA, which contains the PvuII TK-containing fragment from KOS cloned into pBR322 (see appendix). In many cases, none of the isolates examined were the desired recombinant. Fig. 10 shows a typical marker rescue screening blot in which one of the isolates, seen in lane 3, was the desired product. This was evident both from the size of the major TK-containing fragment, which was approximately 1000 bp larger than the wt TK fragment, and from the presence of additional TK-containing fragments diagnostic of segment inversion through the insert.

The diagram in Fig. 9 shows the predicted structure of viral genomes which have undergone segment inversion around the inserted L and S termini. This diagram is almost identical with Figure 3 in Mocarski and Roizman (1982a) which explains the results of their marker rescue experiment. Normal termini are designated L1, L2, S1, and S2. Normal joints, L1S1, L1S2, L2S1, and L2S2, are indicated. None of these sequences will hybridize with a TK probe. New (predicted) termini are designated L3 and S3, and the joint composed of L3S3 is the hybrid TK insert sequence. Recombinant fragments derived from segment inversion

FIGURE 10

Sample screening gel used in the search for KOSxF recombinant viruses. TK^- plaques were picked into Linbro plates containing monolayers of Vero cells and allowed to amplify. The resulting stocks were used to prepare Hirt (1967) extracts of viral DNA which were digested with BamHI, electrophoresed through 1.4% agarose gels, and blotted to nitrocellulose. The blots were hybridized with ^{32}P-labelled DNA from a plasmid, pTK173, which contains the PvuII TK fragment from HSV-1. All of the stocks in this gel came from one transfection. The virus in lane 3 displays the predicted restriction profile for a KOSxF recombinant, while all of the rest are indistinguishable from the marker in the lane marked m, which contains KOS DNA, and probably represent point mutations. Many different stocks were screened in this manner. Only KOSxF or FxF recombinants from separate transfections were chosen for further study.


m 2 3 4 5 6 7 8 9 10 11 12

around the insert are designated L3S1, L3S2, S3L1, and S3L2. All of the sequences containing L3 or S3 will hybridize with a TK probe, although the intensity will depend on the degree of homology. For example, the PvuII TK sequence in pTK173 shares only 252 bp of homology with S3 fragments from viruses with the pRB381 (L terminus) insert and with L3 fragments from viruses with the pRB382 (S terminus) insert. The deletion displayed in Fig. 9B simply illustrates one of the recombinational mechanisms which might give rise to S3L2 fragments. As mentioned above, deletions of this sort were observed by Smiley et al (1981).

Analysis of the KOSxF recombinants was facilitated by direct comparison with similar recombinants made at the same time in which the strain F termini were inserted into strain F DNA. These viruses are identical with the viruses analysed by Mocarski and Roizman (1982a) and found to undergo free segment inversion around the inserts. Blots of BamHI digests hybridized with ^{32}P-labelled TK DNA were screened for the presence of the particular phenotype displayed in Fig. 10, lane 3. Laborious screening of a number of stocks prepared as described above yielded 11 different isolates, from 11 different stocks, in which the viral DNA contained multiple bands homologous with TK, one of which comigrated with the plasmid marker. Six contained the strain F S terminus inserted in the KOS TK gene (KOScF1, KOScF2, KOScF3, KOScF4, KOScF5, KOScF6), two contained the strain F L terminus in KOS TK (KOSbF2, KOSbF13), one contained the strain F S terminus inserted into the strain F TK (FcF9), and two contained the strain F L terminus rescued into F TK (FbF14, FbF17). The last three were the positive controls mentioned above. BamHI and PvuII digests of viral DNA from

four representative viruses were analysed for the presence of inversion fragments by hybridization with ^{32}P-labelled TK DNA (Fig. 11). All of the expected fragments were present (Table I). The patterns displayed by the positive controls were indistinguishable from those presented by the heterologous constructs, except in the sizes of some of the inversion fragments, which can be accounted for by the 240 bp difference in size between strains F and KOS "a" sequences.

Segment inversion around the inserts in the heterologous TK$^-$ viruses may have been the result of conversion of the "a" sequences to one of the two types - F or KOS. This possibility was eliminated by examination of the SmaI F subfragments of "a" from the different viruses. (Unfortunately, the terminology is a bit confusing here. "SmaI F" refers to the internal SmaI fragment from "a" as designated by Wagner and Summers, 1978 - see Fig. 14. The SmaI sites in strain F and KOS are homologous and cleave out the internal sequences which bear the heterologous regions of the two "a" sequences.) All of the different isolates were digested with SmaI, electrophoresed, and blotted to Biodyne A membranes. The blots were hybridized with a ^{32}P-labelled SmaI F probe derived from a clone of the KOS joint, pSVODK6 (see appendix), (Fig. 12). Most of the viruses contained two "a" sequences - a KOS "a" sequence and an F "a" sequence. Two heterologous viruses appeared to contain only KOS "a" sequences (KOScF1 and KOSbF13), suggesting that in these isolates the strain F "a" sequence was converted to a KOS "a" sequence. All of the rest, however, contained both forms in roughly equal amounts, although KOScF6 and KOSbF2 appear to have more of the KOS form than of the F form. Remember that the SmaI F

FIGURE 11

Blots of representative KOSxF and FxF recombinant viruses digested with BamHI or PvuII. DNA from four recombinants was digested with BamHI or PvuII, blotted, and probed with ^{32}P-labelled TK DNA (see map in Fig. 9). KOSbF2 is a KOSxF recombinant with the strain F "ba" sequence inserted at TK; KOScF4 is a KOSxF recombinant with the strain F "ca" sequence inserted at TK; FcF9 is an FxF recombinant with the strain F "ca" sequence inserted at TK; FbF17 is an FxF recombinant with the strain F "ba" sequence inserted at TK. The numbers on the left and right of the two panels indicate the position of molecular weight size markers (in kb). Densitometric scans of the "ba" containing viruses (KOSbF2 and FbF17) were performed on various exposures of these blots. The bands in the "ca" containing recombinant viruses were deemed too close together to derive meaningful data from. The data are described in the text. Sizes of recombinant fragments are listed in Table 1. Some of the terminal fragments, not visible on these blots, were detected in longer exposures (not shown).

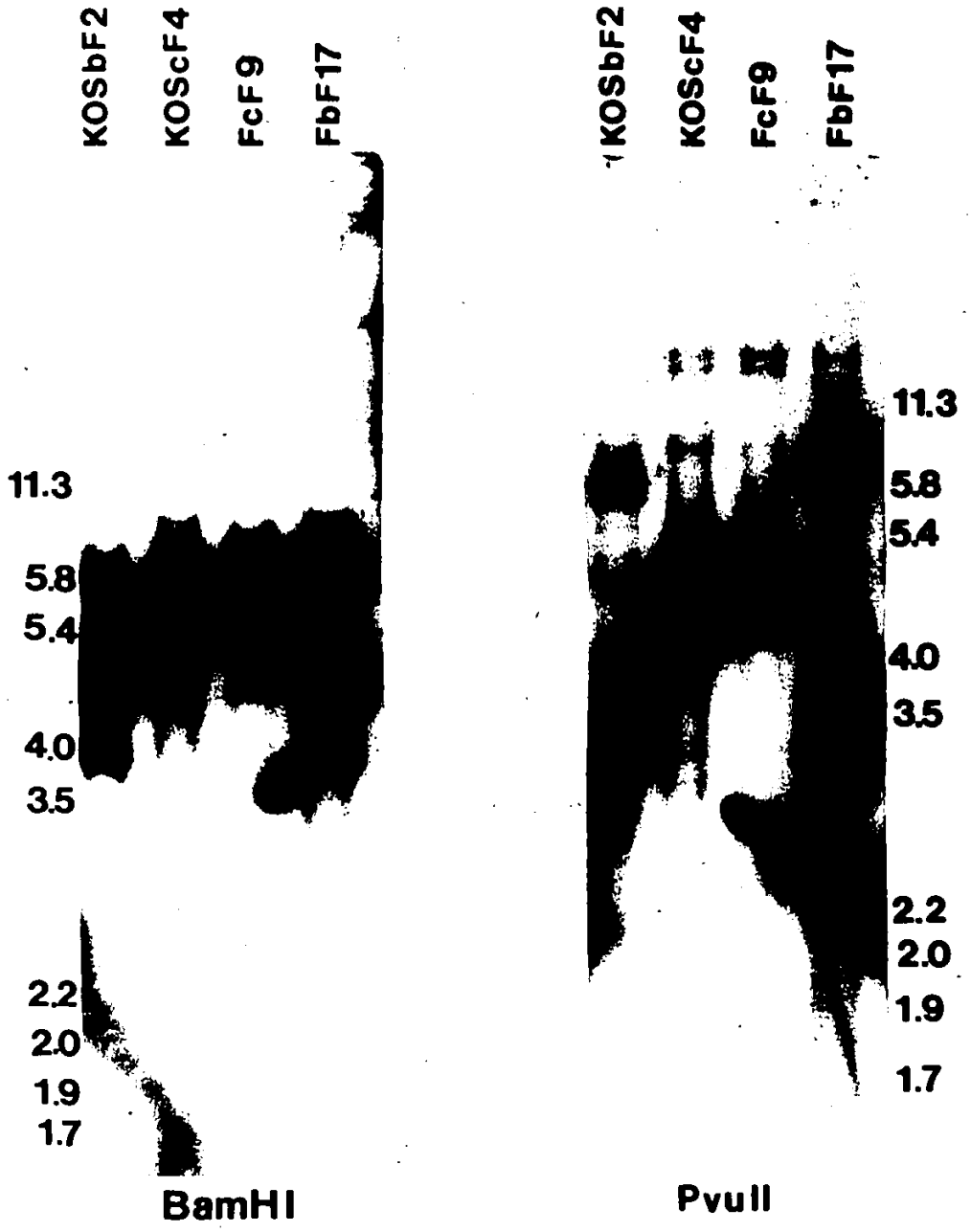

KOSbF2
KOScF4
FcF9
FbF17
11.3
5.8
5.4
4.0
3.5
2.2
2.0
1.9
1.7
BamHI
KOSbF2
KOScF4
FcF9
FbF17
11.3
5.8
5.4
4.0
3.5
2.2
2.0
1.9
1.7
PvuII

TABLE 1

RECOMBINANT TK FRAGMENTS IN KOSxF AND FxF VIRUSES

FRAGMENT		L3	S3	L3S3	L3S1	L3S2	S3L1	S3L2
VIRUS								
				PvuII Digests				
KOSbF2	exp.	2760	802	3012	7460	7460	8800	12600
	obs.	2560	n.d.	2700	7000	7000	n.d.	n.d.
FbF14	exp.	2760	802	3012	7700	7700	9050	12850
	obs.	2500	n.d.	2700	7600	7600	n.d.	n.d.
KOScF5	exp.	802	3950	4100	5500	5500	11950	15750
	obs.	n.d.	3700	4200	5400	5400	n.d.	n.d.
FcF9	exp.	802	3950	4100	5740	5740	12200	16000
	obs.	n.d.	4000	4300	5600	5600	n.d.	n.d.
				BamHI Digests				
KOSbF2	exp.	3480	1300	4230	6680	6680	3400	3400
	obs.	3600	1500	4500	6800	6800	3600	3600
FbF14	exp.	3480	1300	4230	6920	6920	3640	3640
	obs.	3600	1500	4500	8000	8000	3600	3600
KOScF5	exp.	1300	4730	5480	4500	4500	6830	6830
	obs.	1500	5300	6200	4500	4500	8500	8500
FcF9	exp.	1300	4730	5480	4740	4740	7070	7070
	obs.	1500	5300	6300	4500	4500	9000	9000

n.d. - not detected
Note that S3 containing fragments frqm bF viruses and L3 containing fragments from cF viruses share only 252 bp of homology with the probe. Several of the large fragments (which do not transfer well) in the PvuII digests comigrated with minor bands in the KOS marker lanes on some gels.

FIGURE 12

SmaI F digests of KOSxF and FxF recombinants probed with "a" sequence DNA. DNA from all of the recombinants was digested with SmaI, electrophoresed part way through a 2% agarose gel, and blotted to Biodyne A paper. The blot was hybridized with ^{32}P-labelled SmaI F DNA purified from a SmaI digest of pSVODK6. This fragment constitutes most of the "a" sequence.

KOSbF13

FbF14

FbF17

KOSbF2

KOScF1

KOScF2

KOScF3

KOScF4

KOScF5

KOScF6

FcF9

KOS

F

fragment is present in greater than molar quantities bdecause of the iteration of the "a" sequence. Surprisingly, the control viruses - FcF9, FbF14, FbF17 - also contained more than one form of the "a" sequence, as did strain F virus itself. A similar observation was made by Wagner and Summers (1978) in their analysis of KOS joint sequences. These authors found that two forms of the SmaI F fragment - F and F* - which differed in size by 10-12 bp, routinely appeared in their gels. Attempts to trace the origins of these two forms were fruitless (Jim Smiley, pers. comm.). Data presented later in the thesis may shed some light on this problem.

Sequence identity of any of the terminal sequences is clearly not obligatory. However, heterology is apparently unstable, suggesting that frequent exchanges occur, in this region of the genome at least. These could be intramolecular gene conversion-like events, perhaps initiated by segment inversion, or intermolecular exchanges similar to the recombination in bacteriophage T4 which has been shown to be involved in the formation of concatemeric DNA. Much more detailed experiments than those described here will have to be done to establish whether the events resulting in exchanges in HSV DNA are intramolecular, intermolecular, or a combination of both.

The issue of site specific recombination associated with the "a" sequence is unclear at this point. If "a" does contain a cis-acting sequence, then one can argue that both KOS and strain F contain the same signal. However, two observations argue against site specific recombination, at least in a semantic sense. As pointed out in the Introduction, classical site specific recombination systems are rarely

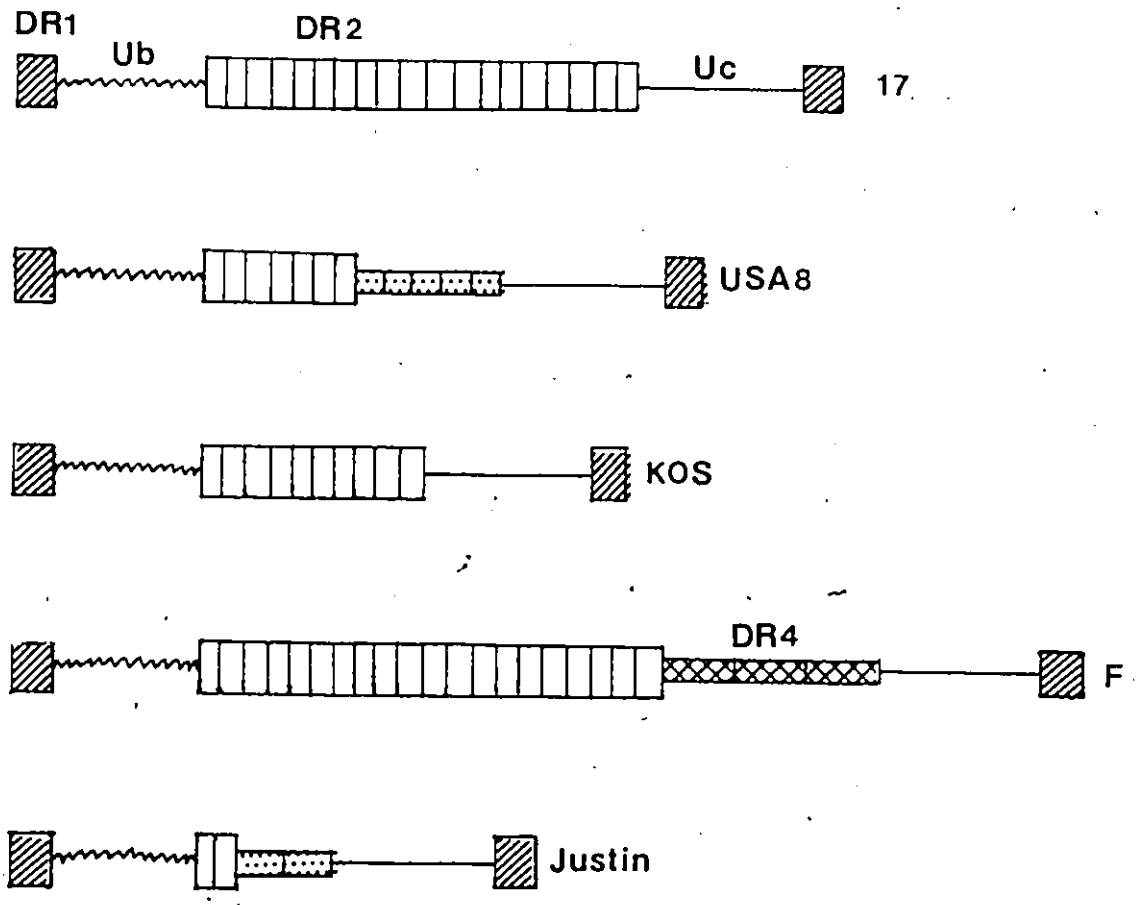
associated with exchange outside of the "site". Yet clearly, hetero-duplexes must form across the heterologous regions not only of the "a" sequence in the recombinants described in this section, but also in the flanking "c" sequences as seen in the previous section. Second, while the interstrain recombinants exhibit similar segment inversion patterns, densitometric scans of several digests have indicated that the L3S1+L3S2 recombinant fragments in KOSxF recombinants are less abundant, with respect to the TK stepladder bands, than their counter-parts in the FxF recombinants, suggesting that the heterology may inhibit exchange somewhat (KOSxF - 1:5; FxF - 1:1).The intermediate behaviour of the KOSxF recombinants from that expected of a site specific recombination system (ie neither fully yes nor no) suggested that alternative explanations, for example generalized recombination, perhaps at a "hot spot", might be in order.

Sequence Analysis of the KOS Joint

The ability of strain F "a" sequences to drive segment inversion in KOS DNA without suffering any apparent change was an interesting observation. At the time that the experiment was performed, the sequence of KOS joint DNA was not known. The large size differences between KOS and F "a" sequences suggested substantial differences in sequence, although Davison and Wilkie's (1981) analysis of several strains of HSV-1 and HSV-2 joints indicated that strain (but not type) differences frequently resulted from variations in the number of repeated sequences, rather than in sequence differences per se. Fig. 13 diagrams the similarities and differences of "a" sequences from several

FIGURE 13

Schematic diagram of "a" sequences from different strains of HSV-1. The scale of the diagram may be taken from the size of DR1 (20 bp), shown here by the diagonally striped boxes. Wavy lines designate Ub sequences, straight lines designate Uc sequences, and open boxes designate DR2 sequences, all of which are conserved among the various strains. The stippled boxes in USA8 and Justin represent a repeat which is conserved between these two strains, but which is absent from the other strains. The crosshatched boxes in strain F represent DR4 (Mocarski and Roizman, 1982a). DR4 is a composite of DR2 and part of Uc. Mocarski and Roizman, 1982a noted that there are four copies of this repeat in strain F. I have chosen to permute the repeat in order to maintain the identity of Uc; this results in the apparent loss of a DR4.


DR1
Ub
DR2
Uc
17
USA8
KOS
DR4
F
Justin

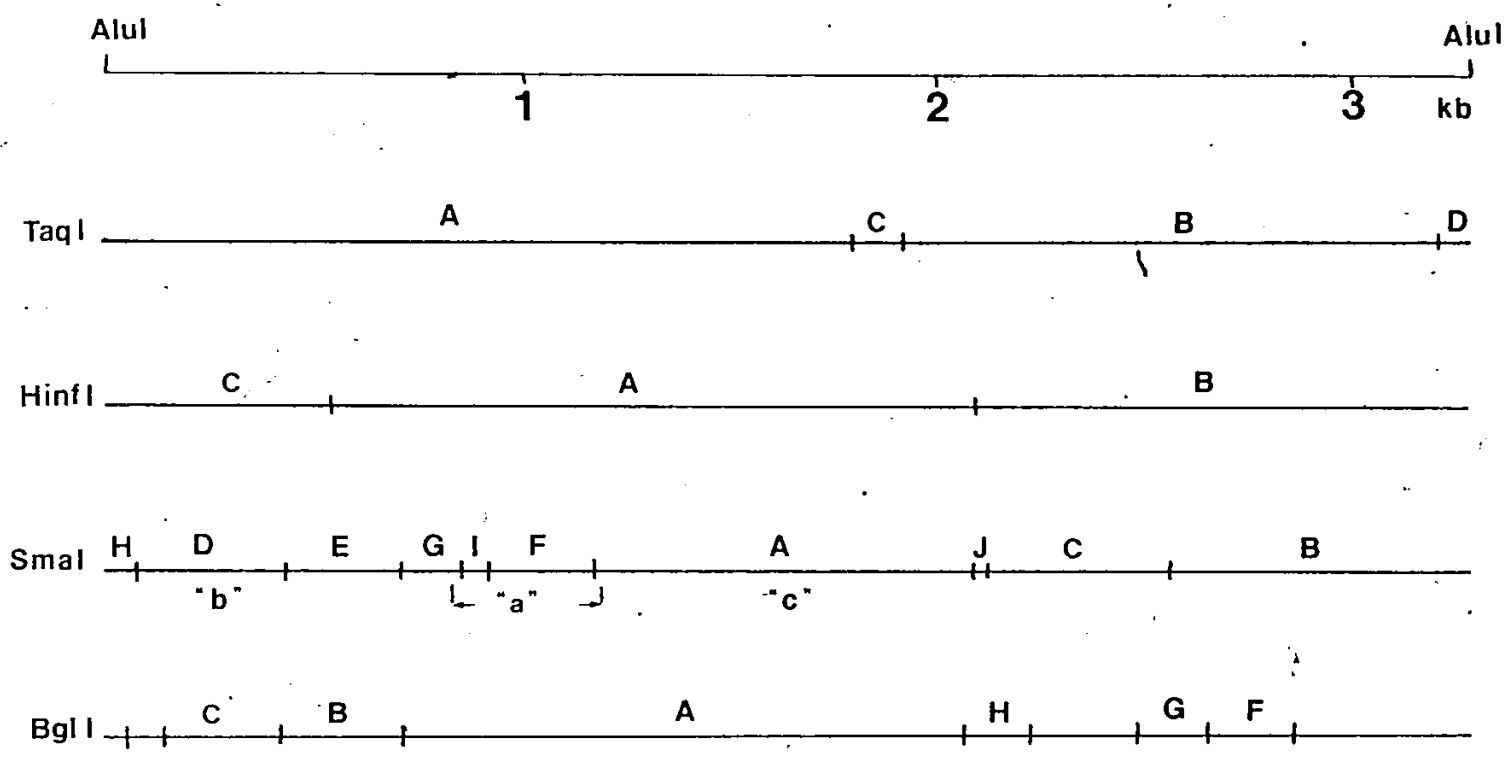
strains of HSV-1, including KOS (this work). Note that the blocks of heterology among the HSV-1 sequences are all internal. This point will become significant later.

The DNA source for the sequence reactions was a plasmid, pSVODK6, which contained the BamHI joint spanning fragment from KOSc1i cloned into the BamHI site on pSVOD (see appendix). KOSc1i was a plaque purified progeny of KOSc1 which contained a homozygous small "c" sequence (see "Gene Conversion in the Inverted Repeats"). As sequence data became available, other clones of joint and terminal fragments were also used. These will be described as they become relevant.

The restriction map of the KOS joint determined by Wagner and Summers (1978) is shown in Fig. 14. The original strategy was to sequence across the SmaI fragments from either direction by labelling the SmaI fragments at their 5' ends, isolating the fragments of interest, separating the strands, and performing the chemical cleavage reactions devised by Maxam and Gilbert (1980). At the outset of the exercise, however, a detailed SmaI map of the cloned BamHI joint, the source of DNA, was not known. Therefore, the first task was to identify the fragments of interest among the various SmaI fragments generated by cleavage of pSVODK6 with SmaI. These were the SmaI A, F, G, E, and I fragments, which span and flank the "a" sequence (see Fig. 14). The HinfI joint spanning fragment A contains all of these sequences (except SmaI E - HinfI cuts E in half). The HinfI joint was identified on acrylamide gels by comparing HinfI digests of various plasmids containing joint spanning sequences, including pINJ2 (SacI joint inserted into the SacI site on pTKex1 - see Smiley et al, 1981), p27K, and pSVODK6

FIGURE 14

Map of KOS junction. This map was prepared by Wagner and Summers (1978). Underneath the SmaI map are the positions of the "b", "a", and "c" sequences with respect to the SmaI sites.


AluI
AluI
1
2
3
kb
TaqI
A
C
B
D
HinfI
C
A
B
SmaI
H
D
E
G
I
F
A
J
C
B
"b"
"a"
"c"
BglI
C
B
A
H
G
F

(see appendix). The pattern produced by HinfI digestion of these plasmids indicated that the HinfI joint from pSVODK6 comigrated with another, unknown, fragment which could be reduced in size by SacI. Thus, SacI/HinfI digestion of pSVODK6 produced a joint spanning fragment which could be separated cleanly from other contaminating fragments on an 8% polyacrylamide gel.

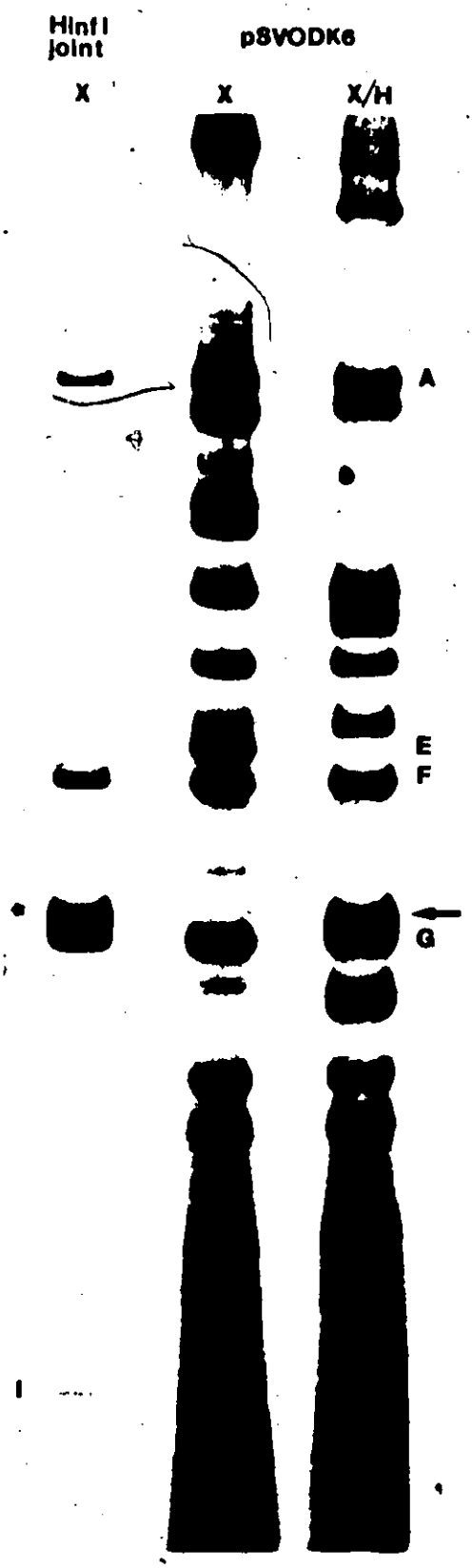
The HinfI joint from pSVODK6 was eluted by the method devised by Maxam and Gilbert (1980) and redigested with XmaI. XmaI is an isoschizomer of SmaI which makes staggered cuts producing 5' overhanging termini, while SmaI makes flush ended cuts. The latter are more difficult to label in vitro with T4 polynucleotide kinase than the former. The XmaI digested DNA was treated with calf intestinal alkaline phosphatase to remove terminal 5' phosphate groups, and then labelled with ^{32}P at the 5' ends with T4 kinase and γ-^{32}P-ATP. Electrophoresis of this DNA with kinased XmaI and XmaI/HinfI digests of pSVODK6 allowed comparison of the known SmaI (XmaI) restriction pattern of the HinfI joint with the unknown pattern of the BamHI joint in pSVODK6 (Fig. 15). Several observations could be made from this gel:

1) While the SmaI (XmaI) A, F, G, I, and terminal XmaI/HinfI fragments (asterisk beside larger one) could be identified in the purified HinfI joint, the presence of several minor bands indicated that the DNA was not digested to completion. This problem is even more evident in the pSVODK6 total DNA digests and was a persistent problem throughout this exercise.

2) pSVODK6 DNA contained a large amount of contaminating RNA which appears as a smear at the bottom of the gel. The SmaI I fragment is

FIGURE 15

Identification of SmaI fragments from pSVODK6. The HinfI joint was purified, digested with XmaI, and end labelled with polynucleotide kinase. pSVODK6 was digested with XmaI or XmaI/HinfI and similarly end labelled. The labelled DNAs were electrophoresed through an 8% polyacrylamide gel and exposed to X-ray film. The strong signals in the SmaI digest of the HinfI joint were identified as the SmaI A, F, G, I, and cleaved E fragments. These comigrated with fragments in the SmaI digest of pSVODK6, allowing identification of the fragments among a set of fragments of unknown map position. The E fragment in the XmaI digest was identified by its absence from the HinfI joint digest, by its disappearance upon digestion with HinfI, and by its relative size. The arrow and asterisk indicate the HinfI cleaved SmaI E fragment in the pSVODK6 digest and the HinfI joint digest, respectively.


HinfI
joint
pSVODK6
X
X
X/H
A
E
F
G

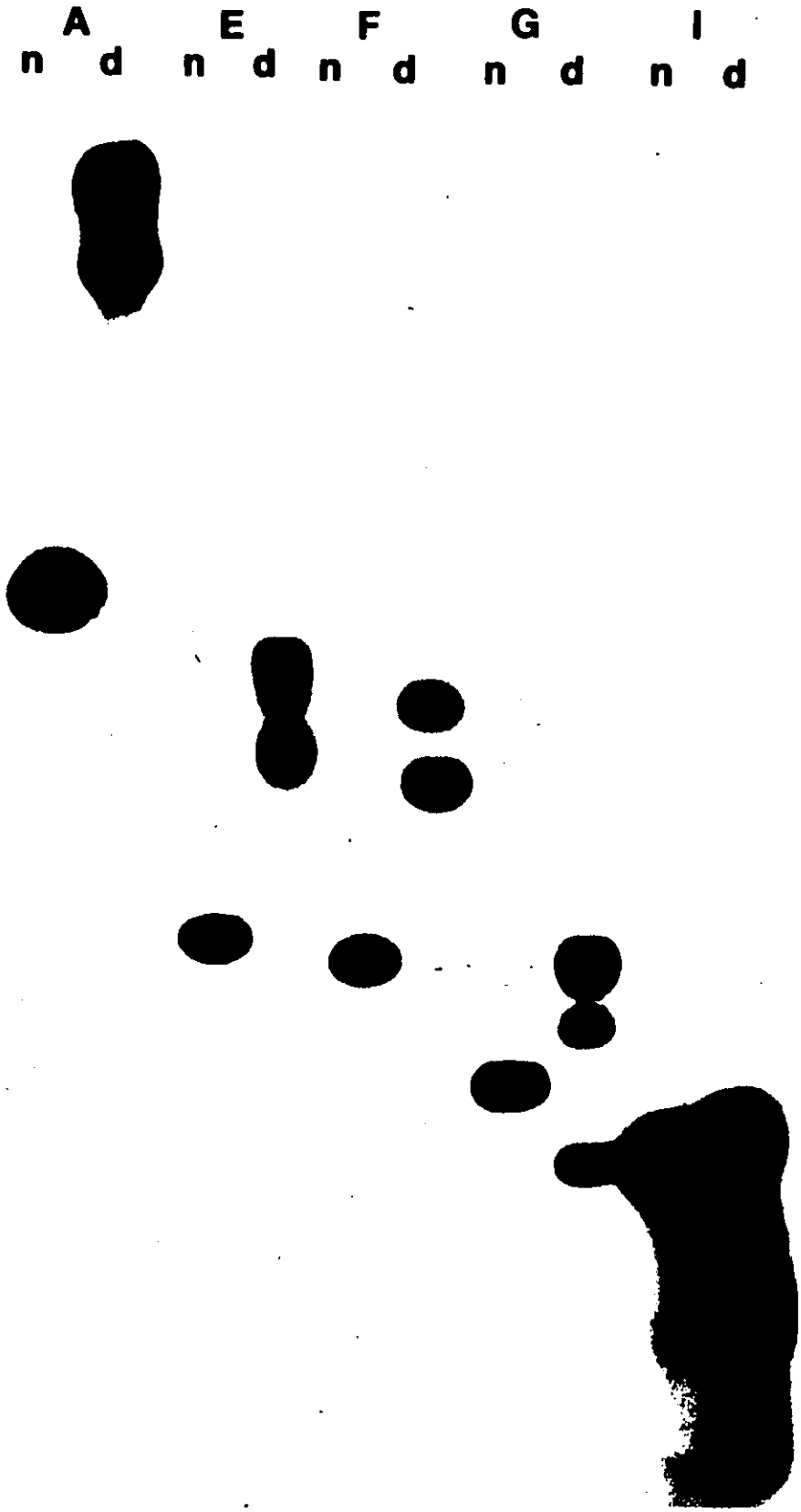
buried in this mess.

3) The SmaI E fragment was tentatively identified by its size and by its possession of a HinfI site - it is slightly larger than SmaI F (see middle track) and is reduced in size by HinfI (see arrow right hand track).

Having identified the fragments of interest from an XmaI digest of pSVODK6, the next step was to determine whether or not they could be efficiently strand separated. End labelled fragments A, F, E, G, and I were eluted from a polyacrylamide gel and heated in the presence of dimethylsulfoxide (DMSO) as described by Maxam and Gilbert (1980). The denatured DNA was electrophoresed beside native marker samples in a strand separating gel (Fig. 16). All of the fragments appeared to separate quite nicely. In particular, the F fragment routinely produced two sharp bands after denaturation. SmaI (XmaI) G produced four separate bands. This fragment regularly labelled twice as efficiently as any of the other bands, as determined by Cerenkov radiation, leading me to believe that it was a mixture of two different fragments of equal size. This suspicion was later confirmed by sequence analysis of the four separated strands. One of them was indeed SmaI (XmaI) G, while the other is an unknown fragment. SmaI (XmaI) I apparently produced two separate bands. However, this fragment also routinely labelled more efficiently than the others. This was attributed to the contaminating RNA in that region of the gel (see Fig. 15). However, subsequent attempts to sequence these fragments were unsuccessful as it was found that SmaI (XmaI) I was also contaminated with another DNA fragment of equal size (see below). Although SmaI (XmaI) A strand separated well on

FIGURE 16

Strand separation of joint fragments. 5' end labelled SmaI A, E, F, G, and I fragments were eluted from a gel containing a SmaI digest of pSVODK6 DNA, and separated in the presence of DMSO at 90°C. One third of each end labelled fragment was retained as a non-denatured marker. The non-denatured (lanes marked n) and denatured (lanes marked d) DNA were electrophoresed through a 5% strand separating gel, and exposed to X-ray film. The bands were numbered according to their rate of migration, with the slowest band always being number 1. Thus, the SmaI F separated strands were designated F1 (top band) and F2 (lower band). The four SmaI G bands were designated G1, G2, G3, and G4 in their descending order on the strand separating gel. This protocol was observed for all strand separations. The first two strands of SmaI G appear as a single band in this gel. However, subsequent preparations revealed two bands which migrated very close to each other. Their separate identity was confirmed by the sequence analysis.


A
E
F
G
I
n d n d n d n d n d

this gel, it did not do so in later attempts, and therefore a different strategy was adopted in determining the DNA sequence of this fragment (see below).

Separated end labelled strands of the various fragments were subjected to four of the chemical modification/cleavage reactions devised by Maxam and Gilbert (1980) - the G, G+A, T+C, and C specific reactions - and electrophoresed through 20% and 8% polyacrylamide sequencing gels. Two 8% gels were prepared so that replicate samples could be electrophoresed to different distances. This allowed the confident determination of up to 150 bp of DNA sequence. The samples were electrophoresed only part way through the 20% gels to allow for sufficient overlap of sequence information between gels.

Figures 17 to 19 show sequence ladders for several fragments. Each one was sequenced several times to ensure confidence in the data. Figures 20 and 21 summarize the data garnered from these (and other, similar) gels. Several observations may be made from these (somewhat imperfect) data. No overlap between E1 and E2 could be found, probably because of their size. Moreover, ambiguities resulting from suboptimal sequence reactions failed to turn up the expected HinfI site in either strand, although the sequence GACCC 115 bp from the 5' end of E1 is a good candidate since it is approximatey in the correct position and would merely reflect a poor T+C reaction, a common occurence. Since SmaI E was not of sufficient interest to this project to warrant more careful analysis, no further attempts were made to resolve these problems.

As mentioned above, SmaI G strand separated into four fragments.

FIGURE 17

Sequence gels of G1, G2, G3, and G4. The four SmaI G strands were subjected to Maxam and Gilbert (1980) sequence reactions and electrophoresed through 20% and 8% sequence gels. Above each track is a letter designating the base specific reaction. The panel on the left shows the 20% gel, in the middle is the short 8% gel, and on the right is the long 8% gel.

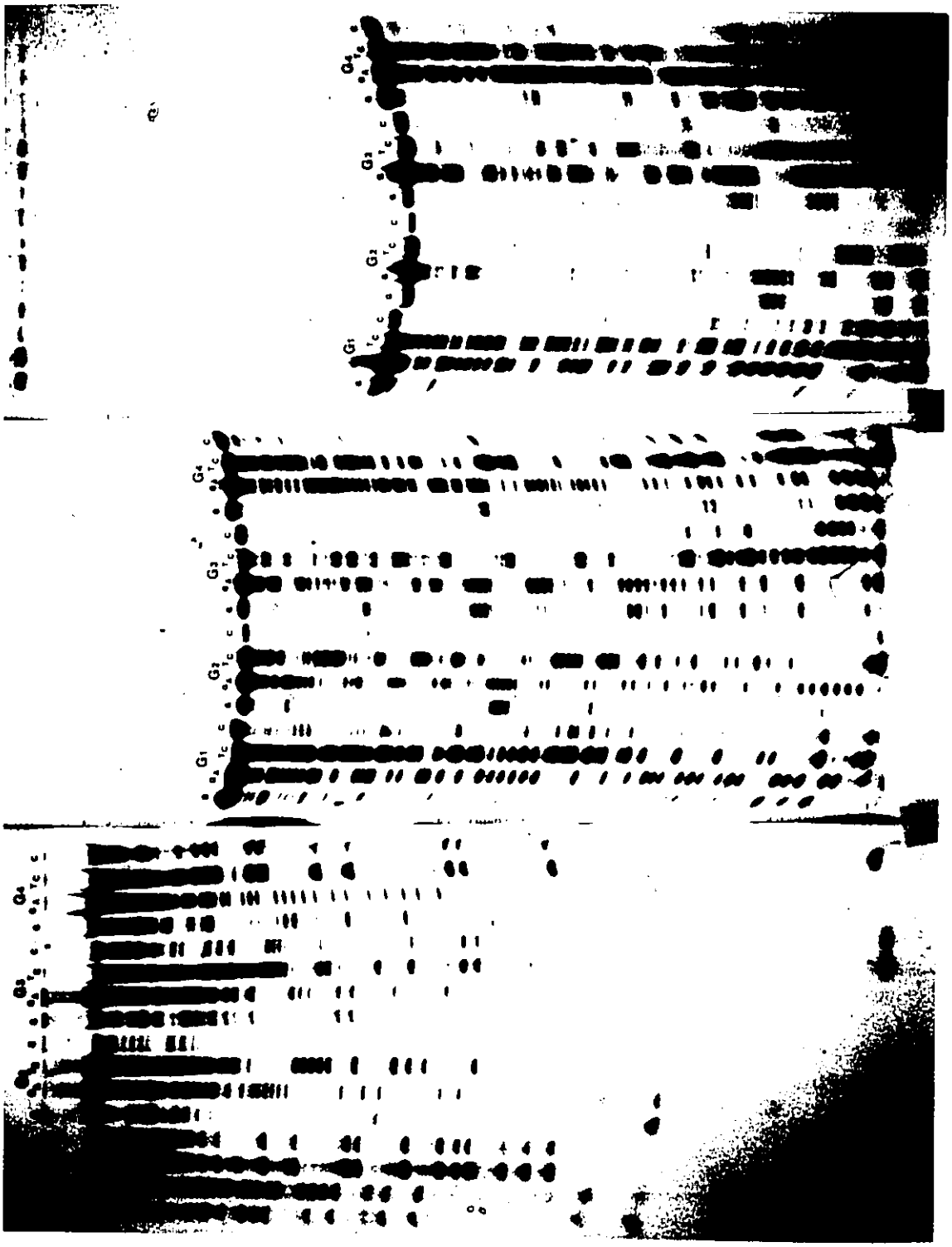


FIGURE 18

Sequence gels of F1, F2, and S terminus from pTKSBS7. SmaI F was end labelled and strand separated, generating F1 and F2. The BamHI S terminus from pTKSBS7 was end labelled at the BamHI sites and redigested with HinfI. The 1300 bp HinfI terminal fragment (S) was eluted from an 8% acrylamide gel. F1, F2, and S were subjected to Maxam and Gilbert (1980) sequence reactions and electrophoresed through 20% (left hand gel) and 8% (middle and right hand gels) polyacrylamide sequence gels.

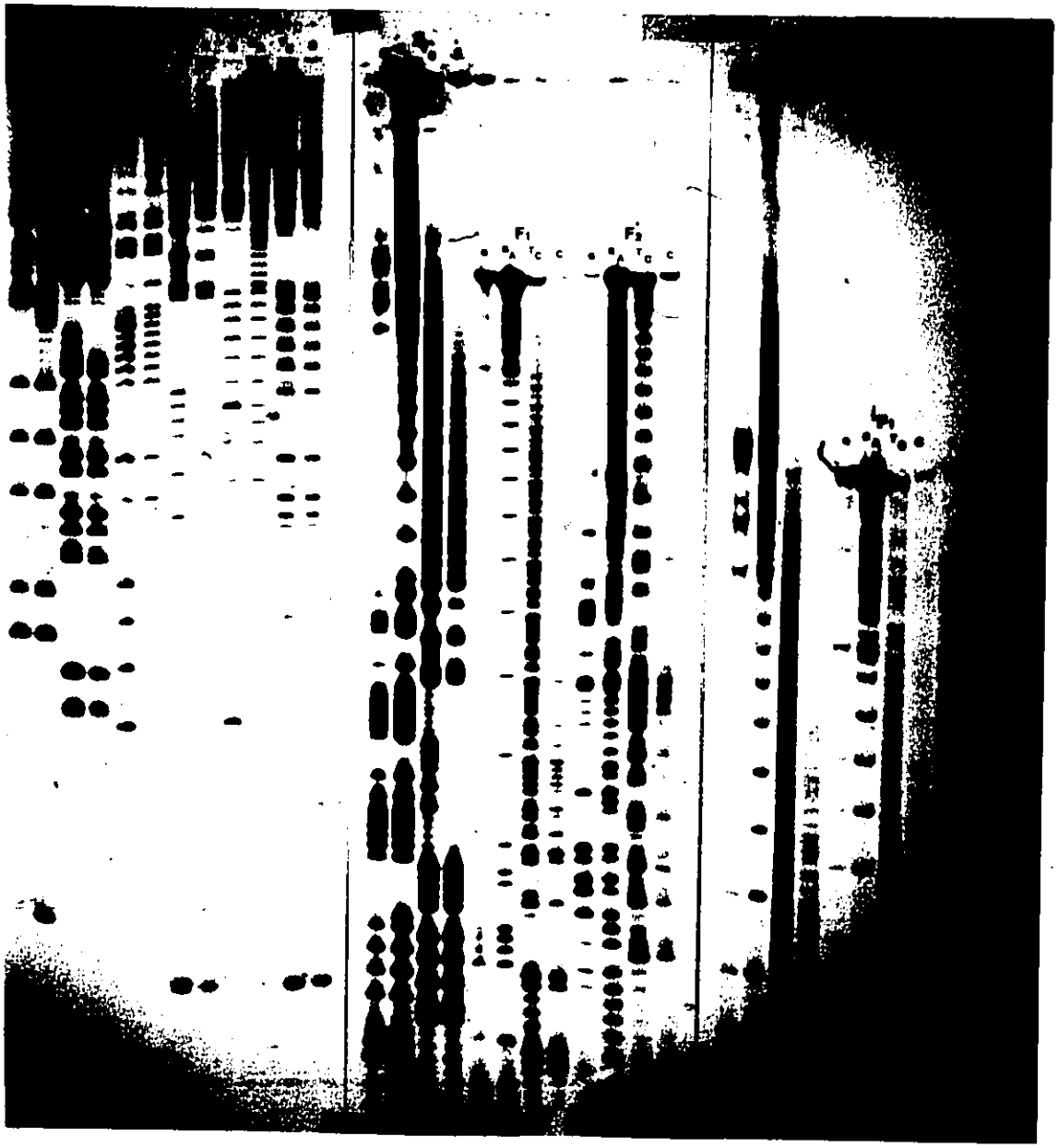

F1
A T C C
F2
A T C C

FIGURE 19

Sequence gels of SmaI A, E1, and E2. SmaI E was end labelled and strand separated to produce E1 and E2. SmaI A was end labelled and redigested with BglI. The large subfragment was used in the sequence reactions after elution from an 8% polyacrylamide gel. The arrow on the left indicates the questionable T_C in DR1.

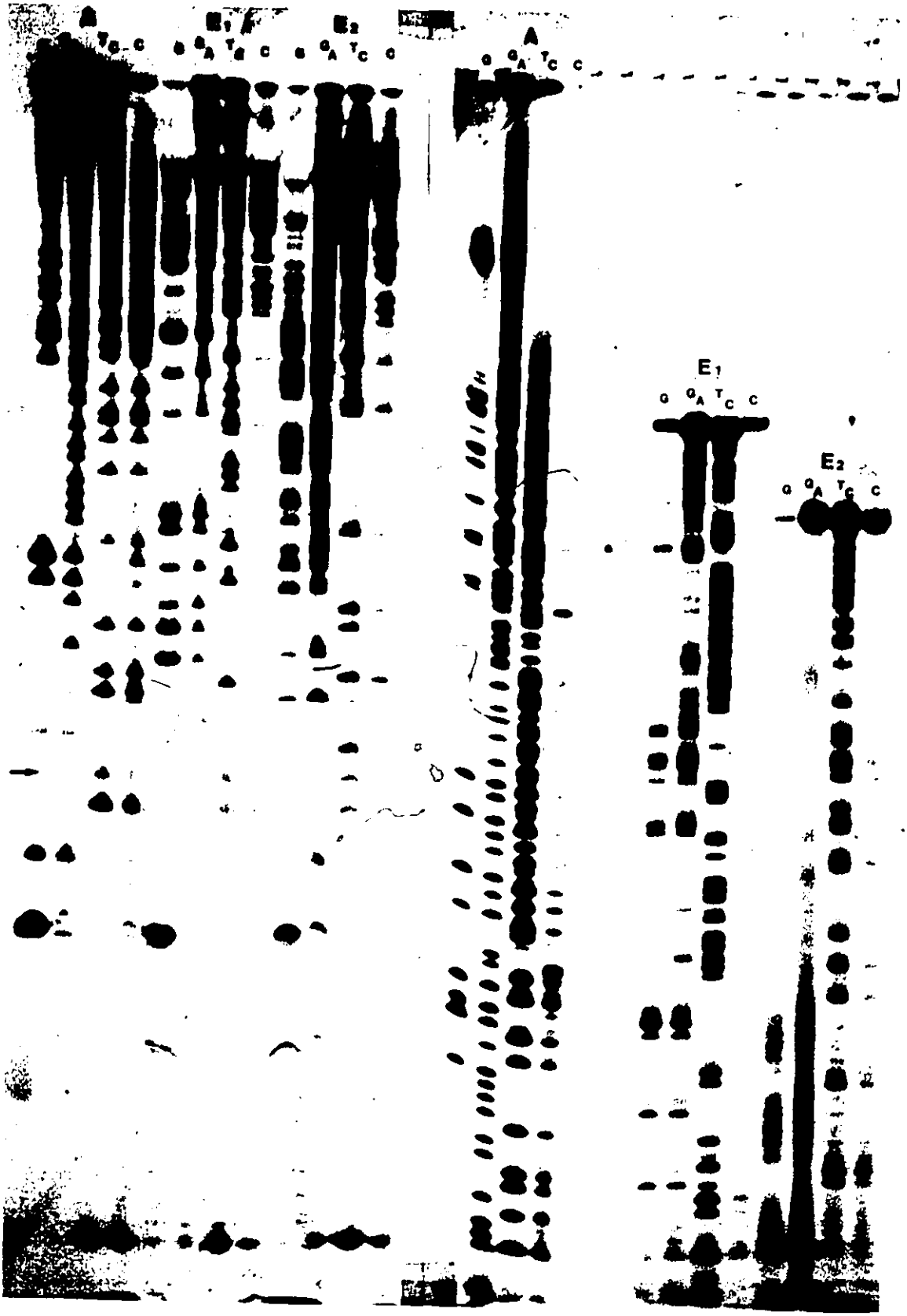

E1
E2
A
G G A T C C
E1
E2
G G A T C C

FIGURE 20

Sequences of E1, E2, F1, F2, G1, G2, G3, and G4. The sequences shown here were deduced from the sequence gels shown in Figures 17-19, and from other gels (data not shown). The first 14 bp of F2 and G1 (underlined) correspond to 14 bp of DR1 from strain F. The sequence underlined in G4 is the BglI site present in SmaI G. G1 and G4 are complementary with each other, while G2 and G3 are complementary with each other. The asterisk above the A in F2 indicates an anomalous sequence which may be subject to compression (see text).

Sequence of E1

CGGGCCCACG GGCGCAATCC CAACCGCACA GTCCTAGGTA ACCTCCACGC CCAACTCGGA
ACCCGTGGTG CAGGAGCGCG CCCGCGGCCG CCCCGCCGCC GCCCCCCGCC AGCGGACCCC
CGCCCNCCCG CGCCGCCGCC GCGCCAG$^{T}_{C}$GG CCCACG...

Sequence of E2

CGGGCCCGCG CCGGGGGGTG GGGACCGGGG GCGGGGAACG GCAACGGTGG ACCGGACCTG
CTGGC$^{G}_{A}$CCGG CCGGGCGGGG GGTCGNCCGG CCAGTCGCGC GCGCGCGGGC GCAGG
CCAGGGACGC...

Sequence of F1

CGGGGTGTGC GCGTTTTGGG GGGGGCCCGT TTCCGGGGTC TGGCCGCTCC TCCCCC
$(GCTCCTCCCCCC)_9$

Sequence of F2

CGGGCCCCCC GCGGGCGCGC GCGCGCGCAA AAAAGGCGAG CGGCGGTCCG AGCAACGCGC
GCGCGCAGAC ARGCGCCGGG GGCAGGACGC GGGAACGGGGGG...iteration

Sequence of G1

CGGGCCCCCC GCGGCCGAGA CTAGGGAGTT AGACAGGCAA GCACTACTCG CCTCTGCACG
CACATGCTTG CCTGTCAAAC TCTACCACCC CGGCACGTTC TGTCTCCATG GCCCGCCGCC
GCCACGCGG

Sequence of G4

CGGGCGGCCG GGGGCGGCGG GGGCCGCGAT GGCAGGCGCG GGCCATGGAG ACAGAGAACG
TGCCGGGGTG GTAGAGTTTG ACAGGCAAGC ATGTGCGTGC AGAGGCGAGT AGTGCCGCCG
CAA..

Sequence of G2

CGGGATCAAC CCGGCGCCCC CAAAGAATAT CATTAGCATG CACGGCCCGG CCCCCGATTT
GGGGAACCAA CCCGGTGTCC CCCAAAGAAC CCCATTAGCA TGCCCCTCCC GCCGACGCAA
CAGGGG..

Sequence of G3

CGGGGCACCG ACGCAGGCCA AGCCCCTGTT GCGTCGGCGG GAGGGGCATG CCAATGGGGT
TCTTTGGGGG ACACCGGGTT GGTTCCCCAA ATCGGGAACC GGGCGTGCAGC...

FIGURE 21

Sequences of SmaI A, I1, I2, and S. The sequences here were read from gels in Figures 18, 19, and 23. The first 6 bp of the S terminus (underlined) were derived from the BamHI linker used to clone the S terminal fragment. Also underlined in this sequence is the SmaI site which separates SmaI I and SmaI F. The 8 bp underlined at the beginning of SmaI A are the part of DR1 which remains after SmaI (XmaI) digestion.

Sequence of S terminus

AICCGGCCCG CCGCCCGCCT.TTAAAGGGCC GCGCGCGACC CCCGGGGGGT GTGTTTTGGG

GGGGGCCCGT TTCCGGGGTC TGGCCGCTCC TCCCCCGCTC CTCCCCCCGC TCCTCCCCCC

....iteration

Sequence of SmaI A

?

CGGGCCGCCA CAGGCAAAAC AACACCAACA AAGCACGGCG CAATCCGCAC GTCACACGTC

ACGTCACCCA CCACACCCGC CCAACAACAC AACTCACAGC GACAACTCAC CG...

Sequence of Large HaeIII Fragment of SmaI I

...GGC$^{T}_{C}$GCCCGCCGCCCGCCTTTAAAGG...

Sequence of Small HaeIII Fragment of SmaI I

...GGGGCGCGCGCG...

The sequences of these four strands indicated that G1 and G4 were complementary, while G2 and G3 were complementary. The presence of the BglI site close to the 5' end of G4 (underlined) indicated that it was SmaI G. This deduction was further confirmed by the presence of 14 bp of DR1 at the 5' end of G1, a feature expected of SmaI G since it flanks SmaI I. This latter fragment is wholly contained within the "a" sequence. Since DR1 itself contains a SmaI site (see Fig. 14), digestion of joint sequences with this enzyme would be expected to place parts of DR1 on adjacent fragments. G2-G3 is an unknown SmaI fragment from the BamHI joint.

F1 and F2 were also too large to obtain satisfactory overlapping sequences. However, the gels indicated a qualitative feature that confirmed their complementarity, namely, the presence in both sets of sequence ladders of blocks of repeats. The first block in F1 was deduced to be the sequence GCTCCTCCCCCC. This sequence, or one very similar to it, was repeated 10 times in F1. F2 also contained a block of 11 - 12 bp which was repeated 10 times. Although the actual sequence was impossible to decipher from the gels, it was clearly composed largely of Gs, and was probably the complement of the repeat in F1.

F2 contained 14 bp of DR1 at it's 5' end. Indeed, it is obvious that these 14 bp are identical with the first 14 bp of G1, the strand abutting SmaI I and containing most of the DR1 derived from the "b" side of "a". Since DR1 is a direct repeat flanking "a", F2 must represent the strand of SmaI F which directly abuts SmaI A at the other end of "a". The sequence of the rest of F2 should therefore contain the Uc sequences of "a" (Fig. 13), and indeed the homology with strain F Uc

is very high. F1, the putative complement of F2, should contain sequences homologous with Ub from strain F, and it does. Thus, the orientation of F1 and F2 can be determined by analogy with their homologues in strain F.

The asterisk beside the string of As in F2 (Fig. 20) indicates a puzzling, but reproducible, anomaly. These fragments always resolved into six bands on the 20% gels, but only five on the 8% gels, and must represent some form of compression in the latter. The issue was resolved later by sequencing the opposite strand from a subclone of SmaI F (see below).

Attempts to sequence SmaI I proved fruitless as it was discovered that the DNA was contaminated with an unknown comigrating fragment. Thus, another strategy was employed to fill in the gaps in the KOS "a" sequence data. The KOS S terminal clone, pTKSBS7 (see appendix) was digested with BamHI, electrophoresed through a 5% acrylamide gel, and the S terminus was eluted. After labelling the 5' ends of this fragment, the DNA was digested with HinfI, electrophoresed through an 8% acrylamide gel, and the S terminus was again eluted. This DNA was subjected to Maxam and Gilbert sequence reactions and electrophoresed through the usual complement of sequence gels (Fig. 18). The sequence of the S terminus deduced from these gels is shown in Fig. 21. The terminus was cloned after ligating BamHI linkers to viral DNA. The linker sequence is CCGGATCCGG. The part of this sequence evident from the sequence gels is underlined in Fig. 21. Also underlined is the SmaI site separating the part of SmaI I still present on the S terminus and SmaI F. Note that the sequence of this strand of the S terminus is

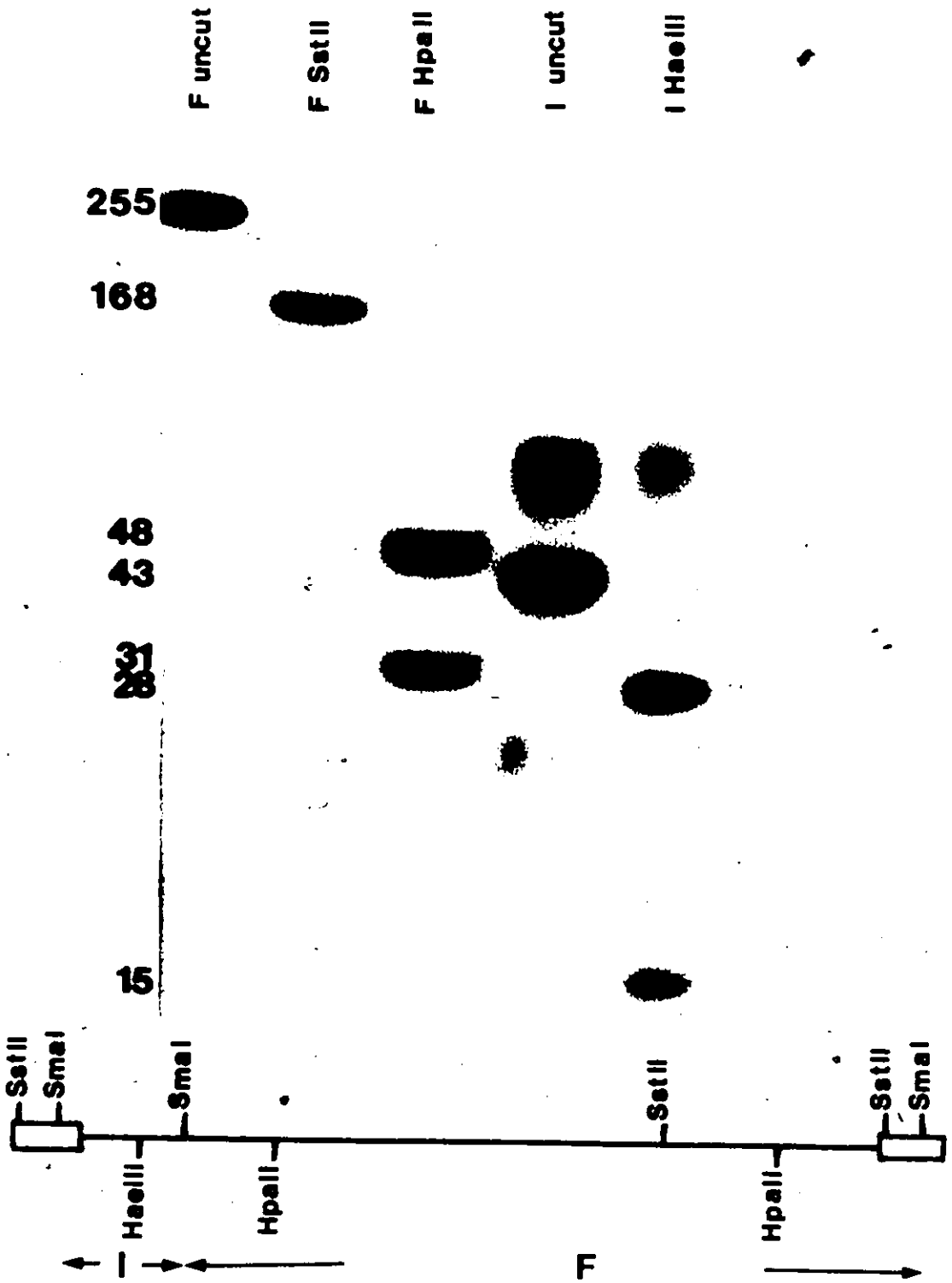
identical with that of F1, as expected from the deduction of the orientation of F1 and F2 with respect to flanking sequences.

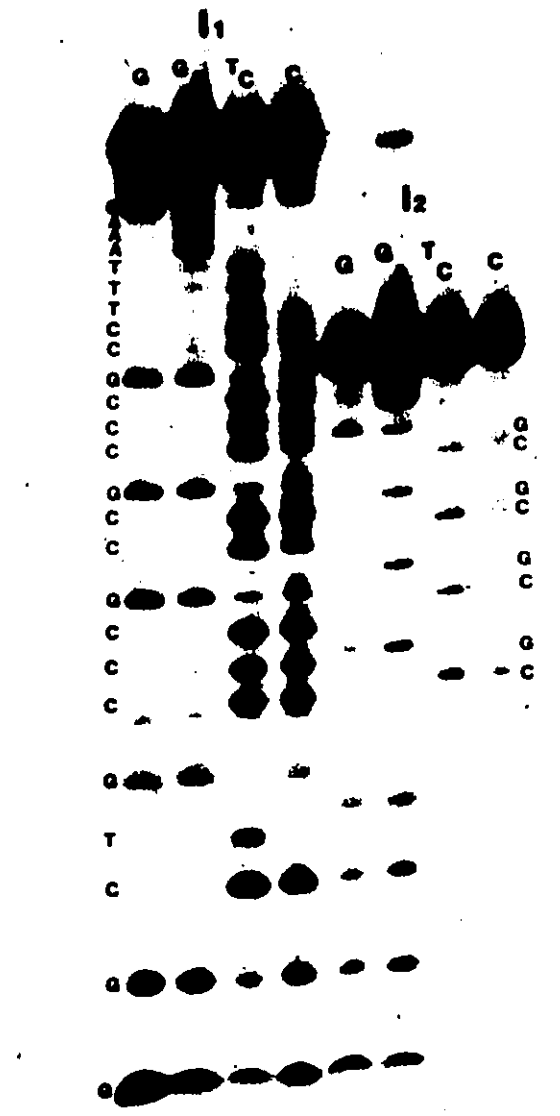
SmaI A proved difficult to strand separate and so a different strategy was adopted in this case as well. The XmaI A fragment was end labelled at the 5' ends and digested with BglI to remove the unwanted end. The sequence ladders are shown in Fig. 19, and the deduced sequence is displayed in Fig. 21. The arrow in Fig. 19 highlights an ambiguous C. This C is part of the DR1 flanking "a" at the border with "c", and may in fact be a T, as would be predicted both by the sequence of DR1 at the "b" border (see below), and by the sequence of DR1 in strain F. This should be confirmed by analysis of the complementary strand, a not untrivial task as there are very few appropriate restriction sites in SmaI A.

The only part of the KOS "a" sequence not determined at all were the four base pairs of SmaI I between the end of the S terminus and the border with SmaI G. Attempts to strand separate SmaI I were unsuccessful, and so an alternative approach was adopted. XmaI I was end labelled at the 5' ends and digested with HaeIII, an enzyme which makes an asymmetric cut in SmaI I (see Fig. 22). Electrophoresis of this DNA on a 20% acrylamide gel revealed, in addition to the two predicted fragments from SmaI I, a fuzzy band migrating above the uncut SmaI I marker. This may have been the contaminating element in previous unsuccessful attempts to sequence this fragment. The two HaeIII subfragments of SmaI I were eluted and sequenced. Fig. 23 shows that these fragments were either contaminated with unknown DNA or that the sequence reactions were exceptionally nonspecific. Given the previously

FIGURE 22

Restriction analysis of SmaI F and SmaI I. SmaI F and I were end labelled and redigested with various enzymes as indicated, electrophoresed through a 20% polyacrylamide gel, and exposed to X-ray film. The map of the "a" sequence below the gel summarizes the data. The open boxes represent DR1. The line underneath the map indicates the positions of SmaI I and F within the "a" sequence. The numbers on the left of the gel are the sizes of the various fragments (in bp). The SmaI I fragment (43 bp) was contaminated with other species of nucleic acids (smears above and below).


F uncut
F SatII
F HpaII
I uncut
I HaeIII
255
168
48
43
31
28
15
SatII
SmaI
SmaI
SatII
SatII
SmaI
HaeIII
HpaII
HpaII
I
F


I1
I2

described uncertainty about the purity of SmaI I, the former explanation seems most likely. Both fragments produced sequence ladders of a major species and contaminating sequence ladders of a minor species. The sequences of the major species agree well with the sequence of this fragment determined from the S terminus (Fig. 21), leading me to believe that they are the legitimate sequence ladders of the HaeIII subfragments of SmaI I. This deduction presents something of a dilemma concerning the sequence of DR1 in KOS. If the major ladders in Fig. 23 are the legitimate sequence ladders for SmaI I, then the sequence of DR1 at the border with SmaI G would be "CCGCGGGGGGCCCGGGCTGC", which is identical with the sequence of DR1 in strains F, 17, and USA8. However, it may not be identical with the sequence of DR1 deduced from the sequences at the border between SmaI F and A, which suggest that the T may be a C, although, as pointed out above, the C in SmaI A is somewhat ambiguous. If these data are to be accepted at face value, either DR1 in KOS is different from DR1 in other strains of HSV-1, or the two copies of DR1 at each end of "a" are different in KOS. Of course, the sequences from these regions should be verified by looking at both strands before concluding finally that either situation is correct. Davison and Wilkie (1981) found that the two copies of DR1 flanking the "a" sequence in their clone of the USA-8 junction differed by one base pair.

Scrutiny of the deduced sequence of the KOS SmaI F fragment indicated that it differed from strain F at several places, including three restriction sites. One of these was an SstII site within the Uc sequence which was apparently missing from KOS; the other two were both

HpaII sites, one in Ub and the other in Uc, which were present in KOS but not in strain F. End labelled SmaI F was digested with SstII and HpaII and electrophoresed through a 20% acrylamide gel (Fig. 22). The apparently missing SstII site was in fact present, and the sequence of SmaI F was changed accordingly. The HpaII site in Ub which is diagnostic of KOS was there. These differences were subsequently confirmed by sequence analysis of subclones of SmaI F, which were made and analysed by Marg Howes in the laboratory. These subclones, f1 and f2, were derived by cleaving a SmaI F clone (pUC7-F, see appendix) with SstII (pUC7-f1, pUC7-f2, see appendix). Her data indicated that a third HpaII site originally observed in Uc, which was diagnostic of KOS, was an artifact.

By combining all of the data generated here and elsewhere in the lab, a composite sequence of the KOS joint region can be deduced (Fig. 24). The differences between KOS and strain F are highlighted, as are other relevant landmarks, particularly the position of the S terminal cleavage site as determined by sequence analysis of the cloned S terminus (see above). The DR2 repeats in "a" are bracketed. Note that some of them consist of "GCTCCTCCCCC", while others are "GCTCCTCCCCCC", a difference of one base pair. At this point it is expedient to speculate on the nature of the F/F* fragments observed by Wagner and Summers (1978). These are probably variants of "a" which contain different numbers of the repeat DR2. Evidence provided in a later section indicates that KOS probably contains "a" sequences with 9 or 10 repeats of DR2.

FIGURE 24

Sequence of KOS joint region. The sequences in Figures 20 and 21 were compiled such that they are arranged in the order "G-I-F-A", or "bac". In some cases the inverse complement of the sequence was deduced (G1 and F2). Some of the sequences were confirmed later by analysis of the cloned SmaI F fragment in pUC7-F (see appendix), and by analysis of the SmaI F subclones, pUC7-f1 and pUC7-f2. This latter work was performed by Maria Booi and Marg Howes. See Figure 45 for a diagram of the sequence strategy. Asterisks highlight bases which differ from strain F. DR1 is underlined, and the S and L terminal cleavage sites are indicated. The L terminal cleavage site is assigned through analogy with strain F, while the S terminal cleavage site is assigned on the basis of the ultimate junction base in the cloned S terminus, adjacent to the BamHI linker in Fig. 21.

```
CCCGGGCGGC CGGGGGCGGC GGGGGCCGCG ATGGCAGGCG CGGGCCATGG AGACAGAGAA
                  20                  40                  60

CGTGCCGGGG TGGTAGAGTT TGACAGGCAA GCATGTGCGT GCAGAGGCGA GTAGTGCTTG
                  80                 100                 120

CCTGTCTAAC TCGCTAGTCT CGGCCGCGGG GGGCCCGGGC TCGCCCGCCGC CCGCCTTTAA
                 140                 160                 180

AGGGCCGCGC GCGACCCCCG GGGGGTGTGT TTTGGGGGGG GCCCGTTTCC GGGGTCTGGC
                 200                 220                 240

*
C(GCTCCTCCC CC)(GCTCCTCC CCCC)(GCTCCT CCCCCC)(GCTC CTCCCCCC)(GC
                 260                 280

TCCTCCCCCC) (GCTCCTCCCC CC)(GCTCCTCC CCC)(GCTCCTC CCCC)(GCTCCT
      300                 320                 340

                                                              **
CCCCC)(GCTCC TCCCCC)GCTC CCGCGGCCCC GCCCCCAACG CCCGCCGCGC GCGCGCGCGC
                 360                 380                 400

     ***** *******                                          ↓
CGCCCGGACC GCCGCCCGCC TTTTTTGCGC GCGCGCGCGC CCGCGGGGGGG CCCGGGCCGC
                 420                 440                 460

CACAGGCAAA ACAACACCAA CAAAGCACGG CGCAATCCGC ACGTCACACG TCACGTCACC
                 480                 500                 520

CACCACACCC GCCCAACAAC ACAACTCACA GCGACAACTC ACCG...
                 540                 560
```

Analysis of the Packaging Signal in SmaI F

Two different lines of evidence had accumulated by this time to indicate that the "a" sequence was invested with interesting biological activities. As mentioned above, there was circumstantial evidence to support the notion that "a" might be a cis-acting signal for site specific recombination, and indeed most people in the field still believe this idea to be true. There was also much more compelling evidence indicating that the "a" sequence contained the packaging signal for HSV-1 DNA. Studies with defective viruses, both naturally occurring and engineered (see Introduction) had indicated that a cis-acting signal residing in the viral termini was required for packaging of viral DNA (Vlazny and Frenkel, 1981; Vlazny et al, 1982; Stow et al, 1983). If you recall, in the Introduction I pointed out that the two alternative cleavage phases embodied in any concatemer of HSV DNA would be sufficient to generate two out of four of the genome isomers. Only one segment inversion event such as recombination or intermolecular ligation is required to provide the substrates for the other two isomers, assuming that both cleavage phases are equally active. Thus, cleavage events could generate the same products as recombination events, and are indistinguishable at the level of the final DNA products.

Were the two "activities" associated with "a" (recombination and packaging) related, and could they each be mapped to a subfragment of "a"? The approach taken to address these questions was to subject various subfragments of "a" to two different assays:

1) The test fragments were cloned on plasmids containing an HSV origin

of replication and cotransfected with viral DNA into Vero cells. The resulting virus stocks were then examined for the replication and packaging of the plasmids as defective genomes. This assay tested for the ability of the test fragment to act as a packaging signal.

2) The test fragments were reintroduced into viral DNA at the TK locus by marker rescue (see Results section "Sequence Identity of "a" is not Obligatory"). Viral DNA from mutants containing the test fragments at TK was examined for the ability of the fragment to drive segment inversion. This tested the recombinational activity of the test sequences in addition to the cleavage/packaging activity.

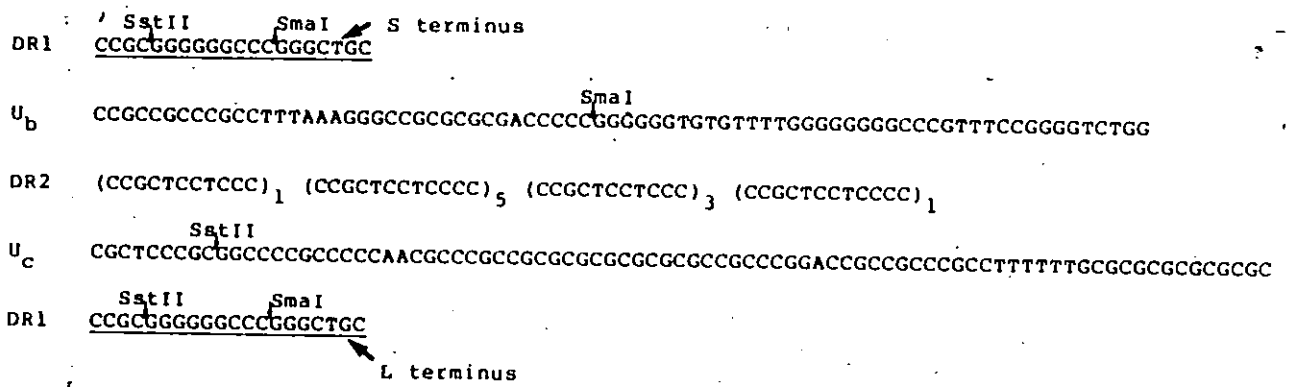
Details of the construction of the various plasmids used in this study, including the cloning procedure used to obtain an HSV origin fragment, are in the appendix. Briefly, various SmaI fragments from pSVODK6 were cloned into the HincII site on pUC-7. HincII, which makes flush ended cuts, is part of the palindromic linker in pUC-7. The HincII sites in the vector are flanked by BamHI and EcoRI sites. Thus, any sequence cloned into the HincII site(s) may then be excised with BamHI or EcoRI. SmaI A, F and I fragments were inserted into pUC-7. A synthesized DR1 (strain F sequence) was also cloned into pUC-7 (Varmuza and Smiley, 1985). Additionally, two subfragments of SmaI F derived by cleavage of the latter with SstII were also inserted into pUC-7. The flanking BamHI sites on these clones allowed me to subclone all of these fragments into pTKSB, a plasmid containing a mutant TK gene in which 200 bp of coding sequence had been replaced with a BamHI linker (see appendix). The pTKSB subclones were used in the marker rescue experiments. After inserting a HindIII origin-containing fragment on

FIGURE 25

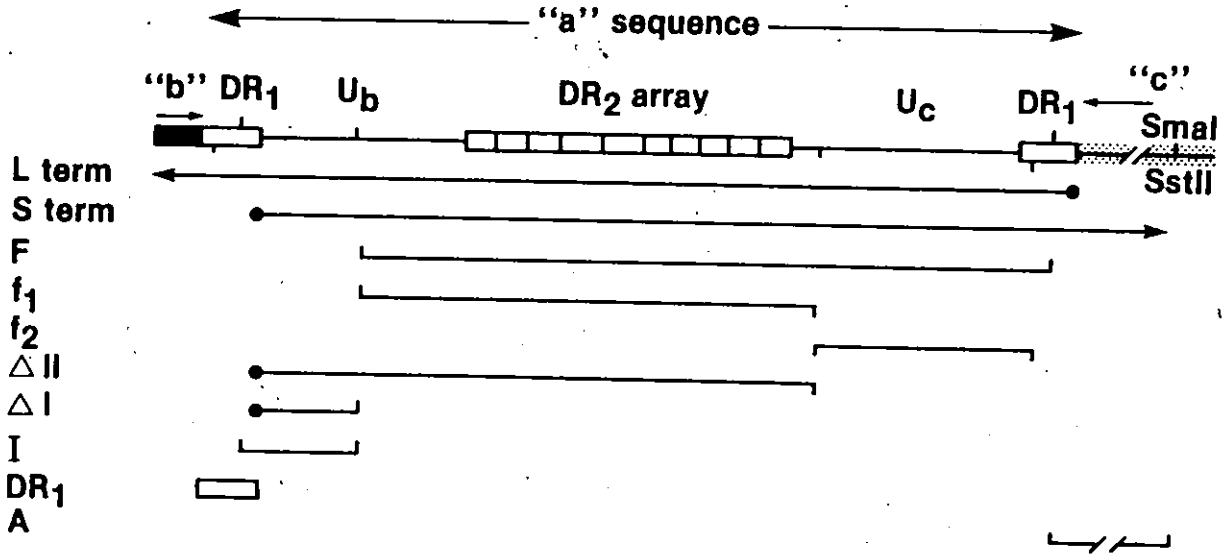
(A) Sequence of the "a" sequence from HSV-1(KOS). This sequence was presented in Figure 24, but has been rearranged here to conform to the designations made by Mocarski and Roizman (1982a), and diagrammed in Figure 13.

(B) Subclones of the "a" sequence tested for packaging activity. The upper line of the figure shows a schematic represehtation of the "a" sequence. The flanking "b" and "c" regions are shown as solid and stippled bars respectively; DR1 and DR2 are shown as open boxes. SmaI sites are shown as vertical lines above the diagram, and SstII sites are displayed below. The various subfragments tested are shown on the lines below. Solid circles indicate viral termini delineated by BamHI linkers, and arrows indicate that the clones extend off the map. SmaI A, 927 bp long, is not drawn to scale. Details of the construction of the various substrates may be found in the appendix: L term (pbaDI), S term (pcaDI), F (pSBF14ori, pSBF14), f1 (pSBf1), f2 (pSBf2), ΔII ($pHR_{\Delta}IIori$, $pSB_{\Delta}II$), ΔI (pHR_{Δ} Iori), I (pSBIori, pSBI), DR1 (pSBDRori, pSBDR), A (pSBA3ori, pSBA3). The "ori" containing plasmids were used in the defective assay, and the other plasmids were used in the marker rescue experiment.

A


SstII
SmaI
S terminus
DR1
CCGCGGGGGGCCCGGGCTGC
SmaI
Ub
CCGCCGCCCGCCTTTAAAGGGCCGCGCGCGACCCCCGGGGGGTGTGTTTTGGGGGGGGCCCGTTTCCGGGGTCTGG
DR2
(CCGCTCCTCCC)1 (CCGCTCCTCCCC)5 (CCGCTCCTCCC)3 (CCGCTCCTCCCC)1
SstII
Uc
CGCTCCCGCGGCCCCGCCCCCAACGCCCGCCGCGCGCGCGCGCGCCGCCCGGACCGCCGCCCGCCTTTTTTGCGCGCGCGCGCGC
SstII
SmaI
DR1
CCGCGGGGGGCCCGGGCTGC
L terminus

B


"a" sequence
"b" DR1
Ub
DR2 array
Uc
DR1
"c"
SmaI
SstII
L term
S term
F
f1
f2
△II
△I
I
DR1
A

the pTKSB subclones, they could then be used in the defective assay. Several other cloned fragments were tested in these experiments. They will be described as they arise in the text.

Figure 25 contains a map of the various test fragments assayed in these experiments. The ΔI and ΔII sequences refer to deletion mutants of pTKSBS7 (see appendix) derived by cleavage of the latter with SmaI and SstII respectively, leaving behind small residues of "a". The S terminus referred to in the map was contained on pTKSBS7, while the L terminus was a clone of the strain F L terminus, pRB143, which was kindly supplied by Dr. B. Roizman. The HindIII origin-containing fragment was cloned onto all of these plasmids to make test replicons. These were then cotransfected into Vero cells with helper viral DNA. The resulting stocks of virus were passaged 4 to 6 times through Vero cells to ensure both adequate build up of defectives in the population and dilution of any contaminating input plasmid DNA. Hirt (1967) extracts of triplicate stocks were digested with HindIII, along with plasmid markers, electrophoresed through 1.4% agarose gels and blotted to nitrocellulose filters. The blots were then hybridized with ^{32}P-labelled pBR322 DNA. Since HSV defective genomes have been shown to consist of tandem, full length, concatemers of the short monomer containing the cis-acting signals (Spaete and Frenkel, 1982; Vlazny et al, 1982), I expected a HindIII digest of the Hirt extracts to contain a monomer-sized fragment comigrating with the plasmid marker, and one or two smaller and less abundant "L" and "S" terminal fragments. (The large or small size of a terminal fragment, coupled with inadequate homology with the probe may preclude its visibility on the blot.)

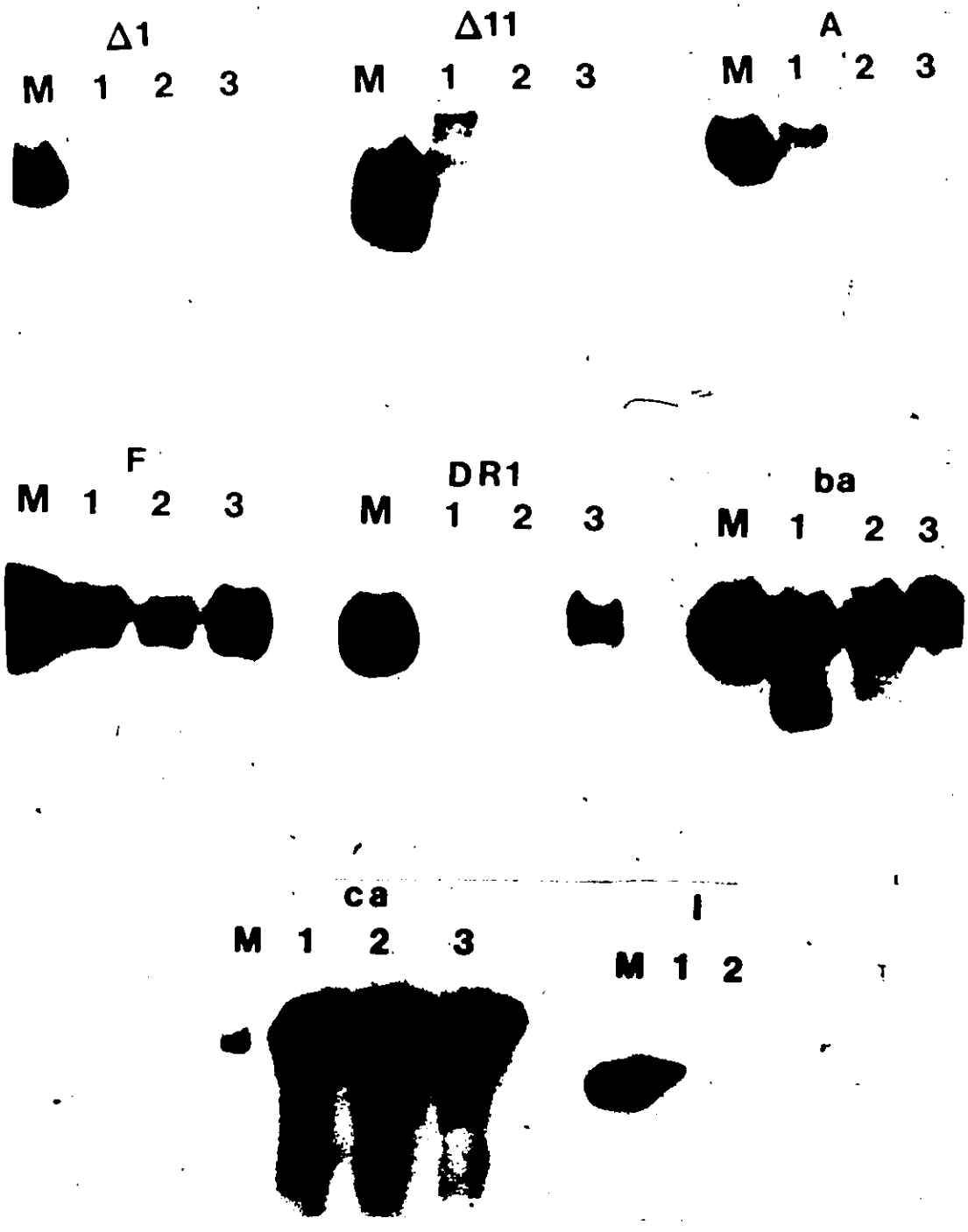
Only SmaI F and the two viral termini (ca=S terminus; ba=L terminus) produced convincing levels of defective viral DNA (Fig. 26). Signals which did not appear in replicate samples, or signals which did not comigrate with plasmid marker were scored as negative in this test. Although the blot contains some apparently anomalous signals, independent confirmation of the result (see below) indicated that the test is fairly reliable as an initial screening method. The irreproducible signals seen in lanes containing, eg. the DR1 test plasmid, probably reflect occasional recombination between the large blocks of HSV DNA on each plasmid, which consist of 1800 bp of TK DNA and 2100 bp of origin DNA. The strong signals seen in the lanes containing SmaI F, ca, and ba defectives are diagnostic of the presence of defective viral DNA. Subsequent assays indicated that strong signals could be seen after one passage, and that the reproducibility of defective production from one stock to another was reliable. Therefore I was able to conclude with confidence that SmaI F contains a viable packaging signal.

The presence of extra bands in some of the defective stocks is apparently a common occurrence,and probably reflects amplification of deleted forms of the seed replicon (Stow, pers. comm.)

All of the accumulated evidence on defective genomes had indicated that the S terminus remains intact, ie, it contains an "a" sequence at the end. Since the seed replicon starts out with one "a" sequence, and since the monomer repeat unit in the defective genome only contains one "a", the question arose as to the appearance of the other, "L", terminus in the defective genomes. Is it naked, or does it also terminate in an "a" sequence? If the latter is the case, where

FIGURE 26

Defective assay. Triplicate cultures of Vero cells were transfected with KOS viral DNA and test plasmid DNA. Resulting virus stocks were passaged four to six times at high multiplicity through Vero cells before Hirt (1967) extracts were prepared. The viral DNA was digested with HindIII, electrophoresed through a 1.4% agarose gel, blotted, and probed with ^{32}P-labelled pBR322 DNA. Lanes marked m contain HindIII digested parental plasmid DNA. ba=pbaDI, ca=pcaDI. Only two of the three cultures transfected with pSBIori and viral DNA produced any virus.


Δ1
M 1 2 3
Δ11
M 1 2 3
A
M 1 2 3
F
M 1 2 3
DR1
M 1 2 3
ba
M 1 2 3
ca
M 1 2 3
I
M 1 2

does the extra "a" sequence come from?

These questions were addressed by examining the defective viral DNA produced by pcaDI. Fig. 27 shows a map of this plasmid with the predicted "L" terminus indicated. The viral S terminus used on the seed replicon was cloned as a BamHI fragment after ligation of BamHI linkers to viral DNA. Thus, the BamHI site in the plasmid mimics the viral cleavage event that generated the original S terminus. Since the S terminus has been shown to contain a cleavage/packaging signal, it is reasonable to suppose that the BamHI site in pcaDI will coincide with the cleavage site in defective DNA that delineates the defective S terminus, which will be identical to a legitimate S terminus. Does it also mimic the cleavage event that will generate the defective "L" terminus? In practical terms, will the HindIII/BamHI fragment abutting the "a" sequence in the plasmid correspond to the HindIII L terminus in defective viral DNA?

Hirt extracts of viral DNA containing caDI defectives were digested with HindIII in two sets, and electrophoresed beside HindIII and HindIII/BamHI digests of pcaDI, the plasmid marker. The gel was blotted, and the blot was cut in half. One half was hybridized with ^{32}P-labelled pBR322, while the other half was hybridized with ^{32}P-labelled SmaI F DNA (Fig.28). The internal monomer from the defective viral DNA comigrated with the plasmid marker. However, the predicted "L" terminus was approximately 250 - 300 bp smaller than the observed "L" terminus. Moreover, this "L" terminus hybridized with SmaI F DNA, stongly suggesting that it had acquired an "a" sequence. At this point I could not say with certainty that the "a" sequence on the defective

FIGURE 27

pcaDI map. The map of pcaDI shows the orientation of the "ca" sequence in the plasmid (stippled and open boxes). Wavy lines represent TK DNA. The box filled with diagonal lines represents the HSV origin fragment. Solid lines represent pBR322 DNA.

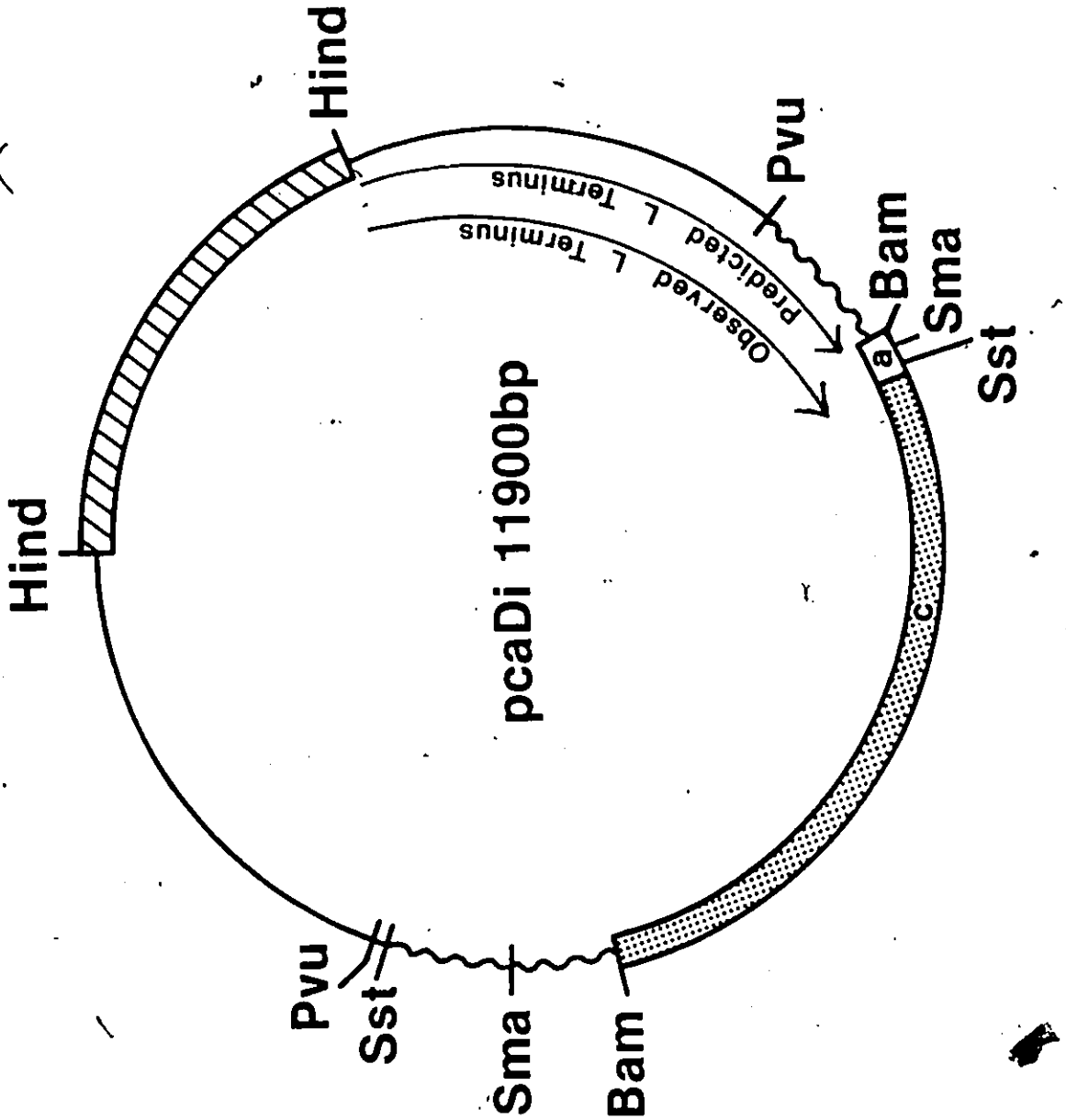
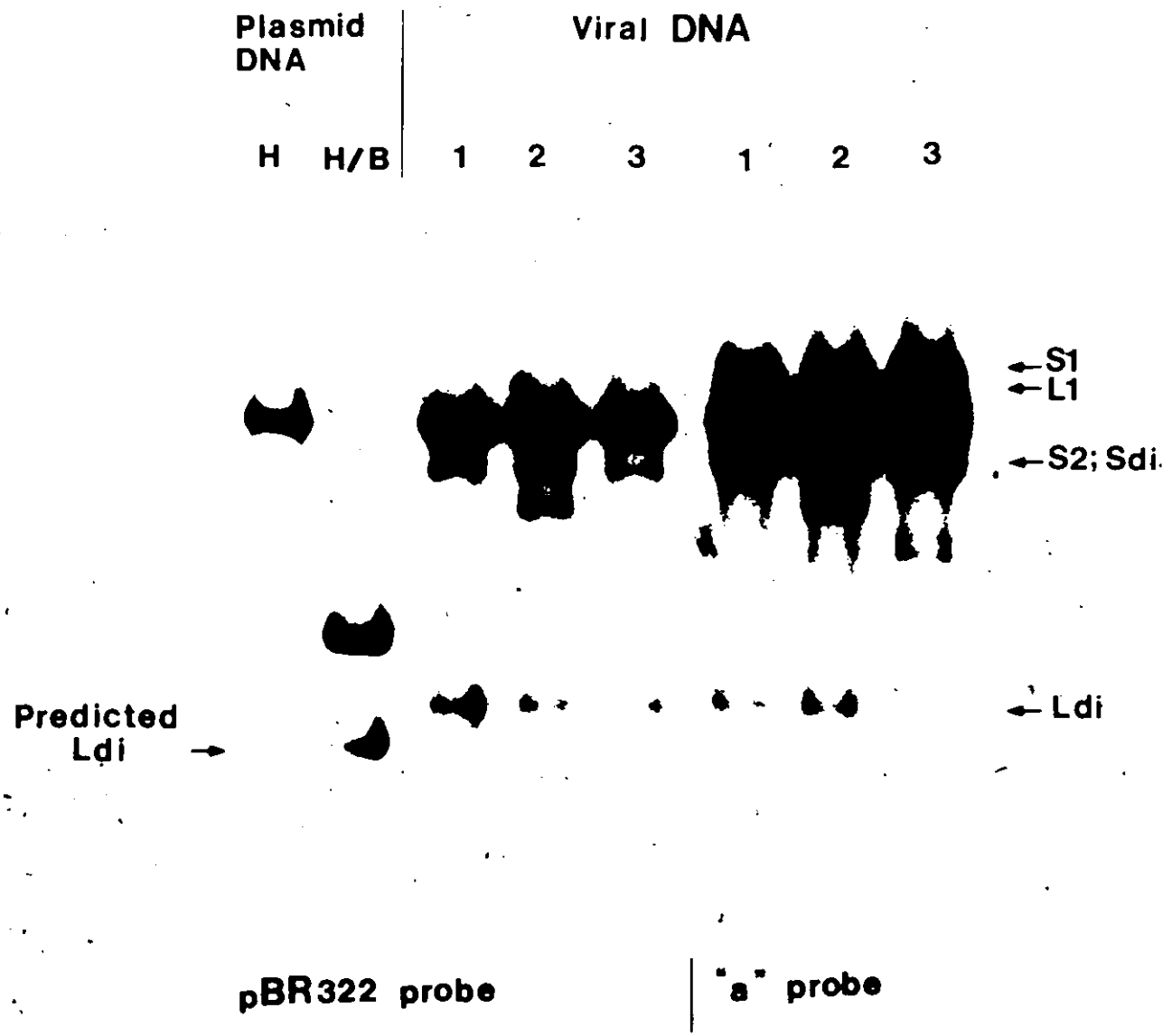

Hind
Pvu
Bam
Sma
Sst
Predicted L Terminus
Observed L Terminus
pcaDi 11900bp
a
c
Hind
Pvu
Sst
Sma
Bam

FIGURE 28

Defectives contain L termini with "a" sequences attached. Viral DNA from triplicate stocks containing defectives made with pcaDI were digested with HindIII and electrophoresed, in duplicate, through a 1.4% agarose gel. pcaDI was digested with HindIII (H), or HindIII/BamHI (H/B) and run on the gel as markers. The gel was blotted to nitrocellulose, and the blot was cut in half. One half (including plasmid markers) was hybridized with ^{32}P-labelled pBR322 DNA, and the other half was hybridized with ^{32}P-labelled SmaI F DNA purified from a BamHI digest of pSBF14 ("a" probe). The numbers at the top represent the replicate stocks. S1, S2 and L1 are legitimate viral termini derived from the helper virus present in the Hirt (1967) extract. Ldi is the "L" terminal HindIII fragment from the defective genomes. Predicted Ldi is the HindIII/BamHI fragment between the HSV origin and the "a" sequence on pcaDI (see Figure 27).


Plasmid DNA
Viral DNA
H
H/B
1
2
3
1
2
3
S1
L1
S2; Sdi
Predicted Ldi
Ldi
pBR322 probe
"a" probe

"L" terminus did not come from the S terminus as the latter was too large to see a reduction in size of 250 - 300 bp, and the legitimate S terminus from the helper viral DNA comigrated with the defective S terminus, and naturally hybridized strongly with the SmaI F probe. However, based on previously published studies of defective viral DNA, it is reasonable to assume that KOS is no different from other HSV-1 strains in this regard, and that the defective S terminus also contains an "a" sequence. This assumption was confirmed later by an independent test. Thus, a seed replicon with one "a" sequence generates two termini which each possess an "a" sequence, without any apparent change in the monomer. In other words, the cleavage event duplicates the "a" sequence such that each terminus ends in one. During the course of preparing this thesis I found confirmation of this observation from a study of defectives produced by HSV-1 (Angelotti) (Knopf et al, 1983). These authors sequenced a cloned L terminus from a defective genome and found that it contained sequences from "a".

Examination of the "a" sequence in Fig. 24 indicates that SmaI F does not contain the cleavage site(s) that generate the L and S termini in legitimate viral DNA. These lie within DR1 at a (single) specific site, as if one double stranded cleavage event were sufficient to produce both termini. This interpretation requires an additional assumption to account for the observation that both termini contain terminal "a" sequences, namely that the double stranded cleavage may only occur between two adjacent "a" sequences. Since sequence analysis of viral termini indicated that they contained residues of DR1 that together added up to a complete DR1, and since it had been determined

by several groups that tandem "a" sequences were separated by a single DR1, the double stranded cleavage model of HSV maturation seemed reasonable (Mocarski and Roizman, 1982a). However, two observations did not fit well into this model. The first is the fact that 50% of joint fragments contain only one "a" sequence. The other 50% contain two, three, four, etc. (see Fig. 2 joint stepladder). The other observation that introduces doubt about the model is that presented above, namely that defectives containing one "a" sequence in the monomer repeat are capable of creating two termini which each contain an "a" sequence. Coupled with the fact that the cleavage site(s) lies outside of the signal-containing fragment, SmaI F, the model becomes somewhat inadequate.

Recombinational Signals in SmaI F

The second part of the experiment described in the previous section required that the target sequences be rescued back into viral DNA at the TK locus. The methodology used was essentially the same as that described in the section "Sequence Identity of "a" is not Obligatory". However, the diagnostic restriction pattern of desired recombinants was different. All of the mutant TK genes were constructed with the same vector, pTKSB. The mutant TK gene on this plasmid had also been rescued back into KOS. The resulting virus, KOSSB, produces a characteristic restriction pattern when digested with BamHI which differs from wild type KOS; KOSSB produces two, smaller, TK fragments while KOS produces a single, larger, TK fragment. Since all of the target fragments were introduced into pTKSB as BamHI fragments, the

BamHI pattern of mutants containing the recombinant TK genes should be identical with that of KOSSB. Fig. 29 shows a sample blot in which TK⁻ mutants were digested with BamHI and hybridized with ^{32}p-labelled pTK173 DNA. There are 6 lanes containing fragments which comigrate with the marker bands in the lane marked "KOSSB" (lanes 2, 3, 5, 6, 7, and 11). These were the desired recombinants. (These viruses were made by cotransfection with pSBA3, a plasmid containing the SmaI A fragment inserted into pTKSB. Similar blots were prepared with viral DNA from TK⁻ mutants prepared by cotransfection with plasmids containing all of the other target sequences. These are diagrammed in the appendix in the maps of plasmids pSBF14, pSBDR, pSBI, pSB II, pSBf1, and pSBf2. Not all of the marker rescue attempts were as successful as the one displayed in Fig. 29, and required exhaustive searches before the desired mutants could be found.)

The BamHI restriction patterns, after hybridization with TK, only indicated the presence of the target sequences (by inference), and did not tell me anything about their biological activities. To examine these, the mutant viral DNAs were digested with a variety of restriction enzymes which cleave the flanking TK sequences, and probed with ^{32}p-labelled TK DNA. Fig. 30 shows two such digests, PvuII and EcoRI. Only three target sequences produced new fragments suggestive of recombination - SmaI F (KOSF10), Δ II (KOSΔII), and SmaI A (KOSA23). Of these, the latter was clearly the most dramatically active, as described below.

The mutant virus containing the SmaI F fragment inserted in the TK gene, KOSF10, produced several new TK-containing fragments. The most

FIGURE 29

Marker rescue screening gel. Vero cells were transfected with KOS viral DNA and one of several plasmids containing test fragments cloned into the BamHI site on pTKSB. TK^- plaques were picked into Linbro plates containing confluent monolayers of Vero cells and expanded. The resulting stocks were used to infect cultures of Vero cells for the preparation of Hirt (1967) extracts. Viral DNA was digested with BamHI, electrophoresed through a 1.4% agarose gel, blotted, and probed with ^{32}P-labelled TK DNA. The pattern displayed by the desired insertion mutants was expected to be identical with the pattern displayed by KOSSB, a TK^- virus which was made by transfecting cells with viral DNA and pTKSB. The expected pattern was observed in six stocks, shown in lanes 2, 3, 5, 6, 7, and 11. The viral DNA in lane 5 appears to be a mixture of at least two forms. The viruses shown here were made with pSBA3.

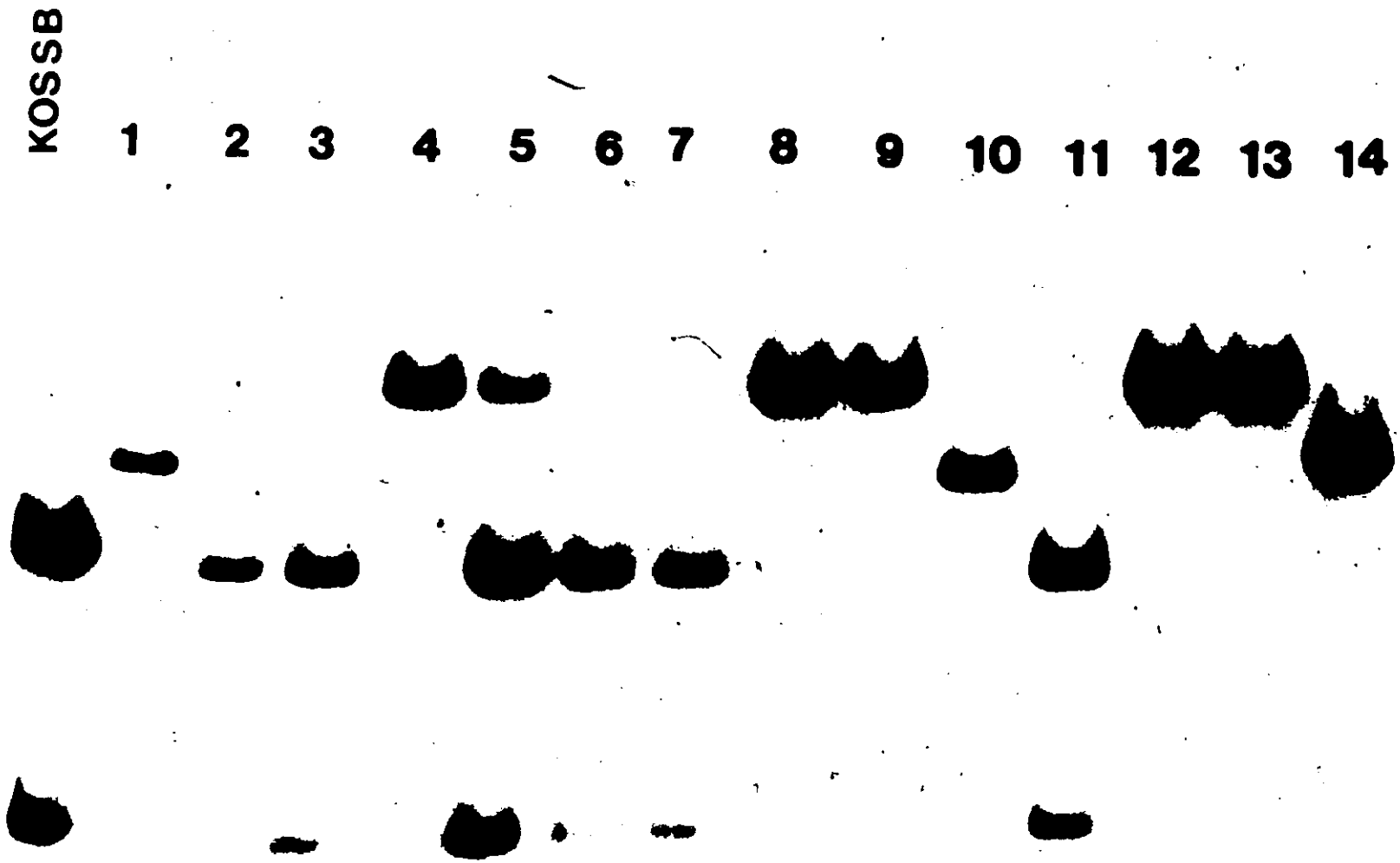
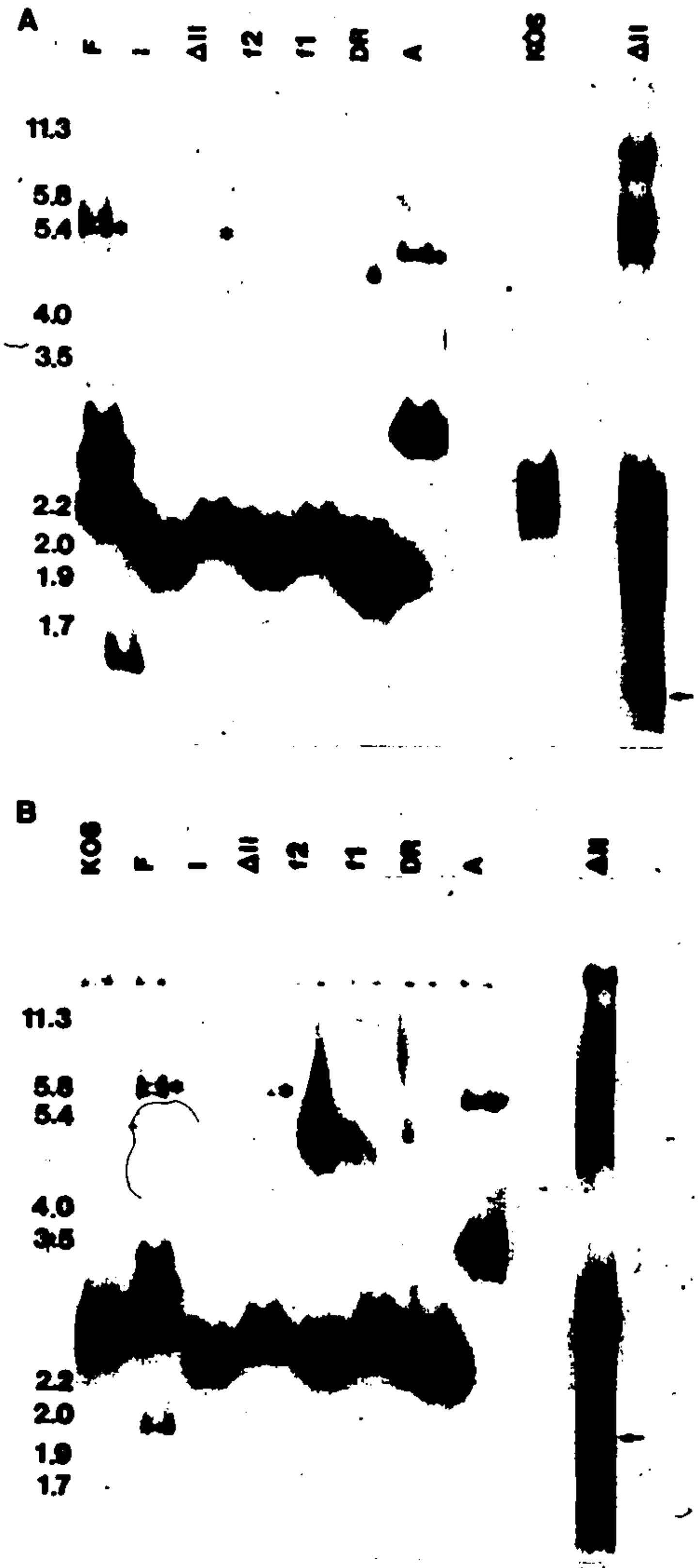

KOSSB
1
2
3
4
5
6
7
8
9
10
11
12
13
14

FIGURE 30

Restriction pattern of TK containing fragments in viruses containing inserts of "a" sequence subfragments. Viral DNA from stocks containing the indicated inserts at TK was extracted by the Hirt (1967) procedure and digested with PvuII (A) or EcoRI (B), electrophoresed through 1.4% agarose gels, blotted to nitrocellulose, and hybridized with ^{32}P-labelled TK DNA from pTK173. The numbers on the left indicate the positions of size markers (in kb). Asterisks indicate recombinant L3S2+ L3S1 fragments (see Table 3). The panels on the right contain longer exposures of lanes marked Δ II; arrows indicate the recombinant S termini produced by this virus. The codes above each lane refer to the insert rescued into each virus. Detailed information on the clones used to prepare these inserts may be found in the appendix and the legend to Fig. 25.

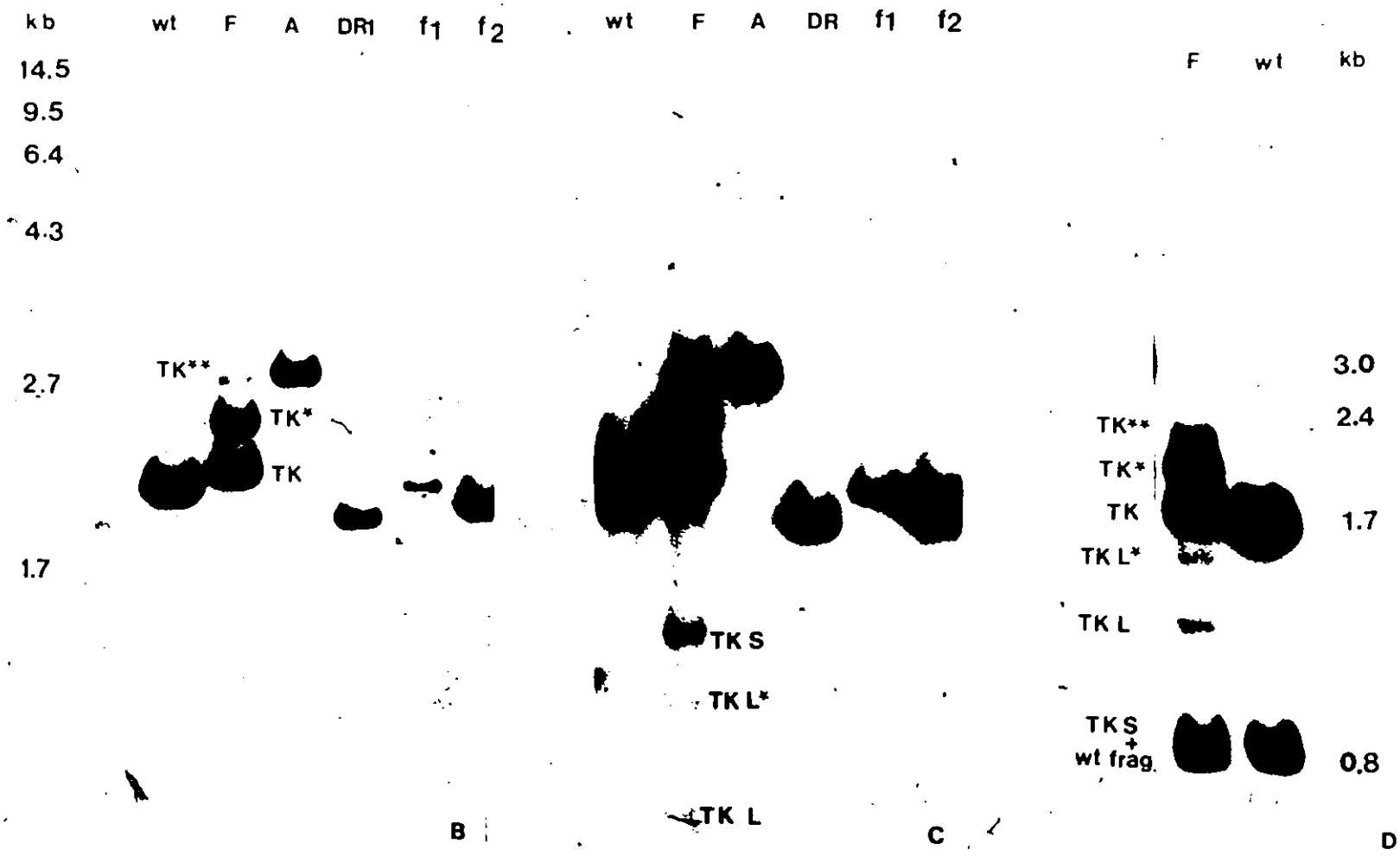

A
F
I
ΔII
12
I1
DR
A
KOS
ΔII
11.3
5.8
5.4
4.0
3.5
2.2
2.0
1.9
1.7
B
KOS
F
I
ΔII
12
I1
DR
A
ΔII
11.3
5.8
5.4
4.0
3.5
2.2
2.0
1.9
1.7

striking was the family of fragments differing in size by 250-300 bp, in which the smallest member was the TK-insert sequence. These are designated TK, TK*, and TK** in Fig. 31. All restriction enzymes tested which cleaved sequences flanking the insert produced this pattern. The other new fragments were smaller than the TK-insert sequence. One of these was more abundant than the others, which also displayed a 250-300 bp size heterogeneity. These smaller fragments were deduced to be new termini composed of TK sequences. The family of smaller fragments displaying the size heterogeneity were designated TK L termini by analogy with legitimate L termini, which contain tandemly iterated "a" sequences, and produce a similar stepladder profile. These are denoted TK L, and TK L* in Fig. 31. The other smaller TK-containing fragment was designated as the TK S terminus by similar logic. As will be seen later on, there is a very compelling reason for this designation.

The pattern of new fragments displayed by KOSF10 was very similar to those produced by the KOSxF and FxF recombinant viruses described above. That is, the insert had become a family of fragments differing in size by 280 bp increments, there were new termini, and there were fragments which were probably the result of recombination between the insert and legitimate junctions (asterisks in Figs. 30 and 31). Indeed, at first sight, I assumed that most of the new fragments might have arisen by recombination, especially if the "a" sequence contained a recombinational "hot spot". The TK stepladder profile in particular indicated that perhaps frequent recombination between the insert and L termini bearing multiple "a" sequences had left behind tandem copies of "a". An alternative, but not deemed very likely,

FIGURE 31

Southern blot analysis of mutant viral genomes. Hirt extracts of cells infected with recombinant viruses bearing the indicated inserts were digested with PvuII (two left hand panels) or SmaI (right hand panel), electrophoresed through a 1.4% agarose gel, blotted to nitrocellulose, and hybridized with ^{32}P-labelled TK DNA. The positions of size markers are indicated. The middle panel is a longer exposure of the blot in the left panel. TK L, TK L*, and TK S indicate the new termini generated by the SmaI F insert. TK, TK*, and TK** indicate internal TK fragments with 1, 2, and 3 tandem iterations of the insert, respectively. Asterisks mark the position of fragments believed to be the products of recombinational inversion through the inserted SmaI F and A inserts.


kb
wt
F
A
DR1
f1
f2
14.5
9.5
6.4
4.3
2.7
1.7
TK**
TK*
TK
B
wt
F
A
DR
f1
f2
TK S
TK L*
TK L
C
F
wt
kb
3.0
2.4
1.7
0.8
TK**
TK*
TK
TK L*
TK L
TK S
+
wt frag.
D

explanation was that the insert had somehow duplicated itself. Fortunately, the construct lent itself well to an analysis which could distinguish these possibilities

The SmaI F insert in KOSF10 is bounded by BamHI sites and contains no SmaI sites. On the other hand, the "a" sequences at the termini contain three SmaI sites and no BamHI sites. Multiple recqmbination between the insert and tandem "a's" at the L terminus (or the L-S junction) would result in a block of "a" sequences bounded by BamHI sites and containing internal SmaI sites (Fig. 32A). On the other hand, some form of de novo amplification at the insert might be expected to duplicate the BamHI sites which flank the insert, and produce a block of "a" sequences punctuated by BamHI sites (Fig. 32B). Thus, in the first case, BamHI would cut out a stepladder of "a" sequences which would be reduced to monomer size by SmaI, while in the second case, SmaI would cut out a stepladder of TK+"a" sequences which would be reduced to monomer size by BamHI. The fragments of interest could be visualized by hybridizing a blot of the gel with ^{32}P-labelled SmaI F. DNA.

The very surprising result is shown in Fig. 33. The latter prediction proved to be the correct one, that is, the stepladder arose by some sort of de novo amplification of the insert. SmaI cleaves out a stepladder which hybridizes with both TK (Fig. 31, right panel) and "a" (Fig. 33) DNA, indicating that the TK steps each contain "a" DNA. BamHI reduces the SmaI stepladder to monomer (ie SmaI F) size. The monomers seen in the SmaI digest correspond to SmaI F fragments from legitimate termini and joints. Moreover, fragments which correspond in size with

FIGURE 32

Two models for the accumulation of tandem iterations of the SmaI F fragment.

(A) Unequal recombination. Multiple cycles of unequal recombination between the SmaI F insert in TK and reiterated "a" sequences present at normal L-S junctions would move SmaI cleavage sites derived from intact "a" sequences into the reiteration block in TK.

(B) De novo amplification. The mechanism shown here is ligation of termini produced by cleavage at the insert. This would result in reiterated inserts lacking SmaI sites, and punctuated with BamHI sites.

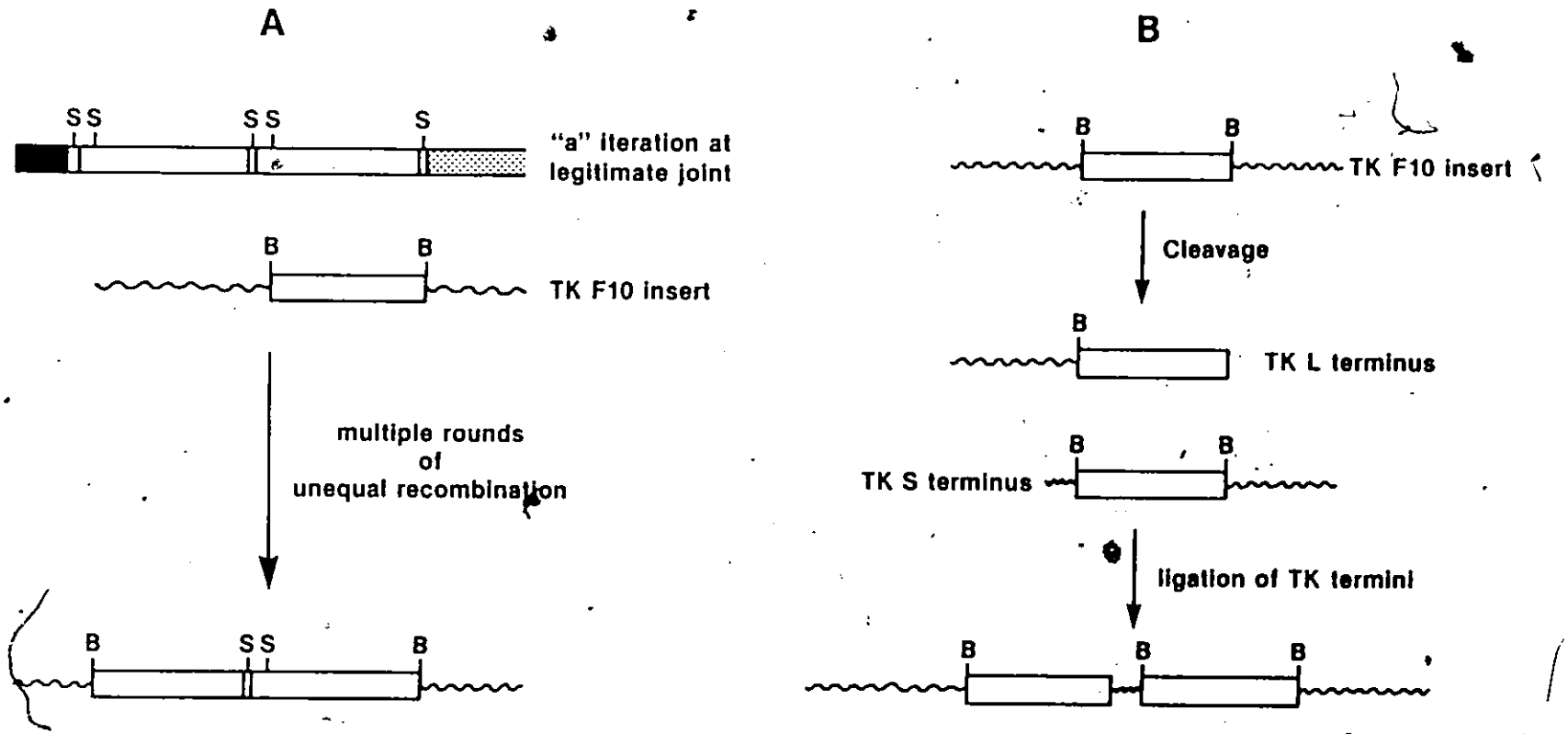

A
S S
S S
S
"a" iteration at legitimate joint
B
B
TK F10 insert
multiple rounds of unequal recombination
B
S S
B
B
TK F10 insert
Cleavage
B
TK L terminus
TK S terminus
B
B
ligation of TK termini
B
B
B

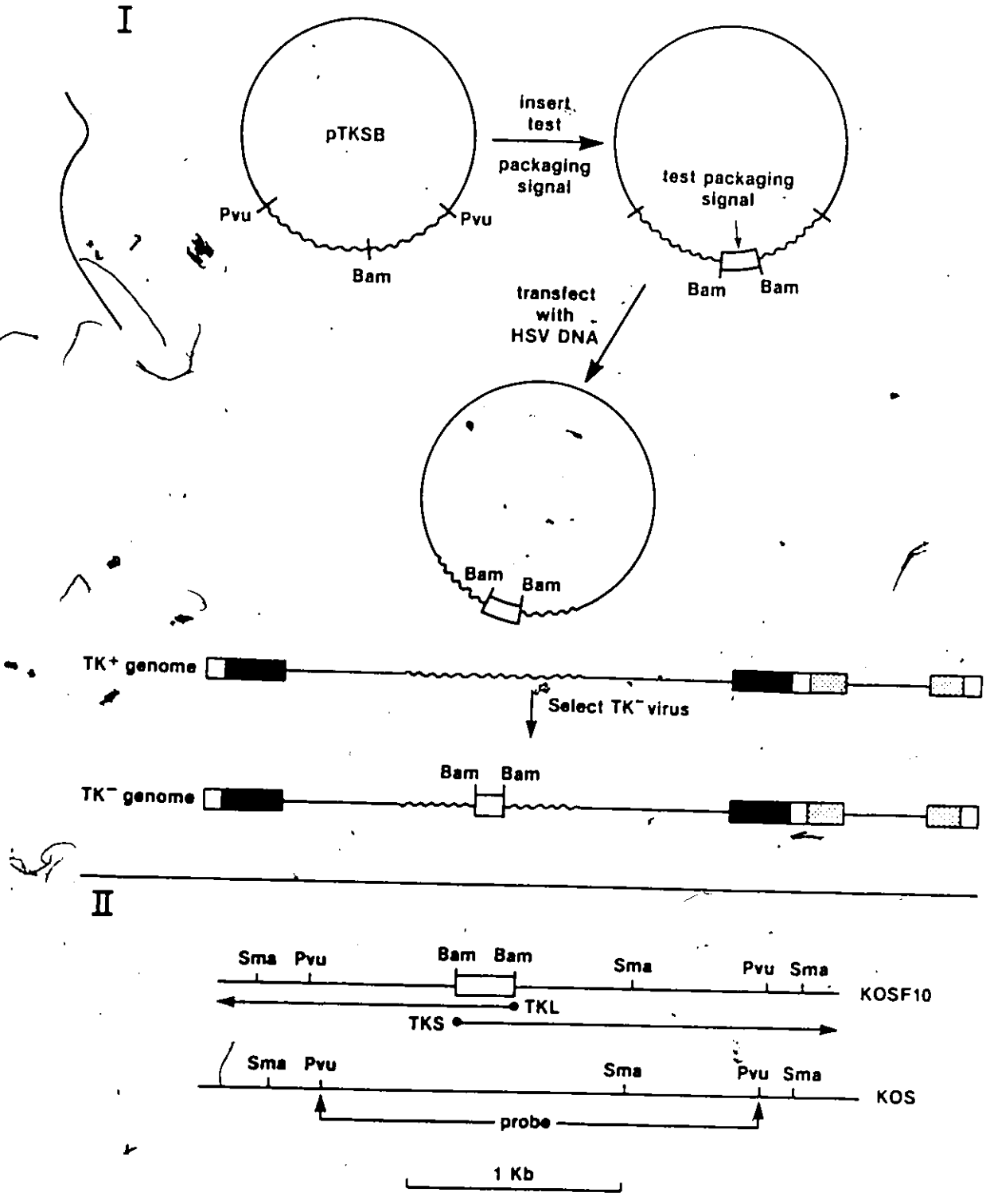
FIGURE 33

SmaI F blot of KOSF10 viral DNA. KOS (wt) or KOSF10 (F) viral DNA was digested with SmaI (S), SmaI/BamHI (S/B), or BamHI (B), electrophoresed part way through a 1.4% agarose gel, and blotted to Biodyne A paper. The blot was hybridized with ^{32}P-labelled SmaI F DNA purified from a BamHI digest of pSBF14.


wt
F
S
S/B
B
S
S/B
B

FIGURE 34

Map of KOSF10 with indicated TK L and TK S termini. Panel I describes the experimental design in making the viruses used in this part of the study, including KOSF10. Panel II shows a map of the TK loci in KOSF10 and wild type KOS. The structures of the TK L and TK S termini are shown under the map of KOSF10.


I
pTKSB
Pvu
Pvu
Bam
insert test packaging signal
test packaging signal
Bam
Bam
transfect with HSV DNA
Bam
Bam
TK+ genome
Select TK- virus
TK- genome
Bam
Bam
II
Sma
Pvu
Bam
Bam
Sma
Pvu
Sma
KOSF10
TKL
TKS
Sma
Pvu
Sma
Pvu
Sma
KOS
probe
1 Kb

the TK "L" and "S" termini also hybridize with "a" DNA (see map in Fig. 34).

This analysis provided the observation which allowed a logical explanation of some rather disjointed data. These were:

1) The ability of a single "a" sequence on defective genomes to drive the cleavage reaction.

2) The appearance of a new "a" sequence on the L termini in defectives, even though only one is present at the cleavage site.

3) The ability of SmaI F to drive cleavage in engineered defectives. This was unusual because the cleavage site in legitimate joints lies outside of SmaI F.

Obviously, de novo amplification of the insert in KOSF10 viral DNA must be related to the latter property of SmaI F, its ability to drive the cleavage reaction. Since the BamHI sites were amplified as well, and since SmaI F lies outside of the cleavage site in legitimate junctions, CLEAVAGE MUST OCCUR AT A DISTANCE FROM THE SIGNAL RATHER THAN AT A SPECIFIC SITE.

This was an important logical step, and led directly into a second major conclusion. If cleavage is a distance specific rather than a site specific event, then how do different strains of HSV-1, with different sizes of "a" sequences maintain the sequence identity of the very ends of their termini (see Fig. 13)? The only logical explanation was that THERE MUST BE TWO SIGNALS - ONE SPECIFYING THE L TERMINUS, AND ONE SPECIFYING THE S TERMINUS. Preliminary confirmation of these two conclusions was made by sequencing several pieces of cloned viral DNA from KOSF10.

Sequence Analysis of TK Termini

Most of the new fragments seen in KOSF10 viral DNA appeared to result from the cleavage/packaging mechanism. It seemed reasonable that the structure of the TK L and S termini would provide a better insight into the details of the cleavage mechanism. Therefore, the TK L and S termini were cloned and sequenced.

In order to facilitate interpretation of TK L and S terminal sequence data, the structure of the SmaI F insert in pTKSB was determined. The orientation had already been found by restriction analysis (see map pSBF14 in appendix), and the sequence of SmaI F had been determined both from XmaI digests of pSVODK6 (see section "Sequence of KOS Joint"), and from BamHI digests of pSBF14 (done by Maria Booi in the lab). The only unknown at this point was the precise location of the deletion end points in pTKSB. This was determined by sequencing the TK flanking regions starting at the BamHI sites.

pSBF14 was digested with BamHI and PvuII, and the TK fragments flanking the SmaI F insert were eluted and labelled at their 5' ends. The 5' flanking fragment was redigested with BglII, while the 3' flanking fragment was redigested with HinfI (see appendix for map). The larger of the end labelled 5' fragments, and the smaller of the end labelled 3' fragments were eluted from an 8% polyacrylamide gel and subjected to Maxam and Gilbert (1980) sequence reactions. The sequence ladders are shown in Fig. 41 while the deduced sequence of the region surrounding the SmaI F insert in pTKSB appear in Fig. 36. HSV sequences are in upper case letters, while polylinker (underlined) and BamHI linker sequences are in lower case letters. The 5' TK sequences shown

in the figure are the inverse complement of the sequences read from the sequence ladders since the BamHI site at the linker/polylinker junction was labelled at the 5'end. This analysis allowed me to determine that exactly 200 bp of TK DNA were deleted during construction of pTKSB.

Although the TK L terminus was almost invisible on ethidium bromide stained gels of EcoRI digested KOSF10 viral DNA, it proved to be easier to clone than the larger, more visible, TK S terminus. The strategy was to digest KOSF10 viral DNA with EcoRI. The reason for choosing this restriction enzyme was that none of the other viral DNA fragments comigrated in the region of the gel occupied by any of the new TK fragments. The TK L terminus was eluted from the agarose gel by electroelution. (Although the fragment could not be directly visualized, size markers on the gel allowed an estimation of the approximate location where it was expected to run.) The TK L terminus was then ligated with pUC-19 DNA which had been digested with EcoRI and SmaI. These sites reside within the pUC-19 polylinker region. The ligated DNA was used to transform JM109 cells to ampicillin resistance on X-gal plates. White colonies were chosen for further analysis. Birnboim preparations of plasmid DNA were digested with BamHI and electrophoresed through an 8% polyacrylamide gel. The polylinker in pUC-19 contained a BamHI site very close to the SmaI site to which the end of the viral terminus was expected to ligate. If the terminus contained both of the BamHI sites from the SmaI F insert site in TK, then BamHI was expected to cleave out the SmaI F insert. If only the 5' TK proximal BamHI site was present in the terminus, then the BamHI site in the polylinker could substitute for the missing BamHI site, and the insert would be

cleaved out (see map of pTKL13 in appendix). Of 16 plasmids screened, two contained a BamHI fragment which comigrated with the insert from pSBF14. One of these, pTKL13, was analysed further and found to display a restriction profile predicted for the TK L terminus (see maps in appendix and Fig. 34).

pTKL13 was digested with SalI and EcoRI, and the smaller insert-containing fragment was eluted from an 8% poyacrylamide gel. The DNA was ~~end-~~labelled with T4 kinase and strand separated. The separated strands were then subjected to Maxam and Gilbert (1980) sequence reactions. One strand was expected to contain sequences from the SalI site in the polylinker directly linked to the "a" sequences on the end of the TK L terminus. The other strand was expected to consist of TK sequences from the 5' EcoRI site.

Figure 35 shows the sequence ladders of the two strands from pTKL13, while Figure 36 contains the deduced sequences of these strands. The predictions stated above were proved correct. The polylinker sequences from pUC-19 are underlined, while the polylinker sequences from both pUC-19 and pUC-7, the vector into which SmaI F was originally cloned, are in lower case letters. HSV sequences are in upper case letters. The cleavage reaction which generated the TK L terminus occured within the flanking polylinker sequences just short of the BamHI site (asterisk in Fig. 36).

The TK S terminus proved extremely difficult to clone. Repeated attempts to employ the strategy described above, with or without treatment of viral DNA with T4 polymerase, were unsuccessful. The reason for this lack of success remains a mystery. Ultimately, KOSF10 viral DNA

FIGURE 35

Sequence gels of pTKL13 fragments. pTKL13 was digested with SalI and EcoRI, and the TK L terminal fragment was eluted from an 8% acrylamide gel. The eluted DNA was labelled at the 5' ends and strand separated. L1 and L2 refer to the position of the separated strands on the strand separating gel. The separated strands were subjected to Maxam and Gilbert (1980) sequence reactions and electrophoresed through 20% (right panel) and 8% (middle and left panels) sequencing gels. Tracks marked G contain G specific reactions, tracks marked A contain G+A specific reactions, tracks marked T contain T+C specific reactions, and tracks marked C contain C specific reactions.

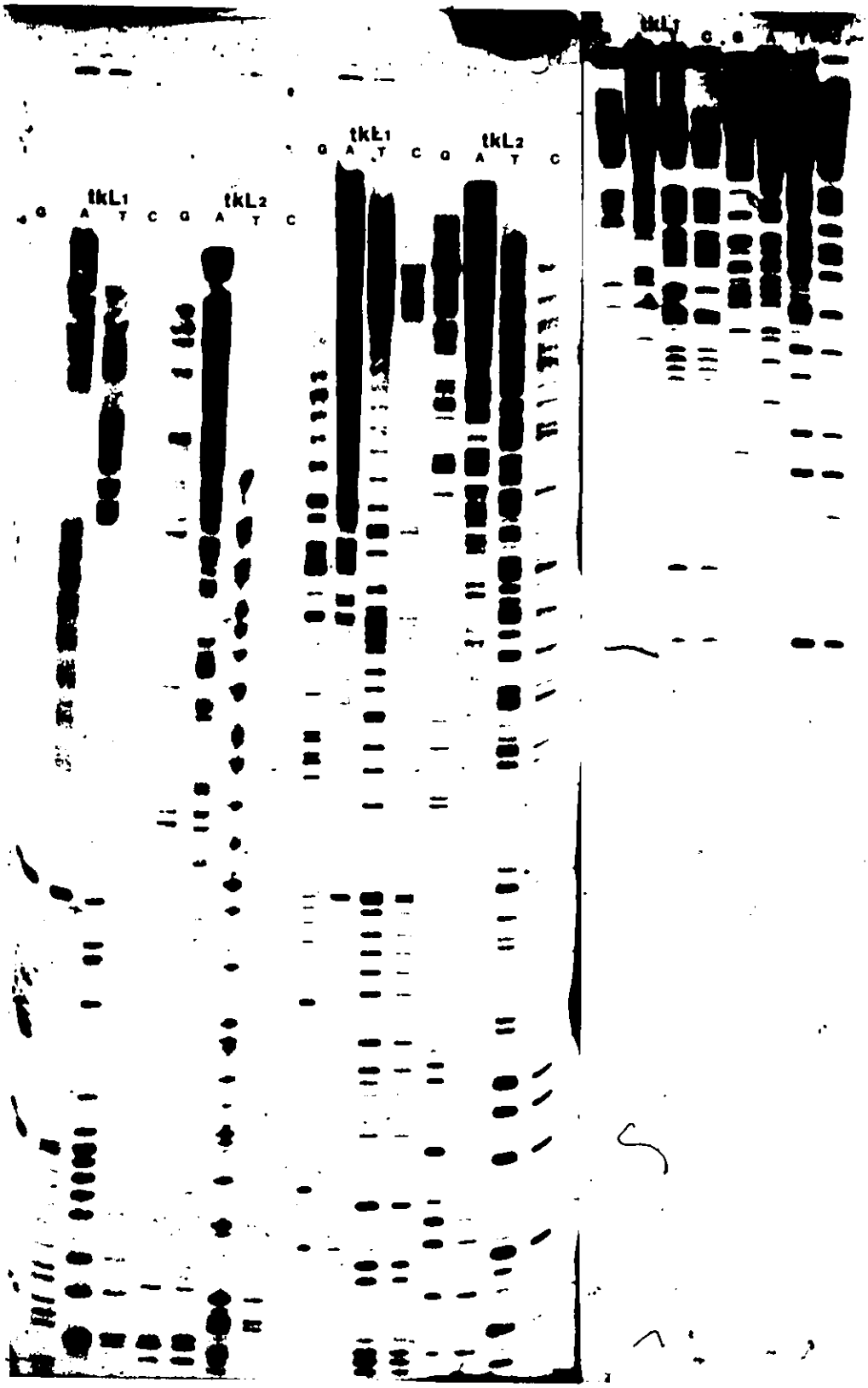

tkL1
G A T C
tkL2
G A T C
tkL1
G A T C
tkL2
G A T C
tkL1
G A T C

FIGURE 36

Sequence of pSBF14. The 5' TK sequences are the inverse complement of sequences read from gels in Fig. 42 and elsewhere (data not shown). The SmaI F sequences were inferred both by the homology between those parts which were readable from the gels and by the pattern of repeating units. See Fig. 45 for the strategy used to derive the sequence for pSBF14. Arrows indicate the sites used in KOSF10 to cleave viral concatemers at this site. BamHI and polylinker sequences are in lower case letters; the latter are underlined.

Sequence of pTKL13. The sequences in brackets are inferred from the structure and sequence of pSBF14, and by the pattern displayed by the repeats in SmaI F. The sequences beyond the bracketted ones are the inverse complement of sequences read from the gels in Figure 35. Polylinker sequences from pUC7 and pUC19 are in lower case letters; the former are underlined. The asterisk indicates the ultimate base from viral DNA.

Sequences from pRIS3. pBR322 and HSV sequences are in upper case letters, while polylinker and EcoRI linker sequences are in lower case letters. The polylinker sequences from pUC7 are underlined. Sequences in brackets are inferred from the sequence of pSBF14 and TK (Wagner and Summers, 1981). The 3' TK sequences beyond the brackets are the inverse complement of the sequences read from the gels in Fig. 37. The asterisk indicates the ultimate base of viral DNA in the clone.

pSBF14 Insert

5' TK sequences Ub sequences

GGTGAGATATCGGCCGGGGACGCGGCGGTGGTAATGACAAGCGCCccggatccgtcGGGGGGTGTGTTTTG

Uc sequences

GGGGGGGCCCGTTT.....SmaI F sequences.....TTTTTTGCGCGCGCGCGCGCCCGCGGGGGG

3' TK sequences

CCCgacggatccggTGGCCCTCATCCCGCCGACCTTGCCCGGCACAAACATCGTGTTGGGGGCGCTTCCGG

pTKL13 Sequences

ATTCGAACACGCAGATGCAGTCGGGGTCGGCGCGGTCCTAGGTCCACTTCGCATATTAAGGTGACGCGTGTG

GCCTCGAACACCGAGCGACCCTGCAGCGACCCGCTTAACAGCGCACACGGCGCCAG...5' TK.....

(GTGAGATATCGGCCGGGGACGCGGCGGTGGTAATGACAAGCGCCccggatccgtcGGGGGGTGTGTTTTG

GGGGGGGGCCCGTTT.....SmaI F sequences....)GCGCGCGCGCGCGCCGCCCGGACCGCCGCC

CGCCTTTTTTGCGCGCGCGCGCGCCCGCGGGGGGCCCgacggatggggatcctctagagtcg

pRIS3 Sequences

pBR322 Sequences * 5'TK Sequences

AGCTTATCGATGATAAGCTGTCAAACATGAGAAttccgCGCGGCGGTGGTAATGACAAGCGccggatccgt

cGGGGGGTGTGTTTTGGGGGGGGCCCGTTTCCGGGGTCTGGCCGCTCCTCCCCC....SmaI F seq...

(GCCTTTTTTGCGCGCGCGCGCGCCCGCGGGGGGCCCgacggatccggTGGCCCTCATCCCGCCGAC....

....3' TK Sequences....)GCACAAACACTGTGTTGGGGGCCCTTCCGGAGGACAGACACATTGAC

CGCCTGTCAAACGCCAGCGCCCCGGCGAGCGGTTTGACCTGCTATGTTGGCCGCGATTCGCCGCGTTTACG

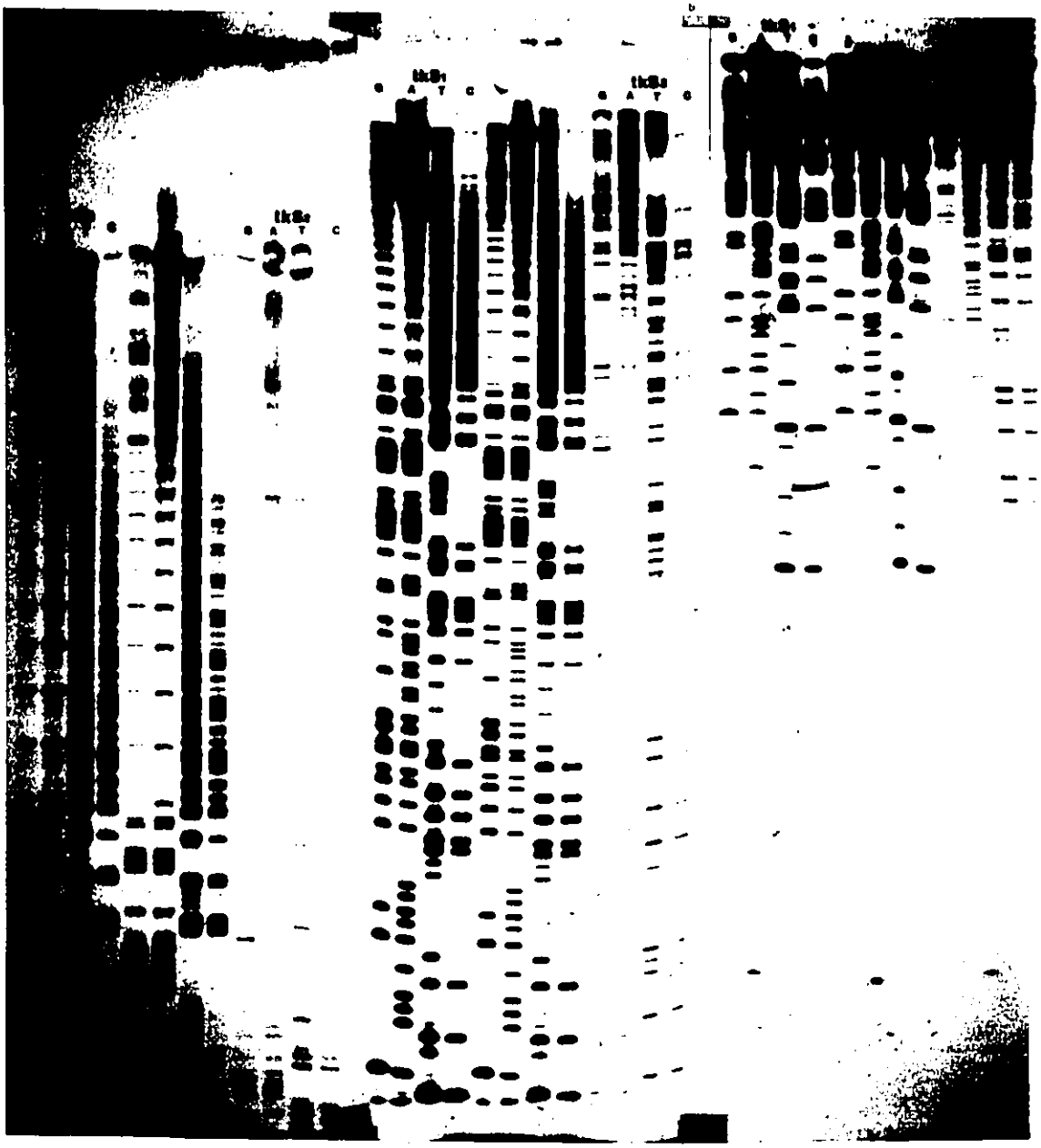
GGCTGCTTGCCAATACGGTGCGGTG

was treated with T4 polymerase to ensure flush ends before EcoRI linkers were ligated to it. The EcoRI TK S terminus was eluted from an agarose gel and ligated with EcoRI digested pBR322 DNA. Ampicillin resistant colonies were replica plated on ampicillin containing Luria plates, and one of each set was transferred to nitrocellulose and hybridized with ^{32}P-labelled TK DNA. DNA from positive colonies was subjected to restriction analysis to verify that they were the correct clones, and to establish the orientation of the insert in pBR322. Six of seven clones proved to be authentic; four were in one orientation (see map of pRIS1 in appendix), and two were in the opposite orientation (see map of pRIS3 in appendix).

pRIS3 was chosen for sequence analysis of the TK S terminus because the HindIII site in pBR322 was determined by restriction analysis to be very close to the end of the cloned terminus which contains the "a" sequence. Since the sequence of TK at the 3' EcoRI site was unknown, and since such a large fragment would have been extremely difficult to strand separate, I chose to digest the plasmid with HindIII and PstI, label the 5' ends and elute the terminus containing fragment from an 8% polyacrylamide gel. The separated strands were subjected to Maxam and Gilbert (1980) sequence reactions. The sequence ladders of the two strands are shown in Fig. 37, and the deduced sequences are presented in Fig. 36. HSV sequences are shown in upper case letters, except 33 bp of pBR322 sequences, which are indicated. Linker and polylinker sequences are shown in lower case letters, and the polylinker sequences from pUC-7 are underlined. Wherever two different sequences are joined by a restriction site (reflecting the

FIGURE 37

Sequence gels of pRIS3. pRIS3 was digested with HindIII and PstI, and the fragment containing the TK S terminus was eluted from an 8% polyacrylamide gel. The eluted DNA was end labelled and strand separated, and the separated strands were subjected to Maxam and Gilbert (1980) sequence reactions. Samples were electrophoresed through 20% (right panels) and 8% (middle and left panels) polyacrylamide sequencing gels. tkS1 is the slower strand and tkS2 is the faster strand. Tracks marked G contain G specific reactions, tracks marked A contain G+A specific reactions, tracks marked T contain T+C specific reactions, and tracks marked C contain C-specific reactions. The middle tracks which are not labelled contain the sequence reaction products of a third strand which is identical with tkS1. For some unknown reason this strand takes on at least two conformations which are distinguishable in a strand separating system.



ligation event), half of the restriction recognition sequence is portrayed as one type of sequence and the other half as the other type of sequence, without regard to the type of cleavage made by the enzyme (ie, flush ended cuts vs staggered cuts). Asterisks lie above the ultimate base in the L or S terminus which corresponds to the sequence in pSBF14, and therefore to the base at the end of the terminus in viral DNA. The cleavage reactions which produced the TK L and S termini occured in flanking TK sequences such that the L terminus contains 7 base pairs of linker DNA from the 3' side of the insert at its end, and the S terminus contains 36 bp of 5' TK plus linker sequence at its end, indicating that the cleavages occur at distances from the signal(s) contained in the SmaI F insert. In addition, except for a GGA trinucleotide, the cleavage sites which generate L and S termini are different, strongly suggesting that two different signals may direct two different cleavage events.

Sequence and Restriction Analysis of TK Stepladder

Preliminary restriction analysis had already suggested that the stepladder at the TK locus in KOSF10 viral DNA was a product of the cleavage/packaging mechanism, rather than a recombinational mechanism (Figs. 32 and 33). Examination of the sequences of TK L and S termini in Fig. 36 reveals that end-to-end ligation of these sequences would produce a TK with two tandem "a" sequences punctuated by BamHI sites. In fact, the two tandem "a" sequences would not be identical since the S terminus contains 36 bp of TK DNA, so the two "a" sequences in the new "joint" would be separated by these 36 bp. Moreover, the BamHI

fragments encompassing the inserted "a" sequences from this new "joint" would be slightly different in size because the distal BamHI site at the TK L terminal cleavage site is destroyed by the cleavage reaction. If the tandem iterations of "a" at TK are indeed the result of ligation of TK L and S termini, then all of these predicted structures should be seen. Consequently, the second step of an EcoRI TK stepladder was cloned into pBR322, and several clones were analysed.

KOSF10 viral DNA was digested with EcoRI and electrophoresed through a 1.4% agarose gel. The second step of the TK stepladder, which was clearly visible after staining with ethidium bromide, was eluted and ligated with EcoRI digested pBR322 DNA. Ampicillin resistant colonies were replica plated as described above, and a nitrocellulose blot of the colonies on one replica of each set was hybridized with ^{32}P-labelled TK DNA (Fig. 38). Plasmid DNA from positively labelled colonies was examined further.

Restriction analysis was carried out on 26 different clones. Six of these contained three tandem copies of "a", while the rest contained two. This was determined by the size of a PstI fragment containing the insert site. The PstI fragments containing three copies were 1400 bp long, while those with two were 1100 bp long. Several clones with one insert also turned up, although they were not analysed further. The PstI insert-bearing fragment in these was 820 bp long. It was reasoned that SmaI should cleave within the insert if the iteration had arisen by recombination (see section "Recombinational Signals in SmaI F"; Figs. 32 and 33). Fig. 39A shows a series of PstI and PstI/SmaI digests of seven different clones. Five of them contain double inserts (3, 6,

FIGURE 38

Colony hybridization. The second step of an EcoRI TK stepladder from KOSF10 was eluted from a 1.4% agarose gel and ligated with EcoRI digested pBR322. Ampicillin resistant, tetracycline sensitive colonies were replica plated onto Ampicillin plates, and one of the replicas was subjected to colony hybridization, using the purified PvuII TK containing fragment from pTK173 as probe. Arrows indicate positive controls (pTK173 containing cells), while the circle surrounds a negative control (pBR322 containing cells). Several of the positfve colonies were picked for further study.

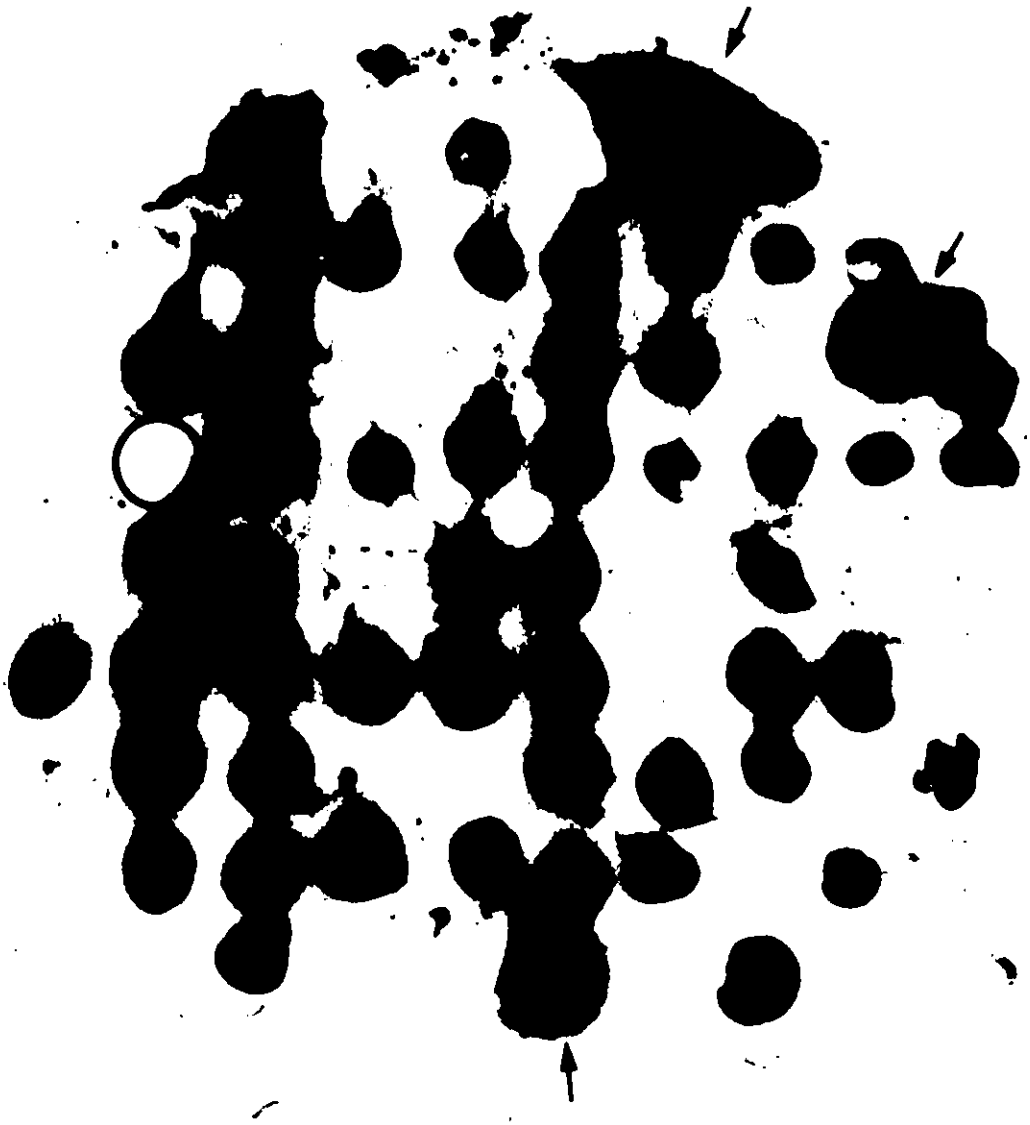
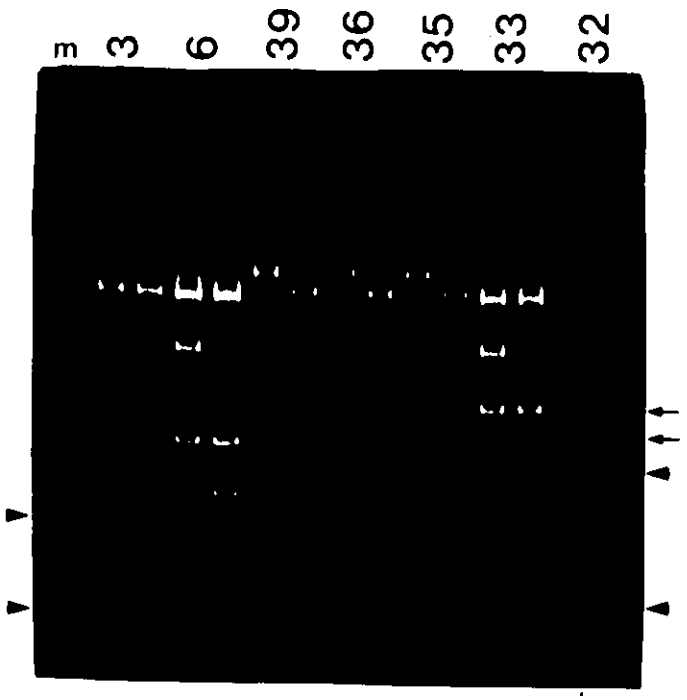


FIGURE 39

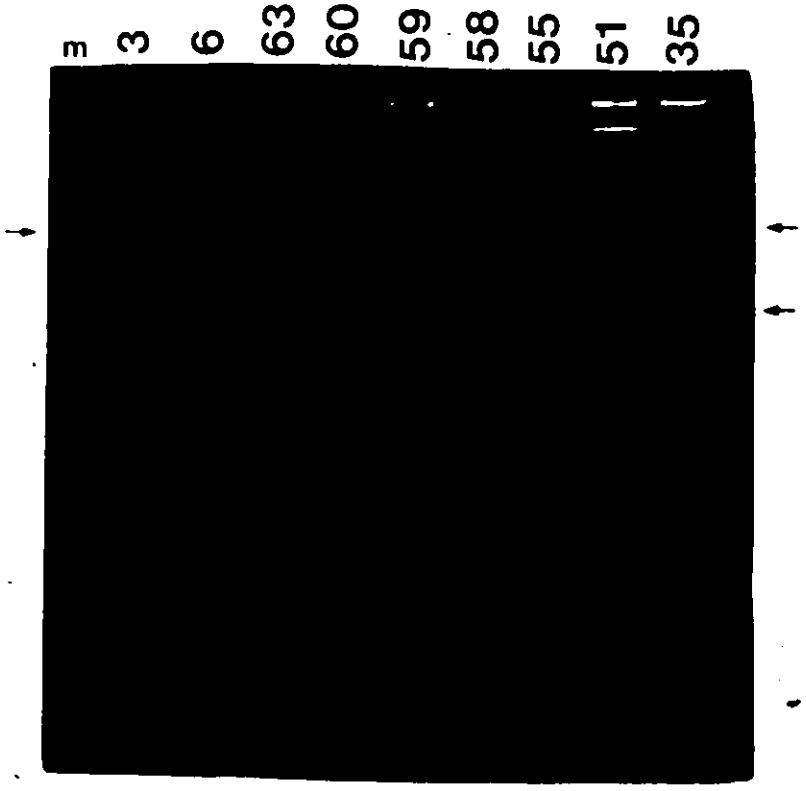
(A) PstI and PstI/SmaI digests of several cloned second and third steps from an EcoRI TK stepladder of KOSF10. The digests of each clone were run beside each other, with the PstI digests on the left of each pair and the PstI/SmaI digests on the right of each pair. The arrows indicate the PstI fragment which spans the insert regions from clones with three tandem inserts (upper arrow) or two tandem inserts (lower arrow); the arrowheads indicate the SmaI subfragments of the PstI insert containing fragment in clones pRIaa3 (arrowheads on left, lanes marked 3) and pRIaa35 (arrowheads on right, lanes marked 35).

(B) BamHI digests of various cloned second and third steps from EcoRI TK stepladder of KOSF10. The arrows indictae the 500 and 295 bp BamHI fragments mentioned in the text. The lanes marked m in both gels contain HinfI digests of pTK173, which generate size markers of 1631, 1327, 1057, 517, 506, 396, 298, 221, 220, 154, and 75 bp.

A



B



39, 36, and 32), while two contain triple inserts (35 and 33), as evidenced by the size of the bands indicated with arrows. Note that clones 39, 36, and 35 contain the insert in opposite orientation within pBR322 than the others, as evidenced by the larger vector fragments (see maps of pRIaa2, pRIaa3, pRIaa4, pRIaa5, and pRIaa6 in appendix). SmaI reduces the size of the insert containing band to 400 bp and 700 bp fragments in pRIaa3 (arrow heads on left), and 400 bp and 1000 bp fragments in pRIaa35 (arrow heads on right), but does not affect the insert containing fragments in any of the other clones. BamHI digests of these clones revealed that those with double inserts lacking SmaI sites contained two BamHI fragments differing in size by approximately 30 bp (295 and 265 bp; smallest bands in most digests - see lane marked "6"), with the smaller fragment comigrating with the BamHI insert in pSBF14 (data not shown). This structure was predicted for double insert-containing fragments derived by ligation of TK L and S termini. Digestion of pRIaa3 and pRIaa35 yielded different patterns; pRIaa3 contained a single BamHI insert fragment of 500 bp, while pRIaa35 contained 500 bp and 295 bp BamHI insert fragments. The latter comigrated with the larger of the two BamHI insert fragments generated by the other clones (Fig. 39B). These two clones therefore exhibited a structure predictive of recombinants as illustrated in Fig. 32A. They were outnumbered 12:1 by cloned fragments of the other derivation, however. Maps of three clones, pRIaa6, pRIaa3, and pRIaa35 are shown in Fig. 40. These were deduced from the restriction patterns shown in Fig. 39, and from the known restriction patterns of TK and pSBF14. Data from restriction analysis of all of the clones is compiled in Table 2. The

FIGURE 40

Restriction maps of insert sites in pRIaa6, pRIaa3, and pRIaa35. The boxes denote the SmaI F inserts, lines represent TK DNA. B=BamHI, R=EcoRI, P=PstI, T=SstII, S=SmaI. The double headed arrows indicate the SstII fragments from pRIaa6 and pRIaa3, and the BamHI fragment from pRIaa6 which were sequenced.

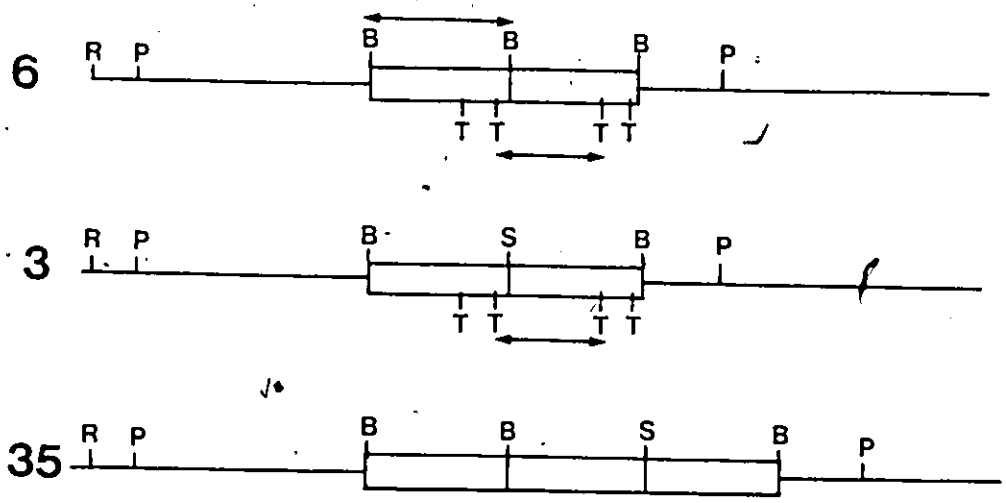

6
R
P
B
B
B
P
T
T
T
T
3
R
P
B
S
B
P
T
T
T
T
35
R
P
B
B
S
B
P

TABLE 2

RESTRICTION ANALYSIS OF KOSF10 CLONED SECOND AND THIRD TK STEPS

CLONE	INSERT	PstI	PstI/SmaI	BamHI
pRIaa2	double >	1100	1100	295 265
pRIaa3	double >	1100	700 400	560
pRIaa4	double >	1100	1100	295 265
pRIaa5	double <	1100	1100	295 265
pRIaa6	double >	1100	1100	295 265
pRIaa19	double >	1100	1100	295 265
pRIaa20	double >	1100	1100	295 265
pRIaa29	double >	1100	1100	295 265
pRIaa31	triple <	1400	1400	295 x 2 265
pRIaa32	double >	1100	1100	295 265
pRIaa33	triple >	1400	1400	295 x 2 265
pRIaa35	triple <	1400	1000 400	560 295
pRIaa39	double <	1100	1100	295 265
pRIaa47	double <	1100	1100	295 265
pRIaa49	double >	1100	1100	295 265
pRIaa51	double >	1100	1100	295 265
pRIaa55	triple >	1400	1400	295 x 2 265
pRIaa58	double >	1100	1100	295 265
pRIaa60	double >	1100	1100	295 265
pRIaa63	double >	1100	1100	295 265
pRIaa64	double <	1100	1100	295 265
pRIaa66	double <	1100	1100	295 265
pRIaa67	double >	1100	1100	295 265
pRIaa68	triple >	1400	1400	295 x 2 265

CLONE	INSERT	PstI	PstI/SmaI	BamHI
pRIaa70	triple <	1400	1400	295 x 2 265
pRIaa76	double >	1100	1100	295 265

symbols ">" and "<" indicate the orientation of the inserts in pBR322 (see maps in appendix for details).

While the restriction analysis was a good indication that the iterated "a" sequences at the TK locus in KOSF10 were derived by ligation of TK L and S termini, more detailed examination of the sequences between the two iterations would remove any doubt of this conclusion. Therefore, the SstII fragments spanning the junctions between the iterations in pRIaa3 and pRIaa6, indicated in Fig. 40, were end labelled and sequenced. (See Fig. 45 for details of the sequencing strategy used to sequence pRIaa6.)

The two plasmids were digested with SstII, labelled at the 5' ends, and electrophoresed through an 8% polyacrylamide gel. The 215 bp SstII fragment spanning the junction was eluted and strand separated. I should mention here that the junction fragment from pRIaa3 was approximately 10 bp shorter than its counterpart from pRIaa6. The reason for this will be revealed below. In addition, as indicated in Fig. 45, the BamHI amplification fragment from pRIaa6 was also end labelled, strand separated and sequenced.

The results are presented in Figs. 41-43 (sequence ladders) and 44 (deduced sequences). The structure of the junction fragment from pRIaa6 is exactly as predicted from the sequences of the TK L and S termini, except for the apparent absence of one base pair from the TK L terminus. The junction fragment from pRIaa3, the recombinant, was slightly different from that predicted in Fig. 32. The presence of six base pairs of pUC-7 linker DNA (underlined) interposed between the two "a" sequences indicated that this molecule must have arisen by ligation

FIGURE 41

Sequence gels of BamHI fragment from pRIaa6, and of flanking TK fragments from pSBF14. The 295 bp BamHI fragment from pRIaa6 was eluted from an 8% polyacrylamide gel, end labelled, and strand separated. The separated strands were subjected to Maxam and Gilbert (1980) sequence reactions and electrophoresed through 20% (left panel) and 8% (middle panel) sequencing gels. 6B1 refers to the slower strand and 6B2 refers to the faster strand. pSBF14 was digested with BamHI and PvuII, and the 680 bp 5' TK fragment and the 1160 bp 3' TK fragment were eluted from an 8% polyacrylamide gel. The fragments were end labelled and redigested with BglII and HinfI respectively. The 430 bp 5' TK fragment and the 140 bp 3' TK fragment were eluted from an 8% polyacrylamide gel and sequenced. Samples were run on a 20% sequencing gel (right panel). Although some samples were run on the 8% sequencing gel, the data from the 20% gel were sufficient to establish the borders of the insert site in pSBF14, and, by analogy, the extent of the deletion in pTKSB. Homologous fragments from pRIaa6 were also sequenced (see Fig. 45 for sequencing strategy; data not shown). Tracks marked G contain G specific reactions, tracks marked A contain G+A specific reactions, tracks marked T contain T+C specific reactions, and tracks marked C contain C specific reactions.

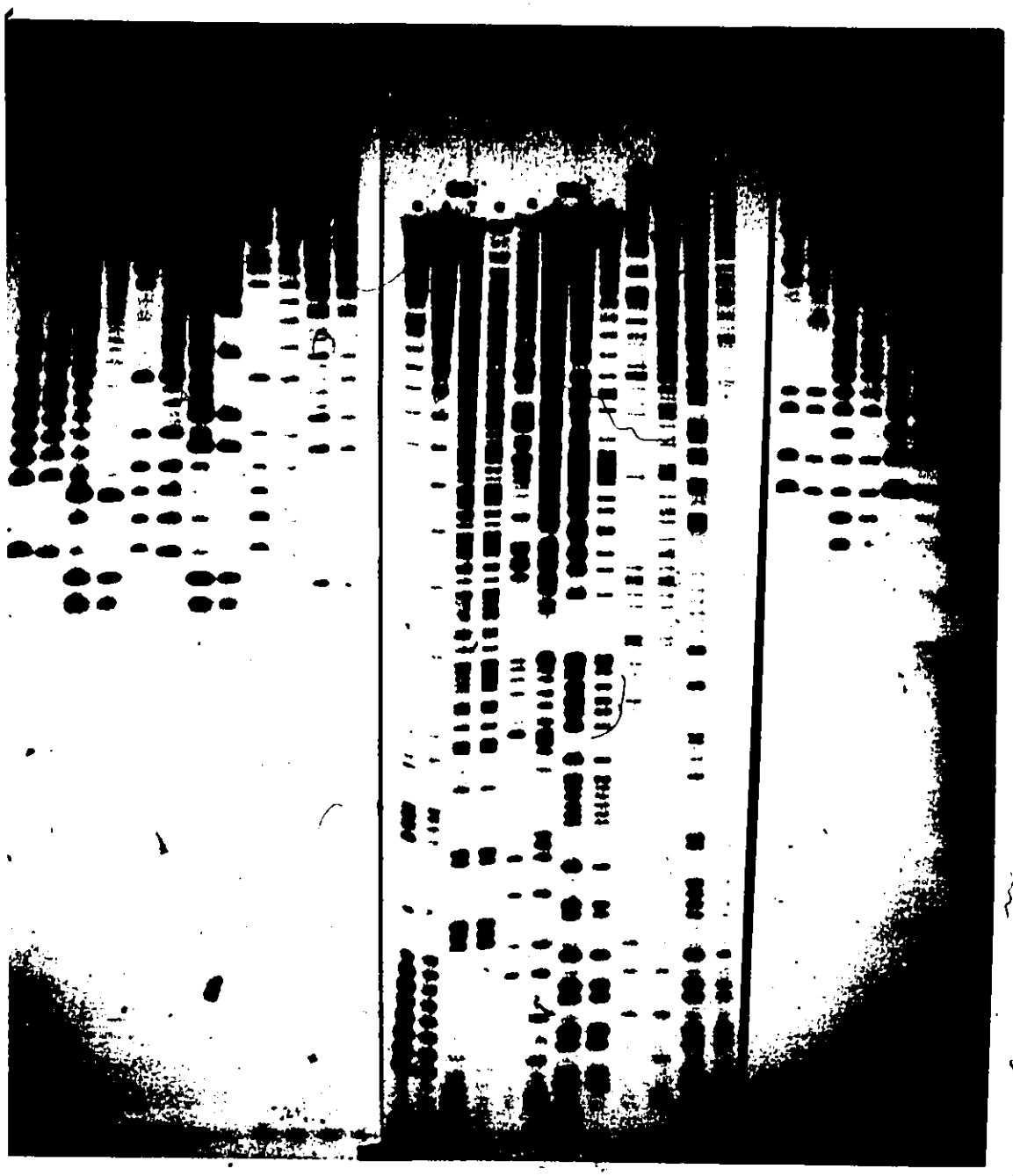


FIGURE 42

Sequence gels of SstII fragment from pRIaa6. The 215 bp SstII fragment spanning the junction between the iterated inserts in pRIaa6 was eluted, end labelled, strand separated, and sequenced. Only one of the strands, 6S1, is shown in this figure. 6S1 is the slower strand. The other strand may be seen in Fig. 43. Samples were run through 20% (right panel) and 8% (middle and left panels) sequencing gels. Tracks marked G contain G specific reactions, tracks marked A contain G+A specific reactions, tracks marked T contain T+C specific reactions, and tracks marked C contain C specific reactions.

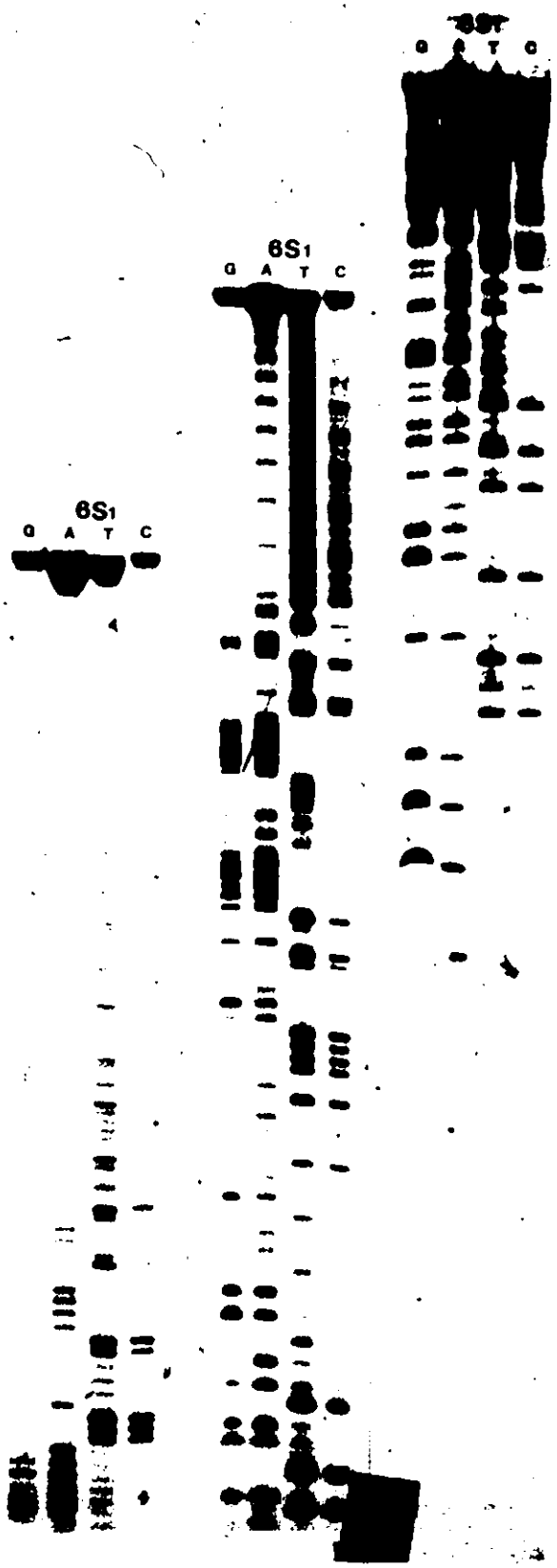

6S1
G A T C
6S1
G A T C
6S1
G A T C

FIGURE 43

Sequence gels of SstII fragments from pRIaa6 and pRIaa3. The 215 bp SstII fragment from pRIaa6, and the 205 bp SstII fragment from pRIaa3 which span the iterated inserts in these two plasmids were eluted, end labelled, strand separated, and sequenced. Samples were run on 20% (left panel) and 8% (middle and right panels) sequencing gels. 6S2 refers to the faster strand from pRIaa6, 3S1 refers to the slower strand from pRIaa3, and 3S2 refers to the faster strand from pRIaa3. Tracks marked G contain G specific reactions, tracks marked A contain G+A specific reactions, tracks marked T contain T+C specific reactions, and tracks marked C contain C specific reactions.

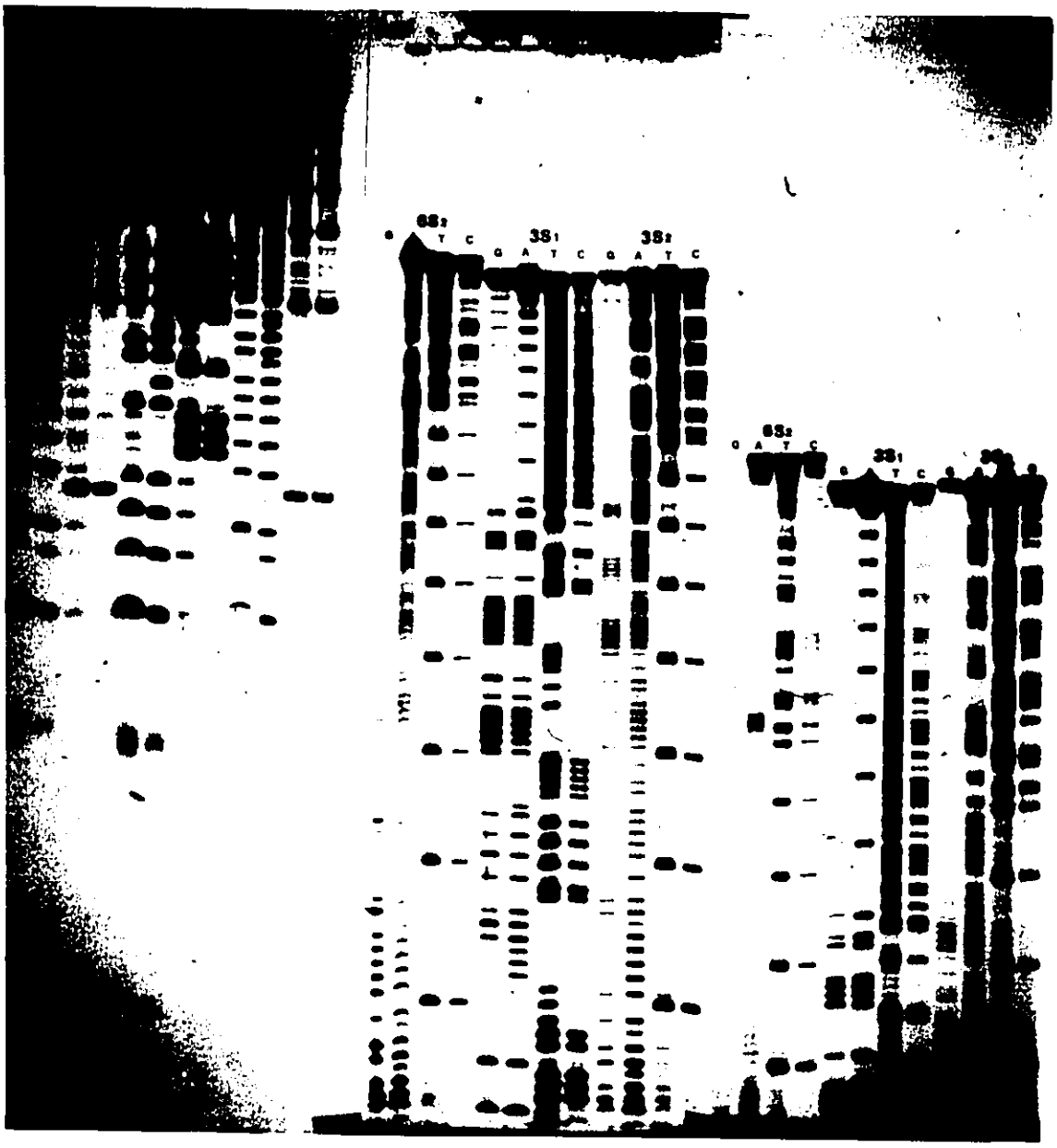

5S2
T C G A T C G A T C
3S1
3S2
5S2
G A T C
3S1
T C

FIGURE 44

Sequences of insert sites in pRIaa6 and pRIaa3. The 5' TK sequences shown here are the inverse complement of sequences read from Fig. 41, which were derived from analysis of pSBF14, and from other gels of sequence analyses of the homologue from pRIaa6. Bracketted sequences are inferred by the homology between sequences in pRIaa6 and those in SmaI F. The SstII sites flanking the junction between the inserts are underlined. Polylinker sequences from pUC7 are also underlined. Asterisks indicate the ultimate bases of the TK L, TK S, and legitimate S termini. Brackets surround sequences inferred from restriction analysis of the clone, and from homology with known sequences.

pRIaa6

5' TK sequences
GGGGCTCCGAGACAATCGCGAACATCTACACCACACAACACCGCCTCGACCAGGGTGAGATATCGGCCGGG

BamHI Ub sequences
GACGCGGCGGTGGTAATGACAAGCGCCccggatccgtcGGGGGGTGTGTTTTGGGGGGGGCCCGTTTCCGG

GGTCTGGCCGCTCCTCCCCCGCTCCTCCCCCC.....(SmaI F sequences).....GCGCCGCCCGG

Uc sequences SstII TK L terminus**TK S terminus
ACCGCCGCCCGCCTTTTTTGCGCGCGCGCGCGCCCGCGGGGGGCCCgacggaCGCGGCGGTGGTAATGACA

BamHI Ub sequences
AGCGCCccggatccgtcGGGGGGTGTGTTTTGGGGGGGGCCCGTTTCCGGGGTCTGGCCGCTCCTCCCCCG

SstII
CTCCTCCCCCC.....SmaI F sequences.....GCTCCCGCGG(CCCCGCCCCCAACGCCCGCCGCG

Uc sequences SstII
CGCGCGCGCGCCGCCCGGACCGCCGCCCGCCTTTTTTGCGCGCGCGCGC)CCGCGGGGGG

BamHI 3' TK sequences
CCCgacggatccggTGGCCCTCATCCCGCCGACCTTGCCCGGCACAAACATCGTGTTGGGGGCCCTTCCGG

AGGACAGACACACCGACCGC

pRIaa3 Sequence

BamHI
(GGTGAGATATCGGCCGGGGACGCGGCGGTGGTAATGACAAGCGCCggatccgtcGGGGGGTGTGTTTTG

SstII
GGGGGGGGCCCGTT.....SmaI F sequences.....GCTCCCGCGGCCCCGCCCCCAACGCCCGCCG

SstII
CGCGCGCGCGCGCCGCCCGGACCGCCGCCCGCCTTTTTTGCGCGCGCGCGCGC)CCGCGGGGG

TK L ter.**Legitimate S terminus SmaI
GCCCgacggaCCGCCGCCCGCCTTTAAAGGGCCGCGCGCGACCCGGGGGGTGTGTTTTGGGGGGGG....

SstII
....9 REPEATS ONLY....CCGCGG(CCCCGCCCCCAACGCCCGCCGCGCGCGCGCGCGCCGCCCGG

SstII BamHI
ACCGCCGCCCGCCTTTTTTGCGCGCGCGCGCGCCCGCGGGGGGCCCgacggatccggTGGCCCTCATCCCG

CCGACCTTGCCCGGC..)

of a TK L terminus with a legitimate S terminus, followed (or preceded) by recombination between the legitimate S terminus and the SmaI F insert. The difference in size between the (analogous although not homologous) SstII junction fragments from pRIaa6 and pRIaa3 was the apparent deletion of one of the 12 bp repeats from the legitimate S terminus involved in the recombination. Whether the deletion occurred during the recombination event leading to the formation of this new molecule, or whether it was pre-existing in the "a" sequence population is a moot point. Wagner and Summers (1978) had noticed a 10-12 bp size heterogeneity in the KOS "a" sequences which they designated F/F*. The 12 bp deletion in pRIaa3 DNA may represent this kind of heterogeneity.

The sequences of the insert region in pSBF14 and the various cloned viral fragments from KOSF10 are summarized in Fig. 45. Also present in this figure is the sequence of the KOS "a" sequence from a legitimate joint with the positions of legitimate terminal cleavage sites indicated. The interesting feature apparent in this figure is the spacing between the SmaI F borders and the terminal cleavage sites in both legitimate joints and the KOSF10 insert site (pSBF14 in Fig. 45); it is identical for both. This strongly suggests that the cleavage mechanism involves a measurement from the cleavage signal within SmaI F to distal sites within flanking DNA. The actual sequence of the latter may be irrelevant.

Cis-Acting Site Specific Recombination Signals Do Not Map Within "a"

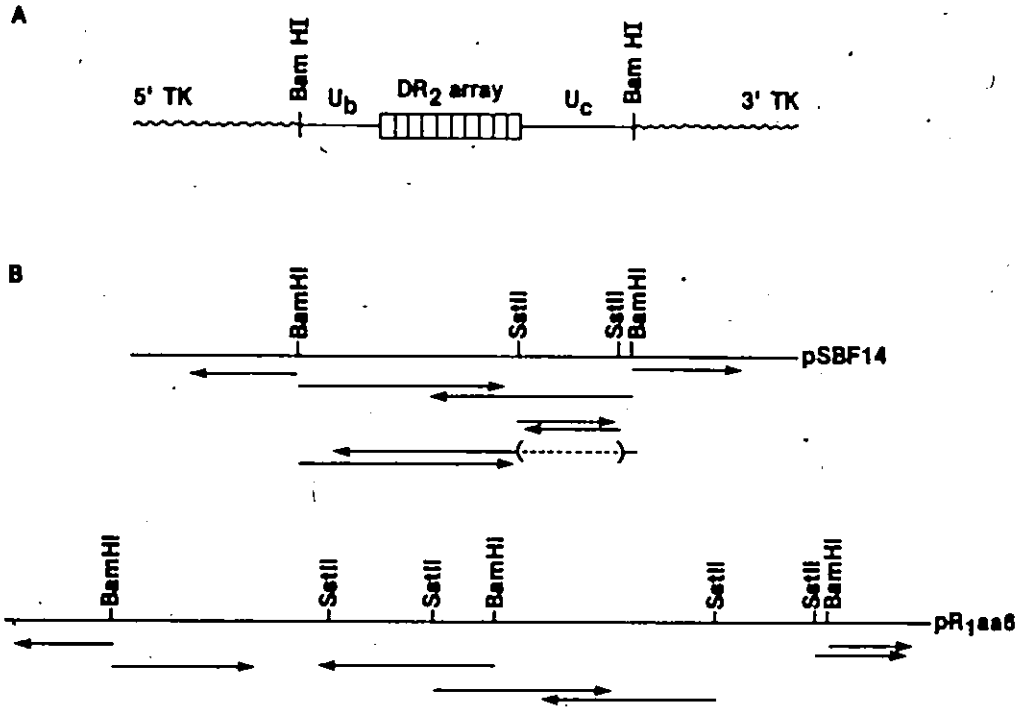
It is clear that one "a" sequence is invested with the ability to direct cleavage of the concatemer. As mentioned above, biphasic

FIGURE 45

(A) Structure of KOSF10 insertion mutation.

(B) Sequencing strategy. Plasmid pSBF14 was the plasmid used to generate KOSF10. Plasmid pRIaa6 is a cloned second step from an EcoRI stepladder of KOSF10, and contains a tandem iteration of the SmaI F insert.

(C) Nucleotide sequence of novel fragments. The sequences of clones representing TK L (pTKL13), TK S (pRIS3), reiteration bearing TK fragments of KOSF10 viral DNA (pRIaa6 and pRIaa3), and the insert site in KOSF10 (pSBF14) are shown aligned with the normal sequence of the KOS L-S junction. The large arrows above the KOS L-S junction and KOSF10 insert indicate sites where cleavage of viral DNA occurs. The small arrow above the KOS L-S junction indicates the predicted cleavage site for normal L termini based on homology with strain F (Mocarski and Roizman, 1982a), and on analogy with cleavage at the KOSF10 site. All three known terminal cleavage sites are the ultimate bases in cloned termini, and do not take into account any bases lost during cloning, as, for example, 3'OH protruding nucleotides may have been in some cases. SmaI F sequences are represented by dashed lines here, but they are aligned directly with each other. Bases on the Ub side of SmaI F are in capital letters, while bases on the Uc side of SmaI F are in lower case letters. Sequences derived from the normal L-S junction are underlined (see for example pRIaa3).



C

```
...GGGCTGCCCGCCGCCCGCCTTTAAAGGGCCGCGCGCGACCCCC--------gggctgc...  KOS  joint
...CCGGGGACGCGGCGGTGGTAATGACAAGCGCCCCGGATCCGTC--------gacggatccggtggccc...  pSBF14
          CGCGGCGGTGGTAATGACAAGCGCCCCGGATCCGTC--------gacggatccggtggccc...  lkS(pRIS3)
...CCGGGGACGCGGCGGTGGTAATGACAAGCGCCCCGGATCCGTC--------gacggat  lkl(pTKL13)
...CCGGGGACGCGGCGGTGGTAATGACAAGCGCCCCGGATCCGTC--------gacggaCGCGGCGGTGGTAATGACAAGCGCCCCGGATCCGTC--------gacggatccggtggcc...  pRIaa6
...CCGGGGACGCGGCGGTGGTAATGACAAGCGCCCCGGATCCGTC--------gacggaCCGCCGCCCGCCTTTAAAGGGCCGCGCGCGACCCCC--------gacggatccggtggcc...  pRIaa3
```

cleavage of the concatemer is sufficient to produce two out of four isomers. The other two must arise from biphasic cleavage of a different sort of concatemer in order to account for all four isomers. Thus, the ability of "a" to direct cleavage is almost sufficient to explain the apparent "site specificity" of recombinational events at the junctions. It was still possible that recombinational activity was a separate biological function mappable to some portion of "a".

I have shown in a previous section that segment inversion does not require identical "a" sequences. However, the sequence differences between strains F and KOS were found to reside in the number of several repeated sequences rather than the sequences themselves. This meant that the putative target sequence for site specific recombination may have been preserved between the two strains. If so, then it should be mappable by standard molecular genetic techniques. I chose to test the various subfragments of "a" for their ability to generate recombinant TK inversion products similar to those produced by the insertions of the L and S termini (see section "Sequence Identity of "a" is not Obligatory", and Fig. 9). PvuII and EcoRI digests of the seven insert mutants tested (SmaI A - KOSA23, SmaI F -KOSF10, SmaI I - KOSI, DR1 - KOSDR, f1 - KOSf1, f2 - KOSf2, and ΔII - KOSΔ II; see Fig. 25) were hybridized with ^{32}P-labelled TK DNA (Fig. 30). All of these viruses contained the insert at the TK locus as described above. The pattern produced by KOSF10 has already been analysed in detail, and most of the new fragments have been found to be the result of the cleavage/packaging mechanism. The faint bands marked with asterisks in Fig. 30 represent inversion products. Three of the inserts were capable of

producing this fragment - SmaI F, ΔII, and SmaI A. While the first two clearly overlap with eachother, neither overlaps with SmaI A, which produces as strong a signal as SmaI F. Three observations may be made from this blot:

1) The segment inversion target sequence cannot be mapped to a single sequence within "a". SmaI A is derived from the flanking "c" sequences. One might argue that there is a cryptic site within "c"; however, by similar logic one would have to argue that SmaI F and ΔII both contain the real site. Since the only sequence they share in common, f1, is inert, the logic of the idea crumbles.

2) Densitometric scans of these digests revealed that the proportion of L3S1+L3S2, with respect to the internal TK band (or family), increased as the size of the insert increased, ie, KOSA23 produced aproximately 1:2, KOSF10 produced approximately 1:4, and KOSΔII produced approximately 1:5 ratios of L3S1+L3S2 TK fragments. This observation is consistent with the idea that recombination frequency varies directly with the length of homology (Rubnitz and Subrameni, 1984)

3) None of these viruses produces all of the expected inversion fragments (Table 3). KOSF10 produces the new termini, as a result of cleavage, not recombination, and only L3S1+L3S2 recombinant fragments in a reproducible manner (both EcoRI and PvuII cleave within the inverted "c" repeat, so that S1 and S2 are identical). KOSA23 also produces only L3S1+L3S2 recombinants, and no new terminal fragments. Since SmaI A, the insert in KOSA23, was shown to be negative for cleavage/packaging function, the lack of terminal fragments is not surprising. KOSΔII presents a different profile in that the S3, but not

TABLE 3

RECOMBINANT TK FRAGMENTS IN KOSF10 AND KOSA23 VIRAL DNA

FRAGMENT		L3	S3	L3S33	L3S1	L3S2	S3L1	S3L2
VIRUS								
				PvuII Digests				
KOSF10	exp.	914	1414	2074	5360	5360	9160	12960
	obs.	840*	1350	2000*	5500	5500	-	(14500)
KOSA23	exp.	1607	2087	2767	5126	5126	9160	12960
	obs.	-	-	2600	5000	5000	-	(14500)
				KpnI Digests				
KOSF10	exp.	1244	3250	4240	8390	8990	6796	6796
	obs.	1160*	3200	4100*	8500	10000	6600	6600
KOSA23	exp.	1917	3927	4917	8136	8736	6800	6800
	obs.	-	-	5000	8200	10500	-	-
				SacI Digests				
KOSF10	exp.	1893	4614	6253	4039	4039	5540	5540
	obs.	1800*	4800	6200*	4200	4200	-	-
KOSA23	exp.	2587	5287	6947	3896	3806	5540	5540
	obs.	-	-	6800	3700	3700	-	-
				EcoRI Digests				
KOSF10	exp.	754	1904	2404	5100	5100	15650	13150
	obs.	754**	2050	2550*	6300	6300	-	-
KOSA23	exp.	1427	2577	3077	4846	4846	15650	13150
	obs.	-	-	3200	6000	6000	-	-

* bottom step of stepladder
** not observed on blot, but cloned and sequenced
\- not detected

the L3, fragment is produced (see arrows Fig. 30), and in very small quantities (1/13 - 1/25 from densitometric scans); again, only a subset of recombinant fragments, L3S1+L3S2, could be seen in digests of this virus. Since the test fragment in KOSΔII was shown to be negative in the defective assay for cleavage/packaging, the S3 seen here could not have arisen by this mechanism. Moreover, the f1 sequence, shared in common with SmaI F and ΔII, ie. the part of ΔII which might contain one of the cleavage/packaging signals, is inert. Therefore, the S3 seen in KOSΔ II DNA must have arisen by some other mechanism, possibly recombination. Taken alone, the data produced by KOSΔII support the idea that the physical ends of the viral genome are recombinogenic as long as there is a homologous sequence available for heteroduplex formation.

The absence from these virus stocks of fragments representing the reciprocal products of recombination between the inserts and junctions, ie the S3L1+S3L2 fragments, suggests that, if the L3S1+L3S2 fragments arose by recombination, it was either not a reciprocal event, or the reciprocal products were lost during amplification of viral DNA. The latter explanation has been invoked to account for non reciprocal recombination in bacteriophage T4 (Stahl, 1979), and there is some evidence to support the idea that HSV DNA also amplifies some sequences at the expense of others. This may occur as a result of recombinational deletion which produces two segments - one containing the S3 sequence and one containing the L3 sequence. As Fig. 46 indicates, the S3 containing segment does not possess an origin of replication (see loop in panel B), while the L3-containing segment possesses three origins (see loop in panel C). Amplification (although not packaging - its too

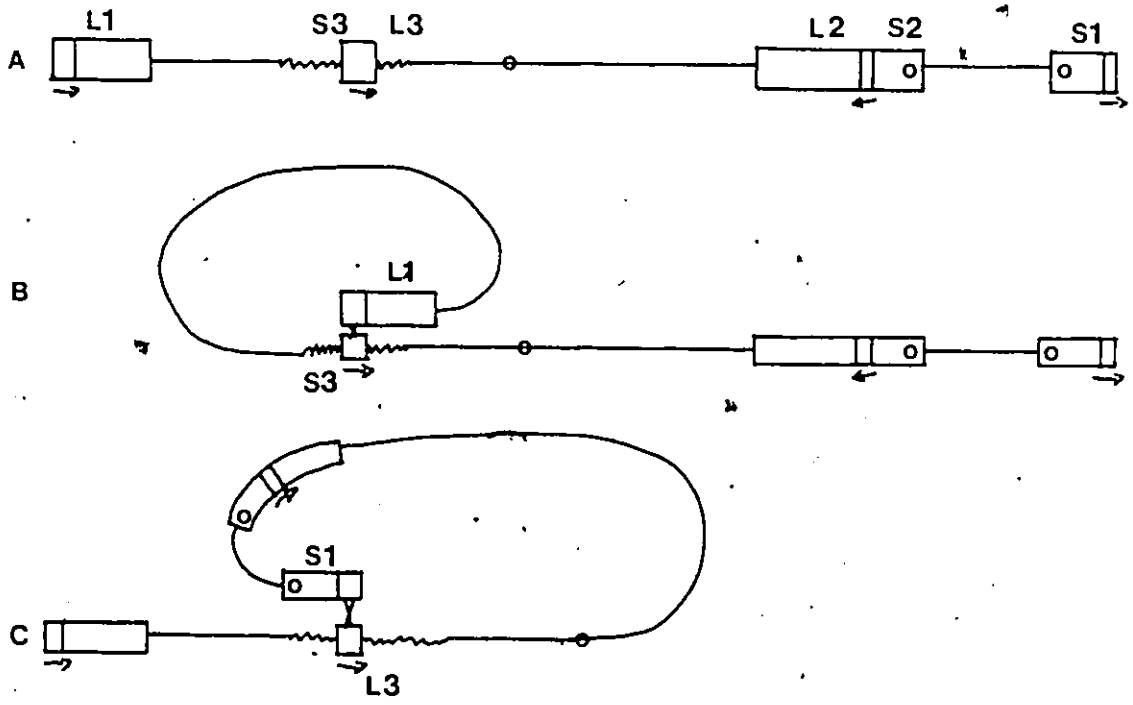
FIGURE 46

Recombinational deletions producing fragments without origins.

(A) A stylized diagram of the KOSF10 genome shows the orientation of the SmaI F insert with respect to the terminal "a" sequences (indicated by arrows). The wavy line represents the TK sequences used to detect recombinant fragments.

(B) Recombination between the L terminus and the insert creates an S3L1 fragment on a circular DNA which lacks an origin (letters o). These fragments were rarely detected in blots of KOSF10 viral DNA probed with TK.

(C) Recombination between the S terminus and the insert creates a circle containing an L3S1 recombinant fragment and three origins of replication. While this structure could not be packaged (the monomer is too small and the dimer is too large), it would be replicated, and would be detected in blots of viral DNA extracted from infected cells by the Hirt (1967) procedure.


A
L1
S3
L3
L2
S2
S1
B
L1
S3
C
S1
L3

large) of the latter combined with removal of the former from the replicating pool of viral DNA would effectively result in apparent non reciprocal representation of the two segments. Such deletions were shown to occur in viruses bearing larger joint-spanning fragments at the TK locus (Smiley et al, 1981).

DISCUSSION

I: Conclusions Drawn From This Work

In the Introduction a number of facts regarding HSV genome arrangement and cleavage were presented, and two older models of genome maturation, originally proposed by Roizman and coworkers, were discussed. Both models were found to be deficient in their explanation of some of the observed facts, specifically the amplification and distribution of the terminal "a" sequence, although both models took pains to justify the other feature of HSV DNA which has been a subject of study, segment inversion.

Several important new findings have come out of my work which may serve to clarify this area of research, and which provide a foundation for a model of HSV DNA maturation which encompasses more of the known facts than the previous two models. At the same time, new questions have arisen which will need to be addressed before a comprehensive picture can be drawn.

There are nine major conclusions which may be ascertained from this work:

1) Sequence identity of the inverted repeats is not obligatory, but heterozygosity in these sequences is unstable.

2) Heterologous "a" sequences are tolerated within a viral stock.

3) No "site" specifying a cis acting signal for site specific recombination maps within the "a" sequence, and in fact a fragment from

the flanking "c" sequence performs as well as or better than the most active subfragment of "a".

4) Both termini of defective genomes contain "a" sequences, even though the seed replicon only possesses one "a".

5) The cleavage signal resides within a subfragment of "a" which does not contain the cleavage site in legitimate viral DNA.

6) Cleavage occurs at a specific distance from the cleavage signal and may occur within sequences different from those at the legitimate junction.

7) There are most likely two cleavage signals, rather than one - one directing cleavage of the L terminus and one directing cleavage of the S terminus.

8) The "a" sequence amplification is a direct consequence of the cleavage mechanism.

9) Intermolecular ligation occurs.

These conclusions, discussed in detail below, provide the foundation for two models of HSV DNA cleavage and packaging which account for some of the previously anomalous data. The two models will be presented below.

1) Sequence identity of the inverted "c" repeats is not obligatory, but heterozygosity in these sequences is unstable such that heterozygous genomes quickly become homozygous. The evidence for this conclusion came from the observation that heterozygous genomes segregated roughly equivalent proportions of homozygous and heterozygous progeny. Unless one invokes a novel mechanism such as conversion-repair via reverse

transcripts (Baltimore, 1985), two explanations of the data are that (i) heteroduplexes across this region are readily formed, or (ii) multiple recombination events between molecules readily shuffle sequences around. This observation is consistent with the readiness with which ICP4 mutants, which map to the "c" repeat, may be isolated, and with Knipe et al's (1979) determination that intertypic marker rescue products contain homozygous ICP4 genes; the heteroduplex may extend quite some distance into the ICP4 coding sequence, or intermolecular recombination throughout the ICP4 gene occurs readily. Thus, mutations in one copy of the ICP4 gene appear at high frequency in the other copy. None of these experiments is capable of determining the mechanism by which the heteroduplex gives rise to a homoduplex since it is not possible to isolate all of the recombination products from a single event. Three possibilities spring to mind: (i) repair synthesis, in which one strand is degraded and replaced with a newly synthesized strand, (ii) "repair" by semiconservative DNA synthesis of the genome, and (iii) multiple crossovers at sites flanking the heterozygous region on either side. The site(s) of initiation of exchange cannot be determined from this work; however, two alternatives can be imagined: (1) initiation is "site specific", but results some of the time in extensive heteroduplex migration, since the deletion studied here does not overlap with any of the ICP4 mutants studied elsewhere (Knipe et al, 1979), or (2) initiation may occur anywhere within the inverted repeat, although there may be preferred spots. If (1) is correct, then the site specific mechanism involved is clearly different from classical site specific recombination mechanisms in other organisms,

where heteroduplex migration rarely occurs. The terminal "a" sequence has already been implicated as a site for site specific recombination (Mocarski and Roizman, 1982a). If this scheme is correct, there is no reason to exclude the possibility that heteroduplexes initiated in "a" migrate into flanking sequences. However, if "a" is a site specific initiator of heteroduplexes, then other sequences must be inert. The assay chosen by several investigators to test the recombinational activity of various sequences, duplication of those sequences by insertion at the TK locus, revealed that some fragments are inert (BamHI F and BamHI J, Mocarski et al, 1980), but that some are active (BamHI L, Pogue-Geile et al, 1985; SmaI A from SacI joint, my work), suggesting that, if initiation is site specific, "a" is not the only site capable of this activity. Explanation (2) satisfactorily accounts for the activity of non "a" sequences, discussed below, without setting any precedents. Generalized recombination, initiated at any site within the inverted repeats, could result in homologous "c" sequences (or homozygous ICP4 genes) by either heteroduplex repair or multiple exchanges on both sides of the heterozygous region.

2) In a similar vein, heterologous "a" sequences are tolerated within a virus stock, although they may be either amplified or eliminated as evidenced by the relative abundance of strain KOS and strain F "a" sequences in various KOSxF recombinants. Davison and Wilkie (1983) also found that intertypic recombinants with heterologous "a" sequences were viable, although unable to undergo frequent segment inversion. Since the heterologous inserts in the KOSxF recombinants supported frequent

segment inversion, some of which apparently resulted in the loss of the inserted "a" from the population (see for example lane marked KOScF1 in Fig. 12), exchanges across, or on either side of "a" must also occur. The same arguments described above apply to this observation. Thus, both "a" and "c" sequences are involved in recombination. As will become evident below, they are both apparently able to initiate exchanges. In addition, since the heterologous "a" sequences were equally adept at driving segment inversion, it could be argued that they share a common "site" for site specific recombination. The sequence data clearly indicate that the differences reside, not within the identity of the sequences, but within the numbers of repeated elements. Site specific recombination, therefore, cannot be discarded on this basis. However, as pointed out below, the "a" sequence is not unique in its ability to drive recombination, and so exchanges mediated by "a" are probably equivalent to exchanges mediated by "c". The apparent amplification of both types of "a" sequence within a heterologous stock indicate that each is capable of amplification, although it is not possible to distinguish whether this is a phenomenon that manifests itself within individual genomes, or whether it is a population effect, such that some genomes contain one type of "a" sequence which is in the correct context for amplification, while some contain the other. In a later section I will discuss the evidence which supports the first possibility, that is, that all "a" sequences, regardless of their position in the genome, are invested with the necessary properties which result in their amplification.

3) No "site" specifying a cis acting signal for site specific recombination maps within the "a" sequence, indicating that there is no site specific recombination mechanism which can be associated with the "a" sequence. As indicated in the Introduction, Mocarski et al (1980) had shown that segment inversion between inverted repeats of junction sequences, but not of other sequences, could occur. Since both termini were also capable of generating the full range of recombinant fragments, and since they shared only the "a" sequence (Mocarski and Roizman, 1982a), these authors concluded that segment inversion is mediated by a site specific recombination mechanism acting on "a", the target sequence. Most people in the field still believe this to be true because the data are compelling. However, new data are now available which do not support this conclusion. Several subfragments of "a" and the flanking SmaI fragment from "c" were tested in a similar assay. No single site could be established as the cis acting signal which drives segment inversion. Indeed, the flanking "c" fragment was more active than any of the "a" subfragments, of which only two, SmaI F and the _II subfragment showed any activity (see Fig. 28). Moreover, the sequence common to both SmaI F and _II, f1, was recombinationally inert, indicating that these sequences alone were not responsible for the activity of SmaI F and _II. Densitometric scans of the gels in Fig. 30 revealed that KOS_II (insert 200 bp) contained recombinant L3S1+L3S2 fragments at roughly 1/5 the frequency of the TK fragment, KOSF10 (insert 250 bp) contained recombinant fragments at 1/4 the frequency of the combined TK stepladder, and KOSA23 (insert 920 bp) contains recombinant fragments at 1/2.5 the frequency of the TK fragment. The data are insufficient to

establish actual rates of recombination, especially in view of the uncertainty about the identity of the recombinant fragments and whether they reside on unpackageable deletion concatemers as postulated in the Results section. However, it is evident that the frequency of their appearance in Hirt extracts of viral DNA can be correlated with the size of the insert. It is interesting to note that KOSf1 (insert 160 bp) contains no discernible recombinants. Rubnitz and Subrameni (1984) determined that the minimum length of homology required for homologous recombination in mammalian cells was approximately 200 bp. The insert in KOS_II lies just within this threshold, while the insert in KOSf1 lies below it. The crucial evidence for site specificity was the inactvity of non "a" sequences, which were derived from U_S (BamHI J, Mocarski et al, 1980; gene for gD, Gibson and Spear, 1983), and U_L (BamHI F, Mocarski et al, 1980; gene for gC, Lee et al, 1982). However, Pogue-Geile et al (1985) recently tested a different fragment from U_L, BamHI L, and found that it readily generated segment inversion. Thus, the new evidence indicates that the specificity associated with segment inversion is not confined to the "a" sequence. Instead, a deeper mystery has been revealed. Why are some sequences in the HSV genome able to drive recombination between non allelic sites while others are not, even though they are all apparently able to undergo recombination when they are in allelic configurations? One possible explanation is that the active fragments contain initiation sites, or "hot spots", for generalized recombination, while the inert fragments do not. The latter would still be able to recombine in an allelic setting because exchange could be initiated by flanking sequences. However, since exchange would of necessity need to be initiated in

homologous sequences, inert fragments at non allelic sites would not be able to participate in exchanges. If this is the case, then the central question becomes "why does HSV₁ DNA contain recombinational hot spots?" Are they fortuitous sequences which interact with cellular machinery or do they serve some purpose in the replication of HSV? Unravelling some of these issues could prove to be a fruitful area of research.

4) Both termini of defective genomes contain "a" sequences, without any apparent change in the seed replicon, which only has a single "a" sequence. Work from Niza Frenkel's lab had shown that the "S" termini of natural defectives contained an "a" sequence (Locker and Frenkel, 1979). Since the seed replicon only contained one "a", derived from the S terminal portion of the defective, it was assumed by most investigators that the cleavage of defective concatemers preserved the integrity of the S terminus, and that the defective L terminus would not contain an "a". No one examined the L termini in detail, and in fact, the apparent heterogeneity of defective L termini was assumed to be the result of heterogeneous starting material (Vlazny et al, 1982). Knopf et al (1983) had sequenced an L terminus from an HSV-1 strain Angelotti defective and found that it was covalently linked to sequences derived from "a", but these authors failed to analyse their defectives in detail, and the observation apparently went unnoticed by most other investigators. I have shown here that the L termini of engineered defectives contain (apparently) full length "a" sequences, even though the seed replicon only contains one (see Fig. 28). Although my data are insufficient by themselves, in conjunction with previous observations

that defective S termini contain full length "a" sequences, it is possible to conclude tentatively that both termini of defectives contain "a" sequences, even though the replicon only has one. This means that one "a" is somehow duplicated during cleavage to ensure that each terminus gets one. Below I discuss evidence from the study of viral insertion mutants that a single "a" sequence is capable of generating two termini each bearing an "a" sequence.

5) The cleavage signal sequence resides within a subfragment of "a" which does not contain the cleavage site in legitimate viral DNA. Work with defectives, mainly from Niza Frenkel's and Nigel Stow's labs (see Introduction; Locker and Frenkel, 1979; Spaete and Frenkel, 1982; Stow etal, 1983), had already shown that the "a" sequence was required for the effective cleavage/packaging of viral DNA into infectious particles. Mocarski and Roizman (1982a) had shown that the sequences of viral termini were consistent with the idea that they arose by cleavage of junctions containing two or more "a" sequences, at a specific site within the DR1 separating two "a"s. This suggested that the cleavage site, DR1, might be a good candidate for the cleavage signal. In fact, my data show that it is inert. Instead, the cleavage signal lies within a subfragment of "a" which does not contain a complete copy of DR1, and is missing the cleavage site in DR1 used by the virus to cleave legitimate junctions. This observation required a realignment of the logic. If the cleavage signal does not contain the cleavage site, then some other physical parameter must determine the apparent specificity of cleavage. The best interpretation of my data is that the physical

parameter must be distance from the signal (see below). However, even though this interpretation now appears to be correct, it does not explain why DR1 exists, and is conserved among HSV-1 strains, both as a sequence block and as the site for viral cleavage events. It is possible that DR1 may focus cleavage, perhaps in an analogous fashion with the activity of TATA boxes in focussing RNA transcription starts (Benoist and Chambon, 1981).

6) Cleavage occurs at a specific distance from the cleavage signal, rather than at a specific site. The evidence for this conclusion came directly from sequence analysis of cloned viral termini produced by cleavage at the insert in KOSF10, although, as pointed out above, the ability of SmaI F to provide the cleavage signal, even though it does not contain the cleavage site, indirectly supports the conclusion. The insert was the SmaI F subfragment of "a" which lacks the legitimate cleavage site in DR1. The cleavages in KOSF10 which generated TK termini were made in sequences flanking the insert at distances analogous to those in a legitimate junction. The sequences where cleavage occured in KOSF10 were quite different from those in legitimate junctions. Thus, the signal in SmaI F directs the cleavage machinery to "reach out and cut" at a distance which may be determined by stoichiometry of the physical machinery, helical pitch of the DNA, or both. At least one restriction endonuclease, HgaI, cleaves DNA in unrelated flanking sequences 5 to 10 bp on one side of the recognition sequence. The L terminal cleavage signal lies 6 to 7 bp down from the SmaI F border, so a mechanism similar to that employed by HgaI may possibly

make the L terminal cleavage, although the L terminal signal may actually be much further away (34-40 bp, see below). The S terminal cleavage occurs in sequences displaced from the SmaI F fragment by at least 36 bp, and again, the signal may be even further away (45-53 bp). These distances suggest that the cleavage machinery must be quite large. E. coli RNA polymerase holoenzyme, a complex of 465 kdaltons, covers only 40 bp of DNA (Kornberg, 1974), and SV40 T antigen tetramer, 376 kdaltons, covers only 50 bp (Meyers et al, 1981). Thus, the interaction between the signal(s) and the cleavage machinery may prove to be highly complex. Whether DR1 "fine tunes" the mechanism remains to be determined.

7) There are probably two cleavage signals, one which directs cleavage of S termini, and one which directs cleavage of L termini. The evidence for this conclusion is three-fold. The first is that two different cleavages are made in KOSF10. Examination of the sequences flanking the SmaI F insert in KOSF10 in which the L and S terminal cleavages are made indicates that, with the exception of a GGA trinucleotide, they are completely different from each other, and from DR1. Second, subfragments of SmaI F, f1 and f2, which represent two segments of SmaI F, are each inert. By itself, this observation is inconclusive since the SstII cleavage site used to separate f1 and f2 might possibly span the cleavage signal. However, preliminary evidence suggests that f1 and f2, if placed on a plasmid in the correct orientation to one another, but separated by spacer DNA, can function as a packaging signal to generate defective viral DNA (Smiley, pers. comm.). Third, "a" sequences from

different strains of HSV, in which the internal iterations vary in number and identity, but in which the outer Ub and Uc sequences are conserved, retain the specificity of terminal cleavage (see Fig.14). I have shown that the cleavage is determined by distance from the signal. If only one signal were operating on the system, then the "a" sequences from different strains might be expected to be uniform in size, but variable in terminal sequence. Instead, they are variable in size, yet uniform in terminal sequence. This argues that the internal variation in the DR2 array has no effect on the cleavage mechanism. Furthermore, f1, which contains the internal DR2 array in its entirety, is inert, indicating that DR2 is not itself the cleavage signal. The best interpretation of the data is that a signal in Uc directs cleavage of the L terminus and another signal in Ub directs cleavage of the S terminus. In normal junctions, the two cleavages are made in the same sequence in DR1, although not necessarily in the same copy of DR1. Thus the final result is that the two cleavages appear to be a single cleavage. In a later section I will speculate on the nature of the two signals, and compare them with possible counterparts in other Herpesviruses.

8) The "a" sequence amplification is a direct consequence of the cleavage reaction. Analysis of the amplified "a" sequences in the TK stepladder in KOSF10 provided the first clue that cleavage, and not recombination, was responsible for this phenomenon (see Results section). Sequence analysis of a cloned second step confirmed the interpretation. As illustrated in Fig. 44, end to end ligation of TK L

and S termini yields the sequence found at the junction between two "a"s in the cloned second step, pRIaa6. Furthermore, restriction analysis of 26 cloned second and third steps indicated that, at least at this gross level, 24/26 clones were identical (in terms of the "a" sequence junctions) with pRIaa6. The original KOSF10 genome must have contained a single insert. After cleavage at this insert, the resulting TK L and S termini each possessed an "a" sequence with flanking TK sequences attached. Religation of those genomes thus cleaved must have generated a TK locus with two tandem "a"s separated by the TK sequences derived from cleavage on either side of the original insert. The third steps would then be produced by similar cleavage/duplication/religation at the second step locus. "a" sequence stepladders in wild type virus are most likely generated by the same mechanism which produces "a" iterations in KOSF10, that is by duplication of "a" during cleavage followed by religation of termini. Below I will present two alternative models of cleavage/packaging which account for the duplication of the "a" sequence, both in "a" sequence stepladders and in defective genomes derived from replicons with a single "a".

9) Intermolecular ligation may occur in HSV DNA. Two of the 26 clones analysed with restriction enzymes for their structure across the "a" junctions displayed a pattern consistent with that predicted for recombinants (see Fig. 35). It was expected that the junction between these "a"s would be identical with the junction between "a"s at legitimate joints as determined by Mocarski and Roizman (1982a), that is, Uc-DR1-Ub. The Ub sequences were expected to contain the SmaI I

sequence missing from the SmaI F insert, and the SmaI site which separates SmaI I and F. The latter two predictions proved correct; however, the junction between Uc and Ub did not contain a complete copy of DR1, and instead contained six base pairs from the 3' flanking region of the SmaI F insert in KOSF10, the same six bases present on TK L termini. There is no way that homologous recombination alone can insert those six base pairs between Uc and Ub. Genomes with TK L termini must also have TK S termini, at least initially, since cleavage occurs at genome length intervals, and since the SmaI F inserts in a concatemer are separated by a genome length of HSV DNA. Homologous recombination between TK termini and legitimate junctions can only occur within SmaI F; thus recombination between a TK terminus and a legitimate junction, which would produce at a gross level a genome with one TK terminus and one legitimate terminus, would be unable to remove the extra bases at the ends of the TK termini derived from the flanking TK sequences. Subsequent intramolecular ligation of such a genome would produce two "a"s whose junction would contain the six bases from the TK L terminus and the thirty six bases from the TK S terminus. The only way the observed structure could have arisen was by intermolecular ligation between a TK L terminus and a legitimate S terminus. The latter must also have recombined, within the SmaI F sequences, with the insert at the TK locus, either before the intermolecular ligation or after. Only one of the two "recombinants" was analysed by sequencing the "a" junctions. The other clone was a third step, ie contained three tandem "a"s. Restriction analysis indicated that only two of the "a"s were recombinant. The restriction pattern was consistent with its

having arisen by intermolecular ligation of a TK L terminus containing two "a"s with a legitimate S terminus (see Fig. 41).

II: Models For Cleavage/Packaging of HSV DNA

1) Old Data

Before presenting the revised model of HSV DNA cleavage and packaging that incorporates my findings, a reiteration of some of the important facts concerning HSV DNA arrangement and maturation is in order:

a) The genome is subdivided into two segments which are flanked by inverted repeats, and which invert with respect to each other to generate an equimolar mix of four isomeric forms of viral DNA. It should be pointed out here that a single pfu must be amplified through at least five rounds of viral replication before sufficient quantities of viral DNA for visual analysis have been produced. Thus, the "rapid" equilibration of viral DNA into the four isomers must be viewed as a cumulative phenomenon spanning many rounds of DNA replication (in which a single pfu is expanded to 10^7 pfu).

b) The terminal "a" sequences are in direct orientation to one another. Moreover, sequence analysis indicates that the "a" sequence displays no dyad symmetry. Instead, the sequence has a directional asymmetry. In this regard, the sequences at the ends of the L and S termini are different from one another.

c) The "a" sequence is amplified in a very stereotyped manner, with the most abundant class of fragments being those with one "a", the second

most abundant being those with two "a"s, etc. Refer to Fig. 2 for a typical "stepladder" profile.

d) The "a" sequence amplification is asymetrically distributed to the L terminus, but not to the S terminus.

e) The cleavage/packaging intermediate of HSV DNA is a head-to-tail concatemer. The mechanism by which it arises is unknown.

f) Maturation of HSV DNA may be coupled to packaging as is the case for Pseudorabies virus, another herpesvirus (Ladin et al, 1980).

2) Model I

With all of these facts now at our disposal, it is possible to deduce a mechanism of HSV cleavage which accomodates most of them. We should start with the HSV concatemer already preformed, as there is no satisfactory data on the mechanism by which it arises, although I will speculate on this matter later in the discussion. The concatemer may possess one or more "a" sequences at any of its junctions, but let us assume for now that all of the junctions in our concatemer contain a single "a". The cleavage mechanism begins by the random choice of a junction by the cleavage machinery. Staggered nicks are made in opposing DNA strands at either end of the "a" sequence, one specified by the S terminal cleavage signal, and one specified by the L terminal cleavage signal. Repair synthesis across the gap in both directions creates two termini, each with an "a", derived from a junction with one "a" (Fig. 47A). Viral DNA is then rolled into the genome from one of these termini until two conditions are fulfilled:

i) the capsid is full, and

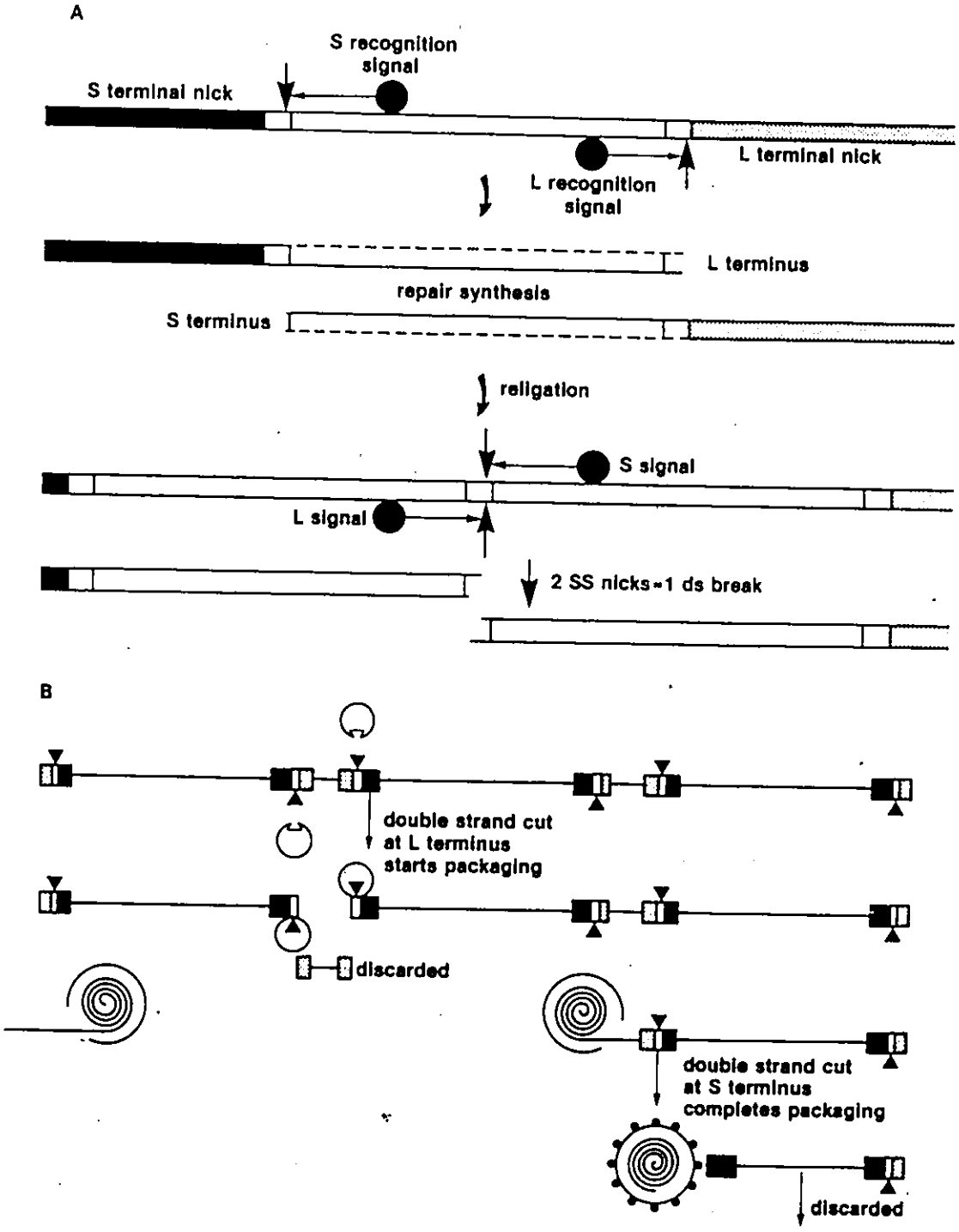
ii) an "a" sequence in the correct orientation is encountered.

FIGURE 47

Cleavage/Packaging Models.

(A) Model I. Cleavage machinery recognizing the S and L recognition signals in SmaI F makes nicks at measured sites from the respective signals. Repair synthesis (dashed lines) using 3'OH ends as primers duplicates the cleavage signal and separates the newly created termini. Subsequent religation creates a tandem duplication of the signal. 3'OH protruding nucleotides may be produced by cleavage at a tandem duplication such that the L terminal nick in the proximal signal and the S terminal nick in the distal signal together make a staggered double strand break with a protruding 3'OH nucleotide.

(B) Model II. A double strand break at the L recognition site is followed by unidirectional packaging starting at the L terminus. When a headfull of viral DNA has been packaged, the cleavage machinery then makes the second double strand break, but this time at the S recognition site. Thus, both termini end up with "a" sequences attached, but at the expense of their nearest neighbours. Segments which donate "a" sequences may be degraded. Solid bars represent "b" sequences, stippled bars represent "c" sequences, and open bars represent "a" sequences.


A
S recognition signal
S terminal nick
L terminal nick
L recognition signal
L terminus
repair synthesis
S terminus
religation
S signal
L signal
2 SS nicks=1 ds break
B
double strand cut at L terminus starts packaging
discarded
double strand cut at S terminus completes packaging
discarded

The cleavage reaction described above is then repeated.

3) Model II

The second model also requires random choice of a junction within a concatemer as the start site. A double stranded cleavage directed by one of the cleavage signals, eg. the L signal, is made. Where junctions with only one "a" are chosen as the cleavage site, such a double stranded cleavage will result in two termini, one with the "a" and one without. Since no termini without "a"s have ever been seen, the model requires that the "a"-less terminus thus produced will be discarded, perhaps by nuclease digestion. The viral DNA is then rolled into the capsid, starting with the terminus containing the "a" produced by the first cleavage, until the two conditions mentioned above, namely a full capsid and presentation of an "a" in the correct orientation, are met. At this point, the second cleavage, eg. the S terminal cleavage, is made (Fig. 47B). The genome segments on either side of the encapsidated genome will be sacrificed as they will not contain an "a" sequence. Since "a"-less segments have never been seen in restriction patterns of viral DNA, flanking segments which are sacrificed must be degraded by nucleases at a rate that ensures their heterogeneity during cleavage of concatemers.

Religation of genomes derived by either Model I or Model II will generate junctions with two tandem "a" sequences. In this way the number of "a" sequences in the population of viral genomes will amplify. Sequence analysis of the amplified "a"s in KOSF10 verified this conclusion. The junction between the two "a" sequences was composed of sequences derived from the TK L and S termini, which in

turn were composed of TK DNA. These models differ from the older ones described in the Introduction in that no constraint is placed on the choice of junctions for cleavage, ie, any junction with an "a" can be cleaved, and, two separate cleavages are made. These two departures allow a rational explanation of (i) the ability of defectives with single "a"s in the seed replicon to be packaged, and of (ii) the tandem amplification of "a", both phenomena which were not satisfactorily explained by previous models. The ability of HSV to use any junction as a cleavage site also provides a partial explanation of segment inversion. As pointed out in the Introduction, biphasic cleavage/packaging will produce two of the four isomers. This will be discussed below.

4) Asymmetric Distribution of Tandem "a" Sequences

Both Model I and Model II are insufficient to explain the asymmetric distribution of tandem "a"s to the L termini but not to the S termini. The asymmetric distribution of tandem "a" sequences is clearly a consequence of the cleavage/packaging mechanism because the TK L and S termini also display this feature. The designations TK L and TK S were originally made on the basis of the asymmetry, but sequence analysis later confirmed that this was the correct designation as the orientation of SmaI F within the TK S terminus is identical with the orientation of SmaI F within the legitimate S terminus, but inverted with respect to both the TK L and legitimate L termini. The positive association between a particular orientation of SmaI F and a particular asymmetry, ie. one or variable numbers of "a", argues in favour of directional or asymmetric cleavage/ packaging of the concatemer. Directionality must be determined by the directional nature of the "a"

sequence cleavage signals, which must interact with a similarly directional or asymmetric cleavage machinery. How does this structural and informational asymmetry impose itself on a concatemer, which probably contains identical junctions placed at genome length intervals? Two possibilities come to mind:

(i) The first cleavage event occurs at random, but the DNA is rolled into the capsid from one end only, the L terminus (Fig. 48A). This must be followed by a non random cleavage at the first encountered "a" to generate the S terminus. Thus, L termini may have a random number of tandem "a"s, but the S terminus will always have only one. Junctions formed by ligation of L and S termini will have a random number plus one of "a" sequences. This scenario accomodates itself much better to Model II than to Model I. The first cleavage event is pictured as random, while the second cleavage event is pictured as non random, indicating a change in cleavage reaction specificity. Model I does not distinguish the cleavage reactions which generate the L and S termini. The substrate is the same, and the reaction itself produces identical products in each case. Thus, the change in specificity from a random L terminal cleavage reaction to a non random S terminal cleavage reaction is not reflected in the molecular apparatus, at least at the level of the substrate and product of the reaction. Model II, on the other hand, requires two different reactions, utilizing two different substrates (although it does not preclude that both signals be present for each reaction, a situation which must apply since both f1 and f2 are inert at causing cleavages in viral DNA in infected cells - Jim Smiley, pers. comm.)

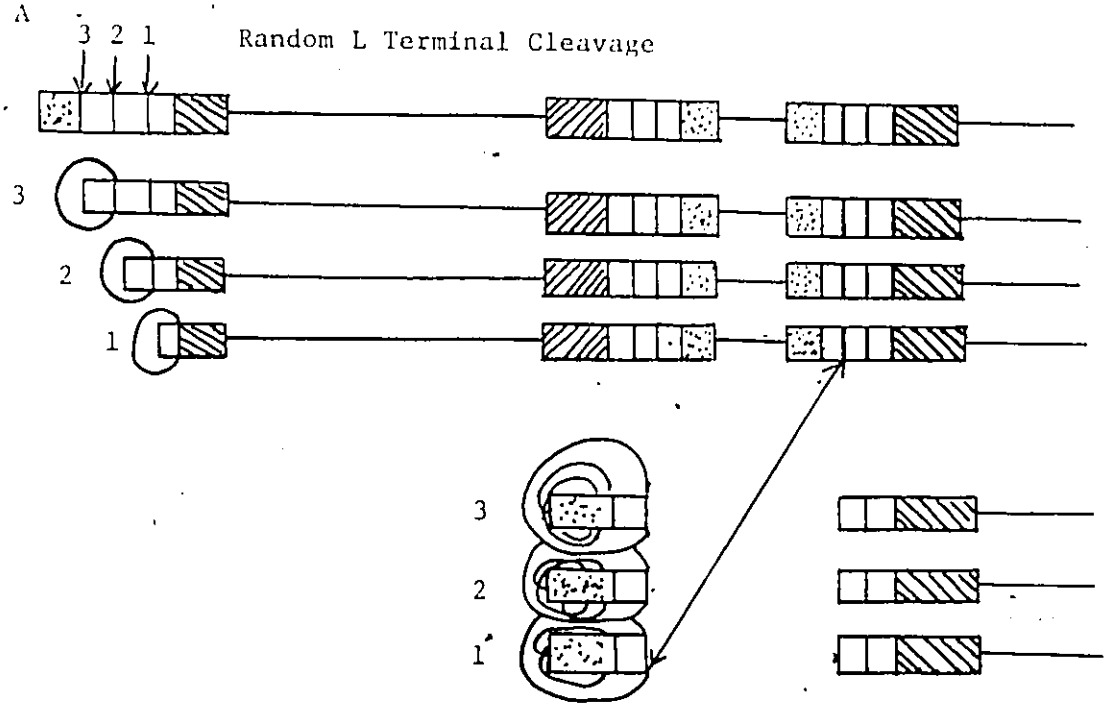
FIGURE 48

Asymmetric Distribution of "a".

(A) The first cleavage reaction may occur at any "a" sequence in a junction (1, 2, or 3), but packaging only proceeds from the L terminus. Thus, the random first cleavage, coupled with the unidirectional packaging ensure that L termini contain a random number of "a" sequences. When a headfull of viral DNA has rolled into the capsid, the S terminal cleavage reaction occurs at the first available "a" sequence (arrow), ensuring that S termini only have one "a" attached.

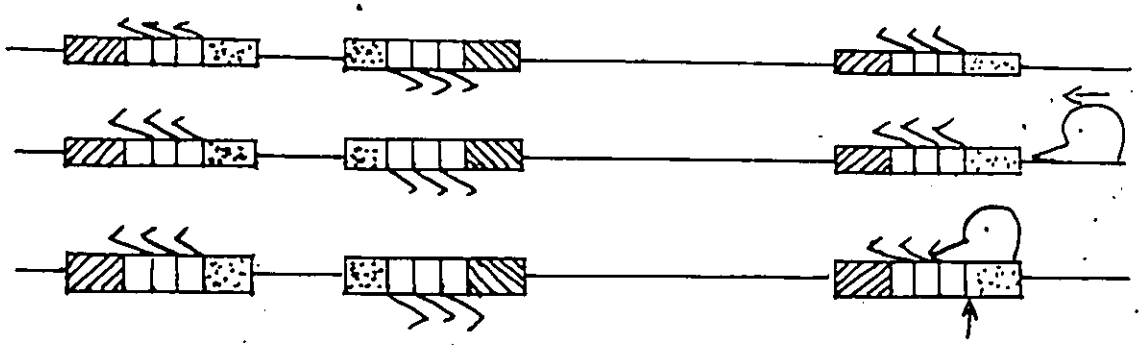
(B) The cleavage machinery may bind to viral DNA randomly, but can only scan along it in one direction (ie. it follows its nose). If it encounters an "a" sequence in the wrong orientation, it will keep on going. However, if it encounters an "a" in the correct orientation, it will immediately stop and initiate cleavage. In order to account for the asymmetric distribution, the correct orientation must be that presented by the "a" sequence to a cleavage machinery moving along the genome from "c" to "a"; that is, the "a" closest to "c" is always the preferred cleavage site, because it is the first one encountered. Bidirectional packaging can be accomodated within the context of this model if the scanning machinery simply serves as a marker. Thus, all of the "a" sequences closest to "c" sequences can be distinguished from their neighbours, and in an asymmetric fashion. Stippled boxes represent "c" sequences, crosshatched boxes represent "b" sequences, and open boxes represent "a" sequences. The barbs on the "a" sequences in (B) represent the directional nature of the signals in "a".

A

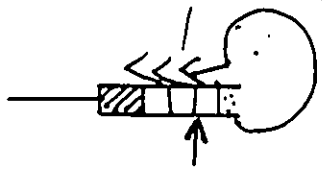


Non Random S Terminal Cleavage

B



L Terminal Cleavage at first "a"



S Terminal Cleavage at first "a"

The idea of directed cleavage/packaging is not unique to HSV. Bacteriophage lambda also appears to prefer to start packaging at one particular end, the left end, and proceed to the right end (reviewed in Lewin, 1977). Lambda concatemers are cleaved at a specific site called cos (for cohesive end site). A 12 bp staggered cleavage results in complementary ends which reanneal upon infection. Linear and circular monomers cannot be packaged, indicating that two intact cos sites are required for the cleavage/packaging reaction. Hence the necessity for concatemeric DNA. Genetic studies have indicated that cleavage and packaging are probably coupled events, since several mutants which prevent assembly of capsids also prevent maturation of concatemeric DNA. The same situation holds for phages T4 (Lewin, 1977) and T7 (Kruger and Schroeder, 1981).

The mechanism described above depends on unidirectional packaging of HSV DNA from the L terminus. Zeller et al (1984) found some evidence that some strains of HSV-1 are indeed packaged unidirectionally, while other strains are not, including KOS. Given the high degree of homology among different strains of HSV-1, it would be profitable to confirm this result before discarding the mechanism described above. If it is true that HSV-1 can package its DNA from either end, then this mechanism is incorrect.

(ii) The viral cleavage machinery may scan concatemeric DNA, in a unidirectional fashion (Fig. 48B). Since the machinery itself must have asymmetric components which recognize the asymmetry in the cleavage signal, it is reasonable to assume that it is capable of movement along a DNA strand in only one direction, although the recognition components

may not be directly involved in the migration. When it encounters an "a" in the correct orientation, that is, the orientation which can interact with the asymmetric cleavage machinery, the first cleavage is made. DNA is rolled into the capsid until the capsid is full, and the second cleavage is then made at the first encountered "a" in the correct orientation. This model necessitates that the first "a" in each case be the one closest to "c" to account for the asymmetry. This scenario as it stands requires that packaging proceed in a unidirectional fashion from the L terminus to the S terminus. However, if one assumes that the scanning apparatus does not itself perform the cleavage, but only serves as a "marker", then the unidirectional constraint can be removed. As pointed out above, a mechanism dependent on unidirectional packaging may not be tenable. Both Model I and Model II may be adapted easily to this mechanism as the two cleavage reactions are equivalent. The scanning mechanism predicts that packaging from the concatemer is random and will result some of the time in the abortive cleavage of subgenomic L and S segments if two initial cleavage events occur within one genome length of viral DNA.

5) 3'OH Protruding Ends On Viral Termini

An observation made by Mocarski and Roizman (1982a) with respect to viral DNA was that the termini each contain 3'-OH protruding nucleotides. If Model I is correct, then the ends which Mocarski and Roizman (1982a) sequenced could not have been derived from junctions containing a single "a", since the repair synthesis could not go beyond the end of the template strand. Since they specifically chose termini with single "a" sequences, they could have been derived by nicks in adjacent "a"s

such that the S terminal nick of the preceeding "a" and the L terminal nick of the next "a" in effect created a staggered double stranded break with protruding 3'-OH ends (Fig. 47A). This kind of cleavage would not require repair synthesis, although how the virus discriminates between staggered nicks that do require repair and staggered nicks that do not (there is no reason to believe that the other set of nicks would not be made at the other ends of the two "a" sequences involved) poses a further mechanistic problem. Moreover, neither of the proposed mechanisms explaining asymmetry can be easily adapted to staggered cleavage of adjacent "a"s since they both depend on a weighted preference for the "a" sequence closest to "c", at least in the generation of S termini. Model II obviously does not need such a complicated explanation of the data; the 3'-OH protruding ends could be created if the two cleavages make staggered double strand breaks. However, this poses a slight dilemma with respect to Model II. Two of the cloned termini, pTK173S7 and pTKL13, were prepared by ligating linkers directly to viral DNA. Either these clones were the result of the serendipitous removal of protruding bases which would have interfered with ligation of linkers, or some of the viral DNA was naturally flush ended. It is difficult to visualize a cleavage machinery which will sometimes make staggered double strand cleavages, and sometimes make flush end double strand cleavages. The first explanation may be closer to the truth since both of the mentioned clones were extremely rare, while cloned TK S termini made by treating viral DNA with T4 polymerase were more abundant (data not shown).

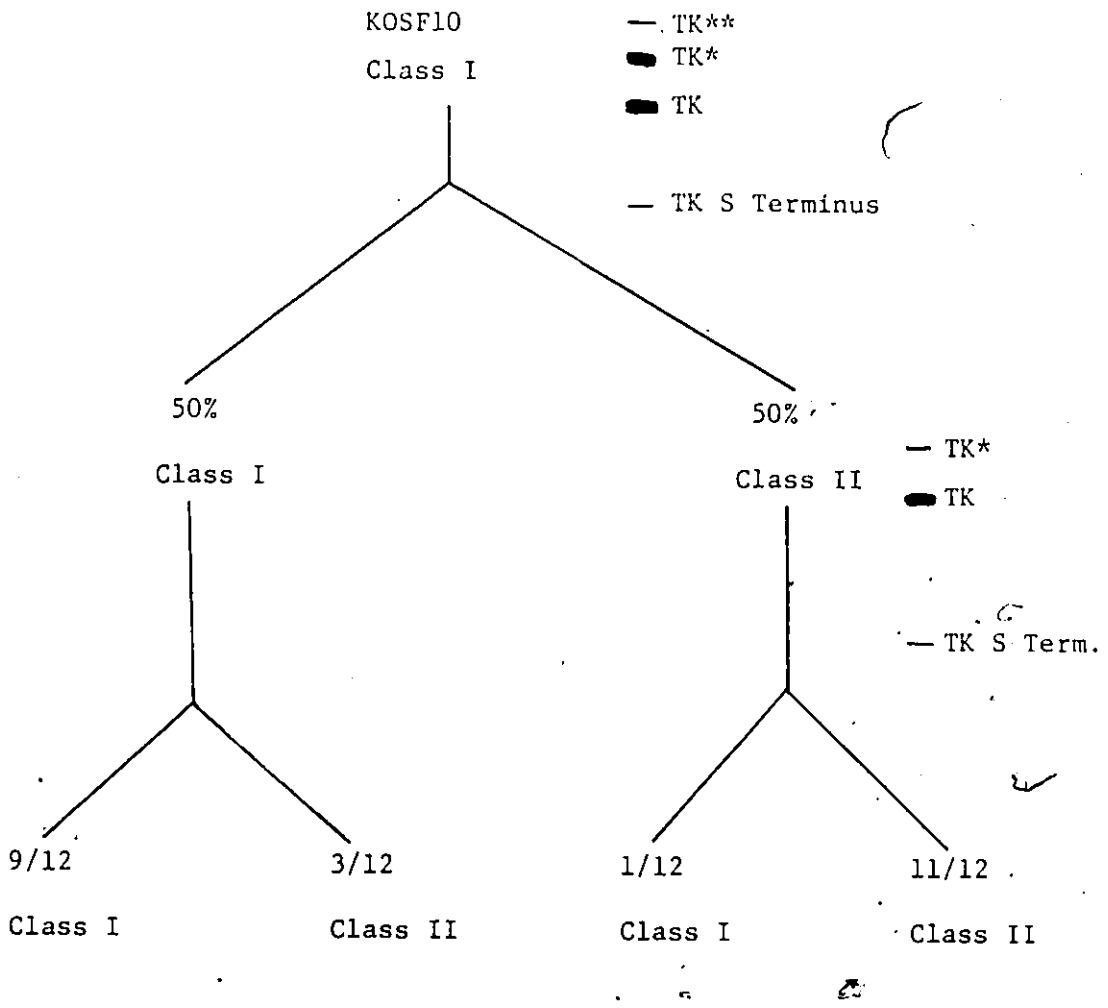
6) "a" Sequence Stereotypic Amplification and Reduction

The basic model described above (ignoring the picayune distinctions between Model I and Model II) neatly accounts for a number of hitherto puzzling features of HSV DNA. The mechanism whereby the "a" sequence amplifies itself can now be accorded the relatively important status it deserves with regard to the mechanism of HSV DNA maturation. The explanation for asymmetric distribution of "a" sequences, while still speculative, is at least logical, and appears to enjoy a precedent in nature. Still to be explained satisfactorily is the stereotypical pattern of "a" sequence amplification. The model described above predicts that, even though one "a" is sufficient to generate termini, the most abundant junctions should be those with two or more "a" sequences, since the cleavage mechanism appears to amplify the "a" copy number. This is clearly not the case (see Fig. 2). Therefore, there must be a mechanism which reduces "a" copy number. The obvious choice is recombination between termini across "a". However, new information indicates that this simplistic explanation may not be correct (Smiley, pers. comm.).

Clonal isolates of KOSF10 segregate into two main groups at roughly equal frequencies (Fig. 49). The first group (class I) is indistinguishable from KOSF10 in possessing a strong TK stepladder. The second group (class II) appears to have lost most of the TK stepladder, retaining only a small amount of the second step containing the amplified insert. In fact, class II isolates show the pattern that would be expected even in the original F10 isolate based on the frequency of cleavage at the insert; the quantities of the TK* fragment should

FIGURE 49

Clonal isolates of KOSF10 fall into two classes. The lineage of clonal isolates derived from KOSF10 is shown here. Stylized representations of the restriction patterns of TK containing fragments is shown. At the top is the pattern displayed by class I isolates, and further down is the pattern displayed by class II isolates. The patterns are observed in blots of restricted viral DNA which have been hybridized with ^{32}P-labelled TK DNA. (Jim Smiley, pers. comm.)


KOSF10
Class I
TK**
TK*
TK
TK S Terminus
50%
Class I
50%
Class II
TK*
TK
TK S Term.
9/12
Class I
3/12
Class II
1/12
Class I
11/12
Class II

reflect the frequency with which the insert is chosen as a cleavage (and subsequent religation) site. Thus, the second step in the step-ladder should be as abundant as the TK S terminus, or the sum of TK L termini. Class II isolates behave as expected, but class I isolates, including KOSF10, contain far greater quantities of second (and third) step TK fragments than would be expected. These isolates must therefore reflect stocks in which the relatively rare cleavage at TK occurred early during amplification of the stock from a single pfu to 10^7 pfu. When a class I clonal isolate was further subcloned, it produced 9/12 class I progeny and 3/12 class II progeny. A class II clonal isolate produced 1/12 class I and 11/12 class II progeny (Fig. 49), indicating that the frequency with which class I isolates, including KOSF10, arise is approximately 1/12.

The observation that class II isolates segregate out mostly parental types (ie. class II), and occasionally class I types was not surprising. In fact, these isolates behaved as predicted from the observed frequency of cleavage at TK. However, plaque purification of a class I isolate was expected to produce viruses with TK (ie. class II), TK*, or TK** fragments only, at frequencies reflecting their relative abundance in the TK stepladder of the class I isolate. Instead, apart from the class II progeny, only viruses with <u>both</u> TK and TK* fragments were found. This suggested that the bottom steps in the class I isolates may have been produced during expansion of a virus derived from a genome with a TK*, or TK** insert. Since all of the class I isolates displayed very strong TK bands, the mechanism by which TK* was reduced to TK must have been rapid in comparison with the mechanism by

which TK was amplified to TK* (ie. cleavage/packaging/religation). This suggests that the reduction mechanism is a separate phenomenon from the cleavage mechanism, and operates with different kinetics, at least with respect to the cleavage kinetics at the KOSF10 insert. One possible reduction mechanism is intramolecular recombination through the internal repeats in the "a" sequence. In legitimate junctions, amplification occurs roughly 50% of the time, ie. each time a junction is chosen for cleavage. In order to account for the stepladder profile, the rate of reduction must be slightly less than the rate of amplification, but large enough to ensure that the most abundant junctions are those with one "a".

7) Multiple Cleavage Signals In HSV and Other Herpesviruses

The rationale for concluding that there are two cleavage signals in HSV has been described above. What might those signals be? Perusal of the base sequence in "a" reveals that, within an unrelieved background of G+C, two blocks of conserved Ts stand out (Fig. 50) - one in Uc (TTTTTT) 33 to 34 bp 5' to the L terminal cleavage site (31-40 bp for HSV-2), and the other in Ub ($TGTGTTT^{T}_{C}$) 41 to 47 bp 3' to the S terminal cleavage site. The TTTAAA sequence also common to HSV-1 strains lies within the dispensable SmaI I fragment. These two blocks of Ts and/or the flanking Gs, especially evident in the Ub region, are good candidates for the two packaging signals, although only genetic analysis can determine this.

Several other Herpesviruses contain similar blocks of Ts surrounded by G+C sequences near their genomic termini. Epstein-Barr Virus (EBV) and Human Cytomegalovirus (CMV) are both terminally redundant,

FIGURE 50

Junction sequences from various strains of HSV-1, and from other Herpesviruses, at which terminal cleavages are made. DR1 sequences are underlined. Where data is available, asterisks indicate the ultimate base in the respective terminus, or clone. The sequences have been aligned so that the putative signals, which are boxed, are lined up. "a" sequences are in capitals, all others are in lower case letters. Sequence data were taken from :

Davison and Wilkie (1981) - strains 17, USA-8, and HSV-2

Mocarski and Roizman (1982a) - strain F

Niza Frenkel (pers. comm.) - strain Justin

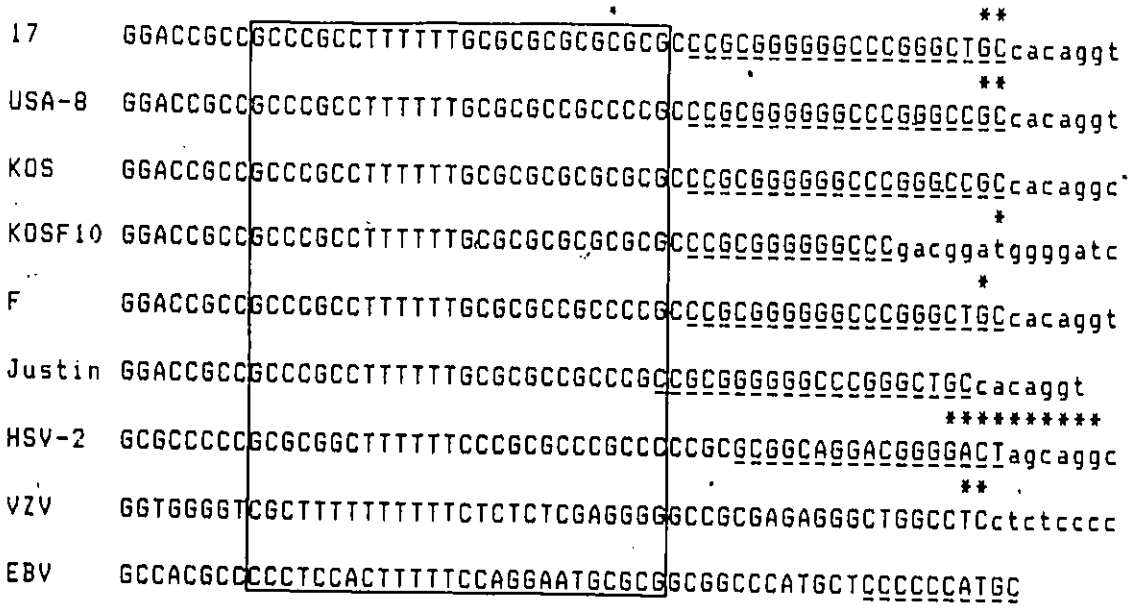
This work - KOS and KOSF10

Davison (1984) - VZV

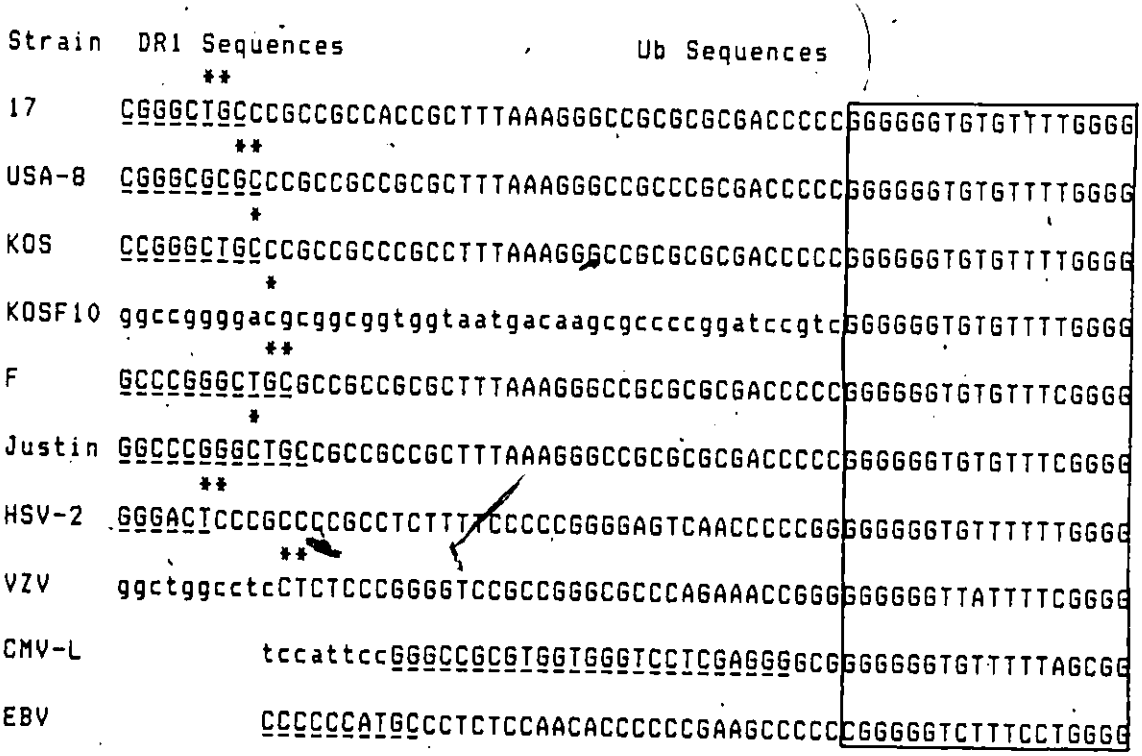
Tamashiro et al (1984) - CMV

Matsuo et al (1984) - EBV

L Termini



S Termini



while Varicella-Zoster Virus (VZV) is not. The terminal redundancies in EBV and CMV are flanked by direct repeats of 10 and 25 bp, respectively. Although there is no sequence data on viral termini available at present, it is tempting to speculate that cleavage probably occurs within these direct repeats. The sequence shown in Fig. 50 for CMV is that at the predicted L terminus. It has been aligned with sequences of S termini from other Herpesviruses because of its similarity to the latter. Tandem copies of the terminal redundancy in CMV accumulate at the S terminus instead of the L terminus. Perhaps the inverted orientation of the signal noted in Fig. 50 plays a role in the directed, asymmetric cleavage reactions. No obvious homologue to the other putative signal could be found in the sequence of CMV DNA. It is interesting to note here that the CMV "a" sequence can be recognized by the HSV cleavage machinery. Spaete and Mocarski (1985) were able to use a CMV terminus as a packaging signal in the production of engineered defectives in HSV-1. The terminal redundancy in EBV becomes tandemly iterated at both ends, suggesting that the polar aspect of HSV and CMV cleavage/packaging signals is not duplicated by the EBV signal. VZV, which is not terminally redundant, also appears to share homology with the HSV cleavage signals. However, both of the sequences noted in Fig. 50 are inverted with respect to flanking junction sequences such that they both "point" inwards at a single cleavage site, rather than outwards at two different cleavage sites.

In spite of what we now know about cleavage being directed to occur at a distance from the (putative) signals, variation among strains in the distance of the termini from the signal correlates well

with the distance of DR1 from the signal. Manipulation of these components in various constructs may help to clarify the role played by DR1.

III) Relationship Between Cleavage Mechanisms and Segment Inversion

The site specific cleavage mechanism described above impinges directly on the mechanism underlying segment inversion by removing any restrictions on the junction used for cleavage. As described above, previous models had proposed that only junctions with one "a" (Roizman, 1979), or two or more "a"s (Mocarski and Roizman, 1982a) could serve as cleavage substrates. My work clearly shows that any junction may be cleaved. This postulate allows biphasic cleavage/packaging, which produces two of the four isomers from one type of concatemer. Thus, only one other type of concatemer need be replicated to produce the other two isomers - one in which either the S segment or the L segment is inverted. Inversion could be accomplished by either intramolecular recombination, or intermolecular recombination or ligation between two genomes of differing isomerization. As noted above, there is evidence that sequences from both "a" and "c" are capable of driving homologous recombination. Since the fragments tested in my work amount to only a fraction of the size of the inverted repeats, and since one of them, SmaI A from "c", produced substantial levels of recombinant fragments within five rounds of viral replication, the probability of equimolar equilibration of recombination mediated by the bulk of the inverted repeats is quite high. Even if some of the sequences within the inverted repeats are not as recombinogenic as SmaI A or SmaI F, these

two sequences together may possess sufficient activity to produce the requisite equimolar levels of recombinant genomes, although the additivity of the recombinogenic activities should be tested directly.

Intermolecular ligation between genomes of differing isomerization was also observed in my work. While the frequency of these molecules in the population is too low to account for all of the segment inversion observed in HSV, intermolecular ligation must also contribute to this phenomenon. The fragments contained in pRIaa3 and pRIaa35 represent molecules which have undergone two rearrangements: intermolecular ligation and intramolecular recombination. Assuming for the moment that the frequencies observed in the population of clones reflect the rate at which such fragments are produced (remember that the data in fact reflect a static proportion of such molecules in a population which has undergone five rounds of viral replication), then the rate at which the molecules arise, 1/13, is a product of the rate at which intermolecular ligation occurs, 1/x, and the rate at which intramolecular recombination through SmaI F occurs, 1/4. (The same assumptions with respect to the rate of intramolecular recombination must also be made.) This produces a provisional rate of intermolecular ligation of 1/3. This kind of analysis of KOSF10 is hampered somewhat by the unusual nature of the virus isolate. If the intermolecular ligation occurred at the initial round of replication which produced the KOSF10 stepladder pattern, then the rate of 1/3 would reflect a real rate per round of replication. If it occurred subsequent to the early round which produced most of the other TK* fragments, then the rate of 1/3 must be spread out over at least five more rounds. By itself, or

even coupled with biphasic cleavage/packaging, this figure is insufficient to account for the equimolar mix of four isomers (Smiley, pers. comm.). However, it must contribute to the accumulation.

Skare and Summers (1977) estimated that intramolecular or intermolecular recombination events would have to occur in 35% of genomes at each round of replication in order to ensure equimolar ratios of the four isomers after amplification of the virus stock from one to 10^7 pfu. However, their calculations assumed that each isomer resulted from individual recombination events, yielding a total of four different events. It is now clear that only one event is actually required. A rough estimate indicates that intramolecular recombination proceeding at a rate of 20% per round is sufficient to produce equimolar isomers, when the effect of biphasic cleavage/packaging is accounted for. Thus, even though the three mechanisms alone, intramolecular recombination, intermolecular ligation , and biphasic cleavage/packaging may be insufficient to generate equal numbers of all four isomers starting from one pfu, together these mechanisms would be quite effective.

Other Herpesviruses also undergo segment inversion. Perhaps the most interesting is VZV, which , like HSV, possesses two segments, L and S. Only the S segment is flanked by long inverted repeats. The L segment is flanked by a very short inverted repeat. The S segment inverts freely, but the L segment inverts only 5% of the time (Davison, 1984). VZV is not terminally redundant, which may be a reflection of the apparent inverted orientation of the packaging signals (see above). However, the sequences displayed in Fig. 50 are contained in the inverted repeats flanking each segment. If they are indeed the packaging

signals, they should be active at both the internal junction and the novel junction formed by ligation of the L and S termini, and should generate I_L containing genomes at a frequency equal to that of P genomes through biphasic packaging. Since this clearly does not occur, it suggests that some feature of the novel junction distinguishes it from the internal junction as the preferred cleavage site. This feature must reside outside of the 89 bp inverted repeat flanking L. Davison (1984) proposed that the 5% segment inversion observed for L probably reflects the activity of a cryptic cleavage signal at the internal junction rather than reduced recombination through the small inverted repeats. Inversion of S must therefore represent intramolecular recombination, since it too would be deprived of the opportunity for biphasic cleavage. Whether or not S inversion in VZV is site specific or generalized remains to be determined. However, experience with HSV suggests that the latter is the more likely explanation.

One final issue regarding HSV DNA replication deserves some attention. The mechanism by which the viral concatemer is formed is still unknown. However, the data presented in this thesis suggests that the original hypothesis formulated by Roizman (1979), a rolling circle, may be correct. There is substantial evidence that the termini ligate together just after infection, and before DNA replication commences (Jacob et al, 1977; Ben-Porat and Veach, 1980; Poffenberger and Roizman, 1985). Ladin et al (1980) showed that cleavage and packaging are coupled events in Pseudorabies virus. If we assume that HSV is similarly constrained, the only opportunity for end to end ligation is shortly after infection. As pointed out above, intermolecular ligation

occurs at a fairly high frequency. What prevents the ligation of S termini with each other or of L termini with each other? One possibility is that the termini are "marked", perhaps with specific proteins, in such a way that only L-S ligation can occur.

Since end to end ligation of termini occurs before DNA synthesis commences, the mechanism of concatemer formation must involve either intermolecular interactions, such as rec A and/or int mediated recombination in bacteriophage lambda, or a rolling circle. The former route now appears to be unlikely. If site specific recombination through "a" were real, this would provide a perfectly logical explanation both of concatemer formation and of the necessity for a site specific recombination mechanism. However, it is not real. Generalized recombination may be the mechanism of concatemer formation; however, rough estimates of the rate of intermolecular recombination based on genetic experiments suggests that it is too low to account for concatemer formation (Honess et al, 1980). Supporting evidence for a rolling circle mechanism was provided by Vlazny and Frenkel (1982). These authors transfected cells with two distinguishable defective monomers and found that the resulting defective genomes were clonal in origin, thus eliminating intermolecular recombination as a mechanism. Interpretation of these results depends on the assumption that transfected DNA tends to enter cells en masse (Wigler et al, 1979). Thus, the rolling circle mechanism seems to be the best candidate. Determining the molecular mechanisms involved may prove to be interesting in the future.

APPENDIX

The appendix contains the maps and details of construction of a number of plasmids. Not all of the plasmids in the appendix were used in the studies described in this thesis. However, a comprehensive catalogue of all the plasmids made during my tenure as a graduate student was deemed advisable. Standard reaction conditions and procedures were not described here in the interests of brevity. Details of this nature may be found in the Materials and Methods section, or in Maniatis et al, 1982.

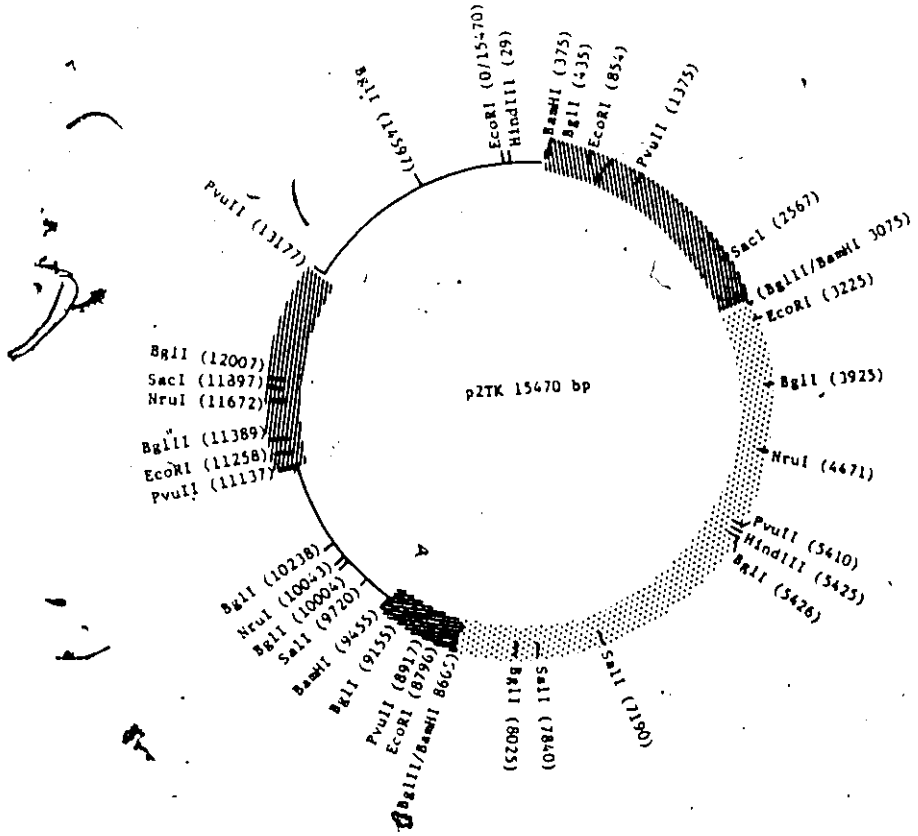
p2TK Amp^R Tet^S

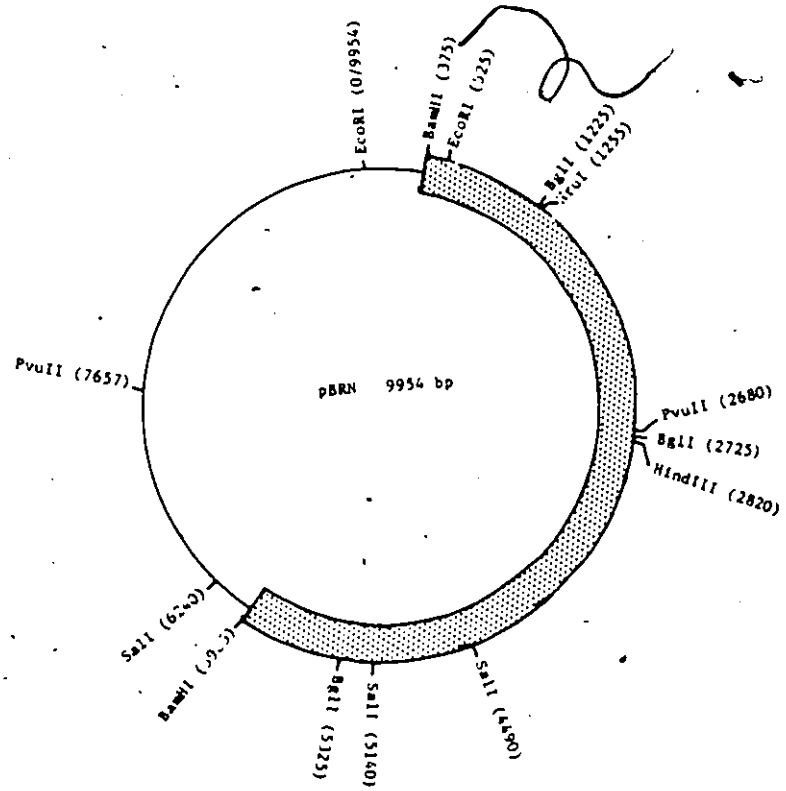
p2TK was constructed by cloning the BamH1 recombinant HSV fragment from pRB316αTK into the BamH1 site on pTK173. pRB316αTK contains the BamH1 N fragment from HSV-1 (F) cloned into the BglII site on a 3.5 kb TK containing BamH1 fragment from HSV-1 (F) such that the ICP4 promoter in BamH1 N drives expression of TK (Post et al, 1981). Individual plasmids were screened by comparing BamH1 digests with pRB316αTK and pTK173. Those with the correct profile were digested with HindIII and BglII to determine the orientation of the 316αTK insert. A plasmid in which the 316αTK insert was in opposite orientation to the PvuII TK insert in pTK173 was identified. The restriction map of p2TK was prepared by compiling information from several sources : a) published sequences of pBR322 (Sutcliffe, 1979) and HSV-1 TK (Wagner and Summers, 1981), b) published restriction maps of HSV DNA (Smiley et al, 1981), c.) restriction analysis of p2TK, pTK173, and pBRN (this work).

pBRN Amp^R Tet^S

pBRN contains the BamH1 N fragment of HSV-1 (KOS) cloned into the BamH1 site on pBR322. It was constructed by subcloning the BamH1 N fragment from p1BJH2, a BglII joint spanning fragment from KOS cloned into pKC7 (W.-C.Leung, pers. comm.). Restriction mapping of sites on the plasmid was performed. The NruI site at 1255 bp on this clone does not correspond with an NruI site on the homologous fragment in p2TK, which is derived from strain F.

pBR322 ——— TK  HSV Joint sequences 


p2TK 15470 bp
EcoRI (0/15470)
HindIII (29)
BamHI (375)
BglI (435)
EcoRI (854)
PvuII (1375)
SacI (2567)
(BglII/BamHI 3075)
EcoRI (3225)
BglI (3925)
NruI (4471)
PvuII (5410)
HindIII (5425)
BglI (5426)
SalI (7190)
SalI (7840)
BglI (8025)
PvuII (8796)
EcoRI (8796)
(BglII/BamHI 8665)
BglI (9155)
BamHI (9455)
SalI (9720)
BglI (10004)
NruI (10043)
BglI (10238)
BglI (11389)
EcoRI (11258)
PvuII (11137)
BglI (12007)
SacI (11897)
NruI (11672)
PvuII (13177)
BglI (14597)


pBRN 9954 bp
EcoRI (0/9954)
BamHI (375)
EcoRI (525)
BglI (1225)
NruI (1255)
PvuII (2680)
BglI (2725)
HindIII (2820)
SalI (4490)
SalI (5140)
BglI (5125)
BamHI (6240)
SalI (6240)
PvuII (7657)

pTK173 Amp^R Tet^S

pTK173 was constructed by Randy McKinnon (PhD thesis 1984, McMaster University). It contains a TK-containing PvuII fragment from HSV-1 (KOS) cloned into the PvuII site on pBR322.

pTK371 Amp^R Tet^R

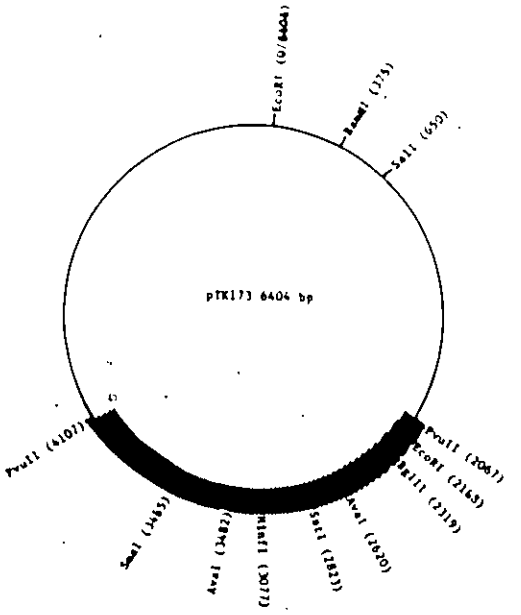
pTK371 was derived from pTK173 by digestion with PvuII followed by religation. Colonies were screened by hybridization with the TK containing PvuII fragment from pTK173 which was labelled *in vitro* with ^{32}P by nick translation. Birnboim preparations of plasmid DNA (Birnboim and Doly, 1979) from positive colonies were analysed by digestion with SalI and EcoRI.

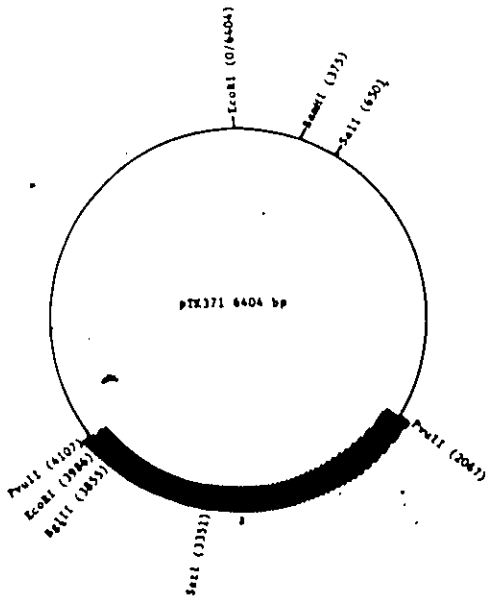
pTK173-B Amp^R Tet^S

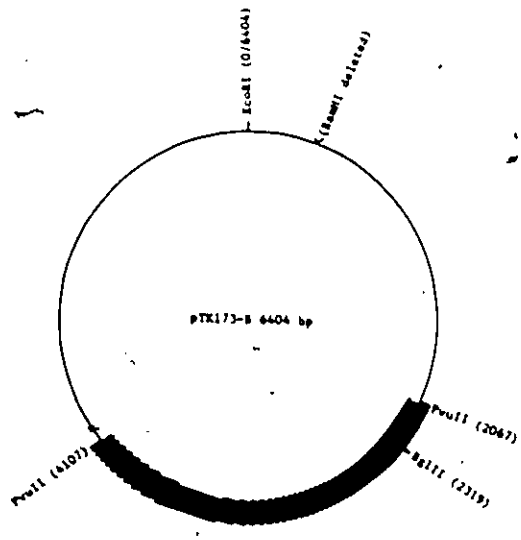
pTK173-B was derived from pTK173 by digestion with BamHI and S_1 nuclease followed by religation. Ampicillin resistant, tetracycline sensitive clones were screened for the absence of the BamHI site. The extent of the deletion in pTK173-B is unknown. However, none of the restriction fragments from this region appear to be grossly changed in size, so it is likely that the deletion is very small, ie only a few base pairs.

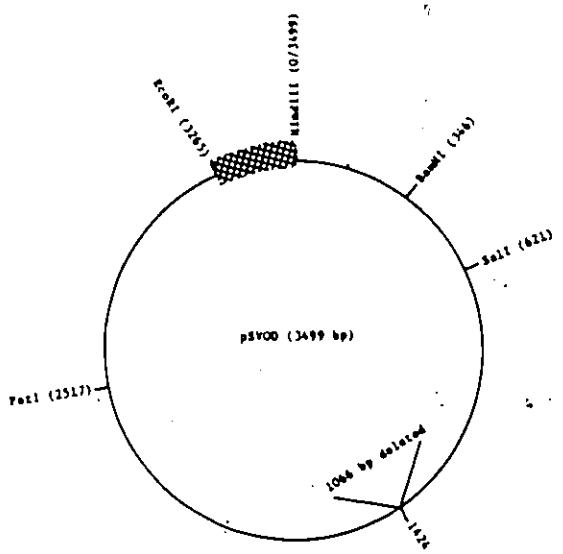
pSVOD Amp^R Tet^S

pSVOD was constructed by Mellon, et al (1981). The hatched area denotes the SV40 origin fragment. The position of the deletion that removes the "poison" sequences is indicated.


pTK173 6404 bp
EcoRI (0/6404)
BamHI (375)
SalI (650)
PvuII (2067)
EcoRI (2165)
BglII (2319)
AvaI (2620)
SstI (2822)
HindIII (3071)
AvaI (3482)
SmaI (3485)
PvuII (4107)


pTK371 6404 bp
EcoRI (0/6404)
BamHI (375)
SalI (650)
PvuII (2067)
SstI (3351)
BglII (3855)
EcoRI (3986)
PvuII (4107)


pTK173-B 6404 bp
EcoRI (0/6404)
BamHI deleted
PvuII (2067)
BglII (2319)
PvuII (4107)

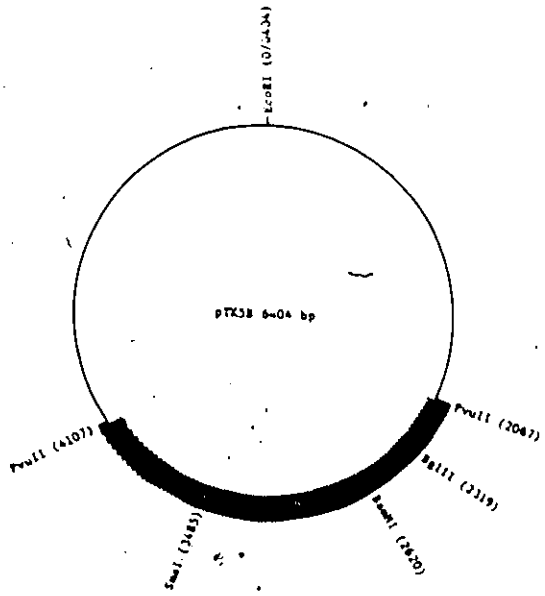

pSV00 (3499 bp)
HindIII (0/3499)
EcoRI (3265)
BamHI (346)
SalI (621)
PstI (2517)
1066 bp deleted
1424

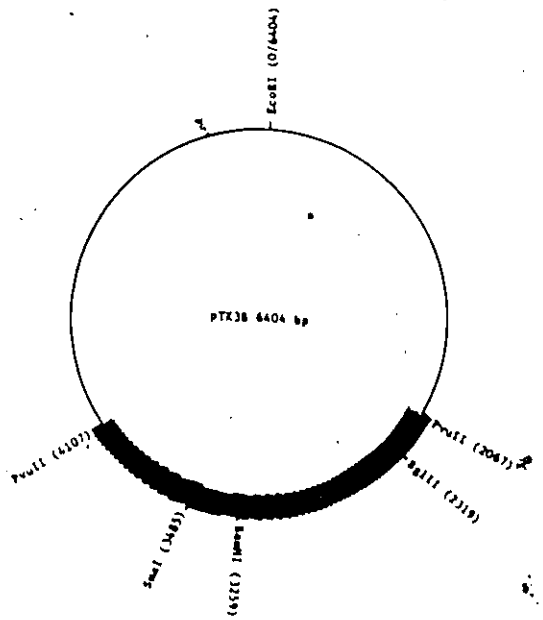
TK3B and pTK5B Amp^R Tet^S

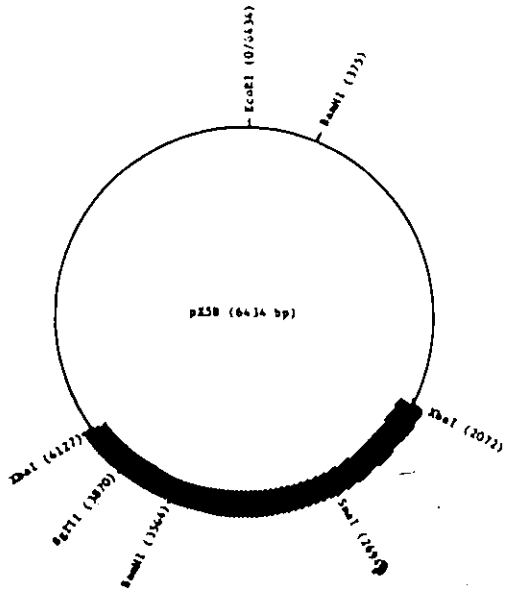
Both of these plasmids were derived from pTK173-B. This plasmid was subjected to partial AvaI digestion and then treated with T4 polymerase. BamHI linkers were ligated to the DNA, which was then digested with BamHI and electrophoresed through a 1.4% agarose gel with a linear pTK173-B marker. Fragments comigrating with the linear marker were eluted and ligated. Birnboim (Birnboim and Doly, 1979) preparations of individual clones were screened for the presence of BamHI sites at the desired locations. Two such clones, pTK5B and pTK3B, had acquired BamHI sites at 2620 bp and 3259 bp respectively. The TK^- phenotype of both genes was confirmed by marker rescue with HSV-1 (KOS). The resulting TK^- viruses are KOSB5 and KOSB3 respectively.

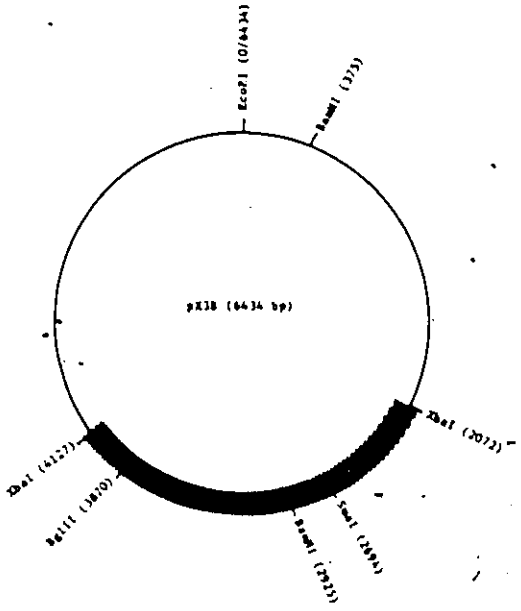
pX5B and pX3B Amp^R Tet^S

pX5B and pX3B were derived from pTK5B and pTK3B respectively. In both cases, the BamHI linker mutation was introduced into pTK123 by replacing the central BglII/SmaI TK fragment from pTK123 with the corresponding fragment from pTK5B or pTK3B. pTK123, which was constructed by Yousef Hajahmad (personal communication), is a derivative of pTK173 in which XbaI linkers were inserted at the PvuII sites. During the cloning procedure the TK gene became inverted such that the orientation in pBR322 is the same as that in pTK371.


EcoRI (0/6434)
pTX38 6404 bp
PvuII (2067)
BglII (3319)
BamHI (3620)
SmaI (3485)
PvuII (4107)


EcoRI (0/6404)
pTX38 6404 bp
PvuII (2067)
BglII (3319)
BamHI (3239)
SmaI (3485)
PvuII (4107)


EcoRI (0/6434)
BamHI (325)
pX38 (6434 bp)
XbaI (2072)
SmaI (2695)
BamHI (3564)
BglII (3870)
XbaI (4127)


EcoRI (0/6434)
BamHI (325)
pX38 (6434 bp)
XbaI (2072)
SmaI (2695)
BamHI (2915)
BglII (3810)
XbaI (4127)

pTKSB Amp^R Tet^S

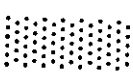
pTKSB was derived from pTK173-B by digestion with SacI followed by brief (30 sec. to 5 min.) treatment with Bal31 exonuclease. After ligation of BamHI linkers, the DNA was digested with BamHI and religated. Clones were screened for the presence of a BamHI site and the absence of a SacI site. A plasmid which had suffered a relatively small deletion was identified. Subsequent sequence analysis indicated that exactly 200 bp of TK DNA, from position 2748 to 2948 on the pTK173 map, had been deleted and replaced with the BamHI linker sequence.

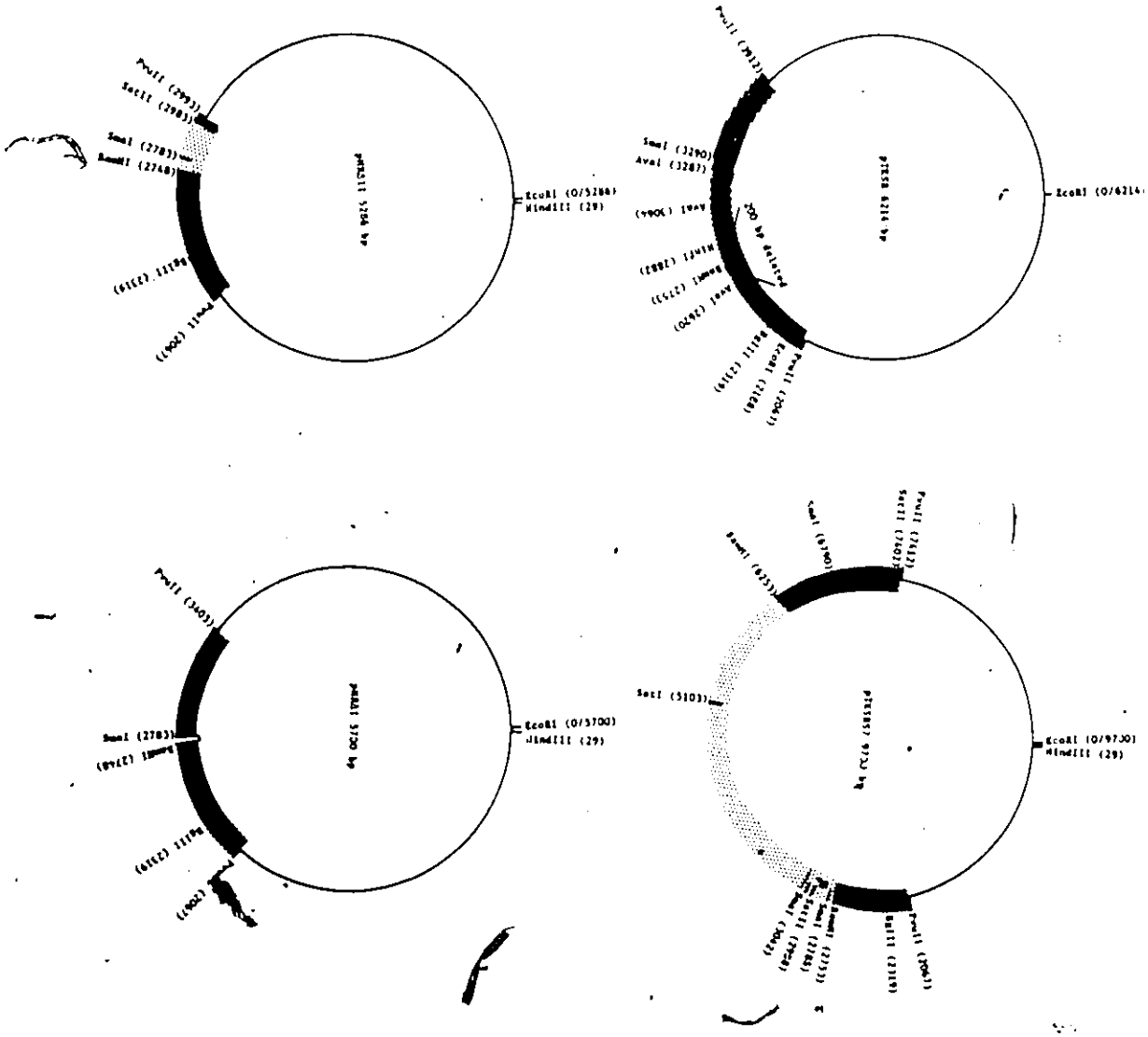
pTKSBS7 Amp^R Tet^S

This plasmid was made by subcloning the BamHI S terminal fragment from pTK173S7 into the BamHI site on pTKSB. This plasmid was the source of a number of additional constructs. The relevant restriction sites have been put on the map.

pHRΔI and pHRΔII Amp^R Tet^S

These clones, constructed by Helen Rudzroga, were derived from pTKSBS7 by digestion with SmaI (pHRΔ I) or SstII (pHRΔII) followed by religation. Birnboim extracts were screened for the appropriate deletions.

pBR322 ——— TK  Joint  or 

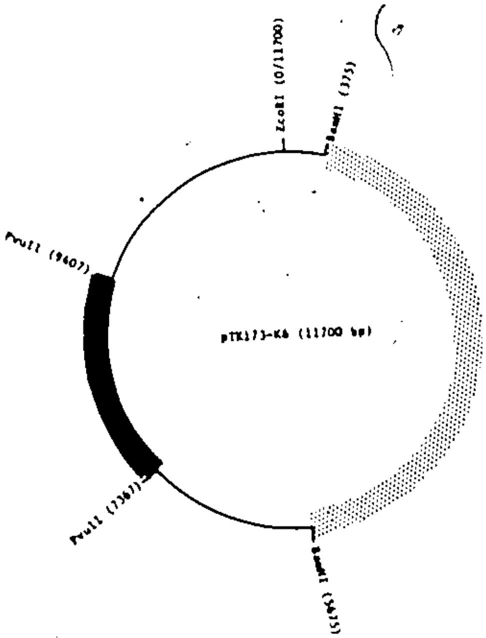

pUC11 5626 bp
EcoRI (0/5288)
HindIII (29)
PvuII (299)
SacII (298)
SmaI (2783)
BamHI (2748)
BglII (2319)
PvuII (1902)
pBR 6214 bp
EcoRI (0/6214)
PvuII (3412)
SmaI (3290)
AvaI (3287)
AvaI (3006)
deleted 66 bp
MluNI (2882)
BamHI (2753)
AvaI (2670)
BglII (2319)
EcoRI (2184)
PvuII (2041)
pBAL 5700 bp
EcoRI (0/5700)
HindIII (29)
PvuII (3403)
SmaI (2783)
BamHI (2748)
BglII (2319)
pBR 9716 (5815)
EcoRI (0/9730)
HindIII (29)
SmaI (10)
BamHI (625)
SacII (5103)
PvuII (2702)
SacII (2702)
BglII (2319)
PvuII (2041)

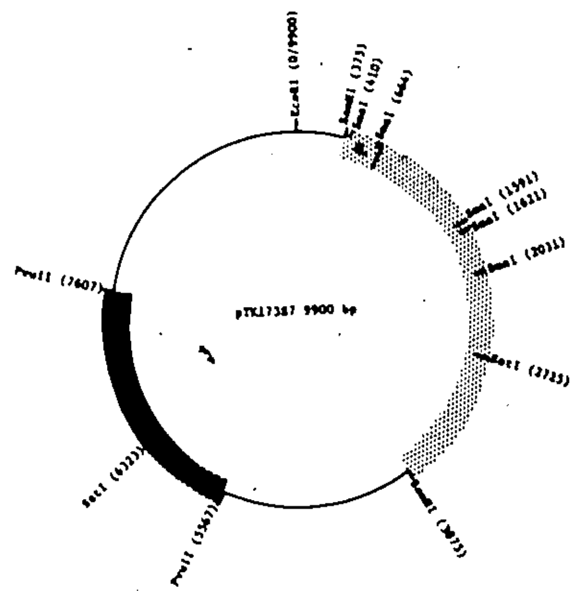
pTK173K6 and pTK173S7 Amp^R Tet^S

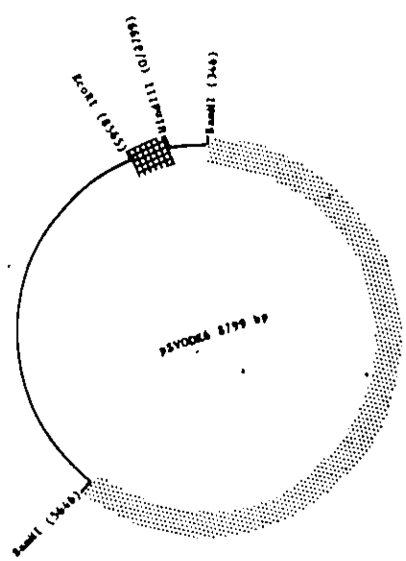
pTK173K6 and pTK173S7 contain the BamHI K and P fragments respectively from HSV-1 (KOS) cloned into the BamHI site on pTK173. The orientation of the inserts has not been determined. HSV-1 (KOSc1i) -DNA was ligated with BamHI linkers, digested with BamHI, and electrophoresed through a 1% agarose gel. The joint spanning K fragment and the S terminal P fragment were eluted and ligated to BamHI digested pTK173. Colonies were screened for the presence of joint DNA by hybridization with a ^{32}P-labelled SacI joint fragment from pINJ-2 (Smiley et al, 1981). Positive clones were verified by restriction analysis and by Southern (1975) blot analysis. In the latter case, the cloned joint was found to comigrate with the bottom step of a BamHI joint stepladder, indicating that it contained one "a" sequence (Wagner and Summers, 1978).

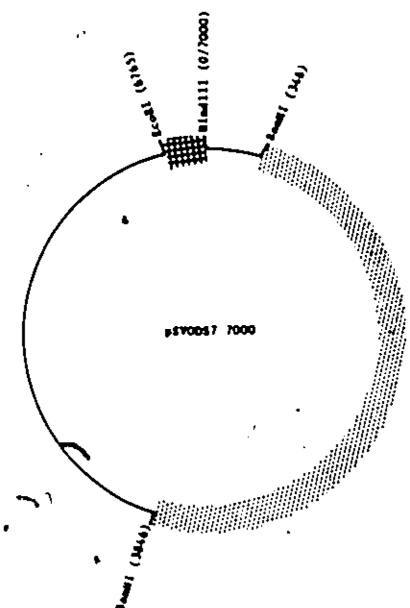
pSVODK6 and pSVODS7 Amp^R Tet^S

pSVODK6 and pSVODS7 were constructed by subcloning the BamHI joint (K6) and S terminal (S7) fragments from pTK173K6 and pTK173S7 respectively into the BamHI site on pSVOD. The orientation of the inserts was not determined.


EcoRI (0/11700)
PvuII (9607)
PvuII (7367)
pTK173-K6 (11700 bp)


EcoRI (0/9900)
PvuII (7607)
pTK17387 9900 bp




HindIII (0/7000)
pSVOD57 7000

pcaDI Amp^R Tet^S

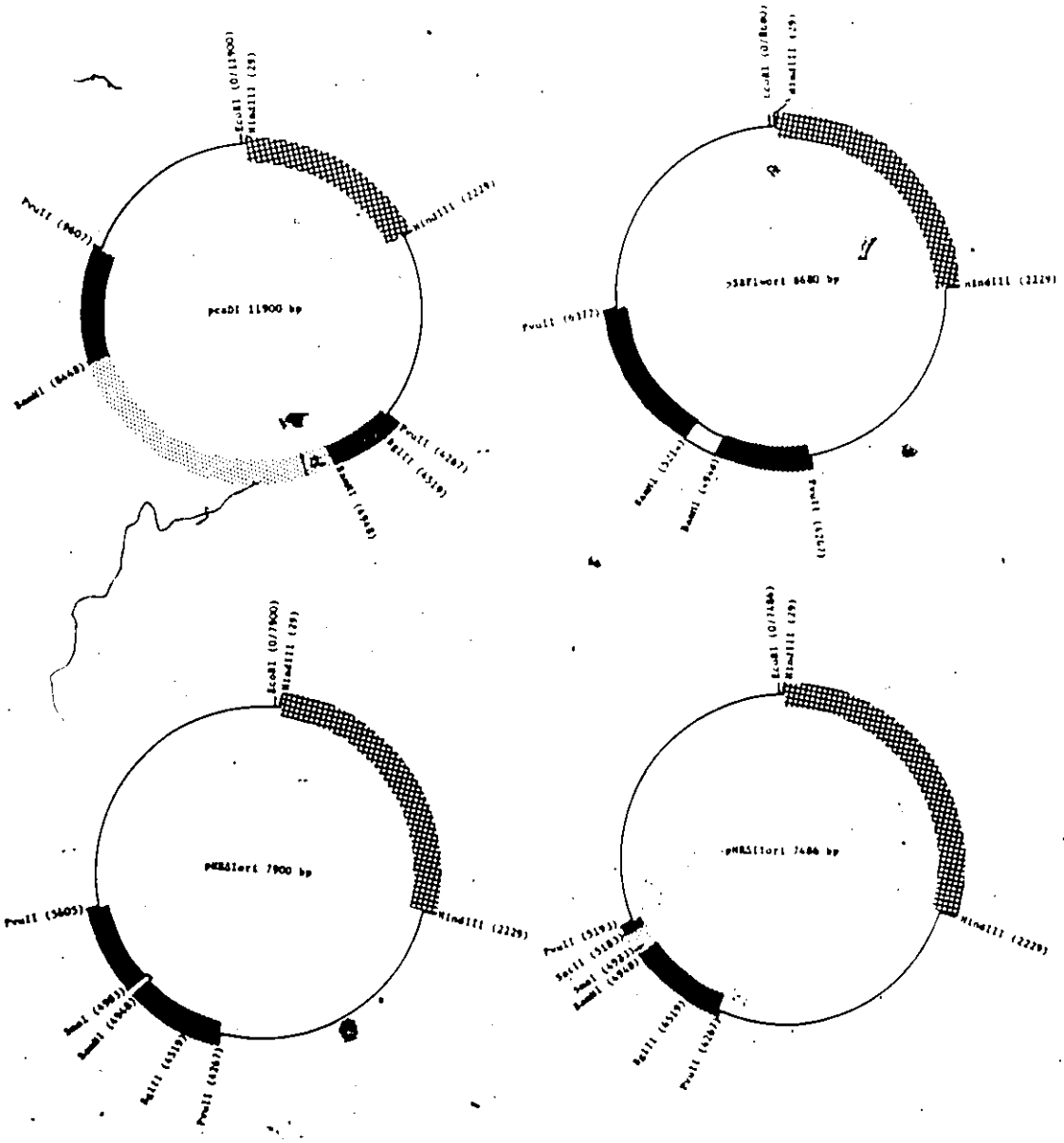
pcaDI (for defective interfering) was constructed by cloning an HSV origin of replication into the HindIII site on pTKSBS7. The origin fragment was derived from p2TK and represents the sequence from the EcoRI site at 3225 bp to the HindIII site at 5425 bp. p2TK was digested with EcoRI and treated with T4 polymerase. HindIII linkers were ligated to the DNA, which was then digested with HindIII. The 2200 bp origin containing fragment was eluted from a 1% agarose gel and ligated with HindIII digested pTKSBS7. Colonies were screened by hybridization with the ^{32}P-labelled EcoRI/HindIII origin fragment from p2TK. Positive clones were further identified by restriction analysis. The orientation of the HindIII origin fragment in pcaDI was not determined. This plasmid served as the DNA source for the HSV origin which was subcloned into several other fragments. It is sometimes referred to as pDL.

pHRΔIori pHRΔIIori and pSBF14ori Amp^R Tet^S

These plasmids were made by inserting the HindIII HSV origin from pcaDI into pHRΔI, pHRΔII and pSBF14 respectively at the HindIII site.

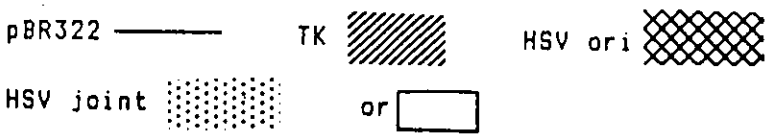
pBR322 ——— TK  HSV ori 

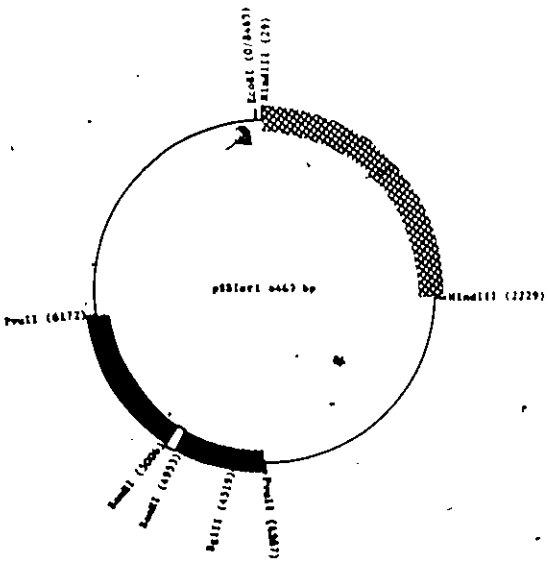
HSV joint  or 

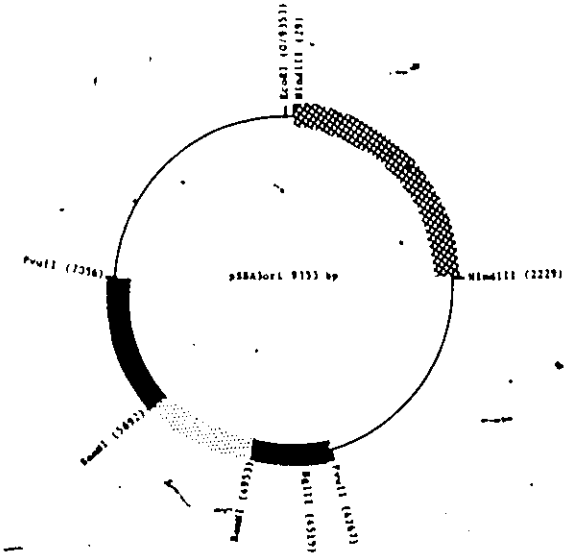

pcaDI 11900 bp
EcoRI (0/11900)
HindIII (25)
HindIII (2229)
PvuII (9607)
BamHI (8468)
PvuII (6287)
BglII (6519)
BamHI (6048)
pSBFlwort 8680 bp
EcoRI (0/8680)
HindIII (25)
HindIII (2229)
PvuII (6317)
BamHI (5214)
BamHI (4968)
PvuII (4525)
pMBAlort 7900 bp
EcoRI (0/7900)
HindIII (25)
HindIII (2229)
PvuII (5605)
BamHI (5083)
BamHI (4968)
BglII (4519)
PvuII (4287)
pMBAlort 7486 bp
EcoRI (0/7486)
HindIII (25)
HindIII (2229)
PvuII (5193)
SacII (5183)
BamHI (5023)
BamHI (4968)
BglII (4519)
PvuII (4287)

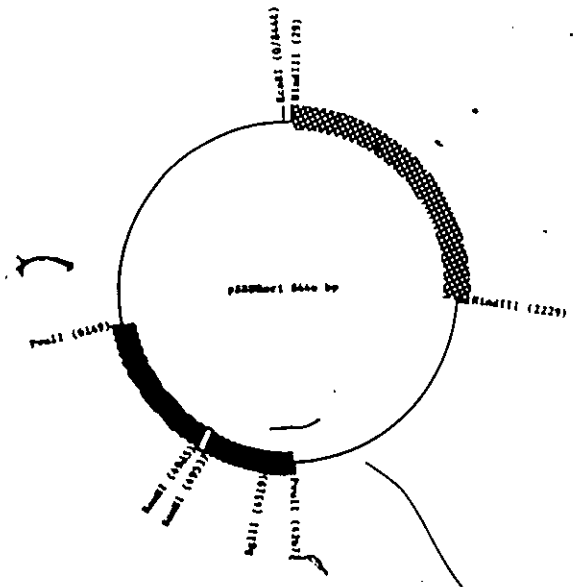
pSBIori pSBA3ori pSBDRori and pbaDI Amp^R Tet^S

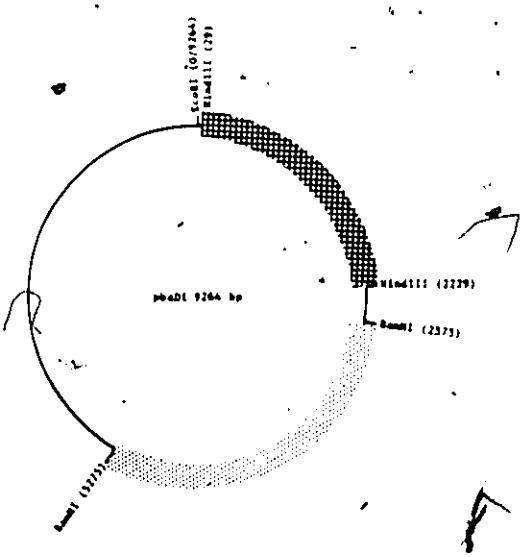
All, of these plasmids were constructed by subcloning the HindIII HSV origin from pcaDI into the HindIII site on pSBI, pSBA3, pSBDR and pRB143. The latter plasmid was kindly donated by Dr. B. Roizman and contains the BamHI L terminus cloned into the BamHI site on pBR322 (Mocarski and Roizman, 1982a).




EcoRI (0/6465)
HindIII (29)
pSBlorI 6465 bp
HindIII (2229)
PvuII (6172)
BamHI (5000)
BamHI (4993)
BglII (4315)
PvuII (3987)


EcoRI (0/9353)
HindIII (29)
pSBAlorL 9353 bp
HindIII (2229)
PvuII (7356)
BamHI (5982)
BamHI (4993)
BglII (4315)
PvuII (3987)


EcoRI (0/6446)
HindIII (29)
pSBBlorI 6446 bp
HindIII (2229)
PvuII (6149)
BamHI (5965)
BamHI (4993)
BglII (4315)
PvuII (3987)


EcoRI (0/9264)
HindIII (29)
pBADI 9264 bp
HindIII (2229)
BamHI (2373)
BamHI (1215)

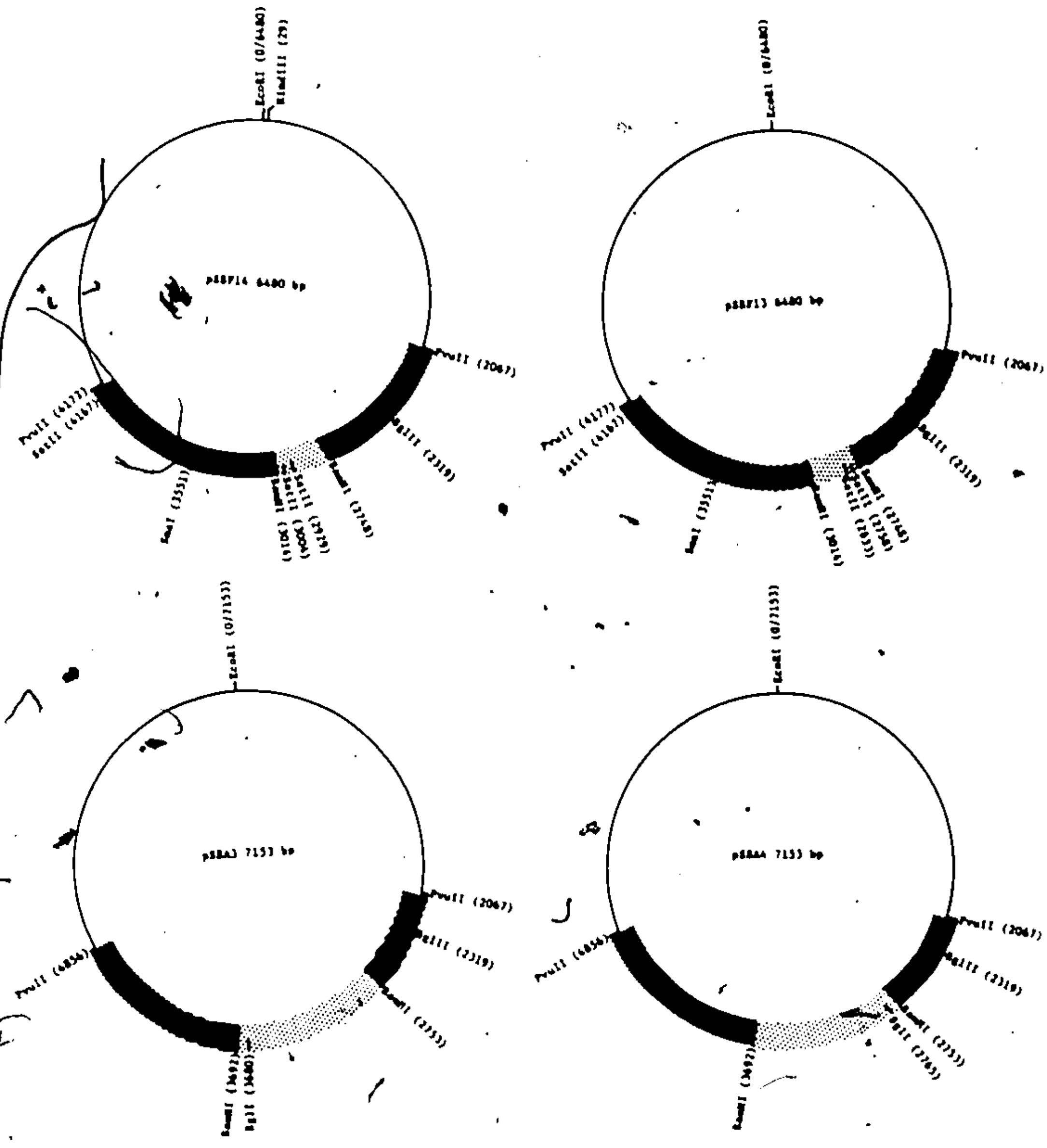
pSBF13 and pSBF14 Amp^R Tet^S

Both plasmids are derivatives of pTKSB. The BamHI insert fragment from pUC7-F was cloned into the BamHI site on pTKSB. pSBF13 and pSBF14 are different clones in which the F insert fragment went into pTKSB in opposite orientations. The orientation was determined by the asymmetricaly placed SstII sites within the insert.

pSBA3 and pSBA4 Amp^R Tet^S

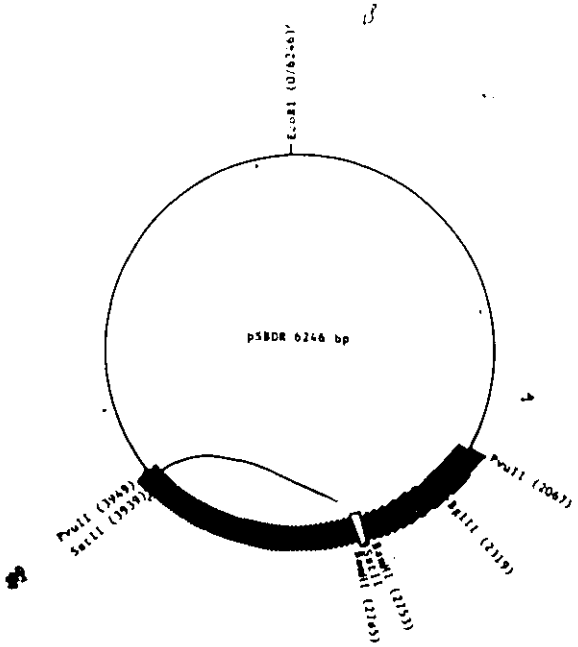
The BamHI insert fragment from pUC7-A was cloned into the BamHI site on pTKSB. The asymmetric BglI site on this sequence was used to orient the insert in the resulting plasmids and two in which the inserts were in opposite orientations were identified. These were pSBA3 and pSBA4.

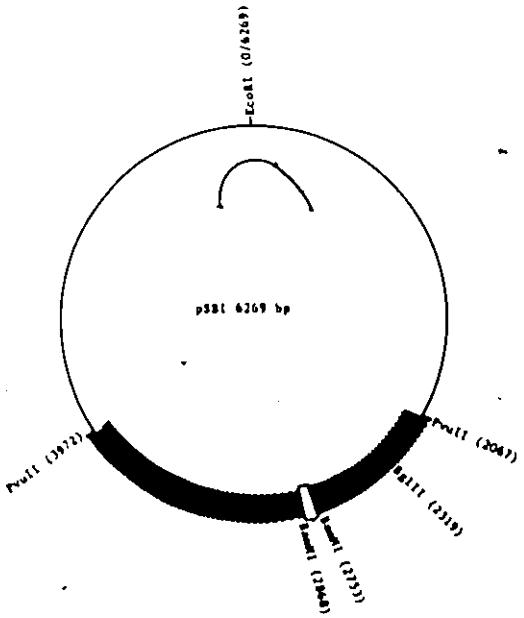
pBR322 ——— TK  HSV joint 

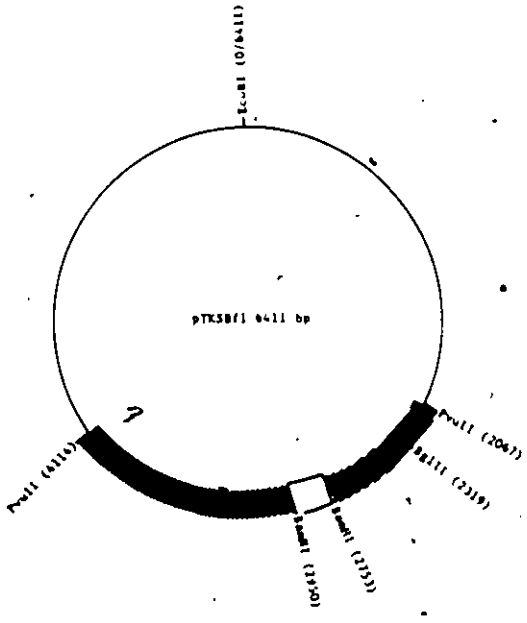

EcoRI (0/6480)
HindIII (29)
pSBF14 6480 bp
PvuII (2067)
PvuII (6177)
BglII (2319)
EcoRI (0/6480)
pSBF13 6480 bp
PvuII (2067)
PvuII (6177)
BglII (2319)
EcoRI (0/7153)
pSBA3 7153 bp
PvuII (2067)
BglII (2319)
PvuII (6856)
EcoRI (0/7153)
pSBA4 7153 bp
PvuII (2067)
BglII (2319)
PvuII (6856)

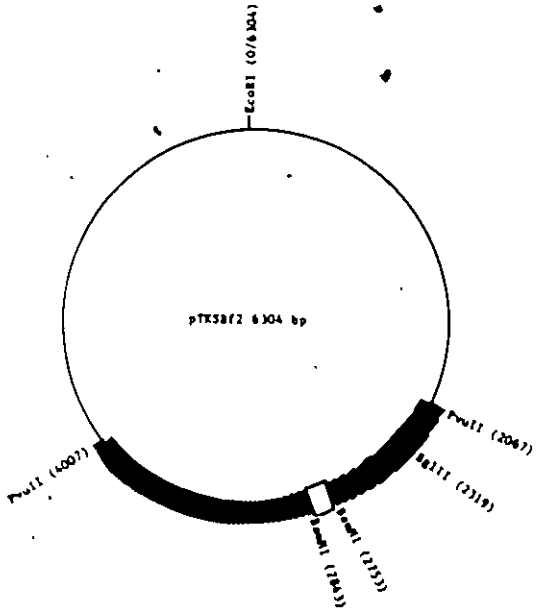
pSBDR pSBI pSBf1 and pSBf2 Amp^R Tet^S

All of these plasmids were made by subcloning the insert containing BamHI fragments from the respective pUC7 clones - pUC7-DR, pUC7-I, pUC7-f1 and pUC7-f2 - into the BamHI site on pTKSB. pSBDR was identified by the presence of the additional SstII site within the DR sequence. pSBI was identified by the presence of the BamHI I insert fragment which comigrated with the homologous fragment from pUC7-I. pSBf1 and pSBf2 were similarly identified. In addition, pSBf1 was also identified by the presence of an extra SstII site.


EcoRI (0/6246)
pSBDR 6246 bp
PvuII (3049)
SacII (3059)
PvuII (2067)
BglII (2319)
BamHI (2155)
SacII (2169)

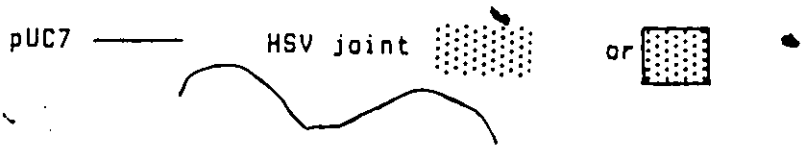

EcoRI (0/6269)
pSBI 6269 bp
PvuII (3972)
PvuII (2067)
BglII (2319)
BamHI (2155)
BamHI (2168)

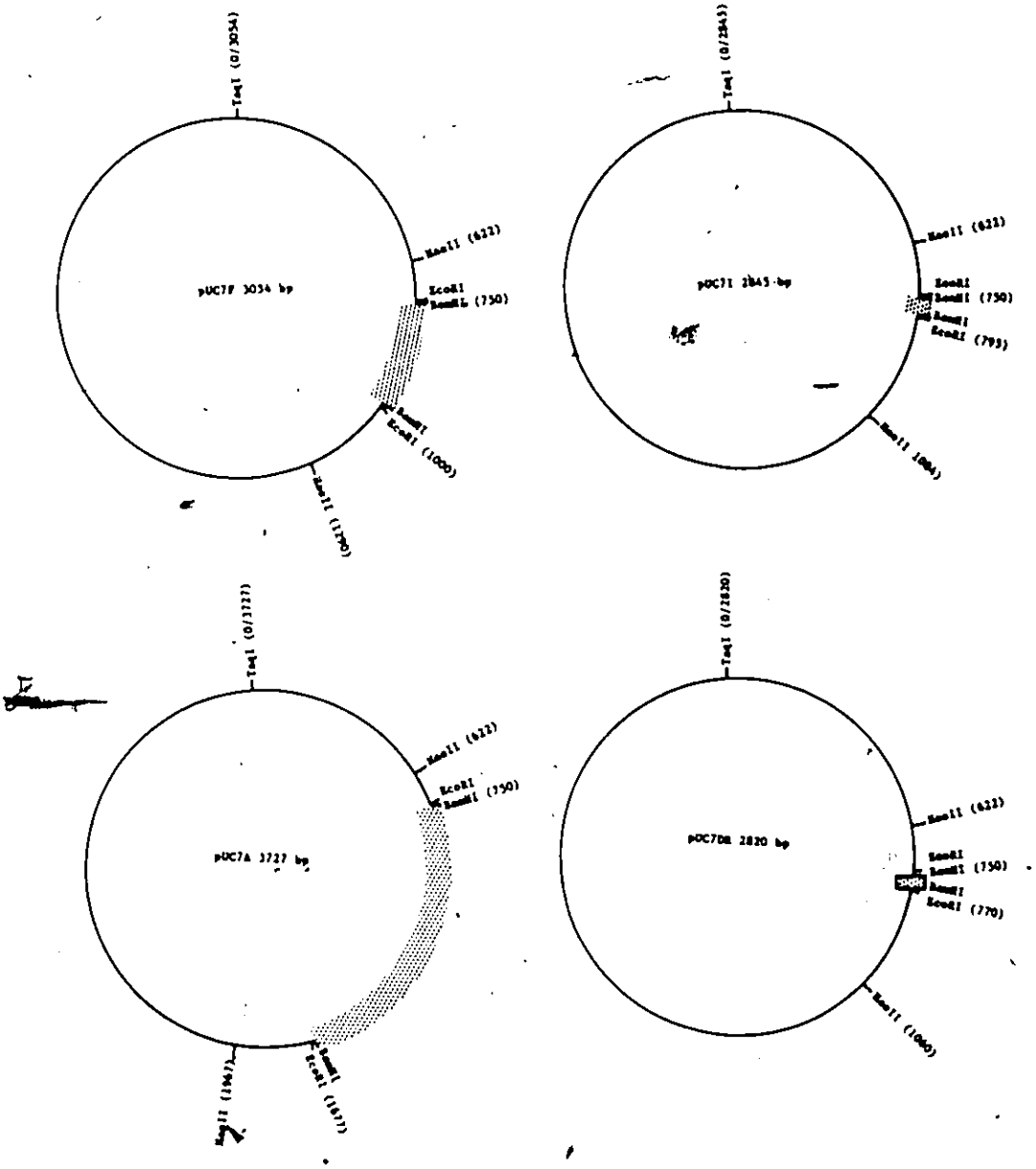

EcoRI (0/6411)
pTK5Bf1 6411 bp
PvuII (4110)
PvuII (2067)
BglII (2319)
BamHI (2155)
BamHI (2169)


EcoRI (0/6304)
pTK5Bf2 6304 bp
PvuII (4003)
PvuII (2067)
BglII (2319)
BamHI (2155)
BamHI (2163)

pUC7-F pUC7-I pUC7-A and pUC7-DR Amp^R lac^-

These plasmids were all derived by cloning flush ended DNA fragments into the HincII site within the polylinker on pUC-7 (Vieira and Messing, 1982). Transformation of JM109 cells with these plasmids results in white colonies when selection is performed on X-gal plates (see Materials and Methods). The DNA fragments used were SmaI A, SmaI F and SmaI I from pSVODK6, according to the nomenclature of Wagner and Summers (1978), and a synthesized sequence, DR1 (Mocarski and Roizman, 1982a). All of these fragments map to the HSV joint region, with SmaI F and I, and DR1 falling within the "a" sequence. SmaI A is the fragment in "c" that directly abuts "a". DR1 was synthesized by Jim Smiley on a SAM synthesizer (Varmuza and Smiley, 1985). The cloning was done by Helen Rudzroga.




TaqI (0/3054)
pUC7F 3054 bp
BamHI (622)
EcoRI
BamHI (750)
BamHI
EcoRI (1000)
BamHI (1700)
TaqI (0/2845)
pUC7I 2845 bp
BamHI (622)
EcoRI
BamHI (750)
BamHI
EcoRI (793)
BamHI (1804)
TaqI (0/3727)
pUC7A 3727 bp
BamHI (622)
EcoRI
BamHI (750)
BamHI (1943)
EcoRI (1677)
TaqI (0/2820)
pUC7DB 2820 bp
BamHI (622)
EcoRI
BamHI (750)
SphI
BamHI
EcoRI (770)
BamHI (1060)

pUC7-f1 and pUC7-f2 Amp^R lac^-

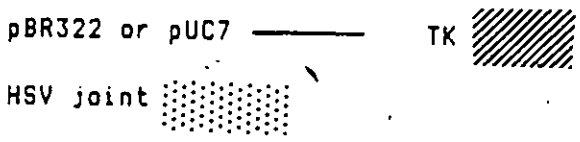
These plasmids were made by digesting pUC7-F with SstII. pUC7-f1 was constructed by religating the DNA after SstII digestion and screening the white colonies for the appropriate deletion. pUC7-f2 was made by treating the SstII digest of pUC7-F with T4 polymerase before electrophoresis through an 8% acrylamide gel. The small fragment was eluted and ligated to HincII digested pUC7. These clones were subsequently used to verify the sequence of SmaI F. They were made by Marg Howes.

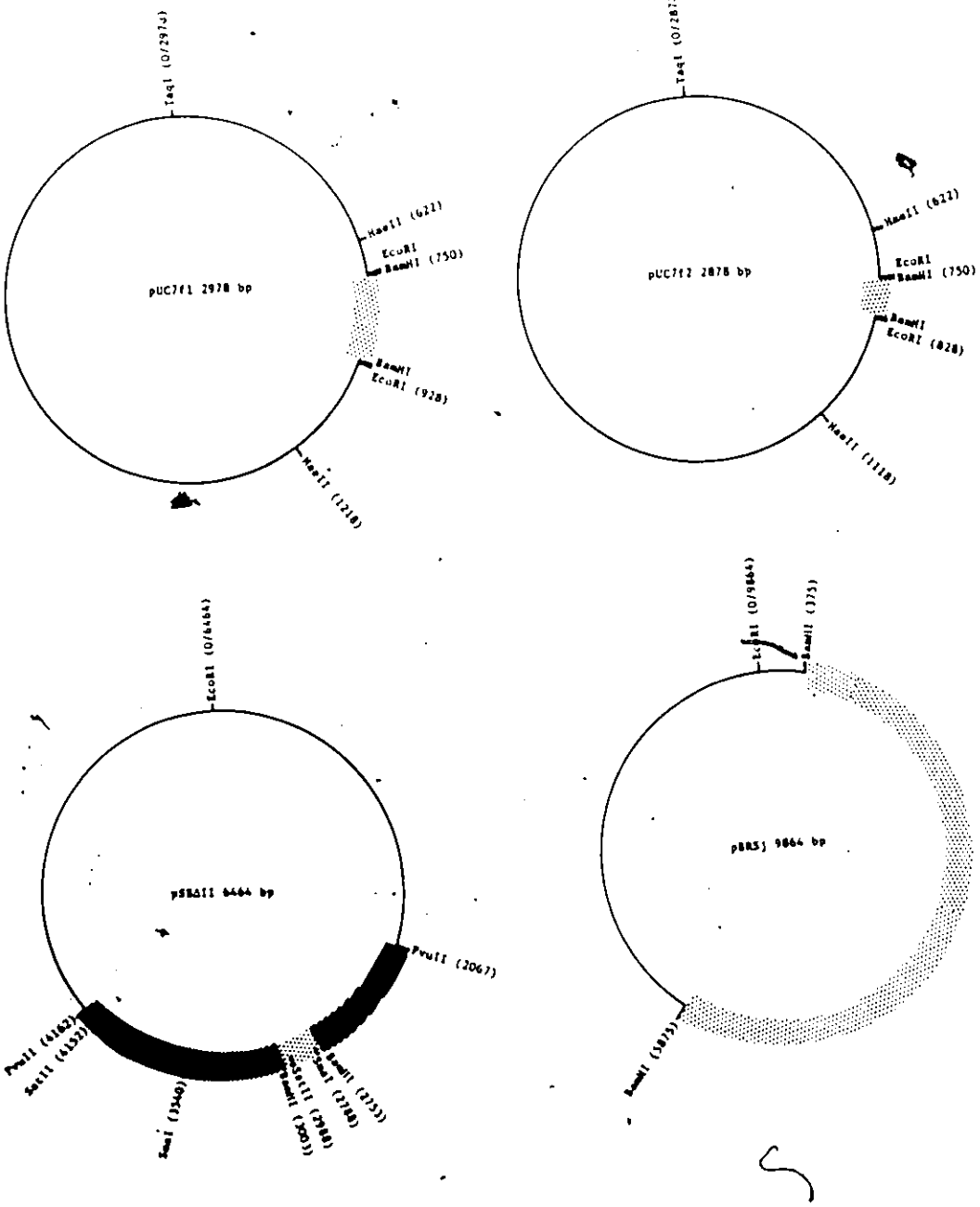
pSB$_\Delta$II Amp^R Tet^S

pSB$_\Delta$II was derived from pHR$_\Delta$ II, which was digested with PvuII and ligated to BamHI linkers. Following digestion with BamHI, the 230 bp fragment corresponding to the S terminal insert in pHR$_\Delta$II plus 10 bp of 3'TK DNA was eluted from an 8% acrylamide gel and ligated to BamHI digested pTKSB. Plasmid DNA was screened for the presence of additional SstII and SmaI sites. Digestion with these enzymes also allowed the orientation of the insert to be determined.

pBRSj Amp^R Tet^S

pBRSj contains the BamHI joint spanning fragment from HSV-1 (KOScls), a different isolate of KOS, inserted at the BamHI site on pBR322. The cloning procedure is essentially the same as that used to make pTK173K6.



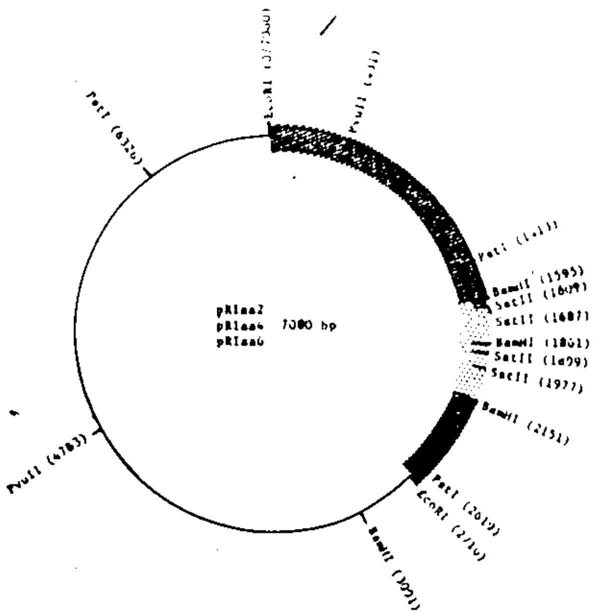

TaqI (0/2974)
HaeII (622)
EcoRI
BamHI (750)
pUC7f1 2978 bp
BamHI
EcoRI (928)
HaeII (1218)
TaqI (0/2873)
HaeII (622)
EcoRI
BamHI (750)
pUC7f2 2878 bp
BamHI
EcoRI (828)
HaeII (1118)
EcoRI (0/6464)
pSBAII 6464 bp
PvuII (2067)
PvuII (6162)
SacII (6152)
SmaI (3340)
BamHI (2755)
SstII (2748)
BamHI (2503)
EcoRI (0/9864)
BamHI (375)
pBR3j 9864 bp
BamHI (5815)

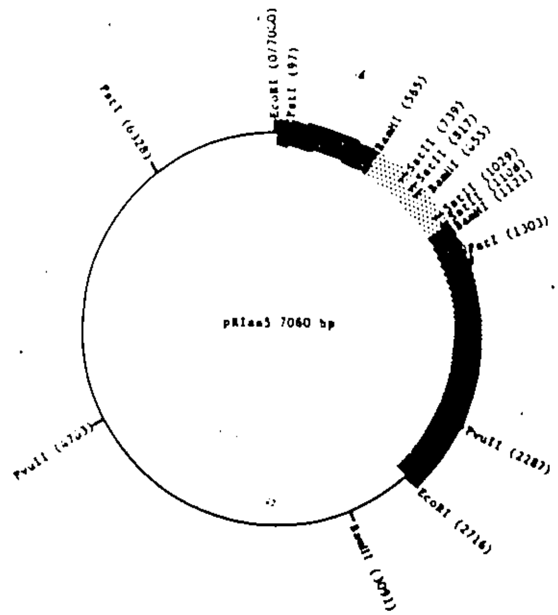
pRIaa2 pRIaa3 pRIaa4 pRIaa5 pRIaa6 Amp^R Tet^S

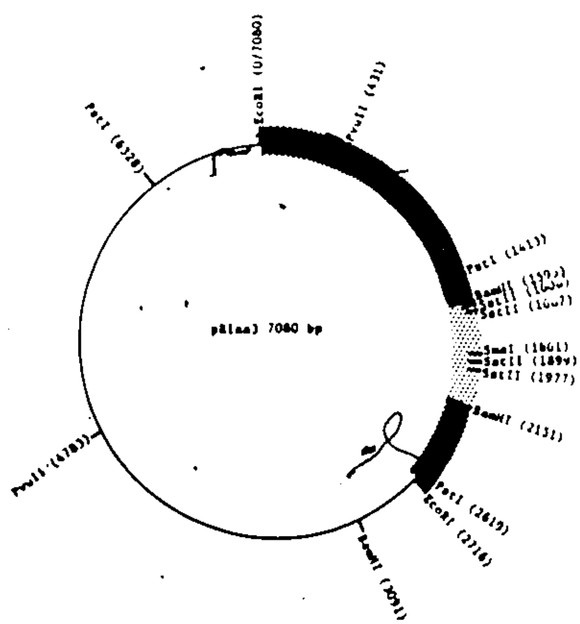
These plasmids were all constructed by cloning the second "step" of an EcoRI TK stepladder from KOSF10 viral DNA into the EcoRI site on pBR322. Colonies were screened by hybridization with ^{32}p-labelled TK DNA from pTK173. Plasmid DNA from positive clones was digested with PstI to determine the number of "a" sequence iterations. Five clones which apparently contained two iterations were further analysed and found to produce the maps shown opposite.

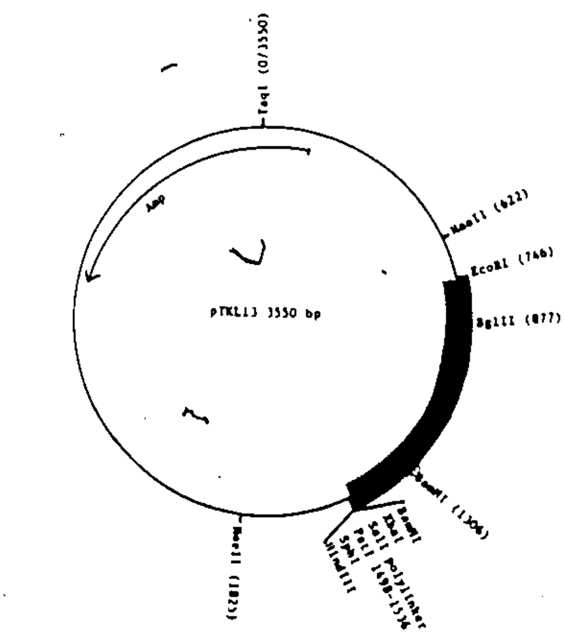
pTKL13 Amp^R lac^-

This plasmid contains the TK L terminus from KOSF10 cloned into the polylinker site on pUC19 (Norrander et al, 1983). KOSF10 viral DNA was digested with EcoRI and electrophoresed through a 1.4% agarose gel. The TK L terminus was electroeluted from the gel and ligated to SmaI/EcoRI digested pUC19. JM109 cells were transformed to ampicillin resistance on solid medium containing x-gal. Plasmid DNA from white colonies was screened for the appropriate restriction pattern. I should note here that repeated attempts to clone the TK S terminus from KOSF10 by this or a similar route (ie pUC18) were unsuccessful for totally obscure reasons.


pRIaa2
pRIaa4 7080 bp
pRIaa6
PstI (6328)
BamHI (1595)
SacII (1609)
SacII (1687)
BamHI (1861)
SacII (1899)
SacII (1977)
BamHI (2151)
PstI (2619)
EcoRI (2716)
BamHI (3091)
PvuII (4783)


pRIaa5 7080 bp
EcoRI (0/7080)
PstI (97)
BamHI (565)
PstI (1303)
PvuII (2287)
EcoRI (2716)
BamHI (3091)
PstI (6328)


pRIaa3 7080 bp
EcoRI (0/7080)
PvuII (431)
PstI (1413)
SacII (1687)
SmaI (1861)
SacII (1899)
SacII (1977)
BamHI (2151)
PstI (2619)
EcoRI (2716)
BamHI (3091)
PvuII (4783)
PstI (6328)

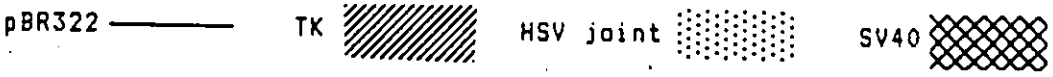

pTKL13 3550 bp
TaqI (0/3550)
Amp
EcoRI (746)
BglII (877)
BamHI (1306)
BamHI
XbaI
SalI
PstI
SphI
HindIII
polylinker

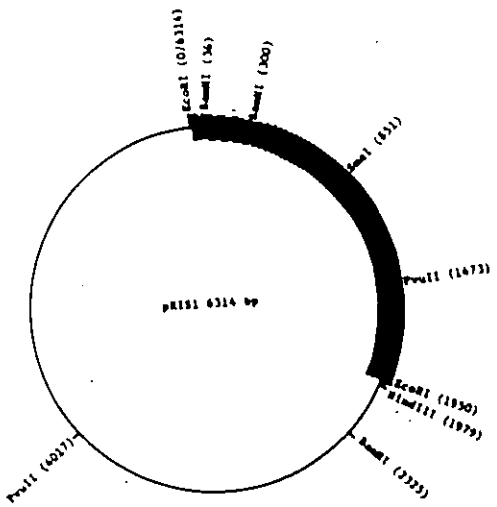
pRIS1 and pRIS3 Amp^R Tet^S

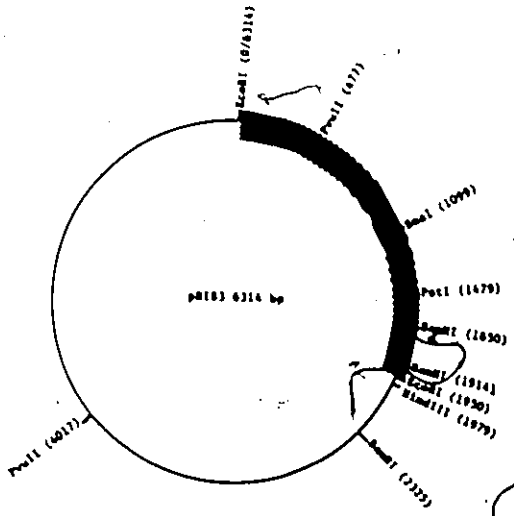
These plasmids contain the TK EcoRI S terminus from KOSF10 cloned into the EcoRI site on pBR322. pRIS1 and pRIS3 contain different orientations of the insert. KOSF10 viral DNA was treated with T4 polymerase and ligated to EcoRI linkers. The viral DNA was digested with EcoRI and electrophoresed through a 1.4% agarose gel. The TK S terminus was electroeluted and ligated to EcoRI digested pBR322. Ampicillin resistant, tetracycline sensitive colonies were screened by hybridizattion with ^{32}P-labelled TK DNA. Plasmid DNA from positive clones was then screened for the predicted restriction pattern.

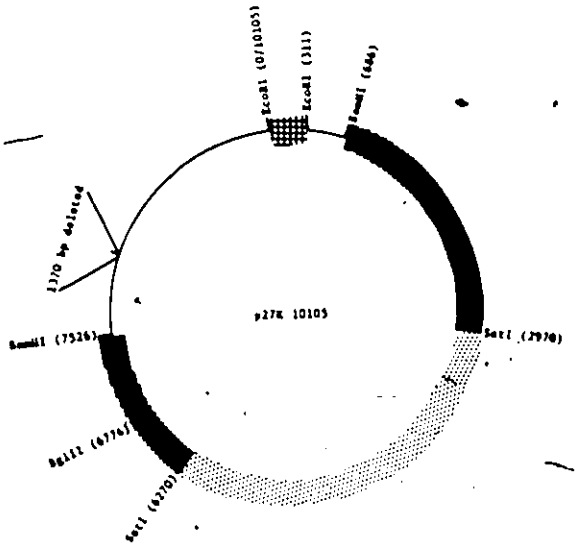
p27K and p27S Amp^R Tet^S

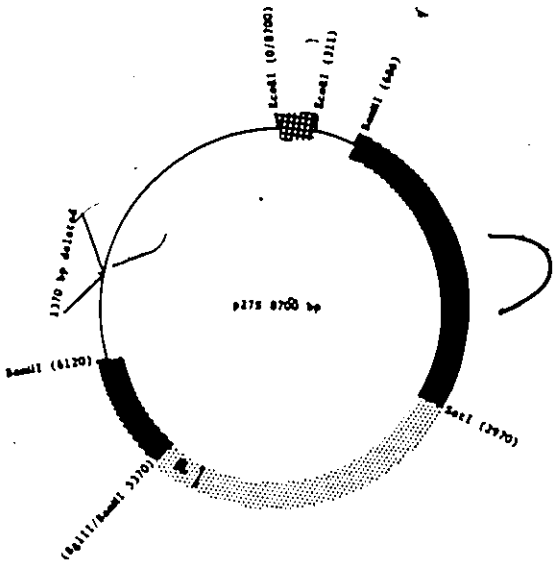
These plasmids were derived from p27 which was kindly supplied by Dr. S. Silverstein. p27 contains the TK containing BamHI Q fragment from HSV-1 (F) cloned into the BamHI site on pSV02 (Meyers et al, 1981). p27K contains the SacI joint spanning fragment from pSVODK6 cloned into the SacI site within the TK gene on p27. p27S contains the BamHI/SacI S terminal fragment from pTKSBS7 cloned into the BglII/SacI deleted TK gene on p27. The orientation of the SacI joint in p27K was not determined.




EcoRI (0/6316)
BamHI (36)
BamHI (200)
BamHI (851)
PvuII (1673)
EcoRI (1950)
HindIII (1979)
BamHI (3225)
PvuII (4017)
pRIB1 6316 bp


EcoRI (0/6316)
PvuII (47)
BamHI (1099)
PstI (1479)
BamHI (1650)
BamHI (1914)
EcoRI (1950)
HindIII (1979)
PvuII (4017)
BamHI (3225)


EcoRI (0/10105)
EcoRI (311)
BamHI (646)
1370 bp deleted
BamHI (7526)
HgiII (8776)
SacI (8370)
p27K 10105
SacI (2970)


EcoRI (0/8700)
EcoRI (311)
BamHI (646)
1370 bp deleted
BamHI (6120)
(HgiII/BamHI 5370)
p27S 8700 bp
SacI (2970)

REFERENCES

Alt, F. and Baltimore, D. (1982), "Joining of Immunoglobulin Heavy Chain Gene Segments: Implications From A Chromosome With Evidence of Three D-J_H Fusions", Proc. Natl. Acad. Sci. 79: 4118-4122.

Alt, F., Yancopoulos, G., Blackwell, T., Wood., C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., and Baltimore, D. (1984), "Ordered Rearrangement on Immunoglobulin Heavy Chain Variable Region Segments", The EMBO J. 3:1209-1219.

Baltimore, D. (1985), "Retroviruses and Retrotransposons: The Role of Reverse Transcription in Shaping the Eukaryotic Genome", Cell 40:481-482.

Barnett, J., Eppstein, D. and Chan, H. (1983), "Class I Defective Herpes Simplex Virus DNA as a Molecular Cloning Vehicle in Eucaryotic Cells", J. Virol. 48:384-395.

Benoist, C. and Chambon, P. (1981), "In Vivo Sequence Requirements of the SV40 Early Promoter Region", Nature 290:304-310.

Ben-Porat, T. and Veach, R. (1980), "Origin of Replication of the DNA of a Herpesvirus (Pseudorabiesvirus)", Proc. Natl. Acad. Sci. USA 77:172-176.

Ben-Porat, T., Kaplan, A., Stehn, B., and Rubenstein, A. (1976), "Concatemeric forms of Intracellular Herpesvirus DNA", Virol. 69:547-560.

Birnboim, H. and Doly, J. (1979), "A Rapid Alkaline Extraction Procedure for Screening recombinant Plasmid DNA" Nucleic Acids Res. 7:1513-1523.

Borst, P. and Cross, G. (1982), "Molecular Basis for Trypanosome Antigenic Variation", Cell 29:291-303.

Broach, J., Guarascio, V., and Jayaram, M. (1982), "Recombination Within the Yeast Plasmid 2 micron Circle is Site Specific", Cell 29:227-234.

Broker, T. and Doermann, A. (1975), "Molecular and Genetic Recombination of Bacteriophage T4", Ann. Rev. Genet. 9;213-244.

Campbell, A. (1983), "Bacteriophage λ", Mobile Genetic Elements, J.Shapiro ed., Academic Press, N.Y., pp 65-103.

Catcheside, D.G. (1977), The Genetics of Recombination, University Park Press, Baltimore, MD.

Clements, J.B., McLauchlan, J., and McGeoch, D. (1979), "Orientation of Herpes Simplex Virus Type 1 Immediate Early mRNAs", Nucleic Acids Res. 7:77-92.

Crick, F. and Lawrence, P. (1975), "Compartments and Polyclones in Insect Development", Science 189:340-347.

Davison, A. (1984), "Structure of the Genome Termini of Varicella Zoster Virus", J. gen. Virol. 65:1969-1977.

Davison, A. and Wilkie, N. (1981), "Nucleotide Sequences of the Joint Between the L and S Segments of Herpes Simplex Virus Types 1 and 2", J. Gen. Virol. 55:315-331.

Davison, A. and Wilkie, N. (1983), "Inversion of the Two Segments of the Herpes Simplex Virus Genome in Intertypic Recombinants", J. Gen. Virol. 64:1-18.

Delius, H. and Clements, J. (1976), "A Partial Denaturation Map of Herpes Simplex Virus Type 1 DNA: Evidence For Inversions of the Unique DNA Regions", J. Gen. Virol. 33:125-133.

Dressler, D. and Potter, H. (1982), "Molecular Mechanisms in Genetic Recombination", Ann. Rev. Biochem. 51:727-761.

Echols, H. and Green, L. (1979), Some Properties of Site-Specific and General Recombination Inferred From Int-Initiated Exchanges by Bacteriophage Lambda", Genetics 93:297-307.

Emmons, S. (1974), "Bacteriophage Lambda Derivatives Carrying Two Copies of the Cohesive End Site", J. Mol. Biol. 83:511-525.

Esperaza, J., Benyesh-Melnick, M., and Schaffer, P., (1976), "Intertypic Complementation and Recombination Between Temperature-Sensitive Mutants of Herpes Simplex Types 1 and 2", Virol. 70:372-384.

Fink, G. and Petes, T. (1984), "Gene Conversion in the Absence of Reciprocal Recombination", Nature 310:728-729.

Frenkel, N., Locker, H., Batterson, W., Hayward, G.S., and Roizman, B. (1976), "Anatomy of Herpes Simplex Virus DNA. VI. Defective DNA Originates from the S Component", J. Virol. 20:527-531.

Frenkel, N., Jacob, R., Honess, R., Hayward, G., Locker, H., and Roizman, B. (1975), "Anatomy of Herpes Simplex Virus DNA III. Characterization of Defective DNA Molecules and Biological Properties of Virus Populations Containing Them", J. Virol. 16: 153-167.

Friedmann, A., Rabkin, S., and Becker, Y. (1982), "Electron Microscopy of Branched HSV DNA Molecules: Possible Recombination Intermediates", in Herpesvirus DNA, ed. Y. Becker, Martinus Nijhoff, publishers, The Hague.

Gibson, M. and Spear, P. (1983), "Insertion Mutants of Herpes Simplex Virus Have a Duplication of the Glycoprotein D Gene and Express Two Different Forms of Glycoprotein D", J. Virol. 48:396-404.

Grafstrom, R., Alwine, J., Steinhart, W., Hill,W., and Hyman, R. (1975), "The Terminal Repetition of Herpes Simplex Virus DNA", Virol. 67:144-157.

Graham, F.L., and Van der Eb, A.J. (1973), "A New Technique for the Assay of the Infectivity of Adenovirus 5 DNA", Virology 52:456-467.

Haber, J. (1983), "Mating-Type Genes of Saccharomyces cerevisiae", Mobile Genetic Elements, J. Shapiro ed., Academic Press, N.Y., pp 559-619.

Hausmann, R. (1976), "Bacteriophage T7 Genetics", Curr. Top. Microbiol. Immunol. 75:77-110.

Hayward, G.S., Jacob, R.J., Wadsworth, S.C., and Roizman, B. (1975), "Anatomy of Herpes Simplex Virus DNA : Evidence for Four Populations of Molecules that Differ in the Relative Orientations of their Long and Short Segments", Proc. Natl. Acad. Sci. U.S.A. 72:4243-4347.

Heffron, F. (1983), "Tn3 and Its Relatives", Mobile Genetic Elements, J. Shapiro, ed., Academic Press, N.Y., pp 223-260.

Henderson, S.A. (1970), "The Time and Place of Meiotic Crossing-Over", Ann. Rev. Genetics 4:295-324.

Hirsch, I., Cabral, G., Patterson, M. and Biswal, N. (1977), "Studies on the Intracellular Replicating DNA of Herpes Simplex Virus Type 1", Virol. 81:48-61.

Hirt, B. (1967), "Selective Extraction of Polyoma DNA from Infected Mouse Cell Cultures", J. Mol. Biol. 26:365-369.

Holliday, R. (1964), "A Mechanism for Gene Conversion in Fungi", Genet. Res. 5:282-304.

Honess, R., Buchan, A., Halliburton, I., and Watson, W. (1980), "Recombination and Linkage Between Structural and Regulatory Genes of Herpes Simplex Virus Type 1: Study of the Functional Organization of the Genome", J. Virol. 34:716-742.

Honess, R. (1984), "Herpes Simplex and 'The Herpes Complex': Diverse Observations and a Unifying Hypothesis", J. Gen Virol. 65:2077 2107.

Iida, S., Meyer, J., and Arber, W. (1983), "Prokaryotic IS Elements", Mobile Genetic Elements, J. Shapiro, ed., Academic Press, N.Y., pp 159-221.

Jackson, J. and Fink, G. (1981), "Gene Conversion Between Duplicated Genetic Elements in Yeast", Nature 292:306-311.

Jacob, R. and Roizman, B. (1977), "Anatomy of Herpes Simplex Virus DNA VIII. Properties of the Replicating DNA", J. Virol. 23:394-411.

Jacob, R., Morse, L., and 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 Arrangements of Viral DNA", J. Virol. 29:448-457.

Jacob, R.J., Morse, L.S., and 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 Arrangmments of Viral DNA", J. Virol. 29:448-457.

Kaerner, H., Maichle, I., Ott, A., and Schroder, C. (1979), "Origin of Two Different Classes of Defective HSV-1 Angelotti DNA", Nucleic Acids Res. 6:1467-1478.

Kaerner, H., Ott-Hartman, A., Schatter, R., Schroder, C., and Gray, C. (1981), "Aplification of a Short Nucleotide Sequence in the Repeat Units of Defective Herpes Simplex Virus Type 1 Angelotti DNA", J. Virol. 39:75-81.

Keil, R.L. and Roeder, G.S. (1984), "Cis-Acting Recombination Stimulating Activity in a Fragment of the Ribosomal DNA of S. cerevisiae", Cell 39:377-386.

Klar, A. and Strathern, J. (1984), "Resolution of Recombination Intermediates Generated During Yeast Mating Type Switching", Nature 310:744-748.

Kleckner, N. (1983), "Transposon Tn10", Mobile Genetic Elements, J. Shapiro, ed., Academic Press, N.Y., pp 261-298.

Klein, H. (1984), Lack of Association Between Intrachromosomal Conversion and Reciprocal Exchange", Nature 310:748-753.

Knipe, D., Ruyechan, W., Honess, R., and Roizman, B. (1979), "Molecular Genetics of Herpes Simplex Virus. The Terminal 'a' Sequences of the L and S Components are Obligatorily Identical and Constitute a Part of a Structural Gene Mapping Predominantly in the S Component", Proc. Natl. Acad. Sci. U.S.A. 76:4534-4538.

Knipe, D., Ruyechan, W., Roizman, B. and Halliburton, I. (1978), "Molecular Genetics of Herpes Simplex Virus : Demonstration of Regions of Obligatory and Nonobligatory Identity Within Diploid Regions of the Genome by Sequence Replacement and Insertion", Proc. Natl. Acad. Sci. U.S.A. 75:3896-3900.

Knopf, C., Strauss, G., Ott-Hartmann, A., Schatten, R., and Kaerner, H. (1983), "Herpes Simplex Virus Defective Genomes: Structure of HSV-1 ANG Defective DNA of Class II and Encoded Polypeptides", J. Gen. Virol. 64:2455-2470.

Kornberg, A. (1974), DNA Replication, W.H. Freeman and Co., San Francisco.

Kruger, D. and Schroeder, C. (1981), "Bacteriophage T3 and Bacteriophage T7 Virus-Host Cell Interactions", Microbiol. Rev. 45:9-51.

Kunz, B. and Haynes, R. (1981), "Phenomenology and Genetic Control of Mitotic Recombination in Yeast", Ann. Rev. Genet. 15:57-89.

Ladin, B., Blankenship, M., and Ben-Porat, T. (1980), "Replication of Herpesvirus DNA V. Maturation of Concatemeric DNA of Pseudorabies Virus to Genome Length Is Related to Capsid Formation", J. Virol. 33:1151-1164.

Langman, L., Paetkau, V., Scraba, D., Miller, R., Roeder, G., and Sadowski, P. (1978), "The Structure and Maturation of Intermediates in Bacteriophage T7 DNA Replication", Can. J. Biochem. 56:508-516.

Lee, G., Pogue-Geile, K., Pereira, L., and Spear, P. (1982), "Expression of Herpes Simplex Virus Glycoprotein C From a DNA Fragment Inserted into the Thymidine Kinase Gene of this Virus", Proc. Natl. Acad. Sci. USA 79:6612-6616.

Lewin, B. (1977), Gene Expression-3, Plasmids and Phages, John Wiley and Sons, New York.

Locker, H. and Frenkel, N. (1979), "Structure and Origin of Defective Genomes Contained in a Serially Passaged HSV Type 1 (Justin)", J. Virol. 29:1065-1077.

Luder, A. and Mosig, G. (1982), "Two Alternative Mechanisms for Initiation of DNA Replication Forks in Bacteriophage T4: Priming by RNA Polymerase and by Recombination", Proc. Natl. Acad. Sci. USA 79:1101-1105.

Mackem, S. and Roizman, B. (1980), "Regulation of Herpes Virus Macromolecular Synthesis : Transcription-Initiation Sites and Domains of Genes", Proc. Natl. Acad. Sci. U.S.A. 77:7122-7126.

Maniatis, T., Fritsch, E., and Sambrook, J. (1982), Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor.

Matsuo, T., Heller, M., Petti, L., O'Shiro, E., and Kieff, E. (1984), "Persistence of the Entire Epstein-Barr Virus Genome Integrated into Human Lymphocyte DNA", Science 226:1322-1324.

Maxam, A. and Gilbert, W. (1980), "Sequencing End-Labelled DNA With Base Specific Chemical Cleavage", Meth. Enzymol. 65:499-560.

Mellon, P., Parker, V., Gluzman, Y., and Maniatis, T. (1981), Identification of DNA Sequences Required For Transcription of the Human 1-Globin Gene in a New SV40 Host-Vector System", Cell 27:279-288.

Messelson, M. and Radding, C. (1975), "A General Model for Genetic Recombination", Proc. Natl. Acad. Sci. U.S.A. 72:358-361.

Meyers, R., Rio, D., Robbins, A., and Tjian, R. (1981), "SV40 Gene Expression is Modulated by the Cooperative Binding of T Antigen to DNA", Cell 25:373-384.

Mizuuchi, K., Kemper, B., Hays, J. and Weisberg, R. (1982) "T4 Endo nuclease VII Cleaves Holliday Structures", Cell 29:357-365.

Mocarski, E., Post, L. and Roizman, B. (1980), "Molecular Engineering of the Herpes Simplex Virus Genome : Insertion of a Second L-S Junction Into the Genome Causes Additional Genome Inversions", Cell 23:243-255.

Mocarski, E. and Roizman, B. (1981), "Site Specific Inversion Sequence of the Herpes Simplex Virus Genome: Domain and Structural Features", Proc. Natl. Acad. Sci. U.S.A. 78:7047-7051.

Mocarski, E. and Roizman, B. (1982a), "Structure and Role of the Herpes Simplex Virus DNA Termini in Inversion, Circularization and Generation of Virion DNA", Cell 31:89-97.

Mocarski, E. and Roizman, B. (1982b), "Herpes Virus-Dependent Amplification and Inversion of Cell-Associated Viral Thymidine Kinase Gene Flanked by Viral 'a' Sequences and Linked to an Origin of Viral DNA Replication", Proc. Natl. Acad. Sci. U.S.A. 79:5626-5630.

Morse, L., Buchman, T., Roizman, B., and Schaffer, P. (1977), "Anatomy of Herpes Simplex Virus DNA IX. Apparent Exclusion of Some Parental DNA Arrangements in the Generation of Intertypic (HSV-1 x HSV-2) Recombinants", J. Virol. 24:231-248.

Norrander, J., Kempe, T. and Messing, J. (1983), "Construction of Improved M13 Vectors Using Oligonucleotide-Directed Mutagenesis", Gene 26:101-106.

Orr-Weaver, T., Szostak, J., and Rothstein, R. (1981), "Yeast Transformation: A Model System For The Study of Recombination", Proc. Natl. Acad. Sci. U.S.A. 78:6354-6358.

Poffenberger, K. and Roizman, B. (1985), "A Noninverting Genome of a Viable Herpes Simplex Virus 1: Presence of Head-to-Tail Linkages in Packaged Genomes and Requirements for Circularization After Infection", J. Virol. 53:587-595.

Pogue-Geile, K., Lee, G., and Spear, P. (1985), "Novel Rearrangements of Herpes Simplex Virus DNA Sequences Resulting From Duplication of a Sequence Within the Unique Region of the L Component", J. Virol. 53;456-461.

Post, L., Mackem, S. and Roizman, B. (1981), "Regulation of Genes of Herpes Simplex Virus : Expression of Chimeric Genes Produced by Fusion of Thymidine Kinase with Gene Promoters", Cell 24:555-565.

Powling, A. and Knippers, R. (1976), "Recombination of Bacteriophage T7 In Vivo", Mol. Gen. Genet. 149:63-71.

Radding, C. (1982), "Homologous Pairing and Strand Exchange in Genetic Recombination", Ann. Rev. Genetics 16:405-437.

Roeder, G.S., and Fink, G. (1983), "Transposable Elements In Yeast", Mobile Genetic Elements, J. Shapiro, ed., Academic Press, N.Y., pp. 300-328.

Roeder, G.S., and Sadowski, P. (1978), "Pathway of Recombination of Bacteriophage T7 DNA In Vitro", Cold Spring Harbor Symposium Quantitative Biology 43:1023-1032.

Roizman, B. (1979), "The Organization of the Herpes Simplex Virus Genome", Ann. Rev. Genetics 13:25-57.

Roizman, B., Jacob, R., Knipe, D., Morse, L., and Ruyechan, W. (1978), "On The Structure, Functional Equivalence, and Replication of the Four Arrangements of Herpes Simplex Virus DNA", Cold Spring Harbor Symposium on Quantitative Biology 43:809-826.

Rubnitz, J. and Subrameni, S. (1984), "The Minimum Amount of Homology Required for Homologous Recombination in Mammalian Cells", Mol. Cell. Biol. 4:2253-2258.

Schaffer, P., Carter, V. and Timbury, M. (1978), "Collaborative Complementation Study of Temperature-Sensitive Mutants of Herpes Simplex Virus Types 1 and 2", J. Virol. 27:490-504.

Sheldrick, P. and Berthelot, N. (1974), "Inverted Repetitions in the Chromosomes of Herpes Simplex Virus", Cold Spring Harbor Symp. Quant. Biol. 39:667-678.

Silverman, M. and Simon, M. (1983), "Phase Variation and Related Systems", Mobile Genetic Elements, J. Shapiro ed., Academic Press, N.Y., pp 537-557.

Skare, J. and Summers, W.C. (1977), "Structure and Function of Herpes Virus Genomes II:EcoRI, XbaI and HindIII Endonuclease Cleavage Sites on Herpes Simplex Virus Type 1 DNA", Virol. 76:581-595.

Skare, J., Summers, W.P., and Summers, W.C. (1975), "Structure and Function of Herpesvirus Genomes I. Comparison of Five HSV-1 and Two HSV-2 Strains by Cleavage of Their DNA with EcoRI Restriction Endonuclease", J. Virol. 15:726-732.

Smiley, J., Fong, B. and Leung, W.-C. (1981), "Construction of a Double Jointed Herpes Simplex Virus DNA Molecule : Inverted Repeats Are Required for Segment Inversion and Direct Repeats Promote Deletions", Virol. 113:345-362.

Southern, E. (1975), "Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis", J. Mol. Biol. 98: 503-517.

Spaete, R. and Frenkel, N. (1982), "The Herpes Simplex Virus Amplicon: A New Eukaryotic Defective-Virus Cloning-Amplification Vector", Cell 30:295-304.

Spaete, R. and Mocarski, E. (1985), "The 'a' Sequence of the Cytomegalovirus Genome Functions as a Cleavage/Packaging Signal for Herpes Simplex Virus Defective Genomes", J. Virol. 54:817 824.

Stahl, F. (1979), Genetic Recombination, Thinking About it in Phage and Fungi, W. H. Freeman and Co., San Francisco.

Stern, C. (1936), "Somatic Crossing-Over and Segregation in Drosophila melanogaster", Genetics 21:625-730.

Stow, N. and McMonagle, E. (1983), "Characterization of the TRs/IRs Origin of DNA Replication of Herpes Simplex Virus Type 1", Virology 130:427-438.

Stow, N., McMonagle, E., and Davison, A. (1983), "Fragments From Both Termini of the Herpes Simplex Virus Type 1 Genome Contain Signals Required for the Encapsidation of Viral DNA", Nucleic Acids Res. 11:8205-8220.

Strathern, J., Hicks, J., and Herskowitz, I. (1981), "Control of Cell Type in Yeast by the Mating Type Locus: the 1- 2 Hypothesis", J. Mol. Biol. 147:357-373.

Sutcliffe, J. (1979), "Complete Nucleotide Sequence of the Escherichia coli Plasmid pBR322", Cold Spring Harbor Symp. Quant. Biol. 43: 77-90.

Szostak, J., Orr-Weaver, T., Rothstein, R., and Stahl, F. (1983), "The Double-Strand-Break Repair Model for Recombination", Cell 33:25-35.

Tamashiro, J., Filpula, D., Friedman, T. and Spector, D. (1984), "Structure of the Heterogeneous L-S Junction Region of Human Cytomegalovirus Strain AD169 DNA", J. Virol. 52:541-548.

Tooze, J. (1980), DNA Tumor Viruses, Cold Spring Harbor Labs.

Toussaint, A. and Resibois, A. (1983), "Phage Mu: Transposition as a Life-Style", Mobile Genetic Elements, J. Shapiro, ed., Academic Press, N.Y., pp 105-158.

Varmus, H. (1983), "Retroviruses", Mobile Genetic Elements, J. Shapiro, ed., Academic Press, N.Y., pp 411-503.

Varmuza, S. and Smiley, J. (1984), "Unstable Heterozygosity in a Diploid Region of Herpes Simplex Virus DNA", J. Virol. 49:356-362.

Varmuza, S. and Smiley, J. (1985), "Signals for Site-Specific Cleavage of Herpes Simplex Virus DNA: Maturation Involves Two Separate Cleavage Events at Site Distal to the Recognition Sequences", Cell, in press.

Veira, J. and Messing, J. (1982), "The pUC Plasmids, an M13Mp7 derived System for Insertion Mutagenesis and Sequencing With Synthetic Universal Primers", Gene 19:259-268.

Vlazny, D. and Frenkel, N. (1982), "Replication of Herpes Simplex Virus DNA: Localization of Replication Recognition Signals Within Defective Genomes", Proc. Natl. Acad. Sci. U.S.A. 78:742-746.

Vlazny, D., Kwong, A. and Frenkel, N. (1982), "Site-Specific Cleavage/Packaging of Herpes Simplex Virus DNA and the Selective Maturation of Nucleocapsids Containing Full Length Viral DNA", Proc. Natl. Acad. Sci. U.S.A. 79:1423-1427.

Wadsworth, S., Jacob, R., and Roizman, B. (1975), "Anatomy of Herpes Simplex Virus DNA II. Size, Composition, and Arrangement of Inverted Terminal Repetitions", J. Virol. 15:1487-1497.

Wadsworth, S., Hayward, G., and Roizman, B. (1976), "Anatomy of Herpes Simplex Virus DNA V. Terminally Repetitive Sequences", J. Virol. 17:503-512.

Wagner, M. and Summers, W.C. (1978), "Structure of the Joint Region and the Termini of the DNA of Herpes Simplex Virus Type 1", J. Virol. 27:374-387.

Wagner, M. and Summers, W.C. (1981), "Nucleotide Sequence of the Thymidine Kinase Gene of Herpes Simplex Virus Type 1", Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445.

Watson, J.D. (1972), "Origin of Concatemeric T7 DNA", Nature New Biol. 239:197-201.

Watson, R., Preston, C. and Clements, J. (1979), "Separation and Characterization of Herpes Simplex Virus Type 1 immediate-early mRNAs", J. Virol. 31:42-52.

Whitehouse, H.K.L. (1977), Genetic Recombination: Understanding the Mechanisms, Wiley, N.Y.

Wigler, M., Sweet, R., Sim, G., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S., and Axel, R. (1979), Cell 16:777 785.

Wilkie, N. (1976), "Physical Maps for Herpes Simplex Virus Type 1 DNA for Restriction Endonucleases HindIII, HpaI and XbaI", J. Virol. 20:222-233.

Wilkie, N., Davison, A., Chartrand, P., Stow, N., Preston, V., and Timbury, M. "Recombination in Herpes Simplex Virus: Mapping of Mutations and Analysis of Intertypic Recombinants", Cold Spring Harbor Symposium Quant. Biol. 43:827-840.

Zeller, H., Zentgraf, H., and Kaerner, H. (1984), "Different Maturation Patterns of Various HSV-1 Strains: HSV-1 Ang Standard DNA id Packaged Starting at the L-End, Strains KOS and Ang Variant Path Packaging Starts at both the L-End and the S-End", Ninth International Herpesvirus Workshop, Seattle (Abstr.).