The Integration of Simian Virus 40 DNA Into Cellular DNA

WILLIAM NMN CHIA

Department of Biochemistry
Imperial College of Science and Technology
London SW7

Submitted for the degree of Ph.D. 1981

Acknowledgements

I am deeply indebted to P.W.J. Rigby for the time, effort, support and advice that he so generously provided during the course of this work.

I wish to acknowledge M.R.D. Scott for his collaboration in the construction of Homer I and to express my gratitude for the many helpful discussions he has provided.

I thank D.M. Glover for teaching me, amongst other things, electron microscopy, H. Roiha for the printing of my photographs, V. Lewis for help and advice on tissue culture, S. Hayman for the typing of the tables, C. Clayton and P.W.J. Rigby for proof-reading, P. Wray for the typing of the text and all the people who have provided me with enzymes and materials.

Finally, I want to express my affection for all the occupants of 608 both for putting up with me and for making it a thoroughly enjoyable place to work.

THESIS ABSTRACT

The fate of Simian Virus 40 (SV40) DNA following the infection of non-permissive BalbC-3T3 mouse cells with SV40 virions has been followed by using the transfer hybridisation technique. A novel high molecular weight form of SV40 DNA was found several days after infection. Restriction endonuclease analysis of this DNA indicates that the majority of it represents large head to tail polymers of viral DNA. Synthesis of this polymeric DNA requires DNA replication and is inhibited by arabinosyl cytosine. By coinfecting cells with two physically distinguishable viral genomes it can be demonstrated that the two types of viral genomes are extensively intermingled in the polymers. However, quantitative experiments indicate that the extent of recombination is insufficient to account for the formation of these polymers by a purely recombinational mechanism. A model for integrative recombination is proposed, based on the premise that these polymeric sequences are precursors to the integrated viral sequences in transformed cells. This model readily explains some of the known properties of integrated viral DNA.

To test this model, it is essential to be able to clone the intermediates involved in integration. To this end, experiments designed to purify double stranded DNA sequences containing integrated SV40 DNA using mercurated SV40 cRNA as an affinity label are described. The results indicate that while this method has great potential as an analytical technique its use as a preparative technique is severelylimited. However, following the advent of *in vitro* packaging systems for bacteriophage λ, with the associated high

efficiencies of molecular cloning, the need to prepurify sequences for cloning is obviated. We describe the development and refinement of a cosmid vector cloning system which allows the cloning of inserts with a mean size of 35Kb at high efficiency with no vector background using recA$^-$ bacterial hosts. Results from experiments designed to clone for integration intermediates using this system are discussed.

Contents

List of Tables

List of Figures

List of Abbreviations

bp	a single nucleotide or nucleotide pair of single or double stranded nucleic acid
BSA	bovine serum albumin
cpm	counts per minute
DNase	deoxyribonuclease I
DTT	dithiothreitol
EDTA	diaminoethanetetraacetic acid
FCS	foetal calf serum
h	hour(s)
Kb	one thousand nucleotides or nucleotide pairs of single or double stranded nucleic acid
mins	minute(s)
MOI	multiplicity of infection
PEG6000	polyethylene glycol (average molecular weight 6000 to 7500)
pfu	plaque forming unit
PI	post infection
Pipes	Piperazine-N,N'-bis (2-ethanesulphonic acid)
rDNA	DNA sequences coding for rRNA
rRNA	ribosomal RNA
RNase	ribonuclease
rpm	revolutions per minute
^{3}H-SAM	S-adenosyl-L-[methyl-^{3}H] methionine
SDS	sodium dodecyl sulphate
sp. act.	specific activity

SV40	Simian Virus 40
SV40(I)	supercoiled monomeric SV40 DNA
SV40(II)	relaxed circular monomeric SV40 DNA
TEMED	N,N,N',N'-tetramethyl-1, 2-diaminoethane
Tris	Tris (hydroxymethyl)-aminoethane

Chapter I

General Introduction

1.1. Introduction

In this chapter I shall concentrate on the process of SV40-induced cellular transformation. Only a brief description of the general and molecular biology of SV40 will be given. This subject has recently been extensively reviewed in Tooze (1980). The topics I wish to specifically discuss are: (1) the biology of viral transformation *in vitro*; (2) the biological effects of viral infection on the host cell prior to stable transformation; (3) the process by which viral DNA becomes stably associated with cellular DNA; (4) the viral functions which are involved in the establishment and maintenance of transformation. The discussion will in general be limited to SV40, except in cases where the only available examples come from the closely related polyoma virus.

1.2 Molecular biology of SV40

1.2.1 structure of SV40

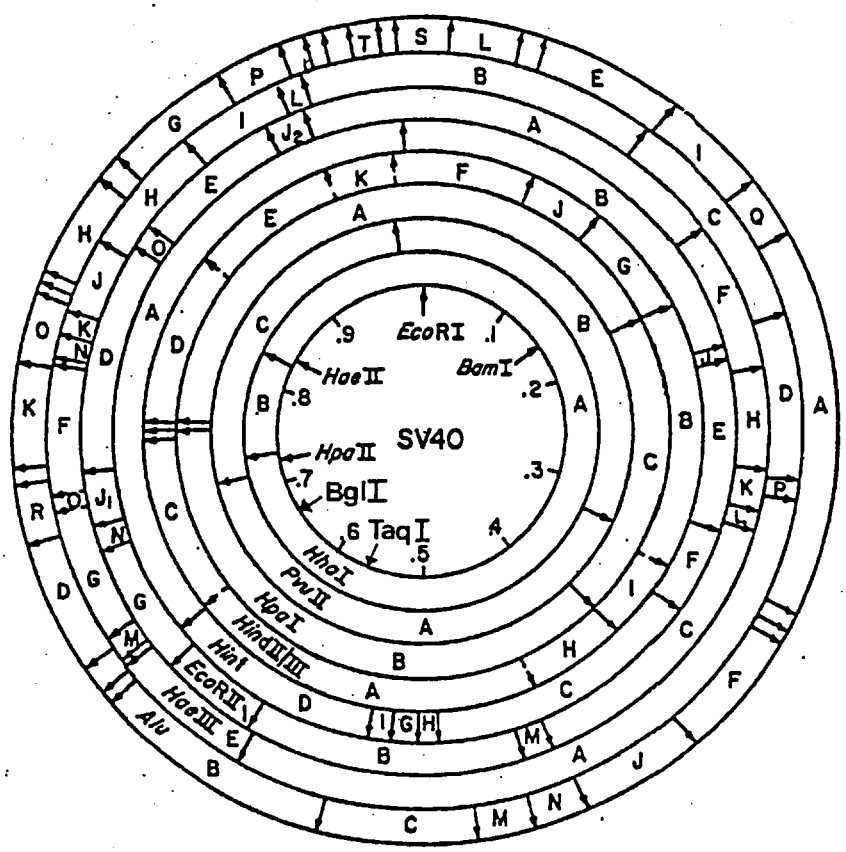
Each SV40 virion consists of a viral chromosome and an icosahedral capsid which surrounds the chromosome. The chromosome contains a single superhelical, double-stranded viral DNA molecule which is associated with four of the host cell's histones, H2A, H2B, H3 and H4. The capsid is made up of three virally coded polypeptides, VP1, VP2 and VP3. The viral DNA molecule has been extensively characterised by restriction endonuclease mapping (Fig. 1.1). As a result of this knowledge, the sequence of the 5243 b.p. molecule has been determined (Fiers *et al.*, 1978; Reddy *et al.*,

Legend to Figure 1.1

Restriction map of SV40 DNA

This figure was taken from Tooze (1980). Enzymes which cleave SV40 DNA at a single site are noted inside the circle. The single EcoRI site defines the zero coordinate. Each concentric ring shows the sites of cleavage by a given enzyme.

Fig.1.1


EcoRI .1
BamI
.2
.3
.4
.5
.6 TaqI
.7 BglI
HpaII
.8
HaeII
.9
SV40
AluI
HaeIII
EcoRII
HinI
HindII III
HpaI
PvuII
HhaI

1978b). The viral genome can be subdivided into 100 map units starting with the single EcoRI endonuclease site which is designated 0 and proceeding clockwise. The origin of replication is located at 0.67 map units, close to the single BglI endonuclease site (Danna and Nathans, 1972; Fareed et al., 1972). It is known that in SV40 chromatin the region around the replication origin (from 0.67 to 0.73 map units) is more susceptible to nucleolytic cleavage than other regions of the viral genome (Scott and Wigmore, 1978; Varshavsky et al., 1979) and is devoid of nucleosomes as determined by electron microscopy (Saragosti et al., 1980).

1.2.2 The sequence of events following viral infection

The very first step in viral infection is the adsorption of viral particles to the plasma membrane. The viral particles then penetrate the cell via pinocytic vesicles and are transported to the cell nucleus. It is thought that viral uncoating occurs in the nucleus. Following uncoating, the viral genome is expressed and this expression is temporally regulated. During the early phase, that is at times prior to the onset of viral DNA replication which occurs approximately 12h after infection, only the nonstructural viral proteins, large T-antigen and small t-antigen, are synthesised. The large T polypeptide is involved in the induction of host functions and the initiation of viral DNA synthesis. The onset of viral DNA replication occurs following the appearance of the large T polypeptide. The late phase is defined by the time of the onset of viral DNA synthesis and lasts until the infected cells die. During this phase the

coding information for the viral structural proteins VP1, VP2 and VP3 is expressed. The viral DNA is packaged inside viral capsids and infectious virus begins to appear at 20h to 25h after infection (Tooze, 1980).

1.2.3 Viral transcription

The early and late mRNAs of SV40 are transcribed in opposite polarities, using the two different viral DNA strands as templates (Lindstrom and Dulbecco, 1972; Sambrook et al., 1972; Khoury et al., 1975). The DNA strand which serves as the template for early transcription is designated the E-strand and the complementary DNA strand is designated the L-strand. The stable early cytoplasmic messages span the region of the SV40 genome from 0.67 to 0.17 map units (counterclockwise) and this half of the SV40 genome is designated the early region. The late messages span the other half of the genome which is designated the late region. The arrangement of the SV40 transcripts and genes is given in Figure 1.2.

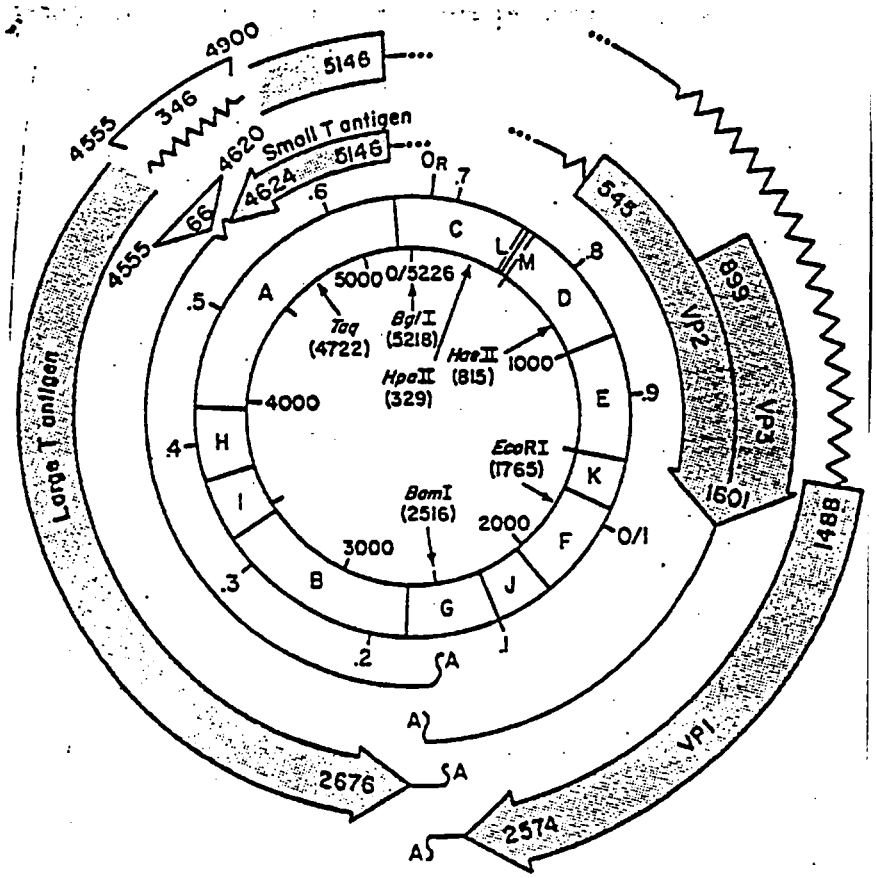
There are two early mRNAs which code for the large T and small t polypeptides (Paucha et al., 1978a; Berk and Sharp, 1978; Reddy et al., 1979). These mRNAs have essentially identical 5' and 3' termini but differ in size by approximately 280 nucleotides due to differences in splicing (Berk and Sharp, 1978; Reddy et al., 1979). The early messages appear not to have unique 5' ends but the four major 5' termini, which are capped (Haegeman and Fiers, 1980), all map within a 50 nucleotide stretch at 0.66 map units, near the origin of DNA replication (Reddy et al., 1979). The 3' ends of the early messages are polyadenylated and

Legend to Figure 1.2

A schematic representation of the SV40 genome indicating the locations of some principal features.

This figure was taken from Tooze (1980). The genome is represented as a circle with the origin of replication at the top. Numbers outside the circle refer to locations in fractional genome length relative to the single EcoRI cleavage site within SV40 DNA. Residue numbers and labels refer to the numbers used in the SV40 sequence (Fiers et al., 1978). Sequence numbers given for codons include termination codons. Shaded arcs with arrowheads indicate coding positions of mRNAs, with arrowheads pointing in the 5' to 3' direction. Outer boxed areas indicate continuous coding segments within the viral mRNA.

Fig.1.2


4900
5146
4555
346
4620
Small T antigen
5146
4624
66
4555
Large T antigen
2676
2574
VP1
1488
1601
VP2
VP3
545
899
OR
.6
.7
.8
.9
.5
.4
.3
.2
A
C
L
M
D
E
K
F
J
G
B
I
H
5000
0/5226
4000
1000
2000
3000
0/1
TaqI (4722)
BglI (5218)
HpaII (329)
HaeII (815)
EcoRI (1765)
BamI (2516)

overlap with the 3' ends of the late messages at 0.17 map units (Dhar et al., 1975). The 3' termini are probably generated by endonucleolytic cleavage (Ford and Hsu, 1978) and the hexanucleotide sequence AAUAAA, which is found close to the site of poly (A) addition of all eukaryotic mRNAs, is found in appropriate locations on the viral genome (Fiers et al., 1978; Reddy et al., 1978b). The mRNA for small t is colinear with the viral DNA sequence from an AUG at 0.66 map units to a UAA at 0.55 map units (Paucha and Smith, 1978; Volckaert et al., 1978; Reddy et al., 1979). The small t splice, which deletes 66 nucleotides from the nuclear precursor to generate the small t mRNA, starts one nucleotide downstream from the small t termination codon at 0.55 map units. The small t coding region on the viral genome is therefore contiguous (Reddy et al., 1979). The splice which generates the large T mRNA deletes 346 nucleotides from 0.59 to 0.54 map units. Since the 5' side of this splice is upstream to the small t splice, the small t termination codon is removed, thereby allowing large T translation to continue to essentially the 3' end of the message at 0.17 map units. Large T polypeptide is therefore coded by two noncontiguous parts of the early region and translated from the smaller of the two early 19S mRNAs, which is approximately 2200 nucleotides in length. Small t polypeptide is translated from the larger, 2500 nucleotide, early mRNA (Paucha et al., 1978a). The large T (M.W. = 94,000) and small t (M.W. = 17,000) proteins share amino-terminal sequences but differ at their carboxyl termini (Paucha et al., 1978a, b; Simmons and

Martin, 1978). A further consequence of the arrangement of the splices in the early mRNAs is that deletions in the region between 0.54 and 0.59 map units have an exclusive effect on the synthesis of small t protein, i.e. they either do not produce a small t polypeptide (Crawford et al., 1978) or they produce shortened small t polypeptides (Sleigh et al., 1978), but they always produce a normal large T polypeptide (Shenk et al., 1976; Sleigh et al., 1978).

The late messages of SV40 belong to two size classes, 16S which codes for VP1 (Prives et al., 1975) and 19S which codes for VP2 and VP3 (see Tooze, 1980, for discussion). Both classes of messages are spliced, capped and polyadenylated. The splicing occurs outside the coding regions so all three coding regions are contiguous on the SV40 genome. All of the 16S species share two features: (1) they contain the same 1209 nucleotide body spanning the region from 0.90 to 0.17 map units (clockwise) which includes the coding sequence; (2) they lack the 936 nucleotide sequence spanning 0.76 to 0.90 map units which is spliced out. The 5' leader sequences which are spliced onto the body are extremely heterogeneous. There are at least seven different 5' leaders, some of which involve a second splice which removes the region from 0.72 to 0.74 map units. The 5' termini of the 16S class are located between 0.70 and 0.75 map units, with the most abundant species having its 5' end at 0.72 map units (Ghosh et al., 1978a, b; Reddy et al., 1978a; Tooze, 1980). There are at least three types of 19S mRNA. The first type is a continuous copy of

SV40 DNA without any splices. The second type contains a 21 nucleotide splice located at 0.76 map units. The leaders are joined to a continuous 2117 nucleotide body spanning the region between 0.76 and 0.17 map units (clockwise). The third type of 19S mRNA has a larger 184 nucleotide splice removing the region between 0.73 and 0.76 map units. The leaders are joined to the same 2117 nucleotide body. The 19S mRNA species show extreme heterogeneity with respect to their leaders and 5' termini. There are at least 17 different 5' termini located between 0.67 and 0.73 map units. The most abundant 19S species shares the same 5' terminus, located at 0.72 map units, as the most abundant 16S species. (Ghosh et al., 1978a, b; Reddy et al., 1978a; Tooze, 1980).

VP2 and VP3 are translated in the same reading frame and the amino acid sequence of VP3 is identical to the carboxy-terminal two thirds of VP2 (Tooze, 1980). In the case of polyoma virus, it is known that an internal AUG codon of VP2 is used for the initiation of VP3 translation (Hunter and Gibson, 1978; Siddel and Smith, 1978). The carboxyl termini of VP2 and VP3 overlap with the amino terminus of VP1 and the initiation codon of VP1 lies within the VP2 coding region (Contreras et al., 1977); However, the overlapping sequences are read in different reading frames (Ghosh et al., 1978b). It should be noted that deletion of large portions of the SV40 DNA coding for the leaders of the late mRNAs does not result in the loss of function (Shenk et al., 1976; Subramanian, 1979; Villareal et al., 1979). However, the splice junctions are absolutely required for the formation of

stable late mRNA. If either or both splice junctions are deleted from SV40 DNA the late transcripts, though synthesised in the nucleus, are rapidly degraded (Lai and Khoury, 1979; Gruss et al., 1979; Hamer et al., 1979; Volkaert et al., 1979).

1.2.4. Regulation of viral gene expression

The rate of L-strand transcription relative to that of E-strand transcription increases from 1:5 to 20:1 from the early to the late phase. During the early phase both E and L-strands are transcribed but the L-strand is transcribed at a very low rate. The transcription of both strands is increased during the late phase, but the increase in L-strand synthesis is approximately two orders of magnitude greater than the increase in E-strand synthesis (Tooze, 1980). If permissive cells are infected with tsA mutants and maintained at the restrictive temperature, the amount of E-strand RNA is increased compared with cells infected by wild-type virus. Similarly, tsA infected cells maintained at the permissive temperature will increase E-strand RNA synthesis following a shift up to the nonpermissive temperature (Reed et al., 1976; Khoury and May, 1977). These observations are consistent with the notion that the large T polypeptide negatively regulates its own transcription. L-strand transcription is also affected by the large T polypeptide. When tsA infected cells are maintained at the nonpermissive temperature, L-strand transcription is not increased at late times (Cowan et al., 1973; Alwine et al., 1977; Khoury and May, 1977). This inhibition of late transcription is not due to the blockage of viral DNA replication by the tsA mutation at the nonpermissive temperature; when cells infected

with wild-type virus are maintained in the presence of DNA replication inhibitors, the amount of L-strand transcription at late times is still increased relative to early transcription (Rosenthal and Brown, 1977). Similarly, if DNA replication inhibitors are added to wild-type SV40 infected cells at late times, both E-strand and L-strand transcription remain unaffected (Manteuil and Girard, 1974; Alwine et al., 1977). Therefore, the switch to increased L-strand transcription at late times may depend on the presence of functional large T polypeptide but not on active viral DNA replication. However, once the switch to increased L-strand transcription has occurred, functional large T polypeptide is not required to maintain the increased rate of transcription; L-strand transcription is unaffected if tsA infected cells are shifted up to the nonpermissive temperature at late times (Birkenmeier et al., 1977; Khoury and May, 1977). It may be relevant that the 5' ends of both early and late mRNAs map in the region of the SV40 genome (between 0.67 to 0.73 map units) which, when isolated as chromatin, is preferentially cleaved by a variety of nucleases (Scott and Wigmore, 1978; Walbeck et al., 1978; Varshavsky et al., 1979). It is reasonable to presume that this region of the SV40 chromatin is more available for interaction with regulatory molecules (See section 1.6).

1.2.5 Viral replication

SV40 replication can occur by two different modes. Theta-form replication is initiated at a unique origin on the SV40 genome and proceeds bidirectionally to a terminus 180° from the initiation site. Pulse labelling (Danna and Nathans, 1972),

electron microscopic (Fareed et al., 1972) and genetic (Cole et al., 1977; Shortle and Nathans, 1979) experiments have shown that the origin of replication is located at 0.67 map units and that only approximately 120 nucleotides encompassing the BglI endonuclease site at 0.67 map units are essential for replication (Gutai and Nathans, 1978a; Subramanian and Shenk, 1978). The initiation of replication is sequence specific; when a segment of SV40 DNA (from 0.67 to 0.73 map units) containing the origin of replication is inserted at 0.17 map units, DNA replication can be initiated from this new position (Shenk, 1978). In contrast, the termination of replication is not sequence specific; using deletion mutants and evolutionary variants it has been shown that termination is purely a consequence of the fusion of the two replication forks; since the two replication forks travel around the genome in opposite directions at approximately the same rate, termination occurs 180° from the origin of replication (Brockman et al., 1975; Lai and Nathans, 1975a). The elongation process share similarities with procaryotic DNA replication; at each replication fork the synthesis of one strand appears to be continuous and the synthesis of the other strand discontinuous; RNA primers and Okazaki-like fragments are involved in the synthesis of the discontinuous strand (Tooze, 1980). Experiments with tsA mutants have demonstrated that functional large T polypeptide is required for the initiation but not for the elongation of theta-form replication; following shift up to the nonpermissive temperature, no newly synthesised viral replicative intermediates can be found; in contrast, pre-existing intermediates can be chased into

monomeric superhelical viral DNA (Tegtmeyer 1972; Chou et al., 1974). The biochemical details of the initiation of viral DNA synthesis are unclear.

While theta-form replicative intermediates comprise the great majority of the total replicating molecules during the early and middle stages of the lytic cycle, late in infection as much as 30% of the total SV40 DNA is present as 'head to tail' polymers (Rigby and Berg, 1978). It is likely that these viral polymers are generated by rolling circle replication and rolling circle replicative intermediates have been observed during lytic infection (Martin and Setlow, 1980).

1.2.6 The large T polypeptide

Large T polypeptide has an apparent molecular weight of approximately 94,000 by SDS gel electrophoresis (Tegtmeyer et al., 1975; Carroll and Smith, 1976). Under nondenaturing conditions this polypeptide can exist in multimeric forms (Carroll et al., 1974; Osborn and Weber, 1975a) and is also found to be associated in transformed and infected cells with a cellular protein (Lane and Crawford, 1979; McCormick and Harlow, 1980; E. Fanning et al., J.Virol., in press). Large T polypeptide is located in the nucleus (Tegtmeyer et al., 1975). That large T is virally coded is clearly demonstrated from a variety of genetic and biochemical studies (see Tooze, 1980, for review). Two enzymatic activities have been found to be associated with this protein, a protein kinase activity and an ATPase activity (Griffin et al., 1979; Tjian and Robbins, 1979); however Tjian et al. (1979) have shown that the kinase activity can be resolved from large

T-antigen; therefore this activity is probably due to an associated protein kinase and is not intrinsic to large T-antigen. The ATPase activity is probably intrinsic to large T polypeptide because: (1) it copurifies with large T to homogeneity; (2) monoclonal antibody directed against large T can inhibit this activity; (3) large T encoded by tsA mutants of SV40 possess a temperature sensitive ATPase activity (Tjian et al., 1979).

Large T polypeptide has been implicated in a number of biological activities: (1) it negatively regulates its own transcription (see section 1.2.4); (2) it is required for the initiation of viral DNA replication (see section 1.2.5); (3) it is responsible for the induction of cellular DNA synthesis and the induction of cellular enzymes (see section 1.4); (4) it is required to switch viral transcription into the late mode (see section 1.2.4); (5) it is required for the establishment and probably for the maintenance of viral transformation (see section 1.5). Precisely how large T polypeptide effects all of these activities is not known. However, recent findings have suggested that large T may act by direct interaction with DNA. It was demonstrated that the large T related polypeptide (D2T) encoded by the adenovirus 2 - SV40 hybrid virus Ad2^{+}D2 (Hassell et al., 1978), which contains all of the large T sequences except for approximately 100 amino acids from the amino-terminus, specifically binds to and protects the origin of SV40 DNA from nucleolytic cleavage (Tjian, 1978). In binding to the SV40 origin successively larger fragments (from 30 nucleotides to 75 nucleotides to 120 nucleotides), progressing in the direction of the late mRNA

caps, are protected with increasing ratios of protein to DNA (Fig.1.3). All of the protected sequences are essential for origin function (Gutai and Nathans, 1978; Subramanian and Shenk, 1978). Similar results have been obtained from nuclease protection experiments using large T polypeptide purified from an SV40-transformed cell line (Shalloway et al., 1980). The specific binding of large T and D2T polypeptides to the SV40 origin of replication strongly supports the notion that the requirement of the protein for the initiation of viral DNA synthesis is because it has a direct role. It may be relevant that the sequential binding of large T polypeptide to tandemly arranged recognition sites is reminiscent of the interaction between the bacteriophage λ repressor and operator (Maniatis and Ptashne, 1973; Meyer et al., 1975). It is possible that saturation of the large T binding sites will block the RNA polymerase entry site for early transcription and facilitate RNA polymerase binding for late transcription. In view of the recent finding that a major class of repeated DNA sequences (the Alu family), found in animal cells, contains a 14 nucleotide sequence homologous to a sequence at the SV40 origin of replication (Jelinek et al., 1980; also see Fig.1.3), and the demonstration that cloned segments of these sequences can bind large T protein (M. Botchan, personal communication), the interaction of large T protein with cellular DNA sequences may also explain its role in the induction of cellular DNA synthesis and the establishment and maintenance of transformation (see section 1.4 and 1.6).

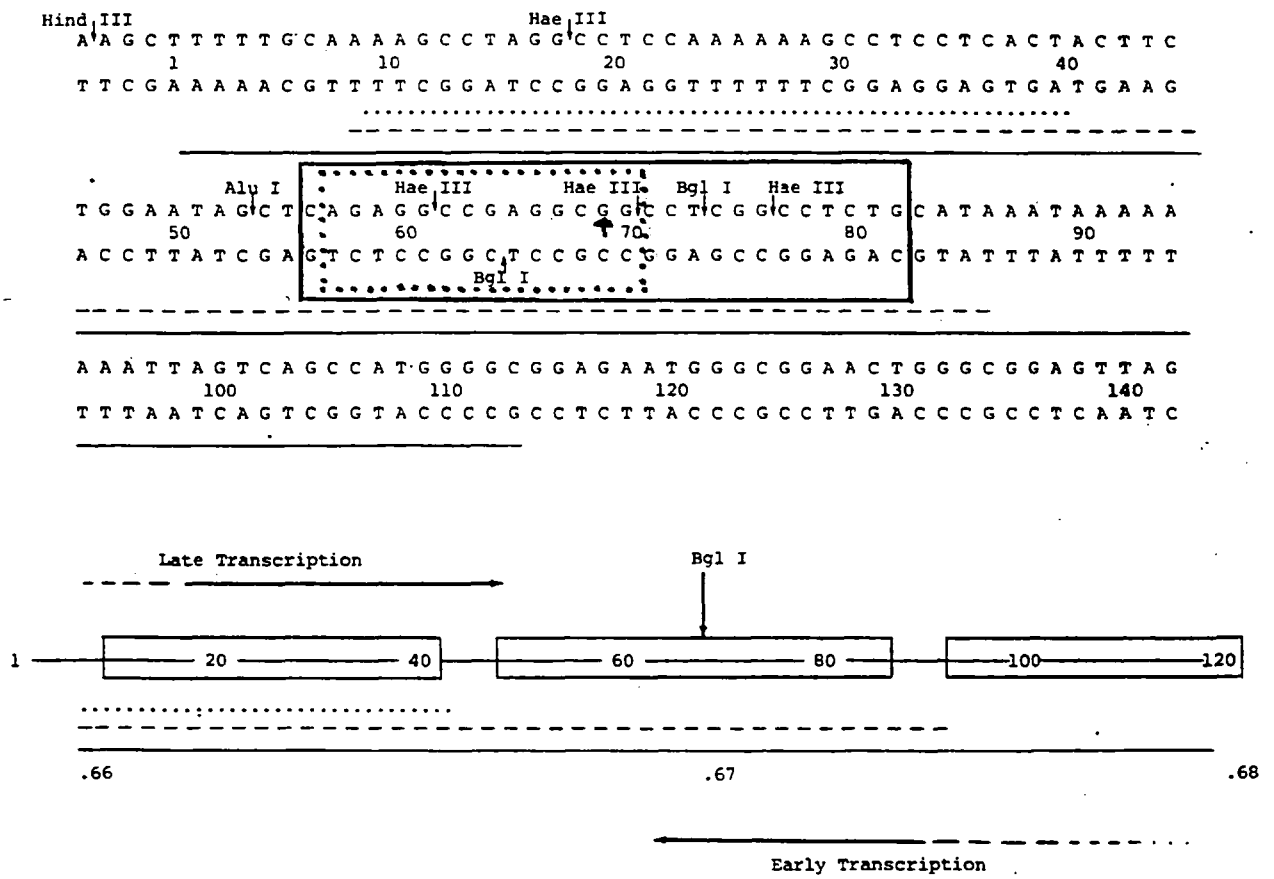
It should be stressed that large T protein exists in

Legend to Figure 1.3

The nucleotide sequence at the binding sites of the D2T hybrid protein

This figure is essentially taken from Tjian (1978) with a few added features. The sequence of 140 nucleotides in the region of the SV40 origin of DNA replication is given. The 30, 75 and 120 bp protected fragments are underlined by dotted lines, discontinuous lines and solid lines, respectively. The 14bp enclosed by the box (with dotted lines) represent the region of the SV40 genome which share extensive homology with the Alu family of interspersed repetitive sequences found in mammalian genomes. The 27bp sequence enclosed by the box (with solid lines) is comprised of a perfect inverted repeated with a 2-fold axis of symmetry centered at the nucleotide indicated by the arrow. The boxed areas in the lower diagram represent a possible arrangement of binding sites deduced from the pattern of protected fragments. The large numbers on the bottom represent coordinates on the conventional SV40 DNA map, and the leftward and rightward arrows represent the direction of early and late transcription, respectively.

Fig.1.3


Hind III
Hae III
Alu I
Hae III
Hae III
Bgl I
Hae III
Bgl I
Late Transcription
Bgl I
Early Transcription
.66
.67
.68

multiple forms: (1) it exists in differing degrees of aggregation (Carroll et al., 1974; Osborn and Weber, 1975a); (2) it contains at least two phosphorylation sites and exists in different states of phosphorylation (McCormick et al., 1979; McCormick and Harlow, 1980; E. Fanning et al., J. Virol., in press); (3) it can be associated with a cellular protein (Lane and Crawford, 1979; McCormick and Harlow, 1980). The possibilities that the various forms of large T may perform distinct biological functions and that the cellular protein(s) associated with large T may mediate some of its function(s) await clarification.

1.3 Biology of *in vitro* transformation

SV40 infection of normal cells *in vitro* can induce the formation of transformants. These transformants exhibit a large number of phenotypic alterations which include altered cell surface properties, intracellular enzyme levels and growth characteristics (Tooze, 1980). The altered growth characteristics, the ability to grow to high saturation density, the reduced requirement for serum, the ability to undergo anchorage independent growth and the ability to overgrow monolayers, allow these transformants to grow under conditions in which normal cells will not. As a result, transformation assays have been developed which allow for the specific selection of transformants (MacPherson and Montagnier, 1964; Todaro and Green, 1964; Smith *et al.*, 1971). It is important to note that any given transformation assay selects on the basis of only a subset of the growth properties exhibited by transformed cells. A study which exemplified the consequence of this point was performed by Risser and Pollack (1974). They obtained clonal lines from SV40 infected mouse cells in the absence of selection. Analysis of these lines demonstrated that transformants can be obtained which have acquired essentially all of the phenotypic alterations associated with transformation or only a subset of these alterations. Some of the transformants, which were called "minimal" transformants, were found to express large T antigen and to be able to grow in the presence of low serum concentrations. These would have been missed if selection had been imposed because they could not overgrow monolayers or grow in semi-solid medium. It is important

to note that transformation encompasses a wide spectrum of abnormal phenotypes.

Transformation by SV40 is a highly inefficient process. Between 10^2 and 10^4 infectious units are required for each transformation event (Tooze, 1980). Moreover, it has been demonstrated that the frequency of transformation is directly proportional to the MOI of infection which implies that only one infectious unit is required per transformation event (Todaro and Green, 1966a). That there is stable association of viral genetic information with the transformants is implied by three lines of evidence. Firstly, it has been demonstrated that stable transformants derived from SV40 infection contain viral DNA sequences (Westphal and Dulbecco, 1968) which are covalently integrated into cellular DNA (Sambrook et al., 1968). Secondly, it has been shown that virally coded early mRNAs and proteins are expressed by SV40 transformants (Oda and Dulbecco, 1968; Khoury et al., 1973; Ozanne et al., 1973; Tooze, 1980). Thirdly, it has been shown that a large number of SV40 transformants release infectious viral particles following fusion with permissive cells (Koprowski et al., 1967; Watkins and Dulbecco, 1967; Botchan et al., 1979).

The most important factor which determines whether or not SV40 infection will lead to the formation of transformants is the genetic properties of the host cell. For example, when African green monkey kidney cells, which are fully permissive to SV40, are infected with wild type SV40 all of the viral genes are expressed,

viral DNA replication occurs, infectious progeny virions are produced and the host cells are killed (Tooze, 1980). Therefore, permissive hosts are, in general, not transformed by SV40 even though it is possible to transform permissive cells by infecting with defective viruses (Shiroki and Shimojo, 1971; Gluzman et al., 1977). On the other hand, 3T3 mouse cells, which are totally nonpermissive to SV40, will survive viral infection. A majority of these infected cells will only transiently express the early viral functions and undergo several rounds of mitosis (Smith et al., 1971). These cells will not produce mature virions and no replication of viral DNA has been observed (Gershon et al., 1966). A small proportion of these cells will become stably transformed as a result of the viral infection. However, when very high MOIs are used (1000 pfu per cell) up to 40% of 3T3 cells can be transformed (Todaro and Green, 1966a). There exists a third type of host, which is semipermissive to SV40. When these cells are infected, some of the cells die because they can support the lytic events leading to the production of infectious virions. Other cells in the population behave as nonpermissive hosts and survive the infection. A proportion of the survivors become stably transformed. Transformants derived from semipermissive hosts will often contain unintegrated viral genomes in addition to integrated forms. Most species, including human and hamster cells, are semipermissive to SV40 (Tooze, 1980). Somatic cell hybridisation experiments have indicated that the cellular genome encodes a factor(s) which determines whether or not a given cell type will be permissive to viral (polyoma) infection

(Basilico et al., 1970). The biochemical basis of these permissivity factors is unknown.

1.4 Early events following nonpermissive infection

SV40 infection of nonpermissive cells will stimulate host cellular DNA synthesis under conditions (i.e. confluence or factor free medium) which do not allow normal, uninfected cells to undergo DNA replication (Smith et al., 1971; Scher, 1971; Setlow et al., 1980). Virally coded functions are required for this to occur since UV inactivated virus fails to induce this round of DNA synthesis (Chou and Martin, 1975; Henry et al., 1966). It is clear that the viral function required is not small t-antigen because small t deletion mutants (Δt), which fail to accumulate detectable amounts of this protein, will induce cellular DNA replication (Setlow et al., 1980). The involvement of large T-antigen in this phenomenon is directly supported by the observation that microinjection of purified large T antigen or the D2T protein into quiescent cells will induce a round of cellular DNA synthesis (Graessmann and Graessmann, 1976; Tjian et al., 1978). However, in apparent contradiction, both tsA and tsA, Δt double mutants will induce DNA synthesis at 40°C, the nonpermissive temperature for viral DNA replication (Chou et al., 1974; Setlow et al., 1980); at 42.5°C tsA mutants will not induce cellular DNA replication (Chou and Martin, 1975), suggesting that there are perhaps two distinguishable large T associated activities involved with DNA synthesis which have different sensitivities to high temperature. It should be noted that since large T synthesis is under autogenous regulation (Reed et al., 1976; Alwine et al., 1977), there

is significant accumulation of the mutant polypeptide at the nonpermissive temperature because the relative instability of the mutant polypeptide is counterbalanced by its increased rate of synthesis (Brockman, 1978; Edwards et al., 1979). Therefore, it does not necessarily follow that all large T functions are defective when infected cells are maintained at the nonpermissive temperature. It has been proposed (Martin and Setlow, 1980) that there are at least two DNA synthesis initiation activities associated with large T: (1) a low affinity, low specificity binding to DNA which can result in the initiation of DNA synthesis at the sites bound and; (2) a highly specific, high affinity binding to the SV40 origin of replication which is responsible for the initiation of DNA synthesis at this site. It was further argued that the mutant large T polypeptides produced by tsA mutants are defective only in the specific binding to the SV40 origin of replication and that the low affinity, low specificity binding activity to DNA is unaffected. Further support for the idea that large T polypeptide has two associated DNA synthesis initiation activities comes from the following two observations: (1) tsA infected permissive cells, when shifted to the nonpermissive temperature, will initiate viral DNA synthesis even though the amount of viral replication is reduced to 2% of the level observed at the permissive temperature; however, the sites of the initiations, observed by electron microscopy, appear to be randomly distributed with respect to the viral genome (Martin and Setlow, 1980); (2) there are two ways in which the large T polypeptide (and D2T

protein) can bind DNA; large T polypeptide can bind specifically to the SV40 origin of replication (Reed et al., 1975), protecting it from nucleolytic attack (Tjian, 1978) and it can bind nonspecifically to a variety of double-stranded DNAs as judged by a filter-binding assay (Carroll et al., 1974). Two observations lend credence to the idea that most existing tsA mutants code for mutant large T polypeptides which are defective for the specific binding of the SV40 origin of replication: (1) 12 out of 13 tsA mutants map in a small localised region of the viral genome, between map units 0.32 and 0.42 (Lai and Nathans, 1975b) suggesting the possibility that most tsA mutants affect the same protein domain; (2) there exist origin defective SV40 mutants which, as a result of their lower rate of DNA replication, produce plaques morphologically distinguishable from plaques produced by wild-type virus (Shortle and Nathans, 1979); second site reversion events which restore wild-type plaque morphology to these origin defective mutants map between SV40 map units 0.43 and 0.50, close to but distinct from the site of the tsA mutations (Shortle et al., 1979). The ability of tsA and tsA, Δt double mutants to induce cellular DNA replication may be attributed to the low specificity binding and DNA synthesis initiation activity of the temperature sensitive large T protein which is known to accumulate at near wild-type levels at the nonpermissive temperature. The available evidence is consistent with the idea that large T-antigen is necessary and sufficient to induce quiescent cells to undergo one round of DNA replication.

Associated with the induction of cellular DNA synthesis is the phenomenon of abortive transformation. Following nonpermissive infection, a large proportion of the infected cells transiently acquire the transformed phenotype, i.e. they express viral antigens and undergo several rounds of cellular division under conditions where normal, uninfected cells will not divide (Smith et al., 1971; Scher, 1971). Some virally coded function is required since UV inactivated virus does not induce this phenomenon. The number of virally induced cellular divisions appears to be directly proportional to the multiplicity of infection (Smith et al., 1971). While a majority of cells infected by wild-type SV40 can exhibit abortive transformation and form microcolonies in agar (Fluck and Benjamin, 1979) or microfoci in factor free medium (Smith et al., 1971), only a small proportion of these will become stably transformed. Small t-antigen may be required because it has been reported that Δt mutants are defective in the induction of abortive transformation (Fluck and Benjamin, 1979); However, this defect can be partially reversed by using a high multiplicity of infection. tsA mutants at the nonpermissive temperature are able to induce abortive transformation at frequencies similar to wild-type virus (Fluck and Benjamin, 1979). While there is no direct evidence, it can nevertheless be argued (see above) that large T-antigen is required for this phenomenon. Small t-antigen presumably plays a role in the subsequent rounds of cellular division following the effect of the initial induction of cellular DNA synthesis by large T-antigen.

It is possible that small t acts by enhancing cellular division in resting cells. These ideas are discussed in greater detail by Martin (1980).

It is tempting to speculate that the reason why a large proportion of cells which undergo abortive transformation do not become stably transformed is due to the failure of viral DNA to become stably integrated into cellular DNA. This is probably not the case. Smith et al. (1972) examined, by hybridisation kinetics, DNA prepared from 3 lines of abortive transformants which were phenotypically indistinguishable from their untransformed parental cell line. Two of the lines contained approximately 5 viral genome equivalents of SV40 DNA per cell. The reannealing kinetics were also monophasic which indicates, at least on a crude level, that the viral DNA was present as complete copies rather than multiple partial copies. This result demonstrates that the integration of SV40 DNA in itself is not sufficient for stable transformation.

It is clear, however, that following infection of nonpermissive cells, some event(s) must take place which allows some of the abortive transformants to become stably transformed. There is evidence that some of these events which fix the transformed state occur during the first round of cellular division following infection; if infected cells are maintained in a non-growing state (confluence), the frequency of transformation as determined by subsequent assay is greatly reduced. However, if infected cells are allowed one round of division and then maintained in a non-growing state, the transformation frequency is not reduced upon subsequent testing

(Todaro and Green, 1966b). Todaro and Green interpreted these results to mean that one round of cellular division is necessary for the fixation of transformation following infection. It is well known that the propagation of infected cells in a non-selective medium prior to assaying for transformation leads to an increase in the frequency of observed transformation (Scher, 1971; Risser and Pollack, 1974; Fluck and Benjamin, 1979). Scher (1971) demonstrated that the round of cellular division necessary to fix transformation must be induced by serum and cannot be substituted by the round of cellular division induced by viral infection; the observation is that even though viral infection induced 80% of the infected cells to divide in depleted medium, when these cells were subsequently assayed for transformation, the frequency of transformation was 50 fold lower when compared with infected cells which were assayed under serum conditions which allowed growth immediately following infection. It is not known whether viral DNA becomes stably associated with the cellular genome during this step or if this fixation process is necessary for the subsequent integration of viral DNA into cellular DNA sites which will allow the expression of viral information. However, it is clear that this fixation process is not merely the activation of viral information which is stably associated with the cellular genome. If that were the case, then one would not have expected to observe the decrease in transformation frequency when infected cells were maintained in a nongrowing state and then assayed for transformation under conditions which allow growth.

1.5 Integrated SV40 DNA in stably transformed cell lines

It is not known when viral DNA first integrates into cellular DNA. However, all stably transformed cell lines derived from SV40 infection contain viral DNA covalently linked to cellular DNA. A number of nucleic acid hybridisation, somatic cell hybridisation and restriction endonuclease mapping experiments have shown that different transformed cell lines contain different amounts of viral DNA integrated at different chromosomal sites (Gelb et al., 1971; Botchan et al., 1976; Ketner and Kelly, 1976; Croce, 1977; Campo et al., 1978; Kucherlapati et al., 1978; Chepelinsky et al., 1980; Rigby et al., 1980). The endonuclease mapping studies further demonstrate that the site on the viral genome which joins to cellular DNA is also not specific, even though the contiguity of the early region of the viral genome tends to be preserved. Cuzin and his colleagues have reported identical integration patterns for one set of three SV40-transformed rat lines and one set of two SV40-transformed mouse lines (Rigby, 1979b). The implication is that there exist preferential chromosomal sites for viral integration. However, this is the only observation of its kind and it is generally accepted that the integration of viral DNA is not specific with respect to either the viral or cellular DNA, at the level of sensitivity of endonuclease mapping. Despite this lack of specificity, each transformed line tends to maintain its characteristic integration pattern if selection is maintained during serial passage of the cells. Moreover, nonpermissive cells transformed by SV40 do not contain viral DNA which is not covalently attached to cellular DNA (Sambrook et al.,

1968). However, there is recent evidence which indicates that cloned SV40 transformed lines derived from nonpermissive cells can segregate progeny cell lines which exhibit nonparental viral integration patterns (M. Lovett, C. Clayton and P.W.J. Rigby; J. Hiscott, D. Murphy and V. Defendi; personal communications). The frequency with which the integrated viral genome undergoes spontaneous mutation, as evidenced by reversion of the transformed phenotype, is reasonably similar to that for cellular gene(s) (Steinberg et al., 1978).

A number of integrated viral genomes along with their adjacent cellular sequences have been cloned (Botchan et al., 1980; Rigby et al., 1980; J. Stringer, personal communication). The DNA sequences spanning the viral-cellular DNA junctions of the rat line 14B (Botchan et al., 1976), which was derived from DNA transfection, indicate that the flanking cellular sequences bear no resemblance to the sequence of SV40 DNA (Botchan et al., 1980). However, sequence data on the cellular-viral DNA junctions from two lines derived from infection with virus have shown short regions of homology between the cellular and viral sequences (Stringer, personal communication). While these results can be interpreted to mean that integrative recombination does not require extensive sequence homology, the validity of this interpretation is not clear. Taking the specific case of 14B, it is clear that sequences have been deleted from both the viral and adjacent cellular DNA. A segment of the viral genome is also inverted with respect to the major viral insertion and is separated from the

major insertion by 40 b.p. of unknown origin which shares partial homology with SV40 DNA (Botchan et al., 1980). While these observations are consistent with integration being the result of several illegitimate recombination events, it is impossible to conclude from structural information when and in what order these events occurred. Since DNA from the untransformed parental cell line corresponding to those flanking the viral insert has not been sequenced, it is also not possible to rule out the possibility that there is homology between viral DNA and the adjacent cellular DNA which is deleted or rearranged in the transformed cell line.

The role of homology in recombination between viral and cellular sequences has also been examined in the permissive system. During serial passage of SV40 in permissive cells, viral variants containing substitutions of cellular DNA arise (Tooze, 1980; Martin, 1980). Since these substituted viral genomes must have resulted from recombination between viral and cellular DNA, the host-viral DNA junctions present in these variants may be generated by a process similar to integrative recombination in nonpermissive cells. The extensive sequence analysis of the viral-cellular DNA junctions present in these variants (Gutai and Nathans, 1978b; McCutchan et al., 1979; Wakamiya et al., 1979), the comparison of these sequences with the known SV40 sequence and, in five fortuitous cases, with the known sequence of the African Green Monkey satellite (Rosenberg et al., 1978) which was the cellular component of five of the junctions, indicate that: (1) extensive nucleotide homology is not required for recombination between viral

and cellular DNA even though discontinuous and patchy homology may be important; (2) recombination occurred at many positions in both viral and cellular DNA although there seemed to be a preference for A.T rich clusters. Here again it is not possible to determine if the joints in question arose from the initial recombination between viral and cellular DNA or if they are the products of secondary events.

Another source of SV40-non SV40 DNA junctions is provided by the Ad2-SV40 hybrid viruses. The nondefective hybrid viruses contain an insertion of SV40 DNA at 0.86 Ad2 map units with a corresponding deletion of Ad2 DNA (Kelly and Lewis, 1973). It is known that DNA corresponding to 0.18 to 0.21 SV40 map units encodes the helper function which allows these hybrid viruses to grow in monkey cells (Grodzicker et al., 1974). The right hand junction of one of the hybrid viruses, $Ad2^{+}ND_1$ has been sequenced along with the corresponding region of the parental Ad2 DNA (Zain and Roberts, 1978). Comparison of the junction sequences with the sequences of the corresponding regions of SV40 and Ad2 DNAs indicate that there is no significant sequence homology between the two parental DNAs in the 95 b.p. sequence which spans the recombinant site. Assuming that recombination mechanisms are similar in monkey, mouse and rat cells, the sequence analyses presented are consistent with the notion that integrative recombination does not require extensive sequence homology.

A large proportion of integrated SV40 genomes found in virally transformed cell lines appear to contain complete or partial tandem duplications of the viral DNA arranged in a 'head to tail'

array (Botchan et al., 1976; Campo et al., 1978; Chepelinsky et al., 1980; Rigby et al., 1980). This effect is dramatic in semipermissive hosts transformed by SV40 DNA in which six to seven tandemly arranged copies of SV40 have been found (Campo et al., 1978); similar results have been obtained for rat lines transformed by polyoma (Basilico et al., 1979; Birg et al., 1979). However, partial tandem duplications occur at high frequency even in transformants derived from low MOI infections of cells which are nonpermissive for SV40 (Botchan et al., 1976; Chepelinsky et al., 1980; Rigby et al., 1980). It is clear that these structures cannot arise as the result of a single recombination event between monomeric, circular viral DNA and cellular DNA. One possible explanation for these tandemly duplicated viral genomes is that they are the result of the sequential integration of two or more monomeric SV40 DNA molecules into the same chromosomal site. This possibility is unlikely because of evidence from two sets of similar experiments. When a revertant cell line derived from 14B, which contains integrated viral DNA at a single chromosomal site and is phenotypically normal, is retransformed with microinjected SV40 DNA, the secondary viral integrations associated with the retransformation events invariably occurred at chromosomal sites distinct from the viral insert present in the revertant (Graessmann et al., 1979). Similar results were obtained when temperature sensitive transformants derived from infection with tsA viral mutants were retransformed with wild-type virus at the nonpermissive temperature (M. Botchan, personal communication).

A second possible explanation is raised by the work of Colantuoni *et al.* (1980). They studied rat lines transformed by the *tsa* mutant of polyoma virus. These lines contain a single viral DNA insertion which is stably maintained when the cells are propagated at the nonpermissive termperature for large T function. When these lines were first passaged at the permissive temperature and single colonies were then isolated and propagated at the nonpermissive temperature, the single integrated viral insert found in several of these "progeny" isolates had amplified part or all of the integrated viral genome found in the parental cell lines. In a majority of the isolates which had undergone amplification, the viral-cellular DNA junctions were identical to the ones found in the parental cell line. A reasonable explanation for this phenomenon is that under permissive conditions large T-antigen stimulates *in situ* replication of the integrated viral DNA; this DNA can then undergo homologous recombination with the viral insertion, thereby generating new tandem repeats. It is possible that the high frequency of partially tandemly duplicated viral genomes found in nonpermissive cell lines transformed by SV40 could have been generated by a similar mechanism. It is important, however, to stress that polyoma is different from SV40 in a number of respects and that rat cells are semipermissive hosts for polyoma. It should also be noted that the integrated viral DNA found in the parental cell lines used by Colantuoni *et al.* contained a partial tandem duplication of the viral genome prior to amplification. It is unclear how this parental tandem duplication was generated. There is also recent evidence which indicates that

polyoma transformed rat lines which contain viral insertions that are not tandemly duplicated cannot undergo amplification (C. Basilico, personal communication). These results raise the possibility that unequal crossing-over may be involved in amplification.

Finally, it is a formal possibility that oligomeric or polymeric forms of viral DNA may play a role in integrative recombination. While no direct evidence existed at the start of this work, there were several observations consistent with the speculation that replicative forms of viral and/or cellular DNA may be involved: (1) SV40 tsA mutants do not transform at the nonpermissive temperature (Tegtmeyer, 1972, 1975); one possible interpretation of this observation is that functional large T-antigen is required to initiate viral DNA replication, even in nonpermissive cells, in order for integration to occur; (2) it has been claimed that transformation and viral DNA integration are stimulated by agents which stimulate cellular DNA replication and repair (Fox and Levine, 1971; Hirai and Defendi, 1974, 1976); (3) oligomeric and polymeric forms of viral DNA are known to be synthesised following lytic infection; approximately 30% of total viral DNA late in lytic infection consists of 'head to tail' polymers of SV40 DNA (Rigby and Berg, 1978); these molecules are probably generated by rolling circle replication since rolling circle replicative forms of both SV40 and polyoma DNAs have been observed late in the lytic cycle (Bjursell, 1978; Martin and Setlow, 1980). If the viral substrates involved in integrative recombination are molecules that contain tandem duplications of the viral genome, one can then easily explain the

high frequency with which tandemly duplicated viral DNA is found integrated in the genomes of virally transformed cells.

1.6 Viral functions required for the transformed state

It has been demonstrated from both DNA transfection experiments using fragments of SV40 DNA and transformation experiments using conditional mutants of SV40 that the late functions of the virus are not required for transformation (Graham et al., 1974; Abrahams et al., 1975; Kimura and Itagaki, 1975; Martin and Chou, 1975; Mueller et al., 1978). That the large T protein is required for transformation is supported by three types of experiments. Firstly, it has been shown that tsA mutants do not transform at the nonpermissive temperature (Tegtmeyer, 1972, 1975). Secondly, transfection experiments using purified SV40 DNA fragments have shown that the early region of the SV40 genome, excepting sequences from 0.54 to 0.59 map units, is both necessary and sufficient for transformation (Graham et al., 1974; Abrahams et al., 1975; Martin, 1980). Thirdly, all of the SV40-transformed cell lines examined to date express all of the large T-antigen coding region (Martin, 1980). There is no doubt that large T antigen is required for transformation.

The requirement for small t is less clear because of complications in the interpretation of results. Whether or not Δ t mutants are defective for transformation depends on at least two factors; (1) the transformation assay used, and (2) the growth state of the cells at and shortly following the time of infection. Thus, as scored by the low density focus assay, in which infected cells are plated at low density under conditions which allow

active growth and dense foci are selected, the Δt mutants transform at frequencies approaching that of wild-type virus (Shenk et al., 1976; Bouck et al., 1978; Sleigh et al., 1978; Martin et al., 1979b), and therefore appear non-defective for transformation. However, if a colony assay which depends on the ability to undergo anchorage-independent growth, or the high density focus assay which depends on the ability to overgrow monolayers, is used, Δt mutants transform at much lower frequencies than wild-type virus (Bouck et al., 1978; Feunteun et al., 1978; Sleigh et al., 1978). Similarly, it has been reported that Δt mutants transform at wild-type frequencies, regardless of the assay used, if cells actively growing at and subsequent to the time of SV40 infection are used (Martin et al., 1979a, b; Seif and Martin, 1979b). Aside from the question of the requirement for small t in transformation, it is clear that small t and large T are functionally distinct, because under conditions where Δt mutants have reduced transformation frequencies, this defect can be complemented by tsA mutants at the restrictive temperature (Bouck et al., 1978; Fluck and Benjamin, 1979).

A second area of confusion concerns the role of small t in the maintenance of transformation. There is no direct way of approaching this question because conditional mutants of small t do not exist. However, a number of investigators have reported that transformants derived from Δt mutants either do not exhibit some of the characteristics associated with transformants derived with wild-type virus [i.e. growth in soft agar (Sleigh et al., 1978; Steinberg and Pollack, 1979), loss of actin cables and

induction of plasminogen activators (Graessmann et al., 1980; Topp and Rifkin, 1980)] or exhibit them at reduced levels, which implies that Δt mutants may be defective for a maintenance function(s). On the other hand, Martin and his colleagues (Martin et al., 1979a, b; Seif and Martin, 1979b) found that transformants derived by infection with Δt mutants of either resting or growing cells do not differ in their growth properties from transformants derived by infection with wild-type virus. The ability of Δt transformants to exhibit anchorage independent growth may depend on the type or batch of serum used in the growth medium (Martin et al., 1979b; Steinberg and Pollack, 1979). While the overall results are confusing, it is reasonable to propose that small t provides at least two functions which are involved in transformation. When resting cells are infected small t may supply a function necessary for the establishment of transformation. It is of interest that some tumour promoting agents, i.e. phorbol esters, can partially complement Δt mutants for this function (Martin et al., 1979a, b). The requirement for the proposed function is obviated when growing cells are used. The second function proposed for small t is involved with the maintenance of transformation. The need for this function may be obviated by certain types of serum.While the role of small t in transformation is unclear, it is clear that there is not an absolute requirement for small t in the transformation process. Under certain conditions, Δt mutants are capable of both transformation in vitro and tumour induction in vivo (Lewis and Martin, 1980).

Two types of experiment have been performed to determine whether large T is required for the maintenance of the transformed state. Firstly, transformants obtained by tsA infection at the permissive temperature were scored for temperature sensitivity of the transformed phenotype. Initial experiments of this type indicated that a majority of tsA transformed rat (Osborn and Weber, 1975b) and hamster (Brugge and Butel, 1975; Martin and Chou, 1975; Tegtmeyer, 1975) cell lines are temperature sensitive for a variety of phenotypes associated with transformation, including anchorage independent growth, reduced requirement for serum, lack of contact inhibition, and growth to high saturation densities. However, whereas Tegtmeyer could not find any temperature sensitive transformants from mouse cells, Brugge and Butel found that all of their mouse transformants were temperature sensitive. Subsequent to these conflicting results, a large number of investigators have reported finding both temperature sensitive (N) and temperature insensitive (A) SV40 transformants from a variety of nonpermissive and semipermissive hosts (Kimura and Itagaki, 1975; Brockman, 1978; Butel and Soule, 1978; Rassoulzadegan et al., 1978; Fluck and Benjamin, 1979; Martin et al., 1979a; Seif and Martin, 1979a; Chepelinsky et al., 1980; Kelley et al., 1980; O'Neill et al., 1980; Perbal and Rassoulzadegan, 1980; Rassoulzadegan and Cuzin, 1980; Robinson et al., 1980). It is clear that both types of transformants can be obtained from tsA infection. Moreover, tsA transformants have been isolated which are temperature sensitive with respect to some transformation phenotypes but not others

(O'Neill et al., 1980). That the temperature sensitivity of the N-type transformants is due to the virally coded function and not a temperature sensitive defect in a cellular gene is likely because these conditional transformants can be retransformed at the nonpermissive temperature by wild type virus (Brockman, 1978). Similarly, the A-type transformants cannot be attributed to trivial causes because virus rescued from these cell lines retains the tsA phenotype (Tenen et al., 1977; Brockman, 1978; Martin et al., 1979a). A variety of parameters have been examined by the investigators cited above including the multiplicity of infection, the particular tsA mutant used, the type of cells used, the growth state of cells during and subsequent to infection, the type of transformation assay used and the number of independent viral insertions in the cell lines, in order to determine the possible causes for these two apparently conflicting effects. However, each of the parameters examined either could not be correlated with the observed phenotype or the correlations observed were the subject of conflicting reports. So, while both A and N type transformants can be generated as a result of tsA infection, it is unclear which parameters or combinations of parameters are important in generating which phenotype. Taken at face value these results indicate that large T is required for the maintenance of transformation in only some of the transformants.

Since there is some evidence that the large T protein produced by tsA mutants is defective only for the recognition of the SV40 origin of replication but not for the ability to initiate cellular DNA replication at the restrictive temperature (see

sections 1.2.6 and 1.4), it can be argued that large T is supplying a function which is absolutely required for the maintenance of transformation. Tenen et al. (1977b) and Martin (1980) have proposed that the reason some tsA transformants are temperature resistant is not because large T is not required for the maintenance of transformation, but because they accumulate greater steady state quantities of large T antigen at the nonpermissive temperature than their temperature sensitive counterparts. While this is an attractive idea, the experimental evidence is not extensive. Edwards et al. (1979) examined the amount of large T-antigen in two tsA transformed hamster lines. One of the lines was temperature sensitive with respect to the ability to overgrow a cellular monolayer while the other line was temperature independent for this ability. Both lines were derived from the same untransformed parental cell line using the same tsA mutant and both lines contained a single viral DNA insert. At the nonpermissive temperature, the temperature insensitive line accumulated approximately three fold higher levels of large T than the temperature sensitive line, but approximately two fold less than a wild-type viral transformant. Similar results were obtained for two tsA transformed mouse lines, one of which was temperature sensitive and the other temperature independent for transformation (Brockman, 1978). These findings support the possibility that there is an absolute requirement for large T for the maintenance of transformation.

The possibility that large T polypeptide maintains the transformed state by directly stimulating a set of host replication

origins which are quiescent in normal cells is supported by the following observations: (1) transformed cells possess more origins of DNA replication than untransformed cells (Martin and Oppenheim, 1977); (2) there exists a major class of repeated sequences in animal cells which contains a 14 nucleotide sequence that is homologous with a sequence at the SV40 origin of replication (Jelinek, 1980; see Fig.1.3); there is recent evidence that cloned segments of human DNA which contain these sequences can bind large T polypeptide (M. Botchan, personal communication). While there is no direct evidence that this family of sequences is involved in cellular DNA replication or serves as cellular targets for large T, the observations that they are dispersed in the genome (Jelinek, 1978), that the 14 nucleotide sequence which is shared with the SV40 origin of replication forms part of the SV40 large T-antigen binding site(s) (Tjian, 1978) and that cloned segments of cellular DNA which contain these sequences can bind large T-antigen in vitro (M. Botchan, personal communication) are consistent with these speculations.

The second type of experiment designed to determine whether large T is required for the maintenance of transformation was reversion studies performed on SV40-transformed cell lines. The rationale was that if large T-antigen is not required for the maintenance of transformation, then selection for the normal phenotype should not select for the loss of the large T protein. The majority of revertants obtained by using negative selection with killing agents exhibit concommitant loss of most of the phenotypic

traits associated with transformation (Pollack et al., 1968; Pollack and Burger, 1969; Pollack et al., 1975), but they still express the large T-antigen. However, the interpretation of these results is complicated by the fact that most transformants contain multiple copies of integrated viral DNA. Since one would expect the viral large T gene to be dominant, if reversion was performed on transformants containing multiple functional viral inserts, one might expect the reversion events to occur by mutations in cellular genes. For this reason, the reversion experiments performed before 1978 are inconclusive because the cell lines used either contained, or probably contained, multiple viral inserts. Therefore, it is not surprising that all the revertants obtained were probably due to mutations in cellular genes as evidenced by the continued expression of large T-antigen by the revertants, the fact that wild type SV40 could be rescued from them following fusion with monkey cells (Pollack et al., 1968; Ozanne and Sambrook, 1971; Culp and Black, 1972; Renger and Basilico, 1973) and the fact that the revertants could not be retransformed by wild-type virus (Ozanne and Sambrook, 1971).

When reversion studies were performed on an SV40-transformed line which was known to contain only a single viral insert containing one copy of the large T gene (Botchan et al., 1976), the results strongly indicated that the transformed phenotype is directly dependent on the presence of a functional viral genome actively expressing large T-antigen (Steinberg et al., 1978). These investigators were able to isolate phenotypically normal revertants using 5-fluorodeoxyuridine as a selective agent. Under

the most stringent selective conditions, all revertants obtained were negative for large T expression. The majority of these revertants had either lost all of the integrated viral DNA or had incurred deletions of the early region of the viral genome. Some of the revertants retained the parental arrangement of integrated viral DNA despite the fact that no viral polypeptides were expressed. All of the revertants tested could be retransformed by SV40 DNA (Steinberg et al., 1978; Graessmann et al., 1979). These results indicate that reversion can occur either by direct mutational alterations of the integrated viral genome or possibly by alterations of cellular genes such that viral gene expression is blocked. In either case, these experiments clearly demonstrate that large T-antigen expression is required for the maintenance of transformation, at least for this particular transformed rat line. The question of whether or not large T is required for the maintenance of the transformed phenotype in all SV40-transformed cell lines cannot be answered until we have a better understanding of the large T protein and the role of this polypeptide in temperature independent tsA-transformants.

The particular aspect of SV40-induced transformation that is central to the work presented here is the integrated viral DNA which is found in all cell lines stably transformed by SV40. Since the integrated viral DNA represents approximately 10^{-6} of the total transformed cellular DNA, the ability to enrich for DNA fragments which contain the viral DNA sequences would greatly facilitate their study. Therefore, the first set of experiments, presented in Chapter 3, were designed to assess the feasibility of

using mercurated RNAs as an affinity probe to enrich for double stranded DNA molecules which contain sequence homology with the RNA probe; the object was to investigate the possibility of using this method to enrich for DNA fragments containing integrated viral DNA. The work presented in Chapters 4 and 5 is concerned with the mechanism by which SV40 DNA integrates into cellular DNA. The structure of the integrated viral DNA possesses certain features which are not readily explained (i.e. the high frequency with which partial tandem duplications of the viral genome are found). It is possible that these structural features arise as a consequence of events which occur at times shortly following viral infection. In Chapter 4 this possibility was investigated by directly following the fate of the infecting viral DNA. The results of Chapter 4 raised the possibility that the initial integration of SV40 DNA may involve 'head to tail' oligomers or polymers of viral DNA. In order to investigate this possibility it was necessary to develop a high efficiency DNA cloning system suitable for cloning large DNA segments containing tandemly repeated SV40 sequences. The work presented in Chapter 5 details the development of such a cloning system and some preliminary experiments designed to investigate the possibilities raised by the work presented in Chapter 4.

Chapter 2

Materials and Methods

2.1 Mammalian cells and viruses

2.1.1 Growth of cells (Mertz and Berg, 1974)

All cell lines were grown at 37°C in 9cm petr. dishes (manufacturer, Nunclon) in an incubator gassed with 5%(v/v) CO_2 in air. The African Green Monkey kidney cell lines, CV-1, CV-1P and MA-134 were grown using DMEM (manufacturer, Flow Laboratories) supplemented with 10% (v/v) calf serum (manufacturer, Gibco), 500 units per ml of penicillin (manufacturer, Sigma) and 100 μg per ml of streptomycin sulphate (manufacturer, Sigma). The mouse Balb/c-3T3 clone A31 cells were grown as above with 10% (v/v) foetal calf serum (manufactuer, Gibco) instead of calf serum.

2.1.2 Virus preparation (Estes et al., 1971)

The viruses used were obtained from Professor P. Berg, Department of Biochemistry, Stanford University. Their origins are described in the references given: wt830 (Mertz et al., 1974); dl(pm)861 (Carbon et al., 1975); dlF884 (Shenk et al., 1976). Virus stocks were prepared by infecting confluent monolayers of MA-134 cells at a MOI of 0.05. Following infection, the cells were fed with DMEM supplemented with 2%(v/v) calf serum. About 11 days PI, when about 90% of the cells exhibited cytopathic effect, the cells together with the medium were scraped into a centrifuge bottle and 0.1 volumes of 1M Tris·HCl, pH8, were added. Following three rounds of freeze-thawing, 0.2 volumes of $CHCl_3$ were added and the mixture was vigorously shaken. The phases were cleared by centrifugation at 3000 rpm for 5 mins at 4°C in a Sorvall GS-3 rotor.

Table 2.1

List of Buffers and Solutions

Buffer A: 20mM Tris HCl, pH8.0, 3mM $MgCl_2$, 1mM EDTA, 0.5% (v/v) β-mercaptoethanol

Buffer M1: 6mM Tris HCl, pH7.5, 30mM spermidine HCl (neutralised with Tris base), 60mM putrescine (neutralised with Tris base), 18mM $MgCl_2$, 15mM rATP (neutralised with NH_4OH), 28mM β-mercaptoethanol

EcoRI dilution buffer: 20mM KPO_4, pH7.5, 50mM NaCl, 10mM β-mercaptoethanol, 0.2%(v/v) Triton X-100, 0.2%(w/v) gelatin

Gel loading buffer: 60%(v/v) sucrose, 1.5%(w/v) Ficoll, 25mM EDTA (neutralised with NaOH), 0.2%(w/v) Bromophenol Blue or Orange G

Hirt extraction buffer: 10mM Tris HCl, pH7.5, 10mM EDTA, 0.6% (w/v) SDS

Phage buffer: 10mM Tris HCl, pH7.4, 100mM NaCl, 10mM $MgCl_2$, 0.1%(w/v) gelatin

PM: 0.02% (w/v) polyvinylpyrrolidone, 0.02% (w/v) BSA, 0.02% (w/v) Ficoll

S1 buffer: 30mM sodium acetate, pH4.5, 1mM $ZnSO_4$, 250mM NaCl and 100 ug per ml of denatured, sonicated salmon sperm DNA

SSC: 150mM NaCl, 15mM trisodium citrate

TB electrophoresis buffer: 89mM Tris, 89mM boric acid, 2.5mM EDTA, pH8

TBS: 25mM Tris HCl, pH7.4, 140mM NaCl, 5mM KCl, 0.7mM Na_2HPO_4, 0.5mM $MgCl_2$, 1mM $CaCl_2$

TE: 10mM Tris HCl, pH8.1, 1mM EDTA

TNE: 10mM Tris HCl, pH7.4, 100mM NaCl, 1mM EDTA

Tris-acetate electrophoresis buffer: 40mM Tris, 5mM sodium acetate, 1mM EDTA, adjusted to pH7.8 with acetic acid

Table 2.2

Media

DMEM	Dulbecco's modification of Eagle's medium
F-agar	L-agar containing Hogness freezing salts
Hogness freezing salts	6.39 K_2HPO_4, 1.8g KH_2PO_4, 0.45g trisodium citrate, 0.09g MgSO4, 0.9g $(NH4)_2SO_4$ and 44g glycerol per litre
L-broth	Luria broth, containing 10g tryptone, 5g yeast extract and 5g NaCl per litre
L-agar	L-broth plus 15g per litre of agar
Top-agar	L-broth plus 9g per litre of agar

Solid PEG 6000 was slowly dissolved into the supernatant until the final concentration was 7% (w/v). Following gentle stirring at 4°C overnight, the viral precipitate was pelleted by centrifugation at 7000 rpm for 30 mins at 4°C in a Sorvall GS-3 rotor and resuspended in 0.1 to 0.05 volumes of 25mM Tris·HCl, pH7.6, 150 mM NaCl. Debris was removed by centrifugation at 10,000 rpm for 40 mins at 4°C in a Sorvall HB-4 rotor. The resultant supernatant was layered over a CsCl step gradient (3ml, ρ = 1.38; 3ml, ρ = 1.30 dissolved in 25mM Tris·HCl, pH7.6, 150mM NaCl and centrifuged at 25,000 rpm for 3h at 4°C in a Beckman SW27 rotor. The visible viral band was collected by side puncture with a syringe, dialysed against 25mM Tris·HCl, pH 7.6, 150mM NaCl, 1 mM $MgCl_2$, sterilised with 0.02 volumes of $CHCl_3$, made 2% (v/v) in foetal calf serum, and stored in liquid N_2.

2.1.3 Plaque assays (Mertz and Berg, 1974)

Confluent monolayers of CV-1P cells grown in 6 cm petri dishes (manufacturer, Falcon) were used. Virus was diluted to 0.2 ml with TBS containing 2% (v/v) foetal calf serum and added to the cellular monolayers from which the growth medium had been removed. Adsorption was allowed to occur at 37°C for 1.5 h with occasional rocking of the plates. The cells were then overlayed with 5 ml of agar medium, consisting of Auto-Pow Minimal Essential Medium (manufacturer, Flow) containing 4% (v/v) foetal calf serum, 1% (w/v) agar, 500 units per ml of penicillin and 100 μg per ml of streptomycin sulphate, and incubated at the appropriate temperature. The cells were overlaid every 5 days with 4 ml per plate of agar medium containing 1% (w/v) rather than 4% (w/v)

foetal calf serum. Agar medium containing 0.01% (w/v) Neutral Red was used as the final overlay to visualise the plaques.

DNA plaque assays were performed using the DEAE-Dextran method of McCutchan and Pagano (1968). Monolayers were twice washed with TBS and incubated for 15 mins at room temperature with 0.2 ml of TBS containing DNA and 500 µg per ml of DEAE-Dextran (average molecular weight = 2×10^6). The cells were washed twice with TBS, overlaid with agar medium and treated subsequently as described above.

2.2 Strains of bacteriophage, bacteria and bacterial plasmids

2.2.1 Phage strains

(1) CI857 Sam 7 was obtained as a lysogen from Professor K. Murray, Department of Molecular Biology, University of Edinburgh.

(2) Charon 4 (Blattner et al., 1977) was obtained from Professor F. Blattner, Department of Genetics, University of Wisconsin at Madison.

(3) The recombinant phage λ SV contains the genome of the SV40 variant ev-1117 (Brockman et al., 1975) cloned in the Hind III replacement vector λ NM763 (Murray et al., 1977) and was made available to us by Professor K. Murray.

All of the phage strains were stored after equilibrium density gradient centrifugation in 4M CsCl at 4°C.

2.2.2 Bacterial strains used:

(1) ED8767 (met$^-$, Hsdk S$^-$ R$^+$ M$^+$, sup E, sup F, recA56; Murray et al., 1977) was obtained from Professor K. Murray.

This strain served as host for the phage λ SV.

(2) K802 (*hsd*$_K$ R^-M^+, *gal*$^-$, *sup*E) was obtained from Professor T. Maniatis, Division of Biology, California Institute of Technology. This strain was used as the host for Charon 4.

(3) LE392 (*hsd*$_K$ R^-M^+, *gal*$^-$, *sup*E, *Sup*F) was obtained from Professor F. Blattner. This strain was used for λ *c*I857 Sam 7 plaque assays.

(4) HB101 (*hsd*$_K$ R^-M^-, *leu*$^-$, *pro*$^-$, *rec* A^-) was obtained from Dr D. Glover, Imperial College. This strain was used for all transformation experiments involving recombinant DNA.

(5) BHB2688 and BHB2690 were obtained from Dr. D. Ish-Horowicz, ICRF, Mill Hill. These temperature-sensitive lysogens were used to make lysates for *in vitro* packaging of λ DNA (Hohn and Murray, 1977; Sternberg *et al*., 1977). Bacterial strains were stored at -70°C after diluting fresh, saturated overnight cultures grown in L-broth with an equal volume of 2X Hogness freezing salts.

2.2.3 Plasmids used:

pBR322 (Bolivar *et al*., 1977b) and pAT153 (Twigg and Sherrat, 1980) were obtained from Dr. S. Kidd, Imperial College and Professor D. Sherrat, University of Sussex. Recombinant plasmids carrying *D. melanogaster* rDNA, cDM103 (Glover *et al*., 1975), cDM219 (Kidd and Glover, 1980) and cDM238 (Glover, personal communication) were supplied by Dr. S. Kidd and Ms. H. Roiha, Imperial College. All plasmids were propagated in HB101.

2.3 DNA preparation

2.3.1 Plasmid DNA

All manipulations involving E.coli strains carrying recombinant plasmids were performed according to the advice received from the Genetic Manipulation Advisory Group. Plasmid DNAs were prepared from saturated overnight cultures essentially as described by Wensink et al., (1974). 100ml of culture was centrifuged at 5,000 rpm for 5 mins at 4°C in a Sorvall GS-3 rotor. The cellular pellet was resuspended in 1.5ml of 25% (w/v) sucrose in 50mM Tris.HCl, pH8.0.0.5 ml of a fresh 10mg per ml aqueous lysozyme solution and 0.5 ml of 0.5 M EDTA (adjusted to pH8 with NaOH) was added and the cellular suspension was gently swirled on ice for 10 mins. The resultant spheroplasts were lysed by addition of 2.5 ml of 50mM Tris·HCl, pH8.0, 62.5mM EDTA, 0.1% (v/v) Triton X-100 followed by gentle swirling on ice for 5 mins. The lysate was cleared by centrifugation at 15,000 rpm for 30 mins at 4°C in a Sorvall SS-34 rotor. Ethidium bromide was added to the cleared lysate at a final concentration of 0.5 mg per ml and the density was adjusted to 1.56g per ml with solid CsCl. Following density gradient equilibrium centrifugation in either a Beckman 60Ti rotor or a 65Ti rotor at 40,000 rpm for 2 to 3 days at 15°C, DNA was visualised with a Blak-ray long wave UV light. The lower band containing supercoiled DNA was collected by side puncture of the tube with a syringe. CsCl was removed by dialysis against TE and ethidium bromide was removed by several extractions with phenol neutralised by equilibration with an equal volume of 1M Tris·HCl, pH8.1.

The above procedure was modified for rapid small scale

preparations of partially pure plasmid DNA. 1 ml saturated cultures were grown in 1.5 ml Eppendorf microfuge tubes. The lysis procedure was scaled down and the mini lysates were cleared by a 10 mins centrifugation in a Beckman microfuge at 4°C. The resultant cleared lysates were phenol extracted, $CHCL_3$ extracted, and precipitated with 0.5 volumes of isopropanol or 2 volumes of ethanol. Following two 70%(v/v) ethanol washes, the DNAs were resuspended in TE. This DNA is mostly restrictable plasmid DNA with some chromosomal contamination.

2.3.2 λ DNA

To prepare DNA from a thermosensitive λ lysogen carrying prophage λ C_{I857} Sam_7, the lysogen was grown in shaking cultures at 32°C to an OD650 of 0.45. The cultures were induced by 30 mins shaking at 42°C. Replication and packaging of λ was allowed to continue for several hours at 37°C with aeration. The cells were concentrated 50 fold by centrifugation at 5,000 rpm for 5 mins at 4°C in a Sorvall GS-3 rotor followed by resuspension in phage buffer. The cells were lysed by addition of 0.05 volumes of $CHCl_3$ and the lysates were digested for 1 h at room temperature with DNase at a final concentration of 10 μg per ml. Cellular debris were removed by centrifugation at 8,000 rpm for 20 mins at 4°C in a Sorvall HB-4 rotor. The density of the supernatant was adjusted to 1.45g per ml with solid CsCl. The phage was purified by two rounds of density gradient equilibrium centrifugation using a Beckman SW41 rotor at 40,000 rpm at 10°C for 24h per round. The visible phage bands were collected by side puncture with a syringe and dialysed against TE to remove CsCl. Phage DNA

was prepared from phage particles by digestion with Proteinase K at a final concentration of 100 µg per ml at 37°C for several hours, followed by several extractions with neutral phenol and extensive dialysis against TE.

To prepare DNA from phage infections, infected cellular lysates, prepared by the PDS method of Blattner et al.(1977), were shaken at 37°C for 20 mins in the presence of 0.002 volumes of $CHCl_3$. Cellular debris was removed by centrifugation at 7,000 rpm for 10 mins at 4°C in a Sorvall GS-3 rotor. The supernatant was made 1M in NaCl and 10%(w/v) in PEG 6000 (Yamamoto et al., 1970). Following gentle stirring at 4°C overnight, the phage precipitate was pelleted by centrifugation at 8,000 rpm for 30 mins at 4°C in a Sorvall GS-3 rotor and gently resuspended in 0.05 volumes of phage buffer. Purification of phage DNA was performed as described above.

2.3.3 Mammalian genomic DNA (Gross-Bellard et al., 1973)

Tissue culture cells were lysed on the plate and incubated at 37°C overnight with 2 to 4 ml of a solution containing 50mM Tris·HCl, pH8.1, 1mM EDTA, 0.5%(w/v) SDS and 100 µg per ml of Proteinase K. The lysate was extracted with neutral phenol and dialysed against TE. Pancreatic RNase was added to a final concentration of 150 µg per ml and the solution was incubated at 37°C for 3 to 6 h. The solution was made 0.5%(w/v) in SDS and 100 µg per ml in Proteinase K and incubation was continued at 37°C overnight. Following two extractions with neutral phenol, the DNA solution was dialysed extensively against TE. In some cases the DNA was

further purified by equilibrium density gradient centrifugation in the presence of ethidium bromide as previously described (section 2.3.1).

2.3.4 SV40 DNA (Hirt, 1967)

CV-1 monolayers were infected at a MOI of 0.05. 10 to 14 days PI, when more than 80% of the cells were exhibiting cytopathic effect, the medium was removed from the cells. 1.3 ml of Hirt extraction buffer was spread over the monolayer. Following 30 mins incubation at room temperature, 0.32 ml of 5M NaCl was added to the plate lysate. After 15 mins at room temperature the extract was scraped off the plate into a centrifuge tube with a rubber policeman. Following storage at 4°C overnight, the extract was centrifuged at 15,000 rpm for 40 mins at 4°C in a Sorvall SS-34 rotor. The supernatant was retained, adjusted to be 200μg per ml in ethidium bromide and 1.56g per ml in density with solid CsCl. Equilibrium density gradient centrifugation and all subsequent steps were performed as described in section 2.3.1.

To prepare DNA from purified virions, virions were prepared as described in section 2.1.2. The virions were further purified by two rounds of equilibrium density gradient centrifugation in TBS adjusted to 1.34g per ml with CsCl. The visible viral band was collected with a syringe and DNA was prepared from the purified virions as described for the preparation of λ DNA (section 2.3.2).

2.3.5 Human adenovirus serotype 2 (Ad2) DNA

Ad2 infected KBG cells were kindly provided by Miss V. Lewis, Imperial College. The cells were disrupted by sonication and the

cellular debris removed by centrifugation at 5,000 rpm for 5 mins at 4°C in a Sorvall HB-4 rotor. Lipids were removed from the supernatant by extraction with an equal volume of Arcton. 0.5g of solid CsCl per ml of supernatant was added and the virus was purified by equilibrium density centrifugation in a Beckman SW50.1 rotor at 45,000 rpm for 24 h at 4°C. DNA was prepared from the purified virions as described (section 2.3.2).

2.3.6 D.melanogaster DNA

D.melanogaster embryonic DNA was a gift from Dr. D. deCicco, Imperial College.

2.4 RNA preparation

2.4.1 D.melanogaster rRNAs

The 18S and 28S rRNAs from the KC line of D.melanogaster tissue culture cells were prepared as described by Glover et al. (1975). Cells from a 100 ml spinner culture were collected by centrifugation at 5,000 rpm for 5 mins at 4°C in a Sorvall GS-3 rotor and the cells were resuspended in 1ml of 50mM sodium acetate (adjusted to pH5 with acetic acid), 1mM EDTA. 10 µl of diethyl pyrocarbonate was added together with 1 ml of 50mM sodium acetate, pH5, 1mM EDTA, 1%(w/v) SDS. The lysate was extracted on ice with an equal volume of phenol equilibrated with 50mM sodium acetate, pH5. Following centrifugation at 9,000 rpm for 20 mins at 4°C in a Sorvall HB-4 rotor, the aqueous phase was collected and layered onto two 37.5 ml, 10% (w/v) to 30% (w/v) sucrose gradients (sucrose solutions were made with 200 mM sodium acetate, pH5, 1 mM EDTA). The gradients were centrifuged in a SW27 rotor at 25,000 rpm for 16 h at 4°C. 0.7ml fractions were collected by puncturing the bottom of the tube and the 18S and 28S rRNA peaks were located either by Cerenkov radiation, if the RNA was labelled with ^{32}P, or by spotting 2 µl of each fraction onto a petri dish containing 1% (w/v) agarose with 1 µg per ml of ethidium bromide and visualising the RNA by UV light. The RNAs from the peak fractions were precipitated with 2 volumes of ethanol and stored at -20°C as ethanol precipitates. 18S and 28S rRNAs prepared in a similar fashion from D.melanogaster embryos were kindly provided by Dr. S. Kidd, Imperial College.

. 2.4.2 SV40 cRNA

SV40 cRNA was synthesised using E.coli RNA polymerase under conditions similar to those described by Westphal (1970). The reaction mixture contained 50mM Tris·HCl, pH7.9, 80mM NaCl, 40mM KCl, 15mM $MgCl_2$, 10mm β-mercaptoethanol, 1 to 5 mM of each ribonucleoside triphosphate and 120 µg per ml of SV40(I) DNA. Following a 1 h incubation at 37°C, the SV40 template DNA was digested with 50 µg per ml of Pancreatic DNase at room temperature for 1 h. The mixture was then extracted with phenol and the aqueous phase was subjected to Sephadex G-50 chromatography, as described in section 2.11.2, to remove oligonucleotides and unincorporated triphosphates. For labelling SV40 cRNA, ^{14}C-rATP or α-^{32}p-rATP was included in the RNA polymerase reaction mixture. Details are given for each experiment. SV40 cRNA prepared in this way is asymmetric, i.e. copied from only one strand of the DNA template.

2.5 Enzyme preparations

SstI and SalI were prepared as described by Rigby and Berg (1978). EcoRI methylase was prepared essentially as described by Greene et al. (1975). Endonuclease S_1 was prepared as described by Shenk et al. (1975).

Restriction endonucleases BstI, HaeIII and HinfI were gifts from Dr. M.S. Neuberger, Imperial College. HpaII was supplied by Dr. M. Fried, I.C.R.F. BglII and HpaI were gifts from Dr. J. Arrand, St. Mary's Hospital Medical School. XmaI was prepared by Miss G. Latos, Imperial College. EcoRI, KpnI, and TaqI were gifts from

Dr. A. Atkinson, CAMA, Porton. HindIII was prepared by Mrs. L. Woods, Imperial College. BglI and HhaI were purchased from Uniscience Ltd., Cambridge. References for all restriction endonucleases used are given in the review by Roberts (1978).

E.coli DNA polymerase I purified by the method of Jovin et al. (1969) was supplied by Dr. A. Atkinson. E.coli RNA polymerase was a gift from Dr. R. Kamen, I.C.R.F. Calf thymus terminal nucleotidyl transferase was a gift from Dr. R. Ratliff, Oak Ridge National Laboratory. Calf intestinal phosphatase, bought from Boehringer Mannheim and further purified by Sephadex G-75 chromatography (Efstratiadis et al., 1977), was provided by Dr. D. deCicco, Imperial College. T4 DNA ligase was a gift from Professor K. Murray, Edinburgh and Mrs L. Woods, Imperial College.

2.6 Restriction enzyme reaction conditions

The reaction conditions are listed below:

BstI; 6mM Tris·HCl, ph7.6, 6mM $MgCl_2$, 6mM β-mercaptoethanol (Bam buffer)

KpnI, HpaII, TaqI, and XmaI; Bam buffer

SalI; Bam buffer plus 100 μg per ml NaCl

SstI; Bam buffer plus 100 μg per ml gelatin

EcoRI; 100mM Tris-HCl, pH7.6, 10mM $MgCl_2$ (EcoRI buffer)

HaeIII, HhaI, HindIII and HinfI; EcoRI buffer plus 50mM NaCl and 1mM DTT

HpaI; 20mM Tris·HCl, pH7.4, 10mM $MgCl_2$, 2mM β-mercaptoethanol, 20mM KCl and 1 mg per ml BSA

BglI and BglII; 50mM Tris·HCl, pH7.9, 50mM KCl, 10mM $MgCl_2$, 10mM DTT

TaqI digestions were performed at 50°C, all other restriction digests were performed at 37°C. When necessary, reactions were terminated either by a 10 mins incubation at 65°C or by extraction with neutral phenol. In general, double and triple digests were performed sequentially starting with the enzyme requiring the lowest ionic strength. After completion of the first reaction, the ionic conditions were readjusted for the subsequent enzyme(s).

2.7 Gel electrophoresis

Agarose gels were cast either in TB or Tris-acetate electrophoresis buffer. The gels, containing 1µg per ml of ethidium bromide, were electrophoresed at 1.5V per cm for 15h using the electrophoresis buffer as running buffer; quite often much higher voltage gradients were used for shorter lengths of time. Polyacrylamide gels were prepared from a stock of 30% (w/v) acrylamide, 0.8% (w/v) bisacrylamide using TB electrophoresis buffer. TEMED and ammonium persulphate were used at concentrations of 0.16% (v/v) and 0.07% (w/v), respectively, in the polymerisation reaction. 0.05 volumes of gel loading buffer was added to all samples prior to loading. Following fractionation, DNA in the ethidium bromide stained gel was visualised under 375 or 254 nm UV light. A 35mm photographic system was used for photography.

2.8 In vitro labelling of DNA and RNA

2.8.1 DNA

DNA was labelled using the nick-translation activity of E.coli DNA polymeraseI essentially as described by Rigby et al. (1977).

Final reaction conditions were 50mM KPO4, pH7.4, 5mM $MgCl_2$, 15μM unlabelled deoxyribonucleoside triphosphates, 7.5μM α-^{32}p -deoxyribonucleoside triphosphates (sp. act. between 100 and 300 Ci per mmole), 15μg per ml of the DNA to be labelled, 10 ng per ml DNase I, and a preoptimised amount of DNA polymerase I. The reaction mixture was incubated at 14°C. When labelling had reached the desired level (10^7 to 2 x 10^8 cpm per μg of DNA), the reaction was terminated by the addition of EDTA to a final concentration of 25mM. Unincorporated triphosphates were removed by Sephadex G-50 chromatography (see section 2.11.2).

Terminal transferase was used to end label Charon 4 DNA. Final reaction conditions were 10mM potassium cacodylate, pH7.5, 0.1mM DTT, 20μg per ml BSA, 8mM $CoCl_2$, 20μm α-^{32}p dTTP and a saturating amount of enzyme. The reaction mixture was incubated at 37°C for 1h and terminated by phenol extraction. Unincorporated triphosphate was removed by Sephadex G-50 chromatography.

2.8.2 RNA

For labelling 18S and 28S Drosophila rRNAs, 5μg of RNA was first partially hydrolysed in 0.1 N NaOH for 1h at 0°C. 0.1 volumes of 3M sodium acetate, pH5.0, was added and the RNA was precipitated at -20°C following the addition of 2 volumes of ethanol. The RNA was resuspended in 50μl of reaction mixture containing 66mM Tris·HCl, pH7.6, 10mM $MgCl_2$, 1mM spermidine, 15mM DTT and 100μCi α-^{32}p rATP (sp. act. greater than 1000 Ci per mmole). 1 to 2 units of T4 polynucleotide kinase were added and the mixture was incubated at 37°C for 1h. EDTA was added to a final concentration of 25mM and

Sephadex G-50 chromatography was performed to remove unincorporated triphosphates (see section 2.11.2).

2.9 Transfer hybridisation and autoradiography

The transfer of DNA from agarose gels to nitrocellulose filters was performed essentially as described by Southern (1975). For some experiments DNA in the gel was first depurinated in 0.25N HCl for 30 mins (Wahl et al., 1979). All gels were denatured in 0.5N NaOH, 1.5M NaCl for 3h and then neutralised in 1M Tris-HCl, pH7.4, 3M NaCl for 1h. Nitrocellulose filters were rinsed in distilled H_2O and then immersed in 3XSSC for at least 30 mins prior to transfer. Transfer was performed with 20XSSC for 24h. Following transfer the nitrocellulose filters were rinsed in 3XSSC and then baked for at least 2h at 80°C.

The hybridisation procedure used for the filters was essentially that described by Jeffreys and Flavell (1977). All of the steps were performed at 65°C in a shaking waterbath. The nitrocellulose filters were preincubated for 3h in 3XSSC, 10XPM (Denhardt, 1966), after which denatured, sonicated salmon sperm DNA and SDS were added to final concentrations of 100 μg per ml and 0.1% (w/v) respectively and incubation was continued for 1h. The filters were blotted dry and hybridised in 3XSSC, 10XPM, 0.1% (w/v) SDS, 100 μg per ml of denatured, sonicated salmon sperm DNA and ^{32}P-labelled probe for 16h. In general nick-translated DNA probes were heat denatured and used at concentrations of between 5 and 25 ng per ml. In some cases, 9% (w/v) dextran sulphate was included in the hybridisation solution defined above (Wahl et al., 1979). Following hybridisation the filters were washed four times for 30 mins in hybridisation solution, then four to five times for 30 mins in

0.1XSSC, 0.1% SDS and finally twice for 20 mins at room termperature in 3mM Tris. The filters were baked dry at 65°C for 1h and autoradiographed at -70°C using preflashed Fuji RX X-ray film and Fuji Mach II intensifying screens as described by Laskey and Mills (1977).

2.10 Gel purification of DNA fragments

Low gelling temperature agarose (Miles Laboratories, melting temperature 65°C, gelling temperature 30°C) gels cast in Tris-acetate electrophoresis buffer, containing 1 µg per ml of ethidium bromide, were used for preparative electrophoresis. The DNA fragments of interest were visualised under 375 nm UV light and collected as gel slices. These gel slices were completely dissolved by a 3 to 4 mins incubation at 67°C and then cooled to 37°C. The agarose solutions were extracted several times with an equal volume of neutral phenol at 37°C. After 10 mins incubation on ice, the extraction mixtures were centrifuged at 9,000 rpm for 40 mins at 2°C in a Sorvall HB-4 rotor. The clear aqueous phases were carefully collected, adjusted to be 1M in NaCl and the DNA was precipitated in a dry ice-ethanol bath following the addition of 2 volumes of ethanol. The precipitates were resuspended in small volumes (50 to 100 µl) of TE and re-extracted with phenol. Following ethanol precipitation and 70% (v/v) ethanol washes the DNA was dissolved in TE and was ready to use. DNA prepared in this way is restrictable and ligatable.

2.11 Velocity sedimentation and gel filtration

2.11.1 Velocity sedimentation

Two types of sucrose gradients were used. To fractionate high molecular weight mammalian DNA from monomeric SV40 DNA,

10 ml 5% (w/v) to 20% (w/v) sucrose (made with TNE) gradients were used. These were centrifuged in a Beckman SW41 rotor at 40,000 rpm for 3 to 4.5h at 4°C. SV40 (I) DNA sediments about half way down the gradient following a 4.5h centrifugation. DNA present in the bottom tenth of the gradient, usually 0.3 to 0.5 of the total DNA, was collected.

To fractionate for 35 to 50 Kb Drosophila genomic DNA fragments following partial restriction endonuclease cleavage, 28 ml 10% (w/v) to 40% (w/v) sucrose gradients (made with 20mM Tris·HCl, pH 8.0, 10mM EDTA, 1M NaCl) were centrifuged in a SW27 rotor at 25,000 rpm for 15.5 to 16h at 20°C. Ad2 DNA (35Kb) and Charon 4 DNA (45Kb) were used as size markers in an identical parallel sucrose gradient. Under these conditions Ad2 DNA sediments about three quarters of the way down the gradient. 0.5ml fractions were collected and DNA contained in the fractions sedimenting between the peaks of Ad2 and Charon 4 DNAs was collected.

2.11.2 Gel filtration

To fractionate labelled nucleic acids from unincorporated triphosphates, 10 to 15 ml Sephadex G-50 columns approximately 15 cm in length were used. The columns were equilibrated and run with TE or TNE. 0.5 to 1 ml fractions were collected. The labelled nucleic acids contained in the excluded fractions were detected by liquid scintillation counting and pooled.

To fractionate DNA from EcoRI linkers (Uniscience Limited, Cambridge) 1.5 ml Biogel A 150 columns were poured into pasteur pipettes. The columns were equilibrated and run in TNE. 3 drop fractions were collected and DNA was detected by ethidium bromide

staining following agarose gel electrophoresis of aliquots of each fraction. Fractions containing DNA were pooled.

2.12 Reactions and methods used in DNA cloning

2.12.1 Phosphatase reaction (Efstratiadis et al., 1977)

In general, plasmid vector DNAs linearised with restriction enzymes which produce cohesive ends were treated with calf intestinal phosphatase to remove the 5' phosphate residues and thus reduce vector background (Ullrich et al., 1977). The reactions were performed in 10mM Tris·HCl, pH 8.0, 1mM EDTA at 37°C for 1 h. The reactions were terminated either by a 10 mins incubation at 65°C or by phenol extraction followed by several ethanol precipitations to remove the residual phenol.

2.12.2 Ligation reactions (Maniatis et al., 1979)

Ligation reactions for joining cohesive ends were used to produce two different types of molecule. To produce covalently closed chimaeric molecules for transformation, a 5 to 6 fold molar excess of target DNA to vector DNA was used at low DNA concentrations (5 to 50 μg per ml) in a reaction mixture containing 70mM Tris·HCl, pH7.5, 10mM $MgCl_2$, 1mM rATP, 15mM DTT and a preoptimised amount of T4 DNA ligase. The reaction mixtures were incubated overnight at 12°C.

When λ or cosmid vectors were used the high molecular weight chimaeric molecules which serve as substrate in the λ in vitro packaging reaction were generated using the above conditions with the following modifications:

(1) total DNA concentration was increased to 250 to 500 μg per ml,

(2) vector DNA was used in large molar excess,

(3) higher concentrations of T4 DNA ligase were used,

(4) the reactions were performed at room temperature for 4 to 6 h.

2.12.3 Reactions used for joining EcoRI linkers to eukaryotic DNA

The commercially available EcoRI linkers (Uniscience Limited, Cambridge) possess 5'-hydroxyl termini (Scheller et al., 1977). Therefore the linkers must be phosphorylated before they will serve as substrates for T4 DNA ligase. The polynucleotide kinase reaction conditions were identical to the ligation conditions described above. 2 units of T4 polynucleotide kinase (Uniscience) were used in a 20 µl reaction containing 5 µg of linkers. The reaction was incubated at 37°C for 30 mins. The phosphorylated linkers were stored in the reaction mixture at -20°C until they were needed.

Only flush ends can be ligated to EcoRI linkers. To generate flush ends from molecules which possessed overhanging ends endonuclease S1 was used under conditions described by Shenk et al. (1975), 30mM sodium acetate, pH4.6, 250mM NaCl, 4.5mM $ZnSO_4$. The DNA concentration in these reactions was 200 µg per ml and endonuclease S1 was used at 0.5 to 50 units per µg of DNA. Incubation was performed at 0°C for 1 h to minimise non-specific cleavage. Reactions were terminated by phenol extraction and the DNA was precipitated and washed in 70% (v/v) ethanol. The DNA was resuspended in ligation buffer (final concentration of DNA was 200 µg per ml), 0.1 volumes of phosphorylated linkers in ligation buffer were added (final concentration of linkers was 25 µg per ml) and the mixture was incubated at room temperature for 4 to 6 h in

the presence of a predetermined amount of T4 DNA ligase. 5 to 10 µl of the reaction mixture was electrophoresed on a 12% acrylamide gel. Successful linking reactions result in the formation of a ladder consisting of oligomers of the oligonucleotide linker. Successful reactions were terminated by the addition of EDTA to a final concentration of 25mM and a 10 mins incubation at 65°C. Monomeric and oligomeric linkers were removed by Biogel A 150 chromatography (see section 2.11.2). Fractions containing linked DNA were pooled, precipitated at -20°C with 70% (v/v) ethanol and stored.

2.12.4 EcoRI methylase reaction (Greene et al., 1975)

The methylation of two of the adenosine residues in the EcoRI recognition sequence by EcoRI methylase renders the EcoRI site resistant to EcoRI endonuclease cleavage (Greene et al., 1975). The reaction conditions used for the methylation reaction were 100mM Tris·HCl, pH8.0, 10mM EDTA and 1.7 to 5µM ^{3}H-SAM. The ^{3}H-SAM (15Ci per m mole, purchased from The Radiochemical Centre, Amersham) was used because of its purity; most commercially available SAM is contaminated with S-adenosyl-L-homocysteine which is a potent inhibitor of methyltransferases (Shapiro et al., 1965). The reactions were assayed either by assessing the resistance of the methylated DNA to EcoRI endonuclease cleavage or by determining the radioactivity incorporated into the substrate DNA using DEAE-81 paper in the filter binding assay previously described (Greene et al., 1975).

2.12.5 Preparation of competent cells and transformation

The protocol used to prepare competent cells was a modification of the Mandel and Higa (1970) procedure (Mr. M. Scott, personal

communication). The majority of the cells used for my experiments were provided by Mr. M. Scott and Dr. D. de Cicco, Imperial College. A saturated 50ml overnight culture was grown from a single colony of E.coli strain HB101 in L-broth. The saturated culture was diluted 1 to100 into 500ml of L-broth. The culture was grown to an OD600 of 0.6 in a shaking 37°C incubator. The cells were pelleted by centrifugation at 6,000 rpm for 10 mins at 4°C in a Sorvall GS-3 rotor. The cells were resuspended in 250ml of ice cold 10mM Pipes, adjusted to pH6.8 with NaOH, 10mM $RbCl_2$, pelleted as before and resuspended once more in ice cold 10mM Pipes, pH6.8, 10mM $RbCl_2$, 75mM $CaCl_2$. The cellular suspension was left on ice for 30 mins, pelleted as before and resuspended gently in 50ml of ice cold 10mM Pipes, pH6.8, 10mM $RbCl_2$, 75mM $CaCl_2$, 15% (v/v) glycerol. The suspension was aliquoted into 1.5ml microfuge tubes, frozen in liquid N_2 and stored at -70°C. These cells remain competent for at least 3 months and give frequencies of greater than 10^6 transformants per μg of supercoiled plasmid DNA.

For transformation reactions, frozen competent cells were thawed on ice for 20 mins. DNA solution, usually from a ligation reaction mixture, was added such that the volume of the DNA solution was less than one half the volume of the competent cells used and the mass of the ligated DNA added was less than 100 ng per 50 μl of competent cells (100ng per 50 μl of cells is saturating). The mixture was incubated on ice for 20 mins and then at 25°C for 10 mins. Three times the volume of L-broth was added and incubation was continued for 50 mins at 37°C to allow the expression of the antibiotic resistance gene(s). Various dilutions of the transformation

mixture were then plated on L-agar plus antibiotics (tetracycline and ampicillin were used at 100 μg per ml).

2.12.6 λ *in vitro* packaging

The protocol used is essentially Blattner's adaptation of the procedure described by Sternberg *et al.* (1977) and is described below. However, by using the strains BHB 2688 and BHB 2690, the apparent requirement for λ A protein (Becker and Gold, 1977) for high efficiencies of packaging is obviated. In my hands, this procedure produces packaging efficiencies that are about ten times higher than those obtained with the Hohn and Murrary (1977) protocol. I wish to acknowledge Dr. D. Ish-Horowicz for making me aware of the advantages of using this protocol.

(1) Preparation of freeze thaw lysate (FTL)

Three 500ml cultures of BHB 2688 were grown in L-broth at 30°C in 2 l flasks with vigorous aeration. At OD600 = 0.3, the cultures were induced by a 15 mins incubation in a 45°C waterbath. The cultures were grown for another hour at 38°C with vigorous aeration. The cells were then pelleted by centrifugation at 9,000 rpm for 10 mins at 4°C in a Sorvall GS-3 rotor. Care was taken to remove all of the supernatant. Each pellet, containing cells from 250ml of culture, was resuspended in 0.5 ml of 10%(w/v) sucrose, 50mM Tris·HCl, pH7.5 using a 1 ml disposable plastic pipette. The resuspended pellets were pooled into ~ 35ml plastic Sorvall tubes. 75 μl of a fresh lysozyme solution (2mg per ml in 0.25M Tris·HCl, pH7.5) was added to the tube and it was gently mixed. The cellular suspensions were frozen in liquid N_2 and stored at -70°C

overnight. The frozen cells were thawed on ice and 75 µl of Buffer M1 was gently mixed into each tube. The tubes were centrifuged at 1700 rpm for 1h at 4°C in a Sorvall HB-4 rotor. The supernatant was aliquoted into precooled 1.5 ml microfuge tubes, frozen in liquid N_2 and stored at -70°C.

(2) Preparation of sonicated extract (SE)

One 500 ml culture of BHB 2690 was grown and induced exactly as described above. The cells were pelleted by centrifugation at 9,000 rpm for 10 mins at 4°C in a Sorvall GS-3 rotor. Each pellet (equivalent to 250 ml of culture) was resuspended in 1.8ml of Buffer A. The pellets were pooled into a 10 ml Sterilin tube, the tube was immersed in an ice bath and the cells were sonicated, without foaming, using a Dawes sonicator in 5 second bursts until the solution was no longer viscous. Cellular debris was removed by centrifugation at 6,000 rpm for 10 mins at 4°C in a Sorvall HB-4 rotor. The supernatant was aliquoted into precooled 1.5 ml microfuge tubes, frozen in liquid N_2 and stored at -70°C.

(3) Packaging reaction and infection

A typical packaging reaction consisted of 7 µl of Buffer A, 1 to 2 µl of DNA (usually 200 to 500 µg per ml from a ligation mixture), 1 µl of Buffer M1, 6 µl of SE and 10 µl of FTL. The precise amount and ratio of SE and FTL used were optimised for each preparation. The reaction mixtures were incubated at 25°C for 1h. Between 100 µl and 500 µl of phage buffer together with two to three drops of $CHCl_3$

were then added and each reaction mixture was vigorously shaken. Debris was pelleted by a 2 mins centrifugation in a microfuge. The supernatants were collected and usually used for infections within 2 to 3 days. However, they can be stored at 4°C with little loss in infectivity.

To infect, fresh overnight cultures of E.coli strains HB101 or LE392 were used. Up to an equal volume of packaged DNA in phage buffer was added to the cells. The infection mixtures were left at room temperature for 10 mins before plating. Where plaques were expected, 3 ml of top agar were added to 50 μl to 500 μl of the infection mixture before plating onto L-agar. Where colonies were expected, between 50 μl to 500 μl of the infection mixture were plated by direct spreading onto L-agar plus 100 μg per ml of ampicillin.

2.13 Colony hybridisation and high density screening

Two types of colony screen were used. Where only a small number of colonies needed testing (i.e. in subcloning experiments), the Grunstein and Hogness (1975) filter screening procedure was used. Where large numbers of recombinants required testing (i.e. for screening a eukaryotic genomic library), the high density screening method of Hanahan and Meselson (1980) was used.

2.13.1 Grunstein-Hogness colony screen

(1) Colonies to be tested were stabbed into two L-agar plus antibiotic plates in an ordered and identical array. The plates were grown overnight at 33°C. One of the plates was stored at 4°C and served as the master plate while a nitrocellulose

filter was placed over the colonies on the other plate. The colonies stick to the filter when it is peeled off the plate. Steps (2) to (6) were performed using blotting paper saturated with the specified solutions.

(2) The filter was placed, colony side up, over 0.5N NaOH for 7 mins.

(3) The filter was transferred to 1M Tris-HCl, pH7.4, for 2 mins.

(4)Step (3) was repeated using freshly saturated blotting paper.

(5) The filter was transferred to 1M Tris·HCl, pH7.4, 1.5M NaCl for 3 mins.

(6) Step (5) was repeated using freshly saturated blotting paper.

(7) The filter was rinsed in 3X SSC and blotted dry.

(8) The filter was baked for 2h at 80°C.

(9) The filter was hybridised and autoradiographed exactly as described for the Southern transfer filters.

2.13.2 High density screen

(1) Plating and growth

(a) A sterile nitrocellulose filter was wetted on a L-agar plus antibiotic plate.

(b) The filter was peeled off the plate and placed onto a fresh L-agar plus antibiotic plate.

(c) Cells (usually from an infection or from a frozen pool of recombinants) were applied to the filter. Up to 500 µl and 25,000 colonies can be applied to a single 82mm diameter filter.

(d) The cells were spread with a glass spreader made from a pasteur pipette.

(e) The filter was incubated until very small (0.1mm diameter), distinct colonies appeared (32°C overnight is usually appropriate). This was used as the template filter for replica plating.

(2) Replica plating

(a) The template filter was peeled off the plate and placed, colony side up, over several sheets of sterile blotting paper.

(b) A wetted nitrocellulose filter was placed over the template filter.

(c) A few sheets of sterile blotting paper were placed over the two nitrocellulose filters.

(d) The two filters were pressed together. Two glass plates were used for this purpose.

(e) The filters were keyed to each other using a syringe needle.

(f) The filters were peeled apart and placed, colony side up, onto fresh L-agar plus antibiotic plates.

(g) The template filter was usually used as the master. It was stored at 4°C. Note that a large number of replica filters can be made off the same template by repeating the above process.

(h) The replica filter was grown at 37°C for 4 to 6 h until small distinct colonies appear.

(i) The filter was transferred to L-agar plus 500 μg per ml of chloramphenicol, incubated at 37°C overnight and

denatured, hybridised and autoradiographed exactly as described in the Grunstein-Hogness procedure.

(j) Positive areas were picked from the master and rescreened at lower density to obtain pure clones.

(3) Freezing master filters

Colonies on master filters can be stored indefinitely at -70°C. For this purpose the following procedure was used:

(a) A sterile nitrocellulose filter was placed on a F-agar plus antibiotic plate. Cells were spread over the filter and incubated as described.

(b) Replica plating was performed as described using a sterile filter which had been pre-wetted on F-agar.

(c) The two nitrocellulose filters were sandwiched between several sheets of sterile blotting paper following replica plating.

(d) Several damp sheets of sterile blotting paper were placed on top of the sandwich and the entire sandwich was wrapped with sterile aluminium foil, sealed inside a sterilin bag and stored at -70°C.

(e) When required, the filters were thawed, peeled apart, placed onto fresh plates, grown until small distinct colonies appeared and then processed as described above.

2.14 Conditions for mercuration of nucleic acids

The direct covalent mercuration of D.melanogaster 18S and 28S rRNAs was performed essentially as described by Dale et al. (1975). The RNA was dialysed extensively against 50mM sodium acetate, pH6.0 before use. The final conditions for the reaction were 50mM sodium acetate, pH6.0, 4 to 12 mM mercuric acetate (when required,

^{203}Hg acetate was used at a specific activity of 3 x 10^7 cpm per μmole of Hg) and approximately 100 μg per ml of RNA. Incubations were performed at 50°C for 0.5 to 2h. Reactions were terminated by addition of 0.4 volumes of 10mM Tris·HCl, pH 7, 1M NaCl, 100mM EDTA and the reaction mixtures were dialysed extensively against TE. 2 x 10^{-5} M β-mercaptoethanol was added to the final dialysis solution. The RNA preparations were stored at -20°C.

Mercurated SV40 cRNA was made by using 5-mercuriuridine triphosphate (PL Biochemicals) as a substrate in the E.coli RNA polymerase reaction as described by Dale et al. (1973).

2.15 Thiol sepharose affinity chromatography (Dale and Ward, 1975)

Activated Thiol Sepharose 4B resin (Pharmacia) was swollen overnight in TNE plus 50mM DTT. 1 to 2 ml columns were poured using pasteur pipettes and the columns were washed extensively with TNE. Samples were applied to the columns and allowed to adsorb to the resin for 10 to 30 mins. The columns were washed with 5 to 10 ml of TNE and eluted with 5 to 10 ml of TNE plus 100mM B-mercaptoethanol. 1 ml fractions were usually collected.

The amount of sulphydryl groups on the commercial resin was 1.24 μmole per ml of resin bed volume. This was determined using the DTNB [5,5'-dithiobis (2-nitrobenzoic acid)] test described by Ellman (1959).

2.16 Hybridisation conditions

2.16.1 Final conditions for RNA-DNA hybridisations in 50% (v/v) formamide were:

50% (v/v) formamide, 100mM Pipes, adjusted to pH 7.4 with NaOH, 300mM NaCl, 1mM EDTA and 0.1% (w/v) SDS. Incubations were performed at 40°C unless otherwise specified.

2.16.2 R-looping (Thomas et al., 1976; White and Hogness, 1977) conditions were:

70% (v/v) formamide, 83mM Pipes, pH7.8, 10mM EDTA.

Temperatures of incubation and nucleic acid concentrations are specified for each experiment.

2.16.3 Final conditions used for experiments designed to allow only RNA-DNA hybridisation with no DNA-DNA hybridisation (Casey and Davidson, 1977) were:

75% (v/v) formamide, 100mM Pipes, pH7.8, 1mM EDTA, 0.1% (w/v) SDS and 300 mM NaCl.

Temperatures of incubation and nucleic acid concentrations are specified for each experiment.

2.16.4 Hybridisation of labelled RNA to DNA immobolised on nitrocellulose discs (G.M. Rubin, personal communication).

DNA samples were added to 1.6ml of TE. Following addition of 1.6ml of 1N NaOH the samples were left at room temperature for 1h. The samples were cooled on ice and 6.4ml of ice cold neutralisation buffer containing 150mM Tris, 2.5M NaCl and 0.4 N HCl were added. DNA in the samples was bound to nitrocellulose discs by slow filtration (approximately 1ml per min). The discs were washed with 6X SSC and baked at 80°C for 2 to 4h. The filters were hybridised in 50% (v/v) formamide, 5X SSC plus labelled RNA at 50°C overnight. Labelled asymmetric RNA was used at greater than 10 fold excess over the total amount of homologous DNA bound.

2.17 Length standards

Agarose gels ranging in concentrationfrom 0.4% (w/v) to 2% (w/v) were used to electrophoretically fractionate DNA fragments

generated from restriction endonuclease digestion. The fragments generated by HindIII, HaeIII, or HinfI endonuclease digestion of SV40 DNA were used as size markers for smaller DNA fragments. Determination of the size of these markers was based on the SV40 DNA sequence (Fiers et al., 1978; Reddy et al., 1978). The fragment sizes given for an EcoRI plus BstI digest of phage λ DNA (Haggerty and Schleif, 1976), HindIII, EcoRI or EcoRI plus HindIII digests of λ DNA (Philippsen et al., 1978) and an EcoRI digest of Ad2 DNA (Lania et al., 1979) were used as size standards for larger fragments.

Open circular monomeric SV40 DNA, generated from supercoiled SV40 DNA by X-ray irradiation, was used as a double stranded DNA length marker for electron microscopy.

2.18 Electron microscopy

The methods described by Davis et al. (1971) were used. Photography was performed using 35mm film in either a Philips EM301 or a Hitachi H-600 microscope. Length measurements were taken by projecting 35mm micrographs with a Leitz Wetzlar projector and tracing the contours of the molecules with a Hewlett-Packard digitizer in line with a 9821 computer.

Chapter 3

Mercurated Polynucleotides as Tools for Gene Purification

3.1 Introduction

The size of the mouse genome is approximately 3 x 10^6Kb. Therefore, if one wants to obtain an *in vitro* recombinant containing 5 Kb of single copy mouse DNA, one has to generate and screen at least 10^6 independently derived recombinants. In 1976 the available methods for generating *in vitro* recombinants allowed efficiencies of approximately 10^3 recombinants per μg of DNA. This low cloning efficiency meant that cloning for a specific single copy DNA sequence from a genome of the complexity of the mouse was impracticable. The obvious way of circumventing this problem was to enrich for duplex DNA molecules containing the sequences one wished to clone. This was the approach taken by a number of investigators. Tilghman *et al.* (1977) were able to clone one of the mouse globin genes by first enriching for this gene by using preparative agarose gel electrophoresis in combination with RPC-5 chromatography. Similarly, Tonegawa *et al.* (1977) were able to clone one of the immunoglobulin light chain variable genes by using preparative agarose gel electrophoresis in combination with density gradient centrifugation of R-loops (see below) formed between this immunoglobulin gene and its purified mRNA. While preparative gel electrophoresis, density fractionation of R-looped DNA and RPC-5 chromatography were all useful, none of these methods give very high degrees of purification. In the experiments that I have done using RPC-5 chromatography most of the endonuclease cleaved genomic DNA fractionated in the main peak. If

the DNA sequence of interest elutes with the main peak, the degree of purification obtained is small.

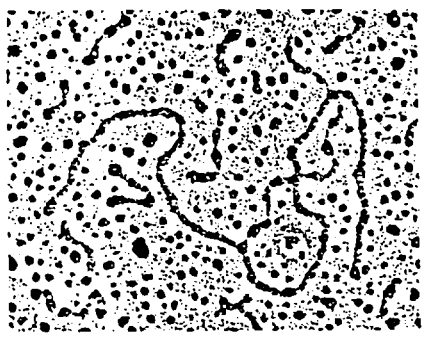
Two findings gave rise to a new idea of how to do the enrichment. Ward and his colleagues (Dale et al., 1973; Dale and Ward, 1975; Dale et al., 1975) showed that polynucleotides can be mercurated in vitro both chemically and by enzymatic incorporation of mercurated precursors. Furthermore, they showed that these mercurated polynucleotides have high affinities for sulphydryl groups coupled to a solid support. White and Hogness (1977) and Thomas et al. (1976) showed that under conditions of high formamide, duplex DNA molecules which contain sequences homologous to a given single stranded RNA will hybridise to that RNA while maintaining the association between the two DNA strands; the region of the DNA molecule which is homologous to the RNA forms an R-loop in which one strand of the DNA pairs with the RNA and the other DNA strand remains single stranded; the regions of the DNA molecule which do not have homology with the RNA remain double stranded. An example of the structures (R-loops) formed as a result of R-looping is shown in Fig.3.1. The idea is, therefore, to R-loop duplex genomic DNA fragments with a specific mercurated RNA and enrich for the DNA fragments containing homology to the RNA by affinity chromatography on thiol-sepharose resin. The work presented in this section is reconstruction experiments designed to assess the feasibility of this idea. SV40 DNA - SV40 cRNA and cDm103 DNA [a cloned segment of D.melanogaster DNA containing a complete rDNA repeat (Glover and Hogness, 1977)] - D.melanogaster 18S and 28S rRNAs were used as model systems.

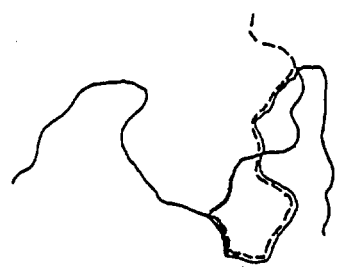
Legend to Figure 3.1

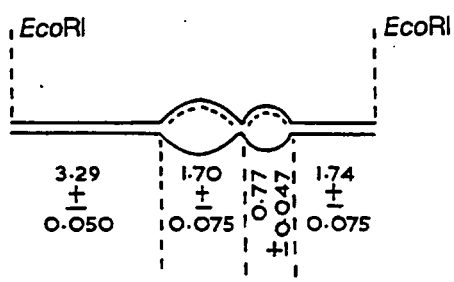
The structure of a R-loop

This figure is taken from Glover (1977). The photograph of a R-loop formed between Drosophila 28S rRNA and a cloned segment of Drosophila rDNA, Dm207, is shown on the left. An interpretative drawing of the photographed structure is shown on the right where the discontinuous line represent RNA and the continuous line represent DNA. A diagram of the structure of the R-looped molecule is also shown.

Fig.3.1







Dm207

3.2 Mercuration of RNA

SV40 cRNA was synthesised by in vitro transcription as described (Section 2.4.2). ^{14}C-rATP was used at a sp. act. of approximately 10^6 cpm per μmole. 5-mercuriuridine triphosphate (Hg-UTP) and non-mercurated UTP were included at various ratios in the reactions. The various SV40 cRNA preparations were tested by thiol-sepharose chromatography (see section 2.15) and the results are shown in Fig 3.2. As expected, the RNAs containing 7.5%, 15% and 30% mercurated nucleotides (the percentages are calculated on the assumptions that Hg-UTP and UTP are incorporated with equal efficiency and that SV40 cRNA contains 30% U residues) were bound to the columns in the absence of sulphydryl groups and could be eluted from the columns with β-mercaptoethanol; unmercurated SV40 cRNA did not bind to the column, indicating that it was the incorporated Hg-UTP which was responsible for the binding. In every case, essentially all of the ^{14}C-SV40 cRNA loaded onto the columns was recovered.

D.melanogaster 18S and 28S rRNAs were chemically mercurated as described (section 2.14). Three incubation mixtures were set up as follows: (1) 100 μg per ml of unlabelled rRNA and 12mM mercuric acetate, (2) 100 μg per ml of unlabelled rRNA and 12mM ^{203}Hg acetate (sp. act. = 9.2 x 10^3 cpm per n mole of Hg), (3) 100 μg per ml of ^{32}P-labelled rRNA (sp. act. = 8 x 10^3 cpm per μg by Cerenkov radiation) and 12mM mercuric acetate. Incubation was performed at 50°C for 1h. Following termination of the reactions, the RNAs were dialysed extensively (6 to 7 changes of dialysis buffer) against TNE. The incorporation of

Legend to Figure 3.2

Thiol-sepharose chromatography of mercurated and unmercurated SV40 cRNAs

SV40 cRNAs were synthesised by *in vitro* transcription with *E.coli* RNA polymerase using the buffering conditions described (section 2.4.2). For the unmercurated reaction:

all four ribonucleosidetriphosphates were used at 1mM and ^{14}C-rATP was used at a specific activity of 1.1 x 10^6 cpm per μmole.

For the mercurated reactions:

rCTP, rGTP and ^{14}C-rATP were used at 2.0mM. The specific activities of ^{14}C-rATP used were between 5 x 10^5 to 10^6 cpm per μmole.

For the 7.5% mercurated reaction: [rUTP] = 1.5mM and [Hg rUTP] = 0.5mM

For the 15% mercurated reaction: [rUTP] = 1mM and [Hg rUTP] = 1mM.

For the 30% mercurated reaction: [Hg rUTP] = 2mM.

The 0.5ml reaction mixtures were incubated at 37°C for 1h. Following the incubation, 5μl aliquots of each reaction mixture were either spotted or acid precipitated onto glass-fibre discs, dried and counted. The amounts synthesised and the specific activities of the SV40 cRNAs were calculated based on the percentage incorporation and the specific activity of the ^{14}C-rATP used: unmercurated SV40 cRNA = 935cpm per μg; 7.5% Hg SV40 cRNA = 755cpm per μg; 15% Hg SV40 cRNA = 715cpm per μg; 30% Hg SV40 cRNA = 860cpm per μg. The reaction mixtures were digested with DNase, extracted with phenol and chromatographed on Sephadex G-50 columns as described (see sections 2.4.2 and

2.11.2). 25µl to 50µl of each RNA preparation containing between 1×10^3 and 1.5×10^3 cpm were subjected to thiol-sepharose chromatography on 1 ml (bedvolume) columns. 1 to 2 ml fractions were collected, acid precipitated and counted. Fractions collected prior to the point indicated by the arrow were flow throughs resulting from washing the columns with TNE. Fractions collected at and subsequent to the arrow were eluant fractions from eluting the columns with TNE plus 100mM B-mercaptoethanol.

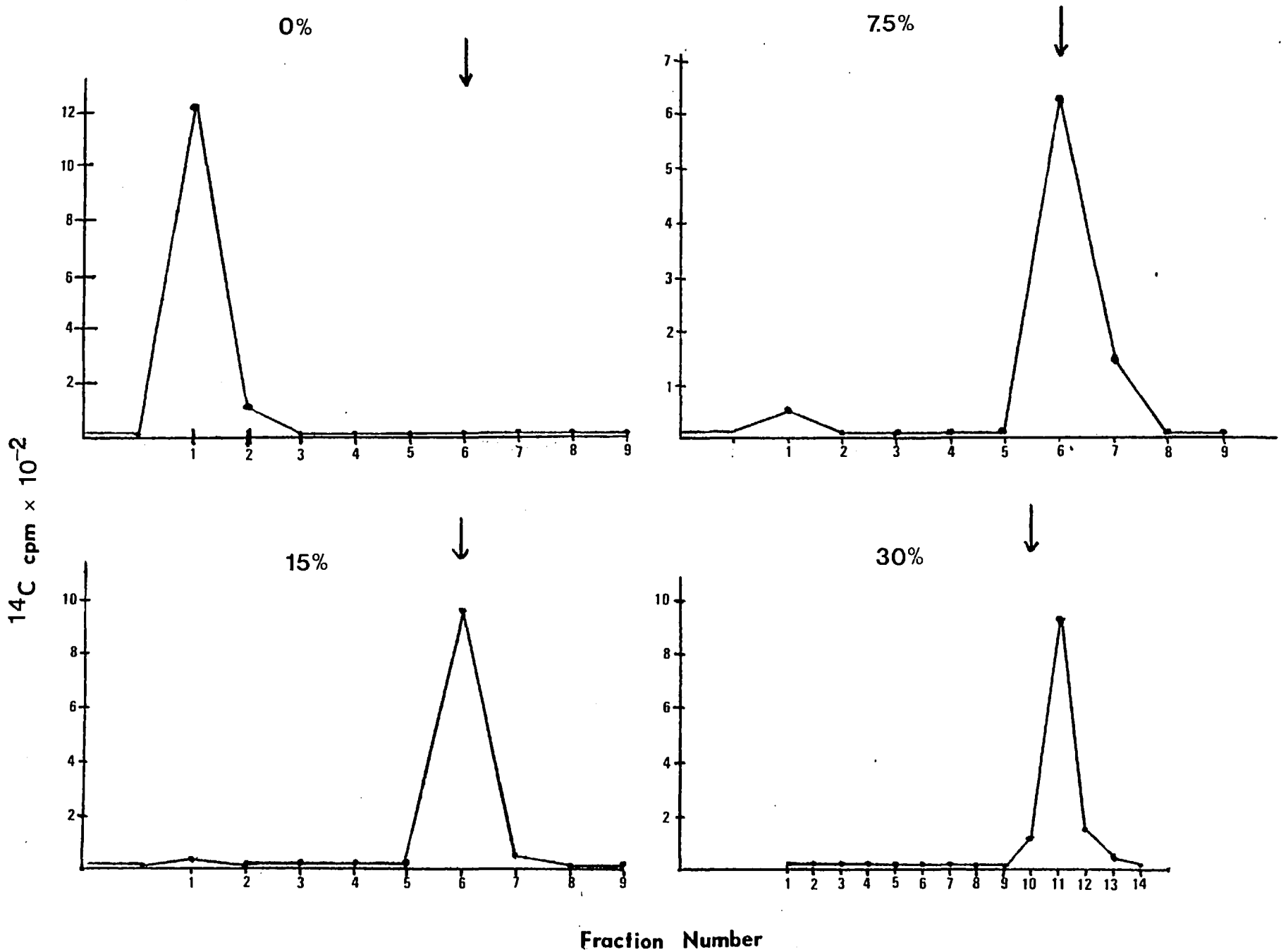


Fig.3.2

^{203}Hg into rRNA (incubation (2), see Fig.3.3), indicates that the above conditions result in 7% mercuration of the ribonucleotide residues. Thiol-sepharose chromatography was performed using the mercurated, ^{32}P-labelled rRNA from incubation (3). The results, shown in Fig.3.3, indicate that essentially all of this chemically mercurated rRNA was bound to the column.

These experiments show that: (1) RNAs can be mercurated either by direct chemical mercuration or by in vitro transcription using Hg-UTP as a substrate in the reaction: (2) these mercurated RNA species will bind to thiol-sepharose in the absence of β-mercaptoethanol and can be eluted from the resin using β-mercaptoethanol; (3) the degree of mercuration does not significantly affect the amount of mercurated RNA bound by thiol-sepharose (i.e. 7% mercuration is sufficient to bind essentially all of the RNA to the resin).

3.3 Hybridisation kinetics of Hg-SV40 cRNA

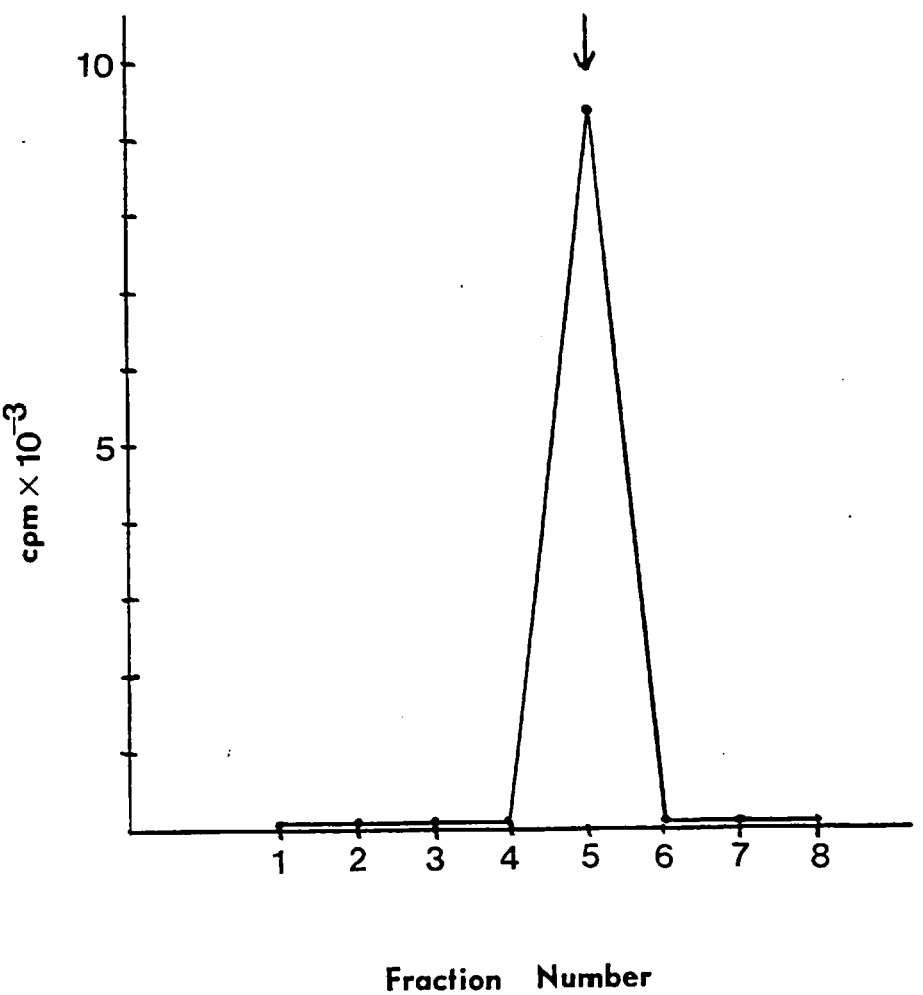
Since the purpose of making mercurated RNAs was to use them as affinity probes, it was important to investigate how these Hg-RNAs behave in hybridisation reactions. The strategy was to hybridise ^{32}P-labelled SV40 DNA of high specific activity, at very low concentrations, to a large excess of SV40 cRNA so that within the time span of the experiment no significant DNA-DNA hybridisation occurs and the only hybridisation occuring is the RNA driven RNA-DNA hybridisation. The calculations for the kinetics are shown below:

Legend to Figure 3.3

Thiol-sepharose chromatography of chemically mercurated rRNA

^{32}P-labelled and unlabelled 18S and 28S rRNA from Drosophila tissue culture cells were prepared as described (see section 2.4.1). These RNAs were dialysed against 50mM sodium acetate, pH6.0 and chemically mercurated in three 0.5ml reaction mixtures containing: (1) 100μg per ml of unlabelled rRNAs in 12mM mercuric acetate and 50mM sodium acetate, pH6.0; (2) 100μg per ml of unlabelled rRNAs in 12mM ^{203}Hg acetate (specific activity = 9.2 x 10^3 cpm per nmole) and 50mM sodium acetate, pH6.0; (3) 100μg per ml of ^{32}P-labelled rRNAs (specific activity = 8 x 10^3 cpm per μg by Cerenkov radiation) in 12mM mercuric acetate and 50mM sodium acetate, pH6.0. The reaction mixtures were incubated at 50°C for 1h and then dialysed extensively against TNE at 4°C. Incubation (2) was used to assess the degree of mercuration under these conditions. The specific activity of the RNA was 1.4 x 10^3 cpm (^{203}Hg) per μg which indicated that 7% of the ribonucleotides in the rRNAs were mercurated. RNA from incubation (3) was subjected to thiol-sepharose chromatography. 20μl of the RNA solution (approximately 10^4 cpm by Cerenkov radiation) was loaded onto a 1ml (bedvolume) column. The column was washed with 8ml of TNE before being eluted with 8ml of TNE plus 100mM β-mercaptoethanol. 2ml fractions were collected and counted by Cerenkov radiation.

Fig. 3.3



$-dC/dt = k_R C_E R + k_D C_E C_L$

C = total concentration of single stranded DNA = $C_E + C_L$

C_E = concentration of unhybridised early strand of SV40 DNA

C_L = concentration of unhybridised late strand of SV40 DNA

k_R = 2nd order rate constant for RNA-DNA hybridisation

R = concentration of unhybridised SV40 cRNA

k_D = 2nd order rate constant for DNA-DNA hybridisation

if DNA concentration is low such that the amount of DNA-DNA hybridisation is negligible, then

$-dC/dt = k_R C_E R$

If $[RNA] \gg [DNA]$

then R = Ro

where Ro = initial SV40 cRNA concentration

$-dC/dt = k_R C_E Ro$

$dC/dt = dC_L/dt + dC_E/dt$

because SV40 cRNA is asymmetric, i.e. will hybridise only to the early strand of SV40 DNA,

C_L is constant = 0.5 Co

where Co = initial total DNA concentration

$-dC/dt = -dC_E/dt = k_R Ro C_E$

$$-\int dC_E/dt = \int k_R Ro\, dt$$

$$-\ln C_E \Big|_{C_E \text{ at time } = 0}^{C_E \text{ at time } = t} = k_R Ro \Big|_0^t$$

C_E at time = t = C_E

C_E at time = 0 = 0.5 Co

$$-(\ln C_E - \ln [0.5\ Co]) = k_R\ Rot$$

$$\ln ([0.5\ Co]/C_E) = k_R\ Rot$$

because endonuclease S1 assay is used to quantitate the amount of duplex DNA (see below),

0.5 Co = 0.5 X total DNA counts

C_E = ([0.5 X total DNA counts] - [S1 resistant DNA counts])

$$\ln \left[\frac{0.5 \times \text{total DNA counts}}{(0.5 \times \text{total DNA counts}) - (\text{S1 resistant DNA counts})} \right] = k_R\ Rot$$

So, k_R is the slope of the line obtained by plotting the left-hand side of the equation against Rot

Endonuclease S1 was used to assay for the amount of duplex DNA (see Fig.3.4). To determine how much enzyme to use, ^{32}P-labelled nick translated SV40 DNA was denatured in hybridisation buffer. An endonuclease S1 titration experiment was performed by taking aliquots of the mixture, diluting them 10 fold into S1 buffer and digesting with endonuclease S1. The amount of enzyme used for assaying the hybridisation time points was sufficient to render approximately 95% of the ^{32}P-labelled, single stranded DNA into acid soluble form.

In the experiment shown in Fig.3.4, five incubation mixtures were set up. Four of the mixtures were hybridisation reactions between denatured, ^{32}P-labelled nick translated SV40 DNA and unmercurated, 7.5% mercurated, 15% mercurated and 30% mercurated SV40 cRNAs. In all four mixtures, the DNA concentration was 0.3 ng per ml; the RNA concentration was 400 ng per ml for the mercurated cRNAs and 150 ng per ml for the unmercurated cRNA; so there is a large

Legend to Figure 3.4

Hybridisation kinetics of mercurated and unmercurated SV40 cRNAs

The final conditions for the five 2ml incubation mixture were 100mM Pipes, pH7.7, 1mM EDTA, 300mM NaCl, 0.1% (w/v) SDS and 50% (v/v) formamide. The mixtures contained: 1, 0.31ng per ml of SV40 DNA; 2, 0.31ng per ml of SV40 DNA plus 150ng per ml of unmercurated SV40 cRNA; 3, 0.31ng per ml of SV40 DNA plus 400ng per ml of 7.5% mercurated SV40 cRNA; 4, 0.31ng per ml of SV40 DNA plus 400ng per ml of 15% mercurated SV40 cRNA; 5, 0.31ng per ml of SV40 DNA plus 400ng per ml of 30% mercurated SV40 cRNA. The SV40 DNA used was labelled with ^{32}P by nick translation and had a specific activity of 10^8 cpm per μg. The SV40 cRNAs used were described in Fig.3.2. DNA in the mixtures was denatured by a 2 mins incubation at 100°C. The mixtures were cooled on ice before the addition of SV40 cRNAs. Hybridisation was performed at 40°C. At 30 mins intervals, starting at time = 0, 100μl aliquots (containing approximately 2000 cpm) were withdrawn, diluted into 1ml of S_1 buffer and digested at 37°C for 1h with a predetermined amount of endonuclease S_1 sufficient to render approximately 95% of denatured, nick translated SV40 DNA acid soluble. The digested samples were acid precipitated and counted with the window of the liquid scintillation counter arranged such that ^{14}C counts were excluded. The value obtained for the 0 timepoint (the 5% S_1 background) was subtracted from the value of all subsequent timepoints. In addition, one 100μl aliquot from each incubation mixture was acid precipitated and counted to determine the total

^{32}P content prior to endonuclease S1 digestion. The data were treated as described in the text.

A gives the result from incubation 1 (DNA alone)

B gives the result from incubations 2 to 5

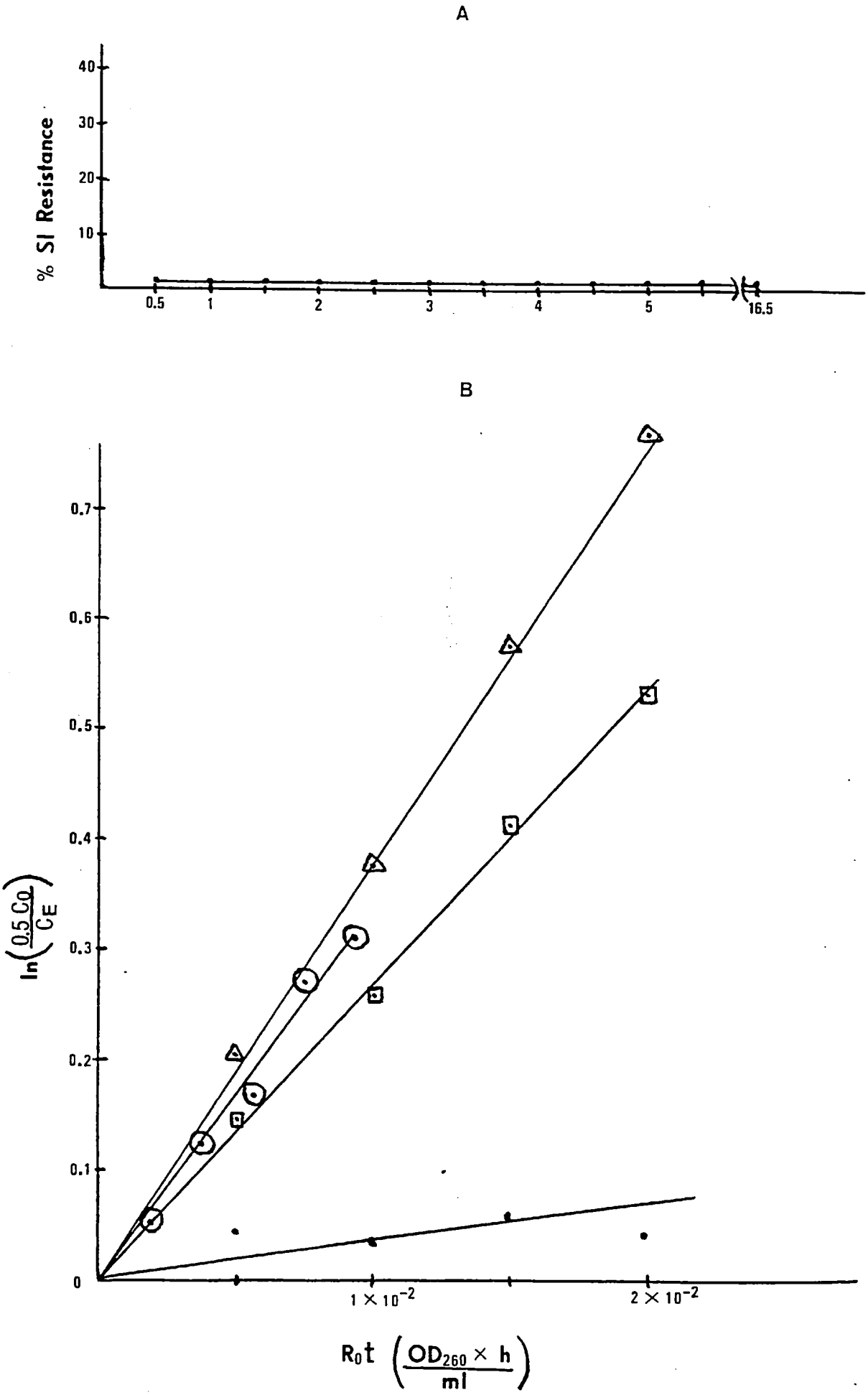
⊙ - Unmercurated (incubation 2)

△ - 7.5% mercurated (incubation 3)

⊡ - 15% mercurated (incubation 4)

• - 30% mercurated (incubation 5)

Fig.3.4


A
% SI Resistance
40
30
20
10
0.5
1
2
3
4
5
16.5
B
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
$\ln\left(\frac{0.5\,C_0}{C_E}\right)$
1×10^{-2}
2×10^{-2}
$R_0 t \left(\frac{OD_{260} \times h}{ml}\right)$

excess of RNA in all four reactions. The fifth hybridisation mixture was a control containing 0.3 ng per ml of DNA only. The data were treated as described above. As can be seen (Fig 3.4A), no DNA reassociation was observed for the reaction containing only DNA. Therefore, the reassociation observed for incubations (1) to (4) was indeed due to RNA-DNA hybridisation. The rate constants obtained for the various cRNA species indicate that 7.5% mercurated cRNA hybridises essentially as well as the unmercurated cRNA; the 15% mercurated cRNA does not hybridise as well as the unmercurated species but well enough to use as an affinity probe; the 30% mercurated cRNA does not hybridise satisfactorily (Fig.3.4B).

Dale and Ward (1975) found that poly Hg-U, the homopolymer of 5-mercuriuridine, will reanneal with poly (rA) or poly (dA) only if a two-fold molar excess of β-mercaptoethanol (or some other mercaptan) is provided as a mercury ligand. This possibility was investigated in the experiment shown in Fig.3.5. In this experiment, an excess of 15% mercurated SV40 cRNA was hybridised to nick translated SV40 DNA with or without a three fold molar excess of β-mercaptoethanol. A similar reaction using unmercurated SV40 cRNA served as control. The results indicate that the mercaptan has no effect on the k_R of the 15% mercurated SV40 cRNA. However, this result does not necessarily conflict with the result obtained by the above investigators. The Hg-S bond has an association constant of approximately 10^{16}; it is essentially covalent. Since the transcription reaction used to synethesise the

Legend to Figure 3.5

The effect of B-mercaptoethanol on the hybridisation kinetics of 15% mercurated SV40 cRNA

The conditions and materials used have been described in Fig.3.4. All three 1ml incubations contained 0.61ng per ml of SV40 DNA labelled by nick translation. In addition: incubation 1 contained 150ng per ml of unmercurated SV40 cRNA; incubation 2 contained 200ng per ml of 15% mercurated SV40 cRNA plus B-mercaptoethanol [a]; incubation 3 contained 200ng per ml of 15% mercurated SV40 cRNA without B-mercaptoethanol. The mixtures were treated as described in Fig.3.

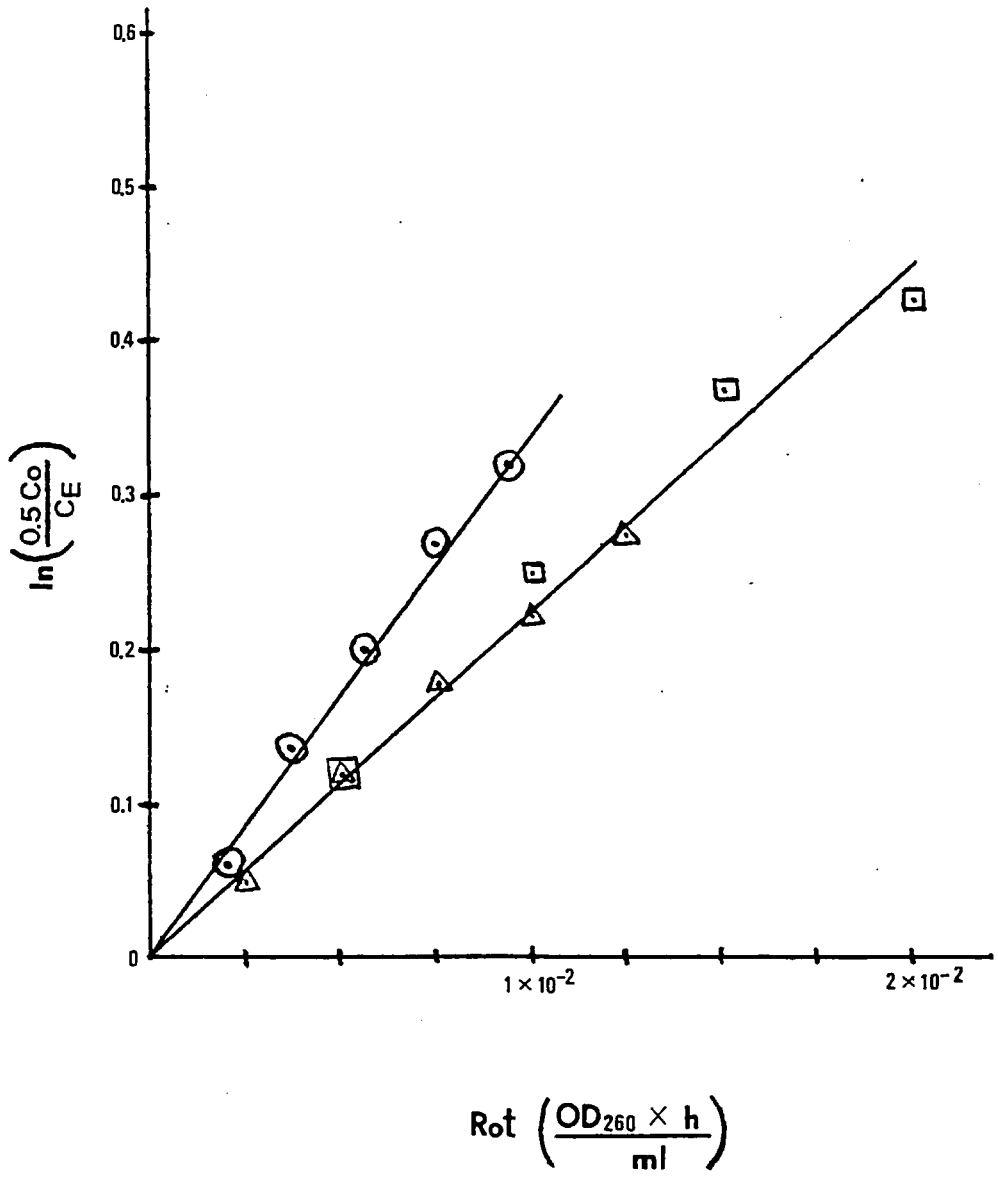
[a] 4 µg of 15% mercurated SV40 cRNA was premixed with 5 nmoles of B-mercaptoethanol before 5 µl of this 100 µl mixture was used in incubation 2.

⊙ - unmercurated, incubation 1

△ - 15% mercurated plus B-mercaptoethanol, incubation 2

⊡ - 15% mercurated without B-mercaptoethanol, incubation 3

Fig.3.5


0.6
0.5
0.4
0.3
0.2
0.1
0
$\ln\left(\frac{0.5\,C_o}{C_E}\right)$
1×10^{-2}
2×10^{-2}
$\text{Rot}\left(\frac{OD_{260}\times h}{ml}\right)$

SV40 cRNA was performed in the presence of an excess of B-mercaptoethanol, it seems reasonable to suggest that all of the mercury residues present in the RNA should already be associated with the mercaptan; therefore, addition of more mercaptan prior to the reannealing reaction should not alter the structure of the Hg-SV40 cRNA.

These experiments indicate that mercurated RNAs, at least the 7.5% and 15% mercurated SV40 cRNA, exhibit fairly normal hybridisation kinetics and should be suitable for use as affinity labels in hybridisation experiments.

3.4 Retention of ^{32}P-labelled nick translated DNA on thiol-sepharose following hybridisation to mercurated RNA

It has been shown that mercurated RNAs will bind thiol-sepharose and that they (at least in the case of the 7.5% and 15% mercurated SV40 cRNAs) can hybridise to homologous DNA with kinetics reasonably similar to those of unmercurated RNA. The experiments described in this section are concerned with the following questions: (1) can DNA hybridised to mercurated RNA be bound to thiol-sepharose? (2) Is this binding specific? In the experiment shown in Table 3.1, the same hybridisation reactions used to determine the k_R of the various Hg-cRNA species were allowed to reanneal for 16.5h. At the beginning and at the end of the reactions, aliquots were assayed with endonuclease S_1 to quantitate RNA-DNA hybridisation. Thiol-sepharose chromatography was performed for each of the hybridisation mixtures at the end of the incubation. Since SV40 cRNA is asymmetric, i.e. copied off only the early strand of SV40 DNA, one would expect a maximum of only 50% of the SV40 DNA to become duplex. At the end of the hybridisation reactions using

Table 3.1

Retention of nick translated SV40 DNA on thiol-sepharose following hybridisation to mercurated SV40 cRNA

RNA used in hybridisation	Acid precipitable ^{32}P cpm		% hybridisation at time = 16.5h	Thiol-Sepharose Chromatography			DNA bound / DNA hybridisation
	Total	resistant at time * = 16.5 h		^{32}P cmp unbound	^{32}P cpm bound	% bound	
None	1908	19	1	--	--	--	--
Unmercurated SV40 cRNA	1709	864	50	1995	8	<1	<0.01
7.5% unmercurated SV40 cRNA	1915	814	43	1310	854	40	0.93
15% mercurated SV40 cRNA	1846	800	43	1370	767	36	0.84
30% mercurated SV40 cRNA	1608	161	10	1800	235	11	1.10

SV40 DNA ^{32}P-labelled by nick translation was hybridised against each of the SV40 cRNAs. The hybridisation mixtures and the conditions used were identical to those described in Fig.3.4. At time = 0 and time = 16.5h, 100µl aliquots were withdrawn from each mixture and assayed with endonuclease S_1 as previously described. In addition, one 100µl aliquot of each mixture was acid precipitated and counted to determine the total ^{32}P content. At the end of the 16.5h incubation, 200µl of each of the mixtures containing SV40 cRNA were subjected to thiol-sepharose chromatography on 1ml columns. The columns were washed with 5ml of TNE and then eluted with 5ml of TNE plus 1% (v/v) β-mercaptoethanol. 100µl of formamide was added to the 5ml eluted fractions and 50µl of β-mercaptoethanol was added to the 5ml wash fractions before they were counted by Cerenkov radiation.

* The numbers given are the values for [(S_1 resistant cpms at time = 16.5h) - (S_1 resistant cpms at time = 0)].

unmercurated, 7.5% mercurated and 15% mercurated SV40 cRNAs, the amount of hybridisation observed was close to this figure (50%, 43% and 43% respectively). In the reaction using the 30% mercurated SV40 cRNA, a smaller amount of DNA became duplex as one would expect from the lower k_R of this RNA. The results also indicate that the great majority (greater than 80% in all cases) of the DNA which had hybridised to the mercurated RNAs could be bound to thiol sepharose. It is clear that this binding is due to the Hg residues in the cRNA because SV40 DNA hybridised to unmercurated SV40 cRNA did not bind to the resin. In view of this result, one additional point needs to be made. It is a formal possibility that in the kinetic experiments (Fig. 3.4 and 3.5) the reassociation observed was due to RNA which had somehow lost its Hg residues in the course of the hybridisation reaction and that mercurated RNA was not actually hybridising to DNA. However, this possibility can be ruled out because the present result suggests that the great majority of the DNA which hybridised to mercurated RNA can bind to thiol-sepharose; therefore, the great majority of the RNA which hybridised to the DNA must have been sufficiently mercurated to bind to the resin.

The results also indicate that while the different mercurated cRNA species differ in their k_R, their abilities to bind hybridised DNA to thiol-sepharose are quite similar (i.e. greater than 80% of all hybridised DNA bound). Therefore, it seemed likely that the reason why only a small proportion (11%) of the DNA was bound in the hybridisation with the 30% mercurated cRNA was because only a small proportion of the DNA was hybridised to the mercurated cRNA, due to

the lower k_R of this RNA. One would expect that if the RNA concentration in the hybridisation reaction was increased, a greater proportion of the DNA would be hybridised to the RNA resulting in a greater proportion of the DNA being bound to thiol-sepharose. The results of such an experiment (not shown), in which the concentration of the 30% mercurated cRNA was increased from 400ng per ml to 16 g per ml, indicate that this is the case.

To show that the binding of hybridised DNA to thiol-sepharose is specific for DNA sequences homologous to mercurated RNA, the experiment shown in Fig.3.6 was performed. In this experiment, an excess of 15% mercurated SV40 cRNA was hybridised either to ^{32}P-labelled nick translated SV40 DNA or to ^{32}P-labelled nick translated ckDm 103#48 [a cloned segment of D.melanogaster rDNA (Glover and Hogness, 1975) which shares no homology with SV40 DNA]. Hybridisation in both reactions was monitored by endonuclease S_1 assay and thiol-sepharose chromatography was performed at the end of the reactions. As can be seen, duplex formation was observed and ^{32}P counts were bound to thiol-sepharose only with the reaction containing SV40 DNA. Very little, if any, of the DNA in the reaction mixture containing the ckDm 103 DNA and 15% mercurated SV40 cRNA bound to the resin. In another experiment, the non-specific binding of DNA to thiol-sepharose was examined using in vivo ^{32}P-labelled E.coli DNA; it was approximately 0.1% (results not shown).

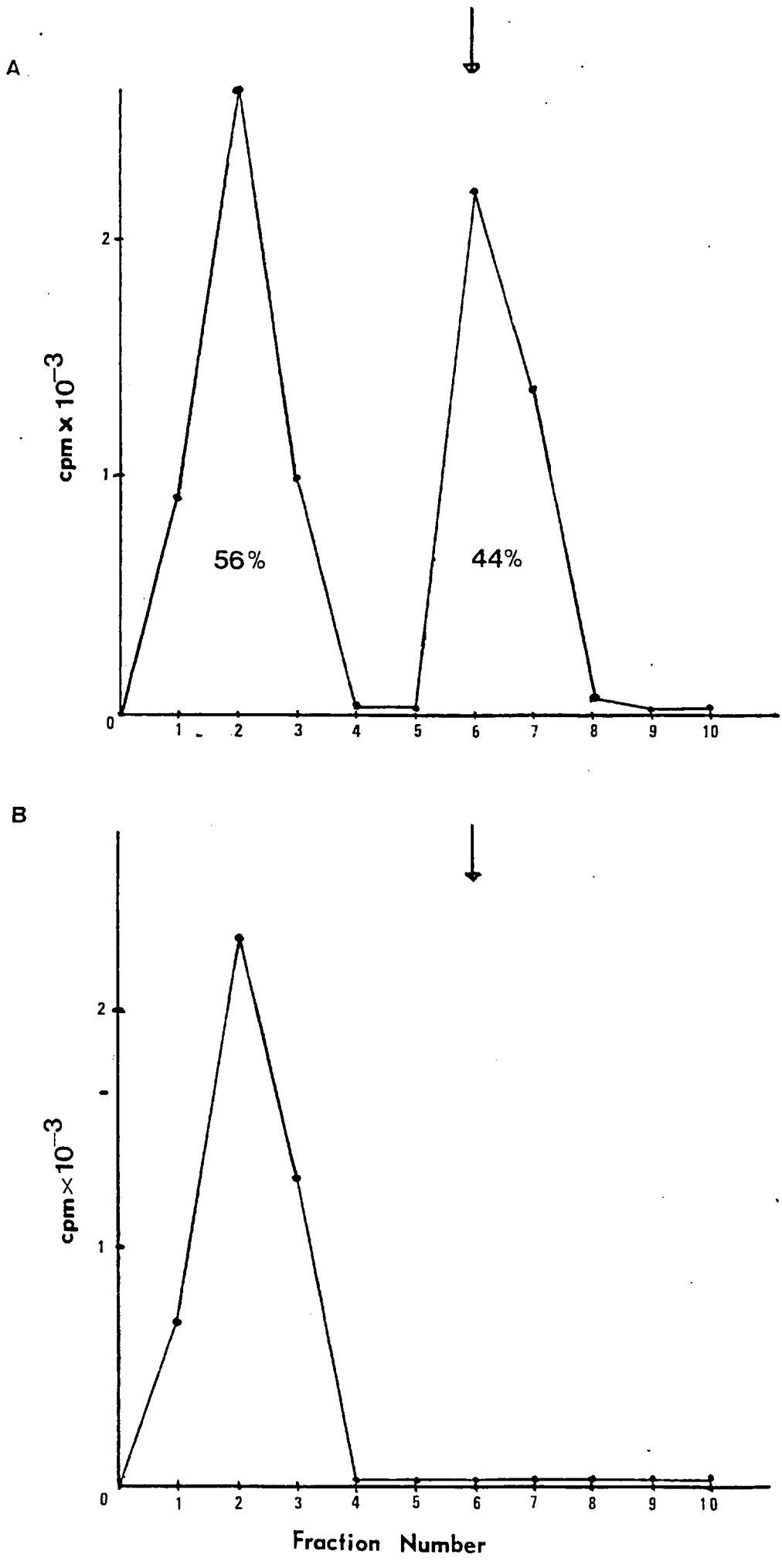
These experiments have demonstrated that nick translated SV40 DNA hybridised with all three species of mercurated SV40 cRNA can be retained on thiol-sepharose. It will be shown in a later experiment

Legend to Figure 3.6

Specificity of binding to thiol-sepharose

The two incubation mixtures used in this experiment contained: 1, 0.3ng per ml of SV40 DNA ^{32}P-labelled by nick translation (specific activity = 4 x 10^7 cpm per μg) and 1μg per ml of 15% mercurated SV40 cRNA; 2, 3ng per ml of ^{32}P-labelled ckDml03#48 DNA (specific activity = 2.3 x 10^6 cpm per μg, also see text) and 1μg per ml of 15% mercurated SV40 cRNA. The final conditions were 100mM Pipes, pH7.7, 1mM EDTA, 300mM NaCl, 0.1% (w/v) SDS and 50% (v/v) formamide. DNA in the incubation mixtures were denatured and the mixtures were cooled on ice prior to the addition of 15% mercurated SV40 cRNA. Incubation was performed at 40°C for approximately 16h. At the end of the incubation period, endonuclease S$_1$ assays were performed. Approximately 50% of the ^{32}P counts in incubation 1 and 5% of the ^{32}P counts in incubation 2 were S$_1$-resistant. 0.5ml of each of the incubation mixtures were chromatographed on 1.5ml (bedvolume) thiol-sepharose columns. Five 1ml TNE wash fractions and five 1ml TNE plus 100mM B-mercaptoethanol elution fractions were collected, acid precipitated and counted. The window of the liquid scintillation counter was set such that ^{14}C counts were excluded. The column profiles for incubations 1 and 2 are shown on A and B respectively.

Fig.3.6


A
cpm x 10⁻³
2
1
0
56%
44%
1
2
3
4
5
6
7
8
9
10
B
cpm×10⁻³
2
1
0
1
2
3
4
5
6
7
8
9
10
Fraction Number

that the chemically mercurated D.melanogaster rRNA shares the same characteristics as the enzymatically mercurated SV40 cRNA. The proportion of DNA retained is considerable (greater than 70% of the theoretical maximum) and this retention is specific for DNA homologous to the mercurated RNA.

3.5 Experiments to bind the full length, linear early strand of SV40 DNA to thiol-sepharose

The experiments performed using nick translated DNA, though informative, do not resemble the real life situation in so far as nick translated DNA has a mean single stranded length of 0.4 Kb., whereas one is interested in purifying large DNA fragments (5 to 20 Kb) useful for cloning experiments. As a step in this direction, some experiments were performed to try and retain the full length, linear early strand of SV40 DNA on thiol-sepharose using hybridised mercurated SV40 cRNA as the affinity probe. Preliminary experiments were inconclusive and emphasised that a number of parameters involved in this type of experiment need to be considered in greater detail.

In hybridisation reactions between linear SV40 DNA and SV40 cRNA two competitive reactions are occuring, DNA-DNA reassociation and RNA-DNA hybridisation. For our purposes, the DNA-DNA reaction needs to be minimised and the RNA-DNA reaction maximised. There are two ways to accomplish this. The first is to use very low DNA concentrations to minimise DNA-DNA reassociation and drive the RNA-DNA reaction with a large excess of RNA. There are two technical problems in doing the experiment in this manner. The first problem

is minor; since *in vivo* labelled SV40 DNA is of relatively low specific activity, filter hybridisation against labelled SV40 DNA or cRNA of high specific activity is necessary to follow the fate of the DNA. The second problem is less trivial; since a large excess of RNA must be used to drive the hybridisation, it is difficult to follow the fate of the RNA in the course of the experiment [particularly because approximately 10% of SV40 cRNA is double stranded (RNase resistant), presumably because transcription is not completely asymmetric]. The second way of optimising for RNA-DNA hybridisation is to do the reaction under the high formamide concentration conditions described by Casey and Davidson (1977). Under these conditions, RNA-DNA hybrids have higher melting temperatures than the corresponding duplex DNA; therefore, it is possible to perform the hybridisation reactions at temperatures where DNA-DNA hybridisation is excluded and only RNA-DNA hybridisation can occur. An additional advantage of using these conditions is that the hybridisation reaction does not need to be performed in large RNA excess, thereby allowing the monitoring of the RNA during the course of the experiment.

The result of an experiment designed to define the temperatures at which only SV40 cRNA - SV40 DNA hybridisation can occur is shown in Fig.3.7. In this experiment, an excess of *in vivo* ^{3}H-labelled SV40 DNA, linearised by digestion with *Eco*RI endonuclease, was hybridised with ^{32}P-labelled SV40 cRNA at a number of temperatures for 2h. Appropriate control reactions were included (see figure legend for details). At the end of the incubation, the reaction

Legend to Figure 3.7

Determination of conditions which allows for the exclusive hybridisation of SV40 cRNA to SV40 DNA

Two incubation mixtures were used for this experiment. They contained: 1, 0.14μg of ^{32}P-labelled SV40 cRNA (specific activity = 10^5 cpm per μg); 2, 0.14μg per ml of the same RNA plus 3.5μg per ml of *in vivo* ^{3}H-labelled SV40 DNA (specific activity = 6.3×10^3 cpm per μg) linearised with EcoRI. The final conditions used were 100mM Pipes, pH7.7, 1mM EDTA, 300mM NaCl, 0.1% (w/v) SDS and 75% (v/v) formamide. 100μl of each mixture was acid precipitated and counted to determine the total ^{3}H and ^{32}P content. 100μl of each mixture was heated to 80°C for 4 mins, diluted into 2ml of S$_1$ buffer and treated either with RNase A (incubation mixture1) or sequentially with endonuclease S$_1$ and RNase A (incubation mixture 2). They were then acid precipitated and counted. These values represented the background.

Five 100μl aliquots were taken from each incubation mixture and placed into microfuge tubes. They were heated to 80°C for 4 mins and then incubated at 40°C, 45°C, 50°C, 53°C or 57°C for 2 h. At the end of the incubation, each aliquot was diluted into 2ml of S$_1$ buffer. The aliquots from reaction 1 were treated with RNase A and the aliquots from reaction 2 were digested sequentially with endonuclease S$_1$ and RNase A. These were then acid precipitated and counted. Background values were subtracted from the values obtained.

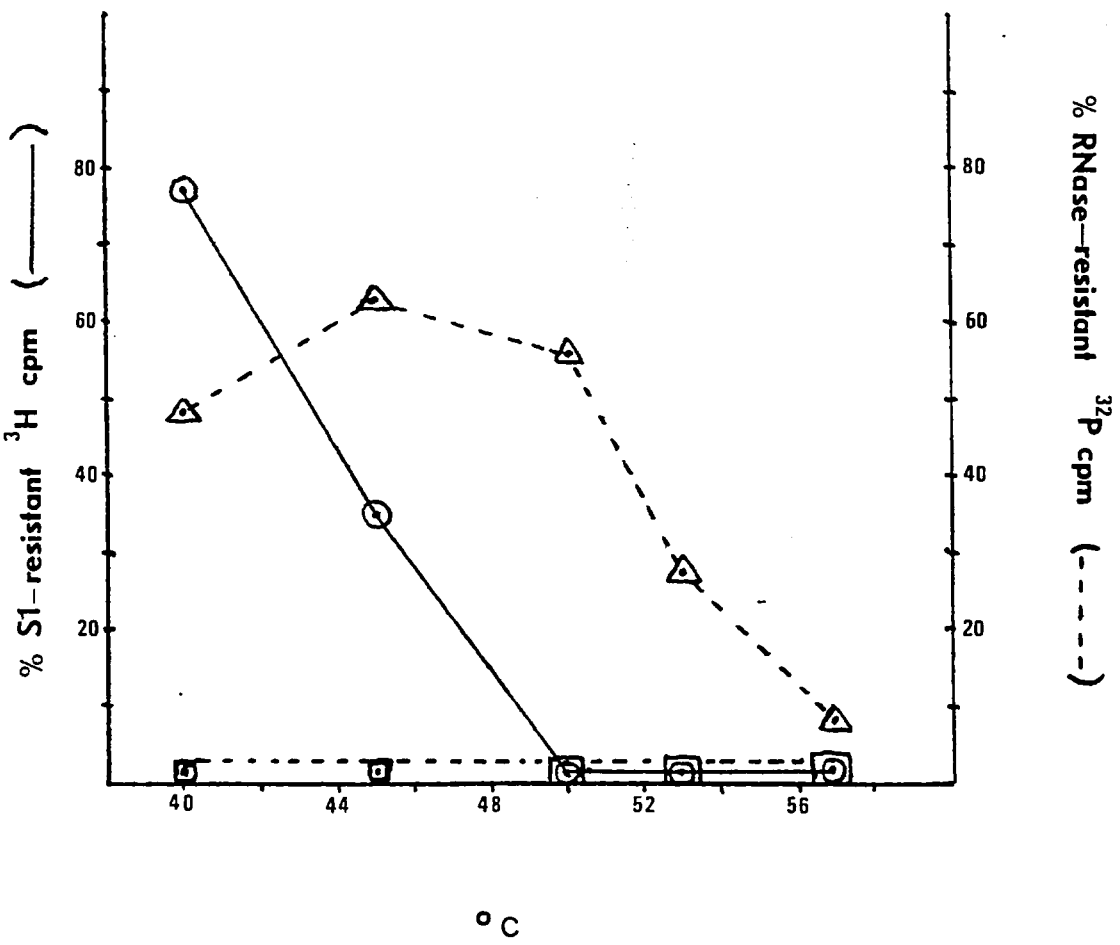
^{32}P counts were read on the ^{14}C + ^{32}P window. The spillover of ^{32}P counts onto the ^{3}H channel was approximately 1% which was left uncorrected.

 ^{32}P counts from mixture 1

 ^{32}P counts from mixture 2

 ^{3}H counts from mixture 2

Fig.3.7


% S1-resistant ^{3}H cpm (——)
% RNase-resistant ^{32}P cpm (- - - -)
80
60
40
20
40
44
48
52
56
°C

mixtures were digested sequentially with endonuclease S_1 and RNase A and the acid precipitable ^{3}H and ^{32}P counts were determined. The results clearly indicate that while no detectable DNA-DNA reassociation occurred at temperatures of 50°C or above, RNA-DNA hybridisation can still be detected at 57°C.

Having determined the appropriate conditions, an experiment was performed by hybridising ^{14}C-labelled 15% mercurated SV40 cRNA to ^{3}H-labelled SV40 DNA linearised by digestion with EcoRI endonuclease. In this experimental hybridisation, the mass ratio of SV40 cRNA to the early strand of SV40 DNA was 1.5:1. Parallel control reactions using identical conditions but containing ^{3}H-labelled linear SV40 DNA only, Hg-SV40 cRNA only or Hg-SV40 cRNA plus ^{3}H-labelled nick translated λ DNA were also included in this experiment. Incubation was performed at 51°C for 6h. The amounts of double-stranded DNA and RNA-DNA hybrid formed during the incubation were assayed by treating aliquots of the reactions diluted into S_1 buffer sequentially with predetermined amounts of endonuclease S_1 and RNase A and quantitating the acid precipitable radioactivity. At the end of the 51°C incubation, thiol-sepharose chromatography was performed on all of the experimental and control mixtures.

The results are shown in Table 3.2. Endonuclease S_1 assays on the control mixture containing only linear SV40 DNA showed that essentially all of the DNA in the reaction mixture is single-stranded. Since this DNA was not denatured prior to incubation at 51°C, this indicate that linear unit length SV40 DNA is indeed denatured at this temperature under the conditions used. In the

Legend to Table 3.2

Four incubation mixtures were used for this experiment:

Incubation A contained 12μg ml^{-1} of 15% mercurated SV40 cRNA plus 16μg ml^{-1} of SV40 DNA linearised by digestion with EcoRI

Incubation B contained 0.75μg ml^{-1} of SV40 DNA linearised by digestion with EcoRI

Incubation C contained 1.5μg ml^{-1} of 15% mercurated SV40 cRNA

Incubation D contained 1.5μg ml^{-1} of 15% mercurated SV40 cDNA plus 3ng ml^{-1} of nick translated λ DNA.

Final conditions used were 100mM Pipes, 1mM EDTA, 300mM NaCl, 0.1% ($^{v}/v$) SDS, and 75% $^{v}/v$ formamide, pH 7.7.

The SV40 DNA used was *in vivo* labelled with ^{3}H-Thymidine (specific activity 6.3×10^{3} cpm μg^{-1}).

The λ DNA was ^{3}H-labelled by nick translation (specific activity 4.3×10^{6}, cpm μg^{-1}). 15% mercurated SV40 cRNA has been described (specific activity 700 cpm μg^{-1}).

Incubations A, C and D were heated to 80°C for four mins. and then incubated at 51°C for 6h. Incubation B was incubated at 51°C for 6h. A 20λ aliquot of incubation A and a 100λ aliquot of incubation B were acid precipitated and counted. 20λ aliquots were taken from incubation A immediately following the 80°C treatment and following the 6h. incubation at 51°C. These were diluted into 1ml. of S1 buffer, digested sequentially with endonuclease S1 and RNase A, acid precipitated and counted. A 100λ aliquot was taken from incubation B following the 6h. incubation at 51°C and similarly treated. Following the 6h. incubation at 51°C, 200λ of incubation A, 350λ of incubation B, 700λ of incubation C, and 350λ of incubation D were subjected to thiol-sepharose chromatography using 1.5ml. columns. The columns were washed with 5ml. of TNE and then eluted with 5ml of TNE + 1% ($^{v}/v$) β-mercaptoethanol. These were acid precipitated and counted.

Table 3.2

Retention of full length linear SV40 DNA on thiol-sepharose columns following hybridisation to mercurated SV40cRNA

Incubation	Hybridisation assays with endonuclease S_1 and RNase A						Thiol-Sepharose Chromatography					
	Acid precipitable ^3H cpm	S_1 resistant ^3H cpm	% S_1 resistant ^3H cpm	Acid precipitable ^{14}C cpm	RNase A resistant ^{14}C cpm	% RNase A resistant ^{14}C cpm	^3H cpm not bound	^3H cpm bound	% ^3H cpm bound	^{14}C cpm not bound	^{14}C cpm bound	% ^{14}C cpm bound
A Hg SV40 cRNA (^{14}C) + Linear SV40 DNA (^3H)	2200[b]	915[a]	42%	159	90[a]	56%	13700	5400	28%	650	1010	61%
B Linear SV40 DNA (^3H)	400	< 20	< 5%	---	---	---	1500	< 10	< 1%	---	---	--
C Hg SV40 cRNA (^{14}C)	---	---	---	---	---	---	---	---	---	<20	580	>95%
D Hg SV40 cRNA (^{14}C) + Nick translated λ DNA (^3H)	---	---	---	---	---	---	3800	< 100[b]	<3%	<20	290	>90%

a. The value shown is the difference between the values obtained from the aliquot taken following the 6h incubation at 51°C and the aliquot taken immediately following the 80°C step.

b. About one third of the ^{14}C counts spilled over into the ^3H channel. However, since the amount of ^3H counts was large compared to ^{14}C counts, this effect was left uncorrected.

experimental reaction mixture, following the 6h incubation, the proportion of S1-resistant DNA was 42% (reasonably close to the expected figure of 50%) and the proportion of RNase A resistant RNA was 56% (the expected value for an RNA to early strand DNA ratio of 1.5:1 is 67%). These results strongly support the assertion that the experimental conditions used produced the desired effect, namely the exclusive hybridisation of the 15% mercurated SV40 cRNA to the early strand of SV40 DNA. Results from thiol-sepharose chromatography showed that:

(1) essentially all of the 15% mercurated SV40 cRNA from the control incubations containing only RNA or RNA plus ^{3}H-labelled nick translated λ DNA bound to the resin,

(2) none of the DNA from the control incubations containing SV40 DNA only or Hg-SV40 cRNA plus nick translated λDNA bound to thiol-sepharose,

(3) approximately 61% of the Hg-SV40 cRNA and 28% of the SV40 DNA from the experimental incubation bound to thiol-sepharose.

It is clear from these results that the proportion of the early strand of SV40 DNA which bound to thiol-sepharose following hybridisation with Hg-SV40 cRNA was approximately half of the theoretical maximum of 50%. This binding is specific for SV40 DNA since none of the nick translated λ DNA bound following hybridisation with Hg-SV40 cRNA. However, why did so much of the early strand of SV40 DNA not bind to thiol-sepharose? The reason cannot be due to the DNA not being hybridised to the Hg-SV40 cRNA because the endonuclease S1 and RNase A assay performed at the end of the hybridisation reaction indicated that most (84%) of the

SV40 DNA complementary to the Hg-SV40 cRNA had hybridised to the RNA. A likely explanation emerges, however, if the experiment is examined from the standpoint of the RNA. In the control incubations where the Hg-SV40 cRNA had no homologous DNA to hybridise to, essentially all of the RNA bound to the resin. In the experimental incubation where the RNA could hybridise to the early strand of the SV40 DNA, a significant proportion of the RNA (39%) did not bind. The simplest interpretation of these results is that the ability of Hg-SV40 cRNA to bind to thiol-sepharose is decreased when it is duplexed with a unit length early strand of SV40 DNA. As a result, approximately half of the early strand of SV40 DNA duplexed with the Hg-SV40 cRNA did not bind to thiol-sepharose.

Other experiments have been performed using the high formamide conditions. These experiments make the following points:

(1) Increasing the excess of 15% Hg-SV40 cRNA in the hybridisation reaction increases the proportion of RNA (to over 90% with tenfold RNA excess) but does not increase the proportion of SV40 DNA bound to thiol-sepharose. This result further supports the claim that it is not the lack of hybridisation to Hg-SV40 cRNA which resulted in some of the early strand SV40 DNA not binding to the resin.

(2) The labelled nucleic acids which did not bind to the resin, following hybridisation, will not bind upon repassage through fresh resin. This formally eliminates the possibility that the capacity of the resin had been exceeded.

(3) The DNA in both the unbound and the bound fractions was full length linear SV40 DNA. This was determined by electrophoresing DNA from the bound and unbound fractions on agarose gels cast in 30mM NaOH, 2mM EDTA and detecting the DNA by Southern transfer hybridisation. Therefore, the

possibility that degradation of DNA occurred and only the smaller degraded fragments were bound to thiol-sepharose can be eliminated.

(4) The content of formamide in the samples loaded onto the thiol-sepharose columns made no difference to the amount of nucleic acids bound.

Together these results are consistent with the previous interpretation of the experiment shown in Table 3.2 and eliminate some of the more trivial alternative explanations.

In summary, the proportion of full length SV40 DNA which binds to thiol-sepharose following hybridisation with Hg-SV40 cRNA is significantly less than for nick translated SV40 DNA. The probable reason for this decrease in binding is that, due to some unknown mechanism, approximately half of the Hg-SV40 cRNA when duplexed with full length SV40 DNA will not bind to thiol-sepharose.

3.6 Binding of cDm103 DNA to thiol-sepharose following hybridisation with mercurated D.melanogaster 18S and 28S rRNAs

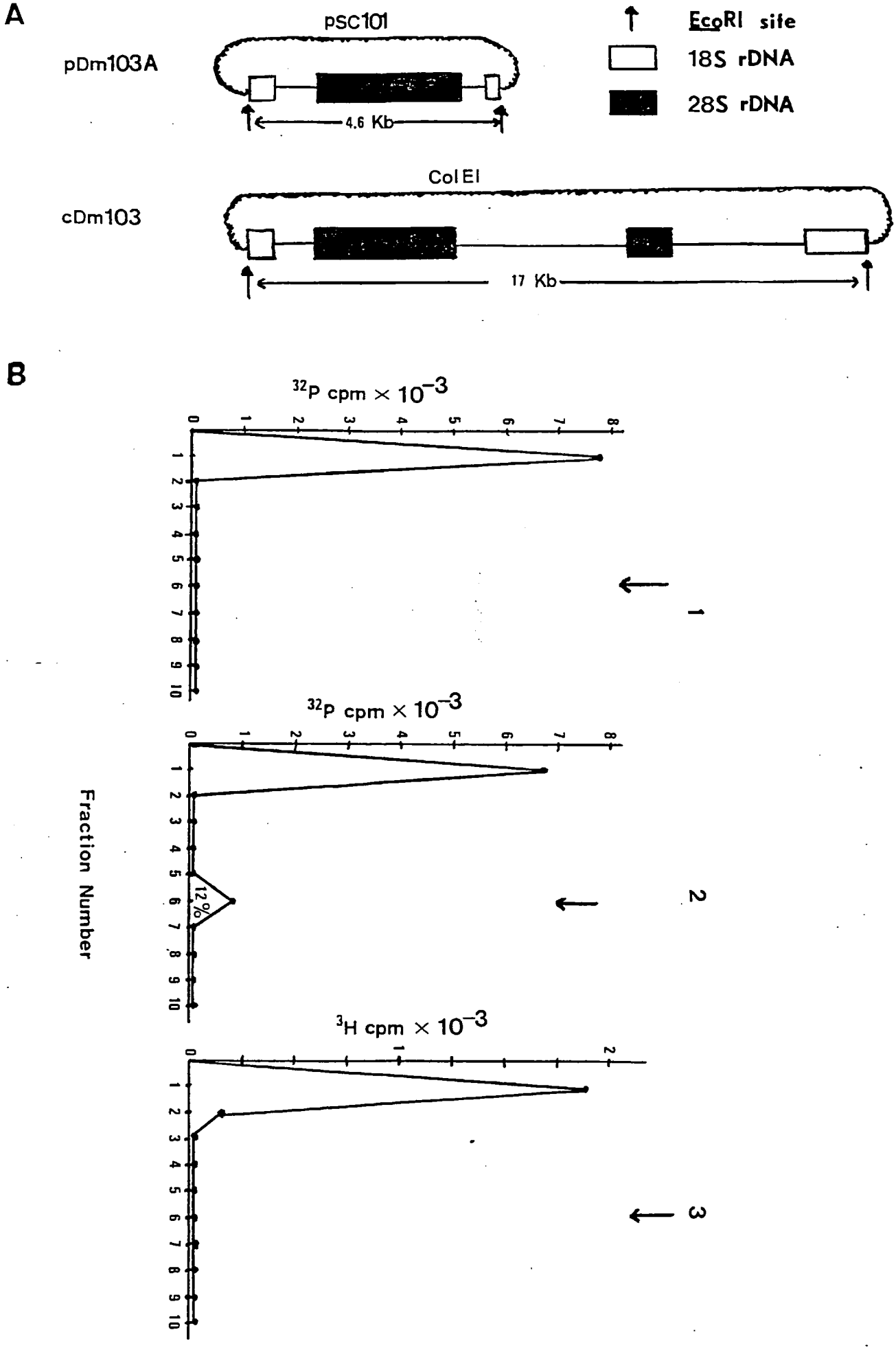
The experiment shown in Fig.3.8 is designed to determine if the 17Kb EcoRI fragment from cDm103 (see figure), which contains a complete repeat unit of D.melanogaster rDNA [containing 1.9Kb of homology to 18S rRNA and 4.2Kb of homology to 28S rRNA (Glover and Hogness, 1977)], can be bound to thiol-sepharose following R-looping of EcoRI endonuclease cleaved, ^{3}H-labelled cDm103 DNA with an excess of mercurated 18S and 28S rRNAs. As a positive control, an excess of Hg-rRNA was hybridised to ^{32}P-labelled nick translated pDm103A (see figure) DNA. Another control incubation containing only nick translated pDm103A DNA was also included. Preliminary experiments were performed, using the electron microscope, to determine the

Legend to Figure 3.8

DNA molecules R-looped with mercurated RNA do not bind to thiol-sepharose

In A, the structures of cDml03 and pDml03A are diagrammed. Three incubation mixtures were used for the experiment shown in B. Incubation 1 contained 2 x 10^4 cpm of ^{32}P-labelled nick translated pDml03A DNA (specific activity $\geq 10^7$ cpm per μg). Incubation 2 contained 2 x 10^4 cpm of ^{32}P-labelled nick translated pDml03A DNA and 15μg per ml of unlabelled, chemically mercurated Drosophila 18S and 28S rRNAs. The final conditions used were 100mM Pipes, pH7.7, 1mM EDTA and 50% (v/v) formamide. The incubation mixtures were denatured by incubation at 90°C for 4 mins. Hybridisation was performed at 40°C for approximately 16h. 200μl of incubations 1 and 2 were chromatographed on 1ml thiol-sepharose columns. Approximately 2ml fractions were collected and counted by Cerenkov radiation. Incubation 3 contained 1.3μg per ml of EcoRI cleaved cDml03 DNA labelled *in vivo* with ^{3}H-thymidine (specific activity = 7 x 10^3 cpm per μg) and 8μg per ml of unlabelled, chemically mercurated Drosophila 18S and 28S rRNAs. The final conditions used were 83mM Pipes, pH7.7, 10mM EDTA and 70% (v/v) formamide. R-looping incubation was performed at 47°C for approximately 16h. 20μl of the incubation mixture was spread for electron microscopy and 200μl of the incubation mixture was subjected to thiol-sepharose chromatography. Approximately 1ml fractions were collected. They were acid precipitated and counted.

Fig.3.8


A
pDm103A
pSC101
4.6 Kb
EcoRI site
18S rDNA
28S rDNA
cDm103
ColEI
17 Kb
B
^{32}P cpm $\times 10^{-3}$
1
^{32}P cpm $\times 10^{-3}$
12%
2
^{3}H cpm $\times 10^{-3}$
3
Fraction Number

optimal temperature for R-looping EcoRI endonuclease cleaved cDm103 DNA with rRNA under the R-looping conditions defined by Thomas et al. (1975). 47° to 49°C were suitable temperatures. Following incubation at 47°C overnight, aliquots of the R-looping mixture containing EcoRI endonuclease cleaved cDm103 DNA and Hg-rRNA were diluted into 50% (v/v) formamide, spread over 15% (v/v) formamide and examined under the electron microscope. Of the clearly interpretable molecules observed, a majority had 28S loops or 18S forks associated with the 17Kb molecule (White and Hogness, 1977). Thiol-sepharose chromatography was performed on all three incubations and the results are shown in Fig.3.8. The Hg-rRNA used was the unlabelled rRNA described in the experiment shown in Fig.3.2. It is clear that this rRNA is mercurated and can bind thiol-sepharose because approximately 12% of the ^{32}P-labelled nick translated pDm103A DNA was bound to thiol-sepharose following hybridisation with an excess of this RNA. Since pDm103A is a 13.6Kb plasmid which contains approximately 0.5Kb of the 18S rRNA coding region and 3.3Kb of the 28S rRNA coding region (Glover and Hogness, 1977), one would expect $14\% \left(\frac{0.5\ (3.3 + 0.5)}{13.6} \right)$ of the plasmid DNA to be complementary to 18S and 28S rRNAs. So the 12% bound represents a very respectable figure. When the R-looping incubation was subjected to thiol-sepharose chromatography, essentially none of the DNA bound to the column. This result indicates that the 17Kb EcoRI fragment of cDm103 does not bind to thiol-sepharose following R-looping with mercurated rRNA.

3.7 Discussion

The results presented here show that mercurated RNA species can be made which bind to thiol-sepharose with high affinity and

specificity. These mercurated RNA species, particularly the ones with lower mercury contents, can hybridise against homologous DNA with kinetics similar to those of unmercurated RNA. Furthermore, it is possible to bind to thiol-sepharose, with high specificity and yield, DNA hybridised against these mercurated RNAs providing that the DNA used is short (i.e. the nick translated DNAs used for these experiments had mean single stranded lengths of approximately 0.4Kb). However, when full length linear SV40 DNA was hybridised to mercurated SV40 cRNA under conditions where only RNA-DNA hybridisation can occur the amount of SV40 DNA bound was reduced. Moreover, experiments with EcoRI cleaved cDm103 DNA R-looped with mercurated rRNA resulted in no binding of the R-looped DNA to thiol-sepharose.

The reasons why full length SV40 DNA and R-looped cDm103 DNA were not retained as well as their nick translated counterparts are unclear. It is possible that when mercurated RNA is duplexed with DNA the Hg groups become sterically unavailable for binding thiol-sepharose; one would then expect RNA-DNA hybrids to bind thiol-sepharose if they contained single-stranded, mercurated RNA tails. Since one would expect a greater proportion of the hybrid molecules to contain single-stranded RNA tails if the DNA used in the hybridisation reaction is short, the observation that nick translated DNA is retained more efficiently than full length DNA is consistent with this speculation. In the case of cDm103 DNA R-looped with mercurated rRNA, branch migration might have played a role in the lack of binding. R-loops are relatively stable because as the RNA is displaced by branch migration of DNA, the displaced

RNA can in turn branch migrate and redisplace the DNA. These two competing reactions will presumably reach a state of equilibrium which would explain the relative stability of R-loops. However, if R-loops are treated with RNase, they quickly disappear (Thomas et al., 1975). The reason for this is that as the RNA is displaced, it is degraded so the reverse reaction of the displaced RNA branch migrating and redisplacing DNA cannot occur. It is conceivable that thiol-sepharose could have a very similar effect on R-loops formed with mercurated RNA; mercurated RNA transiently displaced from R-loops will bind very tightly to thiol-sepharose; once bound, this RNA may not be able to redisplace the branch migrating DNA. If so, DNA branch migration will quickly result in the complete displacement of the R-looped RNA from the DNA molecule.

Whatever the reasons, it is clear that the straightforward approach of using mercurated RNA as an affinity label for purifying double-stranded DNA does not work. A number of other laboratories have come to the same conclusion (R. White, K.B. Smith and D. Ward, personal communications). Further experimentation can probably tell us why and perhaps provide new ideas on how to make the method work. However, in view of the increased cloning efficiencies obtainable with λ in vitro packaging systems (Hohn and Murray, 1977; Sternberg et al., 1977), it was felt that it would be more profitable to develop high efficiency cloning systems which would obviate the need for prepurification.

Chapter 4

The Fate of SV40 DNA Following Infection of Nonpermissive Mouse Cells

4.1 Introduction

Viral transformation involves complex interactions between viral gene products and the host cell (reviewed by Martin, 1980 and Tooze, 1980); it is likely that the stable inheritance of the viral DNA and the continued expression of the early viral functions play an important role in viral transformation (see Chapter 1). It is clear that integration provides an efficient mechanism for the inheritance of viral genes since all SV40-transformed cell lines examined to date contain viral DNA covalently integrated into the cellular genome (Martin, 1980). In order to understand the role of integration in viral transformation, it is important to elucidate the molecular mechanisms by which integration occurs. In cell lines transformed by SV40, the integration of viral DNA is nonspecific with respect to both the viral and cellular genomes (see Chapter 1 Section 1.5). Furthermore, a high proportion of the transformed lines contain integrated viral DNA with partial tandem repetition of the viral sequences (Botchan et al., 1976; Campo et al., 1978; Chepelinsky et al., 1980; Rigby et al., 1980). While the tandem duplications could result from the successive integration of monomeric viral DNAs, the first integration by illegimate recombination and the subsequent integrations by homologous recombination, the experiments of Graessmann et al. (1979) and M. Botchan (personal communication) suggest that this possibility is unlikely (see Chapter 1, Section 1.5). Therefore, if the substrates for integrative recombination are monomeric viral DNA and cellular DNA, one would not expect the integrated viral genomes to contain partial tandem duplications.

Collins and Sauer (1972) have reported that when nonpermissive mouse cells are infected with SV40 (I) DNA, large amounts, approximately 10^3 viral genome equivalents per cell, of viral DNA become integrated into cellular DNA within two days. However, in this study integration was inferred only from velocity sedimentation and equilibrium density gradient centrifugation experiments, conditions which separate large, linear cellular DNA from small, circular viral DNA. Rigby and Berg (1978) have shown that these criteria for integration can be highly misleading if the infected cells contain novel forms of viral DNA produced by DNA replication or recombination. In this section, experiments are described which were designed to directly follow the fate of the infecting viral DNA by using the transfer hybridisation method of Southern (1975). It was our hope that this approach can help elucidate the mechanisms by which SV40 DNA integrates into the chromosomal DNA of mouse cells.

4.2 Infected nonpermissive mouse cells accumulate a high molecular weight form of SV40 DNA

Confluent Balb/c 3T3 clone A31 cells were mock infected and infected with virions of the wt830 strain of SV40 at a MOI of 300 pfu per cell. At 2h and 72h PI total cellular DNA was prepared from the mock infected and infected cultures as described in materials and methods (section 2.3.3). Cellular DNA prepared in this way coelectrophoreses with bacteriophage λ DNA on 0.4% (w/v) agarose gels and, therefore, has an average molecular weight of at least 50 kb. Following agarose gel electrophoresis, the DNA was denatured in situ and transferred to a nitrocellulose filter by the method of Southern (1975). The filter was then hybridised with SV40 DNA labelled with ^{32}P by nick translation and the location, on the filter, of DNA containing SV40 sequences was visualised by autoradiography (Fig.4.1). AT 72h PI, hybridisation was found not only at positions corresponding to the three conventional forms of monomeric viral DNA but also at the position of the cellular DNA, i.e. at the exclusion limit of the gel. Persistence and mild degradation of the infecting viral DNA can account for the three monomeric forms of SV40 DNA but does not explain the high molecular weight form. That this new form of SV40 DNA might be due to artefactual association of unit sized viral DNA with cellular DNA, or due to cross hybridisation between 3T3 cellular DNA and SV40 DNA, is unlikely because the control tracks containing DNA prepared from an infected culture at 2h PI and from a mock infected culture at 72h PI were negative for this band. Therefore at 72h PI, infected 3T3 cells maintained on 10% (v/v) FCS accumulate a novel form of SV40 DNA which coelectrophoreses with cellular DNA. The structure of these molecules will be discussed in section 4.4.

Legend to Figure 4.1

SV40-infected mouse cells contain a high molecular weight form of viral DNA.

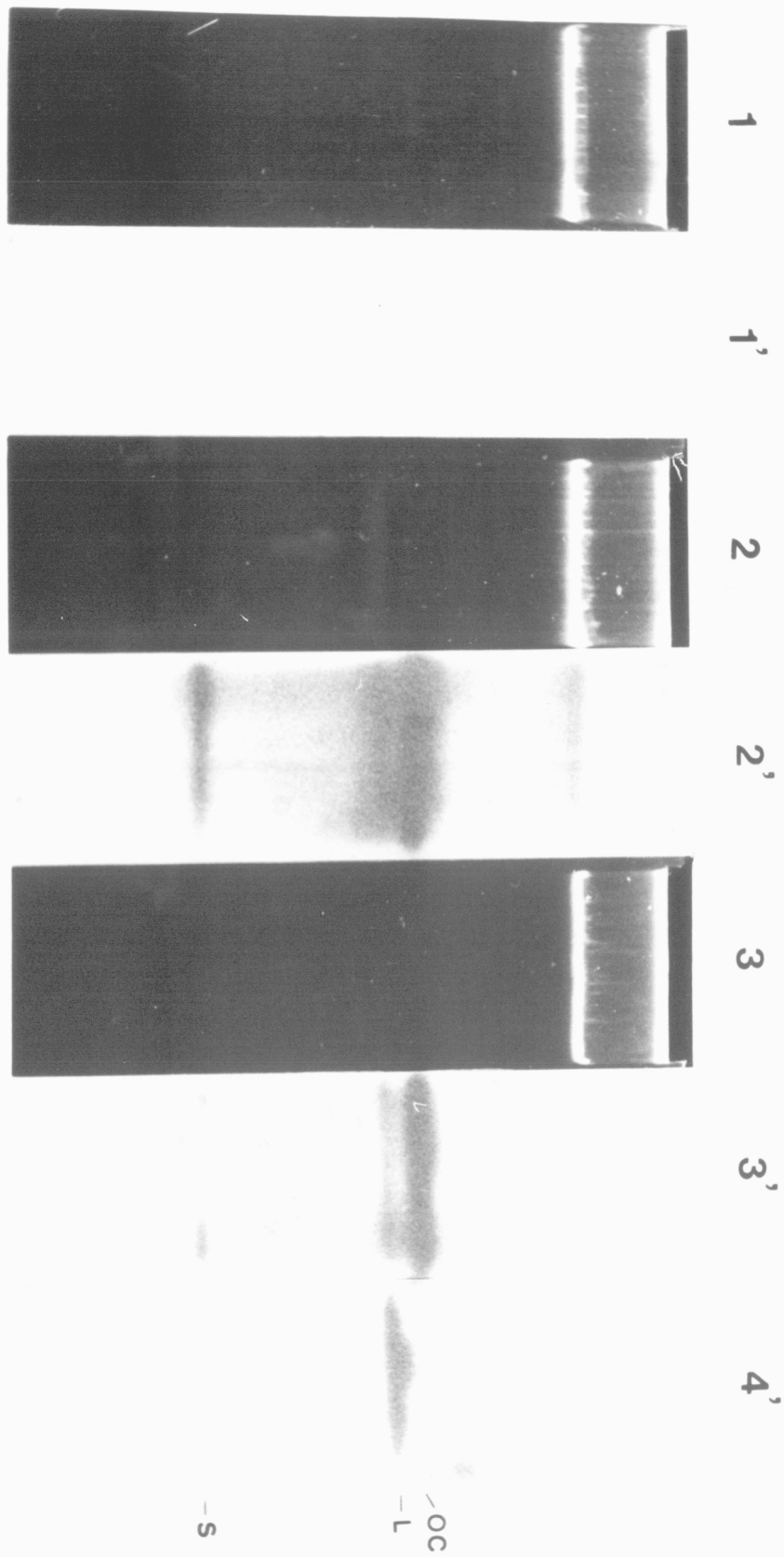
Total DNA was isolated (as described in section 2.3.3) from mock infected cells at 3 days P.I. and from wt830-infected (MOI=300) cells at 2h and 3 days P.I. The DNAs were analysed by gel electrophoresis (0.75% (w/v) agarose gel) and transfer hybridisation. The unprimed tracks are photographs of the DNAs stained with ethidium bromide following fractionation. The primed tracks are autoradiograms of the corresponding unprimed tracks following transfer hybridisation. The contents of the tracks are: track 1, 2 μg of DNA from mock infection; track 2, 2 μg of DNA from wt830-infected culture prepared at 3 days P.I.; track 3, 2 μg of DNA from wt830-infected culture prepared at 2h P.I.; track 4, 0.1ng of linear SV40 DNA. ^{32}P-labelled SV40 DNA was used as the hybridisation probe.

OC = open-circular SV40 DNA

L = linear SV40 DNA

S = superhelical SV40 DNA

Fig. 4.1



4.3 The kinetics of the appearance of the high molecular weight SV40 DNA

The kinetics of the appearance of these molecules in SV40-infected cells maintained under different serum conditions was followed by examining DNA prepared at various times using the transfer hybridisation technique. When infected cells were maintained in medium supplemented with 10% (v/v) FCS, a faint high molecular weight band was observed at 64h PI and the intensity of this band was increased at 85h PI (Fig 4.2). When the infected cultures were taken at 72h PI, when they already contain high molecular weight SV40 DNA, and passaged like normal 3T3 cells, i.e. subcultured at a dilution of 1:3 every three days using medium supplemented with 10% (v/v) FCS, these sequences were lost within six days (Fig 4.3). A very similar result was seen when the infected cells were maintained on medium supplemented with 2% (v/v) FCS without passaging, the conditions often used to select for transformed cell lines. In this case, however, the high molecular weight viral DNA appeared later, at 5 days PI, persisted at least until 10 days PI and was essentially gone by 14 days PI (Fig 4.4). It is clear that cellular transformation could be induced under the experimental conditions used because when infected and mock infected cultures were maintained on medium supplemented with 2% (v/v) or 10% (v/v) FCS, focus formation was observed in the infected cultures but not the mock infected cultures at three weeks PI. These experiments demonstrate that under experimental conditions which allow the induction of cellular transformation, SV40-infected 3T3 cells accumulate a high molecular weight form of viral DNA shortly following infection. The accumulation of this novel form of viral DNA appears to be a transient phenomenon.

Legend to Figure 4.2

The kinetics of appearance of high molecular weight form of SV40 DNA in infected cells maintained on 10%(v/v)FCS

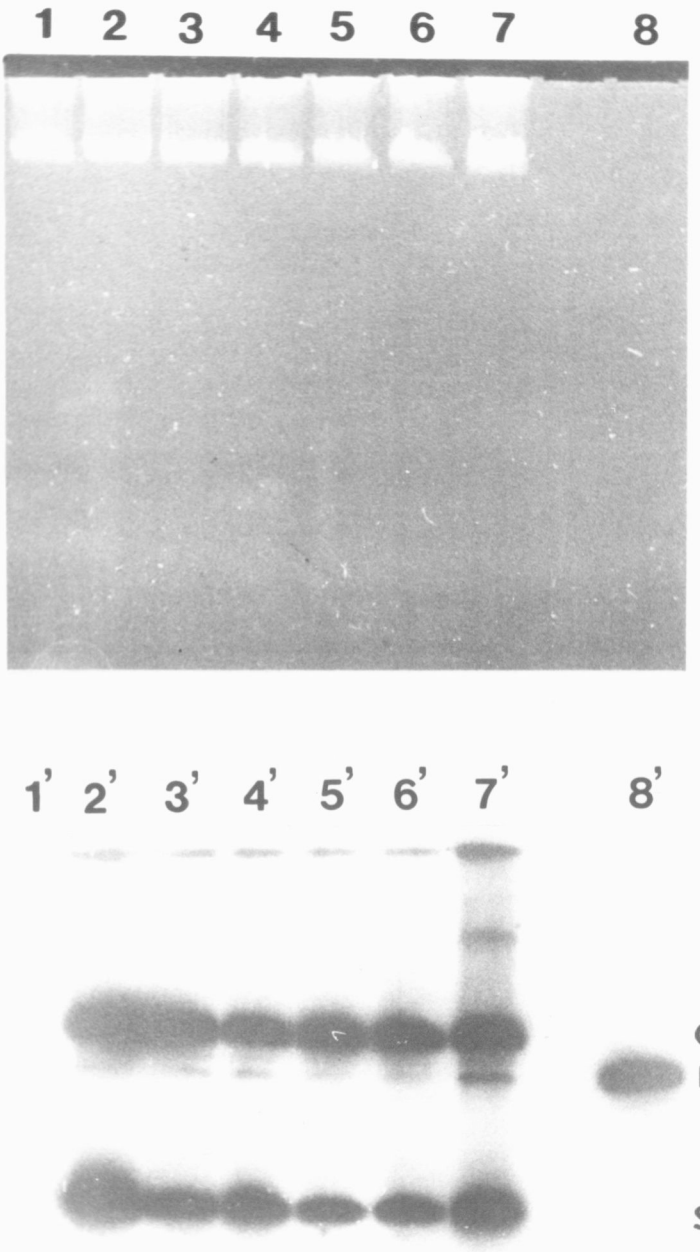
2 μg of each of the DNAs prepared from mock infected mouse cells at 12h P.I. (track 1) and wt830-infected (MOI=100) mouse cells at 12h P.I. (track 2), 24h P.I. (track 3), 36h P.I. (track 4), 48h P.I. (track 5), 64h P.I. (track 6) and 85h P.I. (track 7) were subjected to electrophoresis on an 0.7% (w/v) agarose gel and transfer hybridisation. The top panel (unprimed tracks) is a photograph of the fractionated DNAs stained with ethidium bromide. The bottom panel (primed tracks) is an autoradiogram obtained following transfer hybridisation of the gel. Track 8 contain 0.1 ng of linear SV40 DNA. ^{32}P-labelled SV40 DNA was used as the hybridisation probe.

OC = open circular SV40 DNA

L = linear SV40 DNA

S = superhelical SV40 DNA

Fig. 4.2

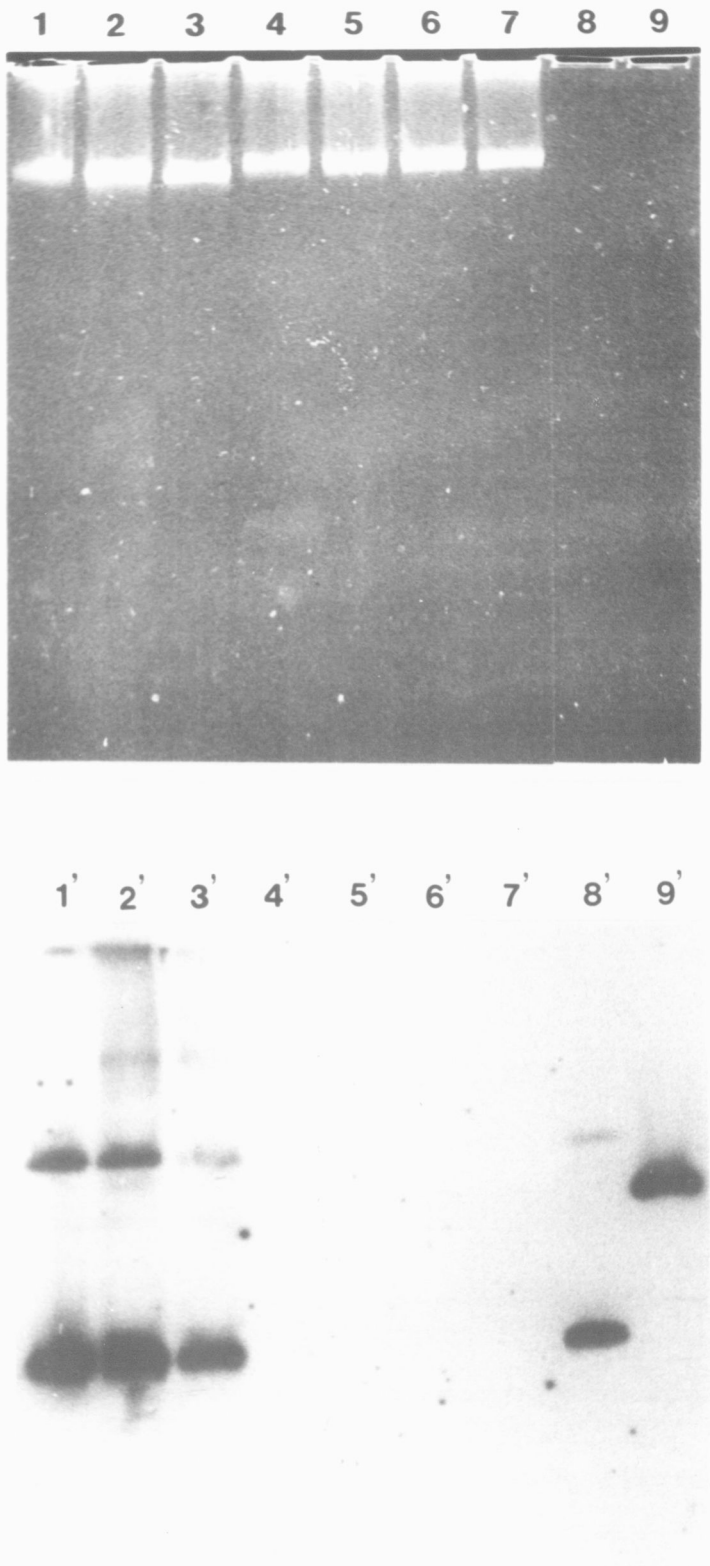


Legend to Figure 4.3

The effect of passaging on the high molecular weight form of SV40 DNA

Total DNA was prepared from mouse cells infected with wt830 (MOI=100) at 6h P.I. (track 1) and 72h P.I. (track 2). An infected culture was also subcultured at a dilution of 1:3 every three days using medium supplemented with 10% (v/v) FCS and total DNA was prepared from the subcultures at 6 days P.I. (track 3), 9 days P.I. (track 4), 11 days P.I. (track 5), 14 days P.I. (track 6) and 17 days P.I. (track 7). The DNAs were analysed by electrophoresis on an 0.7% (w/v) agarose gel and transfer hybridisation. The top panel is a photograph of the gel following ethidium bromide staining. The bottom panel is an autoradiogram obtained following transfer hybridisation of the same gel. Track 8 contains 0.1ng of open-circular (OC) and superhelical (S) SV40 DNA; track 9 contains 0.1ng of linear (L) SV40 DNA.

Fig.4.3

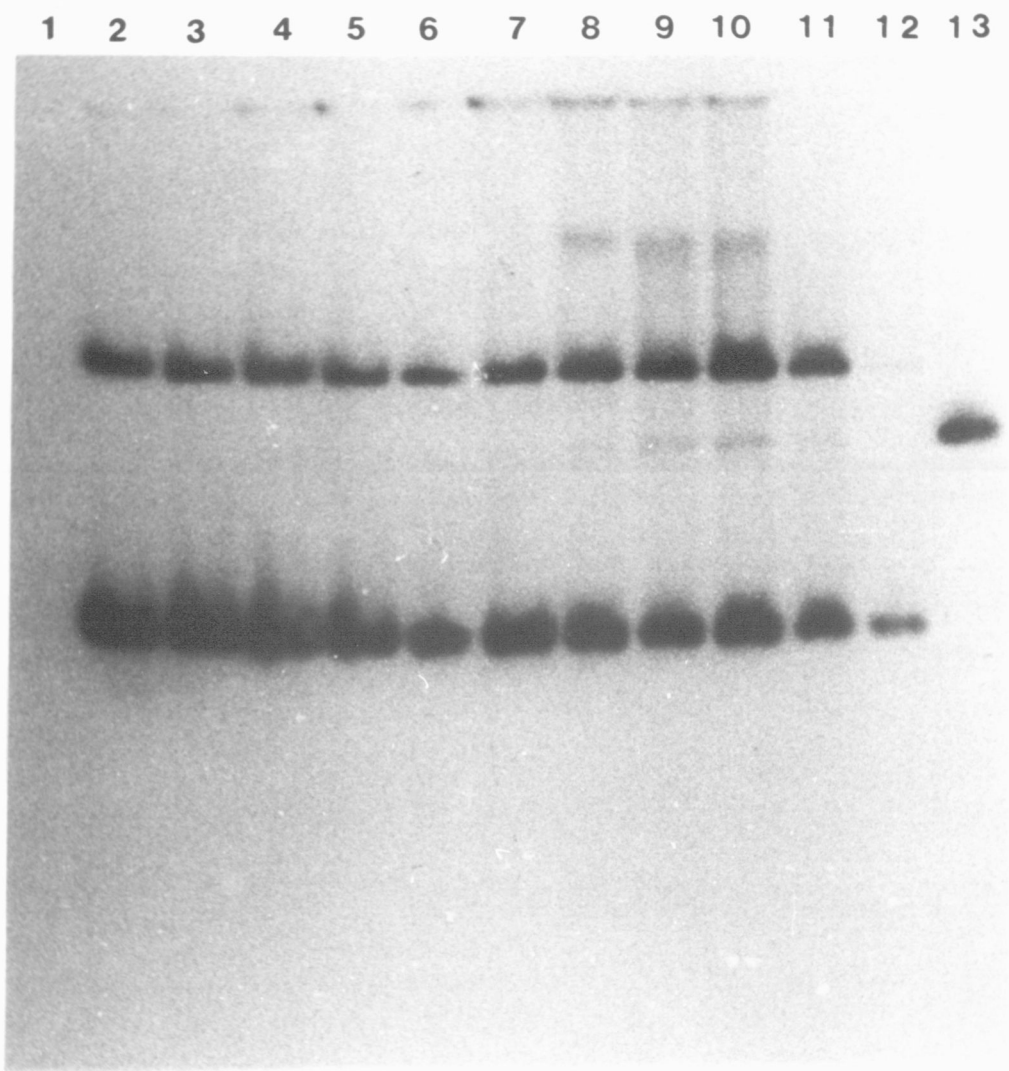


Legend to Figure 4.4

The kinetics of appearance of the high molecular weight form of SV40 DNA in infected cells maintained on 2%(v/v)FCS

Total DNA was prepared from mock infected (track 1) and wt830-infected (MOI=100) mouse cells at 7h P.I. (track 2), 20h P.I. (track 3), 32h P.I. (track 4), 45h P.I. (track 5), 56h P.I. (track 6), 72h P.I. (track 7), 122h P.I. (track 8), 190h P.I. (track 9), 242h P.I. (track 10) and 336h P.I. (track 11). The cells were infected at confluence and maintained on medium supplemented with 2% (v/v) FCS. The infected cells were supplied with fresh medium every 5 days. The DNAs were analysed by electrophoresis on an 0.7% (w/v) agarose gel and transfer hybridisation. An autoradiogram obtained using ^{32}P-labelled SV40 DNA as the hybridisation probe is shown.

Fig.4.4



It is well known that viral infection can induce a round of cellular DNA replication in growth arrested nonpermissive cells (see Chapter 1 section 1.4). The results of a pulse labelling experiment using 3T3 cultures infected at confluence and maintained on medium supplemented with 2% (v/v) FCS is shown in Fig 4.5. The peak of the virally induced cellular DNA replication occurred at 2 to 3 days PI, a time well before the appearance of the high molecular weight form of viral DNA under similar conditions (i.e. at 5 days PI). Therefore, the appearance of the molecules does not coincide temporally with the virus-induced cellular DNA replication.

4.4 The structure of the high molecular weight form of SV40 DNA

The viral sequences associated with the cellular DNA could either be covalently integrated, as suggested by Collins and Sauer (1972), or they could be oligomeric or polymeric forms of SV40 DNA similar to those described in productively infected permissive cells (Martin et al., 1976; Rigby and Berg, 1978). These two possibilities can be clearly distinguished by restriction endonuclease mapping experiments; however, such experiments require that the high molecular weight form of SV40 DNA be separated from the excess monomeric viral DNA found in the infected cells. Total infected cellular DNA was therefore sedimented through neutral sucrose gradients as described in materials and methods (section 2.11.2). The fast sedimenting DNA recovered from these gradients coelectrophoresed with bacteriophage λ DNA and was contaminated to only a slight extent with monomeric viral DNA (see Fig 4.6A).

This purified DNA was digested with a variety of restriction endonucleases and the digests were analysed using the Southern transfer

Legend to Figure 4.5

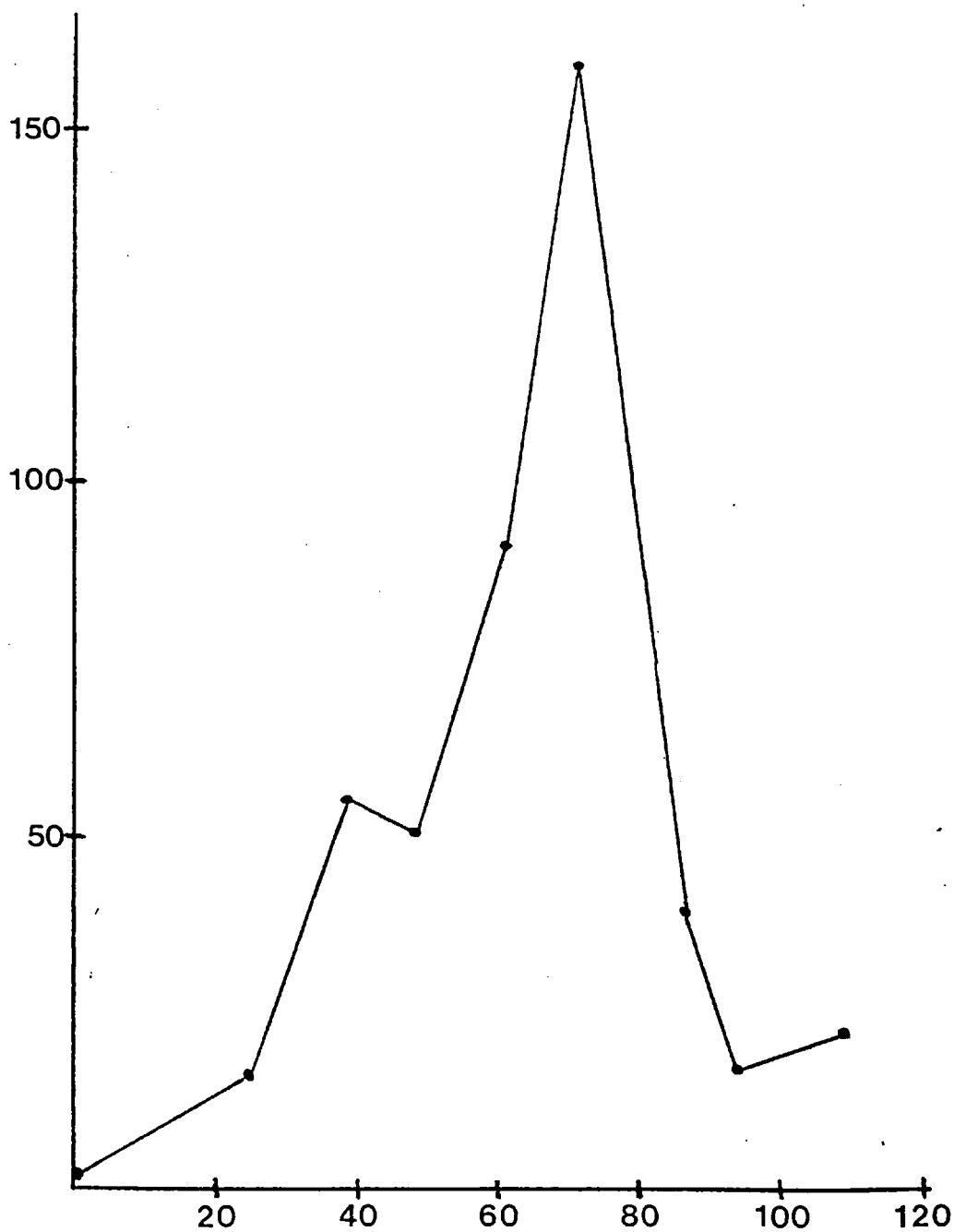
Virally induced cellular DNA replication

Confluent monolayers of Balb/c 3T3 cells were either mock infected or infected with wt830 (MOI=100). The cells were maintained on medium supplemented with 2% (v/v) FCS. At various times P.I., one plate each of mock infected and wt830-infected cells were pulse-labelled for 1h with 25 μCi of ^{3}H-thymidine. Immediately following labelling, total DNA was prepared from the cells as described (see section 2.3.3). The specific activities of the DNAs were determined.

The Y-axis: $\frac{\text{specific activity of wt830 infected DNA}}{\text{specific activity of mock infected DNA}}$

The X-axis: time (in hours P.I.) at which the 1h labelling commenced

Fig.4.5



hybridisation technique. Digestions with EcoRI KpnI or TaqI restriction endonucleases which cleave SV40 DNA once, resulted in the loss of the high molecular weight SV40 band and the appearance of a new band migrating at a position corresponding to monomeric linear SV40 DNA (Fig 4.6B). The great majority of the viral sequences were found at this position. When the autoradiograms were over-exposed, faint bands which could represent linear oligomers of SV40 DNA were seen. Despite the fact that restriction endonucleases were used in excess, these bands could be partial digestion products. Alternatively their presence could be due to small amounts of endonuclease resistant viral mutants contaminating the virus stock. This observation is incompatible with the high molecular weight form of SV40 DNA being due to the independent covalent integration of monomeric viral DNA into cellular DNA. However, this is the expected result if these high molecular weight SV40 molecules represent polymers of viral DNA organised as tandem 'head to tail' arrays. When the purified DNA was digested with SstI, a restriction endonuclease that cleaves cellular DNA but not SV40 DNA, the viral sequences remain at the exclusion limit of the gel despite the fact that the cellular DNA was efficiently digested as judged by ethidium bromide staining of the gel. The same result was obtained with SmaI and SalI, endonucleases which also do not cleave SV40 DNA; however, mammalian DNAs are partially modified against SmaI and contain relatively few SalI recognition sites; therefore, these enzymes do not cleave cellular DNA as efficiently as Sst I endonuclease. If the high molecular weight form of viral DNA results from covalent integration of viral monomers, this latter set of digestions should have produced a broad smear representing a large number of fragments, the sizes of which would

Legend to Figure 4.6

The structure of the high molecular weight form of SV40 DNA

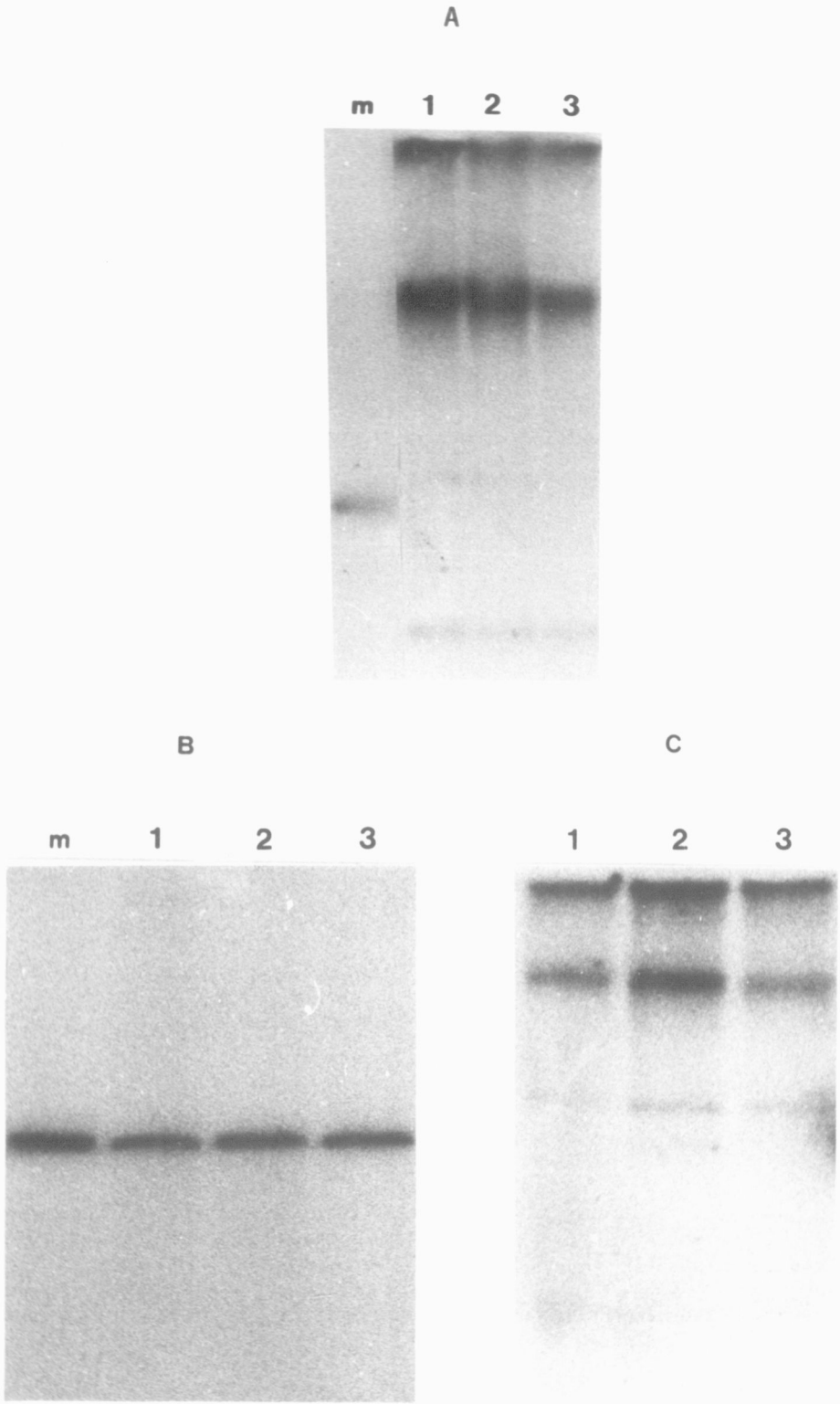
A, total DNA was prepared from mouse cells infected with wt830 (MOI=300), dl(pm)861 (MOI=300) and a 1:1 ratio of wt830: dl(pm)861 (total MOI=300) at 3 days P.I. The DNAs were sedimented through neutral sucrose gradients as described (see section 2.11.2) and the fast sedimenting DNAs were analysed by gel electrophoresis and transfer hybridisation. Track 1, wt830-infected; track 2, dl(pm) 861-infected; track 3, mixed-infected. Besides the intense band of high molecular weight form of SV40 DNA, small amounts of contaminating open-circular and superhelical monomeric SV40 DNA can also be seen (tracks 1, 2 and 3). The marker track (m) contain linear monomeric SV40 DNA.

DNA from wt830-infected cells purified by sucrose gradient sedimentation was digested with restriction endonucleases and analysed by agarose gel electrophoresis (0.6% gel) and transfer hybridisation.

B, the tracks are: 1, 0.3 μg DNA, EcoRI-digested; 2, 0.3 μg DNA KpnI-digested; 3, 0.3 μg DNA TaqI-digested; m, 0.1ng monomeric linear SV40 DNA.

C, the tracks are: 1, 0.3 μg DNA, SmaI-digested; 2, 0.3 μg DNA, SstI-digested; 3, 0.3 μg DNA, undigested.

Fig.4.6



depend on the spacings between the integrated viral sequences and the endonuclease recognition sites present in the adjacent cellular DNA. That the apparent mobility of the high molecular weight viral DNA was unaffected by digestion with enzymes like Sst I (Fig 4.6C) further confirms that these molecules represent polymers of tandemly repeated SV40 genomes arranged in a 'head to tail' array. Despite the fact that these molecules cannot be due to the independent integration of monomeric viral DNA, the possibility that some proportion of these molecules is integrated into the cellular DNA cannot be eliminated; since viral integration is nonspecific and since a heterogeneous population of infected cells is used, the experiments presented would not have detected the heterogeneous population of DNA fragments generated from the postulated integrated viral DNAs by restriction endonuclease digestion.

Therefore, in agreement with Collins and Sauer (1972), a novel form of viral DNA is found shortly following infection of mouse cells; the great majority of these molecules are probably linear as judged by their electrophoretic mobility on agarose gels and the fact that they copurify with cellular DNA by ethidium bromide, cesium chloride density gradient centrifugation. However, in contrast to the interpretation of Collins and Sauer, the restriction mapping data indicate that the great majority of these viral molecules represent 'head to tail' polymers of SV40 DNA and not integrated monomeric viral DNA.

4.5 Quantitation of the high molecular weight SV40 DNA

Transfer hybridisation experiments were performed to quantitate the amount of polymeric viral DNA present in the infected cells. In the experiment shown in Fig. 4.7, total infected cellular DNA and sucrose gradient purified high molecular weight DNA from wt 830-infected

Legend to Figure 4.7

Quantitation of the high molecular weight form of SV40 DNA

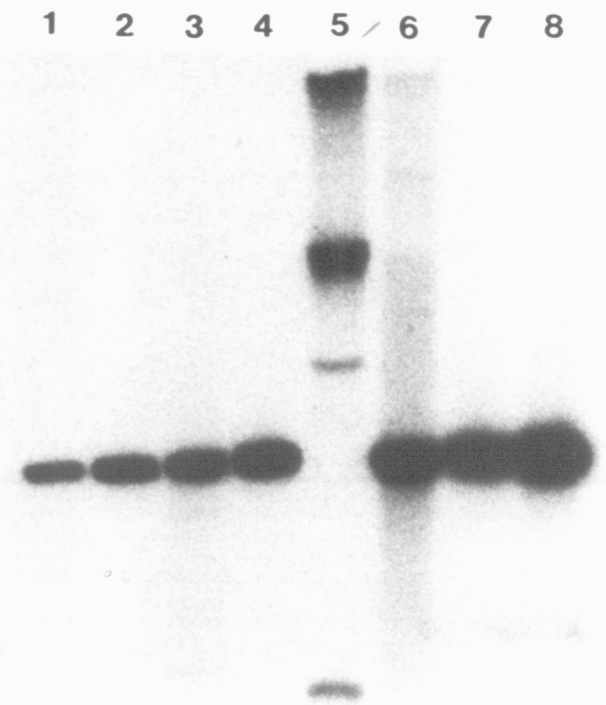
A, the DNAs were analysed by electrophoresis on an 0.7% (w/v) agarose gel. The tracks are: 1, 5pg monomeric linear SV40 DNA; 2, 10pg monomeric linear SV40 DNA; 3, 6.25 ng EcoRI-digested total DNA from wt830-infected cells (MOI=100, 3 days P.I.); 4, 20pg monomeric linear SV40 DNA; 5, 0.1 μg gradient purified high molecular weight DNA from wt830-infected cells (MOI=100, 3 days P.I.); 6, 0.1 μg of same DNA as in 5, EcoRI digested; 7, 40pg monomeric linear SV40 DNA; 8, 80pg monomeric linear SV40 DNA.

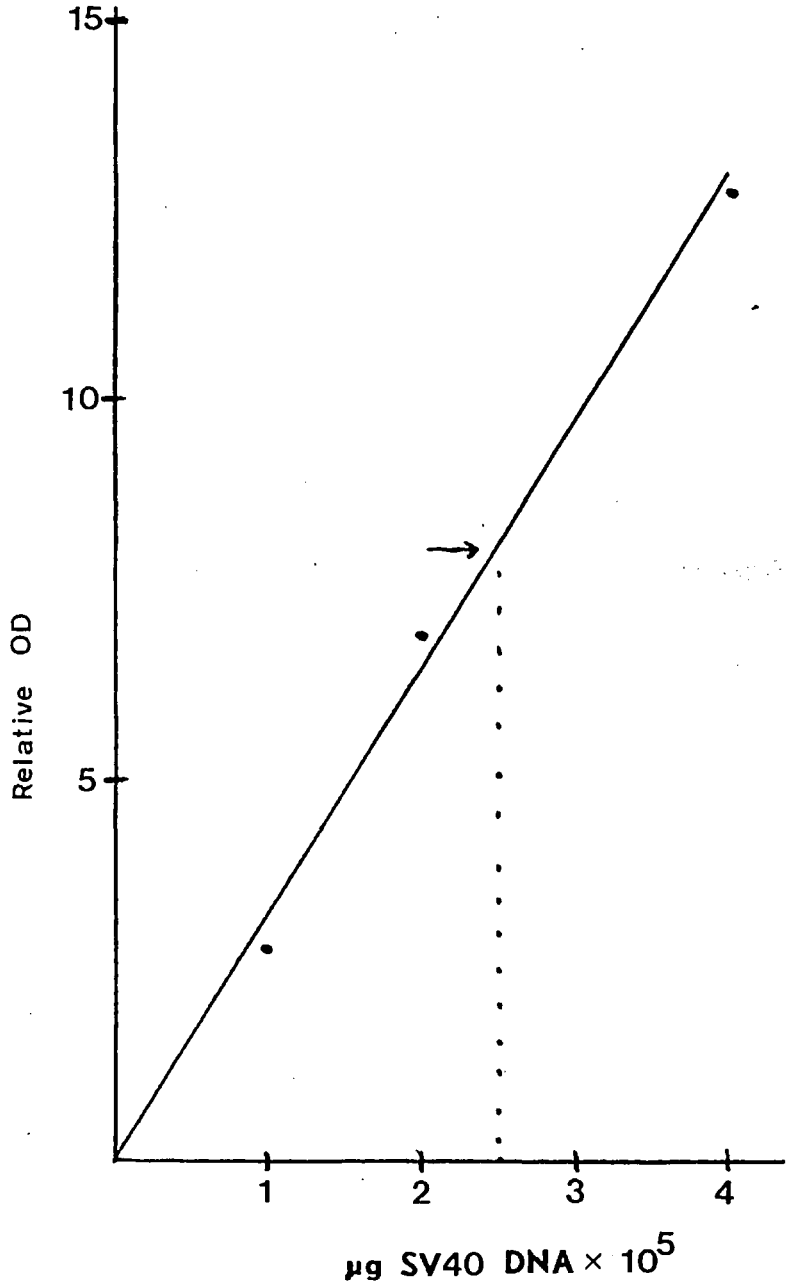
B, microdensitometer tracings of tracks 2, 4, 6 and 7 are shown on the right. The relative optical density of tracks 2, 4 and 7 are plotted against the amount of SV40 DNA present in these tracks to produce the standard curve shown on the left. The amount of SV40 DNA present in track 6 (arrow) can be extrapolated from the standard curve.

C, microdensitometer tracings of tracks 1, 2, 3 and 4 are shown. The amount of SV40 DNA present in track 3 is determined as described in B.

Fig. 4.7

A

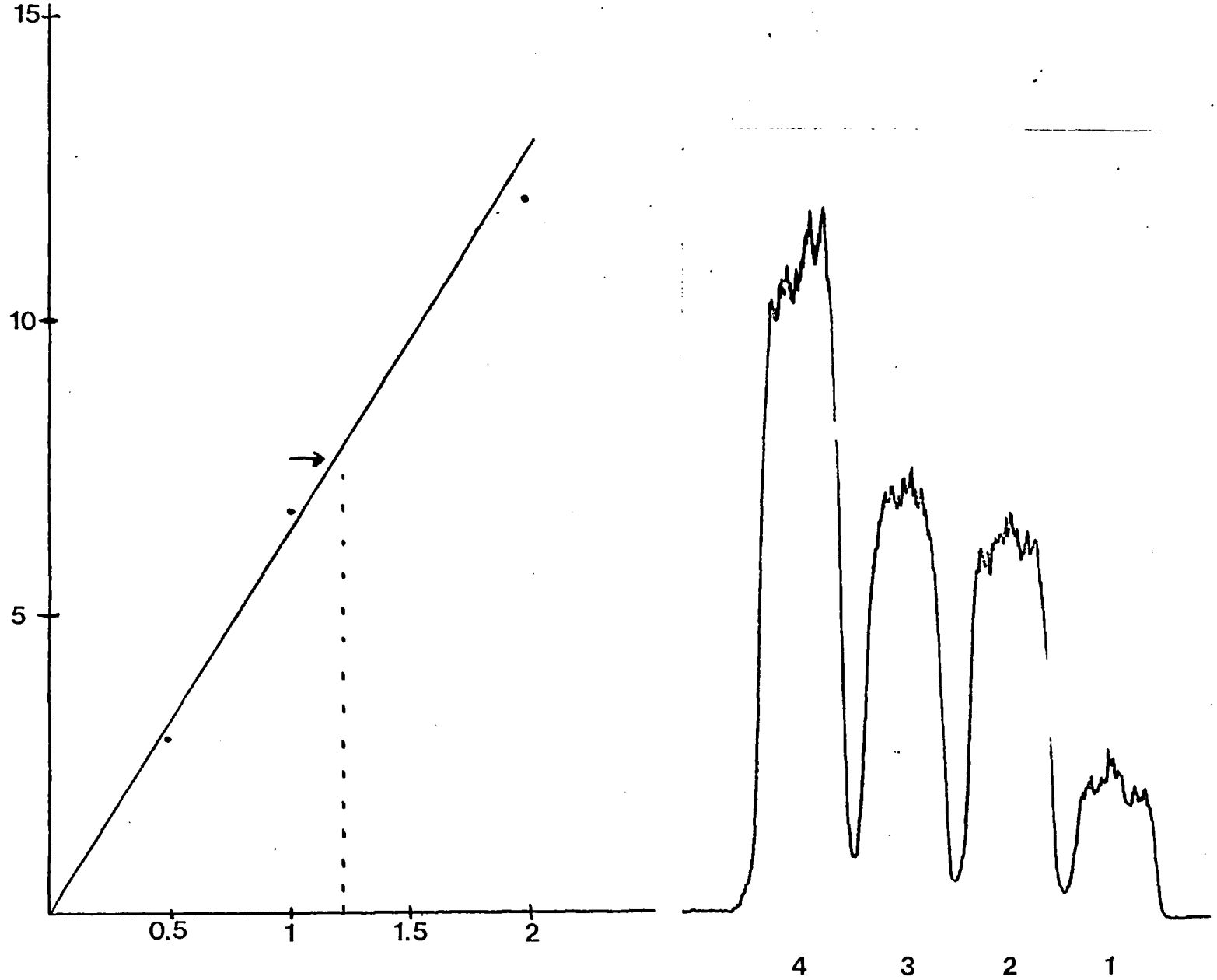




B

4.7

7 6 4 3 2



4.7

cells (MOI=100, DNA extracted at 3 days P.I.) were digested with EcoRI, which converts all of the viral DNA to monomeric linears. Also included in the experiment was a series of standards containing known amounts of linear SV40 DNA. The intensity of the autoradiographic response of both the samples and standards was quantitated by scanning the autoradiogram (Fig 4.7A) with a Joyce-Loebel microdensitometer. Calibration curves were constructed from the standards (Fig 4.7B,C)and the amount of SV40 DNA present in the samples was determined by using the standard curves as described by Lis et al. (1978). The quantitations were performed within the linear response range of the film. The results indicate that there are approximately 2 x 10^{-3} µg of SV40 DNA per µg of total infected cellular DNA and approximately 2.5 x 10^{-4} µg of SV40 DNA per µg of sucrose gradient purified high molecular weight DNA. The polymeric form thus accounts for approximately 12% of the total viral DNA.

4.6 The mechanism of synthesis of the polymeric viral DNA

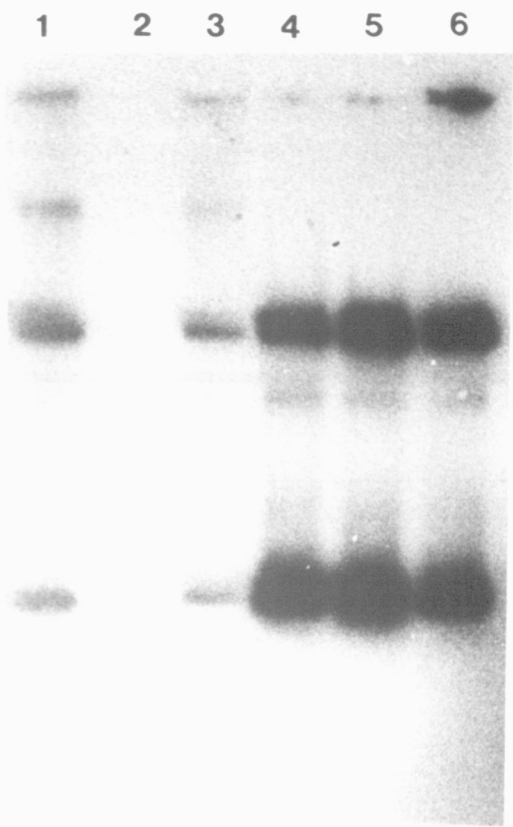
The tandemly arranged 'head to tail' polymers of SV40 DNA could have been synthesised either by recombination between the infecting monomeric viral DNA or by some kind of DNA replication mechanism. The possible involvement of DNA replication was tested directly by maintaining cells infected with wt 830 virus in the presence of varying concentrations of arabinosyl cytosine, an inhibitor of DNA replication (Manteuil et al., 1974). The results shown in Fig. 4.8 indicate that even at the lowest concentration used (1 µg per ml), this inhibitor prevented the appearance of polymeric viral DNA, while control tracks containing DNA from cells infected in parallel with the experimental cultures but maintained in the absence of the inhibitor were positive for the polymeric

Legend to Figure 4.8

The effect of arabinosyl cytosine on the formation of the high molecular weight form of SV40 DNA

Total DNA was isolated from wt830-infected cells at 3 days P.I. and analysed by electrophoresis on an 0.75% (w/v) agarose gel and transfer hybridisation. The observed bands are (from top to bottom) high molecular weight SV40 DNA (only in tracks 1 and 3), monomeric open-circular SV40 DNA, monomeric linear SV40 DNA and monomeric superhelical SV40 DNA. The conditions of infection were: track 1, MOI=100; track 2, MOI=2.5; track 3, MOI=25; track 4, MOI=300 plus 1 μg per ml of arabinosyl cytosine; track 5, MOI=300 plus 5 μg per ml of arabinosyl cytosine; track 6, MOI=300 plus 20 μg per ml of arabinosyl cytosine.

Fig.4.8



viral DNA. The result of this experiment is consistent with the view that the viral polymers are produced either directly by DNA replication or by a mechanism that is in some way dependent on host DNA synthesis.

To assess the role of recombination, confluent 3T3 cultures were coinfected with two mutant viruses with physically distinguishable genomes. Thus confluent 3T3 monolayers were coinfected with dl(pm)861 (Carbon et al., 1975), a viable mutant carrying a deletion of approximately 30 bp including the single Hpa II endonuclease site located at map position 0.725, and dlF884 (Shenk et al., 1976), a viable mutant carrying a deletion of approximately 250 b.p. including the single TaqI endonuclease site located at map position 0.57. Both mutants induce the synthesis of polymeric viral DNA in Balb/c 3T3 cells (see Figs. 4.6A and 4.12). A preliminary experiment was performed by infecting cells with varying MOIs of the two viral mutants and high molecular weight DNA was isolated from the infected cultures. Since the sizes of the mutant genomes differ by approximately 220 bp, the monomeric viral DNA could be resolved into two distinct bands. The ratio of the two types of viral DNA present in the high molecular weight fractions was determined by comparing the intensities of the two distinguishable monomeric linear viral DNA bands produced by EcoRI endonuclease digestion. In this way DNA preparations in which the ratio of the two mutant genomes was close to 1:1 were identified.

These DNA preparations were subjected to restriction endonuclease digestion, agarose gel electrophoresis and transfer hybridisation. Fig. 4.9A shows the result of TaqI endonuclease digestion of purified high molecular weight DNA from coinfected cells. In addition to bands

at the exclusion limit, resulting from TaqI digestion of dlF884 homopolymers, and at the position of monomeric, linear viral DNA, resulting from the digestion of dl(pm)861 homopolymers, there is also a band at the position of dimeric, linear SV40 DNA. If the same experiment was performed using a lower percentage agarose gel, additional bands extending to the exclusion limit of the gel are observed, which represent higher oligomeric forms of linear SV40 DNA (see Fig.4.9C). The presence of the dimeric and oligomeric bands even following digestion with a large excess of TaqI endonuclease can be explained by recombination between the two mutant viral genomes. The occurence of recombination is demonstrated more definitively by the experiment shown in Fig.4.9B and diagrammed in Fig.4.10. Purified high molecular weight DNA from cells coinfected with the two mutants was digested with both TaqI and HpaII endonucleases. As a consequence of recombination between the two mutant genomes in the longer interval (region 2) between the HpaII and TaqI endonuclease sites, the recombinant polymers will generate, following double digestion, not only the two parental linear genomes of 5.19 Kb and 4.97Kb, but also two recombinant fragments of 5.77Kb and 4.39Kb. Similarly, recombination in the shorter interval (region 1) between the two endonuclease sites will lead to the generation of recombinant fragments of 9.33Kb and 0.83Kb. The predicted recombinant fragments result solely as a consequence of recombination regardless of whether recombination occurred between monomeric or polymeric mutant genomes. The results shown in Fig.4.9B indicate the presence of the recombinant fragments of 9.33Kb, 5.77Kb and 4.39Kb in addition to the two parental genomes. The 0.83Kb recombinant fragment was electrophoresed off the end of this

Legend to Figure 4.9

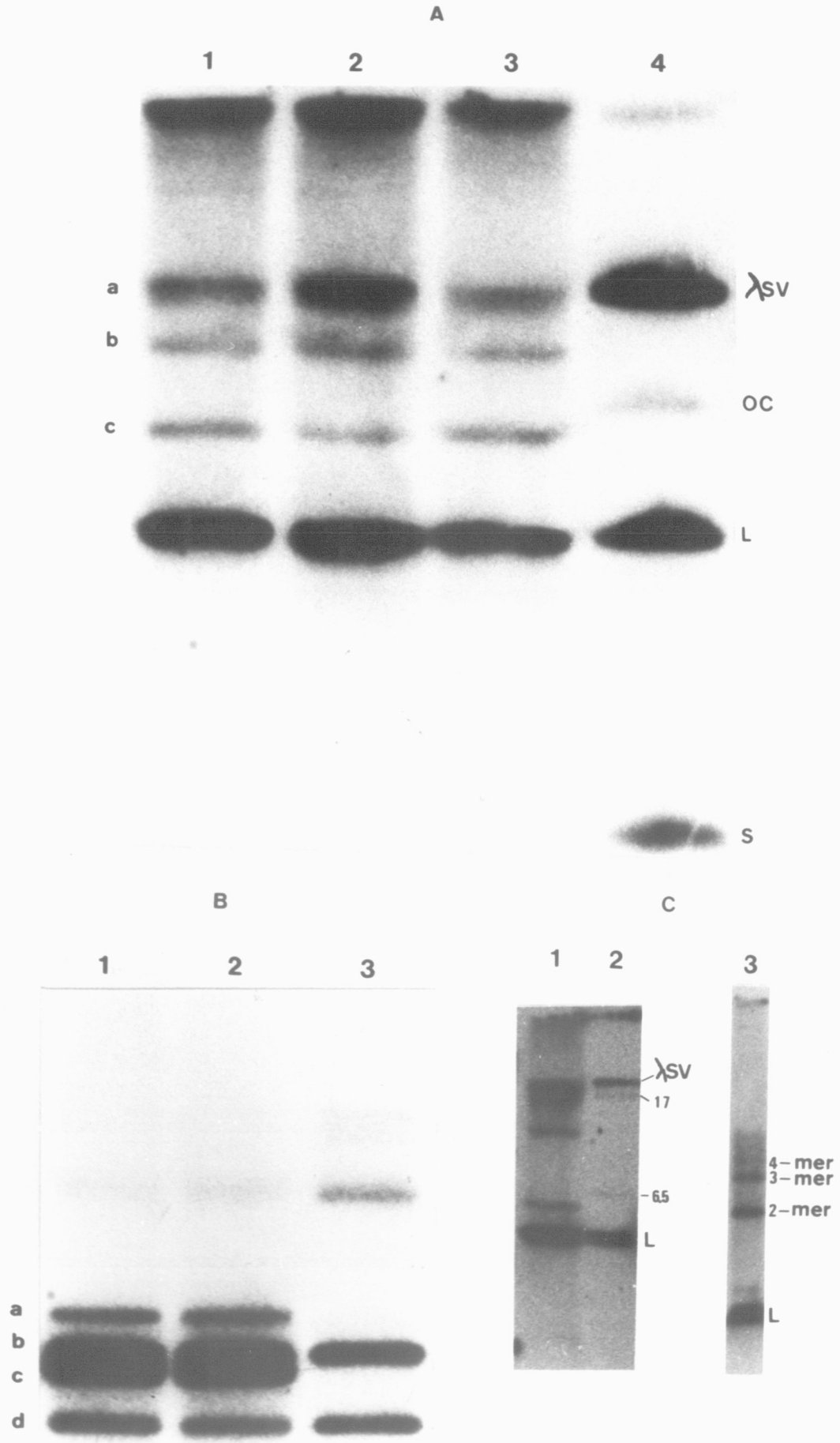
Recombination between infecting SV40 genomes

A, purified high molecular weight DNA from cells coinfected with dl(pm)861 and dlF-884 (MOI=100 for each virus) was digested with TaqI endonuclease and analysed by electrophoresis on an 0.75% (w/v) agarose gel and transfer hybridisation. Track 1, 1 μg DNA, 66h P.I.; track 2, 1 μg DNA, 69h P.I.; track 3, 0.5 μg DNA, 73h P.I.; track 4, markers; in order of decreasing mobility these are: DNA of a recombinant phage (λSV) containing SV40 cloned in λ(approximately 40Kb), monomeric open-circular SV40 DNA (OC), monomeric linear SV40 DNA (L) and monomeric superhelical SV40 DNA (S). The observed bands in tracks 1, 2 and 3 are: a, undigested dlF-884 homopolymers; b, dimeric linear SV40 DNA composed of one dl(pm)861 genome and one dlF-884 genome; c, contaminating monomeric open-circular dlF-884 DNA; monomeric linear dl(pm)861 DNA.
B, DNA prepared as in A above was digested with restriction endonucleases and analysed by electrophoresis on an 0.75% (w/v) agarose gel and transfer hybridisation. Track 1, 0.5 μg DNA, 69h P.I., digested with TaqI and HpaII; track 2, 0.5μg DNA, 73h P.I. digested with TaqI and HpaII; track 3, 0.5 μg DNA, 77h P.I. digested with TaqI only plus 30pg SV40 DNA digested with TaqI and HpaII. The bands in tracks 1 and 2 are: a, 5.77Kb recombinant fragment; b, 5.19Kb parental genome; c, 4.97Kb parental genome; d, 4.39Kb recombinant fragment; the 9.33Kb recombinant fragment can be seen as a faint band; the 0.83Kb recombinant fragment was run off

this gel.

C, purified high molecular weight DNA from cells coinfected with dl(pm)861 and wt830 was digested with HpaII, electrophoresed on an 0.6% (w/v) agarose gel and subjected to transfer hybridisation. In track 1, bands corresponding to monomeric dimeric, trimeric as well as unresolved higher oligomeric forms of linear SV40 DNA can be seen. The marker bands in track 2 are λSV (40Kb), the EcoRI fragments of cDM103 (17Kb and 6.5Kb) and monomeric linear SV40 DNA (L). When the same HpaII digested DNA was analysed on an 0.4% (w/v) agarose gel (track 3), higher linear oligomeric forms of SV40 DNA can be seen.

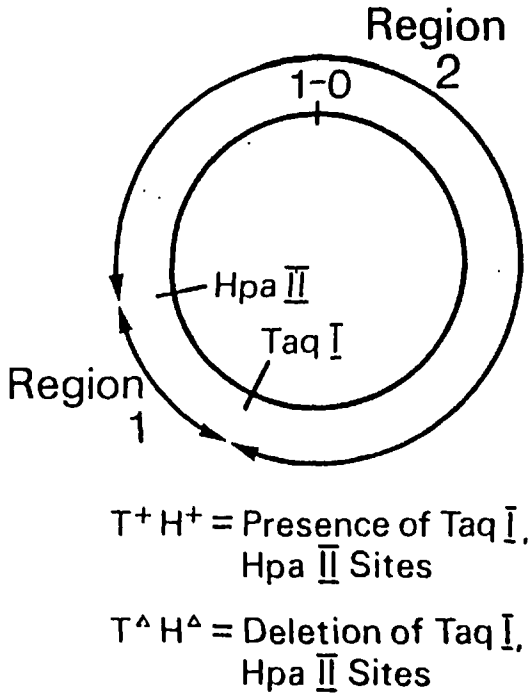
Fig. 4.9

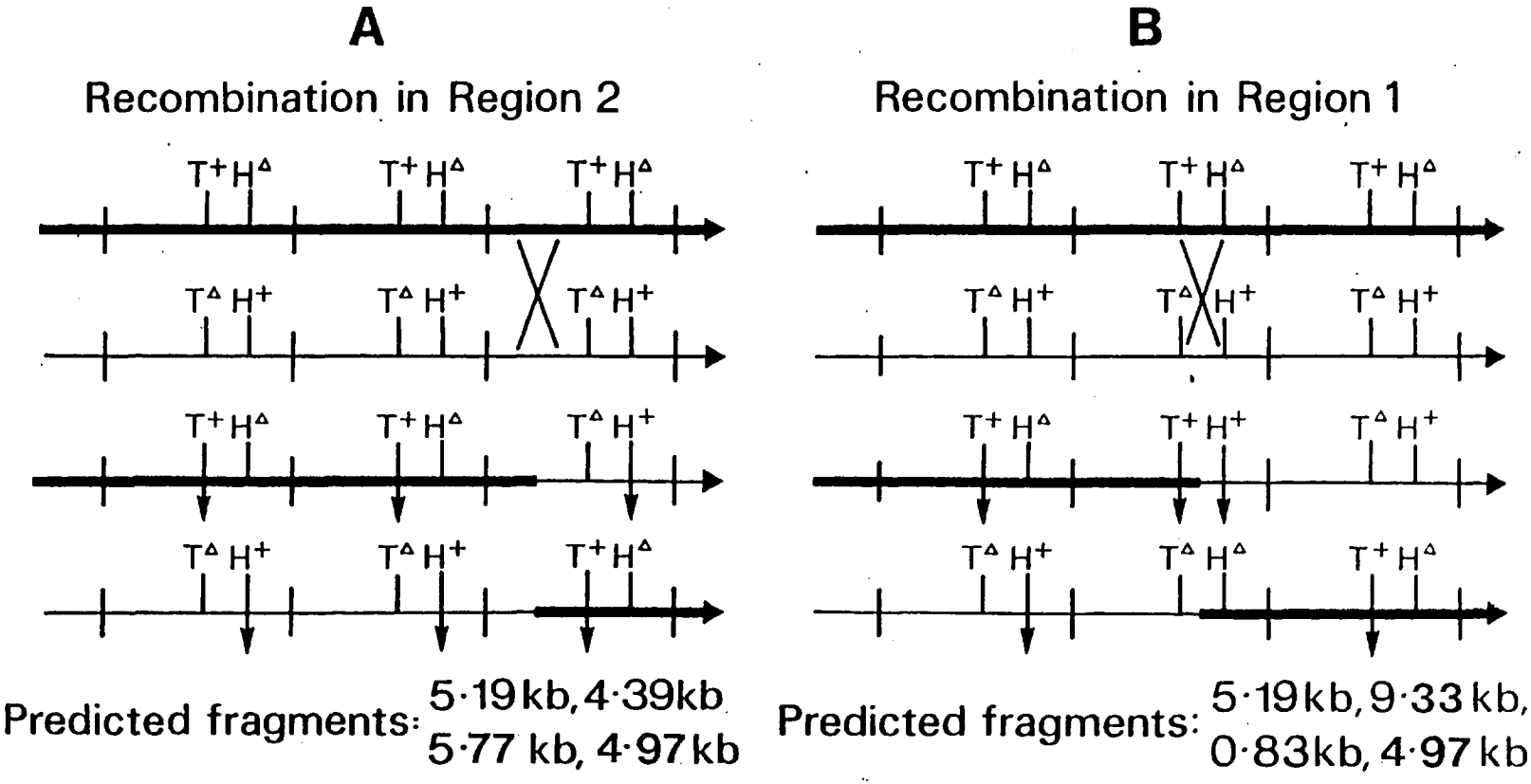


Legend to Figure 4.10

Schematic representation of recombination between infecting SV40 genomes

Fig.4.10


Region 2
1-0
Hpa II
Taq I
Region 1
T+ H+ = Presence of Taq I, Hpa II Sites
T^ H^ = Deletion of Taq I, Hpa II Sites


A
Recombination in Region 2
T+H^
T^H+
Predicted fragments: 5·19kb, 4·39kb, 5·77 kb, 4·97kb
B
Recombination in Region 1
Predicted fragments: 5·19kb, 9·33 kb, 0·83kb, 4·97 kb

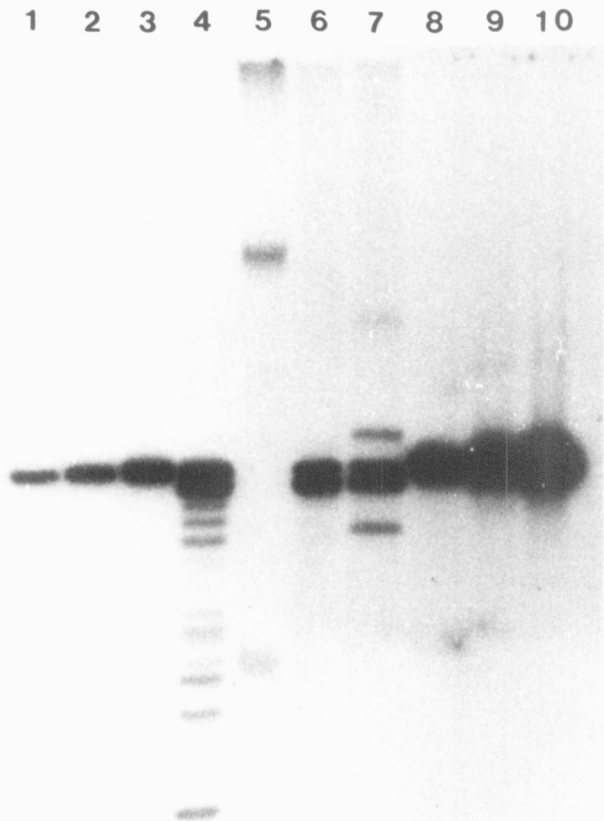
particular gel but has been observed in other experiments. Together these experiments clearly demonstrate the occurrence of recombination over both regions of the viral genome defined by the TaqI and HpaII sites (Fig.4.10). Unsuccessful attempts have been made to identify a time after infection when cells contain only homopolymers and show no evidence of recombination. There are at least two possible explanations for this failure. Firstly, the infections are asynchronous and secondly, the observed recombination events may be coupled to viral DNA replication.

The critical question is whether the observed amount of recombination is sufficiently high to account for the production of the polymeric molecules by a purely recombinational mechanism. A quantative transfer hybridisation experiment was performed, as previously described, using microdensitometry to determine the amount of recombination. High molecular weight DNA purified by sucrose gradient sedimentation from cells coinfected with dl(pm)861 and dlF884 was digested with EcoRI endonuclease or with HpaII and TaqI endonucleases. Microdensitometry performed using an underexposed autoradiograph of the track containing the EcoRI digested DNA indicates that the ratio of dl(pm)861 genomes to dlF884 genomes in the polymeric DNA was approximately 7:8. The data also indicate that there is no preferential site for recombination; the intensities of the recombinant fragments (Fig.4.11) reflect the relative sizes of regions 1 and 2 as defined by the HpaII and TaqI endonuclease sites (see Fig.4.10). In the experiment shown in Fig.4.11, 20% of the total polymeric DNA is in the 3.77 Kb and 4.39 Kb recombinant fragments. Given the relative sizes of regions 1 and 2, these fragments should represent 85% of the total recombination. If the polymers were synthesised purely by recombination then, assuming

Legend to Figure 4.11

Quantitation of recombination between infecting viral genomes

The DNAs were analysed by electrophoresis on an 0.75% (w/v) agarose gel and transfer hybridisation. The tracks are: 1, 6.25pg monomeric linear SV40 DNA; 2, 12.5pg monomeric linear SV40 DNA; 3, 25pg monomeric linear SV40 DNA; 4, 12.5ng of total DNA from cells coinfected with dl(pm)861 and dlF-884 (MOI=100 for each virus, 3 days P.I.) digested with EcoRI. The low molecular weight bands resulted from EcoRI* activity; 5, 0.125 μg of purified high molecular weight DNA (same infection as in 4), undigested; 6, 0.125 μg of the same DNA as 5, digested with EcoRI; 7, 0.125 μg of the same DNA as 5, digested with TaqI and HpaII; 8, 50pg monomeric linear SV40 DNA; 9, 0.1ng monomeric linear SV40 DNA; 10, 0.2ng monomeric linear SV40 DNA. Microdensitometer tracings of tracks 6 and 7 are shown.



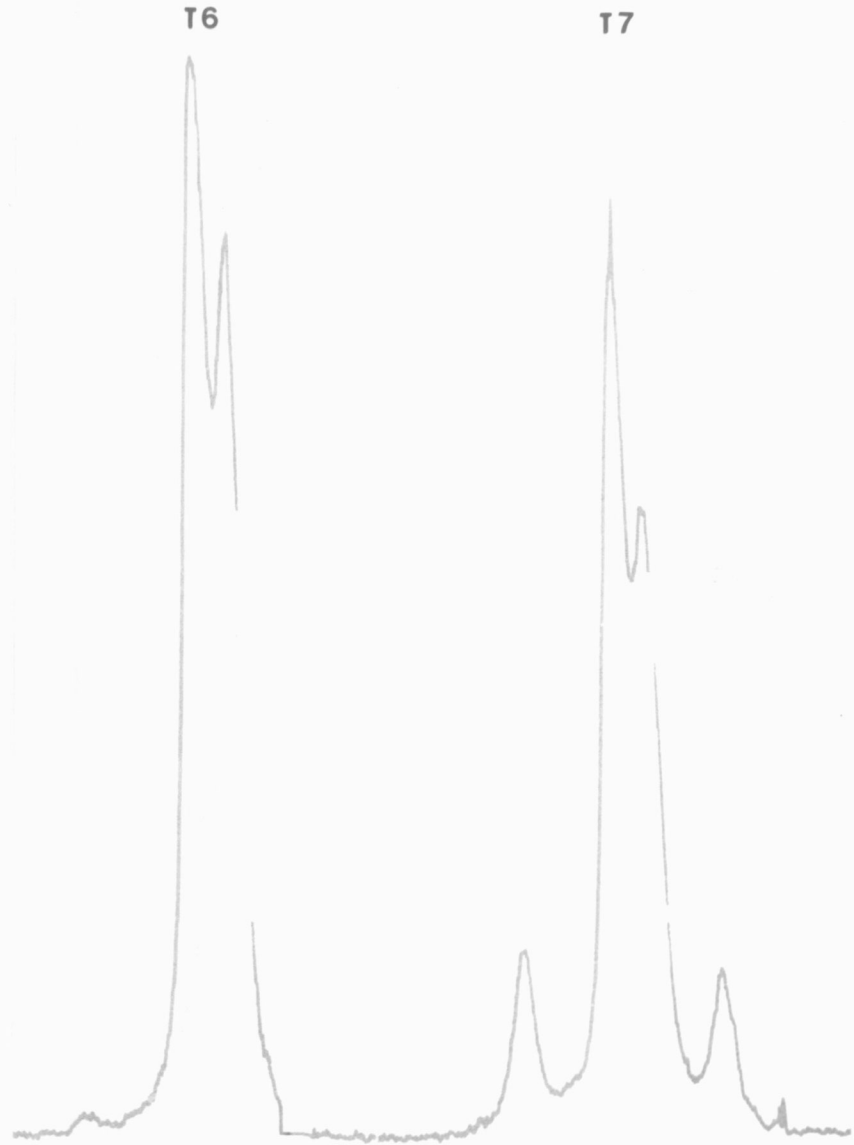


Fig 4.11

that the two mutants are not sequestered in separate pools, total reassortment of the markers would occur, giving a recombination frequency of 50%. However, the recombination frequency is only 24%. This result together with the observed inhibition of polymer synthesis by arabinosyl cytosine, supports the supposition that the viral polymers are synthesised by DNA replication and they then recombine with one another. However, what is demonstrated more definitively is that, in contrast to the exceedingly low levels of viral recombination observed following permissive infection (Goff and Berg, 1977; Rigby and Berg, 1978), the viral DNA molecules observed following the infection of nonpermissive mouse cells are recombinationally active.

4.7 Accumulation of viral polymers in tsA58- and dlF884 - infected cells

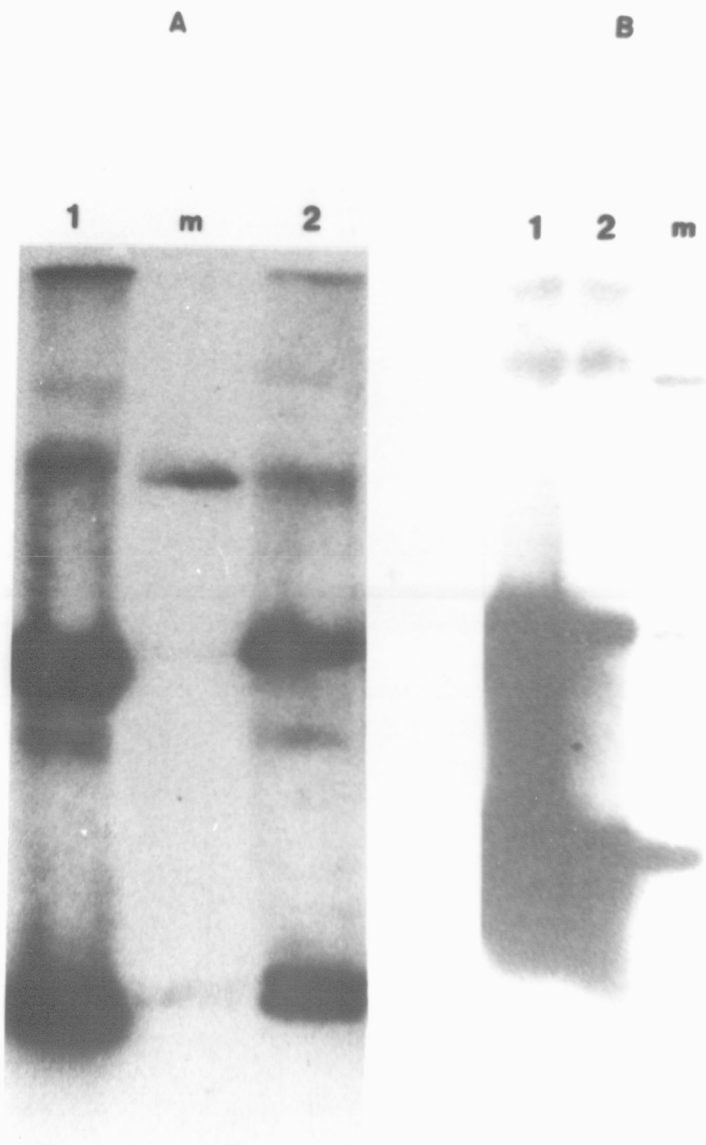
The accumulation of viral polymers has been examined in cells infected with tsA58, a mutant which produces a temperature sensitive large T-antigen and which cannot initiate theta-form replication in permissive cells at 40°C (Tegtmeyer, 1972), and in cells infected with dlF884, a mutant defective for the production of small t-antigen (Shenk et al., 1976). Both mutants induce the formation of viral polymers (Fig.4.12). These results suggest that neither of the early viral functions are required for the formation of the viral polymers. However, because mouse cells do not survive at temperatures above 40°C, the experiment with the tsA58 was performed at 39°C. Interpretation of the tsA experiment is therefore complicated by unanswered questions concerning the leakiness of the mutant. A definitive answer must await an analogous experiment with large T-deletion mutants. The implications of this experiment

Legend to Figure 4.12

Mouse cells infected by tsA58 and dlF-884 accumulated the high molecular weight form of SV40 DNA

A, total DNA isolated from dlF-884-infected cells (MOI=300, 77h P.I.) and wt830-infected (MOI=100, 77h P.I.) was electrophoresed on an 0.7% (w/v) agarose gel and subjected to transfer hybridisation. Track 1, 1 μg of dlF-884-infected DNA; track 2, 1 μg of wt830-infected DNA. The marker track (m) contain 5ng λSV DNA (40Kb) and 50pg monomeric superhelical and open-circular SV40 DNA.
B, total DNA from tsA58-infected cells (MOI=500, 76h P.I., 39°C) and wt830-infected cells (MOI=100, 76h P.I., 39°C) were electrophoresed on an 0.6% (w/v) agarose gel and subjected to transfer hybridisation. Track 1, 1 μg of tsA58-infected DNA; track 2, 1 μg of wt830-infected DNA. The marker track (m) contain 2.5ng λSV DNA (40Kb) and 50pg monomeric superhelical and open-circular SV40 DNA.

Fig.4.12



will be dicussed in Chapter 6.

4.8 Discussion

It has been shown that shortly after infection of Balb/c 3T3 cells by SV40, viral sequences can be found in association with high molecular weight cellular DNA. Restriction endonuclease mapping experiments have demonstrated that this high molecular weight form of viral DNA does not result from the integration of monomeric viral into cellular DNA but rather represents long, tandemly arranged 'head to tail' polymers of SV40 DNA. The accumulation of these viral molecules is dependent on DNA replication and can be blocked by inhibitors of DNA replication. In contrast to the lytic system, viral molecules found in nonpermissively infected cells are recombinationally active. These observations lead to a testable model for the mechanism of SV40 DNA integration which can account for some structural features of the integrated viral DNA found in stably transformed mouse cells. Such a model and its implications will be presented in Chapter 6.

Chapter 5

The development of a cosmid cloning system suitable for cloning integrated SV40 DNA

5.1 Introduction

Three major technical advances have popularised the use of λ cloning vectors. The first of these was the development of EK-2 certified vectors, i.e. the Charon vectors (Blattner et al., 1977) and λ gt WES (Leder et al., 1977). The second was the development of a high density in situ plaque hybridisation assay which allows the rapid screening of a large number of recombinant phage (Benton and Davis, 1977). Finally, the development of in vitro packaging systems (Hohn and Murray, 1977; Sternberg et al., 1977) allowed recombinant λ DNA to be introduced into bacterial cells at much higher efficiencies than methods based on transfection (Mandel and Higa, 1970). However, despite these advantages, there are also a number of limitations associated with the use of these vectors. Since only molecules of a certain size (79% to 109% of wild-type λ DNA) can be packaged into mature phage, and because the lytic functions of the phage are required for the propagation of the recombinants, only relatively small segments of DNA can be cloned (i.e. the upper limits for λ gt WES and Charon 4A are approximately14Kb and 20 Kb respectively). A second consequence of the requirement for phage functions is that it was difficult to build λ vectors suitable for cloning with certain restriction endonucleases. Finally, the existing EK-2 certified λ vectors which can be used to clone reasonably large DNA fragments (i.e. 20Kb) cannot be propagated in recA$^-$ bacteria. This inability to be propagated in a recombination defective host could result in the rearrangment or loss of the cloned

DNA, particularly if tandemly repeated sequences are included in the cloned segment (Cook and Hindley, 1979; Fried et al., 1979). All of these limitations could be circumvented if the recombinants could be propagated as plasmids rather than as phage.

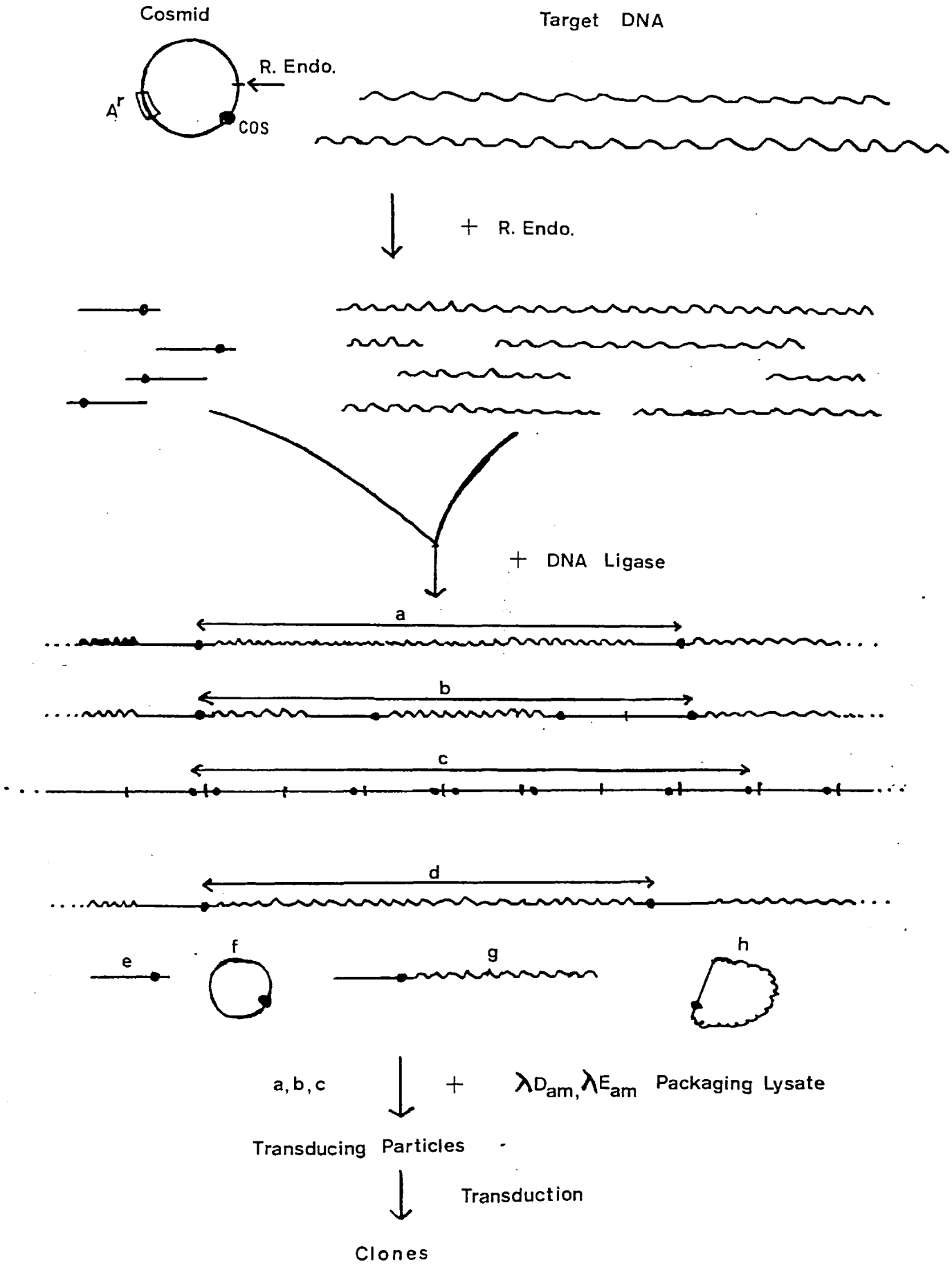
Cosmids, plasmids which contain the cohesive-end site (cos site) of mature λ DNA, which can be used as cloning vectors have been described (Collins and Hohn, 1978; Collins and Bruning, 1978). Previous studies have demonstrated that only a small region in the vicinity of the cos site, which acts as substrate for the packaging-dependent cleavages that produce the cohesive ends of mature λ DNA is required for recognition by the packaging system (Syvanen, 1974). It has also been shown that efficient packaging occurs only if two cos sites are separated by 35 to 50 Kb of intervening DNA (Kaiser et al., 1975; Becker et al., 1977). These results imply that a certain proportion of the molecules formed by the ligation of cosmid DNA to fragments of foreign DNA should serve as substrates for in vitro packaging. The transducing particles thus formed should be able to inject the packaged DNA into λ -sensitive E. coli cells Upon injection, the packaged DNA would presumably circularise via the cohesive ends (like λ DNA) and replicate as plasmid DNA without any further requirement for λ functions. The transductants can then be selected by the antibiotic resistance gene(s) encoded by the cosmid. Collins and his colleagues have shown that the sequence of events described above does indeed occur (see Fig.5.1). Since phage functions are not required, it should be possible to clone DNA fragments of up to 45Kb in size by using small cosmid vectors. Moreover, because only small regions of DNA are required for plasmid replication, it should be relatively simple to construct vectors for use with most

Legend to Figure 5.1

A schematic illustration of gene cloning with cosmid vectors

This figure is similar to one shown by Collins and Bruning (1978) but with several modifications. The cosmid, containing the annealed cohesive ends (cos) of phage λ, a selectable marker (e.g. antibiotic resistance, A^r) and a site for a restriction endonuclease (R.Endo.), is linearised by digestion with the restriction endonuclease. Target DNA to be cloned is similarly digested except in this case cleavage is only partially complete in order to generate large DNA fragments. The cosmid and target DNAs are mixed and ligated at high DNA concentration in order to increase the probability of forming molecules of types a, b, c and d. The molecular form most likely to be packaged and subsequently lead to the formation of transductants is form a, where the distance between the cos sites is between 35 to 50Kb with both cos sites in the same orientation. However, since in vitro packaging systems do not recognise cos sites at 100% efficiency, forms b and c can also be packaged. The transducing particles formed are used to inject their contents (i.e. forms a, b and c) into λ-sensitive E.coli strains. Upon injection, these molecules circularise and replicate as plasmids without expressing any λ functions. Bacteria containing these molecules can be selected using the selectable marker carried by the cosmid.

Fig.5.1


Cosmid
Target DNA
R. Endo.
A^r
COS
+ R. Endo.
+ DNA Ligase
a
b
c
d
e
f
g
h
a, b, c
+
$\lambda D_{am}, \lambda E_{am}$ Packaging Lysate
Transducing Particles
Transduction
Clones

restriction endonucleases. Finally, the recombinants constructed with a cosmid vector can be propagated in rec A$^-$ bacterial hosts. Since a high density in situ colony hybridisation assay has been developed (Hanahan and Meselson, 1980) which can presumably be used to screen recombinants derived from cosmid vectors, it would appear that cosmids, while potentially possessing all of the advantages, are not constrained by the limitations associated with the λ cloning systems.

The experiments presented in this chapter are concerned with: (1) the construction of a GMAG-approved, disabled cosmid vector and its derivatives; (2) the analysis of recombinants derived from this vector; (3) the refinement of the existing cosmid cloning technology; (4) the stability of cloned 'head to tail' polymers of SV40 DNA propagated in a recA$^-$ bacterial host; (5) the attempt to clone integrated SV40 DNA from mouse cells shortly following viral infection. The reason why it was necessary to build these vectors was because the existing cosmid vectors did not satisfy our requirements: (1) they were not biologically disabled; (2) they did not contain an appropriate restriction endonuclease site which is not present on SV40 DNA; (3) they tended to be large in size.

5.2 Construction of cosmid vector Homer I

pAT153 (Twigg and Sherrat, 1980) is a plasmid derived from pBR322 (Bolivar et al.,1977b) by the deletion of two HaeII endonuclease fragments (M. Scott, personal communication) spanning bp 1646 to 2351 in the pBR322 sequence (Sutcliffe, 1978). As a result of this deletion, pAT153 loses its ability to be mobilised by conjugative plasmids even in the presence of a helper plasmid and is therefore a safe vector in terms of biological containment. The

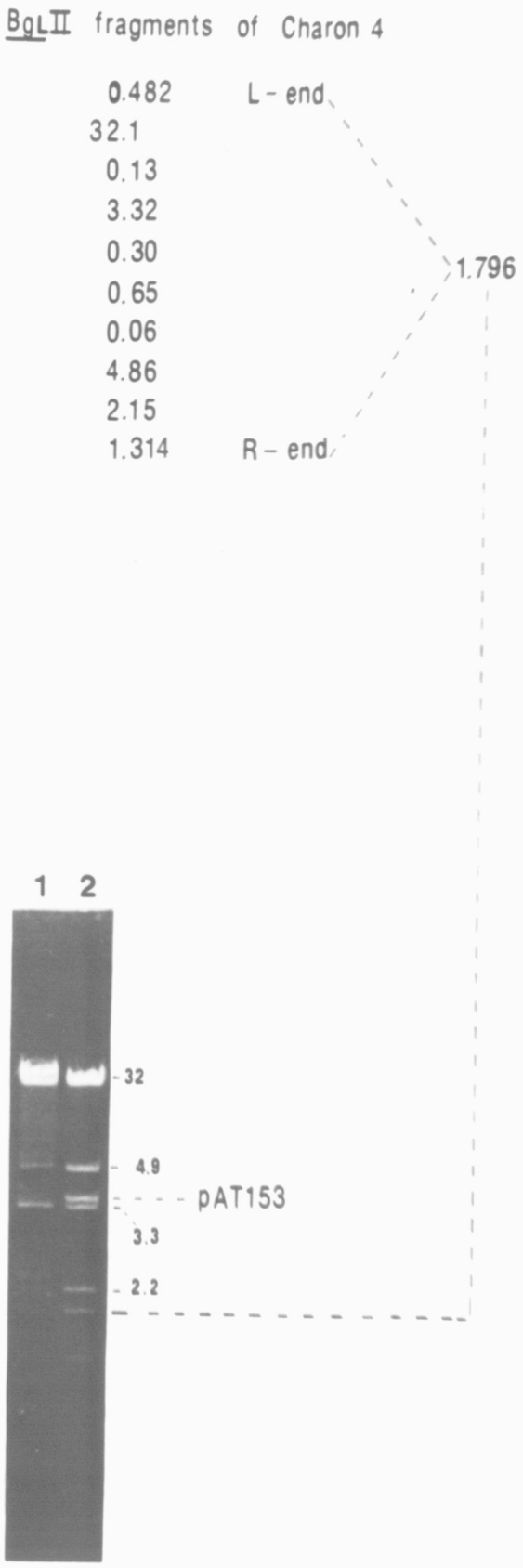
cosmid vector Homer I was constructed by cloning the cohesive-end site (*cos* site) of the λ vector Charon 4 (Blattner *et al.*, 1977) into pAT153. The structures of Charon 4 and pAT153 are given in Fig.5.2. The strategy employed for the construction is as follows: Charon 4 DNA was incubated to anneal the *cos* ends before digesting to completion with *Bgl*II endonuclease. Charon 4 DNA contains nine *Bgl*II endonuclease sites, two of which are 482 bp and 1314 bp from the left and right hand ends of Charon 4 respectively. Since *Bgl*II and *Bst*I endonucleases share the same central tetranucleotide in their recognition sequences, they generate identical cohesive ends which can be ligated to each other. So, an excess of the *Bgl*II fragments of Charon 4, which include a 1796 bp fragment containing the annealed *cos* site, was ligated to pAT153 linearised by digestion with *Bst*I endonuclease. The ligated DNA was redigested with *Bst*I endonuclease and treated with calf intestinal phosphatase. Since the ligation of a *Bst*I cohesive end to a *Bgl*II cohesive end results in the loss of both restriction endonuclease recognition sites, *Bst*I-*Bgl*II joints are refractory to digestion by both enzymes. Therefore, treating the ligated DNA with *Bst*I endonuclease and calf intestinal phosphatase should specifically linearise and dephosphorylate recyclised vector molecules, thereby effectively removing vector background. Competent cells of *E.coli* strain HB101 were transformed with the above DNA. Twenty four ampicillin resistant colonies were tested for tetracycline resistance. They were all tetracycline sensitive. (Because the *Bst*I endonuclease site of pAT153 is located in the gene coding for tetracycline resistance, the insertion of foreign DNA at this site will result in the loss of this function and therefore give a tetracycline sensitive phenotype.)

Legend to Figure 5.2

Cloning the *cos* site of Charon 4 into pAT153.

The *Bgl*II fragments of Charon 4 DNA are listed in topological order in A. Charon 4 DNA was annealed and digested with *Bgl*I endonuclease. pAT153 was linearised with *Bst*I endonuclease. The two DNAs were ligated (see Chapter 2, section 2.12.2 for conditions) at a final total DNA concentration of 50 μg per ml with a molar ratio of *Bgl*II termini to *Bst*I terminiof 8:1. Aliquots of the ligation mixture were fractionated on a 0.8% (w/v) agarose gel before (track 2, B) and after ligation (track 1, B). The 1.8Kb *Bgl*II fragment contains the annealed *cos* site of Charon 4.

Fig.5.2


BglII fragments of Charon 4
0.482
L - end
32.1
0.13
3.32
0.30
1.796
0.65
0.06
4.86
2.15
1.314
R - end
1 2
32
4.9
pAT153
3.3
2.2

These transformants were screened using the Grunstein and Hogness Colony screening method. The hybridisation probe used was generated by first tailing Charon 4 DNA using terminal transferase and α^{32}P dATP. The tailed DNA was then digested to completion with Bgl II endonuclease to generate a hybridisation probe which was specific for the terminal BglII fragments of Charon 4. The result of the screen was not totally convincing because all of the colonies showed some hybridisation, presumably due to the presence of cryptic λ sequences in the E.coli DNA (Kaiser and Murray, 1979). However, isolate 13 gave one of the stronger signals. Twelve of the transformants were screened for the presence of the 1796 bp BglII fragment of Charon 4 using the size of the plasmid DNAs as the sole criterion. Plasmid DNAs were prepared from the transformants using the mini cleared lysate method. These DNAs were digested with EcoRI endonuclease and fractionated by agarose gel electrophoresis. The results shown in Fig.5.3 indicate that isolate 13 contains a plasmid of approximately the size expected for a chimaeric molecule formed between pAT153 and the 1796 bp BglII fragment of Charon 4. Restriction endonuclease mapping (M. Scott, personal communication) confirmed this expectation and the structure of Homer I is shown in Fig.5.4 (courtesy of M. Scott).

5.3 Homer I can be used as a cloning vector in combination with the λ *in vitro* packaging reaction

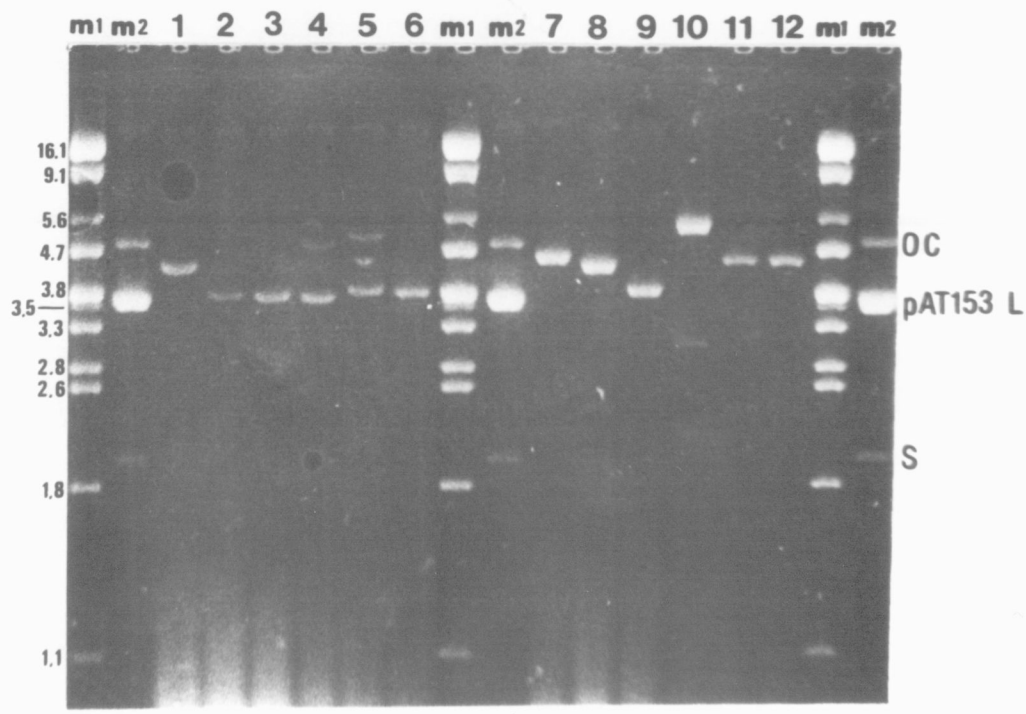
Since Homer I contains the cos site of Charon 4, one would expect this plasmid to possess the properties described for cosmid vectors in the introductory section of this chapter. This was shown to be the case. Drosophila genomic DNA partially digested with EcoRI

Legend to Figure 5.3

Screening for Homer I by gel electrophoresis of mini cleared lysates

The ligated DNA (shown in Fig.5.2B) was heated to 65°C for 10 mins, digested with BstI and treated with an excess of calf intestinal phosphatase (see section 2.12.1) to reduce vector background (see text). The DNA was used to transform competent cells prepared from E.coli strain HB101 as described (section 2.12.5). DNA was prepared from twelve independent ampicillin resistent, tetracycline sensitive isolates using the mini cleared lysate method (see section 2.3.1). These DNAs were digested with EcoRI and fractionated on an 0.8% (w/v) agarose gel (tracks 1 to 12). The marker tracks contain: m1, BstI/EcoRI cleaved λ DNA; m2, a partial EcoRI digest of pAT153 containing open-circular (OC), linear (L) and supercoiled (S) plasmid. Track 10 (isolate 13) contains the putative clone for the 1.8Kb BglII fragment of Charon 4. The faint band running below the 5.5Kb band is probably undigested supercoils.

Fig.5.3


m1 m2 1 2 3 4 5 6 m1 m2 7 8 9 10 11 12 m1 m2
16.1
9.1
5.6
4.7
3.8
3.5
3.3
2.8
2.6
1.8
1.1
OC
pAT153 L
S

Legend to Figure 5.4

The structures of pBR322, pAT153 and Homer I.

This figure was prepared by M.R.D. Scott and taken from Chia, Scott and Rigby, in preparation.

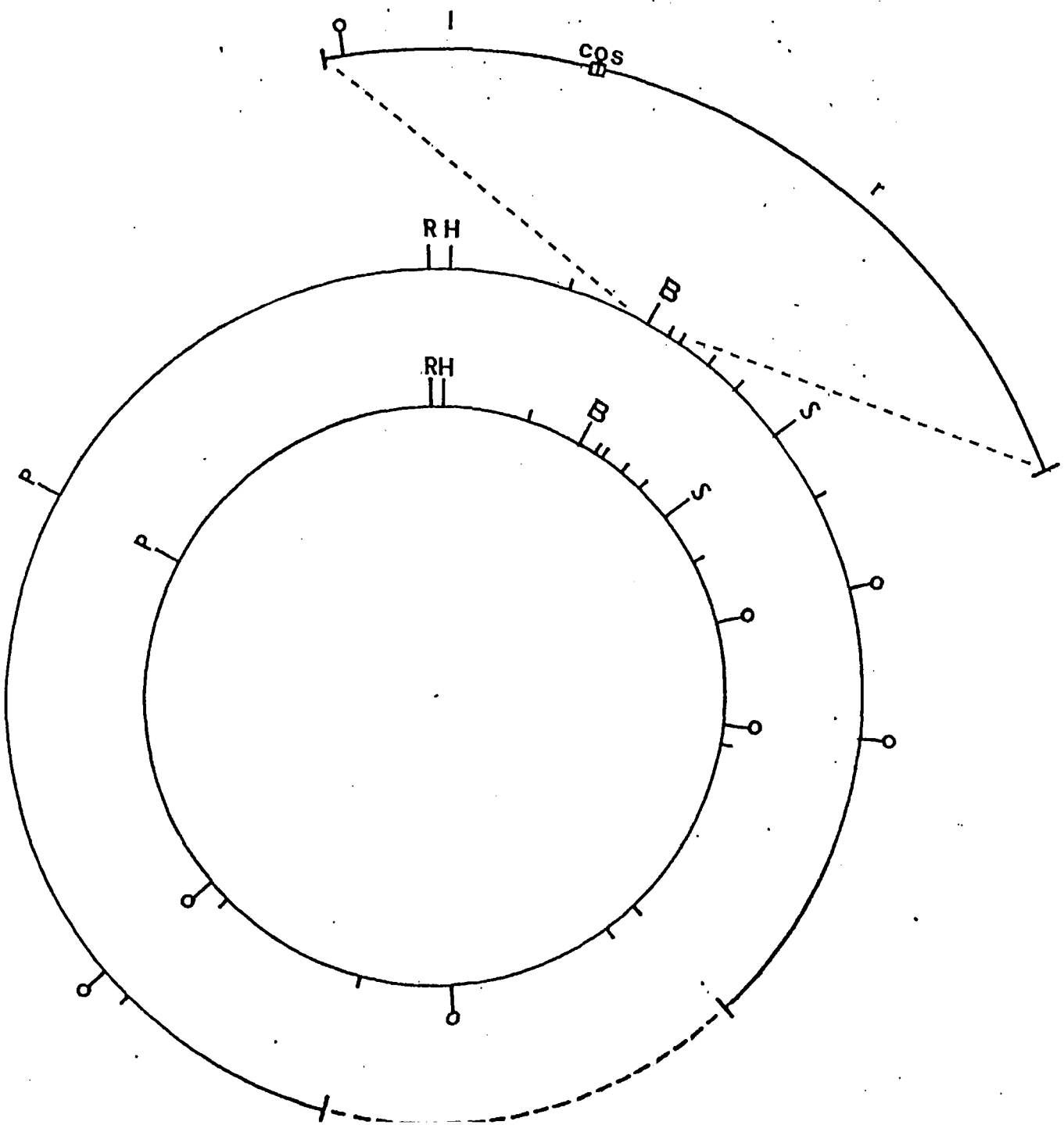


Fig.5.4

Map of Homer I

R	=	*Eco*RI
H	=	*Hind*III
B	=	*Bam*HI
S	=	*Sal*I
P	=	*Pst*I
l	=	*Hae*II
	=	*Bgl*I
cos	=	λ *cos* sites
l	=	left terminal *Bgl*II fragment of Charon 4A
r	=	right terminal *Bgl*II fragment of Charon 4A

The inner circle shows the map of pBR322.
The outer circle shows the map of pAT153.
The dashed line indicates the two *Hae*II fragments deleted during the construction of pAT153 from pBR322.
Circularised Ch4A DNA was digested with *Bgl*II and the 1.8Kb fragment containing the *cos* site was inserted into the *Bam*HI site of pAT153. The λ insert does not contain sites for *Eco*RI, *Hind*III, *Bam*HI, *Sal*I, *Pst*I, *Kpn*I or *Sst*I. We do not know if the insert contains *Hae*II sites.

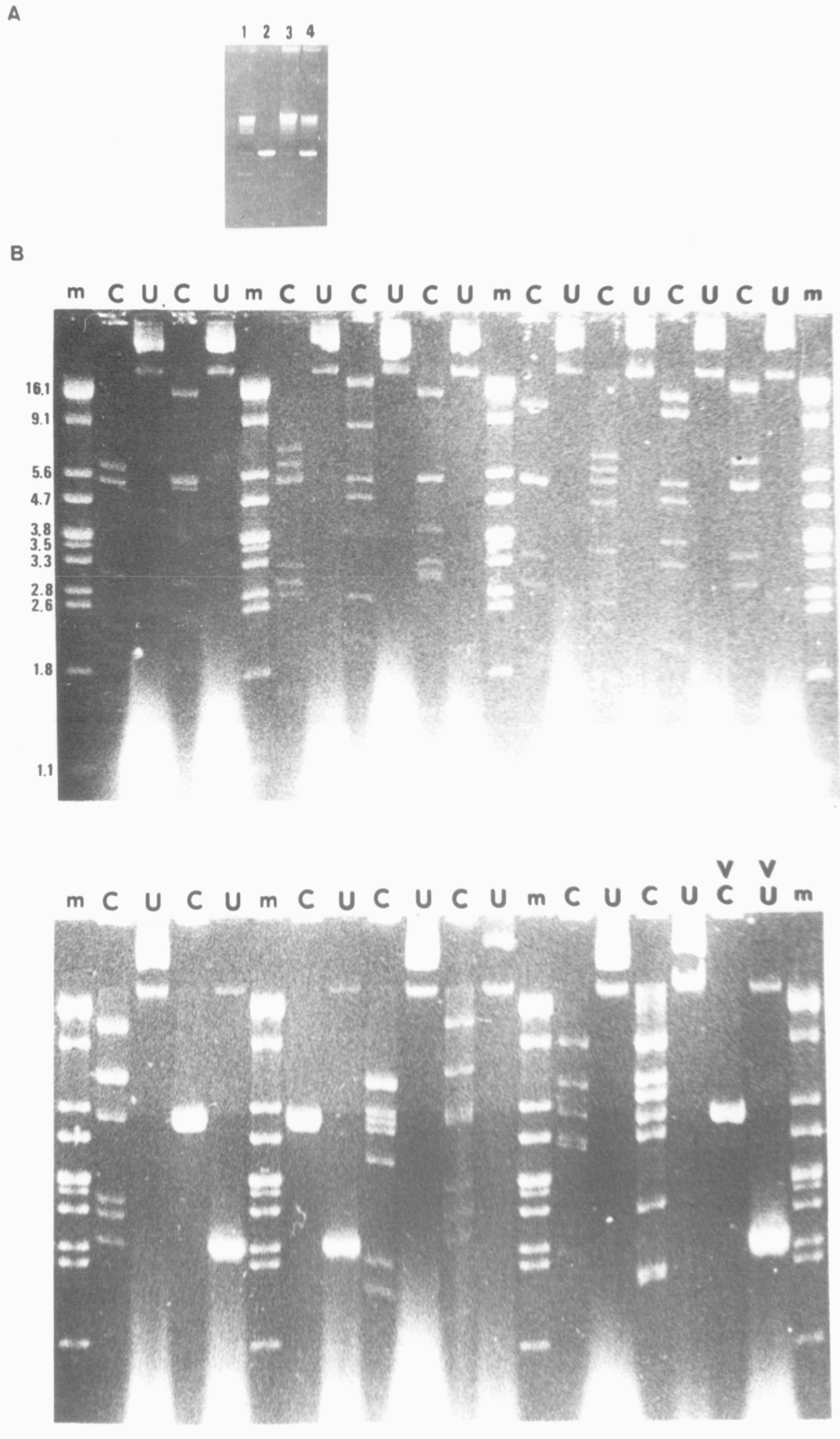
endonuclease (mean length of approximately 30 Kb) was ligated at high DNA concentration with Homer I DNA linearised with EcoRI endonuclease. A parallel control ligation reaction containing Homer I DNA linearised with EcoRI endonuclease was included in the experiment. The ligation reaction mixtures were assayed by gel electrophoresis before and after the reaction. Successful reactions were evidenced by the increased amounts of DNA at the exclusion limit of the gel, the disappearance of the monomeric linear Homer I band, and the generation of a series of faint Homer I oligomers (see Fig.5.5A). Aliquots of successful ligation mixtures were packaged using the λ *in vitro* packaging system described by Hohn and Murray (1977). Following the transduction of fresh overnight cultures of *E.coli* strain HB101 with the packaged DNA, ampicillin resistant bacteria were observed. DNAs prepared from independent isolates using the mini cleared lysate method were subjected to gel electrophoresis either undigested or following digestion with EcoRI endonuclease. The results (Fig.5.5B) show that a majority of the transductants contain large plasmid DNA molecules. Following EcoRI endonuclease digestion, each of these plasmids gave rise to a unique pattern of DNA fragments; however, each set of DNA fragments appears to include a band of approximately the same size as linear Homer I DNA. These results indicate that large concatemeric DNA generated by ligation of *Drosophila* genomic DNA to Homer I DNA will serve as substrates for the *in vitro* λ packaging reaction. A proportion of the transductants (the figure varied between 10% to 40% in a number of experiments) were, however, vector background, i.e. they contain only monomeric Homer I DNA following EcoRI endonuclease digestion. It is clear that the vector DNA present in these background transductants is monomeric Homer I supercoils

Legend to Figure 5.5

Homer I can be used as a cosmid cloning vector.

Drosophila genomic DNA was partially cleaved with *Eco*RI endonuclease to generate fragments with an average length of approximately 30Kb. Homer I was linearised with *Eco*RI endonuclease. Linear Homer I DNA was ligated either by itself (at 100 μg per ml) or with the *Drosophila* DNA (170 μg per ml of Drosophila DNA and 130 μg per ml of Homer I DNA) as described (section 2.12.2). A, aliquots of the ligation mixtures were electrophoresed either before (tracks 2 and 4) or after (tracks 1 and 3) the ligation reaction. The ligated DNA was packaged *in vitro* (see section 2.12.6) using packaging lysates which packed λDNA at an efficiency of 2×10^6 pfu per μg. Transduction was performed as described (section 2.12.6). The cloning efficiency in this experiment was 2.5×10^3 colonies per μg of *Drosophila* DNA. B, DNA was prepared from sixteen independent isolates using the mini cleared lysate method and electrophoresed either undigested (U) or digested with *Eco*RI (C). One isolate obtained from religated Homer I DNA was similarly treated ($^{V}_{C}$ and $^{V}_{U}$). The marker tracks (m) contain *Bst*I/*Eco*RI cleaved λDNA.

Fig.5.5


A
1 2 3 4
B
m C U C U m C U C U C U m C U C U C U C U m
16.1
9.1
5.6
4.7
3.8
3.5
3.3
2.8
2.6
1.8
1.1
m C U C U m C U C U C U m C U C U C V C V U m

rather than supercoils of polymeric forms of Homer I. Since λ *in vitro* packaging systems will only efficiently package DNA molecules in which *cos* sites are located 35-50Kb apart in the same orientation (Collins and Hohn, 1978), this observation is very surprising. It seems likely that the cells were originally transduced by packaged polymeric Homer I molecules, 35Kb to 50Kb in size, which are unstable in *E.coli* strain HB101 and break down to monomers.

5.4 Optimising the efficiency of cloning

While the experiment shown above clearly demonstrated that Homer I can be used as a cosmid vector, the efficiency of cloning was low (approximately 2×10^3 recombinants per ug of *Drosophila* DNA). In order to improve the cloning efficiency, experiments were performed to examine several of the parameters involved in the cloning procedure. The first problem was the low efficiencies obtained from the packaging extracts prepared by the method of Hohn and Murray (1977). In my hands, these packaging lysates rarely packaged at efficiencies above 5×10^6 pfu per ug λ DNA. This problem was alleviated when the packaging lysates were prepared by the method described in materials and methods (section 2.12.6); such lysates package λ DNA at efficiencies of up to 2×10^8 pfu per μg of λ DNA. Using these high efficiency packaging lysates, an experiment was performed to optimise for the ratio of vector to target DNAs in the ligation reaction. The results shown in Table 5.1 clearly indicate that the highest efficiencies, in terms of recombinants per μg of target DNA, are obtained when vector DNA is used in large excess for the ligation reaction. Since target molecules can only be packaged when they are flanked on both sides by Homer I molecules, this is the expected result because conditions

TABLE 5.1

The Effect of Vector to Target Ratio on Cloning Efficiency

Molar Ratio of vector: target	µg/ml Homer I	µg/ml 40Kb *Drosophila* DNA	Recombinants per µg of Homer	Recombinants per µg of *Drosophila* DNA
1:3	12	288	3.0×10^4	1.3×10^3
1:1	33	267	7.5×10^4	9.3×10^3
3:1	82	268	1.2×10^5	4.3×10^4
9:1	159	141	6.1×10^4	6.9×10^4
Vector only	159	-	2.2×10^3	-

Partial *EcoRI* endonuclease cleaved, sucrose gradient fractionated (see Chapter 2 section 2.11.1) *Drosophila* embryonic DNA with an average length of approximately 40Kb was ligated to Homer I DNA linearised with *EcoRI* endonuclease (see Section 2.12.2). The final total DNA concentration in the ligation reactions was 300µg per ml, however, the molar ratio of Homer I DNA to *Drosophila* DNA was varied. The ligation mixtures were *in vitro* packaged, the *in vitro* packaged transducing particles were used to infect fresh cultures of *E. coli* HB101 and ampicillin resistant colonies were selected and scored (see section 2.12.6).

which are optimal for the production of such concatemers would be expected to produce the highest efficiencies of cloning. The frequency with which religated vector DNA was packaged *in vitro* was much lower than when target DNA was included in the ligation. This is to be expected because vector polymers generated by ligation can be packaged only if a number of *cos* sites on contiguous Homer I molecules are missed by the *in vitro* packaging system. Despite the fact that *cos* sites are not recognized at 100% frequency, the frequency of this occurrence should not be high. This experiment shows that when high vector to target ratios are used in the ligation reactions, high cloning efficiencies can be obtained. Moreover, the result with the religated vector DNA implies that vector background should occur at a reasonably low frequency.

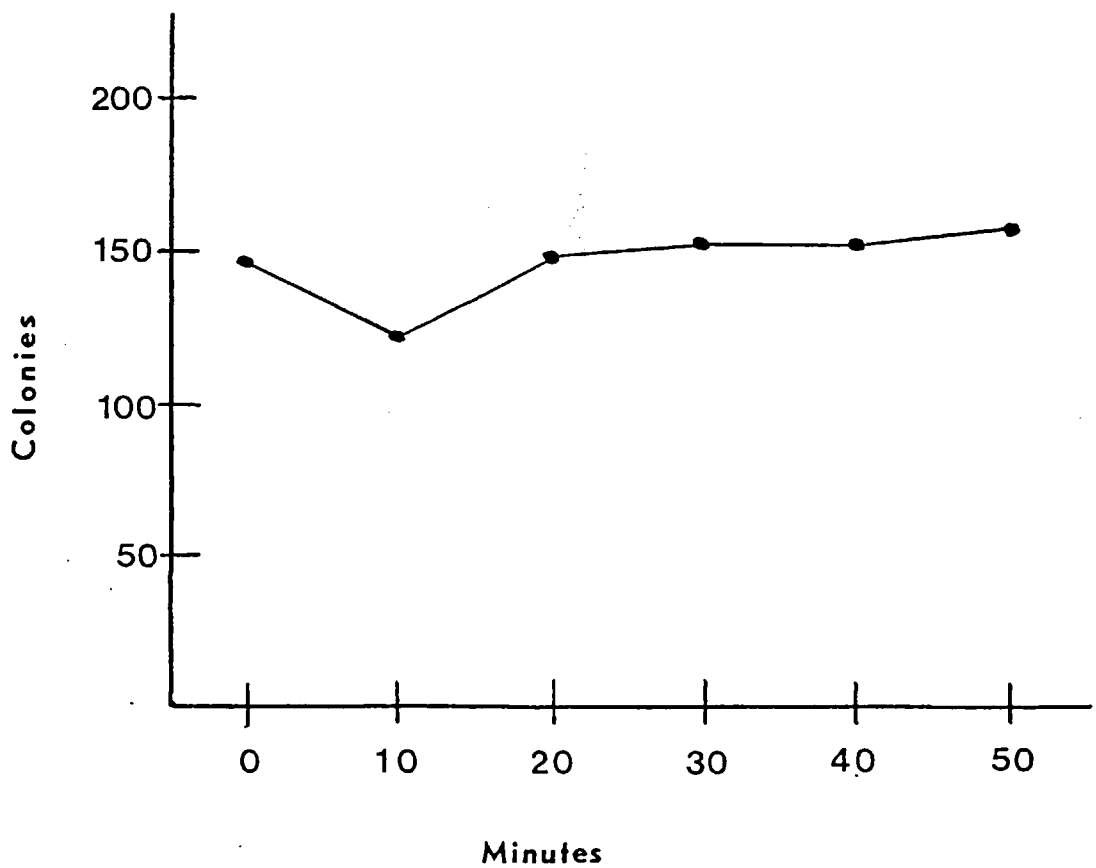
Following the transformation of competent *E.coli* cells with plasmids which code for ampicillin resistance, it is necessary to preincubate the transformants under non-selective conditions to allow the expression of ampicillin resistance before subjecting the transformants to ampicillin selection. It had been assumed that the same applied to transduction with *in vitro* packaged cosmid recombinants. The experiment shown in Fig.5.6 was designed to determine the length of preincubation necessary for activation so as to be certain that transductants were not subjected to ampicillin selection prior to the activation of ampicillin resistance. The results clearly indicate that when fresh overnight cultures are used as the bacterial host, the normal 10 mins adsorption at room temperature was sufficient and it was not necessary to preincubate the transductants under non-selective conditions prior to antibiotic

Legend to Figure 5.6

The effect of activating ampicillinase expression on cloning efficiency

1.5 μg of partial EcoRI cleaved sucrose gradient fractionated CV-1 DNA fragments with an average size of approximately 40Kb (see legend to Fig 5.9B) was ligated to 1 μg of EcoRI linearised Homer I DNA in a final reaction volume of 10 µl. 2µl of the ligation mixture was packaged in a 80µl *in vitro* packaging reaction (see section 2.12.6). The packaging mixture was diluted to 400µl with phage buffer. Transduction was performed by mixing 5µl of the phage buffer containing the transducing particles with 200µl of a fresh overnight culture of *E.coli* HB101. Following adsorption at room temperature for 10 mins, 450µl of L-broth was added to the mixture and the mixture was incubated at 37°C. At various times during the incubation, 100µl aliquots of the mixture was plated onto L-agar containing 100µg per ml of ampicillin. The plates were incubated at 37°C overnight and the number of ampicillin resistent colonies was scored. Time 0 is defined as the time of L-broth addition.

Fig. 5.6



selection. Finally, experiments were performed which indicate that the amount of packaging lysate used under the standard protocol for packaging *in vitro* Homer I recombinants is saturating (data not shown).

5.5 Analysis of recombinants derived from Homer I

Plasmid DNA was prepared from independent isolates obtained from the experiment described in Table 5.1 using the mini cleared lysate method. The DNAs were digested with *EcoRI* endonuclease and fractionated by agarose gel electrophoresis. The results shown in Fig.5.7 have several noteworthy features. All of the recombinants contained an *EcoRI* fragment of approximately the same size as linear Homer I DNA. The observed intensity of this band was disproportionate to the size of this fragment in most of the recombinants, indicating that a majority of the recombinants contained more than one copy of the vector molecule. A very crude estimation of the sizes of the *Drosophila* DNA fragments indicates that on average the recombinants contain only about 20 to 25Kb of *Drosophila* DNA. So it would seem that the conditions which produce high cloning efficiencies also tend to generate recombinant molecules containing several copies of Homer I DNA (this occurred despite the fact that the target DNA used had been size selected by sucrose gradient sedimentation, see legend to Table 5.1. and Fig.5.9B). As a result of having to accomodate several copies of vector DNA, the size of the inserts in these recombinant is smaller than expected. A second undesirable feature was the presence of a significant amount of vector background, i.e. a significant proportion of the transductants appear to contain only Homer I DNA. However, besides the problems of small inserts and vector background, the observation that most

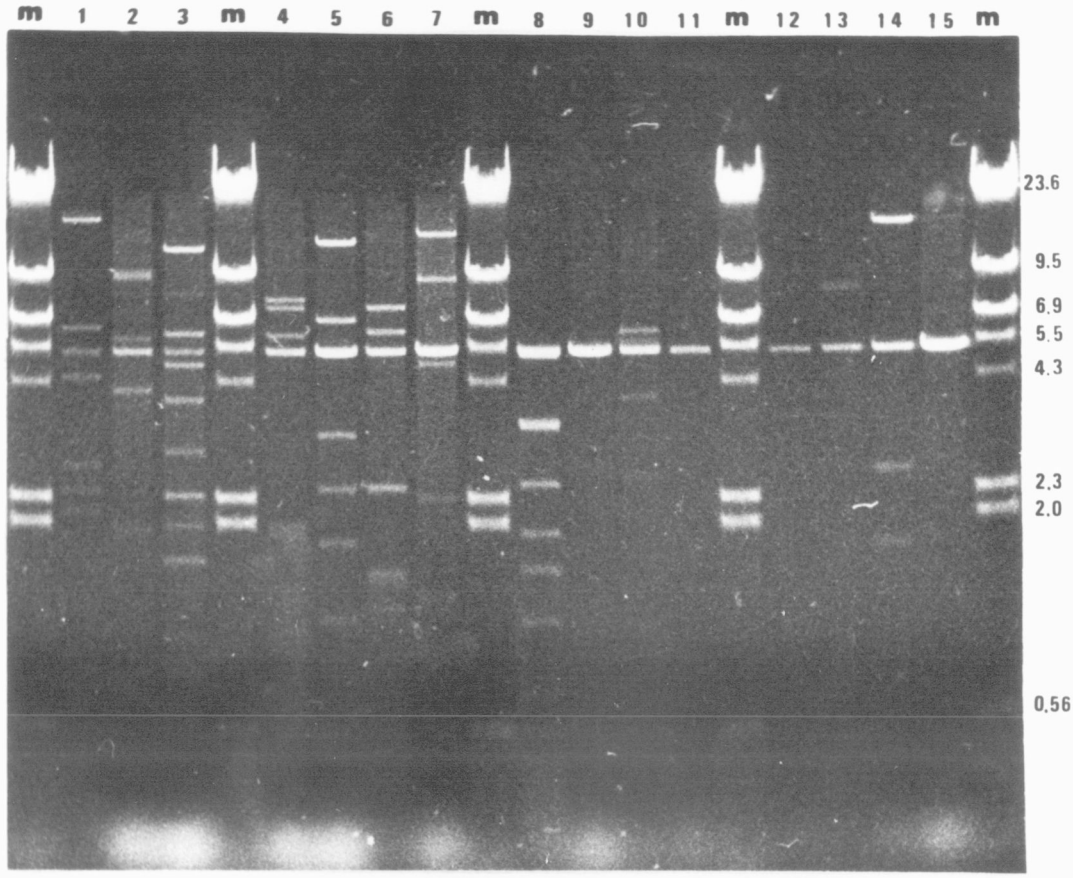
Legend to Figure 5.7

The EcoRI fragments of clones of Drosophila DNA obtained under conditions of optimal cloning efficiency

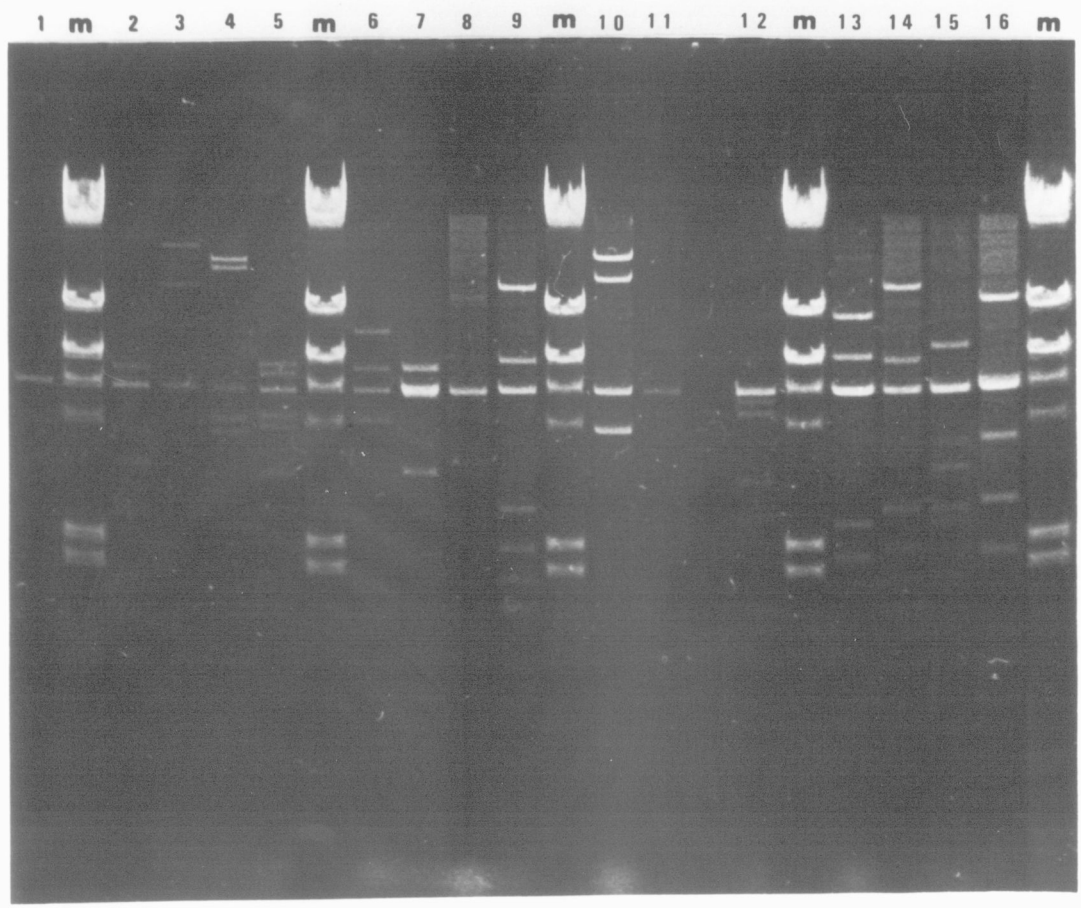
DNA was prepared, using the mini cleared lysate method, from 31 independent isolates obtained from the experiment described in Table 5.1. The DNA preparations were digested with EcoRI endonuclease and electrophoresed on an 0.6% (w/v) agarose gel. In A, the 15 isolates were obtained from the ligation mixture with a 3:1 molar ratio of Homer I to Drosophila DNA. In B, the 16 isolates were obtained from the ligation mixture with a 9:1 molar ratio of Homer I to Drosophila DNA. The marker tracks (m) contain HindIII digested λDNA plus EcoRI digested Homer I.

Fig.5.7

A


m 1 2 3 m 4 5 6 7 m 8 9 10 11 m 12 13 14 15 m
23.6
9.5
6.9
5.5
4.3
2.3
2.0
0.56

B


1 m 2 3 4 5 m 6 7 8 9 m 10 11 12 m 13 14 15 16 m

recombinants contain several vector molecules raised questions with regard to their possible instability. The observation that ampicillin resistant transductants resulting from packaged religated vector DNA contain predominantly monomeric supercoiled Homer I DNA (Fig.5.5B) implies that the packaged polymeric Homer I molecules are unstable in E.coli strain HB101 and break down to monomers. While there is no direct evidence on this point, it is nevertheless conceivable that the recombinants containing multiple copies of Homer I may also not be stable, i.e. monomeric vector supercoils may be generated from the recombinants and will subsequently compete with the recombinant molecules.

5.6 Construction of a Drosophila genomic library using Homer I

In theory, the problems raised in the last section can all be overcome by using dephosphorylated vector molecules in the ligation reaction. Since dephosphorylated vector molecules cannot be ligated to each other, vector polymers cannot be generated in the ligation reaction and thus vector background should be eliminated. Similarly, the use of dephosphorylated vector should ensure that each target end is ligated to only one vector molecule. So by using size selected target DNA, this should effectively eliminate the generation of recombinants containing more than one vector molecule. An additional advantage is that since vector background should be non-existent, dephosphorylated vector molecules can be used in even larger excess in the ligation reaction. The effect would be to minimise the probability of directly joining two target segments which are not contiguous in the genome. Another factor which needs consideration is the randomness with which target molecules are generated. Restriction endonucleases with

hexanucleotide recognition sites, like EcoRI endonuclease, will cleave genomic DNA approximately once every 4 Kb. However, for the purposes of generating a genomic library, it is preferable to clone fragments generated in a more random fashion (i.e. produced by a restriction endonuclease with a tetranucleotide recognition site) in order to increase the probability of including all of the genomic sequences in the library. However, enzymes with tetranucleotide recognition sites will also, in general, cleave the vector at more than one site. An ideal situation is exemplified by BamHI endonuclease and Sau3A endonuclease (MboI). The recognition sequence of Sau3A endonuclease is the central tetranucleotide in the hexanucleotide recognition sequence of BamHI endonuclease. Therefore these enzymes generate identical cohesive ends which can be ligated to each other. So one can ligate vector molecules linearised with BamHI endonuclease to target molecules generated by Sau3A endonuclease. An analogous situation exists for EcoRI (GAATTC) endonuclease and the EcoRI* (AATT) activity associated with the EcoRI endonuclease under certain reaction conditions (Polisky et al., 1975; Bolivar et al., 1977a; Kamp et al., 1977). The various theoretical considerations mentioned above were tested in the generation of a complete Drosophila genomic library.

The strategy employed for the construction of this library is summarised below:

Homer I DNA was linearised with EcoRI endonuclease and dephosphorylated with calf intestinal phosphatase; Drosophila genomic DNA was rendered refractory to EcoRI endonuclease by in vitro methylation with EcoRI methylase. The methylated DNA

was partially cleaved using the EcoRI* activity to generate molecules of approximately 40 Kb; this partially cleaved DNA was fractionated by sucrose gradient sedimentation to select for molecules of 35 to 50 Kb; Ligation was performed at high DNA concentration using the size selected Drosophila DNA and a large excess of dephosphorylated vector DNA; the ligated DNA was packaged in vitro and used to transduce E.coli strain HB101; transductants were plated onto nitrocellulose filters, grown under ampicillin selection and stored at -70°C as described in material and methods (section 2.13.2).

A variety of EcoRI* conditions were tested (Polisky et al., 1975; Bolivar et al.,1977a; Kamp et al., 1977) using SV40 DNA, which contains a single site for EcoRI endonuclease, as the substrate. The conditions optimal for EcoRI* activity were; 25mM Tris.HCl, pH7.7, 6mM $MgCl_2$, 6% (v/v) dimethyl sulphoxide and 20% (v/v) EcoRI dilution buffer (Fig.5.8A). However, even under these conditions, the predominant activity of EcoRI endonuclease was the EcoRI activity, i.e. the EcoRI reaction was complete before any EcoRI* activity could be observed. Therefore, in order to generate 40Kb fragments of genomic DNA with the EcoRI* activity, the DNA must first be made refractory to the EcoRI activity. This was accomplished by modifying Drosophila genome DNA with EcoRI methylase in the presence of ^{3}H-SAM (Greene et al., 1975). Charon 4 DNA was added to two aliquots of the methylation reaction mixture taken before and after the addition of an excess of EcoRI methylase. These aliquots were incubated in parallel with the methylation reaction mixture to monitor the extent of methylation. Following incubation, both aliquots were subjected to EcoRI

Legend to Figure 5.8

Optimising conditions for EcoRI* activity; EcoRI methylation of Drosophila genomic DNA.

A, tracks 1 to 6

0.3 µg of SV40(I) was digested with 3 units of EcoRI endonuclease in 25mM Tris·HCl, pH7.7, 6mM $MgCl_2$ and 0.6% (v/v) DMSO (track 1), 1.8% (v/v) DMSO (track 2), 3.0% (v/v) DMSO (track 3), 4.2% (v/v) DMSO (track 4), 5.4% (v/v) DMSO (track 5), 6.6% (v/v) DMSO (track 6).

tracks 7 to 12,

0.3 µg of SV40(I) was digested with 3 units of EcoRI endonuclease in 25mM Tris·HCl, pH7.7, 6mM $MgCl_2$, 20% (v/v) EcoRI dilution buffer and 0.6% (v/v) DMSO (track 7), 1.8% (v/v) DMSO (track 8), 3.0% (v/v) DMSO (track 9), 4.2% (v/v) DMSO (track 10), 5.4% (v/v) DMSO (track 11), 6.6% (v/v) DMSO (track 12).

Incubations were performed at 37°C for 1h. Track 13 contained undigested SV40(I). Track 14 contains HaeIII digested SV40 DNA. The samples were electrophoresed on a 1% (w/v) agarose gel.

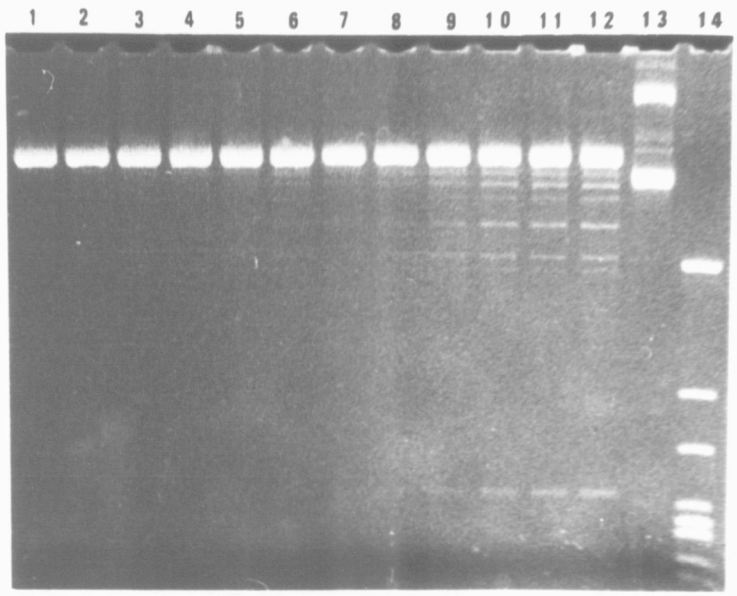
B, 200µg of Drosophila genomic DNA in 1.2ml was methylated with an excess of EcoRI methylase (see section 2.12.4) at 37°C for 1h. 5µl aliquots were removed from the reaction mixture before and after the addition of EcoRI methylase. 0.4µg of Charon 4 DNA, in 2 µl, was added to each aliquot and both aliquots were incubated at 37°C for 1h. Following incubation, DNA in the two aliquots were digested with an excess of EcoRI endonuclease and the samples were fractionated

on a 0.6% (w/v) agarose gel. Track 1, EcoRI cleaved Charon 4 DNA; track 2, aliquot taken prior to addition of methylase; track 3, aliquot taken following addition of methylase.

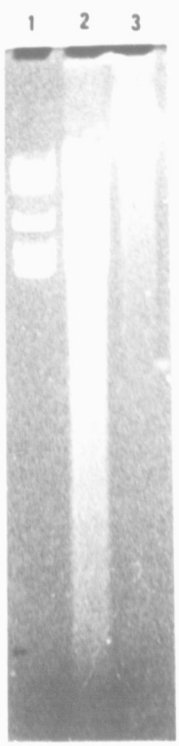
Note: 15% (v/v) DMSO abolishes enzymatic activity and 10% (v/v) DMSO does not improve EcoRI* activity.

Fig.5.9

A



B



endonuclease digestion and agarose gel electrophoresis. As can be seen in Fig.5.8B the methylation reaction was successful because the Charon 4 and Drosophila DNAs in the aliquot taken before the addition of methylase were digested to completion whereas the same DNAs in the aliquot taken following methylase addition were refractory to EcoRI endonuclease digestion.

The methylated Drosophila DNA was digested with various amounts of EcoRI endonuclease under EcoRI* and EcoRI conditions. The results shown in Fig.5.9A illustrate two points. Firstly, the methylated DNA was cleaved by EcoRI endonuclease under EcoRI* conditions but was refractory to enzymatic digestion under EcoRI conditions. Therefore, the methylase, though capable of modifying EcoRI recognition sites, does not appear to modify the EcoRI* sites. Secondly, approximately 2 to 3 units of EcoRI endonuclease per μg of methylated genomic DNA appeared to be optimal for the generation of large (35 to 50Kb) DNA fragments under EcoRI* conditions. 44 μg of methylated Drosophila DNA was digested with 100 units of EcoRI endonuclease. Following termination of the reaction by addition of EDTA to a final concentration of 25mM, the DNA was fractionated by sucrose gradient sedimentation (Maniatis et al., 1979). Ad2 (35 Kb) and Charon 4 (45 Kb) DNAs were used as markers in an identical parallel sucrose gradient. DNA in the marker gradient fractions was precipitated with ethanol, digested with EcoRI endonuclease, and subjected to agarose gel electrophoresis. Since Ad2 and Charon 4 DNAs yield characteristic and distinguishable DNA fragments, fractions containing Ad2 and Charon 4 DNAs could be identified. 5 μg of Drosophila DNA was recovered from the fractions sedimenting between the peaks of Ad2

Legend to Figure 5.9

EcoRI* digestion of Drosophila DNA methylated with EcoRI methylase; Sucrose gradient fractionation of this DNA

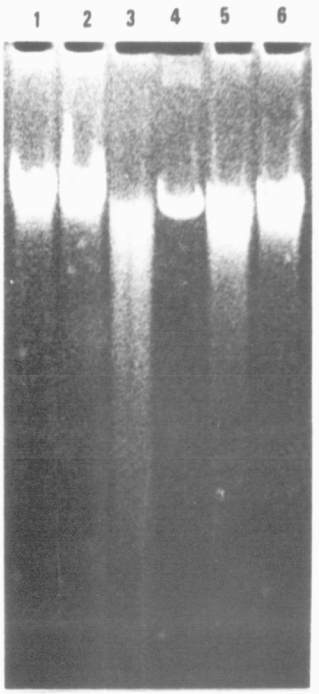
A, 0.55µg of methylated Drosophila DNA was digested with 8 units of EcoRI endonuclease in EcoRI buffer (track 1). 0.55µg of methylated Drosophila DNA was digested in EcoRI* buffer (25mM Tris·HCl, pH7.7, 6mM $MgCl_2$, 6% (v/v) DMSO) with no enzyme (track 2), 8 units of EcoRI (track 3), 4 units of EcoRI (track 5), 2 units of EcoRI (track 6). Track 4 contains undigested Ad2 DNA (37Kb). Incubations were performed at 37°C for 1h and the samples were fractionated on a 0.6% (w/v) agarose gel.

B, 44µg of methylated Drosophila DNA was digested with 100 units of EcoRI endonuclease for 1h at 37°C under EcoRI* conditions. Following the addition of EDTA to a final concentration of 25mM, the sample was fractionated by sucrose gradient sedimentation on a 28 ml 10% (w/v) to 40% (w/v) sucrose gradient (see section 2.11.1). An identical parallel marker gradient contained 25µg each of Ad2 and Charon 4 DNAs. Both gradients were centrifuged in a SW27 rotor at 25,000 rpm for 15.5h at 20°C. 0.5 ml fractions were collected by puncturing the bottom of the centrifuge tubes. DNA from the marker gradient fractions was ethanol precipitated, digested with EcoRI endonuclease and fractionated on a 0.6% (w/v) agarose gel. Tracks 1 to 9 contained EcoRI digested DNA from marker gradient fractions 9 to 17. EcoRI fragments of Charon 4 DNA are present in tracks 1 to

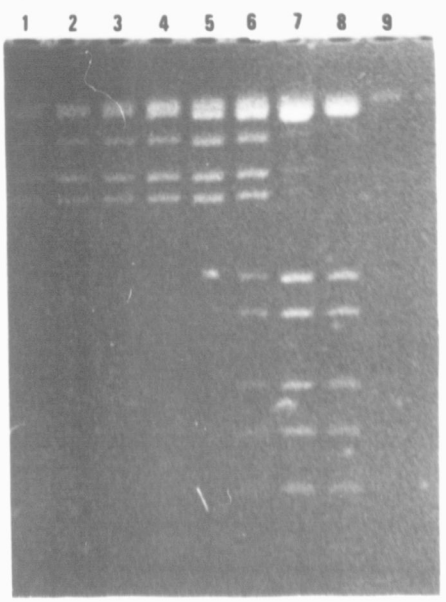
6 and EcoRI fragments of Ad2 DNA are present in tracks 6 to 8. Fraction 11 to 15 (equivalent to tracks 3 to 7) from the sucrose gradient containing EcoRI* fragments of Drosophila DNA were pooled and DNA contained in these fractions was precipitated with ethanol.

Fig.5.9

A



B



and Charon 4 DNAs (Fig.5.9B). Agarose gel electrophoresis of this size selected DNA indicated that the majority of this DNA was in the 35 to 50 Kb range and contamination by DNA fragments less than 25 Kb in size was minimal (Fig.5.10A).

Homer I DNA linearised by EcoRI endonuclease digestion was dephosphorylated with an excess of calf intestinal phosphatase. This vector DNA was indeed incapable of religating to itself (Fig. 5.10B). Following in vitro packaging and transduction, no ampicillin resistant colonies were obtained from religated vector. 1μg of the size selected Drosophila DNA generated by EcoRI* digestion was ligated to approximately 2μg of dephosphorylated vector DNA in a 10 µl ligation reaction mixture. 8 µl of this mixture was packaged in vitro. The packaging reaction mixture was diluted into phage buffer, sterilised with chloroform and used to transduce cells of E.coli strain HB101 from a fresh overnight culture. Approximately 3 x 10^4 independent transductants were obtained. These were plated directly (without prior amplification) onto 20 millipore filter discs (HAWP) at a density of approximately 1.5 x 10^3 ampicillin resistant colonies per disc. The transductants were grown on the filters under ampicillin selection. Replica filters were made and the filters containing the recombinants were stored at -70°C (see materials and methods section 2.13.2). Colonies from 10 of the replica filters were scraped off and stored as a pool. In addition, several filters containing approximately 300 colonies per filter were prepared from the unamplified transductants.

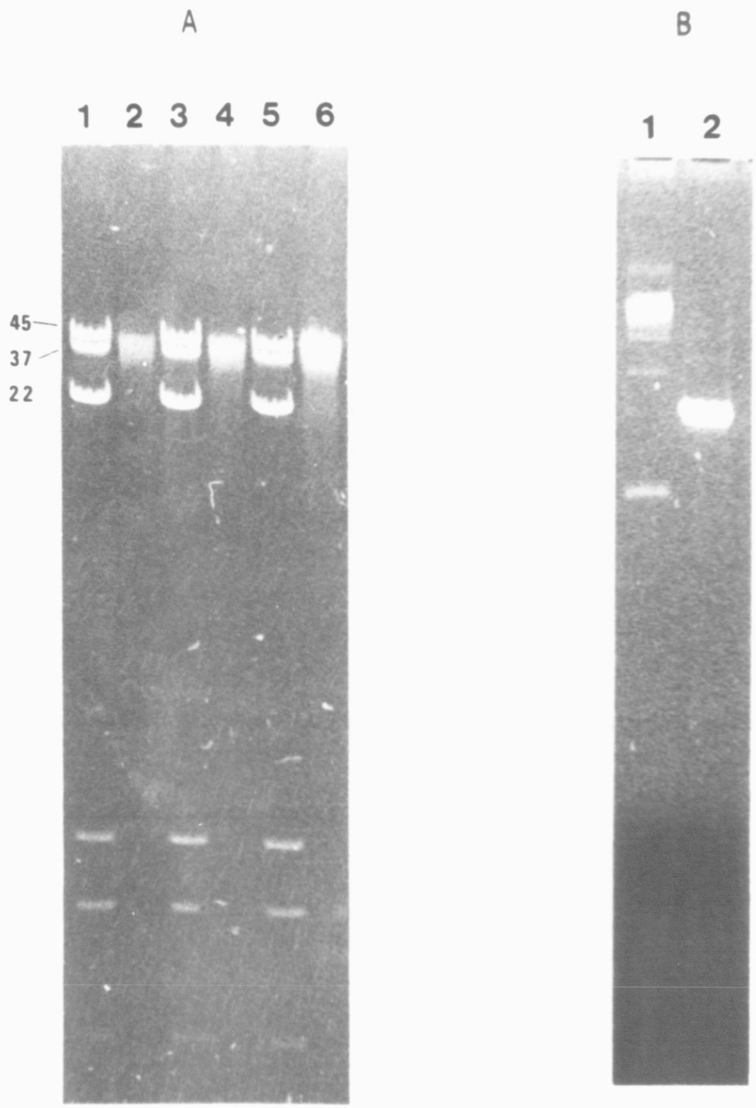
The cloning efficiency in this experiment was approximately 4 x 10^4 recombinants per μg of size selected target DNA. If one

Legend to Figure 5.10

Quality controls on size fractionated EcoRI* fragments of Drosophila genomic DNA and dephosphorylated EcoRI linearised Homer I DNA.

A, tracks 2, 4 and 6 contain increasing amounts of sucrose gradient fractionated ("35 to 50Kb") EcoRI* fragments of Drosophila DNA. Tracks 1, 3 and 5 are marker tracks containing Charon 4 DNA (45Kb), Ad2 DNA (37Kb) and EcoRI cleaved Ad2 DNA (the largest fragment is 21.6Kb). Samples were run on a 0.4% (w/v) agarose gel.
B, track 1 contains Homer I DNA linearised with EcoRI endonuclease religated at a final DNA concentration of 140 ug per ml. Track 2 contains the same DNA which was dephosphorylated with calf intestinal phosphatase prior to religation at a final DNA concentration of 140 ug per ml.

Fig.5.10



is not particularly concerned about avoiding the cloning, in one recombinant molecule, of two or more target segments which are not contiguous in the genome, then smaller target molecules (i.e. from a complete EcoRI endonuclease digestion) can be routinely cloned at efficiencies of above 10^5 recombinants per μg of target DNA.

5.7 Analysis of the *Drosophila* genomic library

Plasmid DNA was prepared from 18 independently derived ampicillin resistant colonies from the *Drosophila* library using the mini cleared lysate method. Following digestion with EcoRI endonuclease and agarose gel electrophoresis, Southern transfer hybridisation was performed. pAT153, ^{32}P-labelled by nick translation, was used as the hybridisation probe in order to identify the DNA fragments containing vector sequences. The results shown in Fig.5.11A indicate that only a single vector-containing DNA fragment was generated from each of the recombinant DNAs by EcoRI endonuclease digestion. For a majority of the recombinant plasmids ($^{13}/_{18}$), the size of this fragment was larger than that of monomeric linear Homer I DNA. The first implication of these observations is that each recombinant molecule contains only a single copy of vector DNA. Secondly, since only $\frac{1}{4}$ of the EcoRI-EcoRI* joints are expected to regenerate the EcoRI recognition sequence, if all of the cohesive ends of the target molecules used for the construction of the library were generated by EcoRI* activity, then one would expect approximately $^1/_{16}$ of the recombinant molecules to yield monomeric linear Homer I DNA following EcoRI endonuclease digestion. The observation that 5 out of 18 recombinants yielded monomeric Homer I DNA is therefore somewhat

Legend to Figure 5.11

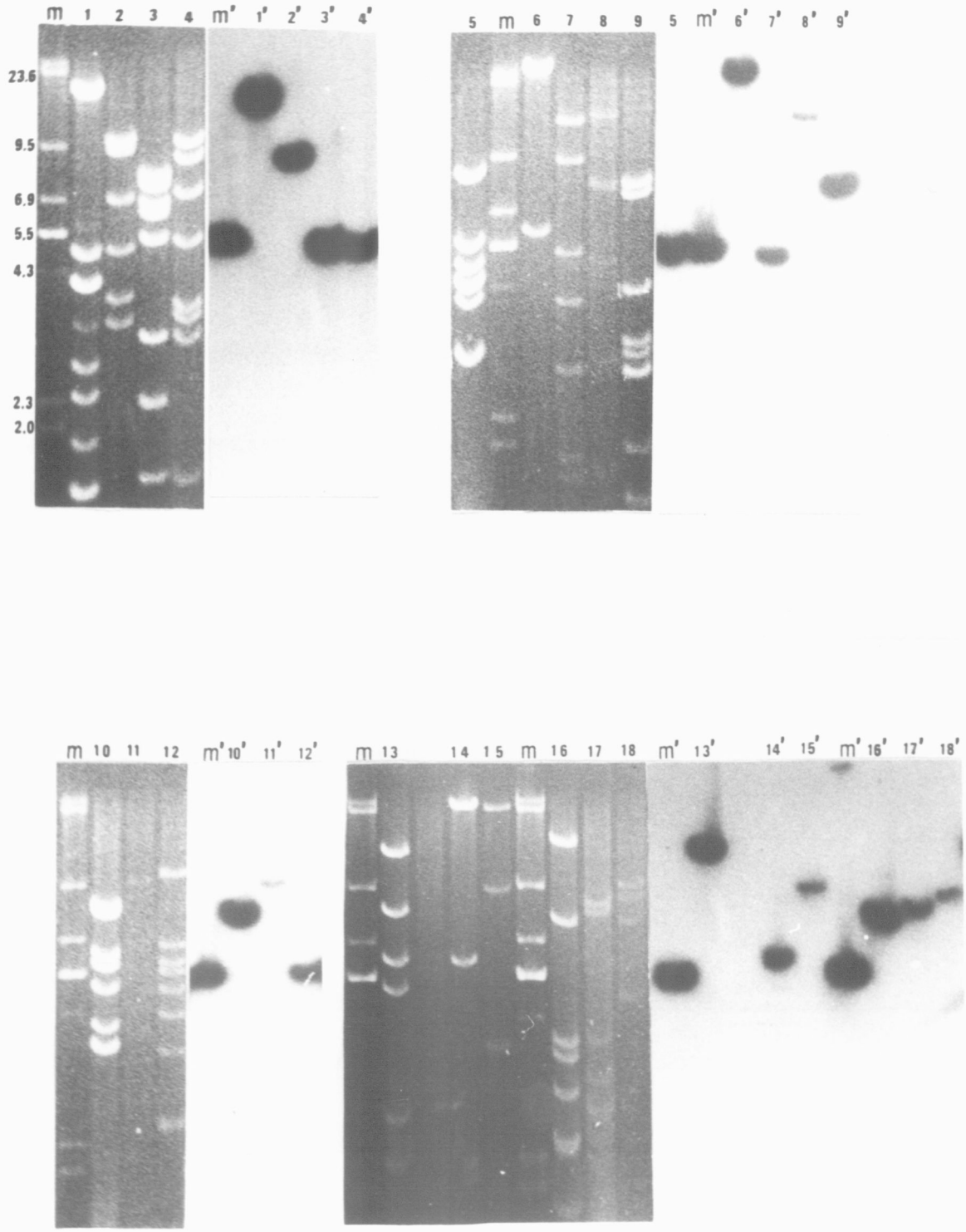
Electrophoretic analysis of cosmid clones from the EcoRI* Drosophila library

A, DNA was prepared from 18 independently derived isolates from the EcoRI* Drosophila library using the mini cleared lysate method. These DNAs were digested with EcoRI endonuclease and fractionated on 0.6% (w/v) agarose gels. Transfer hybridisation was performed as described (section 2.9) using ^{32}P-labelled pAT153 as the hybridisation probe. The unprimed tracks (1 to 18) are photographs of the fractionated DNAs stained with ethidium bromide. The primed tracks (1' to 18') are autoradiograms of the corresponding unprimed tracks following transfer hybridisation of the ethidium bromide stained DNA. The marker tracks (m) contain HindIII digested λDNA plus EcoRI digested Homer I DNA.

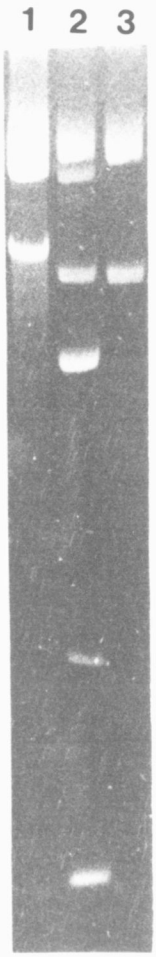
B, track 1 contains DNA prepared, by CsCl ethidium bromide density equilibrium centrifugation, from a pool of approximately 1.5×10^4 independently derived clones from the EcoRI* Drosophila library. Track 2 contains relaxed circular and supercoiled forms of Homer I (5.5Kb), pDM103 (21Kb) and cDM219 (32Kb). Track 3 contains relaxed circular and supercoiled forms of cDM219. The samples were fractionated on a 0.4% (w/v) agarose gel.

Fig. 5.11

A


m 1 2 3 4 m' 1' 2' 3' 4'
23.6
9.5
6.9
5.5
4.3
2.3
2.0
5 m 6 7 8 9 5 m' 6' 7' 8' 9'
m 10 11 12 m' 10' 11' 12'
m 13 14 15 m 16 17 18 m' 13' 14' 15' m' 16' 17' 18'

B



higher than expected. There are two possible explanations. Firstly, it is possible that the EcoRI methylation reaction performed on the Drosophila genomic DNA was incomplete. The assay used to monitor this reaction (Fig.5.8B) would probably not have detected a slightly incomplete reaction. The second possibility is that the EcoRI* activity, rather than cleaving all XAATTX sequences with equal proficiency, may preferentially cleave at GAATTX and XAATTC, i.e. at those sequences most closely related to the EcoRI recognition site. Whatever the reason may be, it is clear that a majority of the target molecule ends were generated by EcoRI* activity. Finally, of the 18 cases examined, no vector background transductant was found. This observation was re-enforced by the finding that total plasmid DNA prepared from a pool of approximately 1.5×10^4 independently derived recombinants by density equilibrium centrifugation did not contain detectable amounts of monomeric Homer I DNA (Fig.5.11B). It was also demonstrated that religated, in vitro packaged, dephosphorylated Homer I DNA yielded ampicillin resistant transductants at low (undetectable) frequencies. Taken together, these observations clearly indicate that the level of vector background present in this Drosphila library is very low.

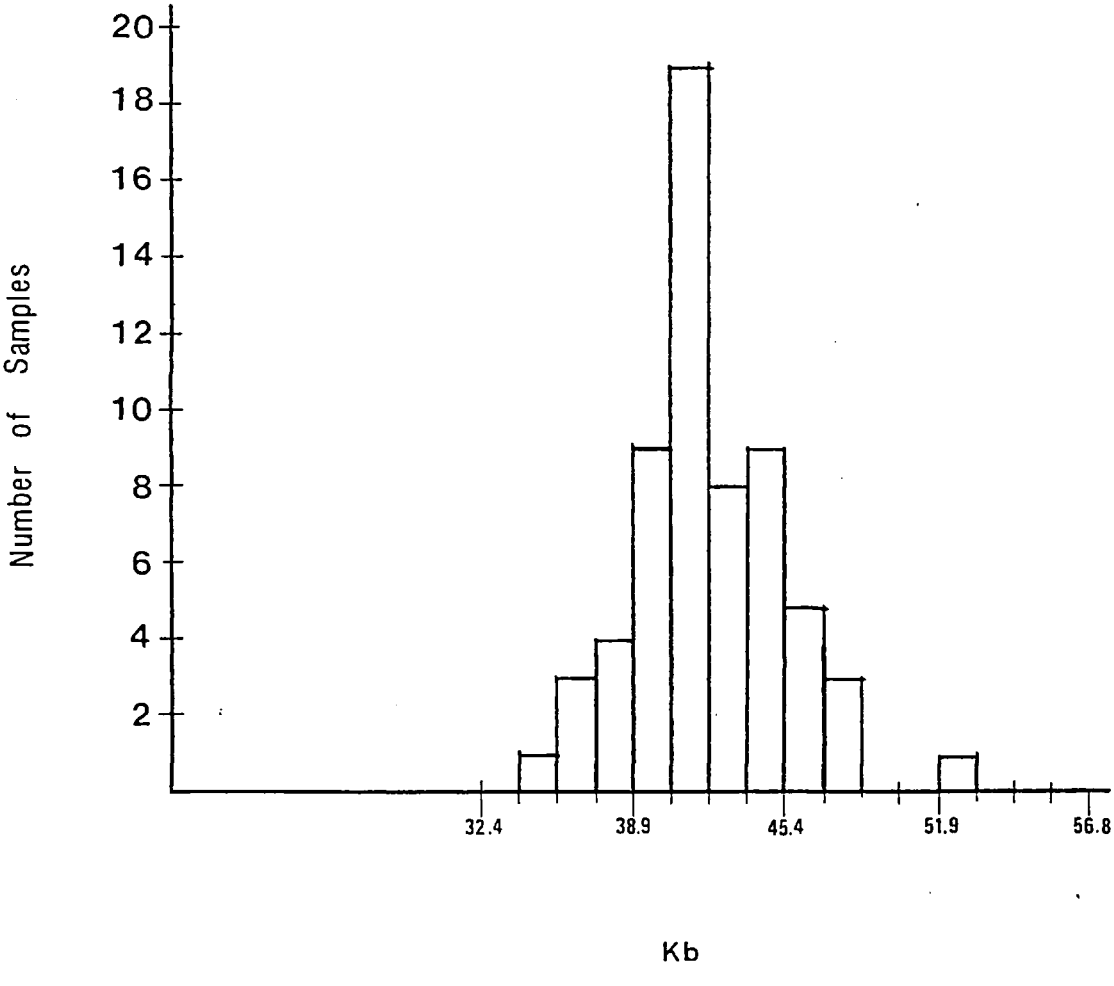
Plasmid DNA prepared from a pool of approximately 1.5×10^4 recombinants was analysed by electron microscopy. The contour lengths of 62 open circular molecules were measured using open circular SV40 DNA as a size standard. The resultant histogram is shown in Fig.5.12. The mean size of the molecules was 41.6 Kb with a standard deviation of 3.0 Kb. The smallest and largest molecules observed were 35 Kb and 52 Kb respectively, reasonably

Legend to Figure 5.12

Electron microscopic analysis of cosmid clones from the EcoRI* Drosophila library

The contour lengths of 62 relaxed circular molecules from DNA prepared from a pool of 1.5×10^4 independently derived clones were measured using SV40(II) DNA as a size standard (see section 2.18). The mean length of the cosmid molecules is 41.6Kb with a standard deviation of 3.0Kb. The mean length of SV40 DNA was given as 5.24Kb. The standard deviation for 18 SV40 molecules measured is 0.19Kb.

Fig.5.12



1 Tic Width = 1.62 Kb

consistent with the known packaging limits of λ DNA. The observed size distribution in the histogram suggests that the recombinants, at least the great majority of them, are stable in E.coli HB101, a recA⁻ stain. This conclusion is strongly supported by the observation that plasmid DNA prepared from a pool of approximately 1.5×10^4 independently derived recombinants contains molecules which are essentially all larger than 32Kb, as judged by gel electrophoresis (Fig.5.11B). Since previous experiments have shown that these recombinants contain only one copy of vector per molecule and given the size of Homer I as 5.5Kb, the mean size of the Drosophila DNA inserts in this library is approximately 36Kb. Having determined the mean insert size and given that the size of the haploid genome of D. melanogaster is 1.6×10^5 Kb (Rudkin, 1972), the completeness of the library can be calculated as described by Clarke and Carbon (1976). The equation used is shown below:

$$N = \frac{\ln (1-p)}{\ln \left(1-\frac{L-X}{m}\right)}$$

where

N = number of independently derived recombinants in the library

P = probability of the library containing any given single copy sequence, assuming the entire genome consists of single copy sequences.

L = mean length of insert = 36Kb for this particular library

X = size of the single copy sequence in question. For this calculation it is assumed to be small $\Rightarrow$ 0

M = size of the genome in question

= 1.6×10^5 Kb for D. melanogaster

For a Drosophila library consisting of 3×10^4 independently derived recombinants with a mean insert size of 36Kb, the probability of finding any given single copy sequence is in excess of 99%. To have a 95% probability, the library has to include only 1.34×10^4 independently derived recombinant clones. So the library described should contain in excess of 99% of all the Drosophila genomic sequences.

Using the method of Hanahan and Meselson (1980; see materials and methods section 2.13.2), 6 nitrocellulose filters each containing approximately 300 independently derived recombinant colonies were screened, in duplicate (12 filters total), with partially base hydrolysed Drosophila 18S and 28S rRNAs, ^{32}P-labelled with T4 polynucleotide kinase. Over 80% of the positive signals obtained were present in duplicate and 61 duplicate signals were obtained. The result obtained from one set of duplicate filters is shown in Fig.5.13A. Consistent with the claims of Hanahan and Meselson (1980), the colonies can be screened at high density without apparently distorting the positive signals. Fig.5.13B gives the results of a filter containing approximately 8×10^3 recombinants screened with ^{32}P-labelled 18S and 28S rRNAs. Using the obtained figures of 61 positives out of 1800 independent isolates screened, the ribosomal gene number can be calculated in the following way:

Legend to Figure 5.13

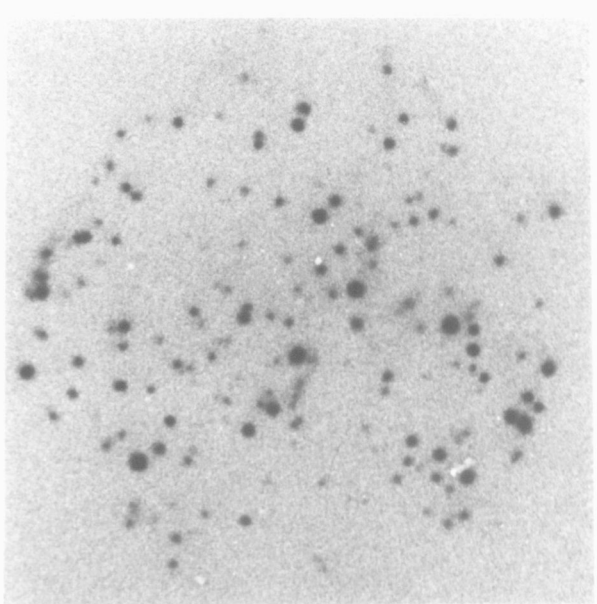
The rDNA representation of the *Drosophila* library

A, one set of two duplicate filters containing 300 independently derived clones from the *EcoRI** *Drosophila* library was screened (see section 2.13.2) with partially base hydrolysed *Drosophila* 18S and 28S rRNAs ^{32}P-labelled using T4 polynucleotide kinase (see section 2.8). The autoradiograms are shown.
B, shows the result of a similarly treated filter containing approximately 8 x 10^3 clones derived from partial *EcoRI* fragments of *Drosophila* DNA cloned in Homer I.

Fig. 5.13

A

B



Given that the rDNA repeats are contiguous in the genome then,

$$\frac{\text{the number of rRNA containing clones}}{\text{number of total clones screened}} = \frac{\text{average size of rDNA repeat} \times \text{copy number of rDNA}}{\text{haploid genome size of } \underline{\text{D. melanogaster}}}$$

Given that:

the haploid genome size of D.melanogaster = 1.6×10^5 Kb (Rudkin, 1972);

the average size of the rDNA unit = 14 Kb (Wellauer et al., 1978);

then according to this calculation,

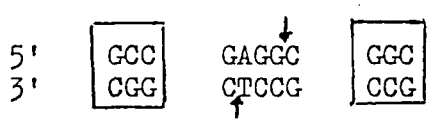
the copy number of rDNA units per haploid genome = 387

This calculated value of 387 agrees well with the published values of between 250 to 400 rDNA repeats per haploid genome (Ritossa and Spiegelman, 1965; Glover et al., 1975; Spear, 1975). This library has also been screened for its content of the zeta elements associated with the type I insertion sequence of D.melanogaster 28S rDNA (Kidd and Glover, 1979). The results (D. De Cicco, personal communication) agree well with the published haploid copy number of 40. These findings indicate that, at least for the two cases mentioned, the representation of D. melanogaster sequences in this library is excellent.

5.8 The stability of cloned 'head to tail' polymers of SV40 DNA

Since one class of molecules that we would like to clone with Homer I is postulated to contain 'head to tail' repeats of SV40 DNA (see chapter 4), an experiment was performed to assess the stability of 'head to tail' polymers of SV40 DNA cloned in Homer I and propagated in the recA$^-$ E.coli strain, HB101. One obvious way of generating SV40 polymers is to religate SV40 DNA which has been linearised with a restriction endonuclease which cleaves SV40

once. However, the problem with this approach is that most restriction endonucleases cleave at sites that contain a two-fold axis of symmetry. As a result, the linear SV40 molecules generated will possess two cohesive ends which are indistinguishable. Therefore, religation of these molecules will generate SV40 polymers that contain not only 'head to tail' joints but also 'head to head' and 'tail to tail' joints. The method used to generate 'head to tail' SV40 polymers relied on using SV40 DNA linearised with BglI endonuclease as the substrate for the ligation reaction. BglI endonuclease belong to a rare class of endonucleases which cleave and generate cohesive ends outside of their recognition sequences (Shortle and Nathans, 1979). In the SV40 sequence shown below,



the recognition sequence is indicated by the boxes and cleavage sites are indicated by the arrows. The resultant three base cohesive ends are not symmetric and can only base pair when the orientation is 'head to tail'. Wake and Wilson (1980) have independently employed this same idea. To test the validity of this idea, polymers of SV40 DNA generated by religation of linear SV40 DNA molecules with EcoRI cohesive ends were digested with BstI and polymers generated by religation of linear SV40 DNA molecules with BglI cohesive ends were digested with EcoRI. The results shown in Fig.5.14A indicate that polymers originating from monomers with EcoRI cohesive ends yield not only monomeric linear SV40 DNA but also the two additional fragments expected to result from 'head to

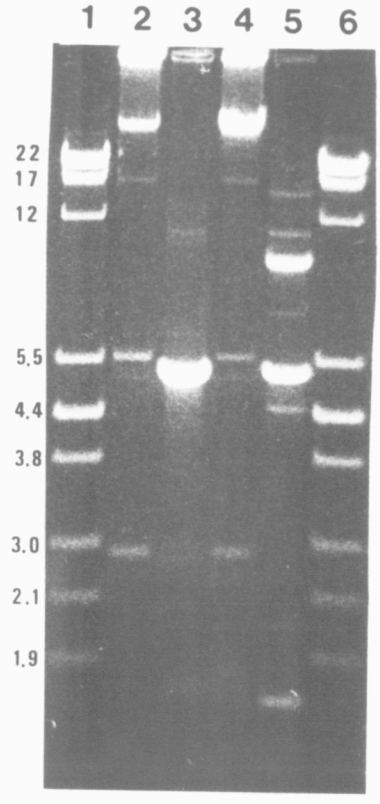
Legend to Figure 5.14

Cohesive termini generated by BglI endonuclease can only be religated in a "head to tail" orientation

A, SV40 DNA linearised with BglI endonuclease was religated at a final DNA concentration of 300 ug per ml. The religated DNA was electrophoresed either undigested (track 2) or following digestion with EcoRI endonuclease (track 3). SV40 DNA linearised with EcoRI endonuclease was similarly religated. The religated DNA was electrophoresed either undigested (track 4) or following digestion with BstI endonuclease (track 5). The markers (tracks 1 and 6) are EcoRI cleaved Ad2, pDM103, pDM238 and Homer I DNAs.
B, a schematic illustration of the rationale for the experiment shown in A. Following EcoRI digestion, religated SV40 DNA with BglI junctions would be expected to produce 5.2Kb fragments from "head to tail" joints (HT), 3.4Kb fragments from "head to head" joints (HH) and 7.0Kb fragments from "tail to tail" joints. The 3.4Kb and 7.0Kb fragments were not seen (A, track 3). BstI digestion of religated SV40 DNA containing EcoRI junctions would be expected to produce 5.2Kb fragments from "head to tail" joints, 1.5Kb fragments from "head to head" joints and 8.9Kb fragments from "tail to tail" joints. The 1.5Kb and 8.9Kb fragments were observed (A, track 5).

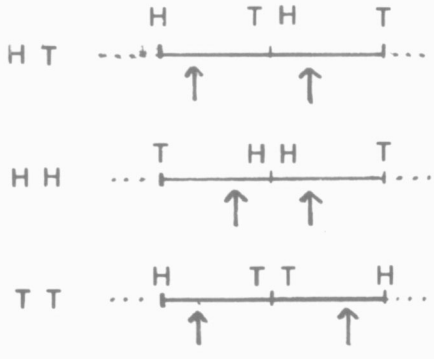
Fig.5.14

A



B

BglI junctions



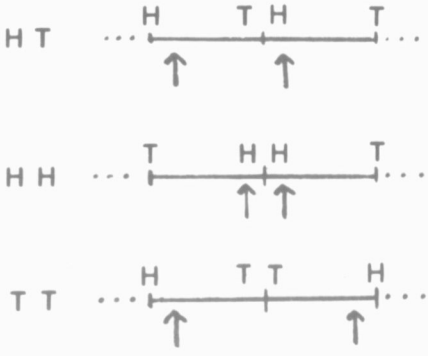
EcoRI Digestion

5.2

3.4

7.0

EcoRI junctions



BstI Digestion

5.2

1.5

8.9

head' and 'tail to tail' joinings of SV40 linears (See Fig.5.14B); whereas polymers originating from monomers with BglI ends yielded only monomeric linear SV40 DNA following EcoRI endonuclease digestion, indicating that the polymers contained only 'head to tail' joints. So the expectations concerning the property of BglI endonuclease were confirmed.

The strategy used to clone 'head to tail' SV40 polymers is as follows:

SV40 DNA was modified by *in vitro* methylation with EcoRI methylase. A mixture containing a 9:1 ratio of methylated to unmethylated SV40 DNA was digested to completion with BglI endonuclease. The linear monomers were ligated at high DNA concentration to generate 'head to tail' polymers. The polymers were digested with EcoRI endonuclease to generate 'head to tail' polymers (average of 10 SV40 repeats per molecule) with EcoRI cohesive ends. These polymers were ligated to an excess of Homer I DNA linearised with EcoRI endonuclease and dephosphorylated with calf intestinal phosphatase. Following *in vitro* packaging, transduction was performed using a fresh overnight culture of *E.coli* strain HB101. Ampicillin selection was used to obtain individual recombinant colonies.

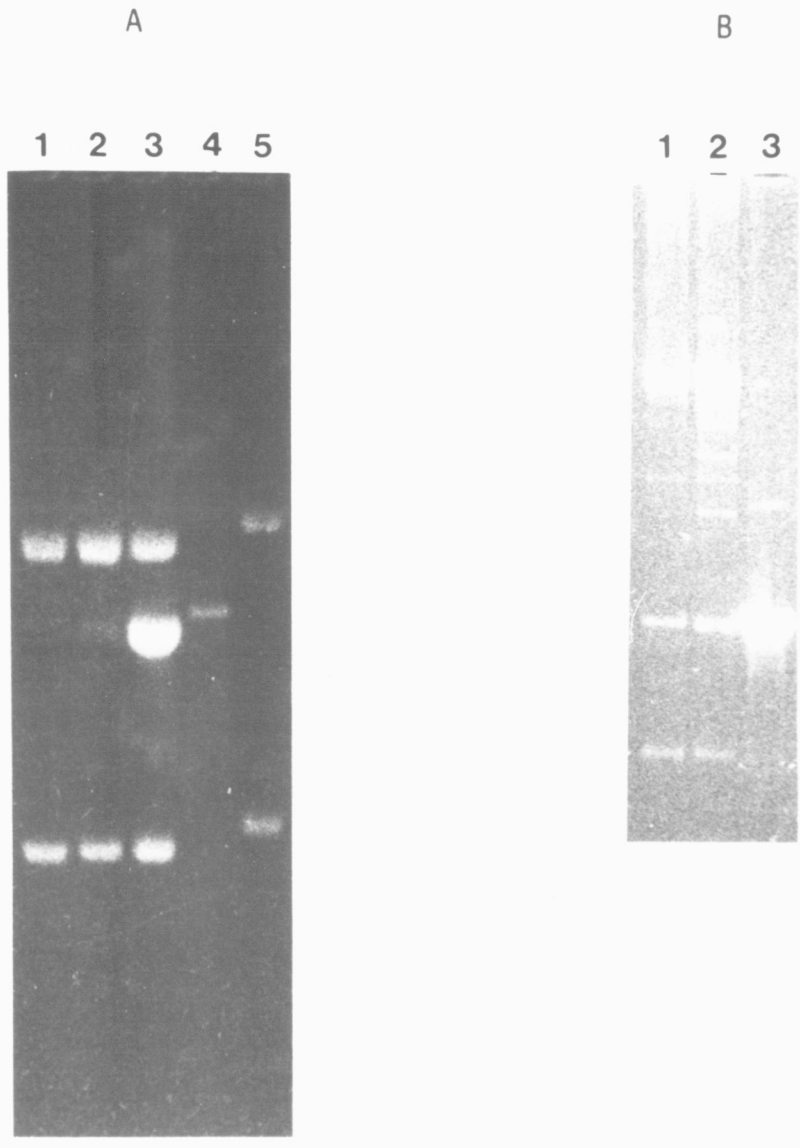
The series of reactions outlined above proceeded as planned. The results shown in Fig.5.15A indicate that supercoiled SV40 DNA was indeed refractory to EcoRI endonuclease digestion following the methylation reaction. The results shown in Fig.5.15B indicate that following religation of SV40 DNA with a 9:1 ratio of methylated

Legend to Figure 5.15

The *in vitro* generation of "head to tail" SV40 polymers with *EcoRI* cohesive termini

A, SV40(I) DNA was methylated with *EcoRI* methylase (see section 2.12.4). Track 1, undigested methylated DNA; track 2, methylated DNA digested with an excess of *EcoRI* endonuclease; track 3, a 1:1 ratio of methylated to unmethylated SV40 DNA digested with an excess of *EcoRI* endonuclease; track 4, linear Homer I DNA; track 5, relaxed circular and supercoiled forms of Homer I DNA.
B, a 9:1 ratio of methylated to unmethylated SV40(I) DNA was linearised with *BglI* endonuclease (track 3). This DNA was religated at a final DNA concentration of 300 ug per ml (track 1). The religated DNA was digested with an excess of *EcoRI* endonuclease (track 2).
All samples were electrophoresed on an 0.55% (w/v) agarose gel.

Fig.5.15



to unmethylated DNA) linearised with BglI endonuclease, high molecular weight polymeric molecules were generated and that EcoRI endonuclease digestion of these polymers resulted in primarily large molecules, presumably with EcoRI cohesive ends. Following ligation to dephosphorylated Homer I DNA with EcoRI cohesive ends, *in vitro* packaging and tranduction, ampicillin resistant isolates were obtained at a frequency of greater than 10^5 per µg of polymer DNA.

Plasmid DNAs prepared from six independently derived isolates using the mini cleared lysate method were digested with BstI endonuclease, which does not cleave Homer I DNA and cleaves SV40 DNA once, and subjected to agarose gel electrophoresis. The results shown in Fig.5.16A indicate that each of the plasmid DNAs yielded two fragments. Consistent with the cloned segments being the result of 'head to tail' joinings, the sizes of the two fragments correspond approximately to those of monomeric SV40 DNA and monomeric SV40 DNA plus Homer I. However, the relative intensities of the two bands are not consistent with the plasmid DNAs being of packageable size, i.e. the monomeric SV40 band is under represented. To investigate this point, plasmid DNAs prepared from three of the isolates by density equilibrium centrifugation were left undigested, digested with SalI endonuclease, which does not cleave SV40 DNA and cleaves Homer I once, or digested with EcoRI endonuclease, and fractionated by agarose gel electrophoresis. The results (Fig.5.16B) clearly demonstrate that DNA prepared from each of the isolates contains a heterogeneous mixture of circular plasmid DNAs. The sizes of these DNAs, as judged from the mobility

Legend to Figure 5.16

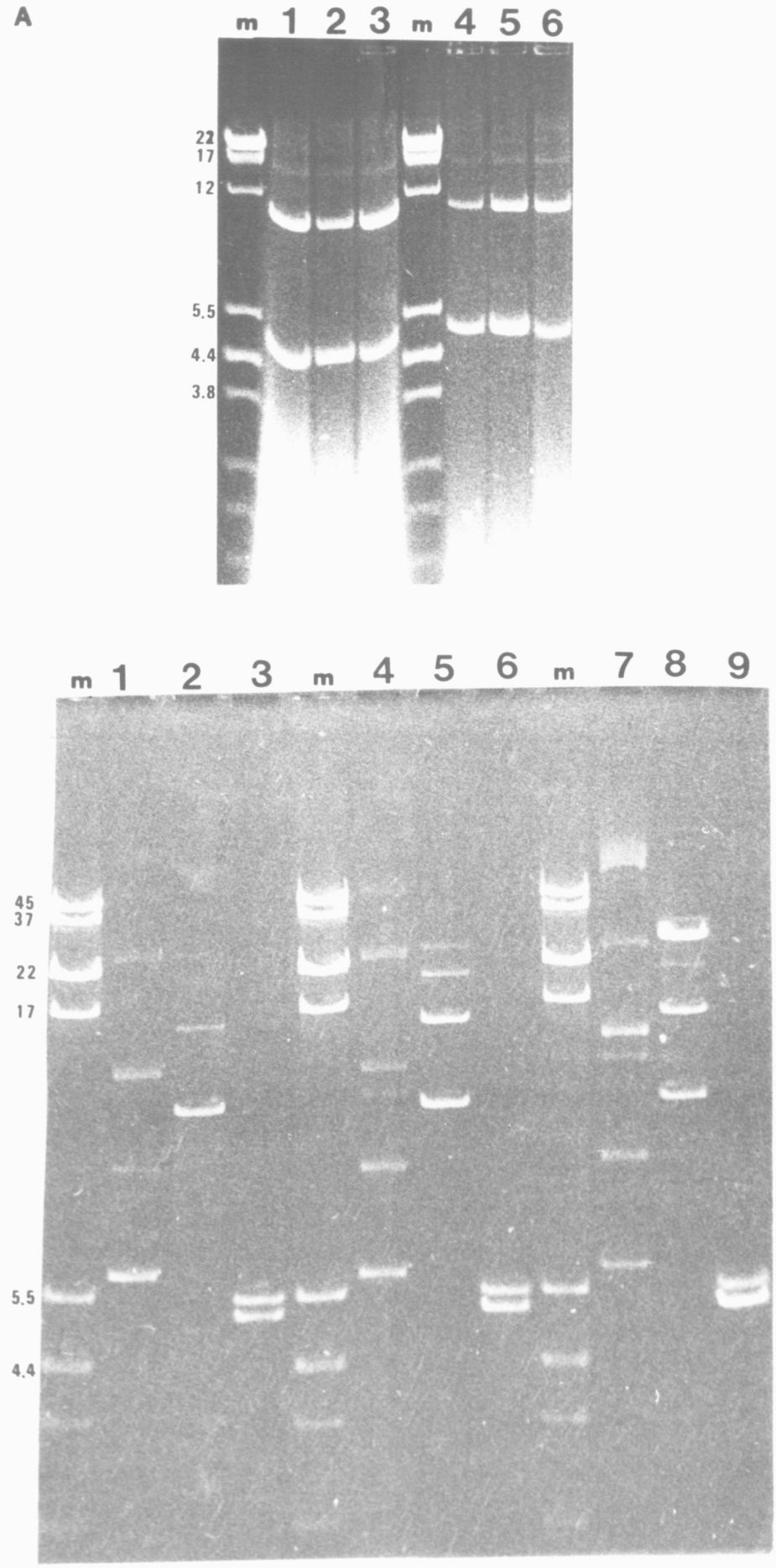
"Head to tail" polymers of SV40 DNA cloned in Homer I are unstable when propagated in E.coli strain HB101

A, 0.6 ug of the in vitroly generated "head to tail" polymers of SV40 DNA was ligated to 1.4 ug of EcoRI linearised, dephosphorylated Homer I DNA at a final total DNA concentration of 200 ug per ml. Following ligation, the ligation mixture was packaged in vitro and the packaging mixture was used to transduce a fresh overnight culture of E.coli strain HB101 (see section 2.12.6). Transductants were obtained at a frequency of approximately 10^5 per ug of SV40 DNA. DNA was prepared from 6 independently derived isolates using the mini cleared lysate method. The DNAs were digested with BstI endonuclease and fractionated on a 0.5% (w/v) agarose gel. Tracks 1 to 6 contain DNAs from cloned isolates 9 to 14 respectively. The marker tracks (m) contain EcoRI digested Ad2, pDM103, pDM238 and Homer I DNAs.

B, DNA was prepared from cloned isolates 10, 12 and 13 by CsCl ethidium bromide density equilibrium centrifugation. The DNAs were left undigested, digested with SalI endonuclease or digested with EcoRI endonuclease and fractionated on a 0.4% (w/v) agarose gel. Tracks 1 to 3 contain DNA from isolate 10 undigested, digested with SalI or digested with EcoRI; tracks 4 to 6 contain DNA from isolate 12 undigested, digested with SalI or digested with EcoRI; tracks 7 to 9 contain DNA from isolate 13 undigested, digested with SalI or digested with EcoRI.

The band in track 7 which migrates slightly faster than the 17Kb marker fragment is probably a cloned tetramer of SV40 DNA (see Fig.5.17).

Fig.5.16



of the linear molecules generated by SalI endonuclease digestion, are consistent with them being cloned monomers, dimers, trimers and tetramers of SV40 DNA. Since only molecules ranging in size from 35 to 50 Kb would have been packaged by the *in vitro* packaging system, the simplest interpretation of this result is that larged cloned 'head to tail' polymers of SV40 DNA are not stable when propagated in *E.coli* strain HB101. Since HB101 reverts to *recA*$^+$ at an appreciable frequency (W.J. Brammar, personal communication), it was necessary to show that the bacterial cells carrying these plasmids are *recA*$^-$. The results shown in Table 5.2 indicate that phage Charon 4, which is genotypically *red*$^-$, *gam*$^-$, cannot plaque on one of the strains carrying the cloned SV40 DNA. Since *red*$^-$, *gam*$^-$ phage can plaque on *recA*$^+$ strains, this result demonstrates that the one bacterial strain tested is indeed *recA*$^-$.

Finally, one of the circular DNA species from the heterogeneous mixture of recombinant plasmids prepared from isolate 13 (see legend to Fig.5.16B) was purified from an agarose gel slice (as described in material and methods, section 2.10). The purified plasmid was probably a cloned tetramer of SV40 DNA as judged by its electrophoretic mobility (see Fig.5.16B). Following transfection of competent cells of *E.coli* strain HB101 with this DNA, ampicillin resistant isolates were obtained. Plasmid DNAs prepared from eleven of the isolates were fractionated by agarose gel electrophoresis. The results shown in Fig.5.17A indicate that this species of SV40 oligmer cloned in Homer I was stably propagated in HB101, i.e. in all eleven cases only a single

TABLE 5.2

One Bacteria Strain Carrying Cloned 'Head to Tail'

SV40 DNA is indeed recA$^-$

Bacterial Strain	Phage Strain							
	Charon 4	10^{-5} dilution of stock			λ vir	10^{-4} dilution of stock		
		1µl	10µl	100µl		1µl	10µl	100µl
KH802	+	48	400	S[a]	+	500	S	C[b]
HB101	-	0	0	0	+	350	S	C
Strain containing 'head to tail' SV40 clone (isolate 12)	-	0	0	0	+	350	S	C

Fresh overnight cultures of KH802, HB101 and one of the 'head to tail' SV40 clones (isolate 12) were infected with various dilutions of either Charon 4 or λvir. The infections were performed by mixing 100µl of bacterial culture in L-broth plus 10mM $MgSO_4$ with 1 to 100µl of phage in phage buffer, allowing phage to adsorb to the cells for 10 mins. at room temperature before adding 3ml of top-agar and plating the mixture onto L-plates. Following incubation at 37°C overnight, plaques were scored.

[a] S = semiconfluent

[b] C - confluent

Legend to Figure 5.17

Cloned "head to tail" tetramer of SV40 DNA is stably propagated in E.coli strain HB101

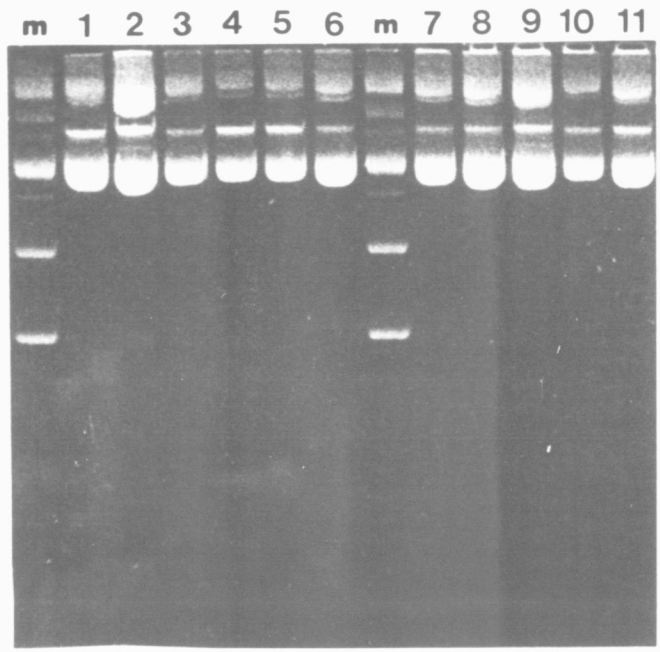
A, the circular DNA band from isolate 13 which possess the mobility expected for a cloned tetramer of SV40 DNA (see legend to Fig.5.16B) was purified from an agarose gel slice (see section 2.10). Competent cells prepared from E.coli strain HB101 was transformed with the purified DNA as described (see section 2.12.5). DNA was prepared from eleven independently derived transformants using the mini cleared lysate method. The DNAs were fractionated on a 0.6% (w/v) agarose gel (tracks 1 to 11). The marker tracks (m) contain DNA prepared from isolate 13.

B, DNA prepared from two of the transformants were digested with SalI endonuclease and fractionated on a 0.55% (w/v) agarose gel (tracks 1 and 2). The marker tracks (m) contain undigested Ad2 DNA plus EcoRI digested Ad2, pDM103, pDM238 and Homer I DNAs. The marker fragment sizes in Kb are: 37, 21.6, 17, 12, 5.5, 4.4, 3.8, 3.0, 2.1 and 1.9.

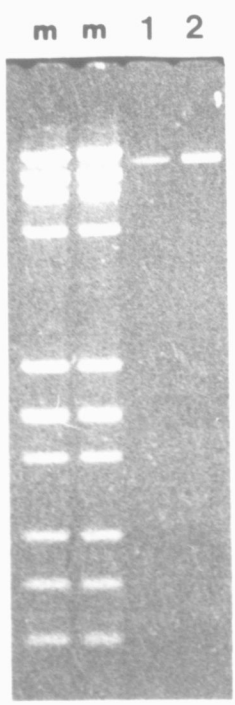
C, DNA prepared from two of the transformants were digested with EcoRI endonuclease and fractionated on a 0.55% (w/v) agarose gel (tracks 1 and 2). The marker tracks (m) contain linear Homer I DNA (5.5Kb) and linear SV40 DNA (5.2Kb).

Fig. 5.17

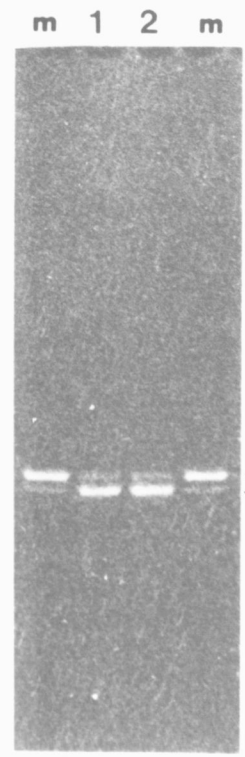
A



B



C



size of circular DNA molecules was observed. Two of the DNA preparations were digested with EcoRI endonuclease and in each case two fragments corresponding in size to linear monomeric Homer I and linear monomeric SV40 DNA were generated (Fig.5.17B). The relative intensity of the two bands suggests that this recombinant plasmid was a cloned oligomer of SV40 DNA. SalI endonuclease digestion of the same two DNA preparations yielded a single fragment the size of which is consistent with that of a tetramer of SV40 DNA cloned in Homer I (Fig.5.17C). These results suggest that whilst large 'head to tail' polymers of SV40 (repeats of 6 to 9 monomers) cloned in Homer I are unstable when propagated in strain HB101, cloned oligomers of SV40 DNA of smaller size can be propagated in a stable fashion.

5.9 Construction of Homer II

For the purposes of cloning integrated SV40 DNA from the genomes of stably transformed cell lines or from DNA prepared from infected nonpermissive cell populations, it is useful to have a cloning vector that contains a single site for a restriction endonuclease which does not cleave SV40 DNA but which will extensively digest mammalian DNA. Such a vector would allow the cloning of the whole of the integrated viral genome together with both viral-cellular DNA junctions in a single recombinant molecule. Homer I does not contain a suitable restriction endonuclease site. λ DNA contains two recognition sites for SstI endonuclease. Following digestion with HindIII and EcoRI endonucleases, the SstI sites are located on two fragments, a 1.95Kb HindIII fragment and

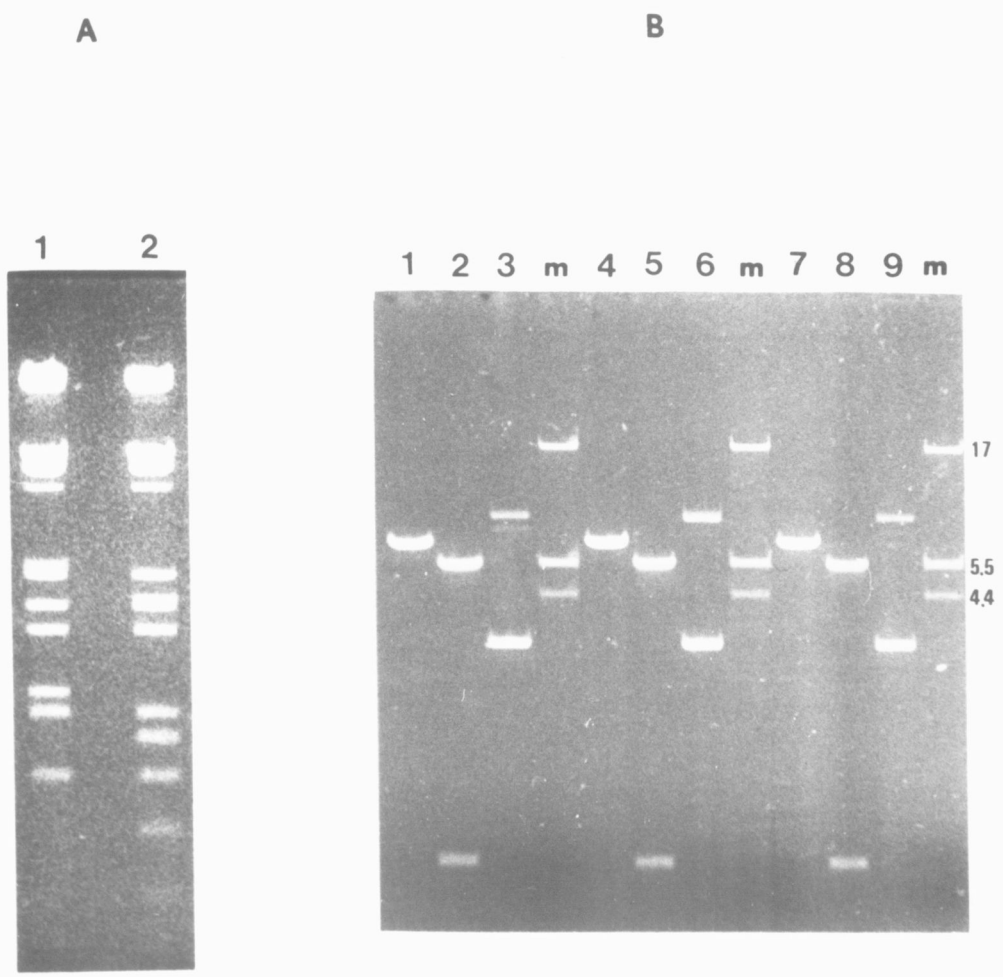
a-0.96Kb fragment with one EcoRI terminus and one HindIII terminus (Fig.5.18A). The 0.96Kb fragment was purified from an agarose gel slice. Homer I was digested with EcoRI and HindIII endonucleases and the digest was fractionated by agarose gel electrophoresis. The larger HindIII - EcoRI fragment (the two enzymatic recognition sites are located 29 bp apart) was purified from a gel slice and ligated to the purified 0.96Kb λ DNA fragment. Following transfection of competent E.coli strain HB101 cells with the ligation reaction mixture, ampicillin resistant colonies were obtained. Plasmid DNA was prepared from three independent isolates. The DNAs were digested with EcoRI and HindIII endonucleases, SstI endonuclease, or left undigested and then fractionated by agarose gel electrophoresis. The results shown in Fig.5.18B indicate that the 0.96Kb λ DNA fragment containing the SstI endonuclease site was successfully cloned into Homer I in all three cases. This cosmid suitable for cloning with SstI is called Homer II.

Legend to Figure 5.18

Construction of Homer II

A, λC1857 Sam7 DNA was digested with EcoRI plus HindIII endonucleases (track 1) or EcoRI plus HindIII plus SstI endonucleases (track 2) and fractionated on a 1.1% (w/v) agarose gel. The smaller of the two HindIII/EcoRI λ fragments which is cleaved by SstI (the third band from the bottom in track 1) was purified from an agarose gel slice as described (section 2.10) and ligated to the larger EcoRI/HindIII fragment of Homer I (see text). Competent cells prepared from E.coli strain HB101 were transformed with the ligated DNA.
B, DNA was prepared from three independently derived transformants using the mini cleared lysate method. The DNAs were digested with SstI endonuclease (tracks 1, 4 and 7), EcoRI plus HindIII endonucleases (tracks 2, 5 and 8) or left undigested (tracks 3, 6 and 9). These samples were fractionated on a 0.6% (w/v) agarose gel. The marker tracks (m) contain EcoRI digested pDM103 and Homer I DNAs.

Fig. 5.18



5.10 Attempt to clone integrated viral DNA at early times following the infection of mouse cells with SV40

The strategy employed was to partially digest with SstI total DNA prepared from infected cells in order to generate large (40Kb) fragments containing integrated viral DNA. These DNA fragments were ligated to a five-fold molar excess of Homer II DNA linearised with SstI at a final total DNA concentration of 300 ug per ml. The ligated DNA was packaged into λ phage particles in vitro. The transductants obtained were screened, in duplicate, with ^{32}P-labelled SV40 DNA using the high density screening procedure of Hanahan and Meselson (1980) at a density of between 1 x 10^4 and 2 x 10^4 colonies per 9cm filter. Approximately 2.5 x 10^5 independently derived transductants were generated from the DNA prepared at 85h P.I. from wt830-infected mouse cells maintained in medium supplemented with 10% FCS (the DNA preparation is the one shown in Fig.4.7, track 7). No duplicate positive signals were obtained and no positive clones were obtained from the few candidates (signal present on only one of the two duplicate filters) which were rescreened at lower density. Since it is a reasonable assumption that only a small proportion of the population of the infected cells would have acquired integrated viral DNA, it is very likely that the number of recombinants screened (which amounts to only 1 to 2 genome equivalents of mammalian DNA) was not sufficiently large to ensure the cloning of integrated viral DNA.

Similarly, 1.5 x 10^5 independently derived transductants were generated and screened, in duplicate, from DNA prepared 10 days P.I. from mouse cells mixedly infected with dl(pm)861(MOI=100) and

dlF-884 (MOI=100) and maintained on medium supplemented with 2% FCS. The rationale of this experiment was that the proportion of cells containing integrated viral DNA would be increased if the infected cells were maintained under serum conditions which do not allow untransformed cells to grow. One positive signal was obtained in duplicate. Following two rounds of rescreening at lower density to obtain a pure isolate, DNA was prepared from this recombinant clone by caesium chloride ethidium bromide density gradient equilibrium centrifugation. This DNA was digested with restriction endonucleases and analysed by agarose gel electrophoresis and transfer hybridisation. The results (Fig.5.19) indicate the following:

(1) following SstI digestion only a single DNA fragment containing SV40 sequences is observed; the size of this fragment is approximately that of monomeric SV40 DNA;

(2) this SstI fragment contains single recognition sites for BstI (0.14 map units) and TaqI (0.56 map units) but does not appear to contain a site for MspI (HpaII).

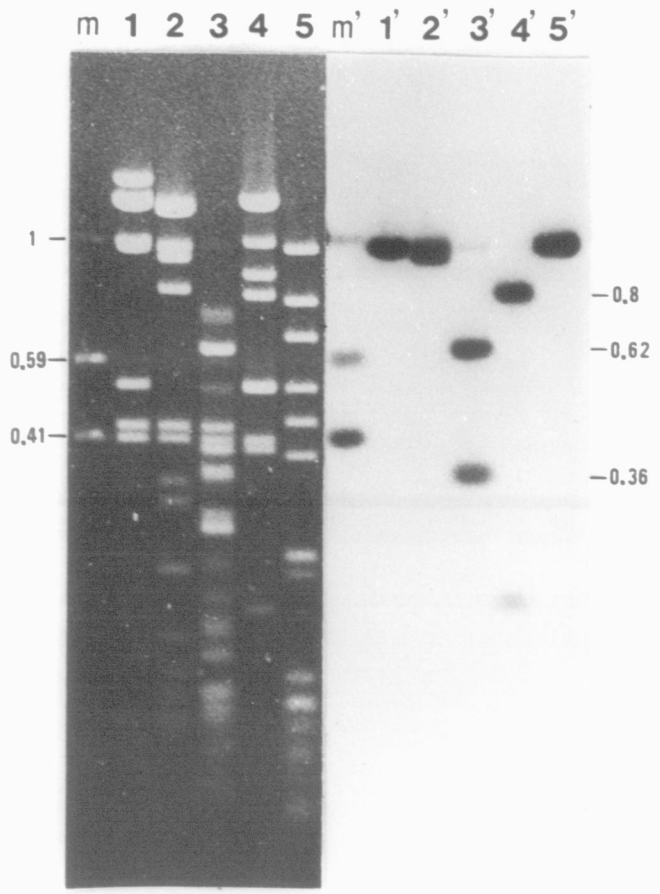
Since the SstI fragment containing SV40 sequences is approximately the size of monomeric SV40 DNA and since dl(pm)861 (one of the viruses used for the infection) is deleted for the HpaII site, these results suggest that the SV40 DNA present in the recombinant clone originated from a defective viral genome which had acquired an SstI site. The size of the SstI/TaqI fragments (0.62 and 0.36 SV40 fractional lengths) and SstI/BstI fragments (0.8 and less than 0.36 SV40 fractional lengths) is consistent with the SstI site being located at approximately 0.93 map units. This was confirmed by transfer hybridisation experiments with HinfI digested (see Fig.1.1

Legend to Figure 5.19

The structure of a cloned SstI fragment containing SV40 sequences cloned from SV40-infected mouse cellular DNA

The cloned DNA (see text) was digested with various restriction endonucleases and electrophoresed on a 0.6% (w/v) agarose gel. The unprimed tracks are photographs of the fractionated DNAs stained with ethidium bromide. The primed tracks are autoradiograms of the corresponding tracks following transfer hybridisation with ^{32}P-labelled SV40 DNA. The tracks are: 1, SstI-digested DNA; 2, SstI/EcoRI-digested DNA; 3, SstI/TaqI-digested DNA; 4, SstI/BstI-digested DNA; 5, SstI/MspI-digested DNA; m, marker track containing monomeric linear SV40 DNA and MspI/BstI-digested SV40 DNA. The sizes of the fragments are given as SV40 fractional lengths.

Fig.5.19



for the locations of cleavage sites) and HinfI plus SstI digested DNA. The results of this experiment (not shown) indicate that SstI cleaves the HinfI fragment A approximately 350bp from one end, which again places the SstI site at 0.93 map units on the SV40 genome. The possibility that this putative defective genome may contain small deletions and/or substitutions was not pursued. The formal possibility that this SstI fragment does represent integrated viral DNA, with two cellular SstI sites located close to the viral-cellular joints, cannot be eliminated. However, these preliminary experiments strongly suggest that the recombinant clone in question does not contain an integrated viral genome but rather contain a cloned defective SV40 genome which had acquired an SstI site located at 0.93 map units on the SV40 genome.

5.11 Discussion

This chapter describes the construction of biologically disabled cosmid vectors and the technology involved in their use. These vectors can be used in conjunction with λ *in vitro* packaging systems to clone eukaryotic DNA fragments at efficiencies comparable to those obtainable with λ cloning vectors. Analyses of recombinants obtained under a number of conditions have indicated that the best results are obtained when large molar excesses of dephosphorylated vector molecules are used in the ligation reactions. Under optimal conditions cloning efficiencies of 5×10^4 recombinants per μg of target DNA can be routinely achieved with essentially no vector background. The recombinants obtained under these conditions have an average insert size of 36Kb; the size distribution of the clones obtained indicates that the great majority of the recombinant molecules

can be stably propagated in E.coli HB101. Analysis of the rDNA content of a complete D.melanogaster genomic library constructed with Homer I indicates that the representation of these DNA sequences in the genomic library is in excellent agreement with the representation of these sequences in the D.melanogaster genome. This method of cloning large fragments of genomic DNA should be generally applicable.

One of the reasons for developing this cloning system was the advantage of being able to propagate recombinants of large size in recA⁻ strains of E.coli. We hoped that, by using recA⁻ strains, recombinants containing tandemly repeated SV40 sequences could be stably propagated. This was important because some of the sequences which we wanted to clone were thought to contain tandemly arranged "head to tail" oligomers of SV40 DNA (see Chapter 4). The experiments performed using in vitro generated "head to tail" polymers of SV40 DNA suggest that cloned "head to tail" SV40 polymers of packageable size (6 to 9 repeats of SV40 DNA) are not stably propagated in E.coli strain HB101 but break down to form a heterogeneous population of smaller "head to tail" oligomers of SV40 DNA. However, a cloned tetramer of SV40 DNA can be stably propagated. Attempts were made to clone integrated SV40 genomes and their flanking cellular sequences from infected mouse cells at early times following infection. This experiment was performed with Homer II using SstI endonuclease which does not cleave SV40 DNA and cleaves Homer II at a single site. However, despite having taken the usual precaution of preparing virus stocks from low MOI (0.05) infections, the only recombinant containing

SV40 sequences obtained, from approximately 4 x 10^5 independently derived clones screened, appears to contain a defective SV40 genome which has acquired a SstI site. Further discussion of these experiments will be given in Chapter 6.

Chapter 6

General Discussion

Restriction endonuclease mapping studies have shown that mouse cells stably transformed by SV40 contain viral DNA covalently integrated into their chromosomal DNA and that the integration of viral DNA is not specific with respect to either the viral or cellular genomes. Moreover, the integrated viral genomes tend to contain tandem repeats of full length viral DNA in transformants derived from semipermissive infection and partially tandemly duplicated viral genomes in viral transformants derived from nonpermissive infection (see section 1.5). The high frequency with which these tandem duplications occur raises questions about the mechanism of viral DNA integration, since it is not obvious how recombination between monomeric, circular viral DNA and linear, chromosomal DNA could generate this type of structure.

There are several models which can potentially explain the formation of tandemly duplicated integrated viral genomes. Firstly, homologous recombination between monomeric integrated viral genomes and free copies of viral DNA could lead to the formation of tandemly duplicated viral genomes; however, the results of the retransformation experiments of Graessmann *et al.* (1979) and M. Botchan (personal communication) argue against this possibility (see section 1.5). Secondly, *in situ* replication of integrated unit-sized viral DNA followed by recombination between the newly replicated and the integrated viral DNA (amplification) can generate these tandemly duplicated structures, as suggested by Colantuoni *et al.* (1980); however, since

only transformed lines which contain integrated viral DNA with partial tandem duplications can undergo amplification (see section 1.5), it seems more likely that the events observed by Colantuoni et al. (1980) are due to unequal recombination between the repeated portions of the viral insert and not amplification of the integrated viral DNA. Therefore, their results do not provide evidence for amplification as a mechanism for the generation of tandemly duplicated viral DNA from an initial monomeric integrated viral genome. Thirdly, it is possible that the viral DNA which initially integrates into cellular DNA is oligomeric; however, while the accumulation of viral oligomers and polymers has been observed late in permissive infection (Martin et al., 1976; Rigby and Berg, 1978), such forms of viral DNA had not, prior to this work, been reported following non-permissive infection.

In this study the mechanism of viral DNA integration was approached by following the fate of infecting viral DNA using the transfer hybridisation technique of Southern (1975). The data show that shortly after the infection of mouse cells with SV40 there appears a novel form of viral DNA that copurifies with high molecular weight cellular DNA. Approximately 12% of the total SV40 DNA present in the infected cells is in this high molecular weight form. Restriction endonuclease mapping experiments clearly indicate that the great majority of these high molecular weight viral sequences are not the result of covalent integration of monomeric SV40 DNA into chromosomal DNA; rather they represent large linear polymers of viral DNA in which the genomes are organised in tandem, "head to tail" arrays.

Experiments with arabinosyl cytosine, an inhibitor of DNA replication, demonstrate that these viral molecules are either synthesised directly by DNA replication or that their production depends indirectly on a process involving DNA replication. Furthermore, by coinfecting cells with viable deletion mutants of SV40 with physically distinguishable genomes it was shown that there is significant recombination between the two SV40 species. These results invoke two explanations. Either the polymers are synthesised by DNA replication and there is then recombination between the polymers, or the polymers are produced by recombination between the infecting monomeric viral DNA and the involvement of DNA replication is indirect. Quantition of the frequency of recombination favours the former possibility.

It is my belief that viral DNA can replicate in SV40-infected Balb/c 3T3 cells and I consider it most likely that this replication occurs by some form of rolling circle replication. It has been previously demonstrated that similar polymers are synthesised in productively infected permissive cells (Rigby and Berg, 1978); Martin and Setlow (1980) and Bjursell (1978) have demonstrated, by electron microscopy, the existence of rolling circle replicative forms in cells productively infected with SV40 or polyoma virus. Mouse cells have been previously considered to be absolutely nonpermissive for SV40 DNA replication. However, this conclusion was based on experiments, designed to detect theta-form replication, which would not have detected replication of the type proposed here (Henry et al., 1966).

The observation that mouse cells infected with a tsA mutant of SV40 accumulate the high molecular weight DNA at 39°C requires

explanation and there are several possibilities. Firstly, this could simply be due to the leakiness of the mutant, particularly since high MOI and a relatively low temperature were used (mouse cells will not survive at 41°C, the true nonpermissive temperature for the mutant). Analogous experiments with large T deletion mutants could provide a definitive answer; however, the preparation of such deletion mutant virus stocks has only very recently been made possible by the development of monkey kidney cell lines transformed by SV40 DNA which has been rendered origin-minus by *in vitro* manipulation; these cell lines can provide a trans-acting large T helper function (Y. Gluzman, personal communication). Secondly, the work of Martin and Setlow (1980, see Chapter 1) suggests that, even under nonpermissive conditions, *tsA* mutants of SV40 can initiate DNA synthesis at different, and possibly random, sites on the viral genome in productively infected permissive cells; these authors propose that large T-antigen can initiate DNA synthesis both site-specifically, at the SV40 origin of DNA replication, and also non-specifically, and that the primary defect in *tsA* mutants is in the site-specific initiation function. It is possible that the synthesis of the viral polymers depends on the non-specific initiation function which is not (completely) inactivated by *tsA* mutations even at the nonpermissive temperature. Thirdly, large T function(s) may not be required for the formation of viral polymers in cells nonpermissive for SV40 DNA replication. Harland and Laskey (1980) have shown that all DNA sequences tested, including circularised fragments of SV40 DNA, can replicate when microinjected into *Xenopus laevis* eggs regardless of whether they include a "eukaryotic origin of DNA replication".

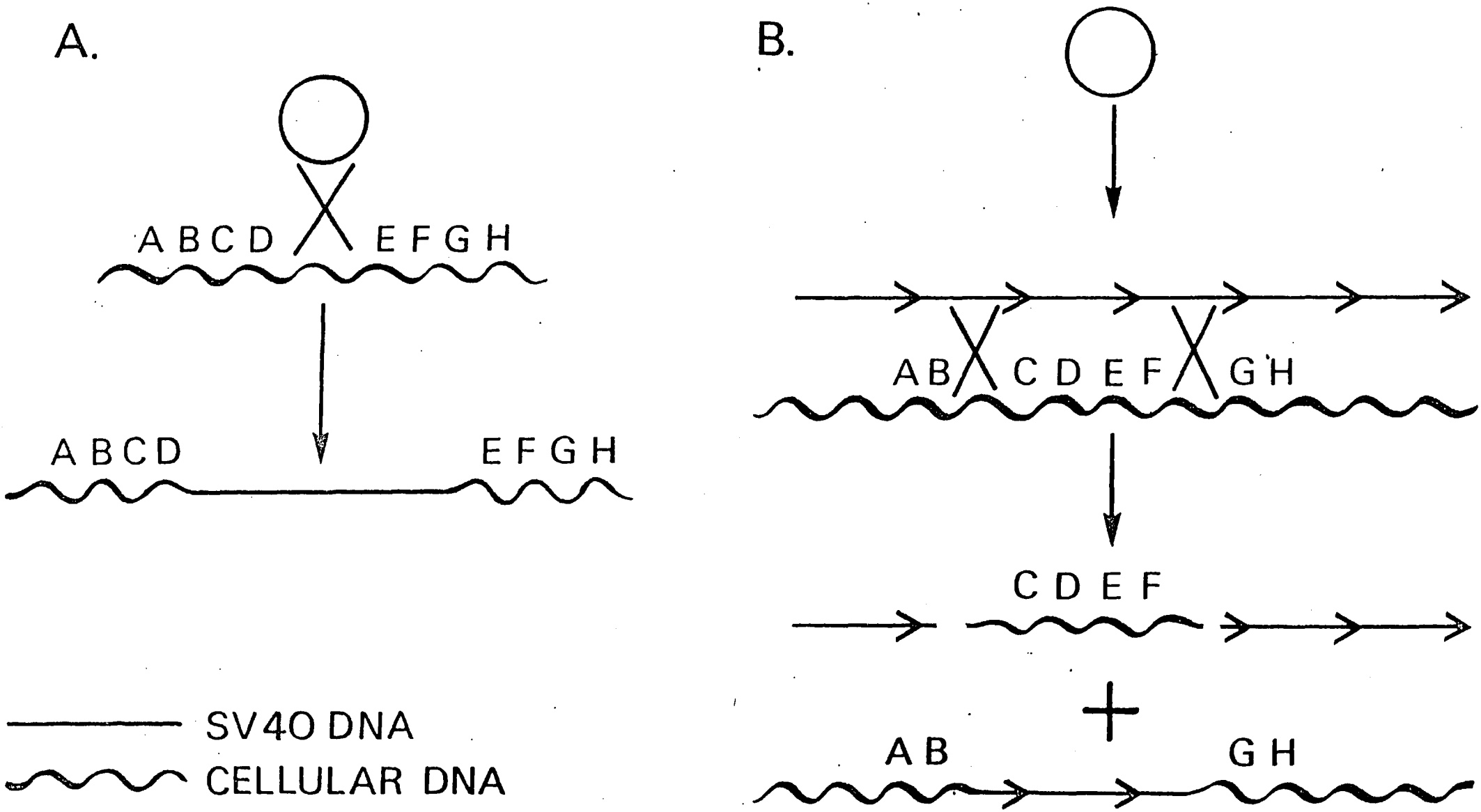
Moreover, the replication of the injected DNAs appears to be under normal cellular control, i.e. only one round of replication of the injected DNA is observed per cell cycle. Therefore, at least in this case, the eukaryotic DNA replication machinery can initiate and regulate DNA replication on a variety of DNA sequences without the requirement for an "origin of replication" or, in the case of micro-injected, cyclised fragments of SV40 DNA, without a requirement for large T-antigen. These authors further propose that perhaps the role of the viral origin of replication and the large T-antigen is to uncouple viral replication from cellular control such that viral DNA replication can be reinitiated many times in the course of a single cell cycle, as is the case in productively infected permissive cells. Following this line of reasoning, it is possible that in infected cells that are nonpermissive for SV40 DNA replication some proportion of the infecting viral DNA may be able to replicate via the normal cellular replication machinery without requiring large T-antigen; the viral polymers observed may be the result of this type of replication.

Regardless of the precise manner by which these viral polymers are produced, the experimental evidence presented is consistent with the following model for the mechanism of the integration of SV40 DNA into the chromosomal DNA of mouse cells (Fig.6.1). Shortly after infection long polymers of viral DNA are synthesised. These polymers then recombine effectively with each other and with the chromosomal DNA. Thus the immediate precursor to covalently integrated viral DNA is not monomeric, circular viral DNA but polymeric viral DNA which we believe to be linear. Recombination between the polymers

Legend to Figure 6.1

Model for the integration of SV40 DNA into cellular DNA

Fig.6.1


A.
A B C D
E F G H
A B C D
E F G H
SV40 DNA
CELLULAR DNA
B.
A B
C D E F
G H
C D E F
A B
G H

and chromosomal DNA can directly explain the high frequency with which viral insertions containing tandem duplications are found because the precursor to the integrated viral DNA is already tandemly duplicated. It should be stressed that there is no direct evidence that the observed viral polymers are involved in viral integration. However, this model does make specific testable predictions. Recombination between two linear molecules requires two cross-over events in contrast to the single cross-over required to integrate a monomeric circle. It is therefore possible to distinguish between these two models by analysing the organisation in untransformed cells of the sequences that flank the integrated viral genomes in transformed cells. If integration involves a single cross-over then the sequences, "CD" and "EF" in Fig.6.1, flanking the integrated viral DNA in the transformed cell should be immediately adjacent in the parental untransformed cell. However, recombination involving two cross-overs will necessarily lead to a deletion of chromosomal DNA at the site of the viral insertion. Thus the sequences flanking the integrated viral genome, "AB" and "GH" in Fig.6.1, will be non-contiguous in the DNA of the parental untransformed cells. It is of interest that the cellular DNA sequences flanking the left hand side of the cloned viral insertion of the SV40-transformed rat line 14B are indeed rearranged when compared with the corresponding DNA of untransformed rat cells (Botchan et al., 1980). In this particular case, it is unclear whether the rearrangement involves deletion, inversion or translocation. However, a number of integrated viral genomes, together with their flanking cellular sequences, have now

been cloned and the generality of this prediction of the model has been found to be true, i.e. there appear to be deletions of flanking cellular DNA in all of the cases examined (J. Sambrook; C.E. Clayton and P.W.J. Rigby, personal communications). It should be stressed that these results which were obtained from stably transformed cell lines must be interpreted with care. Since there is recent evidence which indicate that SV40 transformed cell lines can segregate progeny cell lines that exhibit nonparental viral integration patterns (M. Lovett, C.E. Clayton and P.W.J. Rigby; J. Hiscott, D. Murphy and V. Defendi, personal communications), it is unclear whether the observed deletions of cellular sequences arose as the direct result of viral integration or if they resulted from rearrangement events which occurred subsequent to viral integration.

A further prediction of this model is that there should be integrated viral DNA containing complete tandem duplications (at least two complete viral copies) of the viral genome following SV40 infection of nonpermissive cells. However, all of the tandemly duplicated integrated viral DNAs that have been examined in stable transformants derived from nonpermissive cells contain only a partial tandem duplication of the viral genome. It should be stressed that a general problem associated with any attempt to deduce the "primary" viral integration events from the structure of the integrated viral genomes found in stable transformants is that the stable transformants are temporally very far removed from the "primary" events. Therefore, any event which occur subsequent to the primary integration events would complicate the interpretation of the results obtained

from the study of stable transformants. Thus it is possible that the initial integration of viral DNA into the genomes of nonpermissive cells does generate complete tandem duplications of viral DNA, but that these structures are unstable and are modified, possibly by selective factors, such that by the time they are examined in stable transformants only partial tandem duplications of the viral DNA remain. In order to overcome this problem and to test this prediction of the model, it is necessary to examine the products of these "primary" events as soon as possible after they have occurred. This approach necessarily confines one to working with a heterogeneous population of infected nonpermissive cells.

In order to study integrated viral sequences in such a heterogeneous population, it is necessary to clone these sequences. However, any cloning system to be used for this purpose must be able to accomodate relatively large fragments of DNA and be able to stably propagate sequences containing tandemly duplicated SV40 genomes, because the DNA sequences to be cloned are expected to have these properties. Furthermore, the vector used must be able to accomodate the use of a restriction endonuclease which does not cleave SV40 DNA but which extensively cleaves mammalian DNA. None of the existing λ cloning systems satisfied these criteria and I therefore developed my own cloning system which is applicable not only to this experiment but should be generally useful (see Chapter 5). However, despite taking the usual precautions, the only recombinant containing SV40 sequences that was obtained from approximately 4×10^5 independently derived clones generated from SV40 infected mouse cellular DNA appeared to contain a defective SV40 genome which had acquired an SstI site.

While this particular experiment failed, this approach for attempting to clone the products of "primary" integration events is worth pursuing. In experiments of this kind there are at least two factors which must be treated with care and these were not sufficiently well handled in the particular experiment that was described (see section 5.10). Firstly, the probability of cloning defective viral genomes containing SstI sites must be minimised. Besides the obvious precaution of using virus stocks prepared from low multiplicity infections, the DNA fragments to be cloned should also be dephosphorylated prior to ligation. The dephosphorylation step should prevent the cloning of defective monomeric viral genomes containing SstI sites. The problem of cloning unintegrated viral polymers which include defective viral genomes containing SstI sites cannot be eliminated. However, since only approximately 10% of the total viral DNA is in this form and since defective viral genomes containing SstI sites should be rare, this should not be a major problem. Secondly, DNA used for these experiments should be prepared from an infected cell population in which the proportion of cells containing integrated viral DNA is optimal. While it is unclear how this can be accomplished, it is reasonable to assume that growth conditions which maximise both the transformation frequency and the growth rate of transformed cells relative to untransformed cells will also by optimal for this variable. In practice, this means infecting subconfluent cells with high MOIs of virus and growing the infected cells in medium supplemented with 2% (v/v) FCS. These steps should increase the probability of obtaining clones containing integrated viral DNA at early times following infection.

Of the many experiments which should have been performed but were not, two, in particular, stand out. Density labelling experiments should have been performed to determine whether the observed viral polymers were newly synthesised. Transfer hybridisation experiments should also have been performed to examine the structure of the integrated viral DNA from stable transformants derived from mixed infections using physically distinguishable mutants of SV40. If integrated viral DNA consisting of recombinant viral genomes was found, one could make a stronger case for the observed viral polymers being precursors to integrated viral DNA.

Nevertheless, I have characterised a novel form of viral DNA which is synthesised shortly after the infection of nonpermissive mouse cells with SV40 virions. These tandemly arrayed viral polymers are recombinationally active and I believe that they also recombine with chromosomal DNA and thus are the immediate precursors to integrated viral DNA. If this model is correct it readily explains the high frequency of tandemly duplicated integrated viral genomes found in stably transformed cells. The model also makes specific predictions which can be tested experimentally by using the cosmid cloning technology which I have developed.

Bibliography

Abrahams, P.J., Mulder, C., Van De Voorde, A., Warnaar, S.O. and van der Eb, A.J. (1975). J. Virol. 16, 818.

Alwine, J.C., Reed, S.I. and Stark, G.R. (1977). J. Virol. 24, 22.

Basilico, C., Gattoni, S., Zouzias, D. and Della Valle, G. (1979). Cell 17, 645.

Basilico, C., Matsuga, Y. and Green, H. (1970). Virology 41, 295.

Becker, A. and Gold, M. (1975). Proc. Natl. Acad. Sci. USA. 72, 581.

Becker, A., Marko, M. and Gold, M. (1977). Virology 78, 291.

Becker, A., Murialdo, H. and Gold, M. (1977). Virology 78, 227.

Benton, W.D. and Davis, R.W. (1977). Science 196, 180.

Berk, A.J. and Sharp, P.A. (1978). Proc. Natl. Acad. Sci. 75, 1274.

Birg, F., Dulbecco, R., Fried, M. and Kamen, R. (1979). J. Virol. 29, 633.

Birkenmeier, E.H., May, E. and Salzman, N.P. (1977). J. Virol. 22, 702.

Bjursell, G. (1978). J. Virol. 26, 136.

Blattner, F.R., Williams, B.G., Blechl, A.E., Thompson, K.D., Faber, H.E., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L., and Smithies, O. (1977). Science 196, 161.

Bolivar, F., Rodriguez, R.L., Betlach, M.C. and Boyer, H.W. (1977a). Gene 2, 75.

Bolivar, F., Rodriguez, R.L., Green, P.J., Betlach M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977b). Gene 2, 95.

Botchan, M., Stringer, J., Mitchison, T. and Sambrook, J. (1980). Cell 20, 143.

Botchan, M., Topp, W. and Sambrook, J. (1976). Cell 9, 269.

Botchan, M., Topp, W.C. and Sambrook, J. (1979). Cold Spring Harbor Symp. Quant. Biol. 43, 709.

Bouck, N., Beale, N., Shenk, T., Berg, P. and di Mayorca (1978). Proc. Natl. Acad. Sci. USA 75, 2473.

Brockman, W.W. (1978). J. Virol 25, 860.

Brockman, W.W., Gutai, M.W. and Mathews, D. (1975). Virology 66, 36.

Brugge, J.S. and Butel, J.S. (1975). J. Virol. 15, 619.

Butel, J.S. and Soule, H.R. (1978). J. Virol. 26, 584.

Campo, S.M., Cameron, I.R. and Rogers, M.E. (1978). Cell 15, 1411.

Carbon, J., Shenk, T.E. and Berg, P. (1975). Proc. Natl. Acad. Sci. USA. 72, 1392.

Carroll, R.B., Hager, L. and Dulbecco, R. (1974). Proc. Natl. Acad. Sci. USA. 71, 3754.

Carroll, R.B. and Smith, A.E. (1976). Proc. Natl. Acad. Sci. USA. 73, 2254.

Casey, J. and Davidson, N. (1977). Nucleic Acid Research 4, 1539.

Chepelinsky, A.B., Seif, R. and Martin, R.G. (1980). J. Virol. 35, 184.

Chou, J.Y., Avila, J. and Martin, R.G. (1974). J. Virol. 14, 116.

Chou, J.Y. and Martin, R.G. (1975). J. Virol. 15, 145.

Clark, L. and Carbon, J. (1976). Cell 9, 91.

Colantuoni, V., Dailey, L. and Basilico, C. (1980). Proc. Natl. Acad. Sci. USA. 77, 3850.

Cole, C.N., Landers, T., Goff, S.P., Manteuil-Brutlag, S. and Berg, P. (1977). J. Virol, 24, 277.

Collins, J. and Bruning, H.J. (1978). Gene 4, 85.

Collins, J. and Hohn, B. (1978). Proc. Natl. Acad. Sci. USA. 75, 4242.

Collins, C.J. and Sauer, G. (1972). J. Virol. 10, 425-432.

Contreras, R., Rogiers, A., van de Voorde, A. and Fiers, W. (1977). Cell 12, 529.

Cook, H.J. and Hindley, J. (1979). Nucleic Acid Res. 6, 3177.

Cowan, K., Tegtmeyer, P. and Anthony, D.D. (1973). Proc. Natl. Acad. Sci. USA. 70, 1927.

Crawford, L.V., Cole, C.N., Smith, A.E., Paucha, E., Tegtmeyer, P., Rundell, K. and Berg, P. (1978). Proc. Natl. Acad. Sci. USA. 75, 117-121.

Croce, C.M. (1977). Proc. Natl. Acad. Sci. USA. 74, 315.

Culp, L.A. and Black, P.H. (1972). J. Virol. 9, 611.

Dale, R.M.K., Livingston, D.C. and Ward, D.C. (1973). Proc. Natl. Acad. Sci. USA. 70, 2238.

Dale, R.M.K., Martin, E., Livingston, D.C. and Ward, D.C. (1975). Biochemistry 14, 2447.

Dale, R.M.K. and Ward, D.C. (1975). Biochemistry 14, 2458.

Danna, K.J. and Nathan, D. (1972). Proc. Natl. Acad. Sci. USA. 69, 3097.

Davis, R.W., Simon, M., and Davidson, N. (1971). Methods in Enzymology 21D, 413.

Denhardt, D.T. (1966). Biochem Biophys Res. Commun. 23, 641.

Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J. and Weissman, S.M. (1975). Cold Spring Harbor Symp. Quant. Biol. 39, 153.

Edwards, C.F., Khoury, G. and Martin, R.G. (1979). J. Virol. 29, 753.

Efstratiadis, A., Vournakis, J.N., Donis-Keller, H., Chaconas, G., Dougall, D.K. and Kafatos, F.C. (1977). Nucleic Acid Research 4, 4165.

Ellman, G.L. (1959). Arch. Biochem. Biophys. 82, 70.

Estes, M.K., Huang, E.S. and Pagano, J.S. (1971). J. Virol. 7, 635.

Fanning, E., Nowak, B. and Burger, C. J. Virol. In press.

Fareed, G.C., Garon, C.F. and Salzman, N.P. (1972). J. Virol. 10, 484.

Feunteun, J., Kress, M., Gardes, M. and Monier, R. (1978). Proc. Natl. Acad. Sci. USA. 75, 4455.

Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., van de Voorde, A., van Heiverswyn, H., van Herreweghe, J., Volckaert, G. and Ysebart, M. (1978). Nature 273, 113.

Fluck, M.M. and Benjamin, T.L. (1979). Virology 96, 205.

Ford, J.P. and Hsu, M.T. (1978). J. Virol. 28, 795.

Fox, T.D. and Levine, A.J. (1971). J. Virol. 7, 473.

Fried, M., Weissman, C., Murray, K., Murray, N., Greenaway, P.J. and Tooze, J. (1979). Nature 279, 811.

Gelb, L.D., Kohn, D.E. and Martin, M.A. (1971). J. Mol. Biol. 57, 129.

Gershon, D., Sachs, L. and Winocour, E. (1966). Proc. Natl. Acad. Sci. USA. 56, 918.

Ghosh, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P. and Weissman, S.M. (1978a). J. Mol. Biol. 126, 813.

Ghosh, P.K., Reddy, V.B., Swinscoe, J., Choudary, P.V., Lebowitz, P. and Weissman, S.M. (1978b). J. Biol. Chem. 253, 3643.

Glover, D.M. (1977). Proc. Natl. Acad. Sci. USA. 74, 4932.

Glover, D.M. and Hogness, D.S. (1977). Cell 10, 167.

Glover, D.M., White, R.L., Finnegan, D.J. and Hogness, D.S. (1975) Cell 5, 149.

Glutzman, Y., Davison, J., Oren, M. and Winocour, E. (1977). J. Virol. 22, 256.

Goff, S.P. and Berg, P. (1977). J. Virol. 24, 295.

Graessmann, M. and Graessmann, A. (1976). Proc. Natl. Acad. Sci. USA. 73, 366.

Graessmann, A., Graessmann, M., Tijan, R. and Topp, W.C. (1980). J. Virol. 33, 1182.

Graessmann, A., Graessmann, M., Topp, W.C. and Botchan, M. (1979). J. Virol. 32, 989.

Graham, F.L., Abrahams, P.J., Mulder, C., Heijneker, H.L., Warnaar, S.O., de Vries, F.A.J., Fiers, W. and van der Eb, A.J. (1974). Cold Spring Harbor Symp. Quant. Biol. 39, 637.

Greene, P.J., Poonian, M.S., Nussbaum, A.L., Tobias, L., Garfin, D.E., Boyer, H.W. and Goodman, H.M. (1975). J. Mol. Biol. 99, 237-261.

Griffin, J.D., Spangler, G. and Livingston D.M. (1979). Proc. Natl. Acad. Sci. USA. 76, 2610.

Grodzicker, T., Anderson, C., Sharp, P.A. and Sambrook, J. (1974). J. Virol. 13, 1237.

Gross-Bellard, M., Oudet, P. and Chambon, P. (1973). Eur. J. Biochem. 36, 32.

Grunstein, M. and Hogness, D.S. (1975). Proc. Natl. Acad. Sci. USA. 12, 3961.

Gruss, P., Lai, C.J., Dhar, R. and Khoury, G. (1979). Proc. Natl. Acad. Sci. USA. 76, 4317.

Gutai, M.W. and Nathans, D. (1978a). J. Mol. Biol. 126, 259.

Gutai, M.W. and Nathans, D. (1978b). J. Mol. Biol. 126, 275.

Haegeman, G. and Fiers, W. (1980). J. Virol. 35, 955.

Haggerty, D.M. and Schleif, R.F. (1976). J. Virol. 18, 659.

Hamer, D.H., Smith, K.D., Boyer, S.H. and Leder, P. (1979). Cell 17, 725.

Hanahan, D. and Meselson, M. (1980). Gene 10, 63.

Harland, R.M. and Laskey, R.A. (1980). Cell 21, 761.

Hassell, J.A., Lukanidin, E., Fey, G. and Sambrook, J. (1978). J. Mol. Biol. 120, 209.

Henry, P., Black, P.H., Oxman, M.N. and Weissman, S.M. (1966). Proc. Natl. Acad. Sci. USA. 56, 1170.

Hirai, K. and Defendi, V. (1974). Cold Spring Harbor Symp. Quant. Biol. 39, 325.

Hirai, K. and Defendi, V. (1976). Virology 69, 229.

Hirt, B. (1967). J. Mol. Biol. 26, 365.

Hohn, B. and Murray, K. (1977). Proc. Natl. Acad. Sci. USA. 74, 3259.

Hunter, T. and Gibson, W. (1978). J. Virol. 28, 240.

Jeffreys, A.J. and Flavell, R.A. (1977). Cell 12, 429.

Jelinek, W.R. (1978). Proc. Natl. Acad. Sci. USA. 75, 2679.

Jelinek, W.R., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. (1980). Proc. Natl. Acad. Sci. USA. 77, 1398.

Jovin, T.M., Englund, P.T. and Bertsch, L.L. (1969). J. Biol. Chem. 244, 2996.

Kaiser, D., Syvanen, M. and Masuda, T. (1975). J. Mol. Biol. 91, 175.

Kamp, D., Kahmann, R., Zipser, D. and Roberts, R.J. (1977). Molec. Gen. Genet. 154, 231.

Kelley, S., Bender, M.A.R. and Brockman, W.W. (1980). J. Virol. 33, 550.

Kelly, T.J. and Lewis, A.M. (1973). J. Virol. 12, 643.

Ketner, G. and Kelly, T.J. (1976). Proc. Natl. Acad. Sci. USA. 73, 1102-1106.

Khoury, G., Byrne, J.C., Takemoto, K.K. and Martin, M.A. (1973). J. Virol. 11, 54.

Khoury, G., Howley, P., Nathans, D. and Martin, M. (1975). J. Virol. 15, 433.

Khoury, G., Martin, M.A., Lee, T.N.H. and Nathans, D. (1975). Virology 63, 263.

Khoury, G. and May, E. (1977). J. Virol. 23, 167.

Kidd, S.J. and Glover, D.M. (1980). Cell 19, 103-119.

Kimura, G. and Dulbecco, R. (1972). Virology 49, 394.

Kimura, G. and Itagaki, A. (1975). Proc. Natl. Acad. Sci. USA. 72, 673.

Koprowski, H., Jensen, F.C. and Steplewski, Z. (1967). Proc. Natl. Acad. Sci. USA. 58, 127.

Kucherlapati, R.S., Hwang, P., McDougall, J.K. and Botchan, M.R. (1978). Proc. Natl. Acad. Sci. USA. 75, 4460.

Lai, C.J. and Khoury, G. (1979). Proc. Natl. Acad. Sci. USA. 76, 71.

Lai, C.J. and Nathans, D. (1975a). J. Mol. Biol. 97, 113.

Lai, C.J. and Nathans, D. (1975b). Virology 66, 70.

Lane, D.P. and Crawford, L.V. (1979). Nature 278, 261.

Lania, L., Griffiths, M., Cooke, B., Ito, Y. and Fried, M. (1979). Cell 18, 793.

Laskey, R.A. and Mills, R.D. (1976). FEB Letters 82, 314.

Leder, P., Tiemeier, D. and Enquist, L. (1977). Science 196, 175-177.

Lewis, A.M.J.R. and Martin, R.G. (1979). Proc. Natl. Acad. Sci. USA. 76, 4299.

Lindstrom, D.M. and Dulbecco, R. (1972). Proc. Natl. Acad. Sci. USA. 69, 1517.

Lis, J.T., Prestidge, L. and Hogness, D.S. (1978). Cell 14, 901.

Macpherson, I. and Montagnier, L.M. (1964). Virology 23, 291.

Mandel, M. and Higa, A. (1970). J. Mol. Biol. 53, 159.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstradiadis, A. (1978). Cell 15, 687.

Maniatis, T. and Ptashne, M. (1973). Proc. Natl. Acad. Sci. USA. 70, 1531.

Manteuil, S. and Girard, M. (1974). Virology 60, 438.

Manteuil, S., Kopecka, H., Caraux, J., Prunell, A. and Girard, M. (1974). J. Mol. Biol. 90, 751.

Martin, M.A., Howley, P.M., Byrne, J.C. and Garon, C.R. (1976). Virology 71, 28.

Martin, R.G., Advances in Cancer Research, In Press.

Martin, R.G. and Chou, J.Y. (1975). J. Virol. 15, 599.

Martin, R.G. and Oppenheim, A. (1977). Cell 11, 859.

Martin, R.G. and Setlow, V.P. (1980). Cell 20, 381.

Martin, R.G., Setlow, V.P. and Edwards, C.A.F. (1979a). J. Virol. 31, 596.

Martin, R.G., Setlow, V.P., Edwards, C.A.F. and Vembu, D. (1979b). Cell 17, 635.

Mertz, J.E. and Berg, P. (1974). Virology 62, 112.

Mertz, J.E., Carbon, J., Herzberg, M., Davis, R.W. and Berg, P. (1974). Cold Spring Harbor Symp. Quant. Biol. 39, 69.

Meyer, B.J., Kleid, D.G. and Ptashne, M. (1975). Proc. Natl. Acad. Sci. USA. 72, 4785.

Mueller, C., Graessmann, A. and Graessmann, M. (1978). Cell 15, 579.

Murray, N.E., Brammar, W.J. and Murray, K. (1977). Molec. gen. Genet. 150, 53.

McCormick, F. and Harlow, E. (1980). J. Virol. 34, 213.

McCormick, F., Chaudry, F., Harvey, R., Smith, R., Rigby, P.W.J., Paucha, E. and Smith, A.E. (1979). Cold Spring Harbor Symp. Quant. Biol. 44, 171.

McCutchan, J.H. and Pagano, J.S. (1968). J. Nat. Cancer Inst. 41, 35.

McCutchan, T., Singer, M. and Rosenberg, M. (1979). J. Biol. Chem. 254, 3592.

Oda, K. and Dulbecco, R. (1968). Proc. Natl. Acad. Sci. USA. 60, 525.

O'Neill, F.J., Cohens, S. and Renzetti, L. (1980). J. Virol. 35, 238.

Osborn, M. and Weber, K. (1975a). Cold Spring Harbor Symp. Quant. Biol. 39, 267.

Osborn, M. and Weber, K. (1975b). J. Virol. 15, 636.

Ozanne, B. and Sambrook, J. (1971). Lepit. Colloq. Biol. Med. 3, 248.

Ozanne, B., Sharp, P.A. and Sambrook, J. (1973). J. Virol. 12, 90.

Paucha, E., Harvey, R. and Smith, A.E. (1978a). J. Virol. 28, 154.

Paucha, E., Mellor, A., Harvey, R., Smith, A.E., Hewick, R. and Waterfield, M.D. (1978b). Proc. Natl. Acad. Sci. USA. 75, 2165.

Paucha, E. and Smith, A.E. (1978). Cell 15, 1011.

Perbal, B. and Rassoulzadegan, M. (1980). J. Virol. 33, 697.

Philippsen, P., Kramer, R.A. and Davis, R.W. (1978). J. Mol. Biol. 123, 371.

Polisky, B., Green, P.J., Garfin, D.E., McCarthy, B.J., Goodman, H.M. and Boyer, H.W. (1975). Proc. Natl. Acad. Sci. USA. 72, 3310.

Prives, C., Aviv, H., Gilboa, E., Winocour, E. and Revel, M. (1975). INSERM Colloq. 47, 305.

Pollack, R. and Burger, M.M. (1969). Proc. Natl. Acad. Sci. USA. 62, 1074.

Pollack, R., Green, H. and Todaro, G.J. (1968). Proc. Natl. Acad. Sci. USA. 60, 126.

Pollack, R., Osborn, M. and Weber, K. (1975). Proc. Natl. Acad. Sci. USA. 72, 994.

Rassoulzadegan, M. and Cuzin, F. (1980). J. Virol. 33, 909.

Rassoulzadegan, M., Perbal, B. and Cuzin, F. (1978). J. Virol. 28,1.

Renger, H.C. and Basilico, C. (1973). J. Virol. 11, 702.

Reddy, V.B., Ghosh, P.K., Lebowitz, P. and Weissman, S.M. (1978a). Nucleic Acids Res. 5. 4195.

Reddy, V.B., Ghosh, P.K., Lebowitz, P., Piatak, M. and Weissman, S.M. (1979). J. Virol. 30, 279.

Reddy, V.B., Thimmappaya, B., Dhar, R., Subramian, K.N., Zain, B.S., Pan, J., Ghosh, P.K., Celma, M.L. and Weissman, S.M. (1978b). Science 200, 494-502.

Reed, S.I., Ferguson, J., Davis, R.W. and Stark, G.R. (1975). Proc. Natl. Acad. Sci. USA. 72, 1605.

Reed, S.I., Stark, G.R. and Alwine, J.C. (1976). Proc. Natl. Acad. Sci. USA. 73, 3083.

Rigby, P.W.J. (1979b). Nature 282, 781.

Rigby, P.W.J. and Berg, P. (1978). J. Virol. 28, 475.

Rigby, P.W.J., Chia, W., Clayton, C.E. and Lovett, M. (1980). Proc. Roy. Soc. Lond. B. 210, 437.

Rigby, P.W.J., Dieckman, M., Rhodes, C. and Berg, P. (1977). J. Mol. Biol. 113, 237.

Risser, R. and Pollack, R. (1974). Virology 59, 477.

Ritossa, F. and Spiegelman, S. (1965). Proc. Natl. Acad. Sci. USA. 53, 737.

Roberts, R.J. (1978). Gene 4, 183.

Robinson, C.C., Swartzendruber, D.E. and Lehman, J.M. (1980). J. Virol. 35, 246.

Rosenberg, H., Singer, M. and Rosenberg, M. (1978). Science 200, 394

Rosenthal, L.J. and Brown, M. (1977). Nucleic Acids Res. 4, 551.

Rudkin, G.T. (1972). In Results and Problems in Cell Differentiation 4, 59-85.

Sambrook, J., Sharp, P.A. and Keller, W. (1972). J. Mol. Biol. 70, 57.

Sambrook, J., Westphal, H., Srinivasan, P.R. and Dulbecco, R. (1968). Proc. Natl. Acad. Sci. USA. 60, 1288.

Saragosti, S., Moyne, G. and Yaniv, M. (1980). Cell 20, 65.

Scheller, R.H., Dickerson, R.E., Boyer, H.W., Riggs, A.D. and Itakura, K. (1977). Science 196, 177.

Scher, C.D. (1971). Virology 46, 956.

Scott, W.A. and Wigmore, D.J. (1978). Cell 15, 1511.

Seif, R. and Martin, R.G. (1979a). J. Virol. 31, 350.

Seif, R. and Martin, R.G. (1979b). J. Virol. 32, 979.

Setlow, V.P., Persico-Dilauro, M., Edwards, C.A.F. and Martin, R.G. (1980). Virology 101, 250.

Shalloway, D., Kleinberger, T. and Livingston, D.M. (1980). Cell 20, 411.

Shapiro, S.K., Almenas, A. and Thomson, J.F. (1965). J. Biol. Chem. 240, 2512.

Shenk, T. (1978). Cell 13, 791.

Shenk, T.E., Carbon, J. and Berg, P. (1976). J. Virol. 18, 664.

Shenk, T.E., Rhodes, C., Rigby, P.W.J. and Berg, P. (1975). Proc. Natl. Acad. Sci. USA. 72, 989.

Shiroki, K. and Shimojo, H. (1971). Virology 45, 163.

Shortle, D.R., Margolskee, R.F. and Nathans, D. (1979). Proc. Natl. Acad. Sci. USA. 76, 6128.

Shortle, D. and Nathans, D. (1979). J. Mol. Biol. 131, 801.

Siddell, S.G. and Smith, A.E. (1978). J. Virol. 27, 427.

Simmons, D.T. and Martin, M.A. (1978). Proc. Natl. Acad. Sci. USA. 75, 1131.

Sleigh, M.J., Topp, W.C., Hanich, R. and Sambrook, J.F. (1978). Cell 14, 79.

Smith, H.S., Gelb, L.D. and Martin, M.A. (1972). Proc. Natl. Acad. Sci. USA. 69, 152.

Smith, H.S., Scher, C.D. and Todaro, G.J. (1971). Virology 44, 359.

Spear, B.R. (1975). Chromosoma 48, 159.

Southern, E.M. (1975). J. Mol. Biol. 98, 503.

Steinberg, B.M. and Pollack, R. (1979). Virology 99, 302.

Steinberg, B., Pollack, R., Topp, W. and Botchan M. (1978). Cell 13, 19.

Sternberg, N., Tiemeier, D. and Enquist, L. (1977). Gene 1, 255.

Subramanian, K.N. (1979). Proc. Natl. Acad. Sci. USA. 76, 2556.

Subramanian, K.R. and Shenk, T.E. (1978). Nucleic Acid Res. 5, 3635.

Sutcliffe, J.G. (1978). Cold Spring Harbour Symp. Quant. Biol. 43, 81.

Sutcliffe, J.G. (1978). Nucleic Acid Research 5, 2721.

Syvanen, M. (1974). Proc. Natl. Acad. Sci. USA. 71, 2496.

Tegtmeyer, P. (1972). J. Virol. 10, 591.

Tegtmeyer, P. (1975). J. Virol. 15, 613.

Tegtmeyer, P., Schwartz, M., Collins, J.K. and Rundell, K. (1975). J. Virol. 16, 168.

Tenen, D.G., Martin, R.G., Anderson, J. and Livingston, D.M. (1977). J. Virol. 22, 210.

Thoma, M., White, R.L. and Davis, R.W. (1976). Proc. Natl. Acad. Sci. USA. 73, 2294.

Tilghman, S.M., Tiemeier, D.C., Polsky, F., Edgell, M.H., Seidman, J.G., Leder, A., Enquist, L.W., Norman, B. and Leder, P. (1977). Proc. Natl. Acad. Sci. USA. 74, 4406-4410.

Tjian, R. (1978). Cell 13, 165.

Tjian, R., Fey, G. and Graessmann, A. (1978). Proc. Natl. Acad. Sci. USA. 75, 1279.

Tjian, R. and Robbins, A. (1979). Proc. Natl. Acad. Sci. USA. 76, 610.

Tjian, R., Robbins, A. and Clark, R. (1979). Cold Spring Harbor Symp. Quant. Biol. 44, 103.

Todaro, G.J. and Green, H. (1964). Virology 23, 117.

Todaro, G.J. and Green, H. (1966a). Virology 28, 756.

Todaro, G.J. and Green, H. (1966b). Proc. Natl. Acad. Sci. USA. 55, 302.

Tonegawa, S., Brack, C., Hozumi, N. and Schuller, R. (1977). Proc. Natl. Acad. Sci. USA. 74, 3518.

Topp, W.C. and Rifkin, D.B. (1980). Virology 106, 282.

Tooze, J. (1980). DNA Tumor Viruses. Molecular Biology of Tumor Viruses, Second Edition, Part 2. Cold Spring Harbor Laboratory.

Twigg, A.J. and Sherratt, D. (1980). Nature 283, 216.

Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W.J. and Goodman, H.M. (1977). Science 196, 1313.

Varshavsky, A.J., Sundin, O. and Bohn, M. (1979). Cell 16, 453.

Villareal, L.R., White, R.T. and Berg, P. (1979). J. Virol. 29, 209.

Volkaert, G., Feunteun, J., Crawford, L.V., Berg, P. and Fiers, W. (1979). J. Virol. 30, 674.

Volckaert, G., van de Voorde, A. and Fiers, W. (1978). Proc. Natl. Acad. Sci. USA. 75, 2160.

Wahl, G.M., Stern, M. and Stark, G.R. (1979). Proc. Natl. Acad. Sci. USA. 76, 3683.

Wake, C.T. and Wilson, J.H. (1980). Cell 21, 141.

Walbeck, W., Fohring, G., Chowdhury, A., Gruss, P. and Sawer, G. (1978). Proc. Natl. Acad. Sci. USA. 75, 5969.

Wakamiya, T., McCutchan, T., Rosenberg, M. and Singer, M. (1979). J. Biol. Chem. 254, 3584.

Watkins, J.F. and Dulbecco, R. (1968). Proc. Natl. Acad. Sci. 58, 1396.

Wellauer, P.K., Dawid, I.B. and Tartof, K.D. (1978). Cell 14, 267.

Wensink, P.C., Finnegan, D.S., Donelson, J.E. and Hogness, D.D. (1974). Cell 3, 315.

Westphal, H. (1970). J. Mol. Biol. 50, 407.

Westphal, H. and Dulbecco, R. (1968). Proc. Natl. Acad. Sci. USA. 59, 1158.

White, R.L. and Hogness, D.S. (1977). Cell 10, 177.

Yamamoto, K.R., Alberts, B.M., Benzinger, R., Lawhorne, L. and Treiber, C. (1970). Virology 40, 734.

Zain, B.S. and Roberts, R.J. (1978). J. Mol. Biol. 120, 13.