NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

Canada

THE UNIVERSITY OF ALBERTA

Holliday Junctions and DNA Analogs: Metabolism and Structural Dynamics

by

Peter Dickie

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF Doctor of Philosophy

DEPARTMENT OF Biochemistry

EDMONTON, ALBERTA

FALL 1987

ISBN 0-315-40864-2

THE UNIVERSITY OF ALBERTA

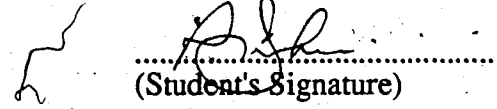
RELEASE FORM

NAME OF AUTHOR: Peter Dickie

TITLE OF THESIS: Holliday Junctions and DNA Analogs: Metabolism and Structural Dynamics

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1987



(Student's signature)

10721 - 80 Ave

(Student's permanent address)

Edmonton Alta

T6E 1V7

Date: Oct 21, 1987

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Holliday Junctions and DNA Analogs: Metabolism and Structural Dynamics", submitted by Peter Dickie in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

H Richard Morgan
.....
(Supervisor)

.....

.....

.....

.....

Date: Oct 8, 1987

ABSTRACT

The telomere sequences of the *Leporipoxvirus* Shope fibroma virus were propagated in bacteria in the form of a 322 base pair inverted repeat insert in the recombinant plasmid pSAlB.56A. The arrangement of the sequences was equivalent to the telomere-telomere concatameric junctions that appear during the replication of poxviruses. Plasmid pSAlB.56A was not viable in wild-type bacteria, but could be isolated intact from recombination-deficient *Escherichia coli*. *In vitro*, the telomere inverted repeat sequences existed in the cruciform configuration, as judged from the plasmids mobility in agarose gels and sensitivity to specific endonucleases, such as the Holliday-resolving enzyme bacteriophage T7 endonuclease I. The cruciform extruded facilely under standard experimental conditions but had a high energy of formation (ΔG_f) estimated to be 44 kcal/mol. The high energy of formation was attributed to the presence of extra-helical bases in the cruciform hairpins.

Bacteriophage T7 gene 3 product, endonuclease I, is a general debranching enzyme that cleaves at or near phosphodiester crossovers. Synthetic Holliday crossovers (fourth-rank junctions) and less complex branched junctions were created and digested with T7 endonuclease I in order to define the substrate specificity of the enzyme. T7 endonuclease I cleaved predominantly one nucleotide 5' of the phosphodiester crossover in all the synthetic "branched" molecules. A preference was shown for crossovers between two duplex branches and junctions that were immobile (did not permit branchpoint migration) but "open" (a gap or nick existed at the junction allowing for greater flexibility). A modest preference for nucleotide sequence was observed. The enzyme did not cleave short, non-complementery single-stranded DNA establishing its

specificity for duplex-containing molecules. In fact, single-stranded DNA interfered with the ability of T7 endonuclease I to resolve the cruciform junction in pSA1B.56A presumably because the enzyme can bind single-stranded DNA.

The digestion of native preparations of pSA1B.56A with T7 endonuclease I resolved the circular plasmid to a linear molecule with nicked hairpin ends. Cleavage positions were mapped to specific regions of telomere sequence that were periodically spaced, on average, ten nucleotides apart. A similar result was observed following the digestion of purified topoisomers suggesting that during branchpoint migration the crossover alternated between sensitive and insensitive conformations. It is argued that conformational isomerization is imposed by the topological implications of having two junctional branches linked in the form of a closed DNA loop. A theoretical consideration of branchpoint movement in plasmid molecules further supports this conclusion.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ΔG_f	free energy of formation
ΔG_s	free energy of supercoiling
Wr	DNA writhe (superturns)
Tw	DNA twist (turns of helix)
Lk	DNA linking number
ΔLk	linking difference of DNA (Lk - Lk °)
Lk°	linking number of DNA when relaxed
form I DNA	covalently closed circular DNA
form II DNA	nicked, open-circular DNA
form III DNA	linear plasmid DNA
r	relaxed
l	linear
scc	supercoiled with cruciform
RF	replicative form
UHA	unperturbed-helix-axis form of Holliday structure
S4	Sobell's pseudo-four-fold version of a Holliday structure
T2	tetrahedrally arranged Holliday structure
oligo	oligonucleotide

N	number of base pairs
SFV	Shope fibroma virus
v	volume
w	weight

I. INTRODUCTION

Eukaryotic genomes, in contrast to the chromosomes of bacteria, are predominantly composed of linear, double-stranded DNA molecules. Exceptions include most native plasmid DNAs and most mitochondrial DNAs, which, like bacterial chromosomes, are double-stranded covalently closed circular DNA molecules. The molecular ends of linear chromosomes are specialized structures called telomeres. A general consideration of the metabolism of DNA molecules has inspired the conclusion of researchers that telomeres evolved to fulfill a special function related to the maintenance of linear chromosomes. Reinforcing their conviction has been the observation that telomere structure has been remarkably well conserved amongst the evolutionary diverse lower eukaryotic organisms that have been studied to date. The origin of this thesis was a discussion of the particular role that telomeres play in the replication of chromosomal ends.

A. Telomere Structure and Function

Telomere function has been inferred from the biology of chromosomal DNA ends (reviewed by Blackburn, 1984; Blackburn and Szostak, 1984). The naked ends of fragmented DNA molecules (ie. following double strand breaks caused by mechanical damage or X-irradiation) are highly recombinogenic (Szostak *et al*, 1983), are prone to nucleolytic degradation and may fuse with other chromosomal ends (McClintock, 1941). By virtue of their unique structure, telomeres act as chromosomal "caps" and protect genomic DNA ends from these deleterious reactions. During meiosis and to a lesser extent during interphase and mitotic prophase, cytological investigations have shown,

telomeres transiently self-associate and possibly mediate the organization of chromosomes into identifiable arrangements (Lipps *et al.*, 1982). More generally throughout the cell life-cycle, telomeres are known to interact with the nuclear envelope (Agard and Sedat, 1983; Mathog *et al.*, 1984). These observations have led to the proposal that telomeres are functional in the nuclear organization of chromosomes, and the segregation of chromosomes during cell division. Telomere complexes (protein-DNA structures) may also act as a topological anchor for DNA molecules *in vivo* and direct, in part at least, the phasing of nucleosomes along the chromosome (Gottschling and Cech, 1984).

A significant role is attributed to telomeres during the replication of linear chromosomes. The capacity to maintain the length of a chromosome during multiple rounds of replication, in spite of the fact that all known DNA polymerases require a 3'-hydroxyl primer from which to initiate DNA synthesis (Cavalier-Smith, 1974; Bateman, 1975), is an aquired property of the telomere structure. Usually, a polynucleotide primer serves to initiate DNA synthesis and, if it is RNA, it is removed following the priming reaction exposing an unreplicated region at the 5' end of the DNA. In the absence of some adaptive mechanism, daughter molecules would become progressively shorter with the accumulation of these 5' terminal gaps (Watson, 1972).

Function is reflected in structure and while certain features of telomere structure are peculiar to individual organisms, there are several outstanding features that are common to all nuclear telomeres. The most intensely studied linear DNA chromosomes are those associated with single-cell eukaryotes. The somatic macronuclei of ciliated protozoa are a convenient source of telomeres for molecular analysis because they contain numerous short linear DNAs, such as ribosomal RNA gene molecules and subchromosomal

molecules derived from the development-related fragmentation of micronuclear chromosomes. The telomeres of these nuclear DNAs consist of tandem repeats of the sequence motif 5'[$C_{1-8}(A/T)_{1-4}$]3'. For example, *Oxytricha* has the repeat (C_4A_4) (Klobutcher *et al.*, 1981; Pluta *et al.*, 1982), *Tetrahymena* the repeat (C_4A_2) (Blackburn and Gall, 1978) and yeast the repeat $C_{2-3}A(CA)_{1-3}$ (Shampay *et al.*, 1984). The number of tandem repeats varies depending on the organism (there are up to 300-500 copies in *Tetrahymena*) and can vary according to the cell cycle stage. Telomeres, therefore, are dynamic structures and increases in length of telomere sequences in propogating *Tetrahymena* have been directly attributed to changes in the number of these tandemly repeated sequences (Greider and Blackburn, 1985; Larson *et al.*, 1987). The G-rich strand forms a 3' overhang at each chromosomal end (3'$(G_4T_4)_2$ in *Oxtricha* and *Stylonychia* (Klobutcher *et al.*, 1981 and Steinhilber and Lipps, 1986)) and in the distal regions of the telomere sequences there are specific arrays of non-ligatable single-strand breaks (Blackburn and Gall, 1978; Johnston, 1980; Katzen *et al.*, 1981; Szostak and Blackburn, 1982; Blackburn and Challoner, 1984). These gaps remain distally located during the lengthening of telomere sequences in vegetative cells and thus are features of the extreme ends of the chromsome. There are a few examples of telomeres studied in higher eukaryotes (including humans) but structures are not yet as well defined (Cooke *et al.*, 1985)

The nucleotides bordering the internal gaps in telomere sequences can be radiolabelled whereas the terminal nucleotides are effectively "blocked" by the telomere structure (Blackburn and Gall, 1978; Emery and Weiner, 1981). Protection of the ends of nuclear

DNAs may be afforded by telomere-specific proteins and some telomere-binding proteins have been identified in nuclear extracts of *Oxytricha*, *Stylonychia* and yeast. These proteins bind to permutations of the poly($C_{1-4}A_{1-4}$) motif and G-rich tail (Gottschling and Zakian, 1986; Steinhilber and Lipps, 1986; Berman *et al.*, 1986) but no specific function has been attributed to them. Blackburn (1984, 1985) has proposed that, in the absence of bound protein, the DNA ends may be "blocked" because telomere DNA adopts a novel secondary conformation. Non-B DNA structures such as Z DNA and, in particular, multi-stranded poly(G) forms (Dugaiczyk *et al.*, 1980) have been proposed. In fact, it has been suggested that the interchromosomal association of poly(G) strands may account for the end-to-end attachment of chromosomes that has been observed (Gottschling and Cech, 1984). Novel fold-back structures involving unusual base-pairing schemes have also been proposed (Blackburn, 1985). Some chromosome ends are "blocked" because the complementary DNA strands are covalently linked in the form of hairpin termini. These have been identified for eukaryotic viruses such as parvoviruses (Bourguignon *et al.*, 1976; Berns *et al.*, 1985), iridoviruses (Gonzalez *et al.*, 1986) and poxviruses (Geshelin and Berns, 1974; Baroudy *et al.*, 1982; DeLange *et al.*, 1986), also in linear yeast plasmids (Kikuchi, *et al.*, 1985), the mitochondrial DNA of *Paramecium* (Pritchard and Cummings, 1981) and in linear plastids of barley (Ellis and Day, 1986). Hairpin termini probably represent a unique class of telomere since those that have been identified do not possess the short tandemly repeated sequences that researchers now believe are responsible for endowing the telomeres of the nuclear chromosomes with functional specificity.

The telomere sequences from a variety of organisms have been cloned (Cavalier-Smith,

1983). Szostak and Blackburn (1982) created a cloning vehicle using the linear derivative of a circular yeast plasmid. They added *Tetrahymena* telomeres to the linear plasmid and discovered that it could be propagated in yeast. This was further evidence that the structure and function of telomeres had been conserved even though during the propagation of similar recombinant plasmids there was an elaboration of host-specific repeats onto the cloned telomere sequences (Shampay *et al.*, 1984). The complexity of telomere structure was emphasized in these experiments when it was shown that an artificial hairpin terminus was unable to stabilize the linear vector *in vivo* (Szostak and Blackburn, 1982). The cloning of telomeres represented a significant advance in the study of telomere structure, the binding of telomere-specific proteins and the replication of telomere DNA.

B. Telomere Replication

A variety of mechanisms exists to overcome the problem of replicating the ends of linear chromosomes. Circular DNA molecules can be continuously replicated to produce multiple genome-length daughter molecules because the DNA templates are uninterrupted polynucleotide chains. The linear DNA of bacteriophage lambda transiently circularizes in infected cells and replicates in this form. The linear genomes of T4 and T7 bacteriophages concatamerize (anneal end-to-end) during replication so that the priming of DNA synthesis is effectively accomplished from neighboring genomes (Broker, 1973). Bacteriophage Ø29 and the mammalian adenoviruses have covalently linked terminal proteins which serve as primers for the DNA polymerase (Harding and Ito, 1980; Rekosh *et al.*, 1977). Another example of a facilitative adaptation might be the presence of hairpin termini since they allow the replicative machinery to circumscribe the uninterrupted,

self-complementary polynucleotide chain of a linear DNA molecule.

Drawing upon these examples of prokaryote and virus replication, general models for the replication of eukaryotic telomeres have been postulated (reviewed in Blackburn and Szostak, 1984). Cavalier-Smith (1974) suggested that the telomere sequences were palindromic (had two-fold symmetry), allowing the unreplicated 3' sequences at the end to fold back into a hairpin structure. The resulting nick would be ligated and the opposite strand nicked to allow the DNA to unfold and replication to be completed off the newly exposed 3'-OH. Bateman (1975) proposed that telomeres naturally possessed hairpin termini and after replication proceeded around the hairpin the "fused" telomeres would be resolved to daughter termini. The representation of the Bateman model illustrated in Figure I.1 demonstrates that the replicative intermediate is, in fact, a large palindrome. The separation of the daughter telomeres would follow (according to the model) the placement of staggered nicks, centered about the two-fold symmetry axis, by a sequence-specific endonuclease. Strand separation, the folding back of the self-complementary overhanging sequences and the sealing of the nicks by ligase would regenerate the native molecular termini. The Bateman model is relevent to the replication of poxviruses and those linear genomes with hairpin termini (Merchlinsky and Moss, 1986). On the other hand, the Cavalier-Smith model is applicable to the replication of the parvoviruses whose linear single-stranded DNA genome has short terminal palindromic sequences which fold-back to form hairpin structures (reviewed in Challberg and Kelly, 1982). Influenced by the cytological evidence demonstrating telomere-telomere interactions, Dancis and Holmquist, (1979) proposed that recombination between telomeres led to the transient fusion of telomeres into the inverted repeat configuration (Figure I.1), a form that also allows the complete replication of the telomere sequences.

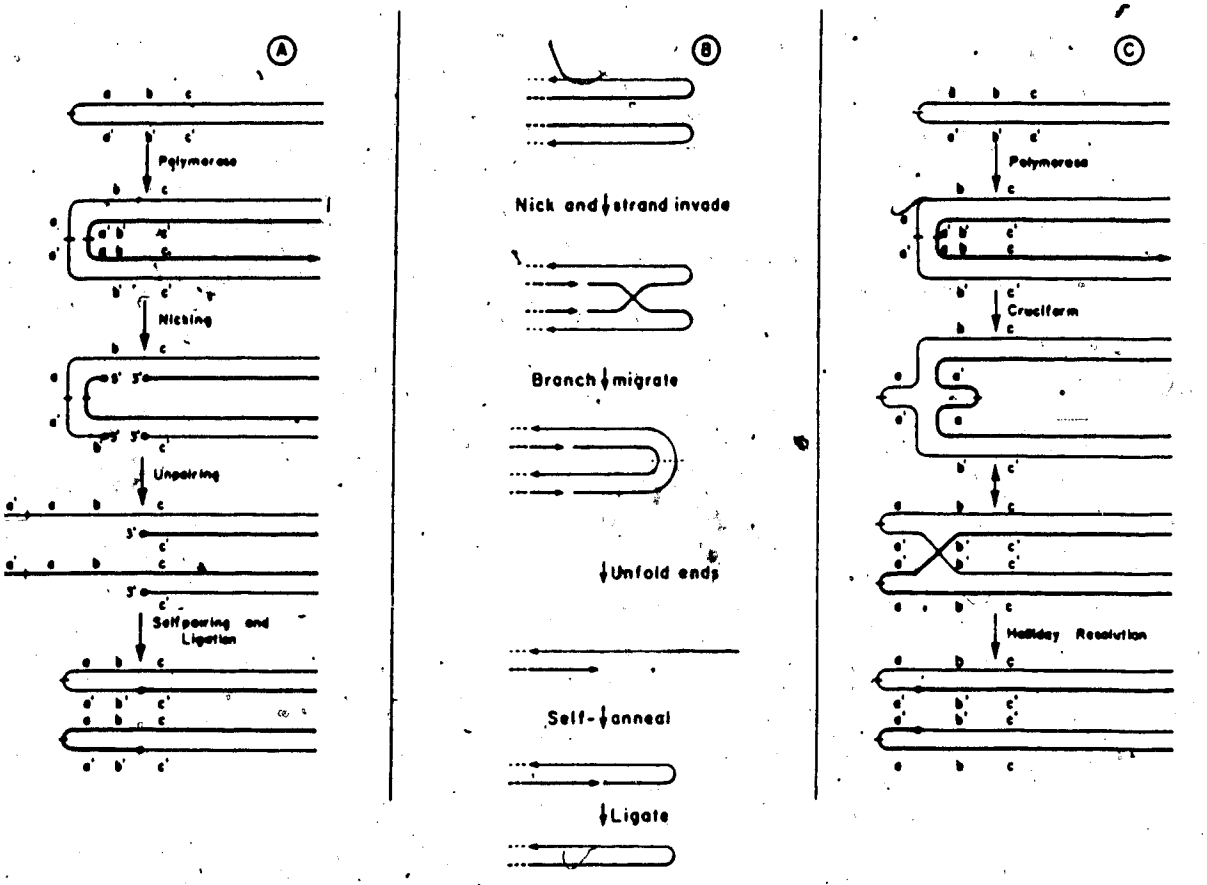


Figure I.1. Models of telomere replication. Panel A: the Bateman model; panel B: the Dancis and Holmquist model; panel C: the cruciform extrusion model. In the three models, replication of hairpin termini (A and C) or recombination between hairpin termini (B) generates an inverted repeat intermediate which is resolved to daughter telomeres by endonucleolytic activity. Complementary sequences are noted as a/a*, etc. (adapted from McFadden and Dales, 1982 and Szostak and Blackburn, 1984).

The resolution of the telomeric arrangement could be accomplished in a manner similar to that described in the Bateman model.

The Bateman model is modified in a more recent model described independently by McFadden and Morgan (1982) and Szostak (1983) such that transiently formed or permanent hairpins are replicated to generate the inverted repeat telomere configuration, but a novel mechanism for the resolution of the telomere sequences is invoked. These authors propose that a cruciform structure is extruded from the telomere sequences and the cruciform crossover is resolved by a debranching, or a Holliday-resolving, endonuclease (Figure I.1). As a consequence of this process, the stem and loop structures of the cruciform become the terminal hairpins of the resolved chromosome. With regard to these models, there is substantial evidence for the appearance of the proposed inverted repeat arrangement of telomere sequences (as depicted in Figure I.1) during the replication of poxviruses (Moyer and Graves, 1981; Moss *et al.*, 1983; DeLange *et al.*, 1984; DeLange *et al.*, 1986), *Paramecium* mitochondrial DNA (Pritchard and Cummings, 1981), yeast DNA (Szostak, 1983) and *Tetrahymena* extrachromosomal rDNA elements (Blackburn and Gall, 1978).

Based on the observation that recombinant linear DNA molecules bearing *Tetrahymena* telomeric sequences acquire yeast telomeric sequences when they are propagated in yeast cells (Shampay *et al.*, 1984), it was proposed that activities exist that attach host-specific telomere sequences to compatible telomeric ends. An activity in *Tetrahymena* cell-free extracts has been identified that adds tandem TTGGGG sequences onto telomere-specific oligonucleotide primers (Greider and Blackburn, 1985) requiring only the primer, dTTP and dGTP. The activity increased during macronuclear development, coinciding with the

demand for telomere synthesis following fragmentation of the micronuclear chromosomes and specified the addition of host repeats to yeast and *Tetrahymena*-related telomere oligomers but not to complementary or non-telomeric oligomer sequences. As part of the replicative process, a telomere terminal transferase could function *in vivo* to elongate chromosomes. As yet, there is no conclusive evidence of terminal hairpin formation in nuclear chromosomes and it remains an open question whether telomere-binding proteins prime DNA synthesis (Cavalier-Smith, 1983).

C. Poxvirus Replication

Insofar as poxvirus replication proceeds through an intermediate in which the telomere sequences are transiently arranged in an inverted repeat configuration, it represents a useful model system to study the replication and resolution of telomeres. As a system its appeal is enhanced by the fact that the viruses replicate automonously in cytoplasmic foci, or "factory areas" (Dales and Pogo, 1982; Holowczak, 1982; McFadden and Dales, 1982; Wittek, 1982; Moss, 1986) and, by implication, under the exclusive control of viral-encoded gene products.

Vertebrate poxviruses have been classified into at least six genera. The prototypic poxvirus, and the first for which the telomeric structure was identified (Baroudy *et al.*, 1983; Moss *et al.*, 1983), is vaccinia virus, a member of the *Orthopoxvirus* genus. Most of what has been written about the replication of poxviruses has been obtained from studies on vaccinia virus. Its genome is approximately 180 kilobase pairs of linear, double-stranded DNA which is covalently linked to form hairpin termini. Hence the molecule is a single continuous self-complementary polynucleotide chain. The sequences

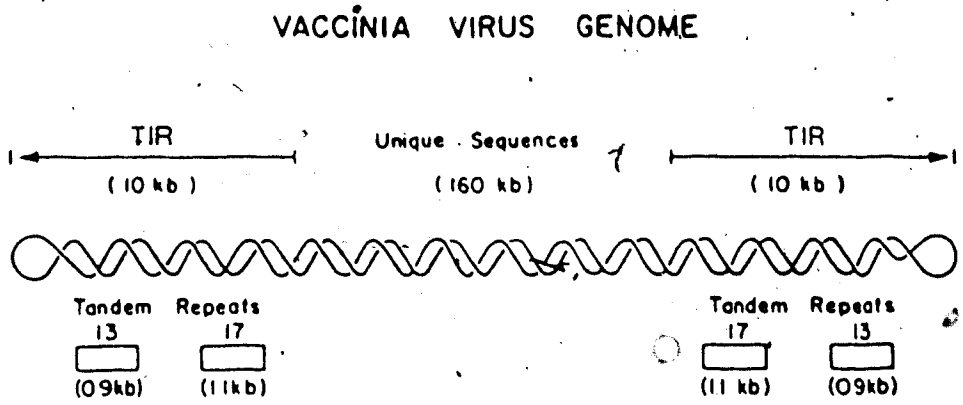


Figure I.2. The structure of the vaccinia virus genome. Characteristic features are the large inverted terminal repeats, the hairpin termini and the tandem repeats. The genome is a continuous polydeoxyribonucleotide (adapted from McFadden and Dales, 1982). TIR: terminal inverted repeat.

at the ends of the genome take the form of large inverted terminal repeats (approximately 10 kilobase pairs long) containing 30 tandem reiterations of a 70 base pair sequence (Figure I.2). The extreme terminal regions are very A+T-rich and possess numerous unpaired bases which are presumably extra-helical (Evans and Morgan, 1982). On account of the unpaired bases (derived from asymmetric base pairs in the inverted repeat configuration), the telomeres exist in two isomeric configurations, "flip" and "flop", that are inverted and complementary (reviewed by Moss, 1986). In relation to the illustrations in Figure I.1, the two daughter hairpins would be representative of these two alternative forms.

The replicative cycle of vaccinia virus begins with the adsorption to and penetration of the host cell, release of the viral core and commencement of early transcription. DNA polymerase is produced in large amounts, along with other enzymes involved in the synthesis and metabolism of DNA. DNA replication occurs 1.5 to 6 hours post-infection and the evidence indicates that replication can be initiated at both ends of the DNA molecule. Little detail is known of the overall replicative mechanism and the form of replicative intermediates except that concatameric arrays bearing the head-to-head and tail-to-tail junctions have been identified (Moss, 1986).

An *in vitro* system has not yet been developed to study poxvirus replication and telomere resolution, and only limited DNA replication has been observed in crude cytoplasmic extracts (Esteban and Holowczak, 1977). A genetic approach to studying poxvirus replication has only just commenced with the acquisition of temperature-sensitive mutants (reviewed by Moss, 1986). The viral DNA polymerase has been purified and characterized *in vitro*, as has a DNA topoisomerase. Another enzyme that has been purified and that may play a very specific role in the replication of vaccinia is a capsid-enclosed deoxyribonuclease. It has the provocative ability to linearize supercoiled plasmid DNA and covalently link the DNA ends (Lakritz *et al.*, 1985). It accomplishes this with no obvious specificity other than nicking at sequences that, when supercoiled, have the potential to form alternative secondary structures featuring single-stranded elements. Should this endonuclease participate in the initiation of replication or telomere resolution then its specificity *in vivo* is most likely to be determined by interactions with accessory proteins.

Vaccinia telomere replication may be explained by either the Bateman model or the

cruciform extrusion model described in the preceding section. In addition, Baroudy *et al.* (1982) have described a self-priming scheme for the replication of poxviruses. It is similar in respects to models of the replication of the parvoviruses that have linear single-stranded genomes with short hairpin termini. In their model, an endonuclease nicks near the terminal hairpin and the hairpin DNA unfolds. The unfolded strand is used as a template for the replication of the complementary strand thereby generating, through *de novo* synthesis, a terminal palindrome of the form Cavalier-Smith (1974) had proposed was generally present at genomic termini. Contrary to the Cavalier-Smith model however, the self-priming model suggests that the terminal palindromic sequences undergo strand exchange to form two inverted and complementary hairpin termini. Replication could then be primed from either end of the genome from the 3'-hydroxyl of one of the fold-back strands.

A significant requirement of resolution that must be accounted for by any model of poxvirus replication is that the "flip" and "flop" forms of the telomeres are generated. The formation of terminal fold-back structures (as in the self-priming model, the cruciform extrusion model or the Bateman model) accomplishes this so long as the hairpin structures consume sufficient sequence to result in the appearance of every potential extrahelical base. Assuming that an endonuclease resolves the fused-telomere intermediate, symmetrically placed staggered nicks must be made outside the sequences bearing all the asymmetrical base pairs. The furthest an extrahelical base is from the genomic terminus in vaccinia virus is 55 nucleotides (Figure I.3). Therefore, the sequence separating the resolving nicks would minimally be 110 base pairs long and strand exchange, mediated by cruciformation, recombination or perhaps catalyzed by some energy transducing

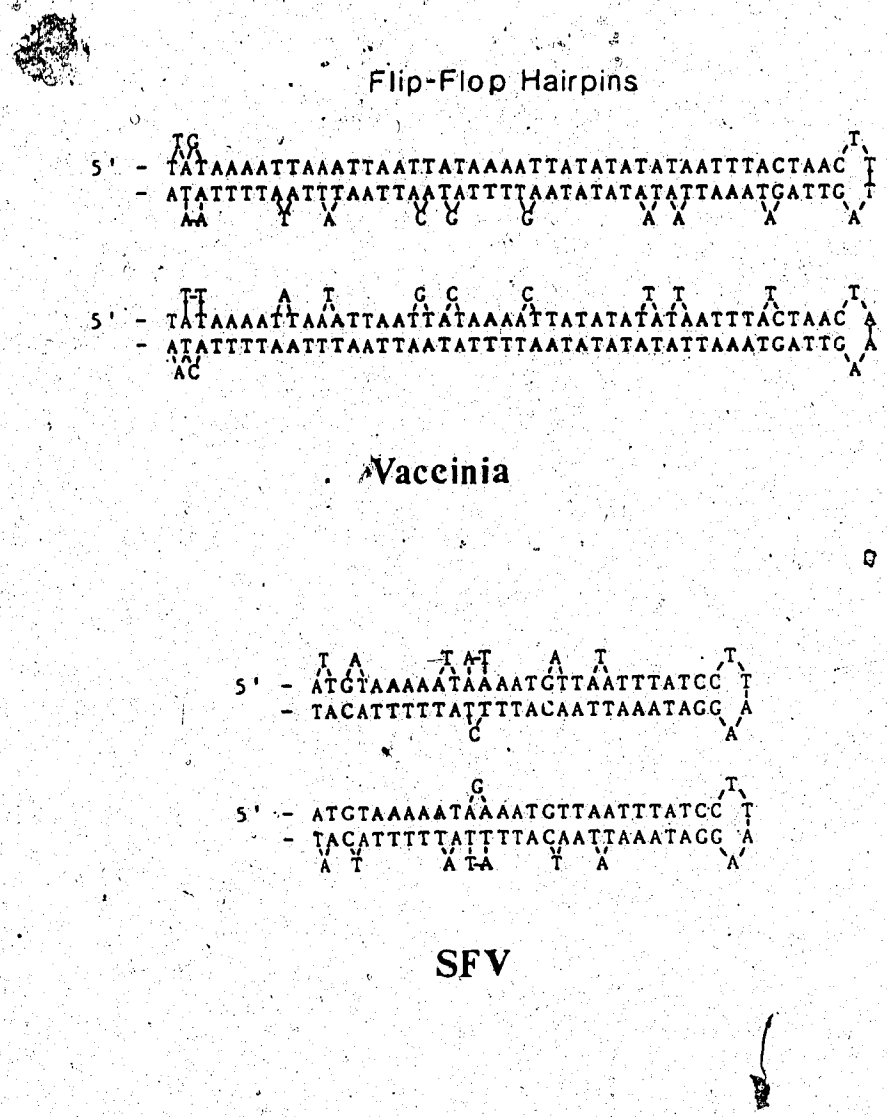


Figure I.3. The hairpin termini of the poxviruses, vaccinia virus and Shope fibroma virus, illustrating the position and number of extra-helical bases.

system which might include a helicase for instance (Warren and Green, 1985), would lead to the formation of telomeres bearing the characteristic extrahelical bases.

Considerable new information regarding the extent and nature of the telomeric sequences

that are critical for viral replication and resolution has, and will continue to be, accrued from studies employing recently developed *in vivo* co-transfection assays. The protocol for these experiments involves the transfection of cloned inverted repeat arrangements of poxvirus telomeric sequences into poxvirus infected cells in order to assess their capacity to be resolved. Vaccinia telomeric sequences were originally cloned according to the methodology of Szostak and Blackburn (1982) (DeLange *et al.*, 1984; Merchlinsky and Moss, 1986). The cloned telomere sequences were propagated in yeast in the inverted repeat arrangement presumably because yeast cellular factors failed to recognize and resolve the vaccinia telomere sequences. The inverted repeat telomere sequences were subcloned into bacterial plasmids for use in the co-transfection assays. The resolution of the cloned vaccinia virus concatameric junctions in vaccinia-infected cells (evinced by the reduction of concatamerie arrays to plasmid monomer size) substantiated the assumption that the inverted repeat configuration is a *bona fide* replication intermediate (Merchlinsky and Moss, 1986). Exogenous plasmid bearing the cloned inverted repeat telomere sequences is replicated in uninfected cells but is not resolved, indicating that the resolving enzymes are viral gene products (DeLange *et al.*, 1986). Furthermore, a telomeric inverted repeat of only 252 base pairs was efficiently resolved, implying that all the sequences critical for resolution lay within 126 base pairs of the molecular terminus.

An *in vivo* co-transfection assay (DeLange *et al.*, 1986; DeLange and McFadden, 1986 and 1987) has been established for the *Leporipoxvirus*, Shope fibroma virus (SFV) a benign fibroma-inducing virus of rabbits (Shope, 1932; reviewed by McFadden, 1987). Like vaccinia virus, SFV telomeres possess numerous extrahelical bases and the sequences immediately downstream of the termini are extremely A+T-rich (Wills *et al.*, 1983; DeLange *et al.*, 1984 and 1986) consistent with these being conserved features of

poxvirus telomeres. They, too, exist in the alternative "flip" and "flop" configurations. The SFV inverted terminal repeat is 12.4 kilobase pairs in length and shares little or no homology with the vaccinia virus inverted terminal repeat outside of a few short sequences proximal to the terminal hairpin. The identity and positions of the extra-helical bases in SFV are shown in Figure I.3. In Figure I.4, the contrasting schemes for the resolution of the replicative intermediates are illustrated, emphasizing the generation of the "flip" and "flop" termini and indicating the relative positions of the conserved sequences. SFV telomere sequences were also cloned in the inverted repeat arrangement (DeLange *et al.*, 1986).

Specific permutations of the SFV palindromic telomere junction were generated using standard molecular biology techniques and then these modifed telomere sequences were tested in the SFV co-transfection assay. In essence, nested deletions were created at the central axis and at either end of the cloned telomere sequences and the effect of these deletions on resolution was scored. The result was an identification of a core target DNA sequence that was required for telomere resolution (DeLange and McFadden, 1987). Interestingly, the central axis A+T-rich sequences containing the extrahelical bases, could be deleted without impeding resolution and thus these conserved features of poxviral telomere structure may be critical for some other viral function. The core target sequence is 76 base pairs long and displaced 45 base pairs from the central axis, beyond the last asymmetric base pair in the SFV inverted repeat (see Figure I.3). It contains the three short regions of homology shared by SFV and vaccinia virus telomeres, so it was not surprising that the SFV concatameric junctions were resolved in vaccinia-infected cells (DeLange *et al.*, 1986).

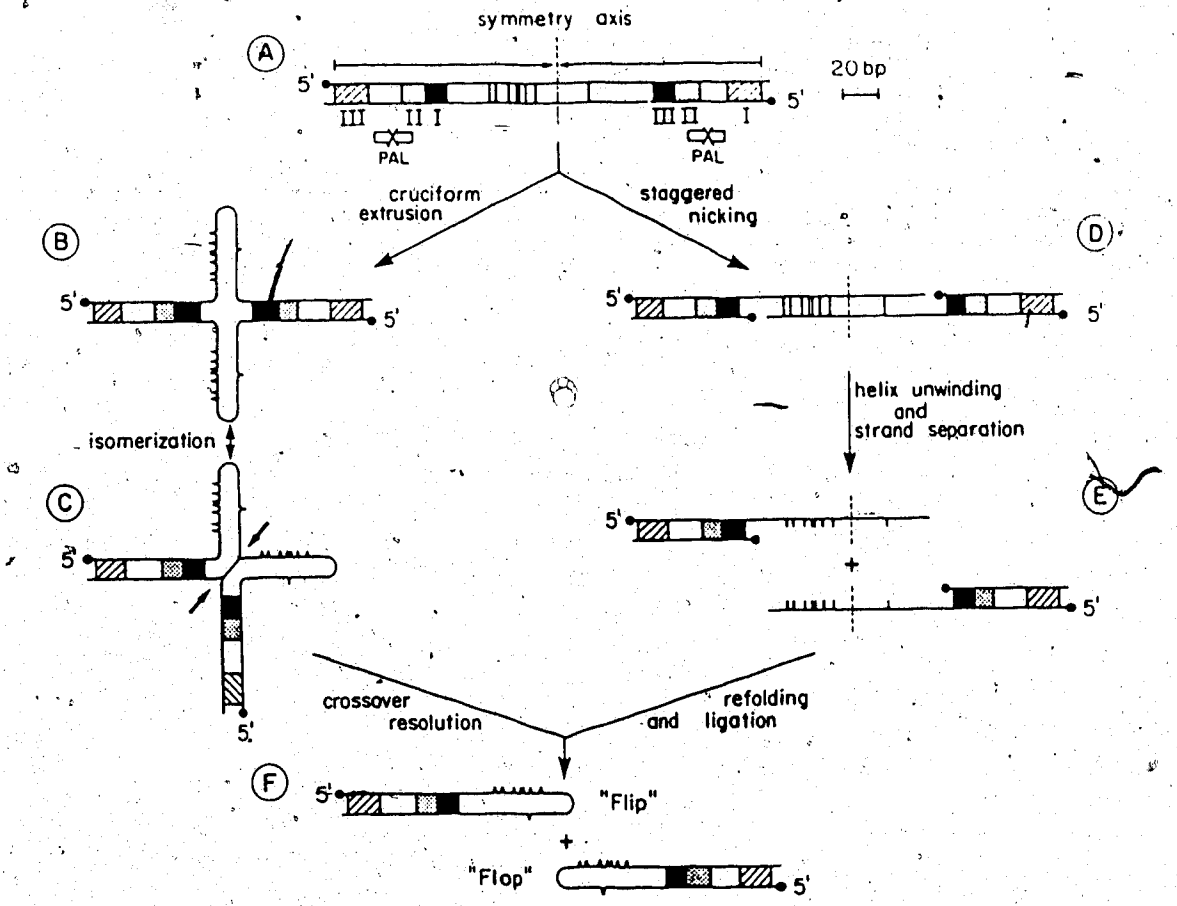


Figure I.4. Generation of "flip" and "flop" viral termini. The two endonucleolytic pathways for the resolution of the SFV telomere inverted repeat arrangement are shown; cruciformation on the left and site-specific cleavage (as per the Bateman model) on the right. Features of the SFV telomeres are emphasized, such as the extra-helical bases shown as spikes on the hairpin structures of B, C and F (vertical lines denote these same nucleotides as members of asymmetric base pairs in the lineform illustrations of A, D and E); the conserved sequences as box regions I, II and III; and a small internal inverted repeat (PAL) which lies in the core target region required for *in vivo* resoluiton.

D. The Biology of Inverted Repeat DNA Sequences

The cloned poxviral telomere sequences have been successfully propogated in *E. coli* in spite of the fact that large inverted repeat sequences are notoriously difficult to clone in bacteria. Inverted repeat sequences, as a rule, have a peculiar biology in prokaryotic systems. They induce changes in cellular DNA metabolism, such that they can affect rates of recombination and the efficacy of DNA replication, and they are themselves prone to metabolism. The implications of inverted repeat sequences in eukaryotic systems will be addressed in later sections.

Inverted repeat sequences are said to be perfect if they are completely complementary and they are palindromic if the repeat sequences are directly adjacent and not interrupted by non-repeated sequences (Mizuuchi *et al.*, 1982b). The cloned poxvirus telomere sequences are examples of imperfect inverted repeats since asymmetric base pairs are interspersed throughout the central axis region. Small inverted repeats are commonly found in prokaryotic systems and they often appear at specific recognition sites for DNA binding proteins. They have been identified at origins of DNA replication in prokaryotes, mammalian viruses and eukaryotic cells (Hobom *et al.*, 1979; Frisque, 1983; Zannis-Hadjopoulos *et al.*, 1984; Weller *et al.*, 1985), operator sequences (Gilbert *et al.*, 1975; Maniatis *et al.*, 1975) and transcription termination regions (Rosenberg and Court, 1979). Usually these are short non-palindromic repeats. Inverted repeats at sites of protein interaction may only reflect the presence of two-fold symmetry in the proteins. Nonetheless, inverted repeat sequences are polymorphic and can take other recognizable forms. For instance, *in vitro*, inverted repeat sequences can be induced to convert to the cruciform configuration when under negative superhelical strain (Mizuuchi *et al.*, 1982), a

property exploited in the cruciform extrusion model for telomere replication. This structural transition compelled some investigators to suggest that the proteins interacting with inverted repeat sequences do so by recognition of the alternative cruciform structure (Gierer, 1966).

No long (exceeding 120 base pairs) perfect inverted repeat has been identified or stably cloned in wild-type bacteria (Mizuuchi *et al.*, 1982). Attempts have been made to clone much larger inverted repeats but, invariably, these attempts have failed. Even a 68 base pair palindrome was unstable in wild-type bacteria (Courey and Wang, 1983). Either large inverted repeats prove to be lethal to the carrier replicon or they undergo RecA-independent deletion to smaller, more stable derivatives (Collins, 1981; Lilley, 1981; Collins *et al.*, 1982; Hagan and Warren, 1982; Mizuuchi *et al.*, 1982; Hagan and Warren, 1983; Leach and Stahl, 1983). A considerable enhancement of the stability of cloned inverted repeats is conferred by propagation in *recA recBC sbcB*$^-$ bacteria (Collins *et al.*, 1982; Leach and Stahl, 1983; Boisy and Astell, 1985; Leach and Lindsey, 1986) suggesting that recombination mechanisms (specifically the *sbcB* and the *recBC* encoded nucleases) contribute to the modification or loss of these sequences. Warren and Green (1985) successfully cloned palindromes of 114 and 147 base pairs in similar recombination-deficient bacteria. Both palindromic inserts inhibited spontaneous multimerization of plasmid DNA. In addition, the presence of the larger palindrome reduced plasmid copy number while the shorter palindrome did not. These observations exemplify the effect that inverted repeat sequences have on recombination and replication.

Two hypotheses have been proposed to explain the instability of palindromic sequences. Deletions could arise during replication if hairpin structures were to form on the lagging

strand and the replicative machinery occasionally skipped over them (Collins, 1981; Courey and Wang, 1983). It appears more likely however, judging from the observations just cited and the specific nature of the deletions, that instability is a consequence of a recombination event. All reported deletions that occur within cloned inverted repeat sequences are flanked by short directly repeated sequences and result in the loss of symmetrically disposed central axis sequences (Collins *et al.*, 1982; Courey and Wang, 1983; Hagan and Warren, 1983; Boisy and Astell, 1985). A complete loss of the inverted repeat would imply that the direct repeats were outside the sequence of the cloned inverted repeat. To generate partial deletions, small internal inverted repeats would have to exist within the arms of the larger unstable inverted repeat. Identical arms of the internal inverted repeats would constitute direct repeats and recombination between these would generate a shorter derivative of the instable inverted repeat. Recombination frequencies are very high in these situations and suggest further that the introduction of an inverted repeat sequence between two directly repeated sequences enhances the rate of illegitimate recombination. It is perhaps related to this observation that inverted repeat sequences stimulate intra- and interplasmidic recombination *in vivo* (Warren and Green, 1985). Warren and Green (1985) make a another point worth repeating. Two palindromes, one 114 base pairs and the other 146 base pairs, shared the same central axis sequences so that, apart from size, they were identical. Nonetheless, the 146 base pair palindrome reduced the copy number of the inverted repeat-bearing plasmid while the 114 base pair palindrome did not. Cruciformation or hairpin formation ("central" effects) in the inverted repeat sequences were presumably as likely in one as the other. Therefore, instability was proposed to be mediated by an "arm" effect; recombination between the inverted repeats, for instance, that would be less likely to occur in the shorter 114 base pair palindrome. This would account for a differential effect based on palindrome size.

The inviability of replicons bearing large palindromic sequences has been accounted for by a variety of proposals. Inviablity may arise because DNA replication is obstructed, by the physical presence of fold-back structures, for instance, or unresolved intermediates of recombination (Bolivar *et al.*, 1977; Leach and Lindsey, 1986). The loss of supercoiling accompanying the extrusion of inverted repeat sequences into a cruciform may be lethal since the free energy of supercoiling is known to support aspects of DNA metabolism (Lilley, 1981; Leach and Stahl, 1983). Were the inverted repeats to be extruded in the form of a cruciform the DNA junction formed at the crossover would be structurally equivalent to a Holliday crossover and may be cleaved by Holliday-resolving enzymes, reducing the DNA to a non-replicative form (Mizuuchi *et al.*, 1982). The likelihood that cruciformation is a determinant of plasmid viability has been questioned by results that showed that the presence of up to 50 asymmetric base pairs at the central axis failed to rescue an inviable plasmid molecule (Warren and Green, 1985). The spontaneous formation of a cruciform from these sequences would be prohibited energetically, although it is feasible that cruciformation could be catalyzed and the cruciform structure stabilized *in vivo*. Consequently, processes that might catalyze cruciformation have been proposed. For example, the action of *recBC* protein has been suggested (Leach and Lindsey, 1986), as has the processes of DNA replication, transcription and general recombination (Warren and Green, 1985). The *recBC* protein is known to unwind duplex DNA and generate large single-stranded loops that could fold back on themselves to form stem-and-loop structures (Taylor and Smith, 1980). Symmetry at the inverted repeat axis would not be as critical in such an extrusion mechanism so long as the stem-and-loop structures were stabilized. Recent reports indicate that inviability is a property of "active", or replicating, DNA substrates (Leach and Lindsey, 1986;

Shurvington *et al.*, 1987). Large palindromes can be maintained in bacteria if the DNA is non-replicative. Entry of the *recBC* protein usually requires a double strand break in the DNA and possibly "active" DNA provides these entry points as well as facilitating the formation of hairpin structures. In overview, it is clear that large inverted repeats are lethal to replicons and the deletion of central axis sequences, probably by illegitimate recombination, can rescue the replicon (Hagan and Warren, 1983).

E. DNA Supercoiling and Cruciform Extrusion: *In Vivo* Implications

As indicated, the biological properties of inverted repeat sequences may be associated with the potential of these sequences to form cruciform structures. Cruciform formation, or cruciformation, involves the disruption and exchange of interstrand base pairing for intrastrand base pairing such that two DNA stem-and-loop segments form adjoining DNA with normal interstrand base pairing (Gieger, 1966; Mizuuchi *et al.*, 1982b). The cruciform structure represents a higher free energy form of inverted repeat sequences by virtue of the presence of unpaired bases in the loop regions and the branched DNA structure formed at the base of cruciform. A minimum of four unpaired bases is presumed to exist in the hairpin loops (Lilley, 1986a). The stabilization of cruciforms therefore requires an input of energy. In plasmid DNA, the existence of cruciform structures *in vitro* has been verified by electron microscopy and use of enzyme probes; and they can be stabilized by the free energy of negative supercoiling (Mizuuchi *et al.*, 1982b).

Negative supercoiling (as a component of DNA writhe, Wr) can exist when there is a deficit of interwinding (or linking number, Lk) of the complementary strands in the DNA

molecule relative to the number of helical turns of DNA (the twist, Tw) in its native form. These three parameters are related by the equation (White, 1969):

$$Lk = Tw + Wr.$$

In other words, a negative Wr (implying negative supercoiling) could arise when the value of Tw exceeds Lk. Negative supercoiling has the tendency to destabilize B-form DNA (ie. reduce Tw to compensate for the linking deficiency) and this promotes cruciformation which removes helical turns from the topologically constrained circular core of plasmid DNA. The value of Lk is unchanged since the DNA strands in the cruciform hairpins are interwound about themselves and not the complementary strand. The free energy of supercoiling, expressed as torsional strain in the circular core of the DNA, decreases during extrusion and the energy difference contributes to the stabilization of the cruciform structure. In Figure I.5, the relationship between cruciformation and DNA topology is emphasized. In instances where the inverted repeat is large, it is possible that extrusion completely relaxes the DNA substrate (as shown in figure I.5). Smaller inverted repeats may only bring about partial losses of supercoiling. Experimentally, changes in DNA topology are detected by shifts in the electrophoretic mobility of the DNA molecule. In general, supercoiled molecules migrate faster than relaxed or partially relaxed derivatives of the same because they are more compact and encounter less resistance in the gel matrix.

Supercoiling *in vivo* is modulated by the catalytic activity of DNA topoisomerases, enzymes that change the linking number of DNA in increments of one or two (reviewed by Gellert, 1981; Drlica, 1984; Wang, 1985). Type I topoisomerases change Lk by ±1, in effect passing one strand of the DNA molecule through the other. Mechanistically, the

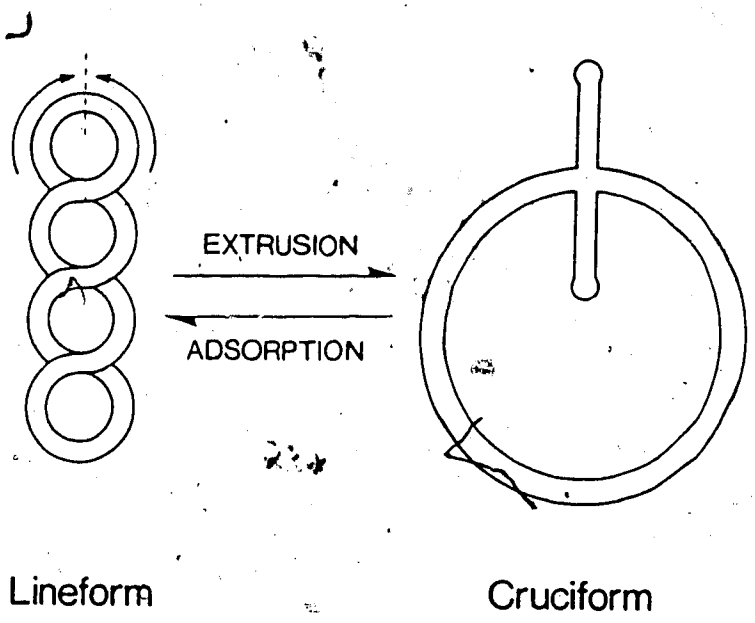


Figure I.5 Topological impact of cruciform extrusion. On the left side of the reaction, inverted repeat sequences (inverted arrows) are present in the lineform in a negatively supercoiled circular plasmid DNA molecule. On the right side of the reaction, the same sequences are shown after cruciform extrusion. There is an accompanying decrease in the number of supercoils in the circular domain of the plasmid. As shown, the inverted repeat sequences were large enough to completely relax the DNA molecule.

enzymes nick the DNA, store that bond energy in the form of a covalent phosphate ester linkage to the DNA, change Lk and then reseal the nick. The actual mechanism of strand passage is unclear. No external energy source is required as the enzymes passively relax positive or negative supercoiling. Type II topoisomerases change Lk by ±2. Energy cofactors are required as they can, in some instances, introduce supercoils as well as remove them. Type II topoisomerases, gyrase of *E. coli*, for example, introduce transient double-strand breaks allowing double-strand passage through duplex DNA.

Bacterial DNA *in vivo* is highly underwound and has a titratable superhelical density of -0.06 to -0.07 (Lilley, 1986b). The superhelical density of DNA is equal to the ratio of its linking difference ($\Delta Lk = Lk - Lk^{\circ}$) to the number of turns of helix in its relaxed state (Lk°). Only about half of the native superhelical density is manifested *in vivo* in the form of torsional strain. The remainder may be constrained by the winding of the DNA about the protein HU (Rouviere-Yaniv and Gross, 1975; Greaves *et al.*, 1985; Peck and Wang, 1985; Lilley, 1986b). The situation is even more unclear in eukaryotic cells where the torsional state of the DNA is greatly influenced by the nucleosomal wrapping of the DNA (Sinden *et al.*, 1980). The amount of native superhelical strain in eukaryotic DNAs may be even less than that in prokaryotes because the DNA is extensively wrapped around histones. In certain states however, chromosomal DNA may behave more like naked supercoils (Ryoji and Worcel, 1984; Luchnik *et al.*, 1985). There is increasing evidence that changes in the degree of supercoiling *in vivo* modulate gene expression (Menzel and Gellert, 1983; Drlica, 1984; Han *et al.*, 1985; Ellison *et al.*, 1987) and it has been argued that this capacity is mediated by changes in DNA secondary structure that influence the interaction of the DNA with transcriptional proteins (Iacono-Connors and Kowalski, 1986; Lilley, 1986b).

Torsional strain can be accommodated by alterations in the secondary and the tertiary structure of the DNA molecule. The balance between modifications in secondary and tertiary structure will be influenced by the nature of the sequences, environmental conditions such as temperature and ionic strength, and the degree of supercoiling. Cruciformation is just one example of a change in secondary structure that absorbs the torsional strain of negative supercoiling. Other possible transitions involve the formation of left-handed Z-DNA (Wang *et al.*, 1979), slippage structures with B-form DNA characteristics (Dickerson and Drew, 1981), the melting of DNA (Vinograd *et al.*, 1968; Walter *et al.*, 1967) such as may facilitate certain DNA-protein interactions (Riggs *et al.*, 1972; Renz and Day, 1976; Lilley, 1986b) or the appearance of an uncharacterized structure associated with polypurine-polypyrimidine stretches (Lilley and Hallam, 1984; Pulleyblank *et al.*, 1985). In theory, any DNA transition that alters the helical twist of the DNA is a potential medium for the modulation of gene expression by manipulating the global influence of supercoiling (Iacono-Connors and Kowalski, 1986) or by providing a recognition-specific alternative structure. Cruciformation has been considered as a biologically significant event in these terms (Sinden *et al.*, 1980) and, furthermore, could influence DNA metabolism in other ways since cruciforms have been shown to bind histone octamers poorly (Nickol and Martin, 1983; Weintraub, 1983) and may disrupt chromatin structure. A recent report indicates that crucifrom structures alter the phasing of nucleosomes in chromatin (Nobile *et al.*, 1986).

Attempts to find direct evidence for the presence of cruciforms *in vivo* have been unsuccessful. In one case, DNA that was cross-linked *in vivo* was isolated and probed for "trapped" cruciforms but no significant presence was discovered (Sinden *et al.*, 1983). Lyamichev *et al.* (1984) and others (Courey and Wang, 1983; Greaves *et al.*,

1985) looked for changes in the supercoiling of a plasmid bearing an inverted repeat sequence *in vivo* but found no significant differences between this plasmid and one without the repeat. The assumption was that a plasmid bearing a cruciform would have a higher titratable superhelical density because *in vivo* mechanisms would have restored the supercoiling lost during cruciformation. Haniford and Pulleyblank (1985) did detect changes in plasmid supercoiling associated with the extrusion of a (dA-dT) repeating sequence into a cruciform. However, for the effects to be observed, these authors had to inhibit protein synthesis by poisoning the host bacteria with chloramphenicol. There is, as described earlier, a possible involvement of cruciformation in the instability of large inverted repeat sequences *in vivo* but no clear experimental example of a biologically significant cruciform extrusion event *in vivo*. There may well be specific situations where a functional role for cruciforms exist. The probability that a lineform-to-cruciform transition is biologically feasible within a given DNA sequence will be a function of the unique thermodynamic parameters associated with the reaction.

F. The Energetics of Cruciformation

Theoretical calculations of the amount of supercoiling energy that would be required to stabilize a cruciform structure extruded from a typical native inverted repeat sequence suggest a value of 25 kcal/mol. This amount of energy, and more, is present in bacterial DNA presuming that the titrated superhelical densities measured *in vitro* for native plasmid DNA reflect *in vivo* values, either locally or transiently (Mizuuchi *et al.*, 1982b; Lilley, 1986b). Lineform-to-cruciform transitions in a number of natural and artificial palindromes have been studied *in vitro*. Typically, the transition is detected by using chemical or enzymatic probes specific for structural elements of the cruciforms or by

observing changes in the electrophoretic behaviour of the DNA associated with the transition.

The distinctive features of the cruciform identifiable by current probes are the single-stranded hairpin loops that consist of the nucleotides present at the two-fold axis of the inverted repeat and the fourth-rank DNA junction at the base of the cruciform stems. The unpaired nucleotides residing in the hairpin loops can be mapped because of their reactivity to bromoacetaldehyde or osmium tetroxide (Greaves *et al.*, 1985; Haniford and Pulleyblank, 1985; Naylor *et al.*, 1986). Bromoacetaldehyde forms adducts with exposed dA and dC residues that are sensitive to N-glycosidic cleavage. Apurinic/apyrimidinic sites created this way are cleavable by treatment with piperidine (Yoshio, 1980). Osmium tetroxide will also form adducts with exposed bases, rendering them permanently single-stranded (Gliken *et al.*, 1984) and susceptible to single-strand-specific nuclease digestion. Cleavage positions following either reaction can be mapped with respect to nearby restriction sites and, indicative of cruciformation, these will map to the axis of dyad symmetry for an inverted repeat sequence. The loops can aso be identified directly by their susceptibilty to single-strand-specific endonucleases such as *Bal 31* nuclease, mung bean nuclease and S1 nuclease (Lilley and Hallam, 1984; Naylor *et al.*, 1986). As a rule, these probes do not detect any single-strandedness in the vicinity of the base of the cruciform structure suggesting that there are no unpaired bases at the junction. NMR evidence supports this conclusion (see section G, this chapter). An alternative method is applicable if the inverted repeat sequence that is being probed has a restriction site at the dyad axis (Mizuuchi *et al.*; 1982b; DeLange *et al.*, 1986). In the cruciform, the site will be disrupted by the formation of the hairpin loops and no cutting would be observed, while cleavage would be observed if the inverted repeat sequences

were in the lineform. The base of a cruciform, on the other hand, can be detected by its susceptibility to the Holliday-resolving enzymes T4 endonuclease VII (Lilley and Hallam, 1984; Lilley and Kemper, 1984; Naylor *et al.*, 1986) and T7 endonuclease I (Panayotatos and Wells, 1981; DeLange *et al.*, 1986) which cleave at or very near to the actual phosphodiester crossover (see section I, this chapter). Once the extrusion of a cruciform has been verified, cruciformation is more readily analyzed through changes conferred upon the electrophoretic behaviour of the DNA (Wang *et al.*, 1983; Gellert *et al.*, 1983, Panyutin *et al.*, 1984).

The free energy of formation of a cruciform structure is estimated from the free energy of supercoiling of the plasmid topoisomer (a plasmid molecule with a unique Lk) at which there is approximately 50% cruciform formation (reviewed by Lilley, 1986b). The identification of the topoisomer at the transition point can be made by observing changes in the gel migration of individual purified individual topoisomers or, more conveniently, by a similar analysis following the two-dimensional agarose gel electrophoresis of complete topoisomer distributions (Wang *et al.*, 1983). The free energy of supercoiling of a topoisomer can be computed given the topoisomers linking difference and size in base pairs, according to the equation given in Chapter II, Materials and Methods (Depew and Wang, 1975; Pulleyblank *et al.*, 1975; Horowitz and Wang, 1984). The energy of formation related to any torsion-induced structural transition approximates the difference in free energy of supercoiling in the topoisomer identified at the mid-point of the transition before and after the transition has occurred. With cruciformation, it is important to realize that once the hairpins have formed the size of the cruciform should have little bearing on its energy of formation. Variably sized cruciforms extruded from the same large inverted repeat sequence are related by branch migration at the cruciform crossover. This implies

an equal exchange of base pairs between the interstranded and intrastranded branches at no energy cost. Branch migration at the cruciform junction should not affect the stability of the structure once it has formed unless the inverted repeat is imperfect and extra-helical bases appear in the cruciform hairpins (Robinson and Seeman, 1987). The appearance of extra-helical bases would shift the equilibrium to the lineform of these sequences since unpaired bases are in a higher free energy form than Watson-Crick base pairs.

Linking differences are determined experimentally by the band counting method of Keller (1975). How this is done is best explained in the context of the two-dimensional agarose gel technique that has been used in experiments described in Chapter II. First, complete topoisomer distributions are prepared either by the limited relaxation of native plasmid DNA (which characteristically consists of a population of topoisomers with a limited range of linking numbers) with topoisomerase I (Lee and Bauer, 1985) or by completely relaxing the DNA with topoisomerase I in the presence of variable amounts of the intercalating drug ethidium bromide (Panyutin *et al.*, 1984). DNA intercalators unwind DNA (ΔTw is negative and ΔWr is usually positive) to a degree that is directly related to the amount of drug bound. Following the enzymatic relaxation of treated DNA, the removal of the drug increases Tw and a degree of negative supercoiling (as a negative ΔWr) is restored to the DNA. A significant range of negatively supercoiled topoisomers can be prepared using variable concentrations of drug.

Next, the topoisomer distribution is electrophoresed and the topoisomers are separated on the basis linking difference and secondary structure (collectively, these parameters will determine the number of superturns in the molecule). The gel is then turned 90° relative to the direction of electrophoresis and re-electrophoresed in a second dimension in the

presence of the intercalating drug chloroquine. Unwinding of the DNA induced by chloroquine binding destabilizes any torsion-induced alternative structures which were present in the first dimension so that separation in the second dimension is solely on the basis of linking difference. A distribution of topoisomers differing by a linking number of one will form a smooth, uninterrupted curve of DNA bands or spots. Topoisomers in which cruciformation (or another transition associated with a ΔTw) has occurred will migrate slower in the first dimension and will be displaced above the profile. The ΔLk of the topoisomer at the threshold of cruciformation, or demonstrating 50% cruciform formation, is determined by counting down the profile beginning with the slowest migrating band whose ΔLk will approximate zero. The amount of sequence extruded in the form of a cruciform can be estimated from the reduction in topoisomer mobility since every supercoil lost will be roughly equivalent to the removal of one turn of helix from the circular core of the topoisomer by cruciformation ($-\Delta Tw = \Delta Wr$, since there is no change in linking number) (Haniford and Pulleyblank, 1985; Nayl[redacted]., 1986). Generally, if the ΔTw characteristic of a structural transition is known, the amount of sequence committed to the transition can be estimated this way. Finally, the energy of supercoiling (torsional energy) remaining in the cruciform-bearing topoisomer is estimated by comparing its mobility with lineform plasmids (of a lower negative ΔLk) and substituting that ΔLk into the ΔG_S calculations.

In close agreement with theoretical calculations, the energy of formation of cruciforms derived from several short palindromes *in vitro* is typically around 18 to 25 kcal/mol (Courey and Wang, 1983; Gellert *et al.*; 1983; Lilley and Hallam, 1984; Panyutin *et al.*, 1984). Native inverted repeat sequences that were perfect yet non-palindromic were

tested, as were some artificial palindromic sequences. Cruciform structures derived from a repeating $(dA\text{-}dT)_{34}$ sequence from a *Xenopus* globin gene had a much lower energy of formation of about 13 kcal/mol (Greaves *et al.*, 1985; Haniford and Pulleyblank, 1985). Modest levels of supercoiling (a superhelical density of around -0.03) are required to stabilize these cruciforms, equivalent to what is expected to be the effective superhelical density of DNA *in vivo* (Lilley, 1986b). Specific features of cruciform structures contribute to their energy of formation. A very small cruciform, for instance, would not be stable since the free energy return provided by partial relaxation would not be sufficient to meet the energy of formation. Imperfect inverted repeats could manifest higher energies of cruciform formation if it leads to the presence of unpaired bases in the cruciform stems or greater than four unpaired bases in the hairpin loops (Mizuuchi *et al.*, 1982b). A minimum of four unpaired bases in the hairpin loops has been indicated (Lilley, 1986a).

A second important thermodynamic parameter governing cruciformation, the energy of activation, is determined by correlating rates of extrusion to temperature in the form of Arrhenius plots (Sullivan and Lilley, 1986). The energy of activation is a measure of the energy barrier to cruciformation (the free energy of the transition intermediate) and reflects the facility of the lineform-to-cuciform transition under the experimental conditions. Early reports indicated that the transition was very slow (measured in hours) at reasonable temperatures, optimal salt concentrations and at superhelical densities sufficient to stabilize the cruciform (Courey and Wang, 1983; Gellert *et al.*, 1983; Sinden and Pettijohn, 1984). The transition was so slow that these authors concluded that cruciformation would be kinetically forbidden *in vivo*. Some sequences, however, can

form cruciforms with relative ease. Again, the repeating $(dA\text{-}dT)_{34}$ sequence was exceptional in this context. No high energy barrier was observed during the extrusion of these sequences into the cruciform (Greaves *et al.*, 1985; Haniford and Pulleyblank, 1985). On the other hand, it appears that the presence of an A+T-rich sequence downstream of an inverted repeat sequence is sufficient to alter the kinetic pathway to the extent that the energy of activation increases dramatically (Sullivan and Lilley, 1986). The introduction of downstream A+T sequences also changed the optimum salt concentration for the extrusion event, indicative of an alternative mechanistic pathway for cruciformation. Considering that repeating (dA-dT) sequences can be extruded into cruciforms *in vivo* (Haniford and Pulleyblank, 1985) and the evidence indicating that cruciform transitions in (dA-dT) sequences occur readily if torsionally stressed, a biologically relevent cruciformation reaction could, with the greatest probability, be associated with A+T palindromic (or near-palindromic) sequences in a context that was otherwise not A+T-rich.

The nature of the high energy intermediate in cruciformation is reflected in the kinetics of cruciform extrusion though its precise structure is not known. Salt concentrations of 0.03M to 0.04M were optimal for extrusion of the 48 base pair palindrome in pUC7 plasmid DNA. Above or below these concentrations, rates fell off precipitously. Spermidine enhanced rates of extrusion 3-fold whereas other multivalent cations like magnesium had no effect. Extrusion rates also increased with increases in negative linking differences. The spermidine effect suggests that neutralization of repulsive electrostatic forces is required in the transition intermediate (to stabilize the initial DNA junction, for instance) and the supercoiling effect indicates that negative supercoiling

stabilizes the reaction intermediate. Gellert *et al.* (1983) have argued that simple denatured loops were not likely the intermediate because significant differences in (G+C) content at the axes of inverted repeat sequences appeared to have little influence on the facility of extrusion. Nevertheless, they did acknowledge that the complex dependence of rates on NaCl concentrations could reflect the involvement of denaturation and renaturation during extrusion. Panyutin *et al.* (1984) tentatively identified what they believed to be an intermediate in the transition to a cruciform. A DNA species with an electrophoretic mobility intermediate to that of the lineform and cruciform plasmid was observed, but it has not been characterized any further. Lilley (1985; and in Sullivan and Lilley, 1986) proposes two mechanistic pathways for the extrusion of cruciforms. Both pathways invoke the denaturation of the DNA; in one, the DNA involved in the transition is completely denatured as an intermediate in the strand exchange process, whereas in the second there is only partial denaturation of the involved sequences and branch migration completes the strand exchange reaction. A tetra-stranded DNA structure has been proposed to mediate branch migration during genetic recombination (Wilson, 1979) and could conceivably participate in cruciform extrusion as well.

To summarize, certain features of inverted repeat sequences such as their base composition and symmetry, their context, environmental milieu and topology can influence cruciform extrusion. The sequences of the inverted repeats characteristic of prokaryotic DNA have the nature that they probably kinetically forbid spontaneous extrusion under physiological conditions. In response to changes in topology A+T-rich sequences are notably polymorphic (Haniford and Pulleyblank, 1985) and perhaps by some related structural transition they influence the extrusion in nearby inverted repeats (see Iacono--Connors and Kowalski, 1986). Barriers to extrusion could be overcome by

catalysis *in vivo*. Any process that promoted the the unwinding or denaturation of DNA sequences, as might occur during replication, transcription or recombination, could facilitate cruciform formation (Courey and Wang, 1983; Warren and Green, 1985). Likewise, certain proteins may catalyze cruciformation if they bind to and thus stabilize the hairpin loops or the high energy intermediate in cruciformation. A specific protein, such as a helicase, could catalyze DNA unwinding. In light of the inability to detect any steady state levels of cruciforms and the proven instability of inverted repeat sequences *in vivo*, if cruciformation is a biologically significant event, it is likely to happen transiently and under specific conditions which regulate its occurrence. It is probably not coincidental that eukaryotic systems tolerate larger inverted repeat DNA sequences and stretches of repeating (dA-dT) sequences. Bulk eukaryotic DNA is believed not to be under significant superhelical strain and this attribute may serve to prevent deleterious, non-specific cruciform transitions (Sinden *et al.*, 1980). At the same time, cruciformation in eukaryotic DNA is potentially more manageable simply because of the ability to sequester and regulate the superhelical strain in chromatin.

G. The Structure of the Cruciform Crossover: A Fourth-ranked DNA Junction

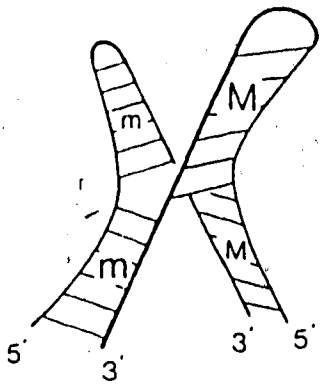
The branched DNA junction formed at the base of a cruciform is locally identical to a Holliday recombination intermediate and thus is potentially a biologically functional structure (Mizuuchi *et al.*, 1982a). The Holliday structure (Holliday, 1964) is widely accepted as the intermediate formed by the exchange of single-strands during recombination between two homologous DNA duplexes (Sigal and Alberts, 1972;

Meselson and Radding, 1975; Szostak, 1983). Recombination enzymes that recognize and resolve Holliday structures *in vivo* will likely recognize cruciform junctions as well (Courey and Wang, 1983; Gellert *et al.*, 1983). Four DNA duplexes converge at the branchpoint in these structures to form DNA junctions of the fourth rank. Strictly speaking, Holliday structures always possess an axis of two-fold symmetry and can branch migrate in either direction whereas this is not true for those cruciforms that have been extruded to the full extent of the inverted repeat sequence and are trapped there by excess supercoiling in the DNA molecules. Beyond what has already been inferred about the structure of four-way junctions from experiments with chemical and enzymatic probes, their actual structure remains uncertain.

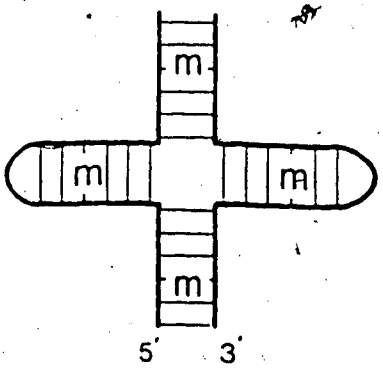
A variety of four-way junctions have been created for physical and enzymatic studies. Cruciform-bearing plasmid DNA molecules have been the most studied. Other substrates have included chi-structures, artificial junctions created from restriction fragments (Hsu and Landy, 1984) and synthetic polynucleotides (Kallenbach *et al.*, 1983; Evans and Kolodner, 1987; Kolodner *et al.*, 1987), as well as artificial cruciforms formed from restriction fragments with limited homology (Gough and Lilley, 1985). These various Holliday structure analogs differ in significant ways. Some possess dyad symmetry while others do not; some are topologically constrained and under torsional strain (most plasmid-borne cruciforms) while others have topologically free DNA branches (chi-structures); and some branches form hairpin loops while others have naked DNA ends. These substrates have been used to study the kinetics of branch migration at Holliday junctions (Thompson *et al.*, 1976; Mizuuchi *et al.*, 1982b), junction conformation and in the characterization of Holliday-resolving enzymes (discussed in the following section).

For clearer definitions of DNA structure, investigators currently rely on the application of high resolution techniques like NMR spectroscopy and X-ray crystallography (Wang *et al.*, 1979; Patel *et al.*, 1985, for example). NMR spectroscopy has been applied to the study of small fourth-rank junctions (Kallenbach *et al.*, 1983) but results have been ambiguous. In solution, either the junction existed in two conformers of comparable stability or certain hydrogen-bonded base pairs existed in more than one unique environment within the same conformer. The data did indicate that all base pairs were intact, including those adjacent to the crossover point, but it was not possible to ascertain whether the duplex bounding the junction was typical Watson-Crick DNA. A monoclonal antibody raised against and specific for cruciform DNA has been obtained, but precisely what structural feature of the junction it recognizes has yet to be established (Frappier *et al.*, 1987).

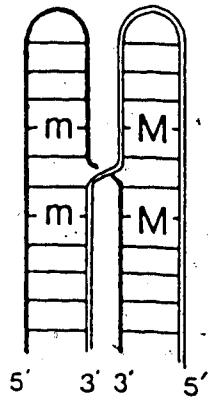
When Holliday (1964) first reported his version of the recombination junction he represented it, perhaps by happenstance, with the four DNA branches arranged tetrahedrally in space. Sigal and Alberts (1972) later proposed that the Holliday structure was planar and the homologous duplexes were aligned parallel to each other. This model has since been called the unperturbed-helix-axis (UHA) form (Seeman, 1981) since pairs of branches (the branches of a homolog, for example) are base-stacked and share the same helix axis. The UHA structure was favored because it was readily accommodated by space-filling models, the energy gained through base-stacking interactions would enhance its stability, and the simple rotation of the helices in the same sense (rotary diffusion) could drive branch point migration. At about the same time Sobell (1972) postulated that the Holliday structure was planar with the four branches related by pseudo-four-fold symmetry. The three structures are illustrated in Figure I.6, modified



Tetrahedral



Sobell (S4)



UHA

Figure I.6. Holliday structure conformation. Three versions of a Holliday structure are shown. The tetrahedrally arranged form has been described by Holliday (1964). The Sobell form has pseudo-four-fold symmetry (S4) and the unperturbed-helix-axis (UHA) form is described in Sigal and Alberts (1972). Helical turns have been removed. m: minor groove; M: major groove.

so as to represent cruciform structures. There is no physical evidence directly supporting one conformation over the others. Electron microscopic observations tend to favor either the Sobell model or the tetrahedral arrangement (Dressler and Potter, 1982) since parallel duplexes are not observed. Gough and Lilley (1985) reported that DNAs bearing cruciform junctions behave electrophoretically much like bent pieces of DNA and suggest that this could only be so if the branches of the junction were tetrahedrally arranged. Clearly the three structures are related by branch flexibility at the junction and many intermediate conformations are possible.

A theoretical consideration of the structure and dynamics of Holliday structures by Robinson and Seeman (1987) has emphasized the significance of junctional flexibility. Energy minimization calculations performed by these authors were used to compare the relative stabilities of the three structural forms described in Figure I.6 and all intermediate forms as well. Their model building relied predominantly on estimates of the electrostatic repulsion and electrostatic torque in the phosphodiester backbone. The preferred structure they identified is most like the planar conformation proposed by Sobell. They suggested however that branch migration was not likely to occur in this conformation since rotary diffusion in one branch would be resisted by viscous drag encountered by the adjoining branches. Instead, they predicted that during branch migration the structure transiently converts to the higher energy UHA form because little resistance would be encountered in branches that shared the same helical axis.

Conformation and conformational flexibility are important parameters governing the interaction of proteins with the Holliday structure. It is not obvious, but of considerable importance, that junction-specific proteins will discern a sidedness to the junction. At a

junction with dyad symmetry in terms of the primary DNA sequence, sidedness is determined by the disposition of the major and minor helical grooves at the branchpoint (Robinson and Seeman, 1987). Presuming that normal B-form DNA surrounds the junction, above the plane of the junction the DNA branches will face each other across either of the two helical grooves. On the opposite face of the junction the opposite grooves will be in apposition. This sidedness can be appreciated from the drawings in Figure I.6 where the grooves at the junction have been labelled. Due to the helical nature of the DNA, the phasing of the grooves will obviously reverse itself every half-turn of helix (and repeat itself every turn of helix) further along the branches. Proteins which bind the junction will most assuredly discriminate between the faces of the junction on the basis of specific interactions it will have with the major and minor grooves. Unfortunately, the only proteins known to specifically recognize a Holliday crossover are the recently discovered monoclonal antibody (Frappier *et al.*, 1987) and the Holliday-resolving enzymes, and their binding interactions have only been considered in gross structural terms.

H. The Structural Dynamics of Holliday Junctions

In models of genetic recombination (Holliday, 1964; Sigal and Alberts, 1972; Meselson and Radding, 1975; Sosztak, 1983) the UHA form of the Holliday intermediate has been endowed with two significant dynamic properties: branch migration and isomerization. Branch migration was imposed on the model to account for the genetic evidence for strand exchange and the formation of heterologous DNA. Isomerization of the Holliday junction accounted for the crossing over of distant genetic markers as would occur when homologs exchanged base-stacked branches (Meselson and Radding, 1975). Branch

point migration has been studied *in vitro* (Thompson *et al.*, 1976; Warner *et al.*, 1979; Mizuuchi *et al.*, 1982b) but there remains no direct approach for testing assumptions made with regards to isomerization mechanics.

Branch point migration implies the processive or random-walk exchange of DNA strands that occurs during genetic recombination between homologs and the extrusion or adsorption of cruciform structures. *In vivo*, branch migration is probably dependent upon interactions of recombining DNA with *rec*A protein and is driven by an external energy source (McEntee *et al.*, 1979). Alternatively, it may be driven by DNA supercoiling (Gellert *et al.*, 1979). In the absence of an energy source, rotary diffusion of the double helices can drive branch movement (Meselson, 1972; Sigal and Alberts, 1972). Theoretical calculations of the rate of branch movement that is driven by rotary diffusion (Meselson, 1972) far exceed rates observed experimentally (Thompson *et al.*, 1976; Warner *et al.*, 1978; Courey and Wang, 1983; Gellert *et al.*, 1983) suggesting that rotary diffusion is sufficient to drive branch movement but is not the rate limiting factor (Robinson and Seeman, 1987). The experimentally observed rate is only 0.3% of the theoretical rate (Thompson *et al.*, 1976).

Rotary diffusion is most effectively transmitted into branch movement when Holliday structures are in the UHA conformation accounting for the wide acceptance of this structural form (Meselson and Radding, 1975; Robinson and Seeman, 1987). In this structure the exchanged strands (the crossing strands) tether two parallel DNA duplexes (see Figure I.6). Rotation, in the same sense, of each DNA duplex about its helical axis moves the crossover up or down depending on the sense of the rotation. Having suggested that the Holliday structure converts from the Sobell form (the planar

conformation with four-fold symmetry) to the UHA form during branch migration, Robinson and Seeman (1987) predicted that mean lifetime for the UHA structure would only allow for the exchange of 2 or 3 base pairs before flexural motion drove it back to the Sobell form. Isomerization of the Sobell form to the UHA form may be one factor contributing to the slower than expected rates observed for branch migration *in vitro* (Robinson and Seeman, 1987). It is known that *recA*-mediated strand exchange can proceed past broad regions of nonhomologous DNA sequence (Flory *et al.*, 1984). Consequently, *in vivo* mismatches do not present the barrier to branch migration that they would for spontaneous, uncatalyzed reactions. The equilibrium would still be shifted to the state where these nucleotides were base paired however (Robinson and Seeman, 1987).

These considerations of branch migration have omitted the influence of DNA topology. Meselson and Radding (1975) assumed that DNA would be flexible enough to tolerate the degree of rotary diffusion required for branch migration *in vivo*. Certainly, if the DNA was topologically constrained, rotation of the DNA would introduce a supercoiling component into the equation. For example, the rotation of a DNA branch with one fixed end (by subcellular attachment or constained by a nucleosomal structure) would alter the linkage of that DNA (change Lk) and since the number of base pairs is fixed, Wr would change accordingly. Supercoiling could oppose branch migration or promote it, as in the case of cruciform extrusion in plasmid molecules, depending on whether it was negative or positive. Similarly, topology could influence an isomerization of the junction by restricting certain movements of the branches through space (like rotations).

A Holliday junction is capable of undergoing a number of specific structural transitions

that have been identified as independent isomerization processes. Traditionally, isomerization has been used in reference to the arm movements that bring the uncrossed strands of a UHA structure into the crossed configuration (and vice versa) as originally described by Meselson and Radding (1975). Branch migration may be regarded as a second form of isomerization but when considered solely in the context of UHA structures, unconstrained by the topology of the DNA, and driven by rotary diffusion, no gross conformational alteration would be imposed upon the junction. The orientation of the DNA branches can remain unchanged, as can the relative disposition of the helical grooves at the crossover point (Robinson and Seeman, 1987). A third form of isomerization can be imagined that reverses the phasing of the helical grooves at the branchpoint. In section (G) of this chapter Holliday structures were described as having two faces. In the Sobell structure, for instance, minor grooves will converge at the branchpoint on one face while major grooves will converge at the crossover on the opposite face. A potential, but as yet uncharacterized, isomerization associated with a change in Tw of the DNA could effect the phase reversal.

The isomerization process described by Meselson and Radding (1975) is reproduced in Figure I.7. They postulated that the Holliday-resolving enzymes would cleave either the pair of crossed or uncrossed strands in a UHA structure and therefore the outcome of the recombination event (the incidence of genetic crossing over) would be determined by the isomeric form of the junction. By comparison, if these enzymes cleaved the Sobell form of the crossover, in which the junction axes are structurally equivalent owing to the pseudo-four-fold symmetry, no isomerization need be imposed to account for crossing over. Crossing over could be dictated, in this case, by the random selection of cleavage axes by the resolving enzyme. According to the scheme of Meselson and Radding (1975)

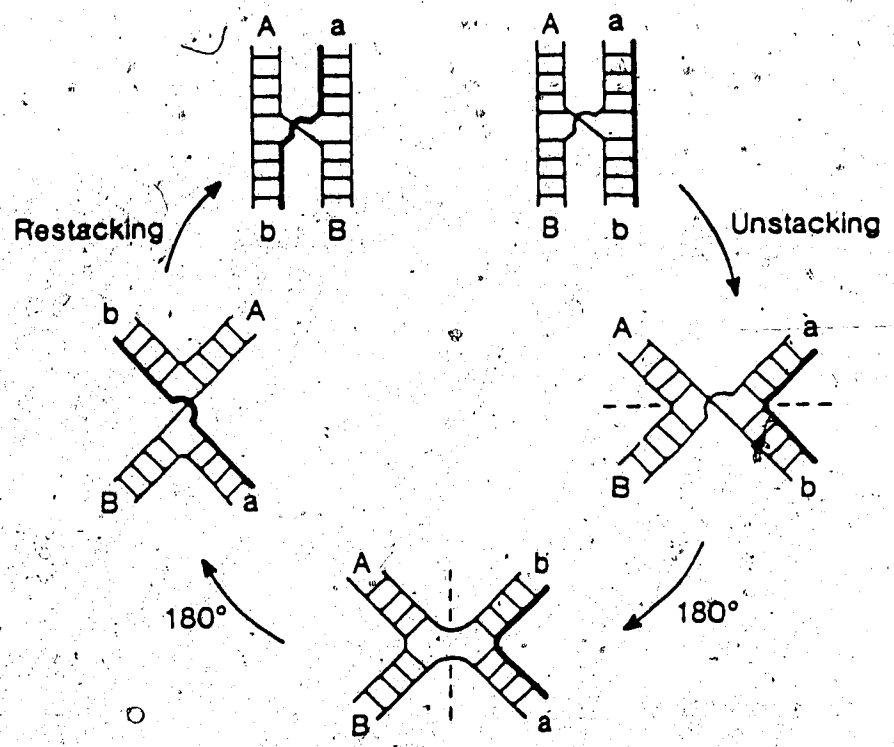


Figure I.7. Isomerization of crossing strands in a UHA form Holliday structure. The Meselson and Radding isomerization is diagrammed. Branch rotations bring alternative pairs of DNA strands into the "crossed" configuration. Axes of rotation are indicated by the greek symbols. Homologous segments are labeled A/a and B/b.

there is no net topological component to the isomerization reaction they describe. In structural terms, the isomerization entails a conformational transition through familiar structures:

$$\text{UHA} \leftrightarrow \text{T2} \leftrightarrow \text{S4} \leftrightarrow \text{T2}^* \leftrightarrow \text{UHA}^*,$$

where UHA and UHA* are isomeric forms, S4 corresponds to the Sobell structure and T2 and T2* correspond to the tetrahedrally arranged isomer forms. An unanswered question is whether or not the topology of DNA *in vivo* or the topology of plasmid-borne cruciform substrates *in vitro* influence the flexibility of the respective junctions and thus the kinetics of the transition.

I. Enzymes that Resolve Holliday Junctions

Endonucleases with a striking specificity for branched DNA structures have been used with increasing regularity to probe for Holliday structures. They are far more specific than the single-strand-specific enzymes (S1 nuclease, P1 nuclease, mung bean nuclease and *Bal* 31) because they cleave very near the crossover point, in regions of duplex DNA (Panayotatos and Wells, 1981; Kemper *et al.*, 1984). Moreover, two coincidental nicks are made on opposite strands at the junction, resulting in the resolution of the branched structure to linear nicked duplex products. In the example of a true Holliday intermediate, resolution products would be the resolved daughter molecules of the recombination event. The equivalent treatment of a plasmid-borne cruciform reduces a circular plasmid to a linear molecule with nicked hairpin ends (Figure I.8) when pair-wise cleavage occurs across one of two junctional axes. The hairpin ends are derived from the cruciform stem and loop structures. The distinguishing feature of the cruciform extrusion model for telomere resolution (McFadden and Morgan, 1982; Szostak, 1983) involves precisely this reaction whereby the "flip" and "flop" cruciform hairpins become the linear poxviruses genomic termini. Three bacteriophage enzymes are known to have Holliday-resolving activity. They are lambda phage Int protein (Hsu and Landy, 1984), T7 gene 3 product, endonuclease I (Panayotatos and Wells, 1981) and T4 gene 49 product, endonuclease VII

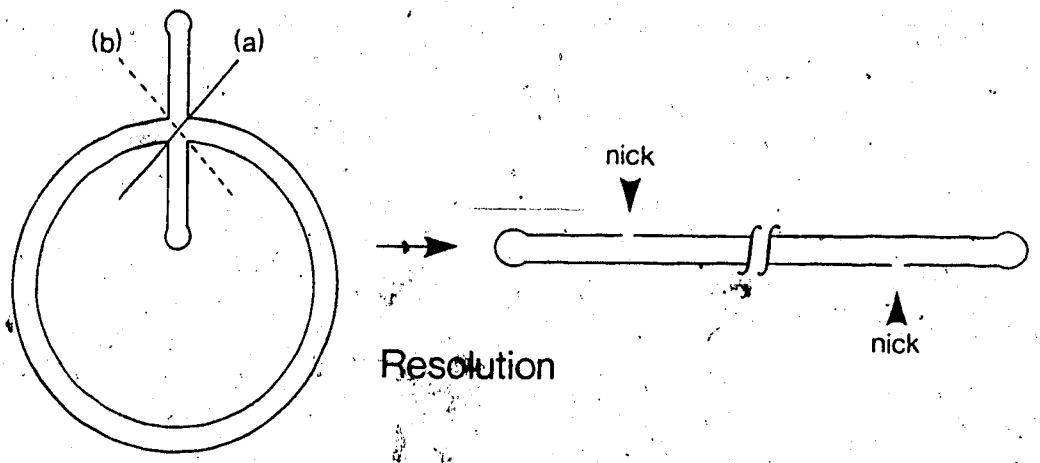


Figure I.8. Resolution of a cruciform junction. The resolving axes are indicated by (a) and (b). Resolution would be accomplished by cleavage of the opposite strands along one of these axes. The product has nicked hairpin termini.

(Mizuuchi *et al.*, 1982a; Lilley and Kemper, 1984). Recently, an activity that resolves cruciform junctions has been identified in yeast extracts (West and Korner, 1986; Symington and Kolodner, 1986; Kolodner *et al.*, 1987). Int protein is sequence specific, mediates integrative recombination in lambda and religates the broken DNA strands when the recombination process is completed (Kikuchi and Nash, 1979).

The bacteriophage T7 and T4 endonucleases have been characterized both *in vivo* and *in vitro* and appear to cleave a spectrum of branched DNA substrates. The T7 and T4 enzymes are required for the maturation of their respective phage DNAs and are involved in genetic recombination (Kerr and Sadowski, 1975; Kemper and Black, 1976). *In vivo*, gene 3 and gene 49 mutants of bacteriophage T7 and T4, respectively, accumulate high molecular weight, multiply branched DNA by-products in infected cells which can not be packaged (Stratling *et al.*, 1973; Kemper and Brown, 1976; Kemper and Janz, 1976; Paetkau *et al.*, 1977; Tsujimoto and Ogawa, 1978). The two enzymes differ in that the T7 enzyme has been shown to have a role in the degradation of host DNA (Center *et al.*, 1970) but nonetheless, T7 gene 3 can complement T4 gene 49 mutants (deMassy *et al.*, 1985). T7 endonuclease I (Center and Richardson, 1970; Sadowski, 1971) and T4 endonuclease VII (Kemper *et al.*, 1981; Kemper and Garabett, 1981) have been purified and they have been shown to demonstrate a strong preference for single-stranded DNA substrates. In addition to cleaving four-way junctions, both enzymes can cleave three-way junctions (Y-structures) that have single nicks at the junction or that have uninterrupted phosphodiester backbones (Jensch and Kemper, 1986; deMassy *et al.*, 1987). In spite of their demonstrated specificity for single-stranded substrates, the T4 and T7 endonucleases do not cleave cruciform structures in the single-stranded loop segments of the hairpin arms.

In vitro, T7 endonuclease I cleaves specific branched DNA substrates on the 5' side of the junction, anywhere from 1 to 4 bases away from the predicted crossover phosphodiester (de Massy *et al.*, 1985; Haniford and Pulleyblank, 1985; de Massy *et al.*, 1987). T4 endonuclease VII, on the other hand, cleaves 3' of the junction, 1 to 5 bases removed from the crossover point (Kemper *et al.*, 1985; Jensch and Kemper, 1986). In both cases

nicks are left unsealed in the product molecules but there is substantial evidence that the nicks can be closed with ligase demonstrating that no single-strand gaps are generated. The relatively broad region of cutting adjacent to the crossover may reflect either the lack of structural specificity in the action of the enzyme or the fact that none of the substrates used had the strictly immobile, or fixed, branchpoint they were designed to have (Jensch and Kemper, 1986). Regardless, there appeared to be little difference in the manner of cutting relative to the rank of the junction, the degree of torsional strain in the DNA branches or the symmetry of the sequences. With T4 endonuclease VII, a modest level of sequence specificity was observed (Kemper *et al.*, 1985).

The observation that resolution by these enzymes generates ligatable nicks demonstrates that there has been symmetrical cleavage across both strands of the junction. Consequently, the variability in the number of sensitive sites near the junction is even more remarkable. In the event that this observation is independent of the mobility of the junction, either (i) there is a strict cooperative interaction between two catalytic moieties to ensure that equivalent cuts are made, or (ii) nicks are made sequentially and the position of the first nick directs the placement of the second (Jensch and Kemper, 1986). Kemper *et al.* (1985) have proposed that the enzyme acts as a dimer and binds to the junction in such a way that it can recognize the pair of crossing strands (as in a UHA structure) and cleaves them cooperatively.

J. The Thesis Proposition

Two general queries are addressed:

1. Does cruciform extrusion from the telomere inverted repeat sequences of poxvirus

replication intermediates mediate resolution to daughter telomeres?

2. What is the specificity of Holliday-resolving enzymes for branched DNA substrates?

The cruciform extrusion model for telomere resolution implies that, (i) cruciformation in the sequences of the telomere concatameric junction is energetically feasible *in vivo*, (ii) strand exchange can proceed past the asymmetric base pairs so that the "flip" and "flop" telomere forms are generated and, (iii) a virus-encoded resolving enzyme exists that can cleave the cruciform crossover at a specified point within the telomere sequences. DNA topology has a profound impact on cruciform extrusion and branch migration. It also affects the conformation of Holliday structures in a way that, as shall be shown, dramatically influences the susceptibility of the junction to resolving cleavage. Thus, the above questions and assumptions are considered with particular attention to DNA topology.

Given access to the cloned telomere sequences of SFV (in plasmid pSAlB.56A; DeLange *et al.*, 1986) and by employing techniques described in sections E and F of this chapter, it was possible to characterize the thermodynamic energy parameters governing cruciformation in the SFV concatameric junction *in vitro* (Chapter II). *In vivo* stability of the cloned inverted repeat was investigated, as was the effect of the cloned inverted repeat on the native superhelical density of the carrier plasmid. The approach was not necessarily intended to demonstrate that cruciformation occurs, or does not occur, *in vivo*. Rather, an understanding of the energetics of cruciformation would provide an insight into the manner by which the cell controls cruciformation, because given the right circumstances cruciformation could indeed occur spontaneously. And depending on the resolution mechanism, either cruciformation within the telomere sequences is necessary or

it is deleterious. By deletion analysis, the contributions that certain sequences and structural features of the telomeres make to the energy parameters of cruciformation were estimated. In particular, the significance of conserved telomere features, like their A+T-richness and the presence of extrahelical bases, was of interest. Sequences distal to the inverted repeat axis were also deleted to test whether there were contextual factors that influenced the thermodynamic parameters. This information would form the basis of arguments for the biological feasibility of cruciformation in what has been identified as a bona fide replicative intermediate in telomere resolution.

Cruciforms were detected using the restriction enzyme *Afl*II, S1 nuclease and T7 endonuclease I as structural probes. The resolution of plasmids bearing the telomere sequences in the cruciform configuration can be accomplished using T7 endonuclease I (DeLange *et al.*, 1986). The T7 endonuclease I cleavage sites were mapped to the telomere sequences. Under the circumstances of the experiment, mapping the base of the extruded cruciform proved to be an accurate way to measure the titratable superhelical density of the plasmid DNA (assuming number of superturns equalled turns of helix extruded). The distribution of supercoiling in the plasmid preparation was typical but the pattern of cleavage was nonetheless unexpected. Rather than cleaving successive phosphodiester positions as would be predicted from established models for branch migration (Meselson, 1972; Sigal and Alberts, 1972; Meselson and Radding, 1975) the enzyme cleaved in discrete regions spaced, on average, 10 nucleotides apart along the telomere sequence. Our immediate failure to acount for these observations emphasized what little was known of the structure of Holliday junctions in topologically constrained DNA and the specificity of the Holliday-resolving enzymes. Despite minor differences in the specificity of Holliday-resolving enzymes, many characteristics are shared by this

class of enzyme and therefore it seemed worthwhile to study T7 endonuclease I more closely. Synthetic junctions with immobile and semi-mobile junctions were created and used as substrates to more clearly define the specificity of the enzyme. It was hoped that the range of substrates prepared would be sufficient to allow considerations of the influence sequence symmetry, branch mobility, rank of the junction and base sequence had on cleavage specificity.

Confident of the specificity of T7 endonuclease I (Chapter III), it was possible to look more closely at the structure of the Holliday crossover associated with the cruciform structure. The inability to cleave every phosphodiester traversed during branch migration suggested that the structure of the junction was periodically changing. The cruciform-bearing plasmid was a unique Holliday-type structure to probe because unrestricted branch migration was possible within a narrow range of superhelical densities and two of the junctional branches were topologically linked in the form of a circle. For the first time it was possible to specifically probe a Holliday junction during branch migration. It was possible to relate structure and structural isomerization to branch migration in response to the topology of the DNA molecule (Chapter IV).

Finally, in Chapter V, which summarizes conclusions, the experimental evidence is assimilated in a discussion that presupposes that cruciformation mediates the resolution of poxvirus telomeres. The influence of topology, Holliday junction conformation and branch migration are imposed on the model with the intention to show what adaptations must exist in infected cells in order to exploit cruciformation as a useful transition. An alternative to the cruciform extrusion model is also proposed.

Bibliography

Agard, D. A. and Sedat, J. W. (1983). *Nature* **302**, 676-681.

Baroudy, B. M., Vankatesan, S. and Moss, B. (1982). *Cell* **28**, 315-324.

Bateman, A. J. (1975). *Nature* **253**, 379.

Berman, J., Tachibana, C. Y. and Tye, B.-K. (1986). *P.N.A.S. USA* **83**, 3713-3717.

Berns, K. I., Muzyczka, N. and Hauswirth, W. W. (1985). p. 414. *In* Virology. B.N. Fields, ed. Raven Press, New York.

Blackburn, E. H. (1984). *Cell* **37**, 7-8.

Blackburn, E. H. (1985). *Trends in Gen.* **1**, 8-15.

Blackburn, E. H. and Challoner, P. B. (1984). *Cell* **36**, 447-457.

Blackburn, E. H. and Gall, J. G. (1978). *J. Mol. Biol.* **120**, 33-53.

Blackburn, E. H. and Szostak, J. W. (1984). *Ann. Rev. Biochem.* **53**, 163-194.

Boisy, R. and Astell, C. R. (1985). *Gene* **35**, 179-185.

Bolivar, F., Betlach, M. C., Heyneker, H., Shine, J., Rodriguez, R. and Boyer, H. (1977). *P.N.A.S. USA* **74**, 5265-5269.

Bourguignon, G. J., Tattersall, P. J. and Ward, D. C. (1976). *J. Virol.* **20**, 290-306.

Broker, T. R. (1973). *J. Mol. Biol.* **81**, 1-16.

Cavalier-Smith, T. (1974). *Nature* **250**, 467.

Cavalier-Smith, T. (1983). *Nature* **301**, 112-113.

Center, M. and Richardson, C. (1970). *J. Biol. Chem.* **245**, 6285-6291.

Center, M., Studier, F. W. and Richardson, C. C. (1970). *P.N.A.S. USA* **65**, 242-248.

Challberg, M. D. and Kelly, T. J. (1982). *Ann. Rev. Biochem.* **51**, 901-934.

Collins, J. (1981). *Cold Spring Harbor Symp. Quant. Biol.* **45**, 409-416.

Collins, J., Volckaert, G. and Nevers, P. (1982). *Gene* **19**, 139-146.

Cooke, H. J., Brown, W. R. A. and Rappold, G. A. (1985). *Nature* **317**, 687-692.

Courey, A. J. and Wang, J. C. (1983). *Cell* **33**, 817-829.

Dales, S. and Pogo, B. G. T. (1982). *In* The biology of poxviruses. Springer-Verlag, New York.

Dancis, B. M. and Holmquist, G. P. (1979). *J. Theor. Biol.* **78**, 211-224.

DeLange, A. M., Futcher, B., Morgan, A. R. and McFadden, G. (1984). *Gene* **27**, 13-21.

DeLange, A. M. and McFadden, G. (1986). *P.N.A.S. USA* **83**, 614-618.

DeLange, A. M. and McFadden, G. (1987). *J. Virol.* **61**, 1957-1963.

DeLange, A. M., Reddy, M., Scraba, D., Upton, C. and McFadden, G. (1986). *J. Virol.* **59**, 249-259.

deMassy, B., Studier, F. W., Dorgai, L., Appelbaum, E. and Weisberg, R. A. (1985). *Cold Spring Harbor. Symp. Quant. Biol.* **49**, 715-726.

deMassy, B., Weisberg, R. A. and Studier, F. W. (1987). *J. Mol. Biol.* **193**, 359-376.

Depew, R. E. and Wang, J. C. (1975). *P.N.A.S. USA* **72**, 4275-4279.

Dickerson, R. E. and Drew, H. R. (1981). *J. Mol. Biol.* **149**, 761-786.

Dressler, D. and Potter, H. (1982). *Ann. Rev. Biochem.* **51**, 727-762.

Drlica, K. (1984). *Microbiol. Rev.* **48**, 273-289.

Dugaiczyk, A., Robberson, D. L. and Ullrich, A. (1980). *Biochem.* **19**, 5869-5873.

Ellis, T. H. N. and Day, A. (1986). *EMBO J.* **5**, 2769-2774.

Ellison, M. J., Fenton, M. J., Ho, P. S. and Rich, A. (1987). *EMBO J.* **6**, 1513-1522.

Emery, H. S. and Weiner, A. M. (1981). *Cell* **26**, 411-419.

Evans, D. E. and Morgan, A. R. (1982). *J. Mol. Biol.* **160**, 117-122.

Evans, D. E. and Kolodner, R. (1987). *J. Biol. Chem.* **262**, 9160-9165.

Flory, S. S., Tsang, J., Muniyappa, K., Bianchi, M., Gonda, D., Kahn, R., Azhderian, E., Egner, C., Shaner, S. and Radding, C. M. (1985). *Cold Spring Harbor Symp. Quant. Biol.* **49**, 33-40.

Frappier, L., Price, G. B., Martin, R. G. and Zannis-Hadjopoulos, M. (1987). *J. Mol. Biol.* **193**, 751-758.

Frisque, R. J. (1983). *J, Virol.* **46**, 170-176.

Gellert, M. (1981). *Ann. Rev. Biochem.* **50**, 879-910.

Gellert, M., Mizuuchi, M., O'Dea, M. H., Ohmori, H. and Tomizawa, N. (1979). *Cold Spring Harbor Symp. Quant. Biol.* **43**, 33-40.

Gellert, M., O'Dea, M. H. and Mizuuchi, K. (1983). *P.N.A.S. USA* **80**, 5545-5549.

Geshelin, P. and Berns, K. I. (1974). *J. Mol. Biol.* **88**, 785-796.

Gierer, A. (1966). *Nature (London)* **212**, 1480-1481.

Gilbert, W., Gralla, J., Majors, J. and Maxam, A. (1975). *In* Protein-ligand interactions, pp. 193-210. H. Sund and G. Blaner, eds. Walter de Gruyter, Berlin.

Glicken, G. C., Vojtiskova, M., Rena-Descalzi, L. and Palecek, E. (1984). *Nucl. Acids Res.* **12**, 1725-1735.

Gonzalez, A., Talavera, A., Amendal, J. M. and Vinuela, E. (1986). *Nucl. Acids Res.* **14**, 6835-6844.

Gottschling, D. E. and Cech, T. R. (1984). *Cell* **38**, 501-510.

Gottschling, D. E. and Zakian, V. A. (1986). *Cell* **47**, 195-205.

Gough, G. W. and Lilley, D. M. J. (1985). *Nature* **313**, 154-156.

Greaves, D. R., Patient, R. K. and Lilley, D. M. J. (1985). *J. Mol. Biol.* **185**, 461-478.

Greider, C. W. and Blackburn, E. H. (1985). *Cell* **43**, 405-413.

Hagan, C. E. and Warren, G. J. (1982). *Gene* **19**, 147-151.

Hagan, C. E. and Warren, G. J. (1983). *Gene* **24**, 317-326.

Han, S., Udvardy, A. and Schedl, P. (1985). *J. Mol. Biol.* **183**, 13-29.

Haniford, D. B. and Pulleyblank, D. E. (1985). *Nucl. Acids Res.* **13**, 4343-4363.

Harding, N. E. and Ito, J. (1980). *Virol.* **104**, 323-338.

Hobom, G., Grosschedl, R., Lusky, M., Scherer, G., Schwarz, E. and Kossel, H. (1979). *Cold Spring Harbor Symp. Quant. Biol.* **43**, 165-178.

Holliday, R. (1964). *Genet. Res.* **5**, 282-304.

Holowczak, J. (1982). *Curr. Topics Microbiol. Immunol.* **97**, 27-7[illegible]

Horowitz, D. S. and Wang, J. C. (1984). *J. Mol. Biol.* **173**, 75-91.

Hsu, P. L. and Landy, A. (1984). *Nature* **311**, 721-726.

Iacono-Connors, L. and Kowalski, D. (1986). *Nucl. Acids Res.* **14**, 8949-8962.

Jensch, F. and Kemper, B. (1986). *EMBO J.* **5**, 181-189.

Johnston, E. M. (1980). *Cell* **22**, 875-886.

Kallenbach, N. R., Ma, R-I., Wand, A. J, Veeneman, G. H., van Boom, J. H. and Seeman, N. C. (1983). *J. Biomol. Struct. Dynamics* **1**, 159-168.

Katzen, L. L., Cann, G. M. and Blackburn, E. H. (1981). *Cell* **24**, 313-320.

Keller, W. (1975). *P.N.A.S. USA* **72**, 2550-2554.

Kemper, B. and Brown, D. T. (1976). *J. Virol.* **18**, 1000-1015.

Kemper, B. and Garabett, M. (1981). *Eur. J. Biochem.* **115**, 123-132.

Kemper, B., Garabett, M. and Courage, U. (1981). *Eur. J. Biochem.* **115**, 133-

Kemper, B. and Janz, E. (1976). *J. Virol.* **18**, 992-999.

Kemper, B., Jensch, F., v.Depka-Prondzynski, M., Fritz, H-J., Borgmeyer, U. and Mizuuchi, K. (1985). *Cold Spring Harbor Symp. Quant. Biol.* **49**, 815-825.

Kerr, C. and Sadowski, P. D. (1975). *Virology* **65**, 281-285.

Kikuchi, Y., Hirai, K., Gunge, N. and Hishinuma, F. (1985). *EMBO J.* **4**, 1881-1886.

Kikuchi, A. and Nash, H. A. (1979). *P.N.A.S. USA* **76**, 3760-3764.

Klobutcher, L. A., Swanton, M. T., Donini, P. and Prescott, D. M. (1981). *P.N.A.S. USA* **78**, 3015-3019.

Kolodner, R., Bishop, D. K., Evans, D., Muster-Nassal, C. and Symington, L. S. (1987). *In* DNA replication and recombination. pp. 637-646. Alan R. Liss, Inc.

Lakritz, N., Foglesong, P. D., Reddy, M., Baum, S., Hurwitz, J. and Bauer, W. R. (1985). *J. Virol.* **53**, 935-943.

Larson, D. D., Spangler, E. A. and Blackburn, E. H. (1987). *Cell* **50**, 477-483.

Leach, D. and Lindsey, J. (1986). *Mol. Gen. Genet.* **204**, 322-327.

Leach, D. R. F. and Stahl, F. W. (1983). *Nature* **305**, 448-451.

Lee, F. S. and Bauer, W. R. (1985). *Nucl. Acids Res.* **13**, 1665-1682.

Lilley, D. M. J. (1981). *Nature (London)*, **292**, 380-382.

Lilley, D. M. J. (1985). *Nucl. Acids Res.* **13**, 1443-1465.

Lilley, D. M. J. (1986a). *Biochem. Soc. Trans.* **14**, 211-213.

Lilley, D. M. J. (1986b). *Biochem. Soc. Trans.* **14**, 489-493.

Lilley, D. M. J. and Hallam, L. R. (1984). *J. Mol. Biol.* **180**, 179-200.

Lilley, D. M. J. and Kemper, B. (1984). *Cell* **36**, 413-422.

Lipps, A. J., Gruissen, W. and Prescott, D. M. (1982). *P.N.A.S. USA* **79**, 2495-2499.

Luchnik, A. N., Bakayev, V. V., Yugai, A. A., Zbarsky, I. B. and Georgiev, G. P. (1985). *Nucl. Acids Res.* **13**, 1135-1149.

Lyamichev, V., Panyutin, I. and Mirkin, S. (1984). *J. Biomol. Struct. Dynamics* **2**, 291-300.

McEntee, K., Weinstock, G. M., and Lehman, I. (1979). *P.N.A.S. USA* **76**, 2615-2619.

McFadden, G. (1987). *In* Virus diseases in laboratory and captive animals. G. Darai, ed. Martinus Nijhoff Publishing, Boston (in press).

McFadden, G. and Dales, S. (1982). p. 173-190. *In* Organization and replication of viral DNA. A. S. Kaplan, ed. CRC Press, Inc., Boca Raton, Fla.

McFadden, G. and Morgan, A. R. (1982). *J. Theor. Biol.* **97**, 343-349.

Maniatis, T., Ptashne, M., Backman, K., Kleid, D., Flashman, S., Jeffrey, A. and Maurer, R. (1975). *Cell* **5**, 109-113.

Mathog, D., Hochstrasser, M., Gruenbaum, Y., Saumweber, H. and Sedat, J. (1984). *Nature* **308**, 414-421.

McClintock, B. (1941). *Genetics* **26**, 234-282.

Menzel, R. and Gellert, M. (1983). *Cell* **34**, 105-113.

Merchlinsky, M. and Moss, B. (1986). *Cell* **45**, 879-884.

Meselson, M. (1972). *J. Mol. Biol.* **71**, 795-798.

Meselson, M. S. and Radding, C. M. (1975). *P.N.A.S. USA* **72**,358-361.

Mizuuchi, K., Kemper, B., Hays, J. and Weisberg, R. A. (1982a). *Cell* **29**, 357-365.

Mizuuchi, K., Mizuuchi, M. and Gellert, M. (1982b). *J. Mol. Biol.* **156**, 229-243.

Moss, B. (1986). *In* Fundamental Virology, pp. 637-[redacted]. N. Fields and D. M. Knipe, eds. Raven Press, New York.

Moss, B., Winters, E. and Jones, E. V. (1983). *In* Mechanism of DNA replication and recombination, pp. 449-461. N.R. Cozzarelli, ed. Alan R. Liss Inc., New York.

Moyer, R. N. and Graves, R. L. (1981). *Cell* **27**, 391-401.

Naylor, L. H., Lilley, D. M. J. and van de Sande, J. H. (1986). *EMBO J.* **5**, 2407-2413.

Nickol, J. and Martin, R. (1983). *P.N.A.S. USA* **80**, 4669-4673.

Nobile, C., Nickol, J. and Martin, R. (1986). *Mol. Cell. Biol.* **6**, 2916-2922.

Paetkau, V., Langman, L., Bradley, R., Scraba, D. and Miller, R. C., Jr. (1977). *J. Virol.* **22**, 130-141.

Patel, D. J., Kozlowski, S. A., Hare, D. R., Reid, B., Ikuta, S., Landes, N. and Itakura, K. (1985). *Biochemistry* **12**, 926-935.

Panayotatos, N. and Wells, R. D. (1981). *Nature* **289**, 466-470.

Panyutin, I., Klishko, V. and Lyamichev, V. (1984). *J. Biomol. Struct. Dynamics* **1**, 1311-1323.

Peck, L. J. and Wang, J. C. (1985). *Cell* **40**, 129-137.

Pluta, A. F., Kaine, B. P. and Spear, B. B. (1982). *Nucl. Acids Res.* **10**, 8145-8154.

Pritchard, A. D. and Cummings, D. J. (1981). *P.N.A.S. USA* **78**, 7341-7345.

Pulleyblank, D. E., Haniford, D. B. and Morgan, A. R. (1985). *Cell* **42**, 271-280.

Pulleyblank, D. E., Shure, M., Tang, D., Vinograd, J. and Vosberg, H-P. (1975). *P.N.A.S. USA* **78**, 4280-4284.

Rekosh, D. M. K., Russell, W. C., Bellett, A. J. D., Robinson, A. J. (1977). *Cell* **11**, 283-295.

Renz, M. and Day, L. A. (1976). *Biochemistry* **15**, 3220-3228.

Riggs, A. D., Lin, S. N. and Wells, R. D. (1972). *P.N.A.S. USA* **69**, 761-764.

Robinson, B. H. and Seeman, N. (1987). *Biophys. J.* **51**, 611-626.

Rosenberg, M. and Court, D. (1979). *Ann. Rev. Genetics* **13**, 319-353.

Rouviere,-Yaniv, J. and Gross, F. (1975). *P.N.A.S. USA* **72**, 3428-3432.

Ryoji, M. and Worcel, A. (1984). *Cell* **37**, 21-32.

Sadowski, P. (1971). *J. Biol. Chem.* **246**, 209-216.

Seeman, N. C. (1981). *In* Biomolecular Stereodynamics, pp. 279-300. R. H. Sharma, ed. Adenine Press, New York.

Shampay, Y., Szostak, J. W. and Blackburn, E. H. (1984). *Nature* **310**, 154-157.

Shope, R. E. (1932). *J. Exp. Med.* **56**, 793-802.

Shurvington, C. E., Stahl, M. H. and Stahl, F. N. (1987). *P.N.A.S. USA* **84**, 1624-1628.

Sigal, N. and Alberts, B. (1972). *J. Mol. Biol.* **71**, 789-793.

Sinden, R. R., Broyles, S. S. and Pettijohn, D. E. (1983). *P.N.A.S. USA* **80**, 1797-1801.

Sinden, R. R., Carlson, J. O. and Pettijohn, D. E. (1980). *Cell* **21**, 773-783.

Sinden, R. R. and Pettijohn, D. E. (1984). *J. Biol. Chem.* **259**, 6593-6600.

Steinhilber, W. and Lipps, H. J. (1986). *FEBS Letters* **206**, 25-28.

Stratling, W., Krause, E. and Knippers, R. (1983). *Virology* **51**, 109-119.

Sullivan, K. M. and Lilley, D. M. J. (1986). *Cell* **47**, 817-827.

Symington, L. S. and Kolodner, R. (1985). *P.N.A.S. USA* **82**, 7247-7251.

Szostak, J. W. (1983). *Cold Spring Harbor Symp. Quant. Biol.* **47**, 1187-1194.

Szostak, J. W. and Blackburn, E. H. (1982). *Cell* **29**, 245-255.

Szostak, J. W., Orr-Weaver, T. L. and Rothstein, R. J. (1983). *Cell* **33**, 25-35.

Taylor, A. F. and Smith, G. R. (1980). *J. Mol. Biol.* **185**, 431-443.

Thompson, B. J., Camien, M. N. and Warner, R. C. (1976). *P.N.A.S. USA* **73**, 2299-2303.

Tsujimoto, Y. and Ogawa, H. (1978). *J. Mol. Biol.* **125**, 255-273.

Vinograd, J., Lebowitz, J. and Watson, R. (1968). *J. Mol. Biol.* **33**, 173-197.

Walter, G., Zillig, W., Palm, P. and Fuchs, E. (1967). *Eur. J. Biochem.* **3**, 194-201.

Wang, A. H-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. and Rich, A. (1979). *Nature (London)* **282**, 680-686.

Wang, J. C. (1985). *Ann. Rev. Biochem.* **54**, 665-697.

Wang, J. C., Peck, L. J. and Becherer, K. (1983). *Cold Spring Harbor Symp. Quant. Biol.* **47**, 85-91.

Warner, R. C., Fishel, R. and Wheeler, F. (1979). *Cold Spring Harbor Symp. Quant. Biol.* **43**, 957-968.

Warren, G. J. and Green, R. L. (1985). *J. Bact.* **161**, 1103-1111.

Watson, J. D. (1972). *Nature (London) New Biol.* **239**, 197-201.

Weintraub, H. (1983). *Cell* **32**, 1191-1203.

Weller, S. K., Spadaro, A., Schaffer, J. E. Murray, A. W., Maxam, A. M. and Schaffer, P. A. (1985). *Mol. Cell. Biol.* **5**, 930-942.

West, S. C. and Korner, A. (1985). *P.N.A.S. USA* **82**, 6445-6449.

White, J. H. (1969). *Amer. J. Math.* **91**, 693.

Wills, A., DeLange, A. M., Gregson, C., Macaulay, C. and McFadden, G. (1983). *Virology* **130**, 403-414.

Wilson, J. (1979). *P.N.A.S. USA* **76**, 3641-3645.

Wittek, R. (1982). *Experimentia* **38**, 285-310.

Yoshio, M. (1980). *Nucl. Acids Symp. Ser.* **8**, 61.

Zannis-Hadjopoulos, M., Kaufmann, G. and Martin, R. G. (1984). *J. Mol. Biol.* **179**, 577-586.

II. CRUCIFORM EXTRUSION IN PLASMIDS BEARING THE REPLICATIVE INTERMEDIATE CONFIGURATION OF A POXVIRUS TELOMERE [1]

A. Introduction

The molecular ends of eukaryotic linear chromosomes are distinguished by the presence of special terminal structures called telomeres. In lower eukaryotes, tandem repeats of specific, short DNA sequences appear to be conserved elements of telomeric regions. Specific proteins are known to associate with these sequences (Berman *et al.*, 1986; Gottschling and Zakian, 1986), forming non-nucleosomal telomere complexes which presumably are required to stabilize the chromosomal ends and to facilitate the complete replication of these termini (reviewed by Blackburn, 1984; Blackburn and Szostak, 1984). The physical attributes of telomere structure which bear directly on the capacity to maintain the integrity and metabolism of linear chromosomes are not yet known. Terminal hairpin structures are genomic features of a variety of eukaryotic DNAs, including *Tetrahymena* rDNA (Blackman and Gall, 1978), *Paramecium* mtDNA (Pritchard and Cummings, 1981), linear plasmids of yeast (Kikuchi *et al.*, 1985) and the linear viral genomes of parvoviruses (Berns and Hauswirth, 1982), iridoviruses (González *et al.*, 1986) and poxviruses (Geshelin and Berns, 1974; Baroudy *et al.*, 1982; DeLange *et al.*, 1986). Vaccinia virus is the prototype virus of the poxvirus family of linear, double-stranded DNA viruses. Telomere involvement in poxviral DNA

1. A version of this chapter has been accepted for publication. Dickie, P., Morgan, A. R. and McFadden, G. 1987. J. Mol. Biol.

replication is particularly amenable to study because poxviruses replicate cytoplasmically in segregated structures called virosomes (Dales and Pogo, 1981). Significantly, poxviruses share with several of these biological systems a common DNA replication intermediate in which the hairpin telomeric sequences become transiently linked in an inverted repeat arrangement (Moyer and Graves, 1981; Pritchard and Cummings, 1981; Moss *et al.*, 1983; Szostak, 1983; DeLange and McFadden, 1986). Consequently, it represents a valuable model system to investigate mechanisms of eukaryotic telomere replication.

Several models have been proposed which attempt to define the molecular events associated with telomere replication. A general feature of a number of related models is the resolution of the covalently-linked telomeres (arising as intermediates in replication) by the action of a site-specific endonuclease that recognizes a telomeric sequence in the normal linear duplex DNA form (Cavalier-Smith, 1974; Bateman, 1975; Dancis and Holmquist, 1979). Since the covalently-linked telomeres generate a large inverted repeat, an alternative model, described by McFadden and Morgan (1982) and Szostak (1983), was proposed suggesting that cruciform extrusion within these sequences may mediate the process of resolution to daughter telomeres. Resolving endonucleolytic cleavage, in this event, would be the property of a structure (cruciform)-specific enzyme and not an enzyme, necessarily,with any sequence specificity. Duplex DNA crossovers formed at cruciform junctions are structurally equivalent to Holliday junctions, thus enzymatic activities normally associated with recombination (Mizuuchi *et al.*, 1982a and 1982b) could effect telomere resolution from cruciform intermediates, should these be generated *in vivo*.

The extrusion of cruciform structures from inverted repeat sequences (cruciformation) was originally demonstrated by Gellert *et al.* (1979). Due to the higher free energy state of cruciform DNA relative to normal lineform DNA, the conversion is driven *in vitro* by the torsional energy of negatively-supercoiled DNA. The specific cleavage of cruciform crossovers is one property of a class of recently described endonucleases (Lilley and Kemper, 1984; Kemper *et al.*, 1984; deMassy *et al.*, 1985). Enzymes like T4 endonuclease VII and T7 endonuclease I cleave very near the base of cruciform crossovers, generating linear nicked molecules with hairpin termini from supercoiled plasmid molecules carrying palindromic sequences. The feasibility of cruciformation as an intermediate in the resolution of telomere structures would be strengthened by the demonstration of a similar enzyme activity associated with poxviruses or the cytoplasmic virosome. Recently, a candidate DNase activity from vaccinia virus has been identified which generated a hairpin linear duplex product from a highly negatively supercoiled plasmid substrate (Lakritz *et al.*, 1985). It is the subject of the present report to analyze the facility with which cruciformation occurs in poxviral telomere sequences *in vitro*. For this purpose, the lineform to cruciform transition within the cloned telomere inverted repeat sequences of the poxvirus Shope fibroma virus (SFV) has been studied.

The linear DNA genome of SFV, a tumorigenic poxvirus, is 160kb in length (Wills *et al.*, 1983; DeLange *et al.*, 1984; Cabirac *et al.*, 1985). It has terminal inverted repeats of approximately 12.4 kb in length from which the viral telomeres have been cloned, first in yeast and later in recombination-deficient *E. coli*, and sequenced (DeLange *et al.*, 1986). Like vaccinia virus telomeres (Baroudy *et al.*, 1982), the telomeric hairpin sequences of SFV are enriched in A-T basepairs and are incompletely base paired in the region adjacent to the hairpin loop. The cloning protocol employed led to the creation of circular plasmid

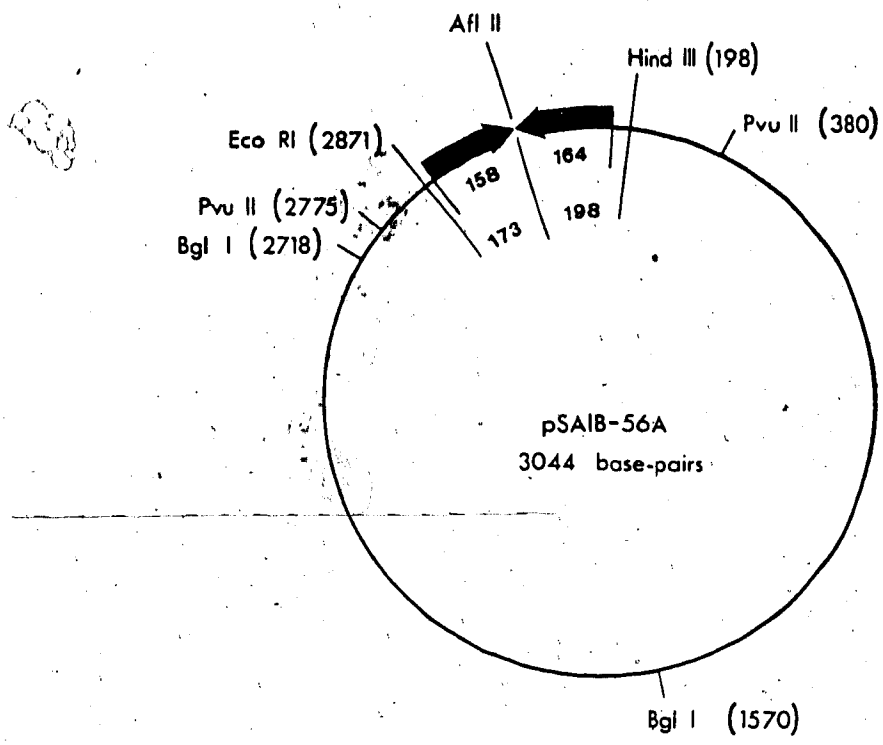


Figure II.1. Plasmid pSAlB.56A. The telomere sequences of Shope fibroma virus were cloned into pUC13 in the form of a 322 base-pair imperfect inverted repeat (heavy arrows). Distances indicated in the brackets are measured in base pairs, clockwise from the *Afl II* site at the axis of the inverted repeat. The numbers beneath the insert sequences refer to distances, in basepairs, between the extended bars.

molecules in which the telomere sequences are propagated in the inverted repeat arrangement. Presumably, neither the yeast nor the bacterial replicative machinery recognized the viral intermediate and left it unresolved. Visualization by electron microscopy of the recombinant plasmid DNA, either as the circular yeast plasmid pYSF-1 containing a 3.8kb telomeric inverted repeat insert or the bacterial plasmid pSAlB.56A with a 322 base-pair inverted repeat insert (Fig. II.1, above), revealed the presence of a

cruciform-like structure (DeLange *et al.*, 1986). Preliminary tests, such as the sensitivity of the viral insert sequences to cleavage by bacteriophage T7 endonuclease I and *Afl*II (the *Afl*II restriction site lies at the central axis of the inverted repeat and exists as a double-stranded site only when the plasmid is in the lineform state) indicated that the cloned viral sequences were extruded into a cruciform. Fine mapping of the cruciform present in the recombinant plasmid pSAlB.56A, as described here, verifies the existence of the cruciform in the viral insert *in vitro*.

Plasmid pSAlB.56A, and various deleted forms of it, have been analyzed in a 2-dimensional agarose gel system (Wang *et al.*, 1983) to determine the energy parameters governing the respective lineform to cruciform transitions. Observations have allowed, for the first time, an estimation of the influence that particular elements of telomere structure, particularly the presence of unpaired nucleotides, have on the generation of cruciforms. The SFV cloned inverted repeat is one of the largest inverted repeats to be studied in this manner. Smaller inverted repeats have been extensively studied (Courey and Wang, 1983; Lilley and Hallam, 1984; Panyutin *et al.*, 1984). Generally, the smaller palindromes have been shown to be extruded into cruciforms with a moderate energy of formation, but a prohibitively high energy of activation is expected to forbid it from occurring *in vivo* (Courey and Wang, 1983; Gellert *et al.*, 1983). Conversely, observations reported here indicate that the conversion of the cloned SFV telomere inverted repeat to the cruciform is characterized by a high energy of formation, reflected in the high negative superhelical densities required for extrusion, and an unusually low energy of activation. The absence of a profound kinetic barrier increases the likelihood that significant cruciformation within these sequences could occur spontaneously *in vivo* at the appropriate superhelical density, or as a consequence of the specific interaction with

putative telomere specific binding proteins.

B. Materials and Methods

Enzymes and reaction conditions Restriction enzymes were purchased from Boehringer Mannheim Biochemicals, Pharmacia Inc., Bethesda Research Laboratories and Amersham Corp. T4 DNA ligase, DNA gyrase and T4 DNA polymerase were obtained from Bethesda Research Laboratories; calf intestinal alkaline phosphatase and DNase I from Boehringer Mannheim and nuclease S1 from Miles Laboratories. Eukaryotic topoisomerase I (Sahai and Kaplan, 1986) and bacteriophage T7 endonuclease I (Sadowski, 1971) were generously donated by Dr. Beni Sahai and Dr. Paul Sadowski, respectively.

DNase I, in conjunction with T4 DNA ligase, was used to prepare relaxed, covalently closed lineform plasmid molecules from relaxed, cruciform-containing molecules that were relatively insensitive to the topoisomerase I. Topoisomerase I treatment was routinely used to reduce the native titratable superhelical density of these molecules (see later). Limited DNase I reactions were carried out on ice as described by DeLange *et al.* (1986) until approximately 80% of the molecules were converted from form I DNA (covalently closed and circular) to form II DNA (nicked and circular) as determined fluorimetrically (Morgan *et al.*, 1979). After the ligation of this material to relaxed form I DNA and deproteinization by phenol extraction, the sample was dialyzed against 50mM Tris.HCl (pH8) and 1mM EDTA (Tris/EDTA buffer).

Nuclease S1 digestions of cruciform-containing plasmid DNA were performed as

described by Greaves *et al.* (1985). Reactions performed to fine map nuclease S1 cleavage positions with respect to 5' end-labelled restriction fragments, were incubated on ice until approximately 50% of the molecules were nicked as determined fluorimetrically. Samples of nicked DNA were deproteinized, then digested further with selected restriction enzymes prior to labelling. Reactions designed to convert form I DNA to form III DNA (linear), for an approximate determination of the S1 sensitive regions, were allowed to proceed to completion.

Bacteriophage T7 endonuclease I reactions were performed as described by de Massy *et al.* (1985), except 4mM spermidine was added to the basic reaction buffer: 50mM Tris.HCl (pH8), 10mM $MgCl_2$, 1mM dithiothreitol and 50µg bovine serum albumin. Reactions were incubated at 37 °C until 50 to 80% of the covalently closed plasmid molecules was converted to a nicked form (approximately 15 minutes). Enzyme was diluted 10-fold into 50% glycerol, 0.2 mM EDTA and 1mM dithiothreitol prior to use. For each 10µl of reaction, which contained approximately 1µg of plasmid DNA, 1µl of diluted enzyme was used. The T7 endonuclease I used was 95% pure based on estimations from SDS-polyacrylamide gel electrophoresis (P. Sadowski, personal communication).

Relaxation of plasmid DNA with eukaryotic topoisomerase I was carried out in 50mM Tris.HCl (pH8), 100mM NaCl, 100mM KCl and 0.1mM EDTA. In 10µl reactions, 1µg of DNA was treated with 0.3µl of titrated enzyme (in 50% glycerol) at 6°C. Limited relaxations yielding broad topoisomer distributions were achieved with 15 to 30 minute incubations under these conditions.

The preparation of topoisomer distributions from relaxed, lineform plasmid DNA required partial reactions with DNA gyrase. Normally, this was accomplished by incubating reactions at room temperature for 5 minutes. Otherwise, the conditions for the DNA gyrase reactions, and all the remaining enzyme reactions, were in accordance with the directions supplied by the manufacturer. Reactions were stopped by deproteinization or the addition of gel electrophoresis loading buffer; 2% (w/v) SDS, 50% glycerol and 0.1% bromophenol blue in Tris/EDTA buffer (2µl per 10µl reaction), for native gels and formamide/dye solutions containing 0.5% (w/v) bromophenol blue and xylene cyanol and 20mM EDTA (10µl per 5µl reaction), for denaturing gels.

The 5' end-labelling of DNA fragments was accomplished using standard procedures (Maniatis *et al.*, 1982). Fragments were dephosphorylated with calf intestinal alkaline phosphatase, deproteinized, ethanol precipitated and labelled with [γ-^{32}P]ATP (3000Ci/mmole) using T4 polynucleotide kinase, under conditions that minimized the labelling of 5' positions at internal nick sites. Labelled fragments were purified away from unincorporated label by gel filtration in Bio Gel A-50m equilibrated in Tris/EDTA buffer containing 0.1M NaCl. Dephosphorylation using bacterial alkaline phosphatase instead of calf intestinal phosphatase resulted in the generation of specific cleavage products in control experiments performed in the absence of T7 endonuclease I. Apparently commercial preparations of the enzyme contained contaminating endonucleases which, interestingly, nicked the DNA in the region of the extra-helical bases in the terminal hairpin sequences. No analogous nicking was observed with calf intestinal phosphatase.

Bacterial strains and plasmids All recombinant plasmids were derived from pUC13

(Messing, 1983) and maintained in *Escherichia coli* DB1256 (*recA recB21 rec22 sbcB15 hsdR F- proA2 his4 thi-1 argE3 lacY1 galK2 ara-14 xyl-5 mtl-1 str-31 tsx-33*) or *Escherichia coli* JM83 (*ara Δlac-pro thi strA Ø80 dlacZ ΔM15*). The former strain was constructed in F. Stahl's laboratory and was obtained from A. Wyman. The transformation of these strains was accomplished using a standard calcium chloride/rubidium chloride method (Maniatis *et al.*, 1982).

Plasmids pSAlB.56A, shown in Figure II.1, and its 42 base-pair deletion derivative, pSD19, have been described (DeLange *et al.*, 1986). pSAlB.56A contains a 322 base-pair insert derived from the telomere sequences of Shope fibroma virus arranged as an imperfect palindrome. The viral insert was cloned into the *Sma*Isite of pUC13. Plasmid pSD19 contains a 280 base-pair insert, as it was derived by deleting 42 bases from one end of the pSAlB.56A insert (see Figure II.8). The creation of the remaining deletion derivatives of pSD19 cited in this report is described in detail elsewhere (DeLange and McFadden, 1987) and is only briefly outlined here. Plasmid pSD19 deletion D21 (pD21) was derived from pSD19 by the further deletion of 41 base pairs from the end of the viral insert opposite the original deletion in pSD19. Plasmids pSD19 D5.16, D6.3 and D5.10 were derived from pSD19 by digesting plasmid DNA with *Afl*III,which cuts at the axis of the cloned viral inverted repeat, followed by limited bidirectional digestions with a combination of exonuclease III plus exonuclease VII. The DNA was blunt end ligated, transformed into *E. coli* DB1256 and individual clones were selected and sequenced. In this manner, bidirectional central axis deletions of pSD19 were obtained. The product viral insert sequences are presented in Figure II.8. The plasmid pSAD-2 was derived from pSD19 D5.10 following the digestion of this latter plasmid with *Hpa*I (cuts at the axis of the cloned inverted repeat) and insertion of a

synthetic oligonucleotide (22 nucleotides in length) by blunt end ligation. The sequence of this 22 residue oligonucleotide, which restores 11 base pairs of viral sequence on either side of the inverted repeat axis, is presented in Figure II.9. Derivatives of pSAlB.56A carrying only one arm of the SFV telomeric inverted repeat were obtained following double digestions with *Afl*II and either *Eco*RI or *Hin*dIII. Religation of the vector and transformation produced the desired plasmids carrying either the long arm (right repeat) of the inverted repeat as shown in Figures II.1, or the short arm (left repeat). An additional plasmid, pSAlB.56A Δ260, appeared naturally in pSAlB.56A plasmid preparations. The Δ260 clone is a recombination product of pSAlB.56A, derived *in vivo* and which represents at least 10% of the DNA in pSAlB.56A preparations from recombination-deficient *E. coli*. It contains a 62 base-pair viral insert and is described in more detail in the Results (this chapter). The plasmid pAT34 (Haniford and Pulleyblank, 1985) was obtained from Dr. David Pulleyblank.

Plasmid DNA was purified following the alkaline extraction of extrachromosomal DNA from bacteria using a scaled-up version of the method of Birnboim and Doly (1979). Plasmid DNA was banded in cesium chloride/ethidium bromide gradients, butanol extracted and precipitated. Further purification of DNA, when required, was achieved following electrophoresis in low melting point agarose as described by Langridge *et al.* (1980). Insert sequences were verified by sequencing with the dideoxy chain-termination method (Sanger *et al.*, 1977) using a pUC specific oligonucleotide primer.

Gel Electrophoresis One-dimensional, native gel electrophoresis was performed in 1% (w/v) agarose gels at room temperature at 2V/cm or in 5% (w/v) polyacrylamide gels at room temperature and 6V/cm. The running buffer was 90mM Tris.HCl (pH8.3), 90mM

boric acid and 10mM EDTA (Tris/borate buffer). Gels were stained briefly with 0.5μg/ml ethidium bromide in Tris/borate buffer, destained in distilled water and photographed. Denaturing gels were standard 7M urea/10% (w/v) polyacrylamide sequencing gels electrophoresed in Tris/borate buffer (Maxam and Gilbert, 1980). Gels were dried onto Whatman #3 paper and autoradiographed at -70°C. Autoradiograms were scanned with a Joyce Loebl Chromoscan 3 densitometer.

Two-dimensional gel electrophoresis was performed in 20cm x 25cm x 0.5cm 1% (w/v) agarose gels (Wang *et al.*, 1983). Sample DNAs containing a broad spectrum of topoisomers (1 μg) were loaded into 0.15cm (d) x 0.4cm (h) circular wells and run in the first dimension in Tris/borate buffer, either at room temperature or 37°C at 2.5V/cm for 26 hours. A sample of plasmid DNA known to contain a mixture of form I and form II DNA was co-electrophoresed as a control. Afterwards, the control lane was stained with ethidium bromide and the limits of the DNA distribution localized and marked. The corresponding region of the gel for each of the experimental plasmids was then excised as a 0.8cm x 10cm gel slab. These were aligned in a second gel such that they were perpendicular to electrophoresis in the second dimension and were embedded in fresh agarose containing 1.28μM chloroquine diphosphate. Electrophoresis in Tris/borate buffer containing 1.28μM chloroquine diphosphate was carried out at room temperature, at 3V/cm for 24 hours. Upon completion, the gels were stained in 1μg/ml ethidium bromide for one hour, and destained in distilled water overnight before being photographed.

Analysis of topoisomer distributions The minimum energy of formation associated with torsion-induced duplex structural transitions in DNA can be derived from the change in

twist (ΔTw) observed for the transition and the critical linking difference (ΔLk) at the threshold of the stable transition. In certain cases, such as those described in this paper, where cruciform extrusion effects the complete relaxation of the DNA molecule (the ΔLk in the circular domain of the plasmid DNA is reduced to zero), the free energy of formation is equivalent to the energy of supercoiling at the transition threshold. The ΔLk of this topoisomer is related to the free energy of supercoiling (ΔG_S) by the following equation (Horowitz and Wang, 1984; Depew and Wang, 1975; Pulleyblank *et al.*, 1975):

$$\Delta G_S = \frac{1100\,RT}{N} (\Delta Lk)^2 ,$$

where R is the gas constant, T is the absolute temperature and N is the molecular size of the DNA in base pairs. Topoisomers were numbered (and ΔLk determined) by the band counting method (Keller, 1975). The slowest migrating band in the first dimension was designated topoisomer 0 (ΔLk ~0), and the remaining topoisomers were numbered consecutively with positive or negative integers (depending on the sign of the writhe), relative to increasing mobilities in the gel (see Fig. 10E, for example). Fractional values of ΔLk, used in the above equation, were estimated from differences in the relative mobilities of several topoisomer pairs to the nearest tenth of a unit (Lee and Bauer, 1985).

C. Results

Cruciformation in plasmid pSAlB.56A. The telomere sequences of Shope fibroma virus cloned into pUC13 comprise a 322 base-pair imperfect inverted repeat within the plasmid pSAlB.56A (Fig. II.1). The precise sequence of the insert is given in Figure II.2, where the eight non-palindromic bases characteristic of SFV are featured in boxes (Delange *et al.*,

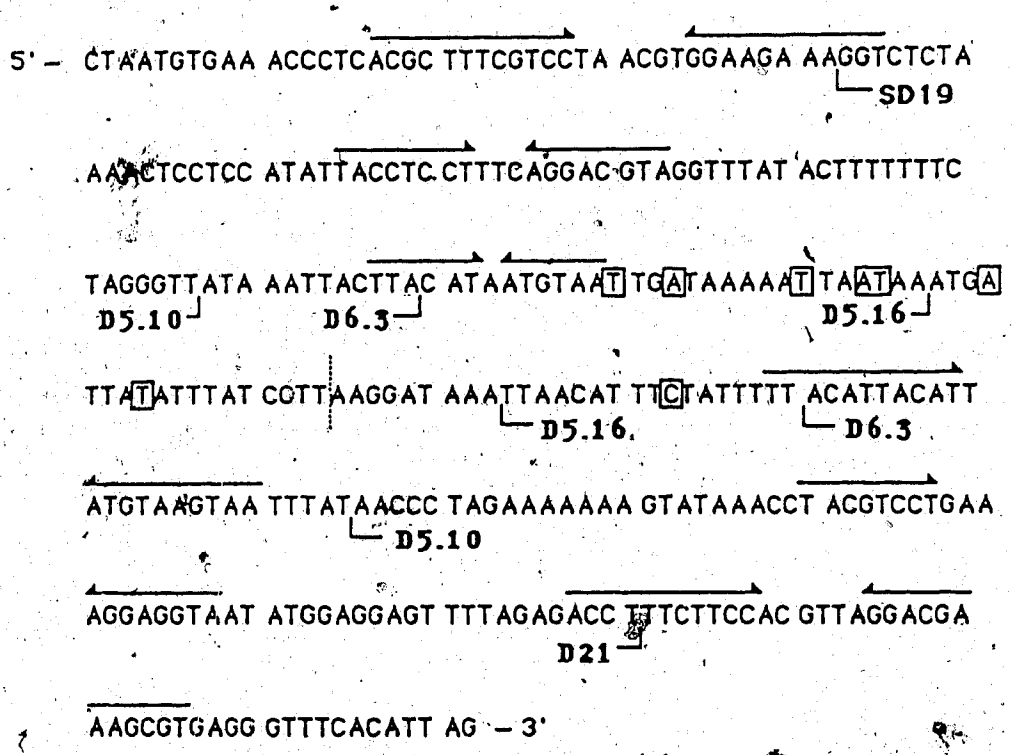


Figure II.2. Sequence of the cloned telomere sequences in pSA1B.56A. The complete viral insert is given, and the associated deletion derivatives are indicated by the subscripted notations. The sequences deleted from pSA1B.56A to form D5.10, D6.3 and D5.16 lie between the subscripts and the central axis of the viral inverted repeat (dotted line). The sequences deleted to form pSD19 and D21 lie between the subscript and the 5' and 3' extremities, respectively. Small inverted repeats within the viral sequence are overlined. Those bases which are extrahelical in the telomere, or in the cruciform stems, are boxed.

al., 1986). Electron micrographs of the purified plasmid have revealed a large cruciform, presumably derived from the extrusion of the viral telomere sequences (Delange *et al.*, 1986). Crude mapping of the extruded sequences as determined by the relative sensitivity of the plasmid to bacteriophage T7 endonuclease I and the restriction endonuclease *Afl* II (the *Afl*II recognition site, CTTAAG, lies at the axis of the viral inverted repeat) placed

the cruciform within the viral insert of the plasmid. The inverted repeat derived from the viral telomere sequences is sufficiently large that cruciform extrusion limited to this region would topologically relax the supercoiled plasmid molecule. Indeed, when the plasmid was purified from *E. coli*, it co-migrates in agarose gels with relaxed plasmid DNA (Fig. II.3, lane a). Two faster migrating DNA species were resolved by these gels. The slower of the two could be shown to be the supercoiled lineform of the plasmid DNA and the faster species was a specific 260 base-pair deletion product of pSAlB.56A (Δ260) which retained a short inverted repeat insert of 62 base-pairs. The latter form represented 10% of the DNA in plasmid preparations and has been sequenced (described in more detail later).

The relaxed cruciform-bearing pSAlB.56A plasmid DNA was purified by preparative gel electrophoresis (lane c, Fig. II.3) and was seen to co-migrate with relaxed lineform plasmid DNA that was prepared by nicking the DNA with DNase I and religating (Fig. II.3, lane b). A characteristic Gaussian distribution of topoisomers in the preparation of relaxed, covalently-closed, lineform plasmid DNA is apparent in Figure II.3, lane b. Plasmid that is topologically relaxed by virtue of cruciform extrusion forms a single band in these gels because the freedom of branch migration at the cruciform junction maintained a linking difference of zero in the plasmid DNA (Fig. II.3, lane c). These two preparations of "relaxed" plasmid DNA (relaxed either by nicking and religation, or by cruciformation) were further distinguished by their different susceptibilities to various enzymes. Treatment of relaxed-lineform DNA with gyrase for thirty minutes at room temperature yielded cruciform-containing, relaxed plasmid DNA. The progression through a supercoiled lineform intermediate was demonstrated in partial reactions (not shown). The treatment of relaxed-cruciform plasmid DNA with gyrase produced a third

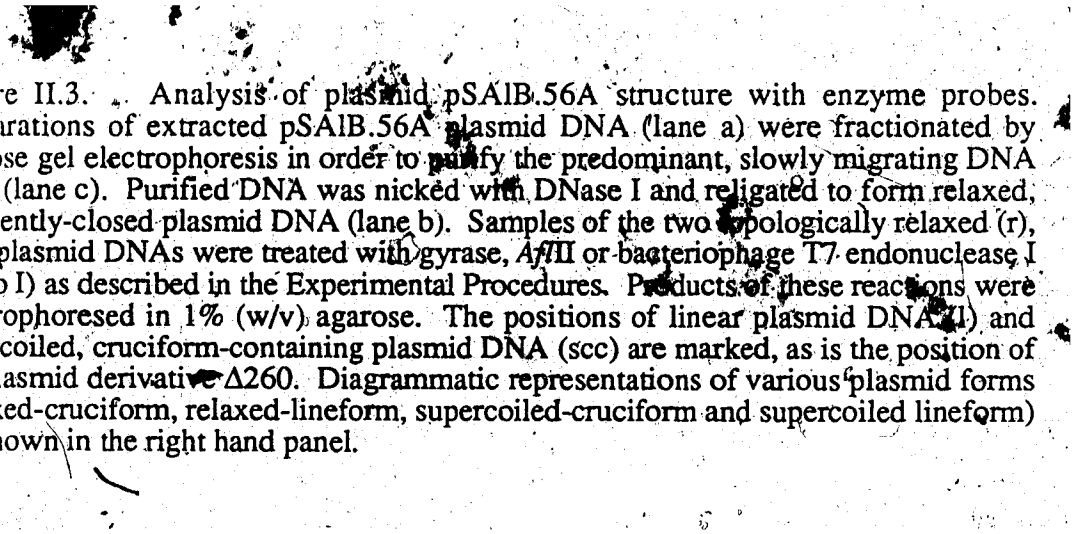
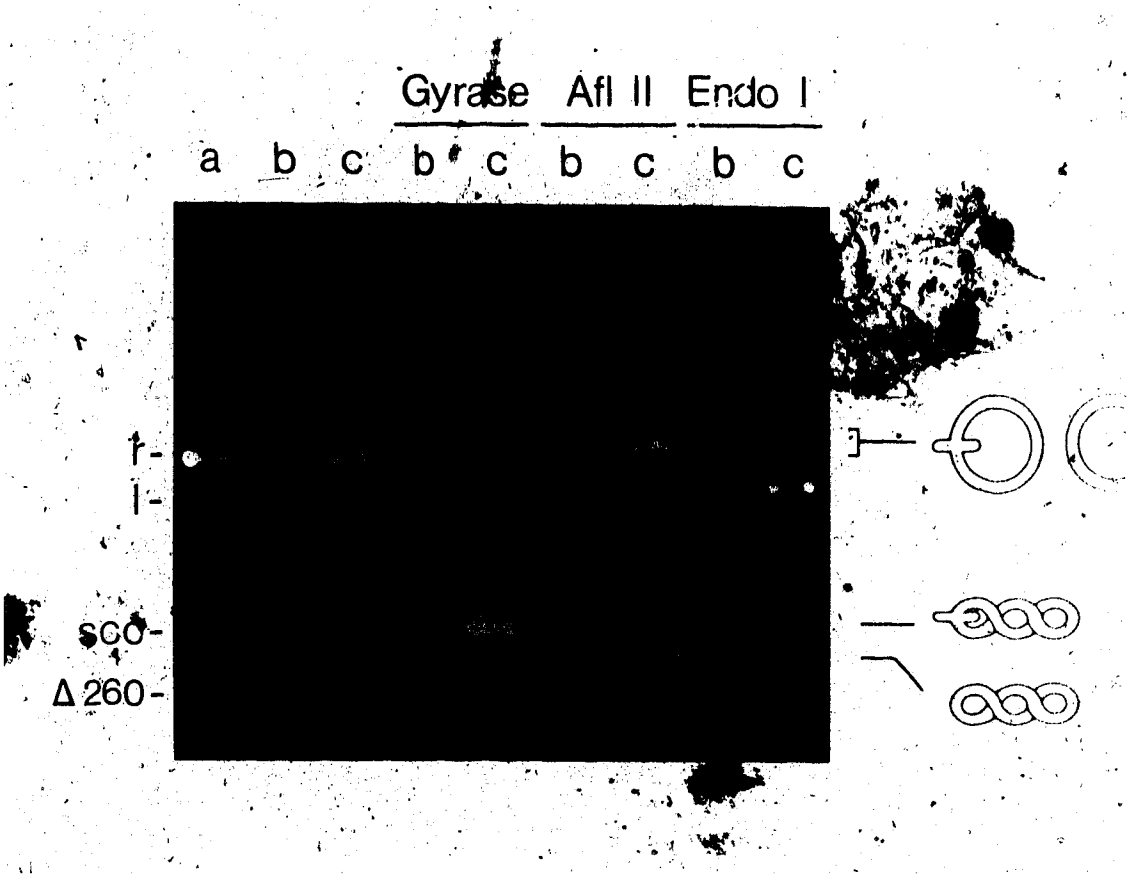


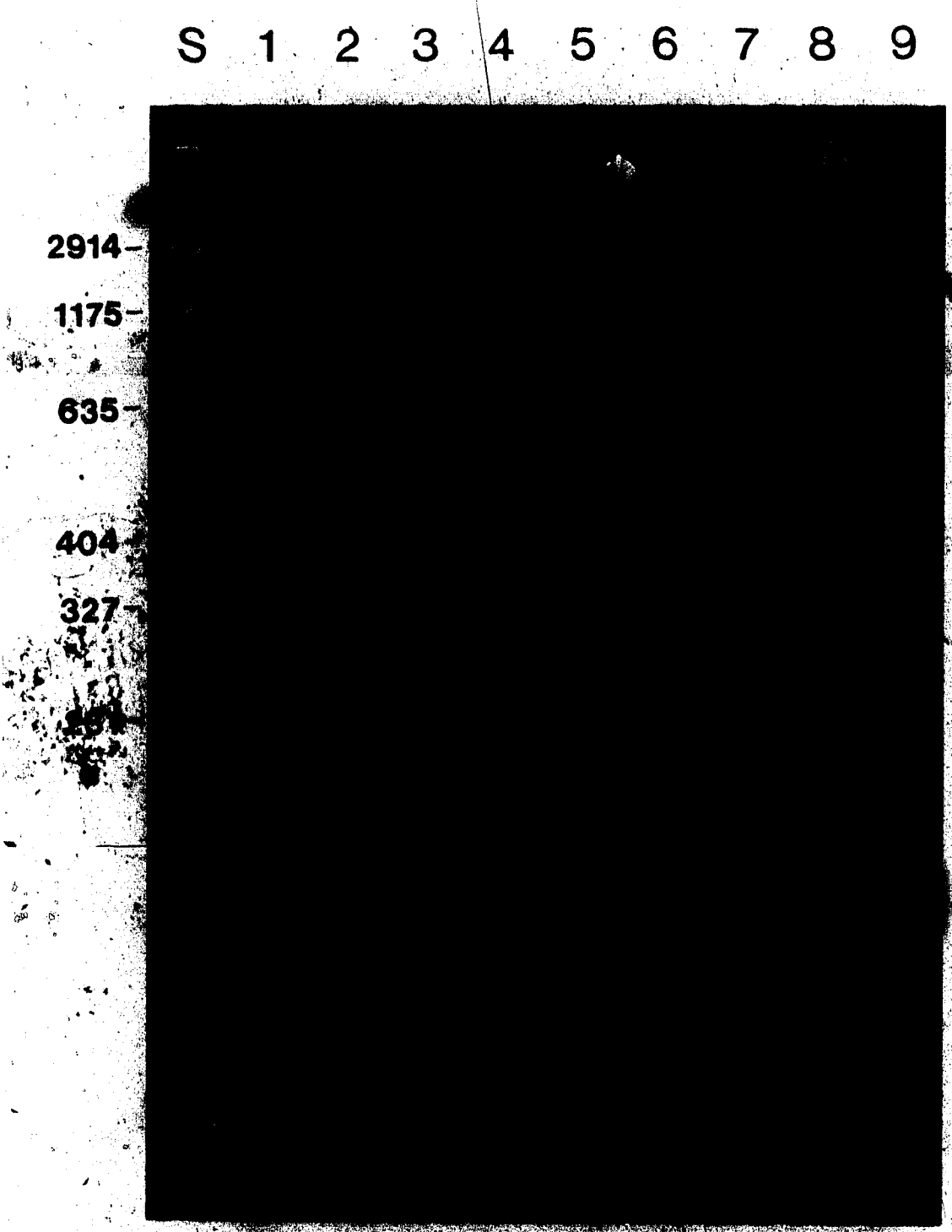
Figure II.3. Analysis of plasmid pSAlB.56A structure with enzyme probes. Preparations of extracted pSAlB.56A plasmid DNA (lane a) were fractionated by agarose gel electrophoresis in order to purify the predominant, slowly migrating DNA band (lane c). Purified DNA was nicked with DNase I and religated to form relaxed, covalently-closed plasmid DNA (lane b). Samples of the two topologically relaxed (r), pure plasmid DNAs were treated with gyrase, *Afl*II or bacteriophage T7 endonuclease I (Endo I) as described in the Experimental Procedures. Products of these reactions were electrophoresed in 1% (w/v) agarose. The positions of linear plasmid DNA (l) and supercoiled, cruciform-containing plasmid DNA (scc) are marked, as is the position of the plasmid derivative Δ260. Diagrammatic representations of various plasmid forms (relaxed-cruciform, relaxed-lineform, supercoiled-cruciform and supercoiled lineform) are shown in the right hand panel.


Gyrase
Afl II
Endo I
a b c b c b c b c
sco-
Δ260-

topological form of the plasmid which can be described as cruciform-containing, supercoiled plasmid DNA. In this form, the inverted repeat sequences are fully extruded and the circular domain of the plasmid has been maximally supercoiled by gyrase. It migrated slower than the supercoiled lineform, presumably because of the retarding effect of the cruciform structure (Fig. II.3).

The presence of a cruciform in the purified plasmid was further established by the plasmids differential sensitivity to T7 endonuclease I and *Afl*II digestion. Bacteriophage T7 endonuclease I, in a site-specific fashion, cleaves the duplex DNA junctions formed at the base of cruciform structures (de Massy *et al.*, 1984; Chapter III) while *Afl*IIcleaves at its recognition site which is present at the axis of symmetry of the inverted repeat in pSAlB.56A, provided that this sequence is in the lineform. The recognition site, CTTAAG, would be lost in the cruciform configuration since it would be extruded in the single-stranded loop region of the cruciform hairpins. Purified plasmid (relaxed-cruciform DNA) was linearized by T7 endonuclease I treatment but not nearly so efficiently by *Afl*II digestion (Fig. II.3, lanes c). Alternatively, the relaxed lineform was sensitive to *Afl*II digestion and largely resistent to T7 endonuclease I. These results support the conclusion that a cruciform exists within the viral insert. Further indication that the cruciform was centered about the axis of symmetry of the viral inverted repeat was provided by the experiments described in Figure II.4. Purified pSAlB.56A plasmid DNA was probed with T7 endonuclease I and nuclease S1, of which the latter would be expected to nick the DNA at or near the single-stranded hairpin regions of the cruciform (Lilley and Hallam, 1984). The ends of the linear molecules produced by these alternative methods were mapped relative to known restriction sites in the pUC13 vector. The ends of the T7 endonuclease I treated molecule, presumably hairpin termini (de Massy *et al.*,

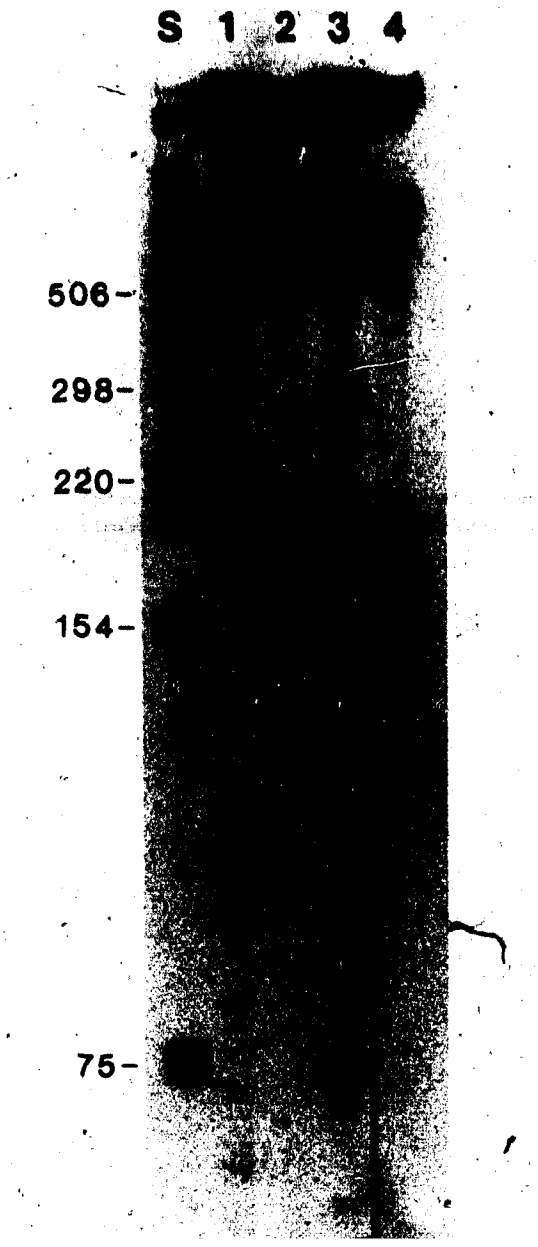
Figure II.4. Site specific cleavage of purified pSAlB.56A plasmid DNA by nuclease S1 and bacteriophage T7 endonuclease I. Cleavage sites have been mapped relative to the restriction sites for *Pvu*II and *Bgl*I (see Figure II.1). ØX174 RF-DNA digested with *Taq*I was used for size markers (S). Lane 1: digestion with *Bgl*I, alone; lane 2: T7 endonuclease I followed by *Bgl*I; lane 3: *Pvu*II digestion, alone; lane 4: T7 endonuclease I followed by *Pvu*II; lane 5: T7 endonuclease I, alone; lane 6 is the same as lane 3; lane 7 is DNA digested with nuclease S1 followed by *Pvu*II and lane 8 is DNA digested with nuclease S1, alone. Untreated DNA is in lane 9. Digestion products were electrophoresed in 5% (w/v) acrylamide.


S 1 2 3 4 5 6 7 8 9
2914
1175
635
404
327

1984), mapped within a 630 base-pair *Pvu*II restriction fragment, since digestions with both enzymes generated, from this larger *Pvu*II fragment, two sub-fragments that were approximately 370 and 270 base-pairs in length (Fig. II.4, lanes 3 and 4). Likewise, the 1.8kb *Bgl*I fragment of pSAlB.56A was cleaved by digestions with T7 endonuclease I and *Bgl*I together to approximately 1380 and 320 base-pair fragments (Fig. II.4, lanes 1 and 2). These results are consistent with the T7 endonuclease I linearization of the plasmid, yielding hairpin termini that map in the vicinity of the central axis of the inverted repeat (see Fig. II.1). Experiments with nuclease S1 produced identical gel patterns, consistent with cleavage at the the axis of the viral inverted repeat, within sequences that form the single-stranded turnaround regions of the cruciform hairpins.

The results of fine cleavage mapping of nuclease S1 and T7 endonuclease I digested pSAlB.56A are presented in Figure II.5. Purified plasmid DNA was treated first with nuclease S1 or T7 endonuclease I and subsequently digested with either *Eco*RI or *Hin*dIII. The 5' ends of the resulting fragments were labelled with [γ-^{32}P] ATP and T4 polynucleotide kinase under conditions that minimized the labelling of 5' ends at single-stranded nick positions. The cleavage positions for nuclease S1 and T7 endonuclease I were mapped relative to both restriction sites, following the analysis of labelled, single-stranded product molecules electrophoresed in sequencing gels. One such gel is presented in Figure II.5. The predominant S1 cleavage sites mapped approximately 195 base-pairs from the *Hin*dIII site and 175 base-pairs from the *Eco*RI site (Fig. II.5, lanes 2 and 4, respectively). The only other specific cleavage site, which was significantly weaker, appeared about 180 base-pairs from the *Hin*III site. The predominant cleavage sites mapped precisely to the inverted repeat axis, which would form the hairpin termini of the cruciform branches. The minor cleavage position appeared

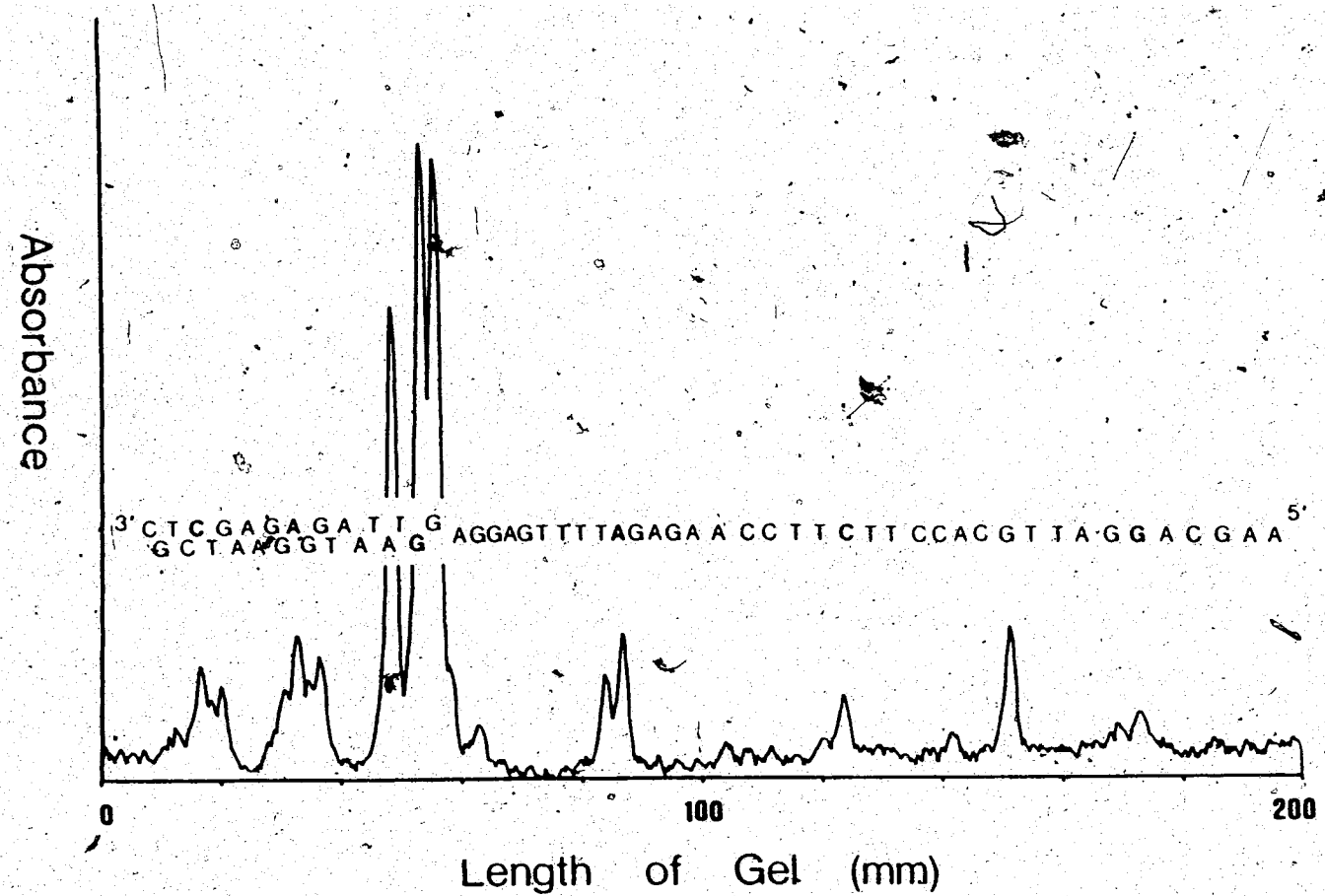
Figure II.5. Fine mapping of nuclease S1 and T7 endonuclease I cleavage positions on pSAIB.56A. Purified plasmid DNA was partially digested with either T7 endonuclease I (lanes 1 and 3) or nuclease S1 (lanes 2 and 4) followed by digestion with either *Eco RI* (lanes 3 and 4) or *Hind III* (lanes 1 and 2). Overhanging 5' ends were phosphorylated with [γ-^{32}P]ATP and T4 polynucleotide kinase and labelled fragments were detected by autoradiography at -70°C following electrophoresis under denaturing conditions in 10% (w/v) acrylamide. pBR322 plasmid DNA digested with *HinfI* was used as size markers. The bracketed portion of lane 3 in the gel was scanned by densitometry and the density profile is presented in Figure II.6.


S 1 2 3 4
506-
298-
220-
154-
75-

to fall in the region of the extrahelical bases. The single-stranded product fragments generated by T7 endonuclease I digestion formed a complex but predictable pattern. First, it is significant that the patterns generated by double digestions with T7 endonuclease I and either *Hin*dIII or *Eco*RI were identical but for the constant difference in fragment lengths of 19 bases which was the distance separating the respective restriction sites from the axis of the inverted repeat (Fig. II.5, lanes 1 and 3, respectively). This would only occur if the T7 endonuclease I cuts were symmetrically paired across either axis of the cruciform DNA crossover. Second, the periodicity of the nicks made proximal to the dyad axis was mirrored by the periodicity of nicks made distal to the dyad axis (such as on the 5' and the 3' sides of the cruciform structure). In agreement with published observations, the enzyme cleaved across either diagonal axis, presumably very near the base of the cruciform (de Massy *et al.*, 1984; Chapter III).

The periodicity of the fragments generated between *Eco*RI and T7 endonuclease I cleavage positions (bracketed region, Fig. II.5, lane 3) is emphasized by the densitometer tracing shown in Figure II.6. It is construed from the regular periodicity of the cutting pattern and the relative intensities of the bands that, generally, the position of cleavage was determined by the linking difference of the plasmid DNA molecules. The T7 endonuclease I/*Eco*RI cleavage fragments formed an approximate Boltzmann distribution which most probably was a reflection of the topoisomer composition of the purified plasmid preparations. Since T7 endonuclease I cleaves artificial Holliday junctions with considerable specificity (one nucleotide removed from the crossover point; see Chapter III), the mapped cut sites identify the various crossover positions of the cruciform junctions in the population of molecules, these crossovers being largely determined by the linking difference of individual plasmid molecules. The major cleavage sites, within

Figure II.6. Bacteriophage T7 endonuclease I cleavage positions within the viral inverted repeat of pSAlB.56A at native superhelical densities. The portion of the autoradiogram bracketed in Figure II.5, which assigns enzyme cleavage positions proximal to the *Eco RI* restriction site with respect to the cruciform structure, has been scanned by densitometry. The viral sequence corresponding to the region of cleavage is superimposed over the scan and is phased with an accuracy of ±1 nucleotide position. Nucleotides are numbered consecutively, according to the overlaying scale, starting from the inverted repeat axis. The assignment of the viral sequence to cleavage postions was assisted by repeating the experiment in the presence of marker DNA generated by cleaving ØX174 DNA with *Hin fI*. Absorbance is in arbitrary units.


Absorbance
3' C T C G A G A G A T T G AGGAGTTTAGAGAACCTTCTTCCACGTTAGGACGAA 5'
G C T A A G G T A A G
0
100
200
Length of Gel (mm)

the sequence CTCCATAT, were approximately 99 and 105 base-pairs from the inverted repeat axis (depending on the arm of the inverted repeat - see Figures II.2 and II.6). Assuming regular B DNA, the topological relaxation of a plasmid with a cruciform extruded to this position within the inverted repeat would have minimally required a superhelical density of -0.067, roughly equivalent to the absorption of about 20 supercoils. This superhelical density is typical of plasmid DNA purified from *E. coli* (Mizuuchi *et al.*, 1982; Lilley and Hallam, 1984). Because there was a unimodal distribution of cleavage sites, and it could be related to a population of topoisomers with an average superhelical density characteristic of plasmid DNA in bacterial cells, it most likely reflected the presence *in vivo* of a uniform population of plasmid DNA that persisted in the lineform. A significant steady-state proportion of cruciform-containing plasmid DNA *in vivo* would have been evidenced by a bimodal distribution of topoisomers, presuming that intracellular gyrase would act to overwind any cruciform-relaxed plasmid molecules in the manner demonstrated *in vitro*.

Additional evidence for the existence of a uniformly supercoiled population of plasmid DNA *in vivo* was obtained by the electrophoresis of purified DNA in agarose containing an amount of chloroquine diphosphate which partially relaxed the negatively supercoiled plasmid DNA molecules. Individual topoisomers were resolved in this way and only one Boltzmann distribution of molecules was observed (data not shown). It must be noted that the supercoiled, lineform plasmid DNA that could be observed in crude plasmid preparations had been removed by preparative gel electrophoresis for most of these experiments. It could be argued that it represents a second population of molecules with a lower, but distinct, average superhelical density. Results from the two-dimensional gel analysis of crude pSAlB.56A plasmid DNA (see Fig. II.10D) indicate that the supercoiled

lineform molecules simply represented the least supercoiled fraction of a single population of plasmid DNA and there was insufficient torsional energy in these molecules to stabilize the cruciform. The apparent ten nucleotide spacing between regions of T7 endonuclease I cleavage and the possible influence that sequence has on the cleavage efficiency of T7 endonuclease I are features of the enzyme that will be considered in Chapter IV.

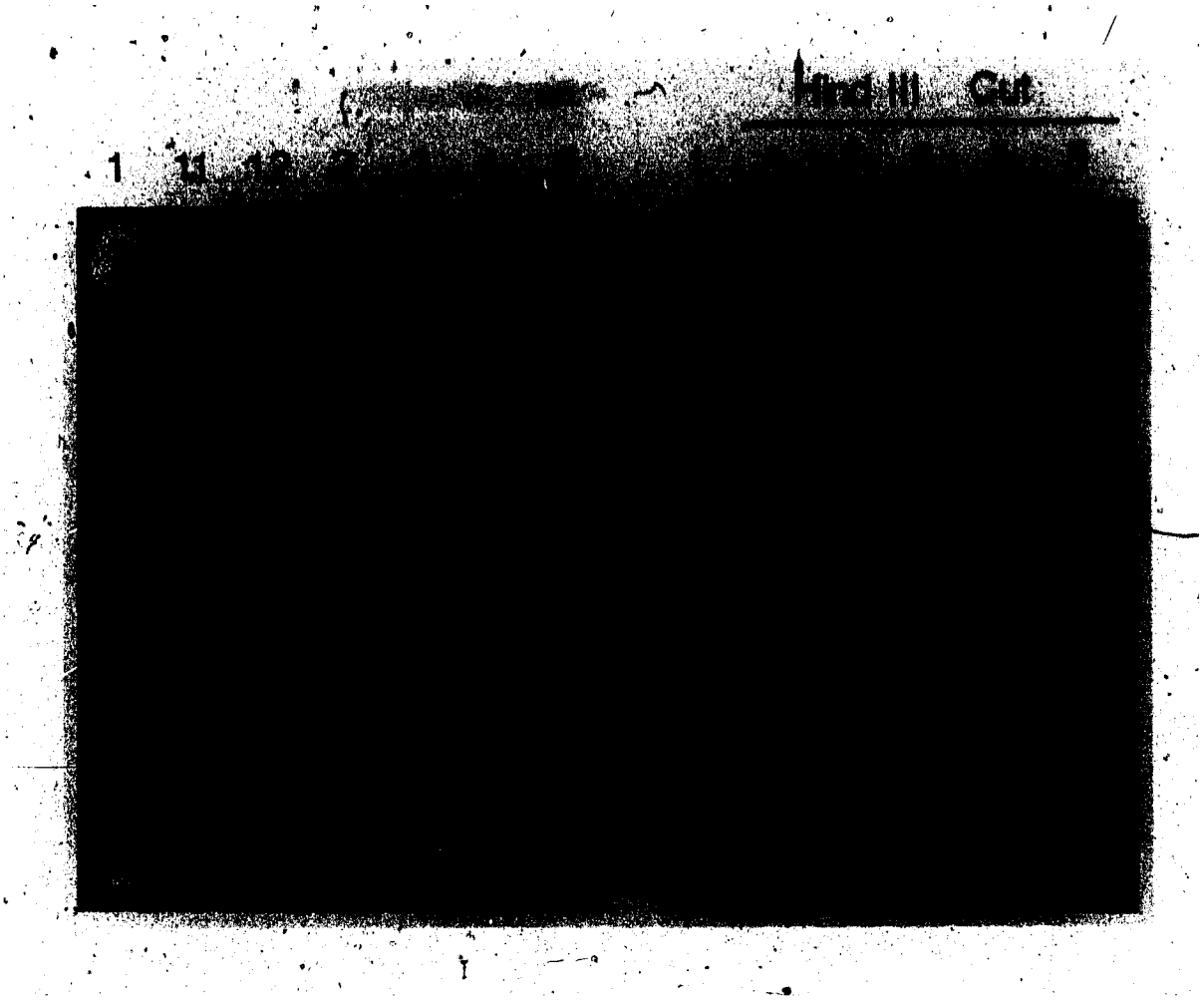
The migration of pSAlB.56A plasmid DNA in agarose gels, its sensitivity to the enzymes described and the position of the T7 endonuclease I cleavage sites, support the interpretation that the native superhelical density of the plasmid, as purified from *E. coli*, is, on average, sufficient to maintain the cloned viral sequences in a cruciform *in vitro*. Cruciformation in these sequences topologically relaxes the DNA molecule. In this relaxed form, there is only one identifiable cruciform and this is extruded from a point at the central axis of the viral inverted repeat insert. While there exist multiple smaller inverted repeats within the cloned sequences (see Fig. II.2), none of these were observed to be in the cruciform.

Stability of the cloned viral sequences in vivo Generally, large palindromic sequences are unstable in wild-type *E. coli* (Lilley, 1981; Mizuuchi *et al.*, 1982; Collins *et al.*, 1982; Warren and Green, 1985)). Palindromic sequences much smaller than the viral insert described here have been shown to be unstable when cloned into *E. coli* (Courey and Wang, 1983; Boissy and Astell, 1985). For example, spontaneous central axis deletions within one specific 68 base-pair inverted repeat have been identified (Courey and Wang, 1983). The instability of inverted repeats *in vivo* is most likely related to the ability of these sequences to adopt unusual DNA secondary structures, such as cruciforms. The presence of cruciforms may adversely affect the replication of these

sequences (Bolivar *et al.*, 1977), or they may be the target for intracellular enzymes which effect their degradation (Leach and Stahl, 1983). The plasmid pSAlB. 56A has been successfully cloned in DB1256 (Fig. II.7, lanes 2 and 3), a *recA recBC sbcB* strain of *E. coli*. The sequence of the insert in purified plasmid pSAlB.56A derived from this host has been verified by sequence analysis (DeLange *et al.*, 1986). The telomeric sequences were clearly unstable in the recombination competent host *E. coli* JM83 (Fig. II.7, lanes 4 and 5). The transformation of DB1256 with purified pSAlB.56A yielded the three plasmid DNA bands identified in Figure II.3 as cruciform-relaxed, supercoiled lineform and the Δ260 deletion. Transformation of JM83 only produced a DNA species which co-migrated with the 260 base-pair deleted plasmid derivative of pSAlB.56A identified by sequence analysis. The insert in this recombinant plasmid is 62 base-pairs long, and is itself an inverted repeat generated by deletion of the central region of the 322 base-pair insert and joining together the distal thirty-one bases from each end of the pSAlB.56A insert (see Appendix 1). Consequently, the breakage and rejoining site for the Δ260 deletion on the pSAlB.56A insert falls in the vicinity of the central axis of a 29 base-pair inverted repeat within the insert sequences (see Fig. II.2 and also inverted repeat 'c' in Fig. II.8). *Hin*dIII digests of these plasmids clearly showed the difference in size between the native and derivative plasmids (Fig. II.7B). Also, it is clear from these experiments that the appearance of the deleted plasmid derivative was a persistent product of pSAlB.56A transformations, even in *recA recBC sbcB E. coli*, where it constituted around 10% of the total plasmid DNA extracted from the DB1256 host.

The nature of the recombination event giving rise to the derivative form is reminiscent of the results of Biossy and Astell (1985), Hagan and Warren (1983) and Collins *et al.* (1982) who also observed that the termini of central axis deletions were found to map

Figure II.7. *In vivo* stability of pSA1B.56A plasmid DNA. Crude preparations of pSA1B.56A (lane 1) were fractionated by agarose gel electrophoresis. Purified full-length plasmid (lanes 1.1) was separated from a 260 base-pair deleted recombination derivative (lanes 1.2) and used to re-transform the recombination-deficient *Escherichia coli* strain DB1256 (lanes 2 and 3) and the recombination-competent strain JM83 (lanes 4 and 5). Individual clones were selected, and the plasmid DNA was extracted and electrophoresed in 1% (w/v) agarose. Panel A: untreated plasmid DNA; panel B: *HindIII* digests of the same, except for the untreated control plasmids in lanes 1.1 and 1.2.


Hind III Cut

near small direct repeats present within or flanking larger inverted repeats. These authors have postulated that the recombinational event responsible for the deletion was related to the ability of the cloned inverted repeats to form cruciforms *in vivo* and thereby juxtaposing the direct repeats which are believed to mediate the recombinational event. Given the topoisomer distribution of pSAlB.56A observed in the bacterial cell, the most highly supercoiled plasmid molecules could conceivably extrude cruciforms that branch migrated to the position where the limits of the deletion have been identified. The presence of a recombination-derived deletion product of pSAlB.56A, with the characteristics described, suggests that the viral sequences may have been extruded into the cruciform in *E. coli*, but in all probability such extruded forms would persist only transiently because there was no evidence for any steady-state pool of topoisomers analogous to the supercoiled, cruciform-containing plasmid DNA demonstrated in Figure II.5. Neither was there any direct evidence for the steady-state presence of extruded cruciforms *in vivo* generally (Sinden *et al.*, 1983).

Energetics of cruciformation and the influence of telomere structure In the past, several features of inverted repeat DNA sequences have been shown to have an influence on the energy parameters governing cruciformation. Important features are base composition (Greaves *et al.*, 1985; Lilley, 1985; Haniford and Pulleyblank, 1986; Naylor *et al.*, 1986), the size of the inverted repeat (Lilley, 1981; Gellert *et al.*, 1983) and the extent of sequence symmetry at the central axis (Lilley, 1985). The Shope fibroma virus telomere is rich in A+T base-pairs and has perfect sequence symmetry at the inverted repeat axis. These are features that might be expected to contribute to lowering the energy barrier to cruciformation in these sequences. In addition, all known poxvirus telomeres contain sequences that have unpaired, and presumably extrahelical, bases close to the telomere

hairpin which arise because the inverted repeat sequences are not completely symmetrical in this region (see Fig. II.9). The presence of eight extrahelical bases in SFV telomeres would be expected to raise the energy of cruciform formation in the cloned inverted repeat sequences, since these bases would be stably paired only in the lineform. The energy cost of extruding asymmetric sequences into cruciforms with extrahelical bases had not been directly examined prior to this investigation. Application of the two-dimensional gel electrophoresis technique for the calculation of the energy parameters governing structural perturbations in duplex DNA has been useful and was therefore applied here to determine the influence of partial asymmetry on cruciformation within SFV telomere sequences.

Experimentally, a series of related deletions of pSAlB.56A were constructed. These constructs, which include three central axis deletions of pSD19 (D5.16, D6.3 and D5.10), one central axis addition of D5.10 (pSAD-2) and one uni-directional deletion of pSD19 (pD21) are summarized in Figure II.8. Their precise sequences can be obtained from Figures II.2 and II.9. In each construct, the palindromic sequences native to pSAlB.56A (and the virus) have been uniquely modified while maintaining complete or near-complete inverted repeat symmetry. The determination of the energy of cruciformation in each was accomplished using a standard two-dimensional gel technique (Wang *et al.*, 1983). The only structural transition detected in the various recombinant plasmids (these are listed in Table II.1) was related to the cruciform extrusion event already characterized in plasmid pSAlB.56A. The supercoiling-dependent transition to the cruciform configuration was visible in the various two-dimensional gel profiles of Figure II.10 as a profound discontinuity in the smooth C-shaped distribution of topoisomer molecules. For example, cruciformation was evident in pSAlB.56A plasmid DNA by the torsionally-induced reduction in mobility of topoisomer -15 in the first

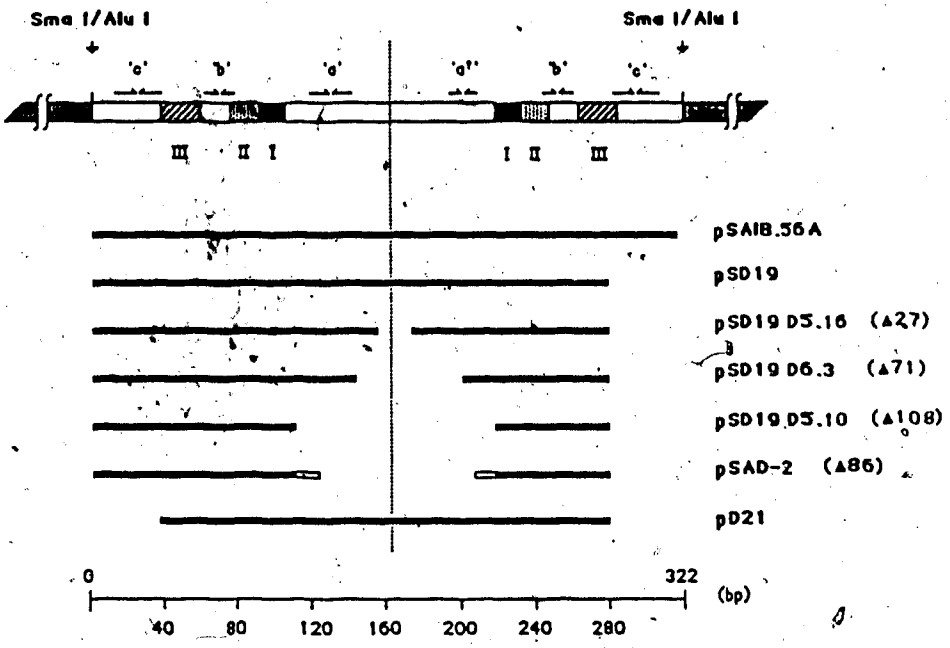


Figure II.8. The relationship between the pSD19 deletion series and important sequences within the viral telomere. Those telomere sequences that have been identified as critical for the *in vivo* resolution of pSAlB.56A are denoted as boxes I, II and III (DeLange *et al.*, 1986). Six small inverted repeat regions are also shown as "a", "a†", "b" and "c". The dotted line represents the axis of symmetry of the cloned viral inverted repeat. The open-bar region of pSAD-2 represents the 22 nucleotides of wild-type viral sequence added to the central axis of pSD19 D5.10. Plasmids pSAlB.56A, pSD19, D5.16, D6.3, pSAD-2 and pD21 are all competent for resolution to intact hairpin telomeres *in vivo*, while D5.10 is resolution-deficient (see Results, Chapter II).

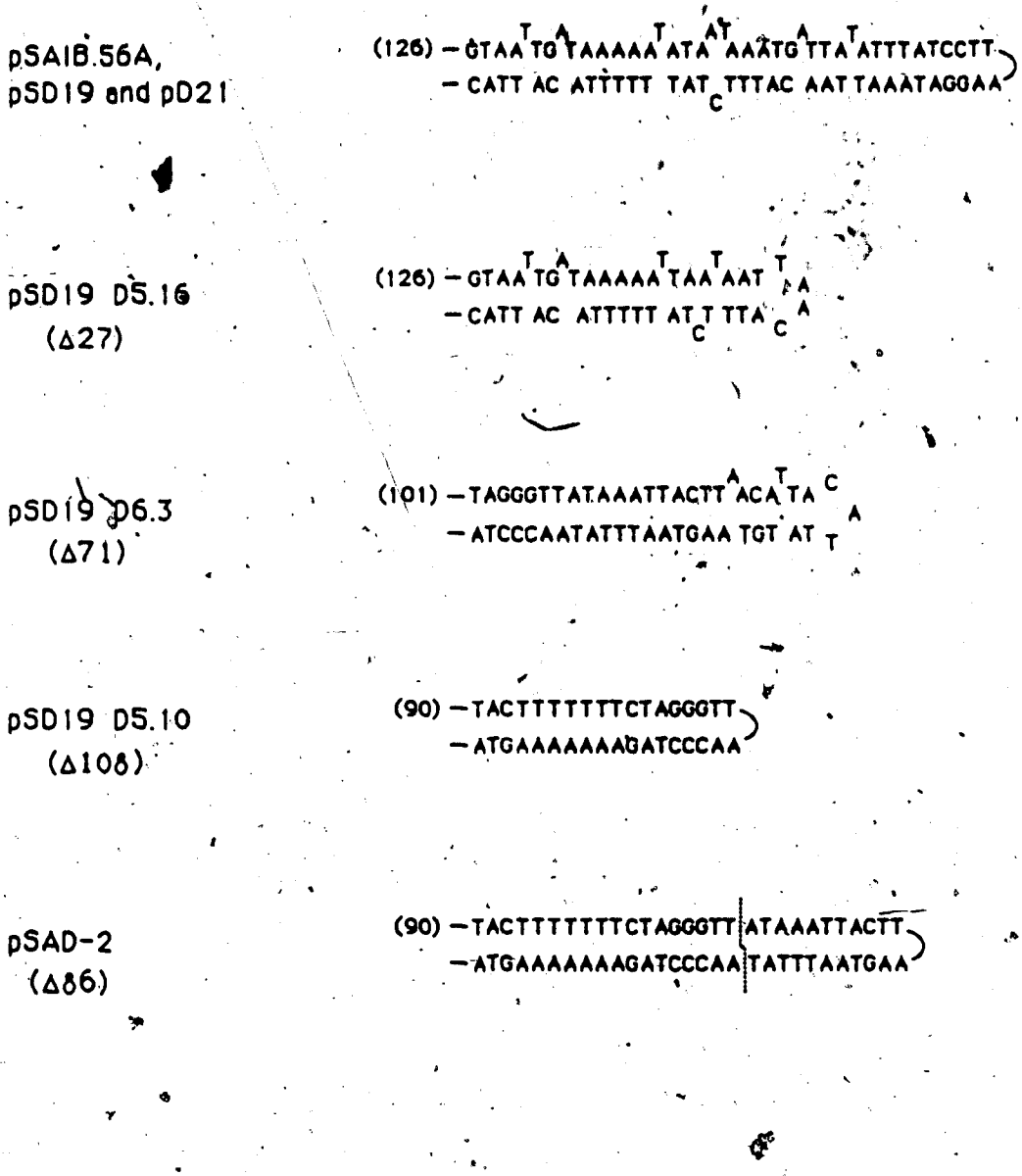
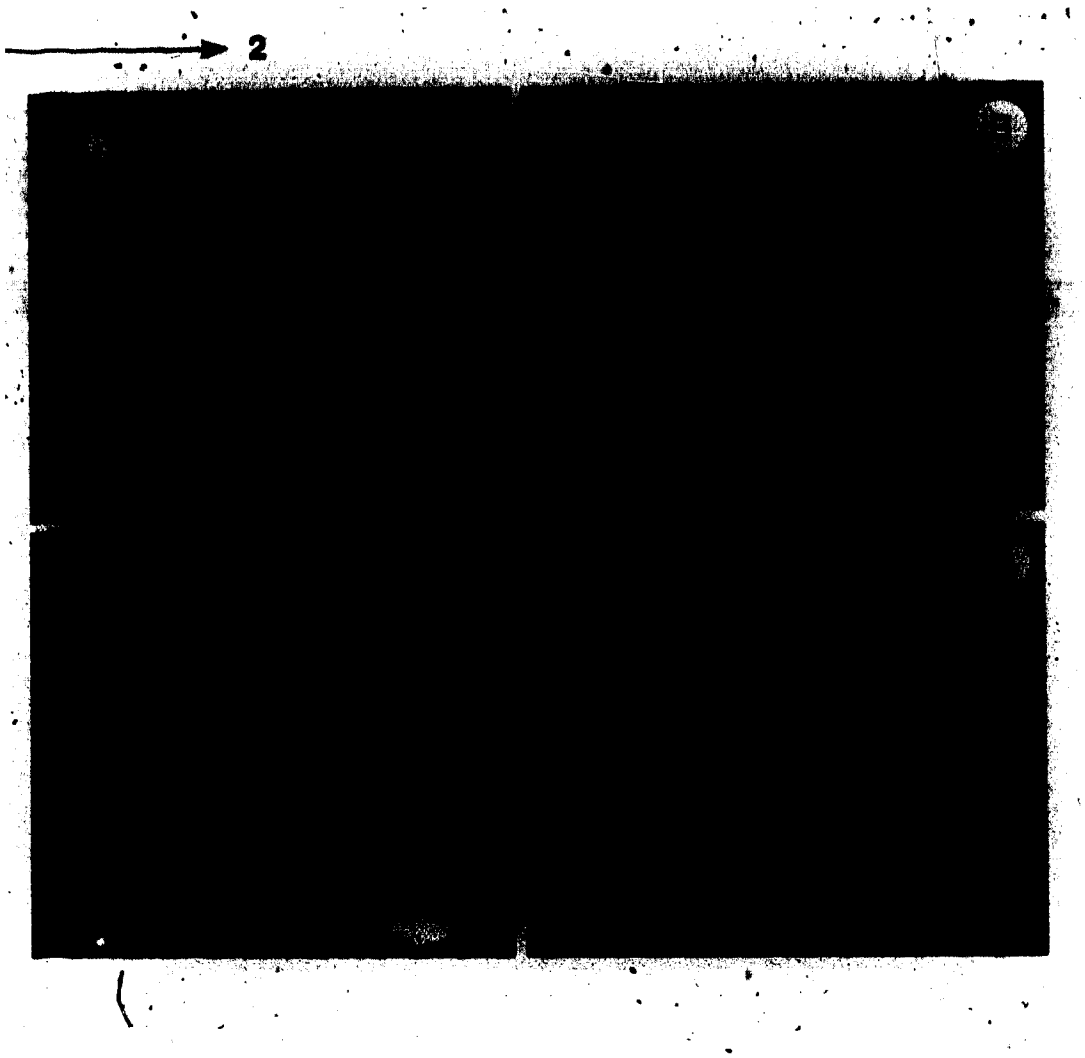
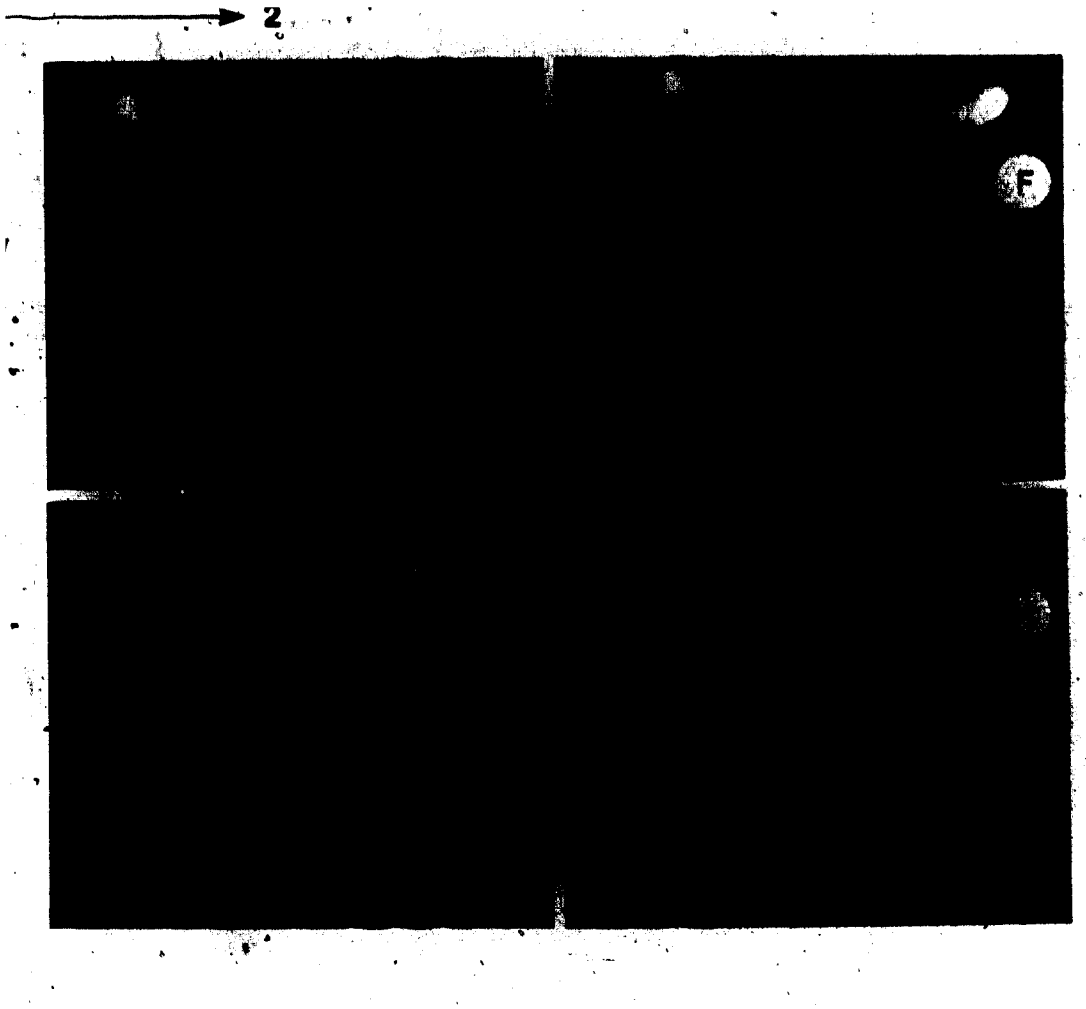
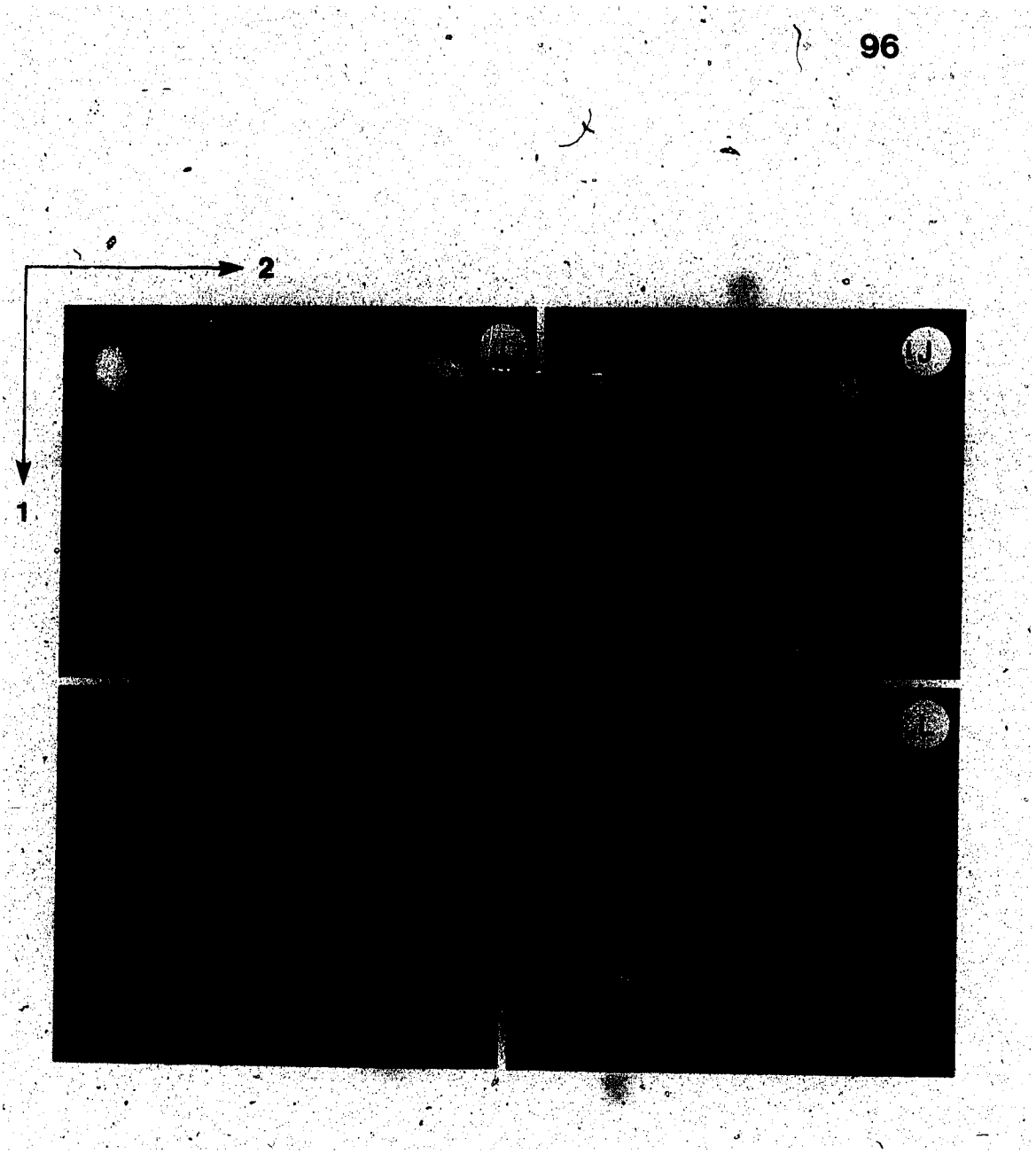


Figure II.9. Cruciform hairpin structures expected for pSAIB.56A and each of its deletion derivatives. The total number of nucleotides deleted from pSD19 is noted in parentheses below the plasmid name. The number of the nucleotide, counting out from the cloning site, is noted in parentheses at the beginning of the sequence. For plasmids D5.10 and pSAD-2, the two hairpins of the cruciform are identical because the inverted repeat sequences are perfectly symmetrical. For pSAIB.56A, pSD19, pD21, D5.16 and D6.3 only one of the two flip-flop hairpin sequences is shown.

Figure II.10. Two-dimensional agarose gel analysis of the various plasmid DNAs. Topoisomer distributions of the respective plasmids were prepared either by partial relaxation with eukaryotic topoisomerase I (plates A, B, C, F, G, H, I and J) or by the partial supercoiling of relaxed, covalently-closed plasmid DNA circles with gyrase (E, K and L). In plate D, plasmid pSAlB.56A is untreated. All gels were 1% (w/v) agarose and were run in the first dimension at room temperature except for those in plates B and E which were run at 37°C. Second dimensions were at room temperature in the presence of 1.28μM chloroquine diphosphate. The respective dimensions are given by the arrows in the upper left. The brightly staining spot in the upper left corner of most plates is nicked plasmid DNA. All gels have been stained with ethidium bromide. Plates A and B: pUC13; plate C: pAT34; plates D, E and F: pSAlB.56A; plates G, H, I, J, K and L are plasmids pSD19, pSAD22, pSD19 D21, pSD19 D5.16, pSD19 D6.3 and pSD19 D5.10, respectively. The topoisomers in plate E are partially numbered according to the procedure desribed in Experimental Procedures. One topoisomer pair, -4 and +4, is shown. The relative mobilities of topoisomers were determined from photographs such as these and used to calculate the respective energy of cruciform formation for each plasmid, as summarized in Table II.1.


2


2
F


2
1

dimension of the agarose gels (Fig. II.10E and II.10F). The numbering of topoisomers is related to linking difference (ΔLk), the slowest migrating lineform molecule being designated as 0. Positively supercoiled topoisomers form the top of the "C" proceeding down to the right. Negatively supercoiled topoisomers extend to the bottom of the gels on the left forming the image of the "C". The transition from lineform to cruciform in pSAlB.56A clearly occurred at topoisomer -14, such that those topoisomers with a higher negative superhelicity co-migrated with nicked plasmid DNA (the intensely staining spot in the upper left corner of the photographs) in the first dimension, just as was observed in the one-dimensional gel of Figure II.3. Migration in the first dimension of these gels distinguishes molecules on the basis of molecular structure, which is determined by the native torsional strain. Migration in the second dimension, in the presence of chloroquine, distinguishes topoisomers solely on the basis of ΔLk. Thus, the linking difference of cruciform-containing DNA (or other duplex-perturbed forms) can be readily identified.

The extrusion of the viral sequences of plasmid pSAlB.56A into a cruciform occurred at a calculated energy of supercoiling (ΔG_S) in the plasmid of 44 kcal/ mole, computed using the published relationship between the energy of supercoiling and linking difference (Wang *et al.*, 1983). This value is equivalent to the minimum energy of cruciform formation (ΔG_f) (Table II.1). Cruciform-relaxed pSAlB.56A purified from *E. coli* was largely insensitive to the action of the topoisomerase I used to convert plasmid DNA to the lineform by the reduction of the titratable superhelical density of the molecule. Prolonged incubations of pSAlB.56A with topoisomerase I at 37 °C failed to "relax" the plasmid DNA sufficiently. In order to obtain the topoisomer distribution shown in Figure II.10F, it was necessary to pretreat the plasmid with ethidium bromide and subsequently

Table 1

Cruciform energies of formation (ΔG_f) in pSAIB.56A and deletion derivatives

Plasmid	Length of the Inverted Repeat Insert (Base pairs)	Bases at axis without dyad symmetry†	Unpaired bases in cruciform stem†	Cruciform ΔG_f* (kcal/mole)
pSAIB.56A	322	0	8	44
pSD19	238	0	8	40
pD21	238	0	8	34
pSD19 D5.16	211	4	5	39
pSD19 D6.3	167	3	2	32
pSD19 D5.10	130	0	0	27
pSAD-2	150	0	0	27
pAT34	34	0	0	12

† The total number of unpaired bases in the stem of the cruciform hairpin is based on predictions of the maximum possible base pairing. The number given for bases without dyad symmetry at the inverted repeat axis denotes the minimum number of obligatory unpaired bases in the hairpin loop, based on sequence. It does not reflect an obligatory number of unpaired bases based on hairpin structure, which may be four to six bases (Lilley, 1986).

* As an average of at least two experiments; with an estimated error of ± 2 kcal/mole.

cold-extract the drug with butanol (Panyutin *et al.*, 1984) in an attempt to remove the pre-existing cruciform from the plasmid. Surprisingly, when the ethidium bromide treated plasmid was maintained at 4°C and electrophoresed at 4°C for analytical purposes, most of the molecules were observed to be in the cruciform state (not shown). Eventually, it proved necessary to treat the ethidium bromide-induced lineform plasmid DNA with topoisomerase I at 4 °C immediately following the extraction step. A complete relaxation (average superhelical density reduced to zero)of the plasmid could be obtained, if desired, indicating that the plasmid had been quantitatively converted to the lineform. The ΔG_f values for cruciformation in pSAlB.56A obtained following the gyrase treatment of nick-relaxed pSAlB.56A plasmid DNA (Fig. II.10E) were indistinguishable from those obtained with the topoisomerase I treated plasmids, proving that the calculated value of ΔG_f was independent of the means used to generate the topoisomer distribution. Inherent in these observations, the extrusion event in pSAlB.56A does not appear to be restricted by as large a kinetic barrier as has been observed for certain other inverted repeats (Lilley, 1985; Gellert *et al.*, 1983). In the time span of a few hours required for the extraction of the drug and the electrophoresis of the DNA, both performed at 4 °C, the lineform was converted to the cruciform. This implies that while the energy of formation may be relatively high, the kinetic barrier must be relatively modest. The difficulty in maintaining the plasmid in the lineform has prevented efforts to calculate precisely the energy of activation for the transition. A linking difference of -14 roughly corresponds to a superhelical density of -0.048 in this plasmid. Above this density, the cruciform was the thermodynamically stable form of the viral sequences in pSAlB.56A, under the experimental conditions imposed.

The computed energies of formation for this major lineform-to-cruciform transition in

each of the deletion plasmids are listed in Table II.1. Plasmids pSD19, D21 and D5.16 were relaxed with topoisomerase I without ethidium bromide pretreatment, whereas pSAD-2 required the pretreatment. Included in Figure II.10 are topoisomer profiles for pUC13 DNA generated either at room temperature or at 37°C (Fig II.10A and B, respectively) as controls. A minor perturbation was apparent in the vector sequences at room temperature at topoisomer -13 and this transition was observed to be unstable due to the coexistence of unperturbed and perturbed forms of the same topoisomer (note the spur in lower right of Fig. II.10A). At 37°C there appeared to be a temperature-induced transition, also apparent at topoisomer -13. The profile of the temperature induced effect is similar to that described by Lee and Bauer (1985) and may correspond to early melting regions in the plasmid DNA. Whether or not these two transitions corresponded to the same sequences in the vector is not known. A distinctive transition was observed with the control plasmid pAT34 (Fig. II.10C.). The repeating (dA.T) sequence in this plasmid is known to be readily extruded into a cruciform (Haniford and Pulleyblank, 1985). The structural transition is limited to a short stretch of duplex whose size can be inferred from differences in topoisomer migration between the native lineform and perturbed DNA forms. The ΔG_f for this transition was calculated to be 12 kcal/mole and is very close to reported values for $(dA.T)_{34}$ from a *Xenopus* globin gene (Greaves *et al.*, 1985). Figure II.10D shows the profile of untreated plasmid pSAlB.56A DNA . This clearly shows that the bulk of the extracted DNA was in the cruciform while only a small proportion of less supercoiled plasmid remains in the lineform.

The respective values for ΔG_f are related to the nucleotide sequences of the various cruciform hairpins in Table II.1 and Figure II.9. In Figure II.9 the actual sequences are

illustrated in the hairpin telomere form which correspond to the extruded cruciform configuration. The illustrated structures have been selected on the basis of a determination of the maximum possible base pairing and the basic features of these structures have been included in Table II.1. The extrusion of viral sequences into the cruciform in pSA1B.56A was marked by the unusually high ΔG_f of 44 kcal/mole. Without having altered any part of those sequences which are extruded in the cruciform, this value was significantly reduced in the deletions pSD19 and pD21. In these plasmids, a part of either one or both of the distal internal palindromes (denoted as "c" in Fig. II.8) has been deleted resulting in reductions of ΔG_f to 34 kcal/mole. One possible explanation is that the deleted sequences may be involved in another transition, or interaction, that competes with the major cruciform transition for available superhelical energy (see Discussion, this chapter). A series of bidirectional central axis deletions in pSD19 were created that alter the number of extrahelical bases that would form in the cruciform hairpins and which also would affect the extent of symmetry at the axis of the inverted repeat. Reductions in the number of extrahelical bases appearing in the cruciform hairpins reduces the value of ΔG_f, as suggested from comparisons of the values for pSD19, pSD19 D6.3 and pSD19 D5.10 (Table II.1). Perhaps offsetting this effect somewhat, a reduction in the extent of dyad symmetry precisely at the inverted repeat axis appeared to increase the value of ΔG_f, as demonstrated by a comparison of pSD19 and pSD19 D5.16. Such an effect is surprising given recent estimates that there are at least four obligatory unpaired bases in hairpin loop structures (Lilley, 1986). Whatever arguments one might propose to rationalize the different ΔG_f values obtained for these variants, an alteration in A+T content is not expected to be a major contributing factor considering the A+T-richness of the telomere sequences as a whole.

For the most part, the cruciforms appearing in the derivative plasmids were stable and formed readily under the conditions of the experiments. The one exception was pSD19 D5.16, which appeared to be facilely extruded into a cruciform but it did not appear to be a stable transition. Topoisomers -15, -16 and -17 all coexist in these preparations in both the lineform and cruciform (Fig. II.10J). The absence of streaking between the two DNA forms is consistent with a slow interconversion of forms during electrophoresis, in terms of the time scale of the experiment. The thermodynamic instability of the cruciform, though, remains unexplained although it could be related to the presence of numerous unpaired bases concentrated at the extreme hairpin ends.

The gel patterns displayed by the plasmids pSD19 D5.10, D6.3 and pSAD-2 (Figure II.10L, II.10K and II.10H, respectively) are distinguished from the others by demonstrating two discrete profiles. These topoisomer patterns were generated by nicking plasmids with DNase I, religating and treating the relaxed covalently-closed molecules with gyrase. The nicking reaction did not cause complete resorption of the cruciforms. Thus, the subsequent treatment with gyrase concomitantly supercoiled lineform DNA and cruciform DNA species producing the different superimposed profiles. In these profiles the supercoiled, cruciform DNA was notably free of any additional transition events and served as a useful internal control.

It was possible that the major cruciform extrusion event obscured or prevented potential minor transitions from occurring elsewhere in the viral telomere sequences. Therefore, it was of interest to investigate the perturbability of viral sequences in the absence of cruciform formation derived from the central axis of the cloned inverted repeat in pSAlB.56A. Single arm derivatives of the pSAlB.56A inverted repeat insert were

prepared by linearizing pSAlB.56A (after conversion to the lineform by nicking with DNase I) with *Afl*II followed by cleavage with either *Hin*dIII or *Eco*RI, blunt-end ligation and transformation into *E. coli* JM83. Plasmid DNAs from selected clones were analyzed, and all those tested carried inserts of the expected size. Thus, the right and left arms (pSAlB.56A-RA and -LA, respectively) of the inverted repeat could be individually analyzed in the two-dimensional gel system. Preparations of both gave results which were indistinguishable from the vector itself (not shown). In the absence of visible duplex perturbations within these single repeat inserts, it is unlikely that secondary viral sequence-specific transitions, such as cruciformation, explain the effect that deletions pSD19 and pSD19 D21 have on lowering the ΔG_f of the major cruciform extrusion event in pSAlB.56A. For instance, the potential for the inverted repeat "c" (Fig. II.8) to be extruded into the cruciform can not be demonstrated and thus it can not be argued that such an event competes with the major cruciformation event for helical torsion energy. In support of this interpretation, a 260 base-pair deleted derivative of pSAlB.56A, which almost completely maintains the deleted sequences of pSD19 (and pD21) in the form of a 62 base-pair imperfect inverted repeat, displayed no transition over and above that perturbation ascribed to vector sequences, when tested at either room temperature or at 37°C (data not shown). This is significant since the inverted repeat in Δ260 is much larger than the inverted repeat 'c' as its repeat sequences extend fully to the limit of the viral insert. Cruciformation in these small, imperfect palindromes appeared to be thermodynamically forbidden under the test conditions imposed here.

D. Discussion

In the plasmid pSAlB.56A, SFV telomere sequences have been cloned in an inverted

repeat arrangement equivalent to their organization as observed in poxvirus replicative intermediates (DeLange and McFadden, 1986; DeLange *et al.*, 1986). At superhelical densities native to bacteria, the viral sequences in the purified recombinant plasmid are extruded into a stable cruciform structure *in vitro*. This investigation of cruciformation within the cloned telomeric sequences was undertaken as an *in vitro* assessment of the potential for extrusion within the same telomere arrangement that appears transiently in replicative intermediates of SFV DNA *in vivo*. Cruciform extrusion has been postulated as an event which could lead to the resolution of these intermediates to hairpin termini *in vivo* (McFadden and Dales, 1982; McFadden and Morgan, 1982; Szostak, 1983), in which case the transition would presumably have to be temporally and spatially regulated. Alternatively, cruciformation may be deleterious to the viral replicative process and there may be safe-guards resident to the telomere sequences which prohibit its occurrence in infected cells. In this study special consideration has been given to the presence of extrahelical bases in the viral telomere hairpins. They have a profound influence on the energetics of cruciformation, as indicated by the Results, and this may be significant as they are a conserved feature of poxvirus telomeres and their *in vivo* function is currently unknown.

The results presented here and previously (DeLange *et al.*, 1986) describe a single duplex DNA transition that occurs within the inverted repeat configuration of the SFV hairpin telomeres. The topological consequences of the transition influence the electrophoretic behaviour of recombinant plasmids bearing these sequences, and mapping data defining the location and limits of the transition are evidence of a major lineform to cruciform conversion specifically within the viral sequences. The amount of supercoiling energy necessary for cruciform extrusion to occur is high relative to the other shorter inverted

repeat sequences studied to date, but the specific linking difference at which cruciformation can occur is well below the average superhelical density normally present in plasmid DNA isolated from bacteria.

In vivo, the viral inverted repeat insert confers upon the plasmid biological properties that are commonly associated with the presence of large palindromic sequences in plasmid DNA. These properties include inviability in recombination-proficient bacteria and plasmid instability, even in recombination-deficient host bacteria. Plasmid inviability and instability have been attributed to the ability of palindromic sequences to adopt cruciform stuctures which are labile or reactive in the cell (Collins *et al.*, 1982; Hagan and Warren, 1983; Boissy and Astell, 1985). However, there is not yet an established correlation between the observed biological fate of palindromic sequences and the spontaneity of cruciformation within these sequences *in vitro* (Hagan and Warren, 1983; Warren and Green, 1985). As a result, other explanations for the observed biological properties of cloned palindromic sequences have been proposed which involve a non-spontaneous, or catalyzed, version of cruciformation (Warren and Green, 1985; Leach and Lindsey, 1986). The plasmid, pSAlB.56A, is unstable in wild-type *E. coli*, and is observed to cast off a specific recombinant form even in recombination-deficient strains of *E. coli* (Fig. II.7). It appears though, that pSAlB.56A is more stable in a *recF*+ background than were the smaller parvovirus-derived palindromes studied by Boissy and Astell (1985). Whether this is, in part, a function of the high ΔG_f for cruciformation in pSAlB.56A is not known. Among other parameters that may influence this peculiar recombination event is the distance separating the direct repeats that clearly are involved in the deletion reaction (Hagan and Warren, 1983). Nonetheless, the point can be made that the viral inverted repeat, despite being imperfect and having a large ΔG_f for cruciformation, retains the

biological properties characteristic of other palindromic sequences. The *in vitro* deletion derivatives of pSAlB.56A are stable, in spite of having lower ΔG_fs but note, however, that these derivatives were all prepared from plasmid pSD19 in which one of the recombination-dependent direct repeats (arms of the inverted repeat 'c', Fig. II.8) has been removed.

Several small inverted repeats are known to be stable in *E. coli* and these have ΔG_f values for cruciformation that are low by comparison to pSAlB.56A, yet which remain substantial (about 18 to 25 kcal/mole) (Courey and Wang, 1983; Lilley and Hallam, 1984; Panyutin *et al.*, 1984). However, *in vivo*, this transition may be forbidden by a very large kinetic barrier (Courey and Wang, 1983). Interestingly, cruciformation in the poxviral telomeric sequences described here does not, according to several observations, appear to be substantially inhibited by a large kinetic barrier *in vitro*. The supercoiled-lineform of pSAlB.56A was remarkably unstable in the relaxation reaction buffer and the electrophoresis conditions used, even at reduced temperatures. In conclusion, it would be unwarranted to preclude cruciformation within the viral sequences *in vivo* on the basis of an unfavorable energy of formation or on the existence of an insurmountable kinetic barrier. In fact, the combination of a high ΔG_f and a low energy of activation for cruciformation in these sequences could be exploited, from a biological viewpoint, in a regulatable viral process. An *in vivo* transition would be expected to be rapid (have a low activation energy), yet controlled (by a high energy of formation), perhaps by manipulation of the local helix torsional energy near the threshold for cruciformation. The spatial and temporal modulatation of the process could, in theory, be achieved through the action of specific DNA binding proteins interacting either at sites

within the extruded sequences or at sites flanking the extruded sequences such that cruciformation would be sensitive to regulation exclusively within the telomeric sequences. Clearly, flanking sequences have a bearing on the energetics of cruciformation, as demonstrated by the differences in ΔG_f values for pSAlB.56A, pSD19 and pD21 (Table II.1). Previous reports (Hagan and Warren, 1983; Lilley, 1986; Sullivan and Lilley, 1986) have cited contextual phenomena, that is, the influence of nearby sequences on aspects of cruciformation within palindromic DNA. Protein interactions with the telomere sequences is consistent with the observation that resolution of SFV telomeric replicative intermediate sequences to daughter hairpins of viral minichromosomes *in vivo* specifically requires inverted copies of a 58 to 76 base-pair core target sequence, spanning the regions I, II and III as indicated in Figure II.8 (DeLange *et al.*, 1986; DeLange and McFadden, 1987).

Unpaired bases are a conserved feature of poxvirus telomeres (Baroudy *et al.*, 1982; DeLange *et al.*, 1986) and probably take the form of extrahelical bases (Evans and Morgan, 1982; Evans and Morgan, 1986). The results indicate that a profound contribution to the high ΔG_f value for cruciformation in pSAlB.56A is made by the presence of extrahelical bases. Other reasonably large cloned palindromes have been shown to have ΔG_f values around 25 kcal/mole (Mizuuchi *et al.*, 1982; Table II.1, this paper: pSAD-2 and pSD19 D5.10). Consideration of the plasmids listed in Table II.1 suggests that the effect of palindrome size and base composition near the central axis are, at best, secondary to the effect of the extrahelical bases. Significantly, there is no obvious correlation between the reported ΔG_f values for the various plasmids and the ability of those plasmids to be resolved to daughter hairpins in an *in vivo* assay system. Using the

in vivo transfection assay for resolution to minichromosomes described earlier (DeLange *et al.*, 1986), the deleted plasmids pSD19 D5.16, D6.3 and pSAD-2 are capable of being resolved to telomere hairpins in virus-infected cells (DeLange and McFadden, 1987). Note that all unpaired bases native to the SFV telomere have been deleted in plasmid pSAD-2 suggesting that extrahelical bases make no vital contribution to resolution in this system. The plasmid pSD19 D5.10 is not resolved in the *in vivo* transfection assay, even though it has the same ΔG_f as pSAD-2 and thus, these deletion clones have served to identify a region of primary viral sequence vital for resolution. Other parameters, such as the activation energy for cruciformation and/or the binding constants of putative accessory proteins, may turn out to be more critical to the replicative process.

Poxvirus telomeres are enriched in A-T basepairs. No attempt has been made in this study to measure the influence of base composition on cruciformation. Sequences within inverted repeats almost certainly influence cruciformation, but in this study differences in sequence brought about by the deletions have not significantly altered the overall base composition of the telomere sequences and their effect is presumed to be minor. Certainly, repeating (dA.T) sequences (as in the *Xenopus* globin gene) typically have lower than average ΔG_f values for cruciformation (Greaves *et al.*, 1985) and a much lower activation energy (Haniford and Pulleyblank, 1985) but this may be attributable more to the repeating nature of the sequence and not simply the presence of an A+T-rich region. Sequences rich in A-T basepairs are conformationally flexible and can affect cruciformation in a profound if not yet clearly understood fashion (Sullivan and Lilley, 1986). Coupled with the presence of perfect symmetry precisely at the axis of the telomeric inverted repeat, these are two features of SFV telomere sequences that conceivably contribute to reduce the activation energy for cruciformation in the SFV

telomere sequences (see Lilley, 1985; Lilley, 1986).

Within the major telomeric inverted repeat cloned in pSA1B.56A there are six small inverted repeat sequences ("a", "b" and "c" in Fig. II.8). Experiments described here have consistently failed to demonstrate the occurrence of secondary, torsion-induced transitions in the cloned sequences under the imposed conditions. Therefore, there appear to be no obvious transitions which compete with the major cruciformation event for available torsional energy. The internal inverted repeats almost assuredly remain in the lineform at physiological superhelical densities. This is not surprising since all these palindromes are imperfect and possess repeats no longer than twelve base-pairs in length. Those marked 'a', 'a†' and 'c' in Figure II.8 can be deleted without adversely affecting resolution in the *in vivo* assay (DeLange and McFadden, 1987), while the small inverted repeats designated 'b' in the same figure fall within the domain shown to be critical for resolution (Delange *et al.*, 1986). A considerable contribution to the understanding of telomere replication in poxviruses will accompany the identification of the proteins that participate in the resolution of the telomere inverted repeat arrangements. A vaccinia virus encoded DNA cross-linking protein has been identified that may participate in this action (Lakritz *et al.*, 1985). In the future, it will be instructive to compare the binding of proteins to SFV telomere sequences as they exist in the lineform and the cruciform, as well as to measure the influence that telomere binding proteins bring to bear on the facility of cruciformation. Hopefully, it will then be possible to assess whether cruciformation is required for telomere replication *in vivo* or whether it is an event that must, in fact, be prohibited for efficient replication to proceed.

Bibliography

Baroudy, B. M., Venkatesan, S. and Moss, B. (1982). *Cell* **28**, 315-324.

Bateman, A. J. (1975). *Nature, Lond.* **253**, 379.

Berman, J., Tachibana, C. Y. and Tye, B.-K. (1986). *Proc. Nat. Acad. Sci. U.S.A.* **83**, 3713-3717.

Berns, K. I. and Hauswirth, W. W. (1982). In *Organization and Replication of Viral DNA*. (Kaplan, A. S. ed.) pp. 3-36. CRC press, Boca Raton, Fla.

Birnboim, H. C. and Doly, J. (1979). *Nucl. Acids Res.* **7**, 1513-1523.

Blackburn, E. H. (1984). *Cell* **37**, 7-8.

Blackburn, E. H. and Gall, J. G. (1978). *J. Mol. Biol.* **120**, 33-53.

Blackburn, E. H. and Szostak, J. W. (1984). *Ann. Rev. Biochem.* **53**, 163-194.

Boissy, R. and Astell, C. R. (1985). *Gene* **35**, 179-185.

Bolivar, F., Betlach, M. C., Heynèker, H., Shine, J., Rodriguez, R. and Boyer, H. (1977). *Proc. Nat. Acad. Sci. U.S.A.* **74**, 5265-5269.

Cabirac, G. F., Strayer, D. S., Sell, S. and Leibowitz, J. L. (1985). *Virol.* **143**, 163-170.

Cavalier-Smith, T. (1974). *Nature, Lond.* **250**, 467.

Collins, J., Volckaert, G. and Nevers, P. (1982). *Gene* **19**, 139-146.

Courey, A. J. and Wang, J. C. (1983). *Cell* **33**, 817-829.

Dales, S. and Pogo, B. G. T. (1981). *The Biology of Poxviruses*. Springer Verlag.

DeLange, A. M., Macaulay, C., Block, W., Mueller, T. and McFadden, G. (1984). *J. Virol.* **50**, 408-416.

DeLange, A. M. and McFadden, G. (1986). *Proc. Nat. Acad. Sci. U.S.A.* **83**, 614-618.

DeLange, A. M. and McFadden, G. (1987). *J. Virol.* **61**, 1957-1963.

DeLange, A. M., Reddy, M., Scraba, D., Upton, C. and McFadden, G. (1986). *J. Virol.* **59**, 249-259.

deMassy, B., Studier, F. W., Dorgai, L., Appelbaum, E. and Weisberg, R. A. (1984). *Cold Spring Harbor Symp. Quant. Biol.* **49**, 715-726.

Depew, R. E. and Wang, J. C. (1975). *Proc. Nat. Acad. Sci. U.S.A.* **72**, 4275-4279.

Evans, D. E. and Morgan, A. R. (1982). *J. Mol. Biol.* **160**, 117-122.

Evans, D. E. and Morgan, A. R. (1986). *Nucl. Acids Res.* **14**, 4267-4280.

Gellert, M., Mizuuchi, K., O'Dea, M. H., Ohmori, H. and Tomizawa, J. (1979). *Cold Spring Harbor Symp. Quant. Biol.* **43**, 35-40.

Gellert, M., O'Dea, M. H. and Mizuuchi, K. (1983). *Proc. Nat. Acad. Sci. U.S.A.* **80**, 5545-5549.

Geshelin, P. and Berns, K. I. (1974). *J. Mol. Biol.* **88**, 785-796.

Glickman, B. W. and Ripley, L. S. (1984). *Proc. Nat. Acad. Sci. U.S.A.* **81**, 512-516.

González, A., Talavera, A., Amendral, J. M. and Viñuela, E. (1986). *Nucl. Acids Res.* **14**, 6835-6844.

Gottschling, D. E. and Zakian, V. A. (1986). *Cell* **47**, 195-205.

Greaves, D. R., Patient, R. K. and Lilley. D. M. J. (1985). *J. Mol. Biol.* **185**, 461-478.

Hagan, C. E. and Warren, J. G. (1983). *Gene* **24**, 317-326.

Haniford, D. B. and Pulleyblank, D. E. (1985). *Nucl. Acids Res.* **13**, 4343-4363.

Horowitz, D. S. and Wang, J. C. (1984). *J. Mol. Biol.* **173**, 75-91.

Keller, W. (1975). *Proc. Nat. Acad. Sci. U.S.A.* **72**, 2550-2554.

Kemper, B., Jensch, F., v. Depka-Prondzynski, M., Fritz, H-J., Borgmeyer, U. and Mizuuchi, K. (1984). *Cold Spring Harbor Symp. Quant. Biol.* **49**, 815-825.

Kikuchi, Y., Hirai, K. Gunge, N. and Hishinuma, F. (1985). *EMBO J.* **4**, 1881-1886.

Lakritz, N., Foglesong, P. D., Reddy, M., Baum, S., Hurwitz, J. and Bauer, W. R. (1985). *J. Virol.* **53**, 935-943.

Landridge, J., Landridge, P. and Berquist, P. L. (1980). *Anal. Biochem.* **103**, 264-271.

Leach, D. and Lindsay, J. (1986). *Mol. Gen. Genet.* **204**, 322-327.

Leach, D. R. F. and Stahl, F. W. (1983). *Nature, Lond.* **305**, 448-451.

Lee, F. S. and Bauer, W. R. (1985). *Nucl. Acids Res.* **13**, 1665-1682.

Lilley, D. M. J. (1981). *Nature, Lond.* **292**, 380-382.

Lilley, D. M. J. (1985). *Nucl. Acids Res.* **13**, 1443-1465.

Lilley, D. M. J. (1986). *Biochem. Soc. Trans.* **14**, 211-213.

Lilley, D. M. J. and Hallam, L. R. (1984). *J. Mol. Biol.* **180**, 179-200.

Lilley, D. M. J. and Kemper, B. (1984). *Cell* **36**, 413-422.

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

Maxam, A. M. and Gilbert, W. (1980). *Methods Enzymol.* **65**, 499-560.

McFadden, G. and Dales, S. (1982). In *Organization and Replication of Viral DNA*

McFadden, G. and Morgan, A. R. (1982). *J. Theor. Biol.* **97**, 343-349.

Merchlinsky, M. and Moss, B. (1986). *Cell* **45**, 879-884.

Mizuuchi, K., Kemper, B., Hays, J. and Weisberg, R. A. (1982a). *Cell* **29**, 357-365.

Mizuuchi, K., Mizuuchi, M and Gellert, M. (1982b). *J. Mol. Biol.* **156**, 229-243.

Morgan, A. R., Lee, J. S., Pulleyblank, D. E., Murray, N. L. and Evans, D. H. (1979). *Nucl. Acids Res.* **7**, 547-569.

Moss, B., Winters, E. and Jones, E. V. (1983). In *Mechanism of DNA Replication and Recombination*. (Cozzarelli, N. R. ed.) pp. 449-461. A. R. Liss, Inc., N.Y.

Moyer, R. N. and Graves, R. L. (1981). *Cell* **27**, 391-401

Naylor, L. H., Lilley, D. M. J. and van de Sande, J. H. (1986). *EMBO* **5**, 2407-2413.

Panyutin, I., Klishko, V. and Lyamichev, V. (1984). *J. Biomol. Struct. Dynamics* **1**, 1311-1324.

Pritchard, A. E. and Cummings, D. J. (1981). *Proc. Nat. Acad. Sci. U.S.A.* **78**, 7341-7345.

Pulleyblank, D. E., Shure, M., Tang, D., Vinograd, J. and Vosberg, H-P. (1975). *Proc. Nat. Acad. Sci. U.S.A.* **72**, 4280-4284.

Sadowski, P. (1971). *J. Biol. Chem.* **246**, 209-216.

Sahai, B. M. and Kaplan, J. G. (1986). *Anal. Biochem.* **156**, 364-379.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). *Proc. Nat. Acad. Sci. USA* **74**, 5463-5467.

Sinden, R. R., Broyles, S. S. and Pettijohn, D. E. (1983). *Proc. Nat. Acad. Sci. U.S.A.* **80**, 1797-1801.

Shope, R. E. (1932). *J. Exp. Med.* **56**, 793-802.

Sullivan, K. M. and Lilley, D. M. J. (1986). *Cell* **47**, 817-827.

Szostak, J. W. (1983). *Cold Spring Harbor Symp. Quant. Biol.* **47**, 1187-1194.

Wang, J. C., Peck, L. J. and Becherer, K. (1983). *Cold Spring Harbor Symp. Quant. Biol.* 85-91.

Warren, G. J. and Green, R. L. (1985). *J. Bact.* **161**, 1103-1111.

Wills, A., DeLange, A. M., Gregson, C., Macaulay, C. and McFadden, G. (1983). *Virol.* **130**, 403-414.

III. THE SITE-SPECIFIC CLEAVAGE OF SYNTHETIC HOLLIDAY JUNCTION ANALOGS AND RELATED BRANCHED DNA STRUCTURES BY BACTERIOPHAGE T7 ENDONUCLEASE I [1]

A. Introduction

Bacteriophage T7 endonuclease I (Center and Richardson, 1970; Sadowski, 1971), T4 endonuclease VII (Kemper and Garabett, 1981) and lambda integrase (Int protein) (Kikuchi and Nash, 1979) are three enzymes which have been shown recently to resolve, by a specific cleavage reaction, branched DNA structures analogous to Holliday crossovers. The Holliday crossover is widely considered to be a key recombinational intermediate that arises from the reciprocal exchange of single-strands between two homologous DNA duplexes (Holliday, 1964; Holliday, 1974; Meselson and Radding, 1975; Potter and Dressler, 1982). The branched junction of cruciform DNA which is structurally identical to a Holliday crossover, is a substrate for the endonucleolytic activity of both T4 endonuclease VII (Mizuuchi *et al.*, 1982; Lilley and Kemper, 1984; Kemper *et al.*, 1985) and T7 endonuclease I (Panayotatos and Wells, 1981; de Massy *et al.*, 1985; de Massy *et al.*, 1987). Both enzymes introduce a pair of nicks at or very near the base of the cruciform hairpins, effectively resolving these branched molecules to linear products. All three enzymes have been shown to resolve chi-form DNA, a Holliday structure prepared from restriction fragments, which characteristically has a limited region of dyad symmetry such that the crossover is free to branch-migrate (Mizuuchi *et al.*, 1982; Hsu and Landy,

1. A version of this chapter has been accepted for publication. Dickie, P., McFadden, G. and Morgan, A. R. 1987. J. Biol. Chem.

1984; de Massy *et al.*, 1987). More recently, the T4 and T7 enzymes have been shown to cleave the three-way junctions of DNA Y-structures (Jensch and Kemper, 1986; de Massy *et al.*, 1987). Endonucleases with a specificity for Holliday junctions, have also been isolated from yeast (West and Korner, 1985; Symington and Kolodner, 1985). It is generally assumed that endonucleases with this specificity function in the resolution of Holliday intermediates during homologous recombination. Therefore, analysis of the cleavage of branched duplex DNA structures by this class of enzyme (and the subsequent rejoining of strands by an enzyme like Int protein) can be instructive both in terms of an understanding of the enzymology of recombination and the stereochemical nature of branched DNA.

T7 endonuclease I and T4 endonuclease VII have been used successfully to probe specifically for cruciform structures in plasmid DNA (Greaves *et al.*, 1985; Haniford and Pulleyblank, 1985; Naylor *et al.*, 1986). Cruciforms can extrude spontaneously from inverted repeat sequences in covalently-closed, circular plasmid DNA that is under sufficient torsional strain imposed by negative supercoiling (Gellert *et al.*, 1979). A pair of nicks placed on opposing strands across one or the other diagonal axis of the cruciform crossover "resolves" the circular molecule to a nicked, linear duplex with hairpin ends. Most of what is known of the activity and specificity of these enzymes towards branched DNA has been derived from the cleavage of cruciform structures. In this, they have demonstrated great structural specificity with little or no sequence specificity, establishing their value as probes. In recent reports, there has been speculation on the resolving mechanism of these enzymes, based on the observed activity towards cruciform-containing DNA (Mizuuchi *et al.*, 1982; Kemper *et al.*, 1985; de Massy *et al.*, 1987). These substrates, however, have limited usefulness in a detailed enzymatic study.

This type of crossover is really a unique structure which possesses two topologically free branches with homologous sequences (the cruciform hairpins) and two branches with unique sequence that are torsionally constrained in the circular, supercoiled domain of the plasmid molecule. It is uncertain to what extent the efficiency and position of cleavage by the endonucleases is influenced by this limited symmetry and configurational strain. Moreover, there remains a limited understanding of the factors that govern the choice between the two cleavage axes at the junction.

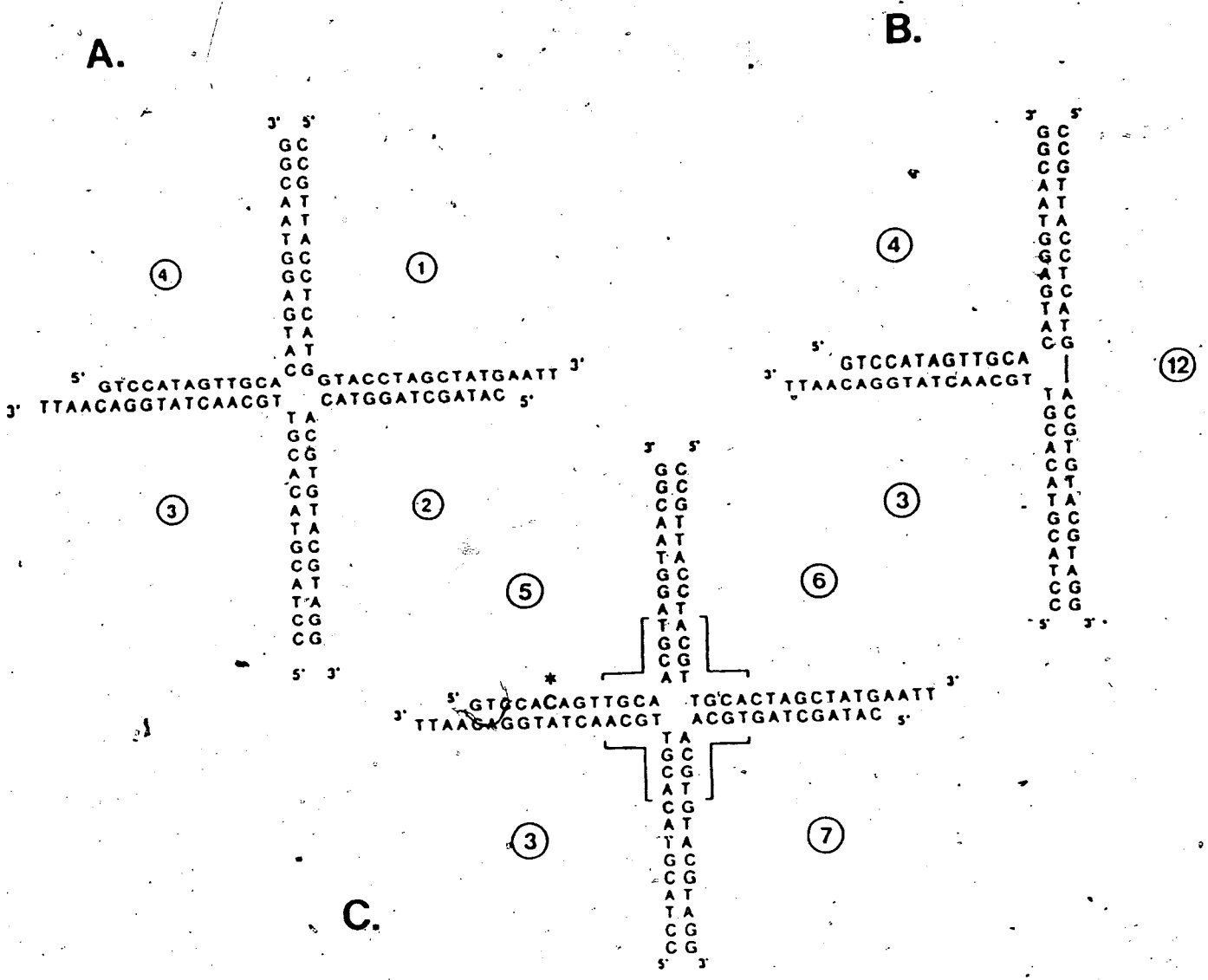
T7 endonuclease I contributes to the breakdown of host DNA during phage infection (Center *et al.*, 1970; Lee and Sadowski, 1981) and is required for phage recombination (Powling and Knippers, 1974; Kerr and Sadowski, 1975; Tsujimoto and Ogawa, 1978). Originally, the enzyme was shown to hydrolyze both single-stranded and double-stranded DNA endonucleolytically, with a 150-fold preference for single-stranded substrates (Center and Richardson, 1970; Sadowski, 1971). Tsujimoto and Ogawa (1978) demonstrated a role for the gene 3 product in the processing of branched DNA during the T7 phage life cycle and suggested that the enzyme could cleave either four-way or three-way branch junctions that may arise during recombination. Consistent with this earlier hypothesis, T7 endonuclease I has been shown to cleave cruciform DNA at the base of the hairpin structures with remarkable specificity (de Massy *et al.*, 1985; Greaves *et al.*, 1985; Naylor *et al.*, 1986), introducing nicks that are usually one or two nucleotides removed to the 5' side of the predicted crossover point, but which may be as many as four nucleotides removed from the crossover. The susceptibility of three-way junctions (Y-structures) to T7 endonuclease I digestion has also been demonstrated recently (de Massy *et al.*, 1985).

To extend the examination of substrate parameters which determine the recognition and cleavage specificity of T7 endonuclease I, this report describes the creation and susceptibility of alternative substrates prepared from partly-complementary, synthetic oligonucleotides. In appropriate combinations, the oligonucleotides were annealed to form three different types of branched DNA structures possessing either three or four duplex DNA branches. The three duplex structures were: (i) a Holliday junction analog without dyad symmetry such that branch migration is prohibited, (ii) a Holliday junction with limited dyad symmetry and, therefore, capable of branch migration and (iii) a Y-junction with a fixed crossover point (see Fig. III.1). Similar synthetic fixed-crossover Holliday structures have been studied by Kallenbach *et al.* (1983) and by Evans and Kolodner (1986) in association with the characterization of a yeast endonuclease. The results presented in this paper demonstrate the formation of the desired DNA structures and establish the sensitivity of each to cleavage by T7 endonuclease I. The cleavage observed was site-specific with respect to the position of the branch point, even more so than reported in the earlier publications. This specificity was equally applicable to the cleavage of DNA structures that possessed single-stranded branches. T7 endonuclease I also demonstrated a definite, but modest, degree of sequence specificity which has not, until now, been identified. This report emphasizes the influence of sequence specificity on the site-directed cleavage of branched DNA forms. Finally, the data are reconciled with the previously described activities of T7 endonuclease I and are compared with the published activity of T4 endonuclease VII.

B. Materials and Methods

Enzymes and Buffers T7 endonuclease I was the generous gift of Dr. Paul Sadowski

Figure III.1. Oligonucleotide sequences and the primary duplex DNA branched structures. Three duplex DNA branched structures were formed under annealing conditions with the appropriate combinations of eight oligonucleotides (oligos 1, 2, 3, 4, 5, 6, 7, 12) as shown here. Panel A: a Holliday junction analog with a fixed, or immobile, crossover point; Panel B: a Y-junction with a fixed crossover point; Panel C: a Holliday junction with limited dyad symmetry (the bracketted sequences) and with the capacity to branch migrate (is semi-mobile). The asterisk over oligonucleotide 5 identifies the one confirmed base mismatch.


A.
B.
C.
GTCCATAGTTGCA
TTAACAGGTATCAACGT
GTACCTAGCTATGAATT
CATGGATCGATAC
GTCCACAGTTGCA
TGCACTAGCTATGAATT
TTAACAGGTATCAACGT
ACGTGATCGATAC
1
2
3
4
5
6
7
12

(University of Toronto). It was obtained at a concentration of 57 units/ml (defined using fd DNA as the substrate; Sadowski, 1971) and was 95% pure as inferred from SDS-polyacrylamide gel electrophoresis (P. Sadowski, personal communication). The enzyme was stored in 50% glycerol (v/v) at -20°C and was titrated against plasmid pSA1B.56A (described in chapter II) known to contain a stable cruciform. The cruciform was extruded from an insert of poxviral telomere DNA sequence which was arranged in an inverted repeat configuration (DeLange *et al.*, 1986; DeLange and McFadden, 1987; Chapter II). In a 10µl reaction, 0.1µl of pure enzyme completely digested 50ng of pure covalently-closed, circular plasmid DNA to a linear form in one minute at 37°C, as assessed by agarose gel electrophoresis of the reaction products. The standard reaction buffer (Endo I buffer) contained 50mM Tris.HCl (pH 8), 10mM $MgCl_2$, 1mM dithiothreitol, 50µg/ml bovine serum albumin and 4mM spermidine. In his original paper, Sadowski (1971) defined a unit of activity as that amount of T7 endonuclease I required to singly nick 0.156 nmole of circular, single-stranded fd DNA in ten minutes (calculated assuming the size of fd DNA to be 6400 nucleotides). By comparison, .0057 units of T7 endonuclease I linearized 0.025 pmole of pSA1B.56A DNA in one minute. By extrapolation, in ten minutes one unit of enzyme would have linearized 0.044 nmole of pSA1B.56A DNA. Considering that the linearization of pSA1B.56A would require two nicks, and that estimations of enzyme activity with this substrate were derived from reaction end-points and not linear reaction rates, the cleavage efficiencies for the two substrates can be regarded as comparable.

T4 polynucleotide kinase was obtained from Bethesda Research Laboratories (BRL) and DNase I was purchased from Boehringer Mannheim. DNase I digestions were performed in 40mM Tris.HCl (pH 7.6), 5mM $MgCl_2$, 50µM EDTA, 0.5mM dithiothreitol and 5%

glycerol (v/v) at room temperature. The chemicals used for DNA sequencing according to the procedure of Maxam and Gilbert (1980) were purchased from NEN/Du Pont. Oligonucleotides were stored in 10mM Tris.HCl (pH 8) and 0.1mM EDTA (TE buffer).

Oligonucleotides: Oligonucleotide synthesis was carried out by the Regional DNA Synthesis Laboratory at the University of Calgary (Canada). Individual oligonucleotides were purified by the supplier either by preparative agarose gel electrophoresis or high pressure liquid chromatography and their sequences were verified by Maxam and Gilbert sequencing (see later). Lyophilized samples were rehydrated in TE buffer at a concentration of 1mg/ml and stored frozen. The oligonucleotides were end-labelled with T4 polynucleotide kinase and [γ^{32}P]ATP (specific activity 3000 Ci/mmole) following standard procedures (Maniatis *et al.*, 1982). The labelled oligonucleotides were then repurified by gel electrophoresis, either in non-denaturing 20% polyacrylamide gels run at room temperature and at 10 V/cm or in standard denaturing gels containing 20% polyacrylamide and 7M urea, with 90mM Tris.borate (pH 8.3) and 1mM EDTA (TBE buffer) as the running buffer in each case. The DNA was located by autoradiography, excised from the gels and eluted, from non-denaturing gels, into a small volume of TE buffer. DNA purified from denaturing gels was eluted into a small volume of 0.3M ammonium acetate, 10mM $MgCl_2$, 1mM EDTA and 10µg/ml tRNA, after which the DNA was precipitated with ethanol and resuspended in TE buffer (Maxam and Gilbert, 1980). Oligonucleotides purified from denaturing gels were greater than 95% pure. When purified from native gels their purity ranged from 50% to 95% (compare Figures III.6 and III.8 with Figure III.7). Examination of the results demonstrated that the identification of cleavage products did not vary significantly between these two classes of oligonucleotide samples. Unless stated otherwise, labelled oligonucleotides were repurified from native

gels and were stored frozen at -20°C.

Annealing and Cleavage Reactions Oligonucleotides were typically combined in 8 or 9µl of TE buffer. Combinations in which all oligonucleotides were unlabelled contained from 50ng to 200ng of each oligonucleotide. Preparations of ^{32}P-labelled oligonucleotide complexes contained 10ng of each unlabelled oligonucleotide and approximately 6ng of labelled oligonucleotide, in the same final volume of TE buffer. The mixtures were pre-warmed to 37°C and cooled to room temperature under annealing conditions provided by the addition of 1µl of 10X Endo I buffer. Cleavage of the annealed products was carried out at room temperature by the further addition of 0.1µl to 1.0 µl of T7 endonuclease I as indicated. Products of the reactions were analyzed by polyacrylamide gel electrophoresis. Non-denaturing gels contained 15% polyacrylamide and were run at 4°C in TBE buffer ± 5mM $MgCl_2$ at 5V/cm. The reactions were terminated by the addition of 2µl of gel loading buffer: 2% (w/v) sodium dodecyl sulfate (SDS), 30% (v/v) glycerol and 0.1% bromophenol blue. The gels were stained with ethidium bromide (0.5 µg/ml) and destained briefly in distilled water before being photographed. The identification of cleavage positions was accomplished by electrophoresing the products of reactions with labelled complexes in standard sequencing gels containing 7M urea and either 12% or 20% polyacrylamide (5% (w/w) of which is N, N'-methylenebisacrylamide) (Maxam and Gilbert, 1980). The electrophoresis buffer was TBE. Digestions of labelled complexes were terminated with an equal volume of 90% formamide, 10mM EDTA and 0.02% xylene cyanol.

Intact Holliday structure analogs, with one constituent oligonucleotide 5' end-labelled with ^{32}P, were purified from non-denaturing 15% polyacrylamide gels run at 4°C in TBE

containing 5mM $MgCl_2$. The specific DNA complexes were located by autoradiography, excised and eluted directly into a small volume of Endo I buffer at 4°C. T7 endonuclease I cleavage of purified complexes was carried out directly on this material as otherwise described.

Sequencing and DNase I Reactions 5' end-labelled oligonucleotides were sequenced using the chemical method of Maxam and Gilbert (1980). Sequencing gels of 20% polyacrylamide/7M urea as described in the preceding section were used. Fragment ladders of 5' end-labelled oligonucleotides were also generated by partial DNase I digestions. For these reactions, 5ng or less of labelled oligonucleotide was diluted into 10µl of DNase I buffer and cleaved with 100ng of DNase I at room temperature for 2 minutes. The reactions were terminated with EDTA (to 10mM) and an equal volume of gel loading buffer, and were analyzed in 12% denaturing gels as previously described.

C. Results

Formation of the DNA Junctions The sequences of the eight oligonucleotides synthesized for this study were chosen such that there would be minimal intra-strand complementarity and only the inter-strand complementarity necessary to permit the formation of the desired branched structures (see Fig. III.1), and that in the resulting structures there would be approximately an equal number of GC and AT basepairs. The structures diagrammed in Figure III.1 exhibit the maximum feasible inter-strand basepairing and, given an equimolar mixture of the individual strands, are the expected major products of the respective annealing reactions. The position of the crossover in the branch-migrating Holliday junction (Fig. III.1C) is presumably variable, but limited to the

region of dyad symmetry. The oligonucleotides were either 26 or 30 nucleotides in length, so that the structures possess either three or four duplex DNA branches which are slightly greater than a turn of helix in length. Aside from the single-stranded extensions on certain oligonucleotides, the complexes were designed to be completely complementary. Sequence analysis identified an error in one of the sequences (oligonucleotide 5, Fig. III.1C), at which point in the Holliday structure analogue a base mismatch would be expected.

Hybridization of combinations of the eight oligonucleotides (hereafter abbreviated oligos 1, 2, 3, 4, 5, 6, 7, and 12 as defined in Fig. III.1) in Endo I buffer resulted in the appearance of novel forms of DNA with a much reduced mobility in polyacrylamide gels in comparison with the individual oligonucleotides. The formation of the immobile-crossover Holliday junction (Fig III.1A) was dependent upon the addition of oligos 1, 2, 3 and 4 to TE buffer and the presence of the duplex-promoting reagents in Endo I buffer (Fig. III.2, lane f). The addition of $MgCl_2$, spermidine or 100mM NaCl separately to TE buffer was enough to stabilize the Holliday structure analog (not shown). The complex did fail to appear when the four oligonucleotides were combined in TE buffer alone (Fig. III.2, lane e) or when any one of the four designated oligos was omitted (Fig. III.2, lanes a through d). The intermediately positioned bands in lanes a through d (Fig. III.2) migrate as though they were dimer complexes of oligonucleotides (see also figure III.4). Although no bands were visible in these gels that corresonded to trimer combinations, in polyacrylamide gels electrophoresed in the presence of $MgCl_2$, trimer complexes were clearly visible (compare with Fig.III.5). The major product of the annealing reaction containing oligos 1, 2, 3 and 4 is composed of stoichiometric amounts all four oligonucleotides, as demonstrated in Figure III.3. Upon the addition of

Figure III.2. Formation of the fixed-crossover Holliday junction analog. Oligonucleotides (oligos) 1, 2, 3 and 4 were combined in various combinations (50ng of each) in T7 endonuclease I buffer (except lane e, see below) and electrophoresed in 15% polyacrylamide in Tris.borate/EDTA (TBE, less magnesium) buffer at 4°C and 5V/cm. The gels were stained with ethidium bromide and photographed. Lane a: mixture of oligos 2, 3 and 4; lane b: oligos 1, 2 and 4; lane c: 1, 3 and 4; lane d: 1, 2 and 3; lanes e and f: all four oligonucleotides combined in TE buffer, or endonuclease I buffer, respectively. Individual oligonucleotides (lanes 1, 3, 2 and 4, denoting the corresponding oligo) were present in 200ng amounts.

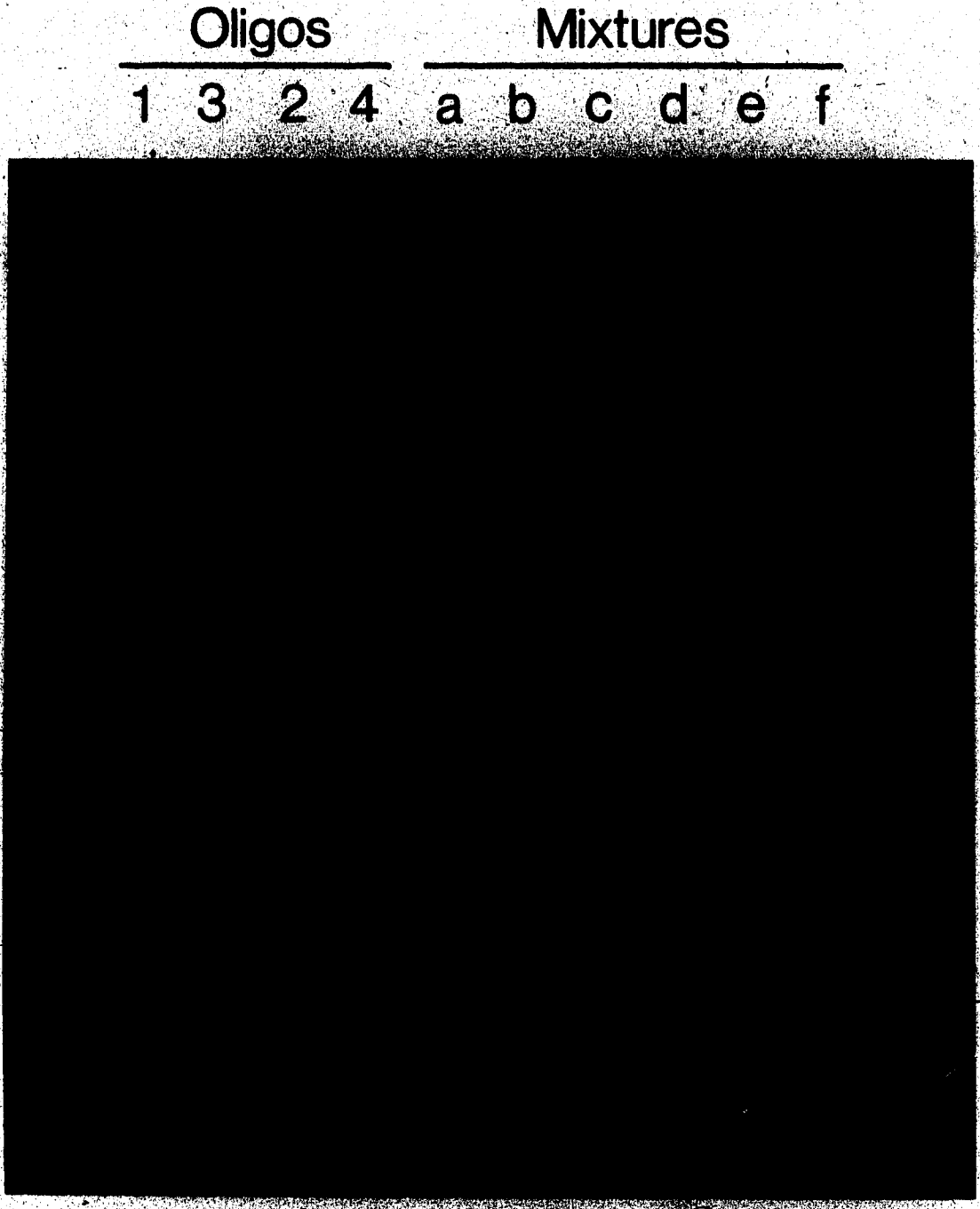
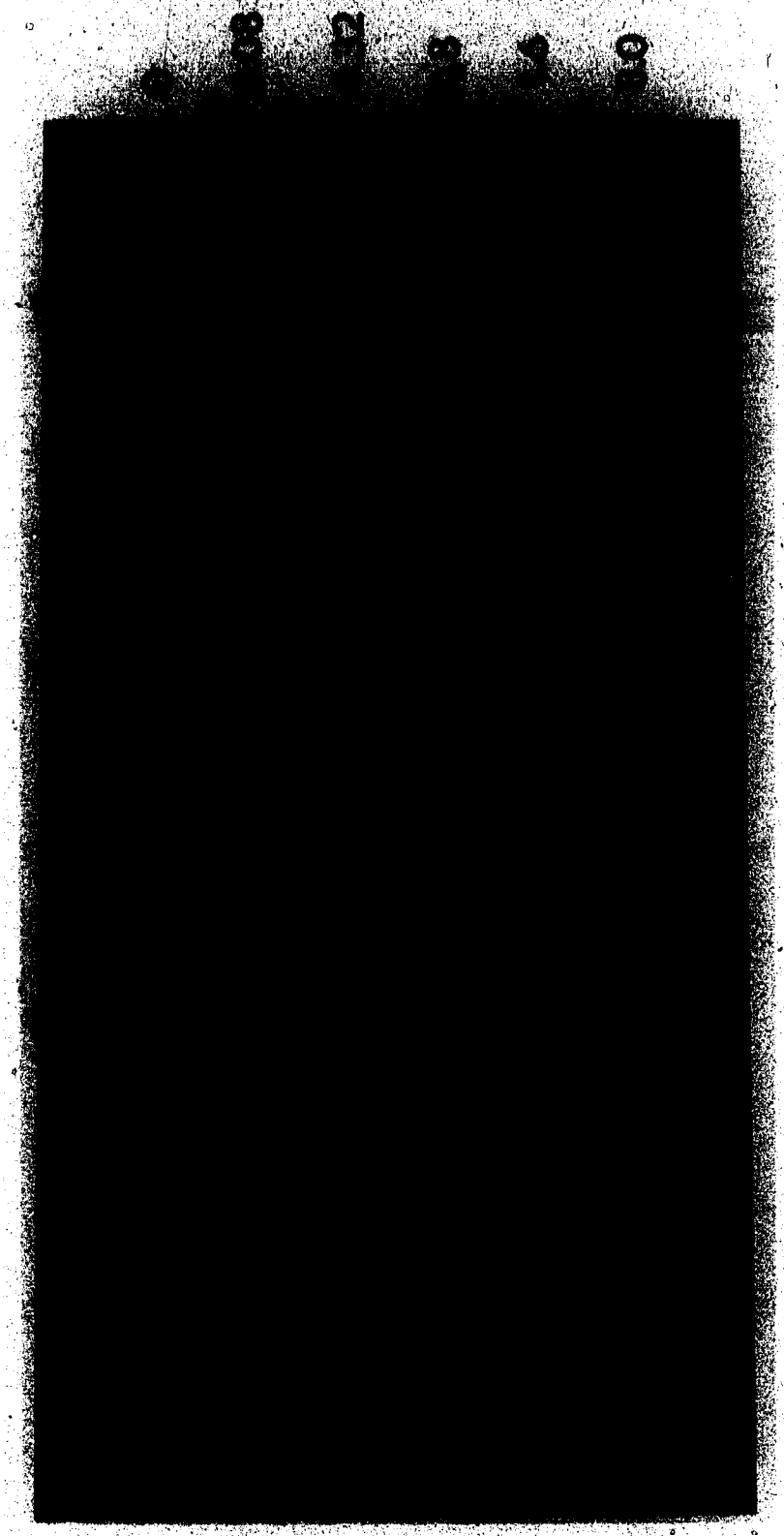

Oligos
Mixtures
1 3 2 4 a b c d e f

Figure III.3. Strand stoichiometry in the fixed-crossover Holliday junction analog. Equimolar amounts of oligos 1, 2 and 3 (0.5μM, determined spectrophotometrically) were combined in T7 endonuclease buffer with increasing amounts of oligo 4 to demonstrate the expected 1:1:1:1 stoichiometry of the component strands in the Holliday structure analog (uppermost band). The rapidly migrating material in the extreme right-hand lane is the excess uncomplexed oligo 4. Electrophoresis conditions were identical to those described in Figure III.2.

Molar Ratio - Oligo 4



increasing amounts of oligo 4 to an equimolar mixture of oligos 1, 2 and 3 (0.5μM per nucleotide) there was an increased appearance of the slowly migrating DNA complex. Maximal levels of this DNA complex were attained when oligo 4 was present at near equimolar concentrations. Hence, these results are consistent with the formation of the tetrameric molecule illustrated in Figure III.1A.

Evidence for the formation of the remaining two branched structural forms in mixtures of the appropriate oligonucleotides is presented in Figure III.4, panels B and C. The mobile-crossover Holliday junction (oligos 3, 5, 6 and 7) appears as the tetrameric complex in Figure III.4B (lane 0), while evidence for the formation of the Y-junction is apparent in Figure III.4C (lane 0). Both these gels and the gels of Figure III.5 were electrophoresed in TBE buffer that contained 5mM $MgCl_2$ because the mobile-crossover Holliday junction was not visible in gels without magnesium and most probably, like the possible trimer combinations of oligos 1, 2, 3 and 4, it was not stable under electrophoresis conditions in the absence of the cation. The low stability of the mobile junction may account for the overall low yield observed even in gels which contained magnesium. The poorer stability of the mobile-crossover Holliday junction may be related to the presence of the single base mismatch identified, or may be attributable to branch migration. The Y-junction, on the other hand, was stable in gels run in the absence of magnesium. Trimer combinations of oligonucleotides, which presumably form structures with two duplex and two single-stranded DNA branches, and therefore are structurally distinct from the duplex Y-junction, co-migrated with the fully duplexed Y-junctions when electrophoresed together in magnesium-containing gels (not shown). The similar mobilities of the various trimer complexes can be appreciated by a comparison of Figures III.4C and III.5A. The relative mobility of dimer complexes, molecules which

Figure III.4. Cleavage of the duplex DNA branched structures by T7 endonuclease I. Appropriate combinations of oligonucleotides (150ng each) in 10µl of T7 endonuclease I buffer were treated with various amounts of T7 endonuclease I at room temperature for various periods of time. The reactions were stopped by the addition of EDTA (to 10mM) and SDS (to 0.1% (w/v)) and electrophoresed at 5V/cm in 15% polyacrylamide containing TBE buffer plus 5mM $MgCl_2$. The gels were run at 4°C then stained with ethidium bromide. Panel A: the fixed-crossover Holliday junction (tetrameric complex of oligos 1,2,3,4), either untreated (0), or digested with 0.2, 0.5 or 1.0µl of enzyme for 30 minutes (see text for activity of enzyme). Panel B: the mobile-crossover Holliday junction (tetrameric complex of oligos 3,5,6,7), similarly treated as A. Panel C: the Y-junction (trimeric complex of oligos 3,4,12), treated with 1.0µl enzyme for 0, 10, 30 or 60 minutes. A dimer complex (oligos 3 and 4) and oligonucleotide 12 are included as markers.

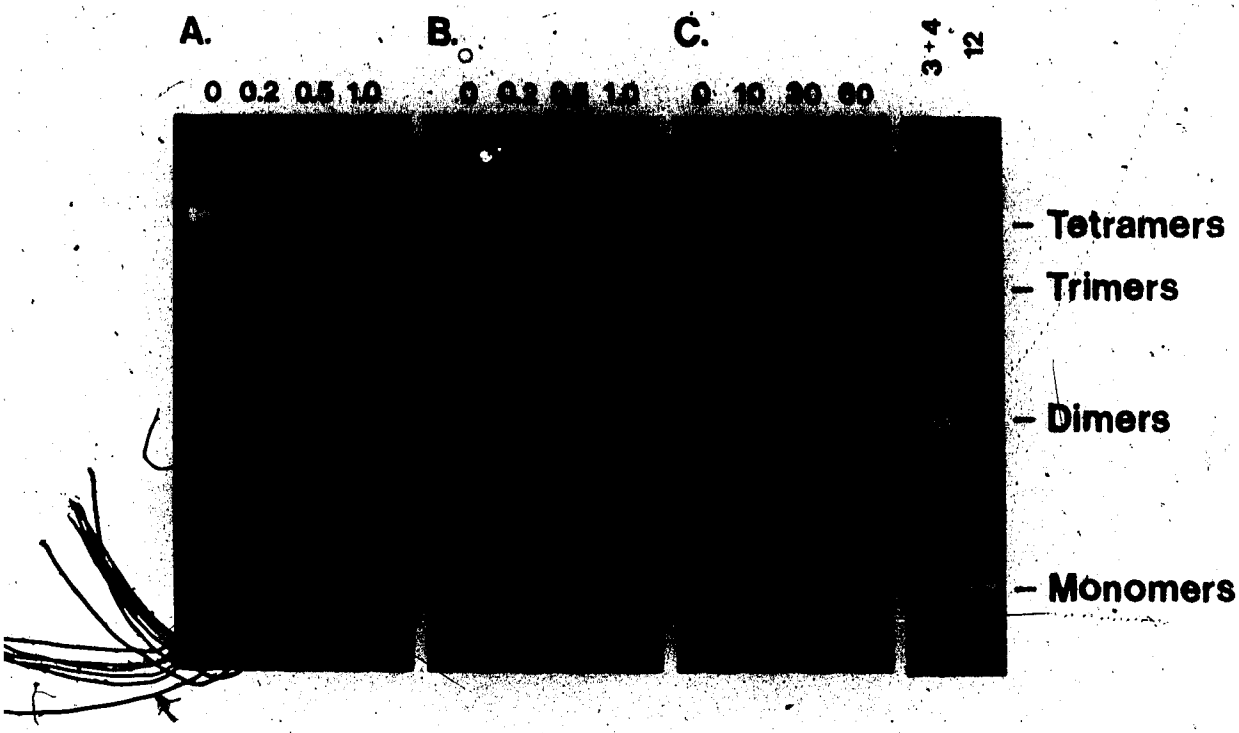
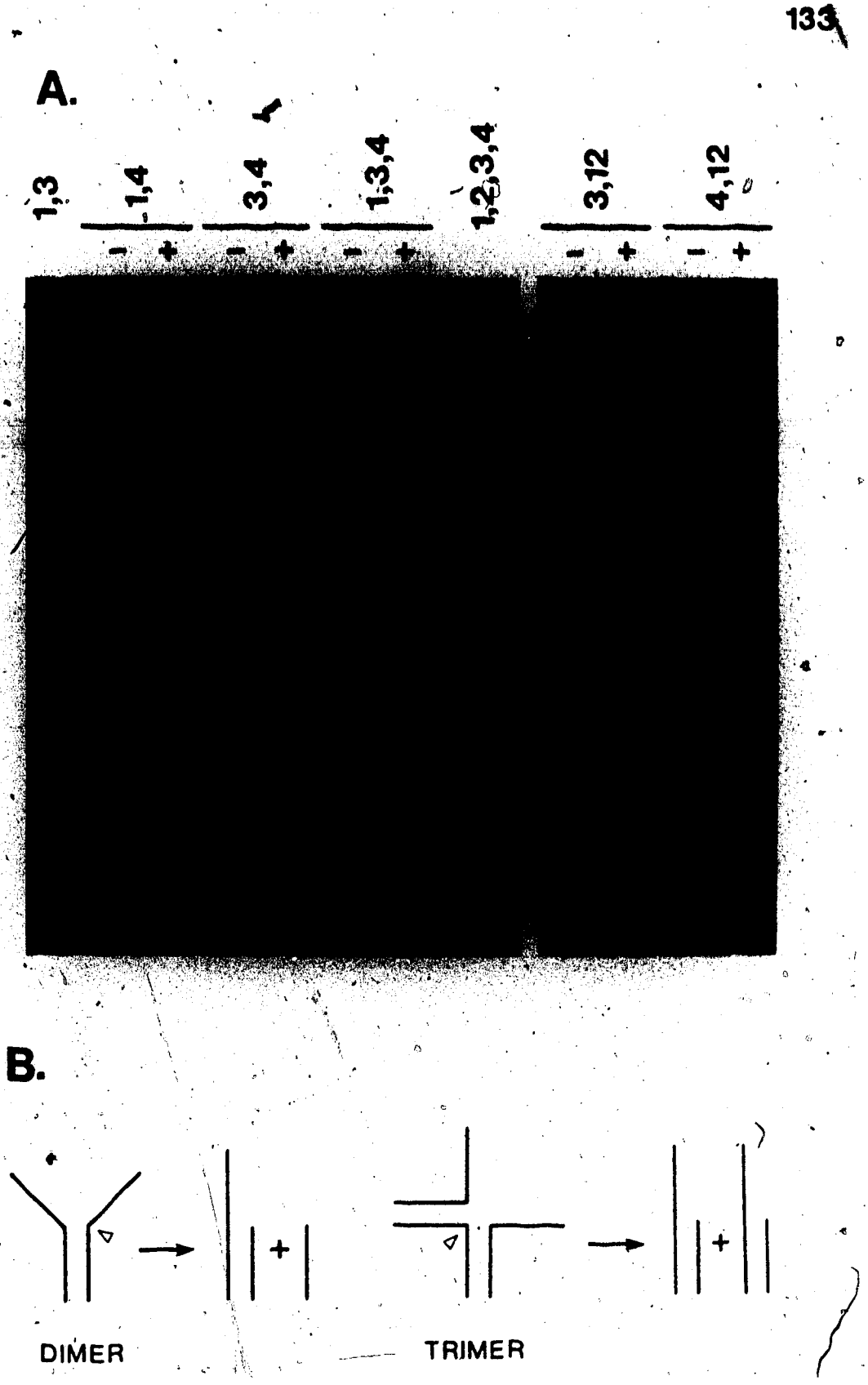

A.
B.
C.
0 0.2 0.5 1.0
0 0.2 0.5 1.0
0 10 20 60
3+4
12
Tetramers
Trimers
Dimers
Monomers

Figure III.5. Cleavage of trimer and dimer oligonucleotide complexes with T7 endonuclease I. Various combinations of oligonucleotides (denoted by the numerals in the lane headings, 150ng each) were annealed in T7 endonuclease I buffer such that dimeric and trimeric branched structures with single-stranded arms would form. Panel A: samples of each were either undigested (-) or were digested with 0.5µl enzyme for 30 minutes (+) under standard conditions. Products were electrophoresed in 15% polyacrylamide gels containing TBE plus 5mM $MgCl_2$, at 4°C and the gels were stained with ethidium bromide. Included as markers are untreated immobile Holliday crossover analog (tetrameric complex of oligos 1,2,3,4) and the noncomplementary oligonucleotides 1 and 3 which migrate as monomer forms. Panel B: hypothetical structures are illustrated for the dimer and trimer complexes as well as the expected cleavage products for each. The proposed cleavage positions are noted by the arrowheads.


A.
1,3
1,4
3,4
1,3,4
1,2,3,4
3,12
4,12
- + - + - + - + - +
B.
+
+
DIMER
TRIMER

may be regarded as having one duplex branch and two single-stranded branches, is shown in Figure III.5, as well.

Cleavage of DNA Branched Structures by T7 Endonuclease I T7 endonuclease I was titrated using cruciform-containing plasmid DNA as the substrate (see Materials and Methods, this chapter). The endonuclease activity demonstrated towards this substrate was comparable to that towards the preferred single-stranded substrates originally described by Sadowski (1971). Cleavage of the substrate DNA was specific for the cruciform structure (see Chapter II) and the resulting nicks were resealable with T4 DNA ligase as indicated by the formation of covalently-closed DNA (determined fluorimetrically; Morgan *et al.*, 1979) (not shown). The enzyme preparation was thus regarded to be free of exonucleolytic activity under the imposed assay conditions. Cleavage of the three different types of branched DNA structures by T7 endonuclease I was observed with the same high level of activity.

The three duplex DNA branch junctions created by the specific combination of various oligonucleotides (Fig. III.1) were cleaved by T7 endonuclease I at more than one site per molecule (Fig. III.4). The fixed-crossover Holliday junction analog was reduced to products which were electrophoretically equivalent to oligonucleotide dimers and which stained intensely with ethidium bromide (Fig. III.4A). This observation is consistent with the "resolution" of the structure by the pair-wise cleavage of opposing strands, ie. cleavage across either axis of the crossover junction, yielding two nicked, linear duplexes half the size of the original molecule (see also Fig. III.12A). Opposing strands in the fixed-crossover Holliday junction would be either oligos 1 and 3, or oligos 2 and 4 (Fig. III.1A). The mobile-crossover Holliday junction was similarly cleaved to products of

dimer size (Fig. III.4B). The Y-junction proved to be significantly less sensitive as a substrate since greater amounts of enzyme were required to achieve significant cleavage in the allotted time. The products of this reaction included species smaller in size than oligonucleotide dimers (one and one-half the size of individual oligonucleotides as illustrated in Fig. III.5B) and thus the introduction of at least two nicks is indicated. For each of the described reactions, it is sufficient that only two nicks be placed near the respective junctions to generate the observed products of cleavage by T7 endonuclease I.

Lower order combinations of oligonucleotides, that is, structures expected to have single-stranded branches, are also cleaved by T7 endonuclease I. Figure III.5 demonstrates the susceptibility of various dimer and trimer combinations. The reactions described contain 150ng of each of the indicated oligonucleotides and 0.5µl of enzyme. Dimers composed of oligos 1 and 4 and that of oligos 3 and 4 were not cleaved in this experiment. However, other dimers (3+12 and 4+12, Fig. III.5) were clearly susceptible. In the latter example, products migrated between monomer and dimer oligonucleotide combinations as if one of the two single-stranded branches was cleaved from these structures. Far more susceptible than dimer complexes were trimer complexes. Only one example is shown in figure III.5A, but all trimer complexes were readily cleaved under identical conditions (see also Fig. III.11). Once again, the products of the reaction migrated as molecules with an estimated size of one and one-half oligonucleotides. A single nick on the fully duplexed strand of trimer molecules (Fig. III.5B), placed near the branch point, would produce two partial duplexes of this size.

These reactions were carried out using relatively high concentrations of oligonucleotides in order to clearly visualize the products of the individual reactions in ethidium stained

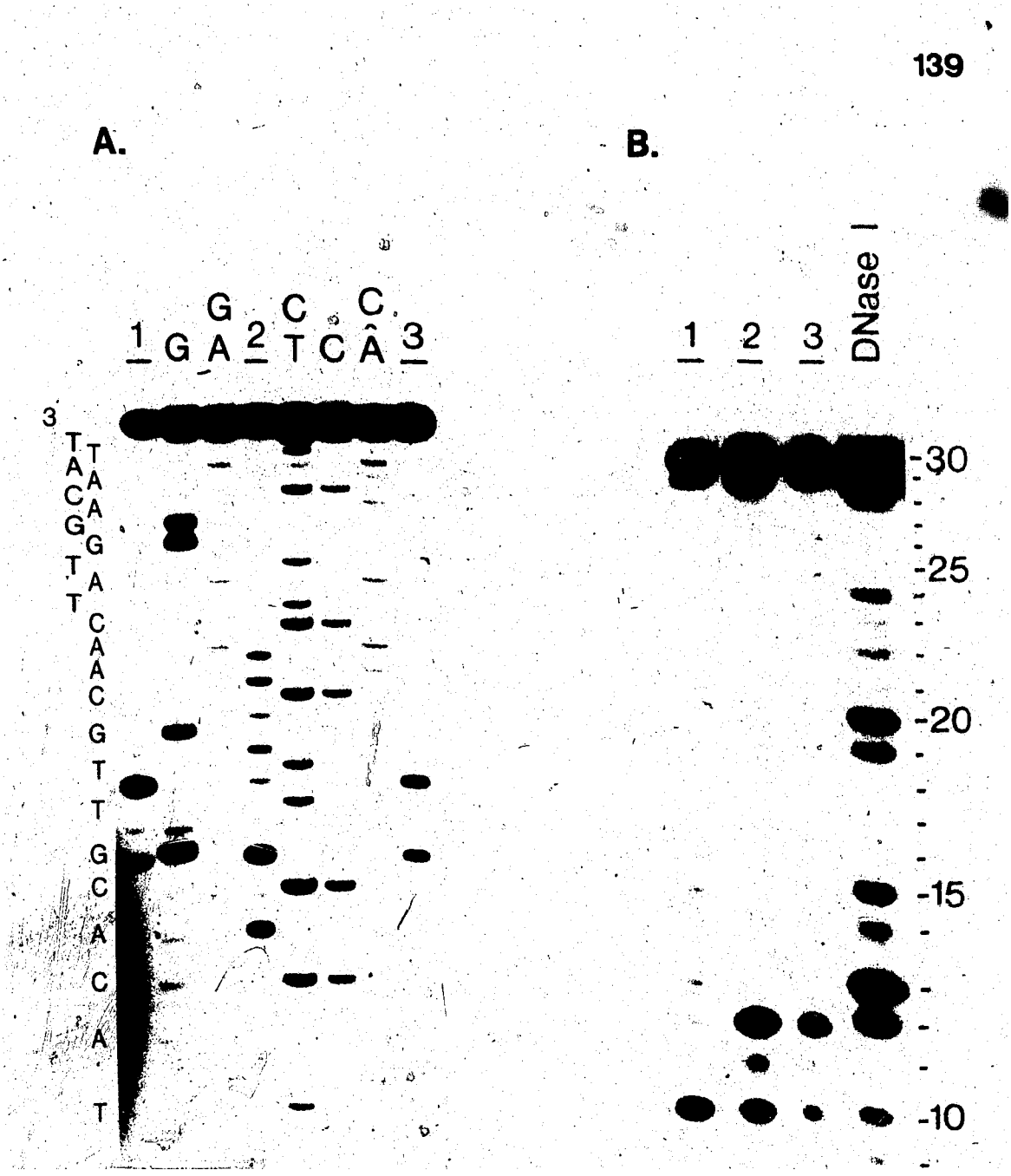
gels. In a typical reaction containing 600ng of oligonucleotides (such as in Fig. III.4A), and assuming that the oligonucleotides are completely complexed in the expected Holliday junction form, it was estimated that one unit of enzyme would cleave 0.03nmole of junction in 10 minutes, comparable to its activity on plasmid DNA. In any one reaction, it is likely that a significant concentration of lower order complexes containing non-duplexed strands, such as trimeric, dimeric and monomeric forms, each apparently with its own characteristic sensitivity, was present. It was observed that with higher concentrations of oligonucleotides, the enzyme exhibited less activity towards the fully-duplexed complexes, necessitating the use of larger amounts of enzyme. Not surprisingly, when T7 endonuclease I was titrated against the cruciform-containing plasmid substrate pSAlB.56A in the presence of increasing concentrations of oligo 3, enzymatic cleavage of the plasmid DNA gradually decreased (see Appendix 2). The presence of the single-stranded oligonucleotide was inhibitory. The nature of the inhibition was not investigated further, although it has been shown (see Fig. III.10) that individual oligonucleotides were not cleaved by T7 endonuclease I under standard reaction conditions.

Mapping Cleavage Positions in Oligonucleotide Complexes Cleavage sites were mapped on each component oligonucleotide present in the three different duplex-DNA branched forms of Figure III.1, as well as for selected lower-order branched forms. Oligonucleotide complexes were formed, as described, wherein one of the component oligonucleotides was 5' end-labelled. After treatment with T7 endonuclease I the products of the reaction were run out on sequencing gels to identify the size of the cleavage products and thus the position of the cleavage site with respect to the 5' end of the individual oligonucleotides. Definition of the cleavage position at the nucleotide level

was accomplished by comparing the mobilities of the product fragments to fragment ladders generated by sequence analysis and DNase I digestions.

Oligo 3 is common to all three duplex DNA branch structures shown in Figure III.1. In Figure III.6, the cleavage products of oligo 3, as a component of the three primary complexes, are shown. In panel A (Fig. III.6), the various product fragments are compared to a Maxam and Gilbert sequencing ladder. The largest and most predominant product fragment resulting from the cleavage of the fixed-crossover Holliday junction (Fig. III.6A, lane 1) and the Y-junction (Fig. III.6A, lane 3) corresponded to cleavage after the twelfth nucleotide, a guanosine residue, (counting from the 5' end) of oligo 3. This was deduced by an adjustment of one and one-half to two base positions in the sequencing ladder to compensate for destruction of one base inherent in the Maxam and Gilbert chemical treatment and the extra phosphate on the 3' end of the ladder fragments which increases the relative mobility of the fragments. Lengths of the oligo 3 product fragments were verified by comparison to a fragment ladder generated by DNase I digestion as shown in Figure III.6B. Counting down the ladder starting at 30 nucleotides (the length of oligo 3), the major cleavage fragment for the fixed crossover Holliday analog (Fig. III.6B, lanes 2 and 3) was determined to be twelve bases long. As demonstrated in Figure III.6, there was no difference in the patterns of oligo 3 cleavage generated by the digestion of gel-purified complex and the digestion of complex formed in complete mixtures (Fig. III.6B, lanes 3 and 2, respectively). The primary cleavage site falls between the nucleotides G and T in the sequence 5'-GT in oligo 3, one nucleotide removed to the 5' side of the crossover phosphodiester. A secondary site was identified in the sequence 5'-AC, three nucleotides to the 5' side of the crossover. The intervening site (5'-CG) was poorly cleaved. Not surprisingly, oligo 3 was cleaved at multiple

Figure III.6. Cleavage of oligonucleotide #3 in T7 endonuclease I digested branched structures. The three different branched structures of Figure III.1 were formed with ^{32}P-labelled with 5' end-labelled oligonucleotide 3 in T7 endonuclease I buffer. The complexes (composed of approximately 10ng of each oligo) were digested with varying amounts of T7 endonuclease I under standard conditons for 45 minutes and were electrophoresed in standard sequencing gels containing either 12% (B) or 20% (A) polyacrylamide. Panel A: the cleavage products of oligo 3 present in the fixed-crossover Holliday junction (1, digested with 0.1µl enzyme), the mobile-crossover Holliday junction (2, digested with 0.2µl enzyme) and the Y-junction (3, digested with 0.2µl enzyme) are compared to an oligo 3 sequencing ladder prepared by the method of Maxam and Gilbert (1980). Labelled oligonucleotide 3 used in this experiment was repurified from 20% denaturing polyacrylamide gels. The nucleotide marked (C, T) should be C according to the prescribed sequence. However, the C-band was faint, making the assignment suspect (see also Fig. III.8). Panel B: the cleavage products of oligo 3 forming part of the mobile-crossover Holliday junction (lane 1) or the fixed-crossover Holliday junction (lane 2), treated as in panel A, are compared to a fragment ladder of oligo 3 digested with DNase I. In addition, gel-purified fixed-crossover Holliday junction, end-labelled on oligo 3 (see Materials and Methods, this chapter), was digested with T7 endonuclease I (lane 3).


A.
B.
1
G
A
2
T
C
Â
3
DNase I
-30
-25
-20
-15
-10

positions in the mobile-crossover Holliday structure. A marked preference was shown for the same 5'-AC position 10 nucleotides from the 5' end of the molecule and also the 5'-AC position 8 nucleotides from the 5' end of oligo 3. In all, eight cleavage positions were identified, extending throughout the region of dyad symmetry, with the notable exception of the 5'-CG phosphodiester at position 11. This range of cutting was expected given the unrestricted branch migration in this molecule and enzyme cleavage in the vicinity of the crossover.

In these reactions, and others to be described later, the enzyme was titrated to effect near complete digestion of the duplex DNA branched complexes. As estimated from the gels shown in Figure III.6, less than 50% of oligo 3 was cleaved in these reactions. For the Holliday structure analogs, 50% might well represent the maximum achievable digestion of oligo 3 since it would be fragmented only when the appropriate axis was cleaved. Moreover, a certain proportion of oligo 3 was assumed to exist in complexes that were much less sensitive to cleavage by the enzyme (see later).

The cleavage sites in all component oligonucleotides for the three primary structures can be discerned from the data presented in Figure III.7. In the fixed-crossover Holliday junction, the primary cleavage position in all the constituent oligos was after the twelfth nucleotide (from the 5' end). The same was true for the Y-junction, where the various product fragments can be compared to the characteristic fragments generated from oligo 3. Due to sequence differences, oligonucleotide fragments of the same length do not necessarily co-migrate. The oligo 4 fragment, which was twelve nucleotides long, migrated slower than fragments of the same size generated by the cleavage of oligos 3 and 12. A broad cleavage pattern was displayed for all four oligonucleotides which composed

Figure III.7. T7 endonuclease I cleavage positions on all component oligonucleotides of the branched DNA structures. The three duplex DNA structures described in Figure III.1 were formed in T7 endonuclease I buffer, singly labelled at the 5' end of each component oligonucleotide, and digested with T7 endonuclease I under standard conditions for 45 minutes. The amount of enzyme used to digest each complex was as indicated for Figure III.6. The products of each reaction were electrophoresed in standard 12% polyacrylamide sequencing gels as described. The numerals in the lane headings designate the labelled oligonucleotide, in each case. The arrows designate fragments that are 12 nucleotides long, as determined separately for each oligonucleotide.

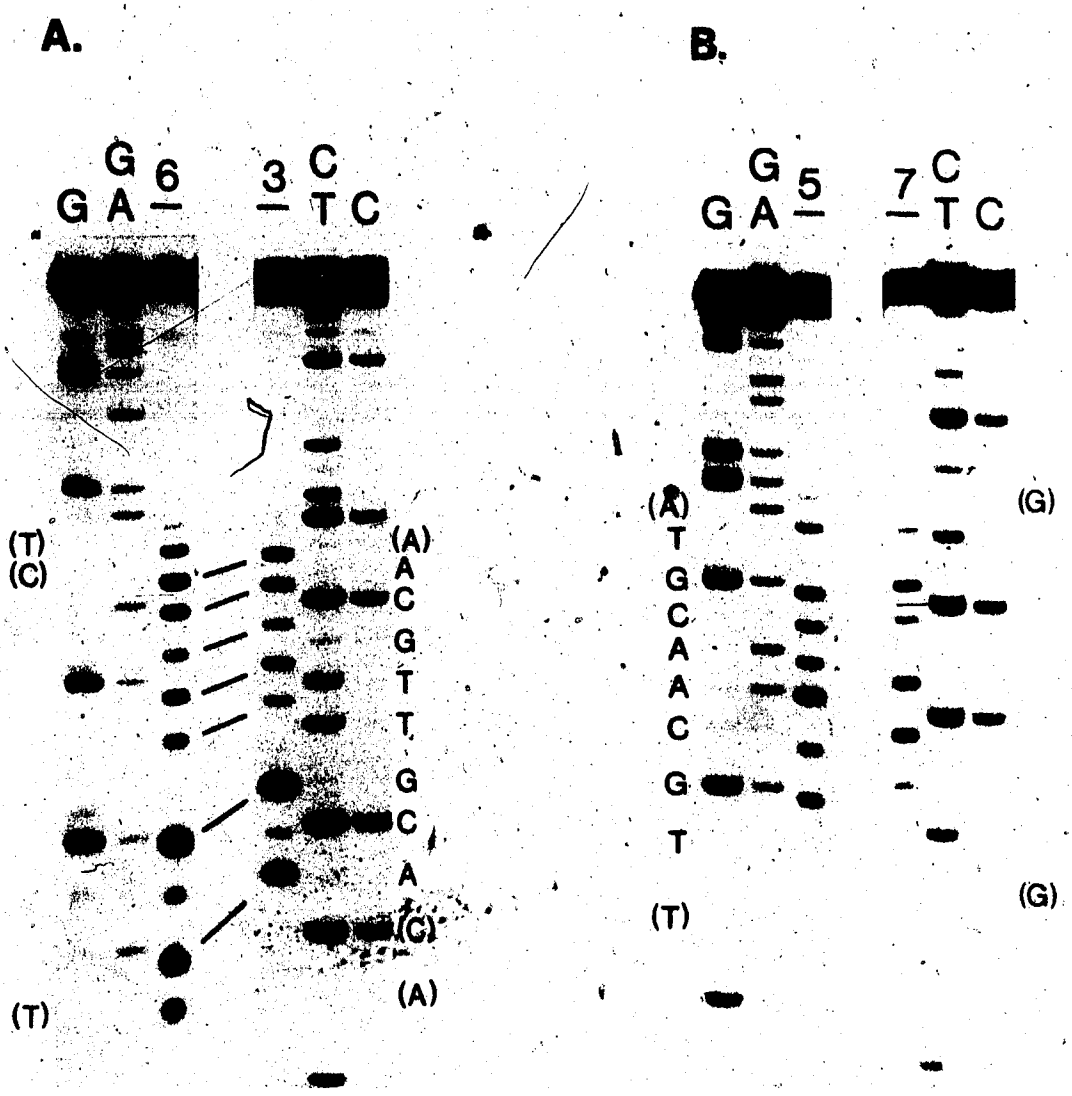
Immobile Crossover			
3	1	2	4

Y-junction		Mobile Crossover			
3	4	12	6	3	7	5

the mobile-crossover Holliday junction, and the cleavage positions are clearly related to the region of dyad symmetry in the molecule and, by inference, the crossover position. The cleavage patterns for oligonucleotides 3 and 6 are similar, as are those for oligonucleotides 5 and 7. These comparisons are most clearly made in Figure III.8, which shows the cleavage profiles of oligonucleotides of 95% purity in polyacrylamide gels with greater resolving power. The two pairs of oligonucleotides lie across, and thus represent, the two axes of the crossover junction in this Holliday structure. Resolution of the Holliday structure would require the placement of symmetrically disposed cuts on opposing strands, within or near the limits of dyad symmetry, and this action would be expected to generate similar cleavage patterns for the diagonally opposed pairs of oligonucleotides. The comparisons shown in Figure III.8 emphasize that this was, in fact, observed. Interestingly, there was a reproducible and significant difference in the manner in which the sequence 5'-TGC was cleaved in oligos 7 and 5. Included in Figure III.8 is one-half of each sequencing ladder to assist in the alignment of cleavage sites with the nucleotide sequence. Both axes of the mobile-crossover and the fixed-crossover Holliday junction were cleaved with equal efficiency since, for each junction, all four oligonucleotides appear to be cleaved equally, or very nearly so. Note, as well, that at the position of the presumed base mismatch in the branch formed by oligo 5 and oligo 3 there was no cleavage of either oligonucleotide (at asterisk in Fig. III.8).

The relationship between the cleavage pattern of each constituent oligonucleotide in the mobile-crossover junction and the nucleotide sequence is demonstrated in Figure III.9, which highlights the observation that certain nucleotide positions within, or adjacent to, the region of dyad symmetry, were preferentially cleaved. For instance, T7 endonuclease I has a certain propensity to cleave the sequence 5'-AC, and to a lesser extent 5'-GC.

Figure III.8. T7 endonuclease I cleavage patterns in oligonucleotides composing the mobile-crossover Holliday junction. The mobile-crossover Holliday junction (Fig. III.1C) was formed, labelled and digested as described in Figure III.7. All four labelled oligonucleotides had been repurified from 20% denaturing polyacrylamide gels prior to these experiments. The cleavage pattern for each labelled component oligonucleotide (as part of the Holliday structure) is shown after electrophoresis in standard 20% polyacrylamide sequencing gels. Panel A: oligos 6 and 3 constitute a pair of opposing strands of the Holliday structure falling on one cleavage axis of the crossover. Their product fragments are compared to partial sequencing ladders of each to assist in the alignment of the cleavage positions to the sequence in and surrounding the region of dyad symmetry. Panel B: cleavage patterns for oligos 5 and 7, which constitute the second cleavage axis, are compared. The sequences shared by the pairs of oligonucleotides are shown next to the sequence ladders and those portions of the sequences which are unique to each strand are included in parentheses. The asterisk marks the position in oligonucleotide 5 of the misincorporated nucleotide (see Fig. III.1).


A.
G
G A 6
3 C
T C
(T)
(C)
(A)
A
C
G
T
T
G
C
A
(C)
(A)
(T)
B.
G
G A 5
7 C
T C
(A)
T
G
C
A
A
C
G
T
(T)
(G)
(G)

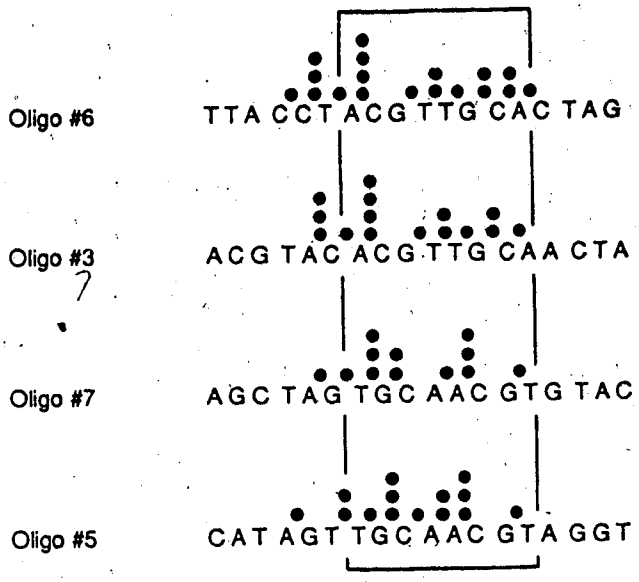


Figure III.9. Sequence specific cleavage of semi-mobile Holliday crossovers. The relative cleavage efficiency at positions within the four oligonucleotides that make up the mobile-crossover Holliday junction are indicated by the filled circles (based on results described in Figures III.7 and III.8). The sequences shown include the region of dyad symmetry (contained within the boxed area) and the immediately flanking sequences. No cleavage positions were ever identified beyond the sequences shown.

Cleaved poorly are the sequences 5'-CG and 5'-CA. This observation is in agreement with Sadowski's original report (1971) that the enzyme primarily leaves pyrimidines at the 5' ends of cleaved DNA strands. Moreover, the pattern of preference is very reminiscent of the sequence preference exhibited by T4 endonuclease VII (Mizuuchi *et al.*, 1982; see Discussion, this chapter). This helps to explain the appearance of a second site

of cleavage for oligo 3 in fixed-junction Holliday molecules (Fig. III.6). The secondary cleavage position is the highly preferred 5'-AC site identified in the mobile-crossover Holliday junction. The disposition of the cleavage site with respect to the crossover position is substantiated by the data presented in Figure III.9. In general, the sites of cleavage are restricted to a region which is offset one nucleotide to the 5' side of the regions of dyad symmetry, consistent with the predominant cleavage position being one nucleotide removed from the crossover on the 5' side.

Cleavage sites were also mapped on selected lower order complexes that were readily cleaved. Shown in Figure III.10 are the reaction products of digested complexes composed of all possible oligonucleotide combinations related to the fixed junction Holliday structure and specifically containing oligo 4 as the labelled strand. Oligo 4 was not appreciably cleaved when it was present alone or in combination with any other single oligonucleotide. However, it was cleaved in a trimeric combination with oligos 1 and 3 (Fig. III.10). The cleavage pattern was identical to that obtained when oligo 4 was part of the tetrameric Holliday structure. No other component oligonucleotide of the fixed-crossover Holliday junction was visibly cleaved, individually or in a dimeric complex under the same reaction conditions (not shown). Included in Figure III.10 (right-hand most lane), are the product fragments of oligo 4 produced from the cleavage of the fixed-crossover Holliday junction with five times the usual amount of enzyme. Even with this excess amount of enzyme, cleavage is predominantly limited to the 5' side of the crossover. A significant amount of fragments 13 and 11 nucleotides long were generated in these reactions. Analysis of the sequence shows that these preferred cleaved sequences are 5'-AC and 5'-GC sites, respectively.

Figure III.10. T7 endonuclease I cleavage of oligonucleotides in lower-order complexes. All possible oligonucleotide combinations, labelled at the 5' end of oligo 4, were formed in T7 endonuclease I buffer and digested with 0.2µl of T7 endonuclease I (except 1.0µl enzyme in lane 4,1,2,3*) under standard conditions for 45 minutes. Products were electrophoresed in a standard 12% polyacrylamide sequencing gel. Lane headings designate the combined oligonucleotides, for each reaction. A mixture of all four oligonucleotides forms the fixed-crossover Holliday junction (Figure III.1A) and lesser combinations would be expected to form structures that have single-stranded branches (refer to Fig III.5B). The arrowhead identifies the position of fragments 12 residues long.

4

4,1

4,3

4,2 .

4,1,3

4,1,2

4,2,3

4,1,2,3 .

4,1,2,3*

The cleavage of various trimeric combinations was revealing. Oligo 1 behaved the same as oligo 4 (Fig. III.11, lane 1, 2, 4), in that only when it constituted the fully-duplexed partner of a trimeric complex was it cleaved. Oligo 3 was cleaved when it was the fully-duplexed partner (Fig. III.11, lane 3,2,4) and, to a lesser but significant extent, it was cleaved when in association with oligos 1 and 2. In the latter example, the cleavage site was assigned to the 5' position (one nucleotide away from the crossover) in the duplex arm of the complex. Titration of the enzyme to lower concentrations prevented the cutting observed when oligo 3 was in combination with oligos 1 and 2 while, at the same time, it was still cleaved when combined with oligos 2 and 4 (not shown). Generally, it was observed that partially-duplexed oligonucleotides (like oligo 3 in association with oligos 1 and 2) were weakly cleaved and only at the 5' site (one nucleotide removed from the branch point) present in the duplex arm of the molecule (observed in Fig. III.11, lanes 4, 2, 3 and 1, 3, 4 as very weak reactions). At the levels of enzyme used, cleavage in single-stranded regions was never observed. Oligo 2 was inefficiently cleaved in these reactions, possibly because it has substantial self-complementarity and could self-anneal in Endo I buffer (not shown).

D. Discussion

Tsujimoto and Ogawa (1978) have proposed that T7 endonuclease I functions during the replicative cycle of T7 bacteriophage by trimming branch junctions that arise in phage DNA recombination intermediates. They believed that the activity of the enzyme was general, cleaving three-way as well as four-way junctions and cleaving simple junctions where single-stranded regions were present at the fork. The observations made in this study are in definite agreement with their hypothesis. A variety of branched DNA forms

Figure III.11. T7 endonuclease I cleavage positions in trimer complexes. All possible trimer complexes (stylized structures are shown in Figures III.5B and III.12C) formed from three of the four oligonucleotides that compose the fixed-crossover Holliday junction (Figure III.1A) were digested with 0.1µl of T7 endonuclease I under standard reaction conditions for 45 minutes. The products were electrophoresed in standard 12% polyacrylamide sequencing gels. The combined oligonucleotides are included in the lane headings with the one labelled oligonucleotide in each complex underlined.

4,1,3
4,1,2
4,2,3
3,2,4
3,1,4
3,1,2
2,1,3
2,1,4
2,3,4
1,2,4
1,3,4
1,2,4

can be cleaved by T7 endonuclease I, though they differ significantly in their sensitivity. The best substrates described in this paper were the trimeric complexes which possess two duplex arms and two single-stranded arms and the tetrameric fixed-crossover Holliday junction. Less efficiently cleaved were the tetrameric mobile-crossover Holliday junction, the fully duplexed three-way junction (Y-junction), and finally the various dimer complexes which may be regarded to have one duplex arm and two single-stranded arms. In light of the failure to degrade individual oligonucleotides and to further degrade the duplex products of the Holliday structure resolution reaction (Fig. III.4), non-branched forms, be they duplex or single-stranded, do not serve as substrates or are relatively poor ones. This observation also confirms the absence of exonuclease activity in the enzyme preparation.

A specificity for branched DNA substrates is substantiated by the remarkable cleavage site specificity displayed by the enzyme. For each DNA branched structure tested, the evidence strongly indicated that the principle cleavage site was in duplex regions one nucleotide removed to the 5' side of the branch point. Five unique oligonucleotides composing two structurally distinct fixed-junction complexes were cleaved with this specificity. Cleavage further from the 5' side of the junction was observed, but consistently at much reduced levels. The enzyme exhibited a moderate degree of sequence specificity, like T4 endonuclease VII (Kemper *et al.*, 1985). Sequence specificity was identified from the cleavage patterns of oligonucleotides composing the mobile-junction Holliday complex. It was assumed that the crossover in these molecules had an equal, or near-equal, probability of existing at any phosphodiester within the region of dyad symmetry. The presence of a preferred cleavage sequence one or two nucleotides further removed from a fixed junction promoted cleavage at these additional

positions. However, cleavage at these sites was always less efficient than cleavage at the site one nucleotide removed from the 5' side of the crossover even when the latter site possessed a non-preferred sequence as defined in this paper. Oligo 3, for example, was predominantly cleaved at the 5'-GT position one nucleotide away from the branch point despite the presence of the much preferred 5'-AC sequence three nucleotides away (Fig. III.6). The sequence 5'-CA is very poorly cleaved in mobile-crossover Holliday junctions (see Fig. III.9), yet oligo 4 of the fixed-crossover Holliday junction was readily cleaved at this sequence which constituted the 5' position one nucleotide removed from the crossover (Fig. III.7). Thus, structural specificity appeared dominant over sequence specificity, at least with this substrate. The site-specificity of cleavage by T7 endonuclease I seems more precise than that observed with T4 endonuclease VII which cleaves from 1 to 5 bases away from the crossover (Kemper *et al.*, 1985; Jensch and Kemper, 1986). There was even more variability in the cleavage of Y-junctions with the T4 enzyme, but it is not clear that the junctions tested were strictly immobile (Jensch and Kemper, 1986). The results with T7 endonuclease I are consistent with the published cleavage positions mapped at cruciform crossover junctions. Numerous reports have indicated that T7 endonuclease I cleaved one to two nucleotides to the 5' side of the cruciform crossover (Panayotatos and Wells, 1981; de Massy *et al.*, 1985; Haniford and Pulleyblank, 1985; Naylor *et al.*, 1986). Deviation from this specificity, particularly with substrates able to branch migrate, may arise due to due the sequence specificity of the enzyme (de Massy *et al.*, 1987). Sequence preferences may also account for the differential cleavage of dimeric complexes.

While the susceptibilty of other non-B DNA structures was not specifically assessed, it was perhaps fortuitous that an unexpected sequence error in oligo 5 generated a base

mismatch outside the region of dyad symmetry in the branch-migrating Holliday structure. This served as an internal control and demonstrated, given the absence of cleavage at this position, that mismatched basepairs or extrahelical bases (Evans and Morgan, 1982) were not targets for enzyme action, at least at the levels of enzyme used. The redundancy of the data, given the various classes of structure studied and the precision of the cleavage with respect to the crossover position, argue strongly against the involvement of DNA structural elements other than a branch point, as already suggested, in the determination of observed cleavage positions. All secondary cleavage positions (those further than one nucleotide from the crossover), can be ascribed to preferred sequences present at these locations. Although sequence dictated a modest degree of freedom in the selection of cleavage positions, cleavage was invariably related to the position of the crossover.

T4 endonuclease VII also exhibits sequence specificity (Kemper *et al.*, 1985). Like T7 endonuclease I, it efficiently cleaved the bases 5'-AC but failed to cut between 5'-CA. Likewise, 5'-CG was poorly cleaved, though 5'-GC was cleaved well. The similarities between the two enzymes are striking and extend beyond this limited comparison. The pattern of cleavage observed here for T7 endonuclease I is also consistent with the original observation by Sadowski (1971) that cleavage with T7 endonuclease I preferentially left pyrimidines at the 5' ends of the cut DNA. In the case of the cleavage of dimer complexes, the variable sensitivity of chosen dimer molecules (Fig. III.5) may, in fact, be related to this sequence preference. Kemper *et al.* (1985) have proposed that because T4 endonuclease VII cleaved in linear double-stranded DNA regions with the same sequence specificity as its activity on Holliday structures, cleavage specified by the presence of the crossover must occur in duplex regions as well. The same conclusion has

been derived from this work with T7 endonuclease I except with a different experimental approach, namely the differential cleavage of component oligonucleotides in trimer complexes. While the enzyme appears to be able to cleave at junctions with single-stranded branches, the sensitive 5' cleavage site appears to fall exclusively in the duplex branches of these complexes. No evidence was found for cleavage in single-stranded regions.

The results presented can be reconciled with published accounts of the single-strand specific endonuclease activity of T7 endonuclease I. Titration of the enzyme with cruciform-containing plasmid DNA as the substrate strongly suggested that branched, double-stranded DNA is as efficient a substrate as is single-stranded fd phage DNA (Sadowski, 1971). Early reports (Center and Richardson, 1970) had indicated that double-stranded substrates were 100 to 200 times less efficiently cleaved. It is reasonable to assume that because of the comparably high activity levels demonstrated towards duplex branch DNA substrates that the fd DNA substrates, as well as the other single-stranded substrates originally tested, contained sufficient secondary structure to provide recognizable junctions upon which the enzyme was acting. A similar explanation was proposed by de Massy *et al.* (1987).

Analyses of the cleavage of the two Holliday structure analogs indicate no obvious preference in the choice of cleavage axis by T7 endonuclease I. Potentially some bias could have been demonstrated, considering the modest sequence specificity of the enzyme. Reviewing the 5' sites surrounding the junction in the immobile Holliday structure, there was no reason to predict a preference for one axis over the other. Neither the mobility of the junction nor the extent of dyad symmetry has influenced the chosen

axis of cleavage.

Recognition, by resolving enzymes, of crossed or uncrossed strands in Holliday structures (Fig. III.12B) has generally been invoked to explain the selection of a pair of strands which, when cleaved, would effect resolution (Sigal and Alberts, 1972; Meselson and Radding, 1975; Kemper *et al.*, 1985). In essence, the resolving enzyme would have to recognize a global feature of the four-way junction. However, it has already been recognized that this structural element is nonexistent in Y-junctions (de Massy *et al.*, 1987) and it does not exist in the trimeric complexes studied here. Therefore, based on the substrate profiles developed in this study, it is proposed that T7 endonuclease I recognizes, as the basic elements of a branch junction, a distortion in the continuous phosphodiester backbone of a DNA strand bridging two branches and, minimally, one duplex DNA branch in which must lie the 5' end of the bridging DNA strand (since the cleaved bond lies in duplex region on the 5' side of the junction). A bridging strand bordered by two duplex branches, typified by a trimeric complex (Fig. III.12C), makes a much better substrate. Holliday structures like the tetrameric complexes studied, when represented in a simple form (Fig. III.12A), can be seen to have four equivalent (ignoring sequence) and independently accessible sites of cleavage which fit the hypothetical binding site requirements modelled upon the trimer complex in Figure III.12C. A planar, or perhaps a tetrahedrally arranged, Holliday structure (Meselson and Radding, 1975; Gough and Lilley, 1985; Robinson and Seeman, 1987) would ideally suit this model because of the presence of four pseudo-equivalent domains meeting the proposed criteria.

The effective resolution of Holliday structures to nicked-duplex daughter molecules implies the placement of coincidental nicks on two opposing strands of the crossover

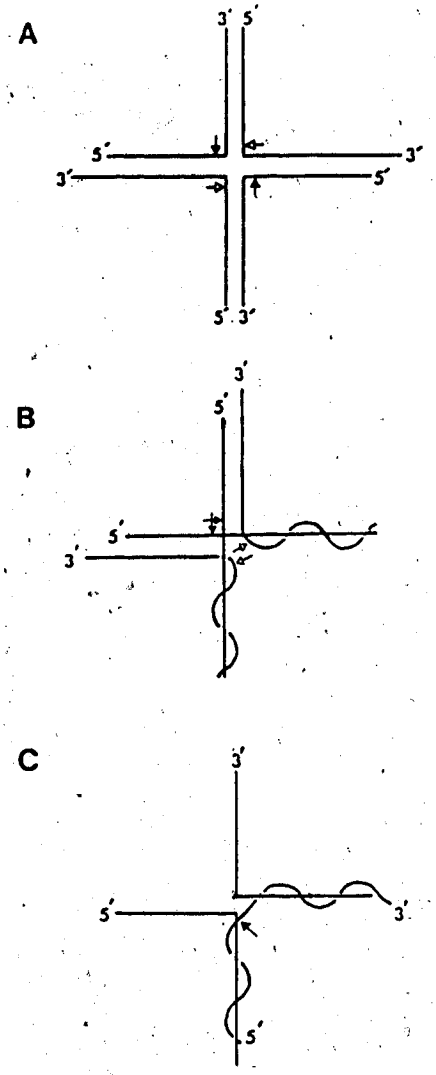


Figure III.12. Disposition of T7 endonuclease I cleavage sites in branched DNA molecules. The Holliday structure (Panels A and B) and the structure of a trimer complex (Panel C) are shown in stylized form. Diagrams A and B differ only in perspective, were the branches of the Holliday structure to adopt a tetrahedral type of arrangement. The two cleavage axes are denoted by the pairs of filled and open arrows. In A, the equivalence of the four cleavage positions is emphasized. In B, with the concept of crossed and uncrossed strands imposed, the alignment of pairs of sites is emphasized and, therefore, two non-equivalent enzyme binding domains are described. In C, the effective reduction of the trimer complex to a non-branched form can be accomplished by one nick and this site is shown. The use of curved and straight strands is solely to assist in the differentiation of the strands.

junction. The involvement of a homodimeric resolving enzyme is a reasonable proposal (Kemper *et al.*, 1985), but the fidelity of resolution could be equally well ensured by the negative cooperative interaction of binding site exclusion. The binding of a catalytic moiety to one cleavage position of a planar Holliday structure, for example, along a chosen axis of the junction (at an open or filled arrow in Figure III.12A), probably involves contacts with the two bordering duplex branches (in order to recognize the bridging strand), as previously argued. This probably would exclude binding at the two neighboring sites, which together define the opposite cleavage axis. Consequently, the only remaining site flanked by unbound duplex DNA would be on the opposing strand along the same axis. A coincidental nick here by a second catalytic unit would effect resolution. Two observations are consistent with an alternative hypothesis that the enzyme-mediated resolution of synthetic Holliday structures may occur without a cooperative interaction between enzyme molecules aside from the proposed binding site exclusion. First, it is clear from the cleavage of trimer complexes (at crossovers between neighboring duplex DNA branches, see Fig. III.12C) that single, specific cuts can be made by the enzyme. Cleavage, in this case, was no less efficient than it was for the synthetic Holliday junctions which might have otherwise benefited, in kinetic terms, from the cooperative interaction of two DNA-bound catalytic units. Second, is the observation of differences in the cleavage patterns observed on opposite strands of the mobile-crossover Holliday junction (Fig. III.8 and III.9). For instance, the sequence 5'-TGC is cleaved with a different specificity in oligo 7 than is the same sequence in oligo 5. The context of the sequences, particularly the region 5' to it which is unique for each strand, may have influenced the specificity by altering the binding of the enzyme. This suggests some independence in the binding and cleavage on opposing strands during resolution, ie. there is freedom for the two simultaneously-acting enzyme molecules to

individually position themselves at the branch junction.

Alternatively, while positive cooperativity may not be a significant factor in the binding of the enzyme, it may be important in the attainment of an activated complex with the tetrameric complexes since these structures will have limited conformational freedom at the crossover compared to trimeric complexes. Such a cooperative interaction would be impossible during the cleavage of Y-junctions, and since the conformational flexibility of these junctions may be severely restricted as well, it may not be surprising that these are much less efficiently cleaved by T7 endonuclease I.

Bibliography

Center, M. and Richardson, C. (1970). *J. Biol. Chem.* **245**, 6285-6291.

Center, M., Studier, F. W. and Richardson, C. (1970). *Proc. Nat. Acad. Sci. U.S.A.* **65**, 242-249.

DeLange, A. M. and McFadden, G. (1987). *J. Virol.* **61**, 1957-1963.

DeLange, A. M., Reddy, M., Scraba, D., Upton, C. and McFadden, G. (1986). *J. Virol.* **59**, 249-259.

de Massy, B., Studier, F. W., Dorgai, L., Appelbaum, E. and Weisberg, R. A. (1985) *Cold Spring Harbor Symp. Quant. Biol.* **49**, 715-726.

de Massy, B., Weisberg, R. A. and Studier, F. W. (1987) *J. Mol. Biol.* **193**, 359-376.

Dressler, D. and Potter, H. (1982). *Ann. Rev. Biochem.* **51**, 727-762.

Evans, D. H. and Kolodner, R. (1986). *J. Cell Biochem. Suppl.* **10B**, 213.

Evans, D. H. and Morgan, A. R. (1982). *J. Mol. Biol.* **160**, 117-122.

Gellert, M., Mizuuchi, K., O'Dea, M. H., Ohmori, H. and Tomizawa, J. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 35-40.

Gough, G. W. and Lilley, D. M. (1985). *Nature* **313**, 154-156.

Greaves, D. R., Patient, R. K. and Lilley, D. M. J. (1985). *J. Mol. Biol.* **185**, 461-478.

Haniford, D. B. and Pulleyblank, D. E. (1985). *Nucl. Acids Res.* **13**, 4343-4363.

Holliday, R. (1964). *Genet. Res.* **5**, 282-304.

Holliday, R. (1974). *Genetics* **78**, 273-287.

Hsu, P. L. and Landy, A. (1984). *Nature* **311**, 721-726.

Jensch, F. and Kemper, B. (1986). *EMBO J.* **5**, 181-189.

Kallenbach, N. R., Ma, R.-I., Wand, A. J., Veeneman, G. H., van Boom, J. H. and Seeman, N. C. (1983). *J. Biomol. Struct. Dynamics* **1**, 159-168.

Kemper, B., Jensch, F., v. Depka-Prondzynski, M., Fritz, H.-J., Borgmeyer, U. and Mizuuchi, K. (1985). *Cold Spring Harbor Symp. Quant. Biol.* **49**, 815-825.

Kemper, B. and Garabett, M. (1981). *Eur. J. Biochem.* **115**, 123-131.

Kerr, C. and Sadowski, P. D. (1975). *Virology* **65**, 281-285.

Kikuchi, A. and Nash, H. A. (1979). *Proc. Nat. Acad. Sci. U.S.A.* **76**, 3760-3764.

Lee, D. and Sadowski, P. (1981). *J. Virol.***40**, 839-847.

Lilley, D. M. J. and Kemper, B. (1984). *Cell* **36**, 413-422.

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

Maxam, A. M. and Gilbert, W. (1980). *Methods Enzymol.* **65**, 499-560.

Meselson, M. S. and Radding, C. M. (1975). *Proc. Nat. Acad. Sci. U.S.A.* **72**, 358-361.

Mizuuchi, K., Kemper, B., Hays, J. and Weisberg, R. A. (1982). *Cell* **29**, 357-365.

Morgan, A. R., Lee, J. S., Pulleyblank, D. E., Murray, N. L. and Evans, D. H. (1979). *Nucl. Acids Res.* **7**, 547-569.

Naylor, L. H., Lilley, D. M. J. and van de Sande, J. H. (1986). *EMBO J.* **5**, 2407-2413.

Panayotatos, N. and Wells, R. D. (1981). *Nature* **289**, 466-470.

Powling, A. and Knippers, R. (1974). *Mol. Gen. Genet.* **134**, 173-180.

Robinson, B. H. and Seeman, N. (1987). *Biophys. J.* **51**, 611-626.

Sadowski, P. (1971). *J. Biol. Chem.* **246**, 209-216.

Sigal, N. and Alberts, B. (1972). *J. Mol. Biol.* **71**, 789-793.

Symington, L. S. and Kolodner, R. (1985). *Proc. Nat. Acad. Sci. U.S.A.* **82**, 7247-7251.

Tsujimoto, Y. and Ogawa, H. (1978). *J. Mol. Biol.* **125**, 255-273.

West, S. C. and Korner, A. (1985). *Proc. Nat. Acad. Sci. U.S.A.* **82**, 6445-6449.

IV. CONFORMATIONAL ISOMERIZATION OF THE HOLLIDAY JUNCTION ASSOCIATED WITH A CRUCIFORM DURING BRANCH MIGRATION IN SUPERCOILED PLASMID DNA [1]

A. Introduction

Interpretive analyses of the molecular mechanisms of genetic recombination rely on the premises originally stated in the Holliday model (Holliday, 1964; Holliday, 1968; Brooker & Lehman, 1971; Sigal & Alberts, 1972; Holliday, 1974; Meselson & Radding, 1975; Dressler and Potter, 1982; Szostak *et al.*, 1983; Hsu & Landy, 1984; Robinson & Seeman, 1987). Central to this model is the definition of a Holliday structure as the branched DNA complex formed by the exchange of single strands between two homologous duplexes (Holliday, 1964 and 1968). Several refinements have since been incorporated into the model describing more rigorously the possible orientations of the branches about the junction and their functional isomerization (Sigal & Alberts, 1972; Sobell, 1974; Meselson & Radding, 1975; Robinson & Seeman, 1987). Basically, three conformations for Holliday junctions have been proposed. In one version, postulated by Sigal and Alberts (1972), the branches are base-stacked and co-planar such that the two recombining molecules have unperturbed helix axes (referred to as the UHA structure by Robinson & Seeman (1987)). Another representation views it as a planar structure, but not base-stacked, with the branches at right angles to each other such that the structure possesses pseudo-four-fold symmetry (Sobell, 1974). Lastly, an intermediate

1. A version of this chapter has been submitted for publication. Dickie, P., Morgan, A. R. and McFadden, G. 1987. J. Mol. Biol.

conformation, wherein the branches of the junction are arranged in a tetrahedral fashion, has been postulated (Gough and Lilley, 1985). The only physical evidence purporting to discriminate between the various forms has come from electron microscopy (Dressler & Potter, 1982) and the electrophoretic behaviour of Holliday-like structures (Gough & Lilley, 1985). Given the likely conformations of the junction, descriptions of the mechanics and dynamics of branch point migration have also been provided (Meselson, 1972; Warner *et al.*, 1978; Thompson, *et al.*, 1986; Robinson & Seeman, 1987).

With respect to the structural dynamics of Holliday junctions; actual experimental evidence is scarce, though there have been studies on the kinetics of branch point migration in topologically unconstrained branched structures (Warner *et al.*, 1978; Thompson *et al.*, 1986). The process of branch migration is biologically significant in that it determines the extent of heterologous DNA formed during recombination, whereas an isomerization of Holliday junctions is postulated to determine the incidence of genetic "crossing-over" during recombination (Meselson & Radding, 1975; Szostak *et al.*, 1983). Mechanistically, both these processes involve the twisting or rotation of the helical DNA branches at the junction, thus both should be influenced by the topological nature of the DNA substrate. Branch migration can be driven by topological strain (supercoiling) in a DNA molecule (Mizuuchi *et al.*, 1982b), but the coincidental impact of superhelical strain on junction conformation and isomerization has not been investigated, largely because of the unavailability of a suitable experimental system. Circular bacterial plasmids bearing large palindromic sequences that can be extruded into stable cruciforms are substrates in which branch migration can occur in response to torsional strain in the circular plasmid (Mizuuchi *et al.*, 1982b). The plasmid pSAlB:56A (Fig. IV.1) bears the telomeric sequences of Shope fibroma virus as a 322 base-pair imperfect inverted repeat (DeLange

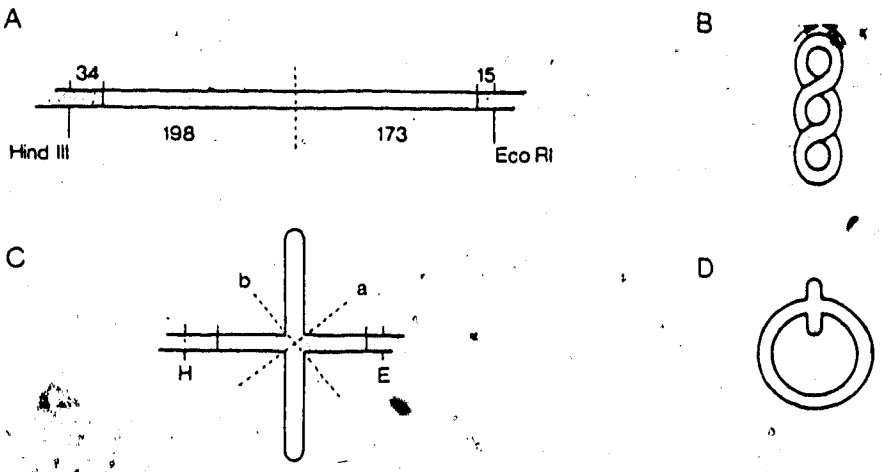


Figure IV.1. The effect of cruciformation on the topology of plasmid pSAlB.56A. Panel A and B: the lineform representation of the inverted repeat insert of Shope fibroma virus telomere sequences is shown (by arrows). Vector sequences are shaded and all distances are given in basepairs. Panel C: the cruciform configuration of the same sequences is illustrated, where 'a' and 'b' denote the two axes of the crossover junction. In panels B and D are schematic representations illustrating the topology of the plasmid when bearing either configuration of the cloned sequences. H: *Hin*dIII ; E: *Eco*RI.

et al., 1986). Cruciform extrusion from the axis of the inverted repeat (Fig. IV.1C and D) has been shown to topologically relax the plasmid (Chapter II) and, in fact, the plasmid isolated from bacteria is in the extruded cruciform, topologically relaxed, conformation. Consequently, branch migration of the cruciform crossover should occur in these molecules within the limits imposed by the superhelical energy in the topologically-closed, circular core of the plasmid.

Bacteriophage T7 endonuclease I (Center & Richardson, 1970; Sadowski, 1971) has been used to probe for the cruciform crossover in pSAlB.56A (Chapter II). It belongs to a class of recombination-specific endonucleases with a specificity for branched DNA forms. Members of this class also include Int protein (Hsu & Landy, 1984), endonucleases purified from yeast (West & Korner, 1985; Symington & Kolodner, 1985) and bacteriophage T4 endonuclease VII (Mizuuchi *et al.*, 1982a; Lilley & Kemper, 1984; Kemper *et al.*, 1984; Jensch & Kemper, 1986). The bacteriophage endonucleases have demonstrated an *in vitro* specificity that makes them potentially useful as structural probes for branched DNA junctions. In fact, several reports have included inferences about the conformation of Holliday structures based on the activity of the T4 and T7 enzymes (Kemper *et al.*, 1984; de Massy *et al.*, 1987; Chapter III).

Bacteriophage T7 endonuclease I has a broad specificity for branched DNA structures (Panayotatos & Wells, 1981; de Massy *et al.*, 1985 and 1987; Chapter III). It is site-specific, cleaving predominantly one nucleotide to the 5' side of the crossover, and shows a modest degree of preference for certain DNA sequences. The site-preference has been defined for freely-branched Holliday structures, but it is similar for fixed cruciform junctions which are structurally equivalent to Holliday crossovers. When endonuclease I was used to digest the plasmid, pSAlB.56A, a linear plasmid molecule with hairpin termini was produced (Fig. IV.2A). The termini produced are equivalent to the telomere hairpins in the native virus (De Lange *et al.*, 1986). Numerous cruciform crossover positions in the plasmid pSAlB.56A were identified using T7 endonuclease I as a probe (Chapter II). These sites corresponded to cleavage across one chosen axis of the junction (axis 'a' as represented in Fig. IV.1C), as mapped from restriction sites flanking either side of the inverted repeat sequences. Unexpectedly the cleavage positions were

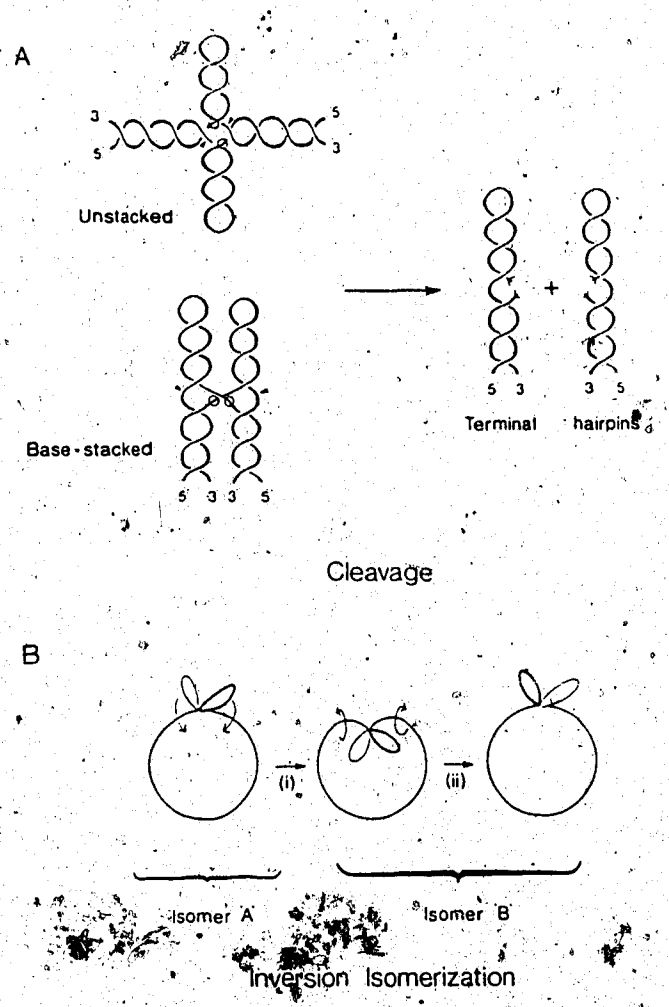


Figure IV.2. Cruciform isomerization and T7 endonuclease I cleavage of cruciform crossovers. Panel A: Potential cleavage sites for T7 endonuclease I are positioned at a cruciform crossover as it may exist in two hypothetical conformations (left side of reaction). Cleavage sites are paired (circles and arrowheads) to denote the two cleavage axes. The top illustration is unstacked with respect to the basepairs at the crossover, and planar. In the lower representation the branches are stacked one on top of the other (the UHA structure first described by Sigal & Alberts, 1972). Cleavage results in resolved hairpin termini (right side of reaction, cleaved at the circled sites). Panel B: Two possible isomeric forms of a cruciform structure are schematically drawn. The curved arrows illustrate the direction of rotation of the junctional branches. Reaction (i) is an isomerization analogous to that described by Meselson & Radding (1975). In the plasmid substrate, the cruciform hairpins are simply inverted with respect to the circular branches of the crossover. Reaction (ii) is a rotation of the cruciform hairpins about the helical axis of the circular branches and effects no isomerization of the junction. Light and heavy loops schematically illustrate the two cruciform hairpins, whereas a single line represents the double-stranded circular core domain of the plasmid.

discovered to be periodically spaced, an average of ten nucleotides apart (the equivalent of one turn of primary helix) along the DNA sequence. The periodicity was surprising since every phosphodiester within the limited region of dyad symmetry of a synthetic Holliday structure capable of branch migration was shown to be sensitive to cleavage by T7 endonuclease I (Chapter III). A spacing of 5 to 6 nucleotides (equivalent to a half-turn of primary helix) might have been predicted based upon the fact that the topoisomers making up the plasmid preparation would differ by a linking number of one (Pulleyblank *et al.*, 1975; Keller, 1975). By way of explanation, consider the analogous example of the relaxation of one supercoil in a pSA1B.56A topoisomer during cruciform extrusion. One-half turn of helix on either side of the cruciform structure would be extruded (removed from the circular core of the molecule) and the cruciform junction would be displaced 5 to 6 nucleotides along the DNA sequence. The observed ten nucleotide periodicity, therefore, suggested that variable structural forms of the cruciform junction existed and that the isomerization event that related them was related to the linking difference in the circular core of the plasmid molecules. The inability to cleave at least every other presumptive crossover position indicated that either (i) the cleavage at consecutive "stable" crossovers (crossovers associated with topoisomers in their lowest free energy form) alternated between the two cleavage axes (axes 'a' and 'b', Fig IV.1C), or (ii) the junctions at these positions were altogether insensitive to T7 endonuclease I digestion.

The conformational isomerization of Holliday structures has not previously been reported to be linked, or coupled, to branch migration. The isomerization of Holliday structures described by Meselson and Radding (1975), through which opposite pairs of strands are brought into the "crossed" configuration, has no apparent topological implications for a

cruciform junction since the process would only involve the inversion of the cruciform structure relative to the circular domain of the molecule (Fig. IV.2B). Therefore, experiments were carried out to determine (i) whether either axis could be cleaved at the "sensitive" crossover positions, (ii) whether the cleavage pattern was exclusively related to the previously characterized cruciform structure in pSAlB.56A (Chapter II) and, (iii) if branch migration in isolated plasmid topoisomers would lead to the generation of the same cleavage pattern. The results were found to be consistent with the interpretation that T7 endonuclease I discriminates between conformational isomers of the cruciform junction that arise as a result of a change in the helical twist of the plasmid circular core during branch migration. This presumptive isomerization appears to be unrelated to the more familiar inversion isomerization process.

B. Materials and Methods

Preparation of DNA The cruciform-containing plasmid substrate was pSAlB.56A, which has been extensively characterized (De Lange *et al.*, 1986; Chapter II). It is a pUC13 derivative bearing the telomere sequences of Shope fibroma virus in the form a 322 basepair inverted repeat. The plasmid, with a stable cruciform extruded from the viral insert sequences, was purified from low-melting agarose (Langridge *et al.*, 1980). The hairpins of the cruciform contain a total of eight extrahelical bases as a consequence of the imperfect nature of the inverted repeat. The creation of two pSAlB.56A derivatives, pSAlB.LA and pSAlB.RA, has been described (Chapter II). These plasmids carry the left and right arms, respectively, of the viral inverted repeat (see Fig. IV.1A).

The viral sequences were excised from pSAlB.56A by digestion with *Hin*dIII and *Eco*RI

(Fig. IV.1). This "bridge" fragment was purified from low-melting agarose and heat-denatured by boiling for 3 minutes in a small volume of 10mM Tris.HCl and 0.1mM EDTA, pH8 (TE buffer). The DNA was immediately chilled on ice and the individual strands self-annealed to form "snap-back" structures which are equivalent to the cruciform hairpins and the original viral telomere hairpins (De Lange *et al.*, 1986).

Enzymes and reaction conditions Restriction enzymes and T4 DNA polymerase were purchased from Bethesda Research Laboratories (BRL). Mung bean nuclease and T4 polynucleotide kinase were obtained from Pharmacia and calf intestinal phosphatase was obtained from Boehringer Mannheim. Bacteriophage T7 endonuclease I was generously donated by Dr. P. Sadowski. It was greater than 95% pure with a reported activity of 57 units/ml (personnal communication, P. Sadowski).

Digestions of plasmid DNA with T7 endonuclease I were performed as previously described (Chapter II) except that the reaction buffer contained 4mM spermidine. In 10µl reactions, 50 to 200ng of DNA was digested with 1µl of diluted enzyme (1:10 in 50% glycerol and 1mM EDTA) at 37°C. The incubation times were adjusted to achieve 50 to 80% linearization of the plasmid (usually 5 to 15 minutes) as analyzed following electrophoresis of reaction products in agarose gels.

Mung bean nuclease digestions were performed in T7 endonuclease I buffer at 37°C for 5 minutes. Reactions contained 200ng of DNA and 12.5 units of enzyme in 20µl. The partial linearization of pSAlB.56A by cleavage with *Ava*II was performed in T7 endonuclease I buffer at 37° C for 10 minutes. All other enzyme reactions were performed according to the manufacturers specifications or standard procedures (Maniatis

et al., 1982). End-labelling of all DNA fragments was accomplished using [γ-^{32}P]ATP (5200Ci/mmole) and T4 polynucleotide kinase or [α-^{32}P]dATP (3700Ci/mmole), 2mM each of dGTP, dTTP and dCTP and T4 DNA polymerase. Radiolabelled ATP and dATP was purchased from ICN Radiochemicals.

Gel electrophoresis conditions Topoisomers of pSAlB.56A were purified from 20 x 25cm, 1% low-melting agarose gels electrophoresed at room temperature in 90mM Tris.borate and 2mM EDTA, pH8.3 (TBE buffer) containing chloroquine diphosphate (Sigma) for 72hr at 2.5V/cm. The DNA bands were visualized by staining with ethidium bromide and were extracted according to published procedures (Langridge *et al.*, 1980). The analysis of the purified topoisomer DNA was carried out under similar conditions using standard agarose gels electrophoresing for 24hr. End-labelled cleavage products were analyzed in standard 10% polyacrylamide/7M urea denaturing gels (Maxam & Gilbert, 1980) electrophoresed in TBE buffer. Poloroid negatives of agarose gels stained with ethidium bromide were scanned with a Joyce Loebl Chromoscan 3 densitometer.

C. Results

Purified pSAlB.56A co-migrates with nicked, open-circular plasmid DNA under standard electrophoresis conditions by virtue of the extrusion of a large cruciform that relaxes the plasmid topologically (Chapter II). When the plasmid is electrophoresed in the presence of a titrated amount of the intercalating drug chloroquine, which partially unwinds the DNA causing the resorption of the cruciform structure, individual topoisomers can be separated on the basis of linking number. As shown in Fig. IV.3A, preparations of

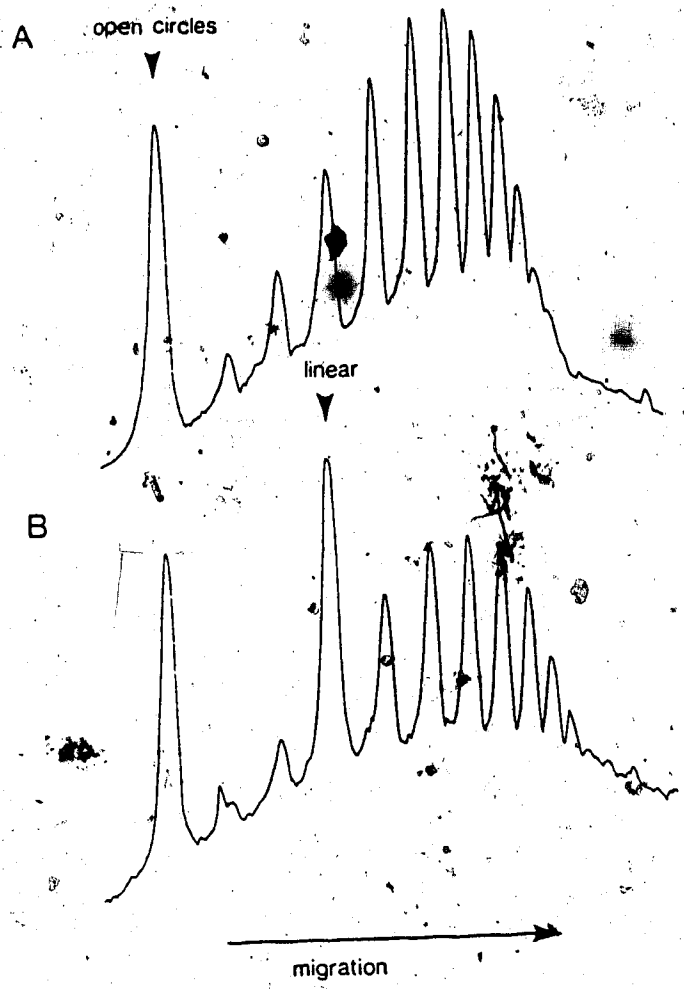
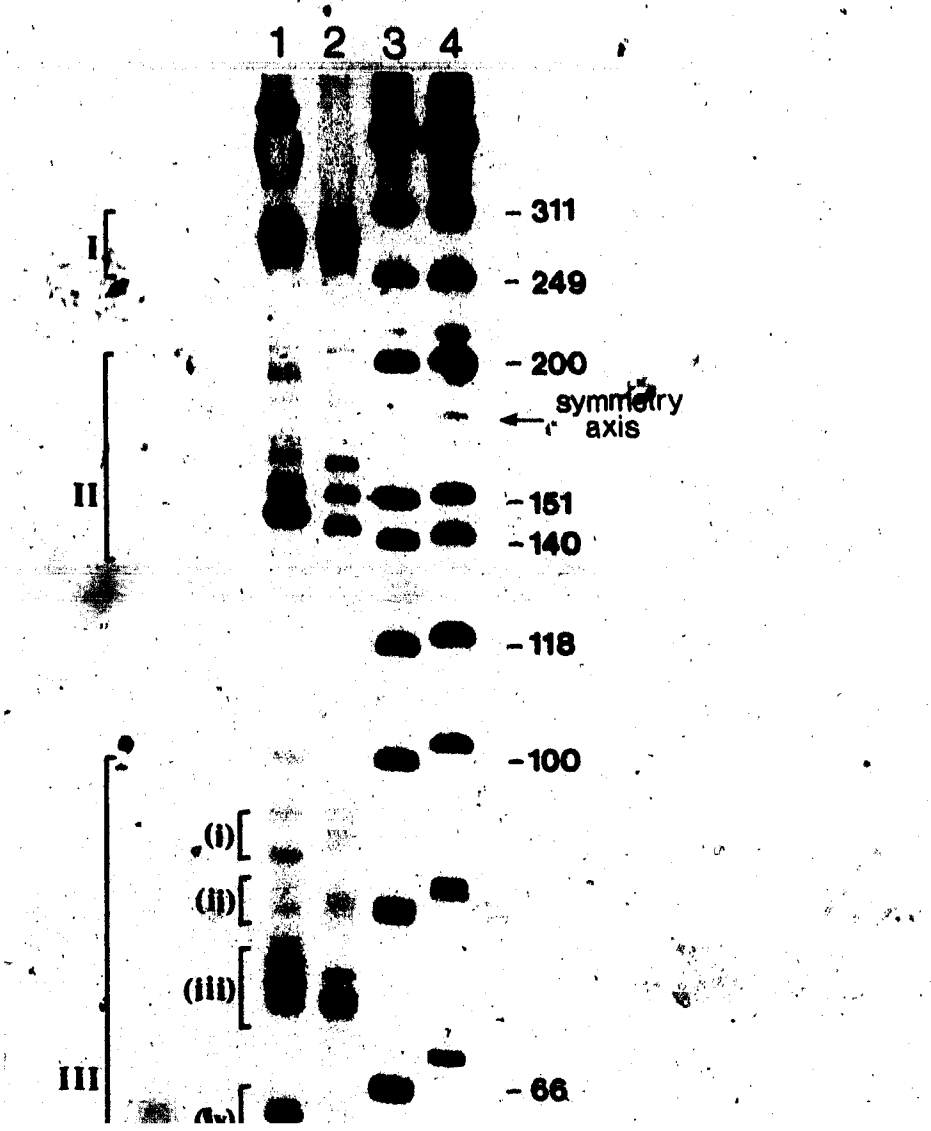


Figure IV.3. Plasmid topoisomer distribution before and after digestion with T7 endonuclease I. Panel A: Undigested pSAlB.56A, electrophoresed in 1% agarose in the presence of 2.1μg/ml chloroquine, stained with ethidium bromide, photographed and scanned with a densitometer. Gel migration was left to right. Open circles refers to nicked plasmid molecules. Panel B: plasmid pSAlB.56A after a partial digestion with T7 endonuclease I at 37°C. The linearized product of the reaction co-migrated with one of the topoisomers in the mixture under these electrophoresis conditions.

purified pSAlB.56A plasmid DNA were demonstrated to contain a normal distribution of topoisomer molecules using this procedure. Partial digestion of pSAlB.56A with T7 endonuclease I at 37°C, analyzed following electrophoresis in chloroquine-containing gels, resulted in the reduction of every topoisomer band, and by inference, showed that each was equally sensitive to cleavage by T7 endonuclease I. The only visible product derived from the eleven discernable topoisomers was linearized plasmid (Fig. IV.3B).

Cruciform crossover positions in the viral sequences of pSAlB.56A were determined by mapping T7 endonuclease I cleavage sites with respect to the *Eco*RI restriction site flanking the insert sequences (see Fig. IV.1). Plasmid DNA was consecutively digested with T7 endonuclease I and *Eco*RI, then end-labelled under conditions that minimized the labelling of internal nick sites. A similar experiment, mapping cleavage sites to one axis (axis 'a', Fig. IV.1C), mapped according to cleavage fragments 5' end-labelled at the *Eco*RI site, has been previously reported (Chapter II). Both 3' and 5' end-labeled cleavage fragments were analyzed here so that resolution across either axis of the cruciform crossover could be identified. A comparison of the cleavage pattern along axis 'a' and axis 'b' (mapped according to the cleavage fragments that are 3' end-labelled) is shown in Fig. IV.4, lanes 2 and 1, respectively. Generally, three broad domains of cleavage were observed within the viral sequences, labelled I, II and III in Fig. IV.4. Domains I and III are symmetrically related and identify the cleavage sites on either side of the inverted repeat axis along the same strand of the viral insert. In domain I are cleavage positions distal to the *Eco*RI site, relative to the cruciform structure, whereas cleavage positions proximal to the restriction site are identified by the fragments in domain III. The two domains, within a given lane (Fig. IV.4, lanes 1 and 2), represent cleavage across both axes of the crossover. On the other hand, the cleavage sites identified in

Figure IV.4. Mapping T7 endonuclease I cleavage sites to the viral insert sequences. Purified pSAlB.56A was partially linearized with T7 endonuclease I, then digested to completion with *Eco*RI. Fragments were either 3'- or 5'-end-labelled at the *Eco RI* termini with ^{32}P and electrophoresed in standard 10% denaturing polyacrylamide gels. Lane 1: 3' end-labelled fragments; lane 2: 5' end-labelled fragments; lanes 3 and 4: 3' and 5' end-labelled fragments of *Hin*fI digested ØX174 RF DNA, respectively. Fragment sizes are given in nucleotides. Three general domains of pSAlB.56A cleavage fragment sizes are identified: I, II and III (for lanes 1 and 2). Within domain III are five sub-regions of cleavage fragments as marked (i) through (v). These regions arise from T7 endonuclease I cleavage at the base of the cruciform on the side proximal to the *Eco*RI site. The inverted repeat axis of the cloned viral sequences is 188 bases from the *Eco*RI site. Fragment sizes are given in nucleotides.


1 2 3 4
311
249
200
symmetry axis
151
140
118
100
66
I
II
III
(i)
(ii)
(iii)

domain II are very near to the inverted repeat axis. Although domain II cleavages are symmetrically placed about the inverted repeat axis, band intensities are far greater on the proximal side of the axis. Cleavage in this region may be indirectly related to the presence of extrahelical bases in the cruciform hairpins and will be discussed later. For the accurate identification of cruciform crossover positions, mapped by endonuclease I cleavage across one or the other axis, only those sites contained in domain III were studied more closely. It was previously demonstrated that the second nick required for effective resolution along one axis could be assigned to the same sequence on the opposite strand by mapping relative to the *Hin*dIII site on the other side of the insert (Chapter II).

Within domain III, five limited regions of T7 endonuclease I cleavage are defined ((i) through (v), Fig. IV.4), the centers of which are spaced, on av[redacted]nucleotides apart along the primary sequence. The relative intensity of cleavage associated with each region reflects a Gaussian distribution of fragment sizes, as might be expected for a natural population of topoisomers, each with a unique crossover position when in its lowest free energy form (ie. the ΔLk in the circular core is near zero). As already suggested, a plasmid molecule with a native superhelical density (around -0.067) would be expected to have the base of its extruded cruciform within this general region of the viral insert sequence (Chapter II).

Cleavage along either junctional axis has been mapped to the same crossover positions. T7 endonuclease I cleaves predominantly one nucleotide to the 5' side of the crossover in synthetic Holliday junction analogs (Chapter III). Consequently, 3' end-labelled fragments should be 2 nucleotides longer than the 5' end-labelled fragments (corresponding to cleavage along axis 'b' and axis 'a', respectively) if cleavage across

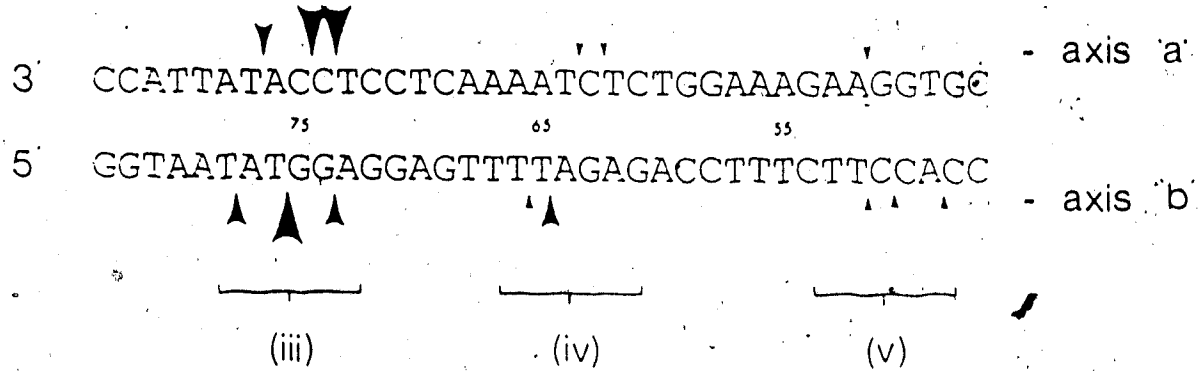


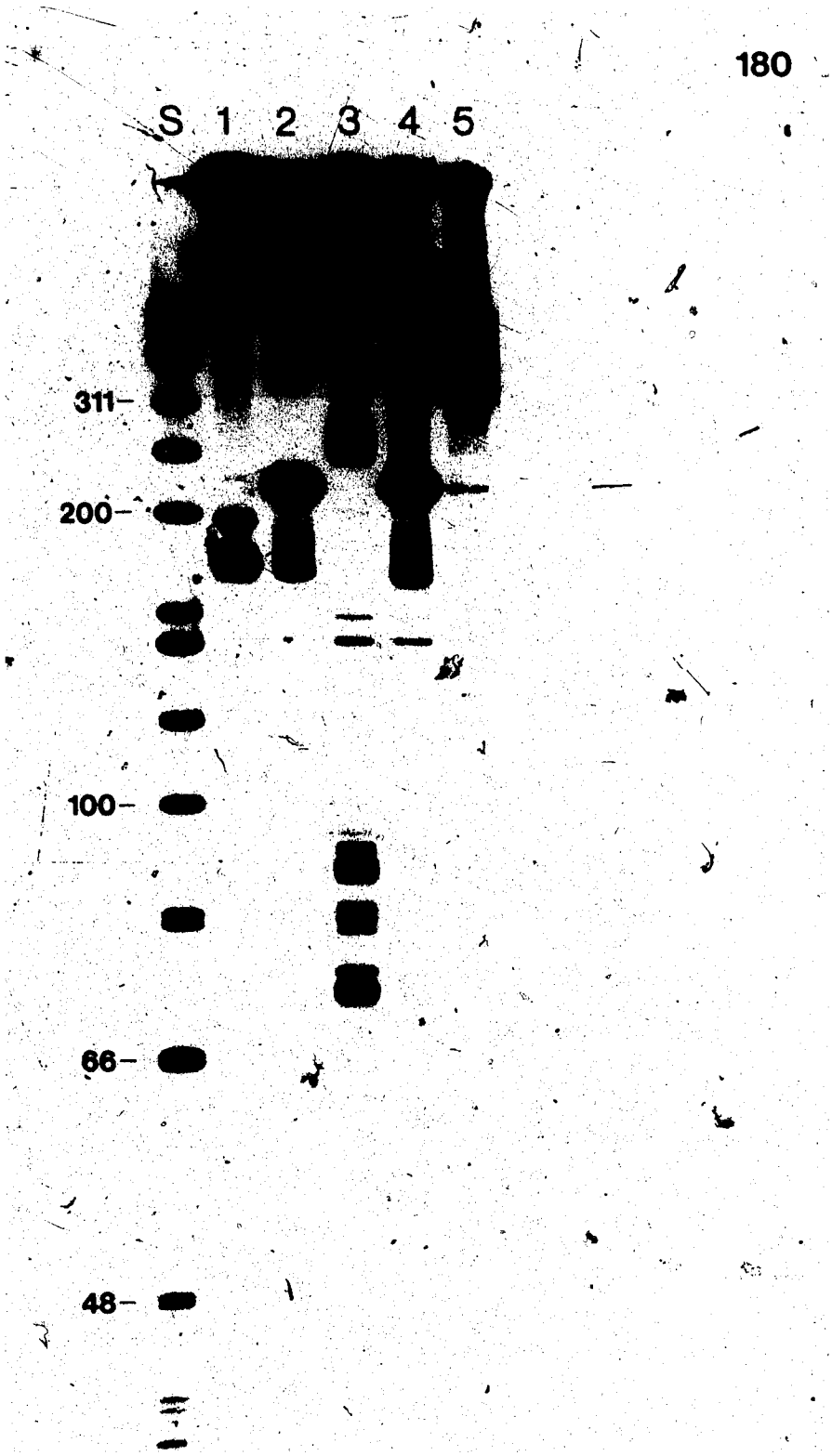
Figure IV.5. Correlation of T7 endonuclease I cleavage positions with DNA sequence. Three cleavage regions ((iii), (iv) and (v)) were mapped, by the direct comparison of pSA1B.56A fragment lengths to known size markers (*Hin*fI digests of ØX174 and pBR322). Cleavage along axis 'a' is read off the top strand labelled at the 5' end, whereas cleavage along axis 'b' is read off the 3' end-labelled lower strand. The numerals between the two strands give the distance, in basepairs, from the labelled *Eco RI* terminus. The cleaved phosphodiesters are indicated with arrowheads, the size of which approximates the relative sensitivity of each site.

either axes occurred at the same crossover position. This was generally observed for regions (iii), (iv) and (v), considering that some variability in cleavage efficiency would be expected because of the sequence preference displayed by the enzyme (Chapter III). In Figure IV.5 above, the cleavage positions have been aligned with the corresponding nucleotide sequence. The enzyme is known to prefer to cleave phosphodiesters 5' to

pyrimidine residues (Sadowski, 1974; Chapter III), particularly cytosine residues. This could explain the greater variability observed in the cutting at crossover position (v), which resides in a polypyrimidine-polypurine stretch (Fig. IV.5). Most significantly, however, within the inter-region sequences there are numerous sites with preferred cleavage sequences that are not cleaved by T7 endonuclease I. For example, the sequence 5'-AC has been identified as a highly preferred cleavage position for the enzyme (Chapter III) and this sequence exists between regions (ii) and (iii), (iii) and (iv) and (iv) and (v). On the other hand, the sequence 5'-CA was poorly cleaved in synthetic Holliday structures capable of branch migration and thus, the absence of cutting at this phosphodiester in region (iii) (top strand, Figure IV.5) is consistent with this demonstrated enzyme specificity and substantiates the sequence alignment.

A series of controls was performed to confirm that the observed endonuclease I cleavage sites did, in fact, correspond to variable crossover positions of a cruciform extruded from a common inverted repeat axis. Isolated hairpins, prepared by the heat-denaturation and "snap-back" of excised inverted repeat insert fragments (see Materials and Methods, this chapter), are not cleaved at all in domain III (lane 5, Fig. IV.6). Mung bean nuclease, used to probe for other non-B DNA structures (Kowalski, 1986), only cleaved at the central axis of the cloned viral inverted repeat (Fig. IV.6, lane 1). These central axis sequences would have formed the single-stranded hairpins of the large cruciform branches in pSAlB.56A. Also, pSAlB.56A was partially linearized with *Ava*II (two *Ava*II sites are positioned 222 basepairs apart far removed from the cloning site in pUC13) and then immediately digested with T7 endonuclease I (lane 4, Fig. IV.6). Linearization destabilizes cruciform structures and would have allowed the molecules to branch-migrate into the more stable lineform configuration. Consistent with the

Figure IV.6. Cleavage of pSAlB.56A by T7 endonuclease I is related to the existence of the cruciform crossover. Plasmid pSAlB.56A was partially digested with various endonucleases and then cleaved with *Eco*RI, end-labelled and electrophoresed as described in the legend for Figure IV.4. Lane 1: pSAlB.56A digested with Mung Bean nuclease; lane 2: pSAlB.56A digested with *Ava*II; lane 3: plasmid digested with T7 endonuclease I; lane 4: plasmid partially digested with *Ava*II, followed by digestion with T7 endonuclease I. In lane 5, isolated cruciform hairpin structures (see Materials and Methods, this chapter) were digested with T7 endonuclease I and then 5' end-labelled. The size markers (S) were 5' end-labelled *Hin*fI fragments of ØX174 RF DNA.


S 1 2 3 4 5
311
200
100
66
48

resorption of the cruciform structure, less cleavage was observed at regions (i) through (iv) following a partial linearization of the plasmid. And finally, when clones bearing the individual arms of the viral inverted repeat (pSAlB.LA and pSAlB.RA) were treated with T7 endonuclease I under similar conditions as digestions of pSAlB.56A, no cleavage was observed whatsoever (data not shown).

Inspection of figure IV.3A reveals that at least eleven topoisomers could be distinguished in the purified pSAlB.56A preparations. In their lowest free energy form, each should have a cruciform of unique size and crossover location. Nonetheless, as evident from the data in Figure IV.4, T7 endonuclease I only identifies five, perhaps six, possible crossover locations. In conjunction with the observed ten nucleotide spacing between the T7 endonuclease I cleavage regions, this suggested that every other topoisomer was not cleaved when in its lowest free energy state. Instead, since every topoisomer has been shown to be cleavable (Fig. IV.3), the cruciforms must have facilely branch-migrated to sensitive crossover positions and been cleaved there. To assess the degree of branch migration in these molecules, individual topoisomers were purified (Fig. IV.7) and digested with T7 endonuclease I (Fig. IV.8). Each topoisomer, after digestion, produced a subset of the cleavage fragments previously identified in the population as a whole. As expected, the greater the ΔLk, the further the average crossover position was from the inverted repeat axis. Most importantly, the crossover in each topoisomer branch-migrated a distance spanning a number of T7 endonuclease I sensitive and insensitive regions. The less efficient cleavage in region (ii), relative to the cleavage at regions (i) and (iii), as demonstrated with the purified topoisomers, was probably not due to a lesser frequency of crossovers in this region, but rather, was attributable to the sequence preference of T7 endonuclease I. Though T7 endonuclease I prefers certain sequences, non-preferred

Figure IV.7. Purification of pSA1B.56A topoisomers. Plasmid DNA was electrophoresed and individual topoisomers were purified from low-melting agarose gels containing 2.1μg/ml chloroquine (see Materials and Methods, this chapter). The purified topoisomers were analyzed following electrophoresis for 24 hours in standard 1% agarose gels containing the same concentration of intercalating drug, and staining with ethidium bromide. In lanes 1 through 8 are the individual purified topoisomers, ranked according to linking difference (Lk) with the most supercoiled species present in lane 8. oc-DNA is the nicked, circular form of the plasmid.

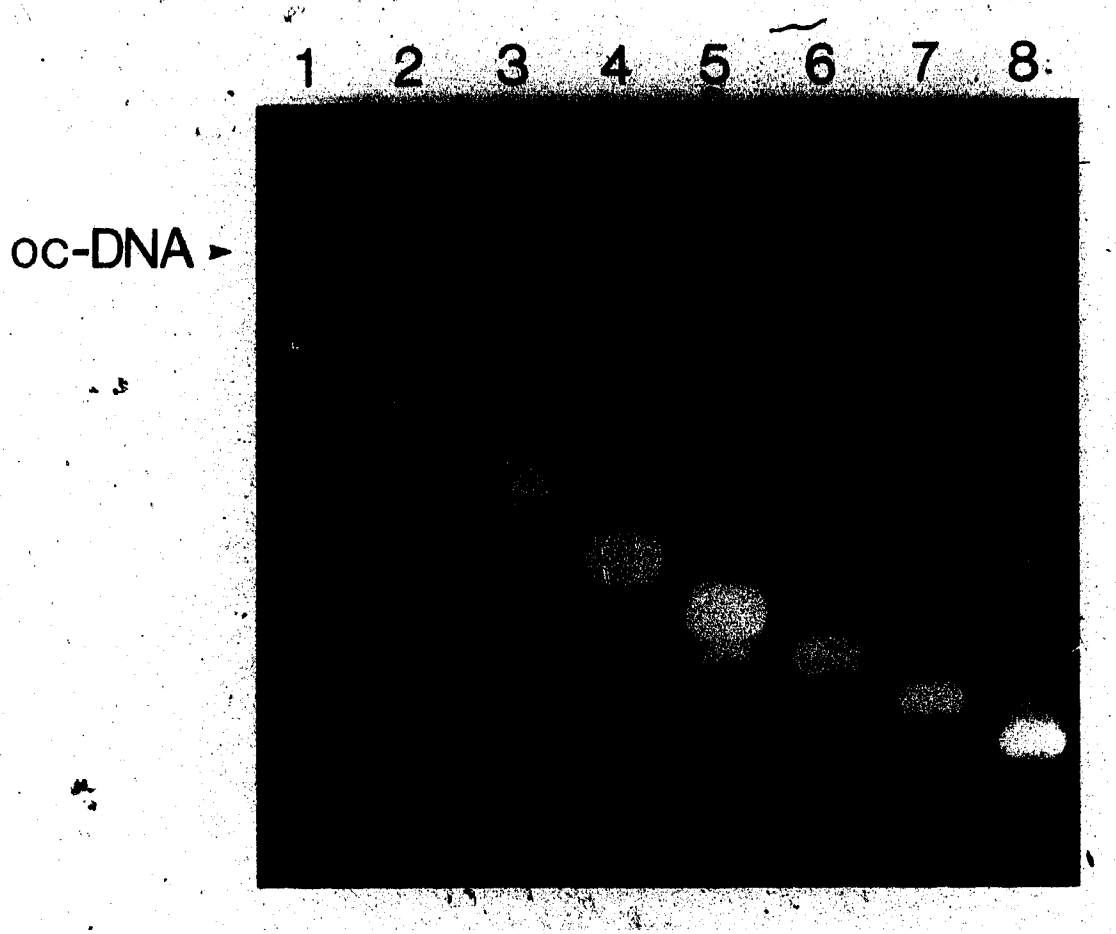
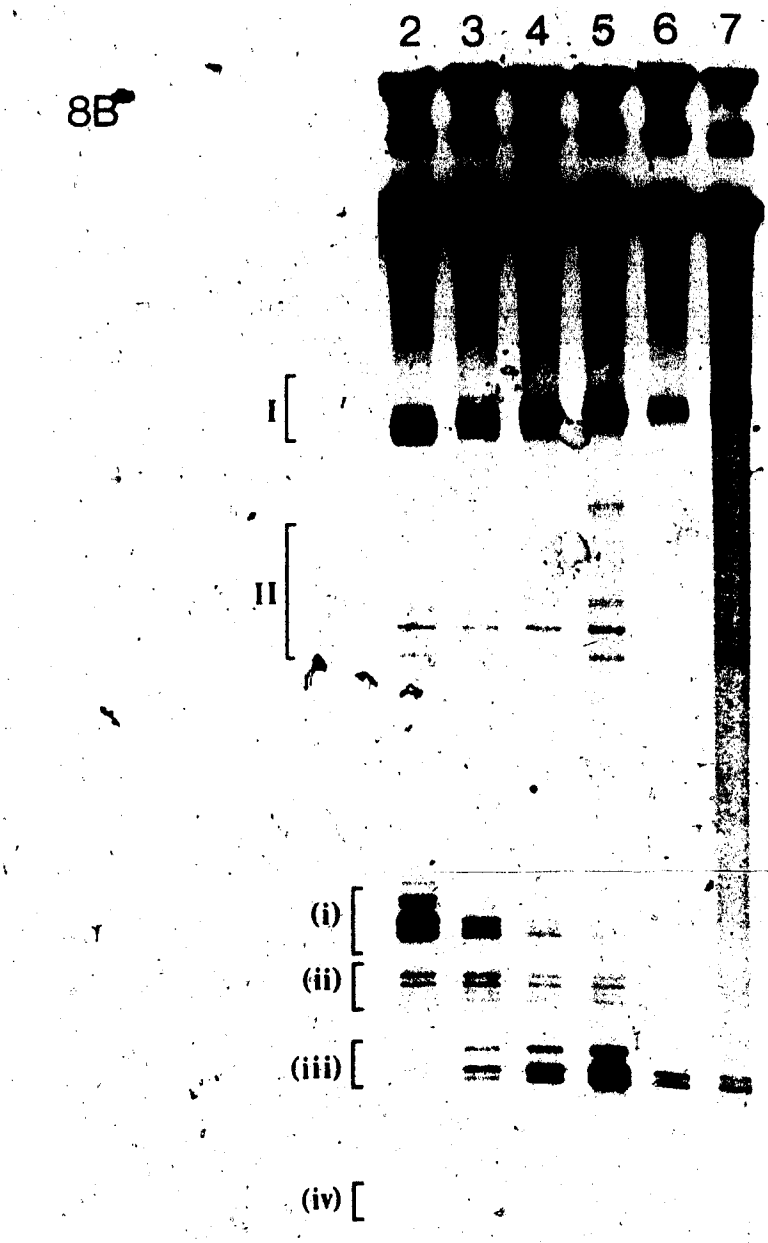

1 2 3 4 5 6 7 8
oc-DNA

Figure IV.8. T7 endonuclease I digestion of purified pSA1B.56A topoisomers. The digestion of plasmid DNA and the analysis of 5' end-labelled cleavage fragments was performed as described in the legend for Fig. IV.4. The lane headings relate directly to the numbering of the purified topoisomers in Fig. IV.7. For example, in lane 7 are the cleavage fragments generated by the digestion of the next to most supercoiled topoisomer purified. Domains and regions of cleavage are the same as described in the legend for Fig. IV.4.


2 3 4 5 6 7
8B
I
II
(i)
(ii)
(iii)
(iv)

cleavage sequences can nevertheless still be cleaved when positioned 5' to a fixed (non-migrating) Holliday structure (Chapter III).

Cleavage very near the inverted repeat axis (in domain II, Fig. IV.4), was also observed in each of the purified topoisomers (Fig. IV.8). The efficiency and position of cleavage in domain II appeared to be independent of linking difference. Moreover, it was not significantly reduced when the plasmid was linearized (lane 4, Fig. IV.6). The cleavage pattern is provocative because the cut sites are phased approximately 10 nucleotides apart beginning from the inverted repeat axis. Their appearance was dependent upon the presence of the inverted repeat and they are absent in digestions of the cruciform hairpins (Fig. IV.6). Their presence is consistent with the extrusion of a small cruciform from the same inverted repeat axis as the major cruciform under study. Consequently, it is possible that small cruciforms have become trapped in this region because the presence of the extrahelical bases in the cruciform hairpins would retard branch migration during the cruciform-to-lineform transition (Robinson & Seeman, 1987).

D. Discussion

Due to the large size of the inverted repeat insert in pSA1B.56A, branch-migration in the plasmid molecules implies that, at any one time, the cruciform crossover could be positioned at any point along the viral insert sequence within a region determined by the range of topoisomer linking differences and the supercoiling energy in the circular cores of the respective topoisomers. Certain crossover positions may be favored, such as those associated with topoisomers that are completely relaxed, topologically, and in their lowest energy form. These crossover positions would be separated along the primary sequence

by the equivalent of one-half turn of helix, attributable to topoisomers whose linking numbers differ by increments of ±1. Mapping crossover positions with T7 endonuclease I resulted in an observed periodicity of ten nucleotides, equivalent to the extrusion of two full turns of helix. All phosphodiester bonds contained within the limits of branch migration in the plasmid molecules were not equally sensitive to endonuclease I cleavage, and those which were cleaved were restricted to defined and periodically spaced regions. Insensitive regions may have reflected positions at which the crossovers were unstable, though this is unlikely because all the topoisomers in the plasmid population were completely relaxed topologically when in the cruciform configuration (Chapter II) and thus, in theory, should possess a unique crossover position. Branch migration in every other topoisomer to a stable crossover position would have introduced at least one superturn into the plasmid which could have been, but was not, detected by agarose gel electrophoresis (Chapter II). Eleven topoisomers were identified in the plasmid population, all of which were cleaved (Fig. IV. 3), and yet only 5 or 6 crossover positions were mapped by T7 endonuclease I digestion (Fig. IV. 5). It follows that certain crossover positions were associated with junctions that were resistant to digestion with T7 endonuclease I.

Plasmid pSAlB.56A is a unique substrate for these experiments in that it is the largest stably cloned palindrome characterized at present (Chapter II). Most cruciforms studied to date were much too small to relax the circular DNAs in which they resided. The plasmid pSAlB.56A contains an average of 20 supercoils and relaxation would require the extrusion of at least 200 base-pairs into the cruciform. The overall size of the inverted repeat insert, 322 base-pairs, exceeds this minimum and permits the migration of the junction subject only to topological constraints. Unfortunately, large inverted repeats are

notoriously difficult to clone, illustrated by the fact that pSAIB.56A can only be stably propagated in recombination deficient bacteria. Thus, it remains to be established that the observed cleavage pattern is generally applicable to cruciform DNAs, although it is highly unlikely that the observed cleavage periodicity was profoundly determined by peculiarities in the sequence of the viral insert of pSAIB.56A. For instance, as shown in Figure IV.5, preferred sequences for T7 endonuclease I (the cleavable sequence 5'-AC, for one) did exist in the insensitive regions. As convincing, the pattern of cleavage was consistent over a span of 60 basepairs in a sequence that is notable only in that it is moderately A+T-rich. Nevertheless, it remains debatable whether the cleavage pattern was influenced by some other structural element present in the poxviral telomeric sequences. However, control experiments, described in this report, repeatedly failed to detect any T7 endonuclease I cleavage in the relevent sequences of pSAIB.56A that was not associated with the presence of the relevent cruciform crossover. Probing the sequences with Mung Bean nuclease also failed to detect any other unusual DNA structural forms which may have been present in the plasmid molecules and which could have influenced the cleavage specificity of T7 endonuclease I.

The T7 endonuclease I mapping of crossover positions in individual purified topoisomers demonstrated that the cruciform crossover could branch-migrate over a range of enzyme sensitive and insensitive regions. Therefore, if enzymatic cleavage is not profoundly affected by the DNA sequence, it is most likely influenced by changes in crossover conformation and the regular periodicity of the cleavage pattern, so provocatively related to the helical repeat unit, suggests that this isomerization of the cruciform structure is coupled to branch migration via the associated ΔTw in the circular core domain of the plasmid molecule. Specifically, it seems reasonable that coinciding with every turn of

helix added to, or removed from, the cruciform structure, the cruciform isomerized. In accordance with the data, one isomer would be sensitive to T7 endonuclease I cleavage across either crossover axis more or less equally, while the other form (or intermediate forms) would be completely resistant. The interconversion of the two forms, proceeding through intermediate conformations, would explain the rather limited regions of cutting observed for the T7 enzyme. Branch-migrating Holliday structures with topologically free branches can be cleaved at every nucleotide position within the region of dyad symmetry (Chapter III). Therefore, the phenomenon of periodic cleavage, and by inference the imposed isomerization of the cruciform structure, would have to be a property of the topological constraint imposed by the circularization of two branches of the junction.

Branch Migration in Supercoiled Plasmid DNA Models for double-stranded branch migration in Holliday structures (Meselson, 1972; Meselson & Radding, 1975; Robinson & Seeman, 1987) have been developed in terms of the rotary diffusion driven movement of a crossover in a base-stacked, Sigal & Alberts-type (1972) molecule (described in Fig. IV.2B). More recently, Robinson & Seeman (1987) have called this version of a Holliday structure the unperturbed-helix-axis (UHA) form. The significant assumption made has been that the DNA branches are sufficiently "flexible" that they are free to rotary diffuse (Meselson & Radding, 1975), ie. they are not topologically constrained. In plasmid-borne cruciform DNA, two of the Holliday junction branches are contiguous and form the topologically closed domain of the molecule. In small plasmid molecules, rotary diffusion of these branches could not occur, to any great degree, independently. Rather, diffusion would rotate the branches in the same direction and merely spin the cruciform hairpins about the helical axis of the circular branches in the manner described in Fig.

IV.2B, (ii). This process neither represents an isomerization of the junction nor would it *a priori* influence branch migration. In small plasmids (and presumably any small topologically closed domain), branch migration will almost exclusively be driven by the free energy of supercoiling (Mizuuchi *et al*, 1982b). Therefore, branch migration will be accompanied by discrete changes in twist (ΔTw) in the circular branches of the molecule, while its direction will be dictated by the sign of the linking difference (ΔLk) of the circular core.

At least three isomerization processes can be described for a cruciform crossover. The most elementary involves the unidirectional movement of the branch point along the DNA sequence. A second is branch inversion (described in Fig. IV.2 and Fig. IV.9) which is comparable to the isomerization process described by Meselson & Radding (1975) and has no apparent topological component. A third isomerization process, not previously alluded to, will accompany a ΔTw in the junction branches. The latter two processes are illustrated and compared in Fig. IV.9. A ΔTw in the branches will change the disposition of the major and minor grooves adjacent to the phosphodiester crossovers. For instance, in one isomer, the minor grooves would face the top of the page (left side of Fig. IV.9), whereas in the alternative isomer (right side of Fig. IV.9) the major grooves would face the top of the page. Note that there is no coincidental reversal in the polarity of the DNA strands. There is no indication in Fig. IV.9 how this process may occur mechanistically. Rather, its association with branch migration in circular plasmid molecules is more fully demonstrated in Fig. IV.10. Figure IV.9 serves only to emphasize that inversion isomerization has no bearing on the disposition of the grooves; it is the topology (ΔTw) of the circular core that influences the phasing of the grooves at the junction. A comparison of the structures labelled [A] and [B] in Fig. IV.9 will show that these are

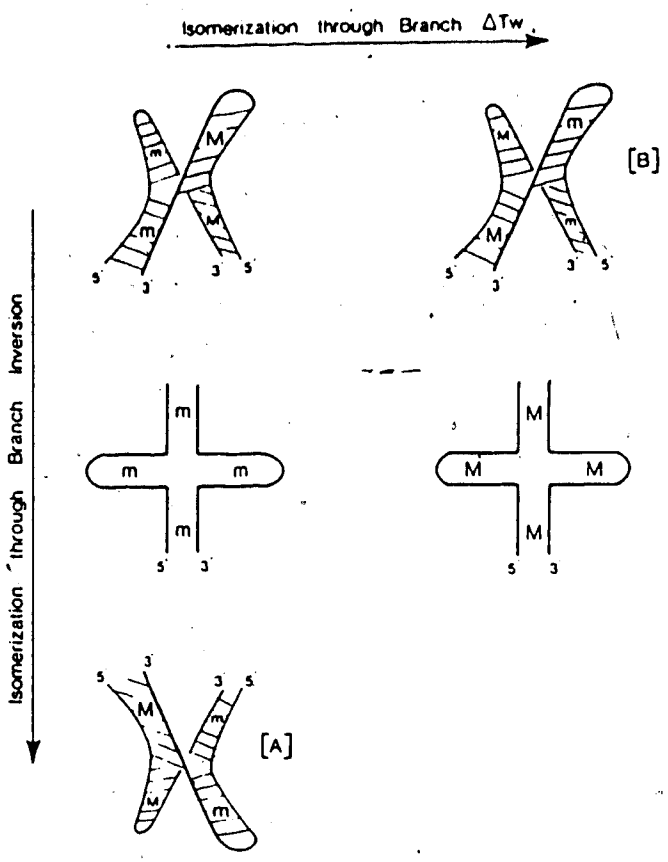


Figure IV.9. Isomerization in terms of the disposition of the helical grooves of the DNA branches adjacent to a cruciform crossover. Two unique isomerization events are schematically represented using the stick figure of a Holliday junction. The helical turns have been removed for clarity and the branches of the junction are shown in the tetrahedrally arranged conformation. Proceeding from top to bottom, junctions undergo an inversion isomerization (as in Fig. IV.2) by passing through a planar form. Left to right in the figure are pairs of isomers that are related by a change in twist (half twist per branch) in the junctional branches. Note that in this case there is a change in the disposition of the grooves relative to the polarity of the DNA strands. In this diagram, open branches, marked with the 5'/3' polarities denote the circular core branches. Structures [A] and [B] are pseudo-enantiomorphic forms of the junction. M: major grooves; m: minor grooves. In the unwound representations shown, the major and minor grooves are on opposite faces of the duplex DNA tracts.

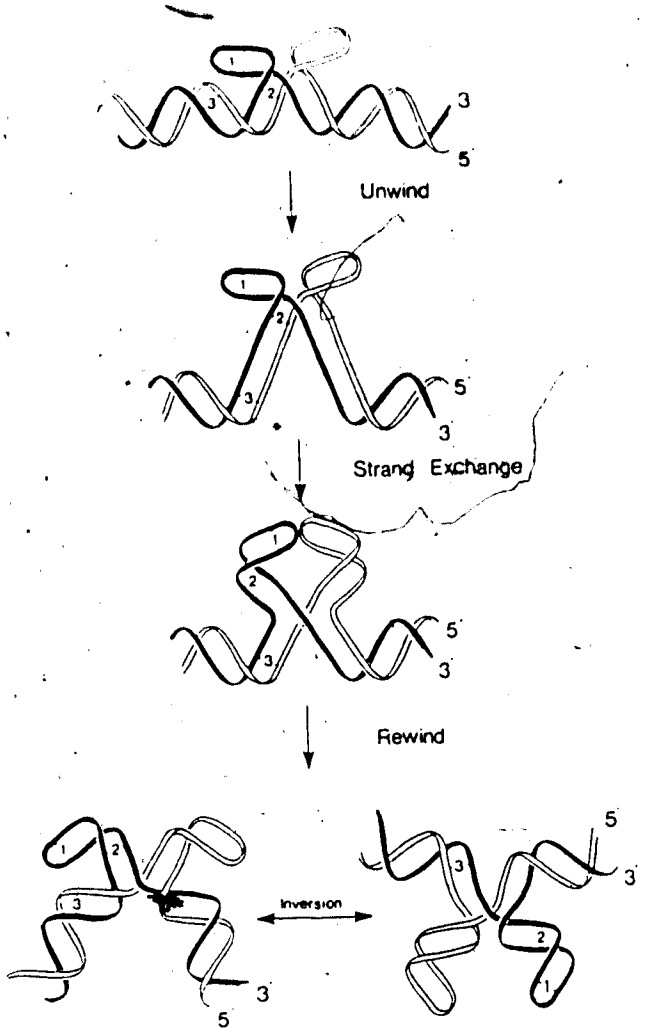


Figure IV.10. Branch migration and isomerization of a cruciform crossover. A plasmid molecule with one helical turn extruded into a cruciform structure (top figure) branch migrates so that a second turn is extruded from the circular core (bottom figures, having a relative ΔLk of +1). The left and right branches of the junction are enclosed to form the circular core of the plasmid. The two bottom figures are inversion isomers of the more extruded molecule. Individual strands are distinguished as light and dark ribbons. Three partial helices are numbered, -1-, -2- and -3- to highlight the net movement of the partial turns during extrusion. As shown, turn -2- is extruded into a hairpin by becoming unwound, undergoing strand exchange and finally being rewound to complete the process. Note that the disposition of the helical grooves has been reversed at the crucform junction. This is the critical difference between the twist-related isomerization and inversion isomerization (bottom reaction).

pseudo-enantiomorphic - a critical consideration when regarding enzyme specificity.

During branch migration, where there is an extrusion of sequences from the circular core domain to the cruciform hairpins, there is an accompanying ΔTw. For every turn of helix extruded, alternative grooves will be brought into apposition at the branch point such that the disposition of the grooves is reversed relative to the polarity of the strands (Fig. IV.10). No extrusion mechanism is implied in Fig. IV.10, rather the illustrations simply diagram the net movement of the DNA strands during the extrusion of one helical turn of DNA. As shown, each hairpin grows by one-half turn of helix for the one turn removed from the circular core and the disposition of the minor grooves changes from being projected 'up' to being projected 'down'. The hairpins maintain their orientation and strand polarities are unaltered. In Fig. IV.10, the junction is depicted as being somewhat tetrahedrally arranged but it is not necessary to impose any structure upon the junction to demonstrate this isomerization. However, it should be stated that branch migration, because of the phasing of the helical twists, can not proceed with the junction exclusively in a base-stacked, UHA form. This conformation could not accommodate the obligatory intermediate forms which would arise with the extrusion of non-integral turns of helix.

The periodic cleavage of branch migrating topoisomers of pSA1B.56A by T7 endonuclease I is consistent with the junctional isomerization coupled to ΔTw. There is little question that the enzyme would differentiate between the pseudo-enantiomorphic forms of the junction. Viewing the planar representations of the pseudo-enantiomers in Fig. IV.9, it can be seen that to properly orient itself with respect to strand polarities, the enzyme would interact with either the major or minor grooves. This specificity would exclude one pseudo-enantiomer and any intermediate forms, consistent with the limited

regions of cleavage observed (Fig. IV.4). It is predicted that with larger circular domains (therefore an increasing "flexibility" in the circular branches), these cleavage regions would broaden since rotary diffusion would have an increasing influence on the movement of the crossover. Rotary diffusion itself does not necessarily impose conformational isomerization on the junction.

Junctional Conformation and Enzyme Specificity An attempt was made in an earlier paper to define the substrate parameters for T7 endonuclease I (Chapter III). It was proposed that the enzyme recognizes a distortion in the phosphodiester backbone of a DNA strand that passes from one duplex branch to another. Based primarily on the cleavage of tri-oligonucleotide complexes (containing two duplex branches and two single-stranded branches), the enzyme was presumed to bind to neighboring duplex branches and cleave the bridging strand one nucleotide 5' to the crossover point. The crossover in a UHA structure (base-stacked form) is fixed by the overall orientation of the four duplex branches. For this reason, the UHA form is probably not a structure that would be a substrate for the T7 endonuclease. Alternatively, Holliday structures could be planar or something in between these two extremes, such as tetrahedral in arrangement (Sobell, 1974; Gough & Lilley, 1985; Robinson & Seeman, 1987). The planar form is attractive simply because it has pseudo-four-fold symmetry and presents four equal cleavage domains, consistent with the observation that both axes of sensitive crossovers were cleaved more or less equally (Fig. IV.4). Robinson & Seeman (1987) have theoretical evidence suggesting that a near planar conformation for Holliday junctions is the most acceptable energetically. Tetrahedrally arranged junctions (and UHA junctions) possess dyad symmetry and the two cleavage axes would be distinct. The cleavage axes could be equated through inversion isomerization but, in circular plasmid molecules,

inversion isomers may not have equivalent conformations. For example, in Fig. IV.10 (bottom illustrations) two inversion isomers are compared. In one form (to left of arrow), the circular branches have their minor grooves in apposition; and in the other form (to right of arrow), the major grooves are in apposition. Quite likely, the orientation of the circular branches will differ in these two forms to minimize the phosphate-phosphate interactions of the backbones. Were this to be achieved at a significant energy cost in terms of the stability of the whole molecule, it is conceivable that the cruciform hairpins could rotate about through the circular core during branch migration (see Fig IV.2B, reaction (ii)) such that either the major or minor grooves were always apposed in the circular branches. The phasing in the two hairpins branches would be less of a concern since these are not topologically constrained and would be relatively free to make the necessary accommodations. It may be significant that the hairpin branches would not have to rotate through the solvent and about the circular cores axis if the Holliday junction were in the planar configuration.

As part of a preliminary discussion of the mechanism of action of T7 endonuclease I, and similar "resolving" enzymes involved in recombination, Fig. IV.11 is included to suggest how catalytic units of the enzyme might approach and bind to a Holliday junction (a cruciform, in this case). It is proposed, as an alternative model to the resolution of UHA forms, or strictly tetrahedral forms, by a dimeric enzyme (Kemper *et al.*, 1985), that individual enzyme units bind separately across one axis of the junction, and ensure the fidelity of the resolving reaction by precluding binding to the opposite axis. A positive cooperative interaction may occur across the junction, in order to promote the attainment of an activated enzyme-substrate complex, for instance. The junction may approach planarity in its most stable form, or planarity may be imposed upon it by the

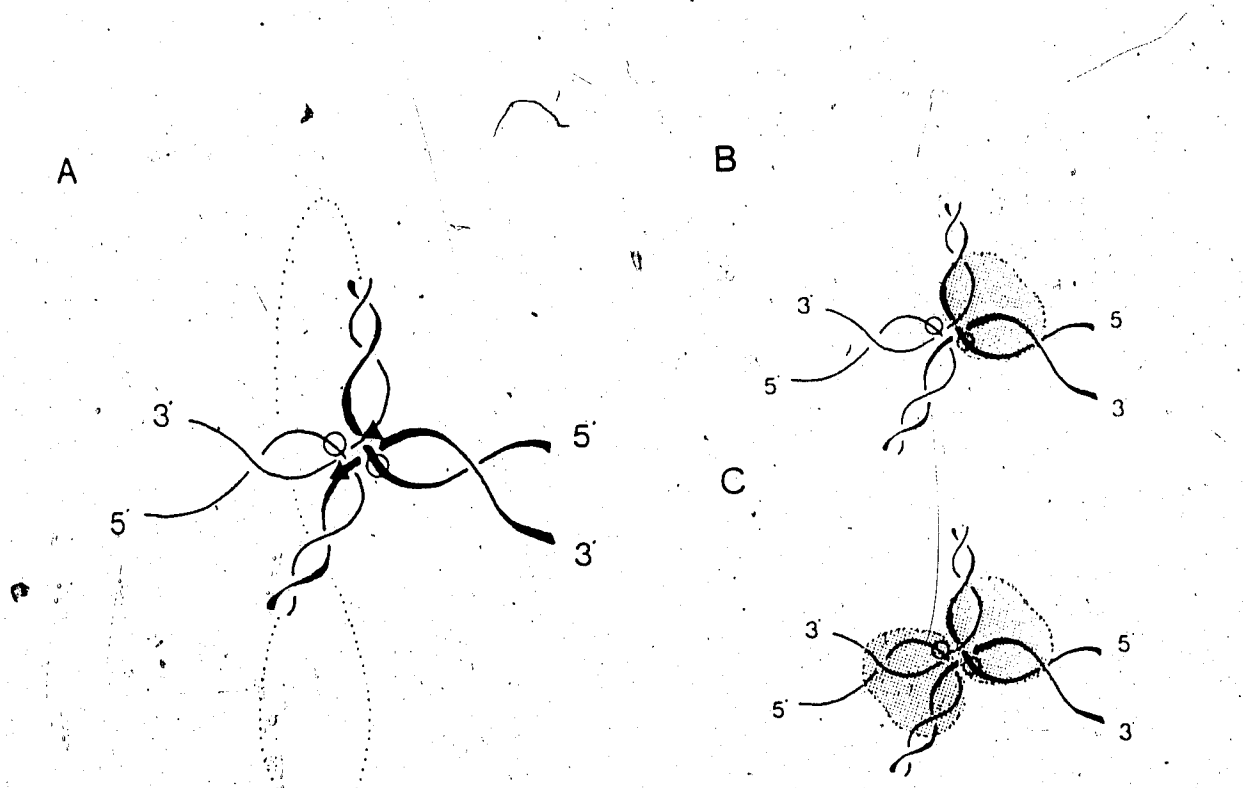


Figure IV.11. A hypothetical cruciform junction and the proposed binding of T7 endonuclease I. Panel A: the branches of the cruciform structure are shown to be arranged in a near-planar form. The circular core region of the molecule is represented by the single dotted line. The cruciform hairpins are only partially drawn and extend to the right and left of the crossover. One hairpin is identified by a 'heavy' strand (to the right) while the other is 'light'-stranded (to the left). The pairs of cleavage sites (representing the two cleavage axes) are denoted by the open circles and the filled triangles. In this representation, the cleavage sites can be observed to be on the crossing-over strands of the component branches of the junction. Panels B and C: the binding of one enzyme molecule (B) and sequentially a second enzyme molecule (C) at the crossover is illustrated. The sequential binding of catalytic units along one cleavage axis was proposed primarily on the basis of earlier results (see Chapter III), which defined the binding domain of the enzyme to include regions of two duplex branches and, most importantly, the one strand which tethers the branches at the junction.

binding of the enzyme.

Bibliography

Brooker, T. R. and Lehman, I. R. (1971). *J. Mol. Biol.* **60**, 131-149.

Center, M. and Richardson, C. (1970). *J. Biol. Chem.* **245**, 6285-6291.

De Lange, A. M., Reddy, M., Scraba, D., Upton, C. and McFadden, G. (1986). *J. Virol.* **59**, 249-259.

de Massy, B., Studier, F. W., Dorgai, L., Appelbaum, E. and Weisberg, R. A. (1984). *Cold Spring Harbor Symp. Quant. Biol.* **49**, 715-726.

de Massy, B., Weisberg, R. A. and Studier, F. W. (1987). *J. Mol. Biol.* **193**, 359-376.

Dressler, E. and Potter, H. A. (1982). *Ann. Rev. Biochem.* **51**, 727-761.

Gough, G. W. and Lilley, D. M. J. (1985). *Nature* **313**, 14-156.

Holliday, R. (1964). *Genet. Res.* **5**, 282-304.

Holliday, R. (1968). Genetic Recombination in Fungi. *In* Replication and Recombination of Genetic Material, W. J. Peacock and K. D. Brook, eds. Canberra: Australian Acad. of Sci. pp.157-174.

Holliday, R. (1974). *Genetics* **78**, 273-287.

Hsu, P. L. and Landy, A. (1984). *Nature* **311**, 721-726.

Jensch, F. and Kemper, B. (1986). *EMBO J.* **5**, 181-189.

Keller, W. (1975). *Proc. Nat. Acad. Sci. U.S.A.* **72**, 4876-4880.

Kemper, B., Jensch, F., v. Depka-Prondzynski, M., Fritz, H.-J., Borgmeyer, U. and Mizuuchi, K. (1984). *Cold Spring Harbor Symp. Quant. Biol.* **49**, 815-825.

Kowalski, D. (1986). *Nucl. Acids Res.* **14**, 8949-8962.

Langridge, J., Langridge, P. and Bergquist, P. L. (1980). *Anal. Biochem.* **103**, 264-271.

Lilley, D. M. J. and Kemper, B. (1984). *Cell* **36**, 413-422.

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

Maxam, A. M. and Gilbert, W. (1980). *Methods Enzymol.* **65**, 499-560.

Meselson, M. (1972). *J. Mol. Biol.* **71**, 795-798.

Meselson, M. S. and Radding, C. M. (1975). *Proc. Nat. Acad. Sci. U.S.A.* **72**, 358-361.

Mizuuchi, K., Kemper, B., Hays, J. and Weisberg, R. (1982a). *Cell* **29**, 357-365.

Mizuuchi, K., Mizuuchi, M. and Gellert, M. (1982b). *J. Mol. Biol.* **156**, 229-243.

Panayotatos, N. and Wells, R. D. (1981). *Nature (London)* **289**, 466-470.

Pulleyblank, D. E., Shure, M., Tang, D., Vinograd, J. and Vosberg, H.-P. (1975). *Proc. Nat. Acad. Sci. U.S.A.* **72**, 4280-4284.

Robinson, B. H. and Seeman, N. (1987). *Biophys. J.* **51**, 611-626.

Sadowski, P. (1971). *J. Biol. Chem.* **246**, 209-216.

Sigal, N. and Alberts, B. (1972). *J. Mol. Biol.* **71**, 789-793.

Sobell, H. M. (1974). "Concerning the stereochemistry of strand equivalence in genetic recombination.". *In* Molecular Mechanisms in Genetic Recombination, R. F. Grell ed., Oak Ridge Symposium held at Gatlinburg Tennessee, April 1-4, 1974 (Plenum Publishing Co., New York) pp. 433-438.

Szostak, J. W., Orr-Weaver, T. L. and Rothstein, R. J. (1983). *Cell* **33**, 25-35.

Symington, L. S. and Kolodner, R. (1985). *Proc. Nat. Acad. Sci. U.S.A.* **82**, 7247-7251.

Thompson, B. J., Camien, M. N. and Warner, R. C. (1986). *Proc. Nat. Acad. Sci. U. S. A.* **73**, 2299-2303.

Warner, R. C., Fishel, R. and Wheeler, F. (1978). *Cold Spring Harbor Symp. Quant. Biol.* **43**, 957-968.

West, S. C. and Korner, A. (1985). *Proc. Nat. Acad. Sci. U.S.A.* **82**, 6445-6449.

V. DISCUSSION AND CONCLUSIONS

Cruciformation and telomere resolution. Our approach has been to characterize cruciformation within the cloned SFV telomere sequences *in vitro* in order to assess the probability that spontaneous cruciform extrusion represents an intermediate reaction in the resolution of replicated telomeres. Having accomplished this, we can make conjectures, on the basis of the available evidence, about *in vivo* circumstances that could exist to regulate or promote cruciformation were it, in fact, to be associated with telomere replication. The critical observations that form the basis of our conclusions are outlined below:

1. Plasmid bearing the cloned telomere concatameric junction (pSAlB.56A) was inviable in *rec+* bacteria and the inverted repeat sequences were unstable even in recombination-deficient hosts. An *in vivo* deletion derivative of pSAlB.56A was identified and was generated by a recombination event between direct repeats within the cloned telomere sequences (see appendix 1). In this respect, the cloned inverted repeat behaved much like any other large inverted repeat studied *in vivo*.

2. Plasmid bearing the telomere sequences characteristically displayed a unimodal distribution of superhelical densities the average of which was typical for plasmid DNAs isolated from bacteria. Therefore, no *in vivo* cruciformation was assumed to have occurred because this would more than likely have led to compensatory supercoiling and a higher titratable superhelical density in those affected plasmids.

3. Cruciformation in the cloned SFV telomere sequences had a high energy of

formation. It was twice that observed for palindromes and the inverted repeats that typically exist in bacteria. There is some doubt whether native superhelical torsion would be high enough to stabilize this cruciform (Lilley, 1986). The cruciform extruded from the concatameric junction of vaccinia virus required even a greater supercoiling energy for stabilization (see Appendix 3).

4. The high energy of formation for the SFV cruciform was attributed to the presence of the extra-helical bases in the cruciform hairpins. The deletion of dyad sequences containing the potential extra-helical bases reduced the energy of formation to typical values while, at the same time, only modestly decreasing the efficacy with which the telomeres were resolved in the *in vivo* assays (DeLange and McFadden, 1987).

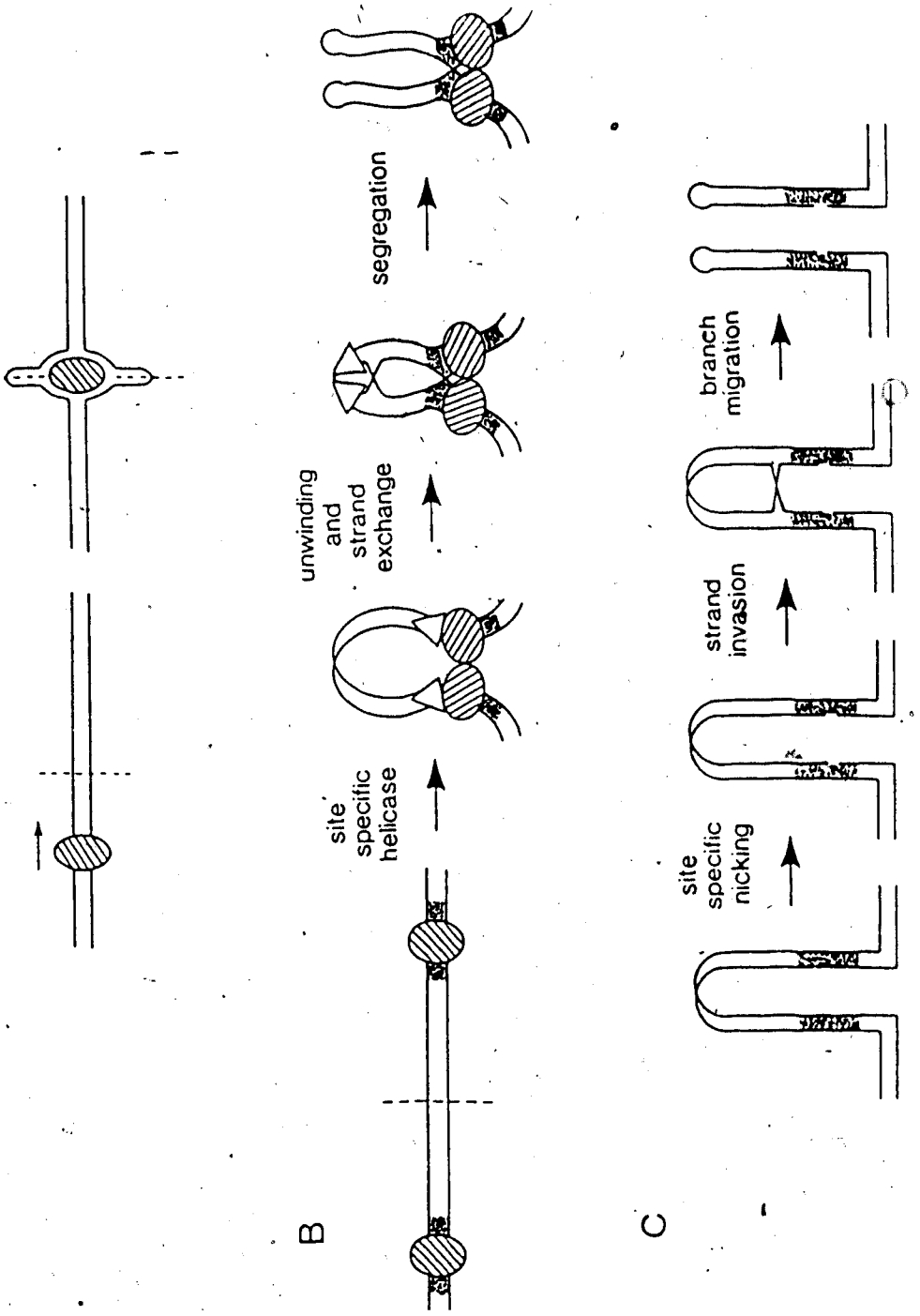
5. Cruciformation in the SFV telomere sequences was facile, indicative of a low energy of activation, though no measureable energy of activation was obtained from our experiments. A low energy barrier to cruciform extrusion is typical of A+T-rich sequences.

Despite models that invoke cruciformation to explain the instability and inviability of large inverted repeats and the similar behaviour of the SFV telomere inverted repeat, the primary conclusion drawn is that cruciformation within the SFV telomere sequences does not occur during propagation in bacteria and is not likely to occur spontaneously in poxvirus-infected cells. The energy of formation related to cruciformation is prohibitively high. Moreover, since equally unstable inverted repeat sequences with more typical energies of formation are known and recombination mechanisms are involved in plasmid metabolism and could mediate the observed biological effects of inverted repeats (Warren

and Green, 1985), it is possible that cruciformation plays no role in determining plasmid instability or inviability. However, this does not necessarily preclude cruciformation *in vivo*. It may occur transiently and may be catalyzed in a manner described in Chapter II. McFadden *et al.* (unpublished results) were able to replace up to 50 base pairs at the axis of the SFV inverted repeat with non-palindromic DNA and still achieve significant levels of resolution *in vivo*. This observation poses a serious challenge to the cruciform extrusion model. However, the result is remarkably like that of Warren and Green (1985) who demonstrated that a similar replacement was not sufficient to prevent plasmid inviablity. The latter authors proposed that cruciformation could still occur by catalysis such that the kinetic barrier to cruciformation could be overcome.

Catalysis could be specific or non-specific. An example of a non-specific mechanism of catalysis would be extrusion that accompanied the helicase-promoted unwinding of DNA during transcription or replication (Figure V.1A). The major consideration with this model is the stabilization of the cruciform once it is extruded. Telomere binding proteins could fulfill this role. Alternatively, catalysis of cruciformation could be the function of a specific helicase (Figure V.1B). With reference to the SFV telomere sequences, protein bound at the core target sequences could sequester the telomeric DNA in a loop and direct a helicase to the telomere sequences. Subsequent unwinding, strand exchange and topoisomerase aetivity to unlink the DNA in the loop would generate a resolvable cruciform structure. The protein bound at the core regions could stabilize the structure or resolution could be effected immediately by a similarly targetted Holliday-resolving enzyme. A similar need for a helicase exists in the Bateman model. As indicated, atleast 100 base pairs of sequence must be unwound between the hypothetical resolving nicks in order to separate the daughter telomeres. In any event, a viral-encoded helicase has yet to

Figure V.1. Models of catalyzed cruciform extrusion and telomere resolution. Panel A: an example of non-specific catalysis of extrusion at an inverted repeat sequence caused by the action of an unwinding protein (thatched egg). Panel B: an example of a specific catalyzed extrusion mechanism acting at an SFV concatameric junction. In this case, proteins bind to the core target regions (shaded sequences) and loop the intervening DNA. A helicase (triangles) binds and unwinds the intervening sequence permitting strand exchange. The cruciform structure is stabilized by the core-binding proteins and ultimately is resolved to daughter telomeres. Panel C: resolution without cruciformation is demonstrated within the SFV concatameric junction. Homologous arms of the inverted repeat pair, nicks are introduced in the core target regions and strand exchange occurs. Branch migration resolves the structure.


segregation
unwinding and strand exchange
site specific helicase
B
branch migration
strand invasion
site specific nicking
C

be identified. Warren and Green (1985) described a recombination pathway for the catalysis of cruciform extrusion. A minor variation to this scheme transforms it into an even simpler model for telomere resolution (Figure V.1C). Homologous alignment of the arms of the inverted repeat, nicking at the core target sequence, strand invasion and branch migration coincident with unlinking by a topoisomerase I resolves the daughter telomeres. Poxvirus-infected cell cytoplasms are extremely active in viral recombination (Evans *et al.*, 1987).

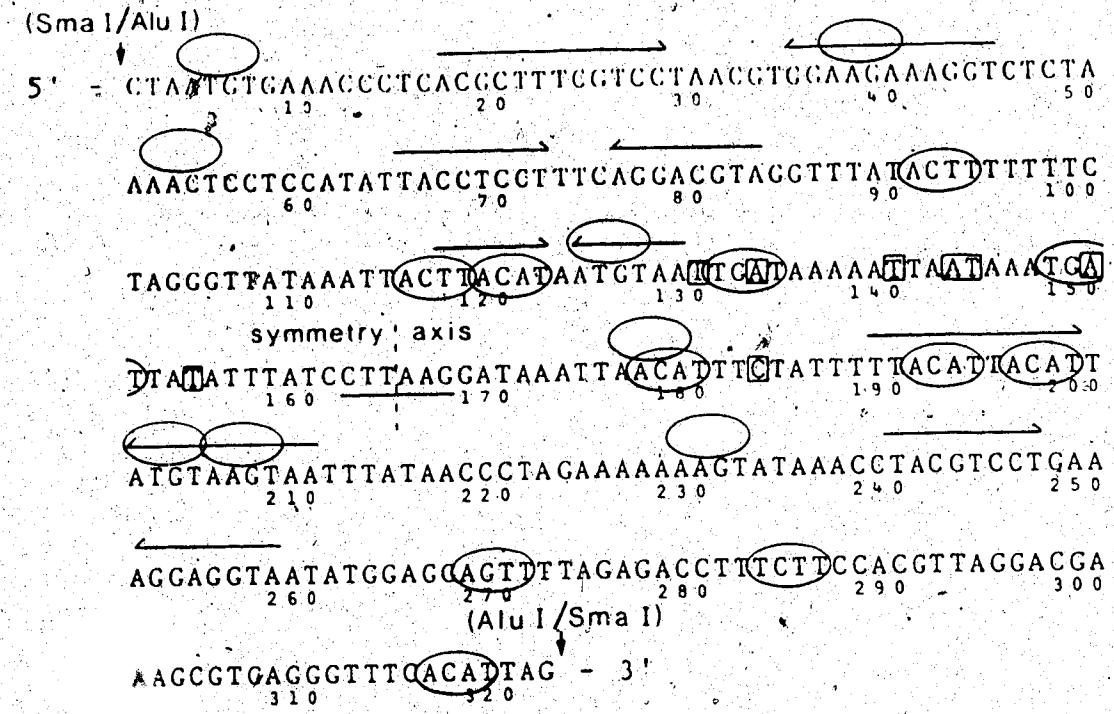
Potentially, the Bateman model could be distinguished from the other two models on the basis of its lack of dependence on topoisomerase activity. No *in vitro* system for poxvirus replication is available and a genetic analysis is not yet possible in order to test this directly, but a survey of the telomere sequences reveals a provocative distribution of eukaryotic type I topoisomerase recognition sites (Figure V.2). In SFV, tandemly repeated copies of the topoisomerse I consensus tetranucleotide 5'-A/T C/G A/T T are found close to the dyad axis and overlapping one of the internal inverted repeat sequences. The association with the small inverted repeat may not be coincidental. In eukaryotic systems this is often the case (reviewed by Wang, 1985; Hyrien *et al.*, 1987). Other topoisomerase I sites apear to be randomly scattered throughout the sequence at expected intervals. In vacccinia virus the distribution of topoisomerase I sites is different. Here there are three conspicuous sites tandemly arranged precisely at the telomere dyad axis. A viral topoisomerase I is encapsulated by vaccinia virus (Moss, 1986).

The requirement for strand separation in long stretches of sequence during resolution may explain the conservation A+T-rich sequences in poxviral telomeres. In regions of DNA where unwinding is a prerequisite to biological function ie. at replication origins and

Figure V.2. Eukaryotic topoisomerase I recognition sites in the SFV and vaccinia virus concatameric junction sequences. The topoisomerase I consensus sequence 5' A/T C/G A/T T is identified for the cloned inserts of pSAlB.56A (SFV) and pVCB.5 (vaccinia). Ellipses above the sequences denote the presence of recognition sites on the complementary strands. Small inverted repeats are indicated. Asymmetric bases are boxed.

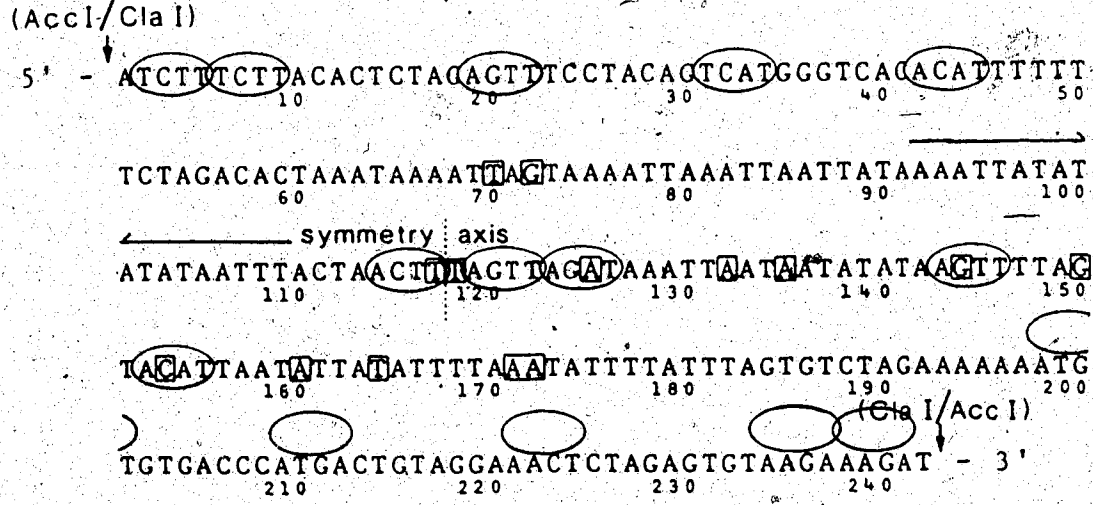
A

pSAIB-56A Insert



B

pVCB-5 Insert



transcription promoters, A+T-rich sequences are often found. Furthermore, recombination hotspots have been identified that are A+T-rich (Hyrien *et al.*, 1987) and A+T stretches have long been associated with early melting regions (Sullivan and Lilley, 1986). Consequently, it is reasonable to assume that these sequences have been conserved to to lower the energy barrier to strand separation. The conservation of extra-helical bases must have no direct bearing on resolution. Rather, they may exist to fulfill some other viral function, such as defining binding domains for proteins that organize or package the DNA. One can not discount the elementary interpretation that they exist purely to discourage cruciformation in the A+T-rich telomere sequences because resolution involves lineform sequences.

Identification of the mechanism of resolution will require the characterization of the proteins involved. A site-specific endonuclease and helicase activity (acting alone or as part of a normal biological process) are required enzymes. Topoisomerase activity, recombination proteins that mediate branch migration and telomere-binding proteins could potentially be involved. It bears repeating that none of these proteins have been identified and proven to be involved in resolution. An *in vitro* system is being sought and assays for specific telomere-binding proteins and resolution-related activities in infected-cell extracts have been undertaken and must continue to be pursued.

T7 endonuclease I. Efforts to characterize T7 endonuclease I were rewarded with a clearer understanding of the specificity of the enzyme. The enzyme cleaves the bridging strand that links two non-base-stacked DNA branches. In the most preferred substrate the branches are composed of duplex DNA. The cleavage site is predominantly one nucleotide 5' from the branch point, regardless of the complexity of the branched

molecule. Nicked linear duplex DNA is not a substrate for the enzyme (de Massy *et al.*, 1987), indicating that cleavage is dependent upon the recognition of a distorted phosphodiester backbone. T7 endonuclease I activity can be inhibited by single-stranded DNA (see appendix 2). This result is consistent with the observation that the enzyme can recognize and cleave single-stranded branches (though the cleaved phosphodiester is, in fact, in a duplex region). It has been suggested by others (Jensch and Kemper, 1986) that the reported specificity of T7 endonuclease I for single-stranded DNA substrates (Center and Richardson, 1970; Sadowski, 1971) is more accurately interpretted as cleavage at the base of stem-and-loop structures present in such DNAs. These structures would be analogous to the oligonucleotide dimers which were cleaved by T7 endonuclease I (Chapter III).

Fixed branchpoint junctions are more readily cleaved by T7 endonuclease I than are semi-mobile junctions. It is hard to predict why given the limited data. Consider though, the predictions of Robinson and Seeman (1986). If the stable form of a Holliday structure is the Sobell conformation and, as we suggest, this is the cleavable form, yet branch migration occurs in the UHA form for freely-branched structures, then at any one time it is obvious that a significant fraction of the semi-mobile Holliday analogs could be in a non-cleavable conformation. The efficient cleavage of trimer molecules suggests that flexibility at the branchpoint is a major-factor. The branchpoints in these structures are "open" in that they are free to rotate or flex at the one bridging strand. On the other hand, the Y-structure will be less flexible because the branchpoint is closed and, not surprisingly, it was cleaved much less efficiently. Flexibility would promote adoption of the most preferred cleavage conformation.

The gene for T7 endonuclease I has been cloned (de Massy *et al.*, 1987) and the enzyme expressed from the cloned DNA has been purified and characterized. We, too, have purified the cloned version of the enzyme and have found that its characteristics are no diferent than those of the preparation obtained form Paul Sadowski. Our use of defined substrates has improved the characterization of the enzyme to the point, we believe, that it is the most defined, general Holliday-resolving enzyme known. Now that it has been cloned, is easily purified and is stable, it is the most attractive probe for branched DNA structures. Moreover, it would be an interesting enzyme to crystallize in order to elucidate the binding interaction with branched DNA substrates.

A more precise definition of the sequence specificity of the enzyme and a measurement of the enzymes ability to cleave further than one nucleotide form the branchpoint are still needed. Manipulation of the sequence at the branchpoint or perhaps the discovery and judicious implementation of non-cleavable phosphodiesters (using phosphorothioate nucleotides) could be beneficial. The use of defined substrates will also permit an analysis of reaction kinetics and minimal recognition attributes of the substrate. They may also be used to more clearly define the resolving mechanism. Are the two cuts made simultaneously, or are they coupled somewhat less rigorously? When we more carefully studied the kinetics of resolution, it did appear that the amount of nicked plasmid increased early in the reactions and was converted to linear, resolved plasmid later in the reaction. This suggested that singly-nicked DNA was an intermediate (reactions were stopped with SDS). A more careful analysis is warranted.

Branch migration in Holliday crossovers. It is perhaps pretentious to suggest that our work has contributed to a greater understanding to the mechanism of branch migration in

topologically constrained DNA. In fact, we have performed only one experiment using one DNA substrate and a probe that has not been completely characterized. But the approach was significant for one reason. Assumptions have been made regarding branch migration which may not be relevent. DNA topology will impinge on branch migration through effects on the structure of the Holliday crossover. Our discussions and model building in Chapter IV draw upon what is known and surmised about structure and branchpoint migration and reinterprets this in light of our one striking observation. Simply by introducing the concept of topology and in particular the change in the twist of the DNA during migration, we have identified a potential isomerization event and present it in the context of a molecular model.

Our model for branchpoint migration in topologically constrained DNA accepts unequivocably that the conformation of the junction isomerizes in direct relation to the change in linking difference accompanying migration. Overlook the implication of topology and infer, as did Meselson and Radding (1975), that the UHA structure exists exclusive of other forms during branch migration and our data is clearly inconsistent. The unique feature of branch migration in pSAlB.56A is exactly that topology can not be ignored and for this reason it is a novel and important experimental observation. In a larger molecule the influence of topology would presumably diminish simply because more freedom for rotary diffusion would exist. Whether this is the situation *in vivo* is not yet known, but the obvious influence of topology on the metabolism of DNA (reviewed by Wang, 1985) and the sequestering of topological domains in eukaryotic DNA suggests that the assumption that recombining DNA is free to rotary diffuse may not be warranted.

The results of Chapters III and IV do not establish what the stable conformation of a

Holliday structure is. We have inferred structure from the cleavage specificity of T7 endonuclease I, but acknowledge that the enzyme could trap or induce a specific conformation on the molecule. Interestingly though, the binding energy of the interaction must not be sufficient to overcome the torsional strain that imposes the ten nucleotide cleavage specificity. The tetrahedral arrangement of branches as proposed by Gough and Lilley (1985) is an attractive model because of the maximal separation of the negatively-charged phosphodiester backbones. But I am tempted to invoke the Sobell model because the relative orientation of the arms need not change during branchpoint migration (they would twist instead) and fewer steric and diffusional complications would have to be considered. I can only answer Robinson and Seeman's concern about viscous drag impeding branchpoint migration by suggesting that it is this influence that makes the calculated rates of branchpoint migration so much less than the theoretically derived rates. Our results must be verified using other DNA substrates. In this way, some confidence can be gained with respect to a more general relationship between topology and crossover structure. Unfortunately we are handicapped by a limited number of probes for the actual DNA crossover. Perhaps the newly discovered monoclonal antibodies directed against cruciform structures(which themselves must be better characterized) will be useful in this respect.

Bibliography

Center, M. and Richardson, C. (1970). *J. Biol. Chem.* **245**, 6285-6291.

DeLange, A. M. and McFadden, G. (1987). *P.N.A.S. USA* **83**, 614-618.

de Massy, B., Weisberg, R. A. and Studier, F. W. (1987). *J. Mol. Biol.* **193**, 359-376.

Evans, D. H., Stuart, D. and McFadden, G. (1987). *Mol. Cell. Biol.* (submitted).

Gough, G. W. and Lilley, D. M. J. (1985). *Nature* **313**, 154-156.

Hyrien, O., Debatisse, M., Buttin, G. and de Saint Vincent, B. R. (1987). *EMBO J.* 6, 2401-2408.

Jensch, F. and Kemper, B. (1986). *Nature* **311**, 721-726.

Lilley, D. M. J. (1986). *Biochem. Soc. Trans.* **14**, 489-493.

Meselson, M. and Radding, C. M. (1975). *P.N.A.S. USA* **72**, 358-361.

Moss, B. (1986). *In* Fundamental Virology, pp. 637-656. B. N. Fields and D. M. Knipe, eds. Raven Press, New York.

Robinson, B. H. and Seeman, N. (1987). *Biophys. J.* **51**, 611-626.

Sadowski, P. (1971). *J. Biol. Chem.* **246**, 209-216.

Sullivan, K. M. and Lilley, D. M. J. (1986). *Cell* **47**, 817-827.

Wang, J. C. (1985). *Ann. Rev. Biochem.* **54**, 665-697.

Warren, G. J. and Green, R. L. (1985). *J. Bact.* **161**, 1103-1111.

Appendix 1

The Δ260 derivative of pSA1B.56A: Plasmid pSA1B.56A, propagated in recombination-deficient bacteria, throws off a deleted form that bears a shorter version of the cloned telomere inverted repeat. The sequence is shown below in Figure A1.1:

5' CTAAT GTGAA ACCCT CACGC

G TA ACGT A
TTTCT TCCAC GTTAG GACGA

AAGCG TGAGG GTTTC ACATT AG

Figure A1.1. The nucleotide sequence of the Δ260 form of pSA1B.56A. Dual assignments reflect the presence of two possible sequences.

Form Δ260 was purifed from low-melting agarose and its insert sequenced by the dideoxy-chain termination method (Sanger, F., Coulson, A. R., Barell, B. G., Smith, A. J. H. and Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161-178). The appearance of alternative sequences reflects the nature of the recombination event. Recombination has occurred between homologous arms of the internal inverted repeat "c" (Figure II.9) in the pSA1B.56A insert. Because the homologous arms of repeat "c" are not perfectly homologous, alternative sequences will arise depending on the position of the crossover. This is why the asymmetric sequence at the dyad axis of repeat "c" is faithfully presented in its two complementary forms.

Appendix 2

Inhibition of T7 endonuclease I cleavage of pSAlB.56A with single-stranded DNA. T7 endonuclease I was titrated using pSAlB.56A as a substrate and measuring its conversion to a linear form. Increasing amounts of synthetic oligonucleotide 3 (Chapter III) were added to the reaction in order to determine its ability to inhibit the resolution reaction. Reactions included 50ng of pSAlB.56A, and varying amounts of oligo 3, in 9µl of resolvase buffer and were initiated with the addition of 1µl of enzyme diluted 1:10 in diluent (as described in Chapter III). Reactions were stopped with SDS and EDTA after 5 minutes, then electrophoresed in agarose gels. Relative amounts of circular and linear plasmid were estimated and the results are shown in Figure A2.1. Oligo 3 was designed not to have any appreciable base complementarity and was not cleaved by T7 endonuclease I under similar conditions but in the absence of pSAlB.56A (Chapter III). The inability to detect any inhibition at low levels of oligo 3 could reflect the fact that excess enzyme was used and reactions had gone to completion.

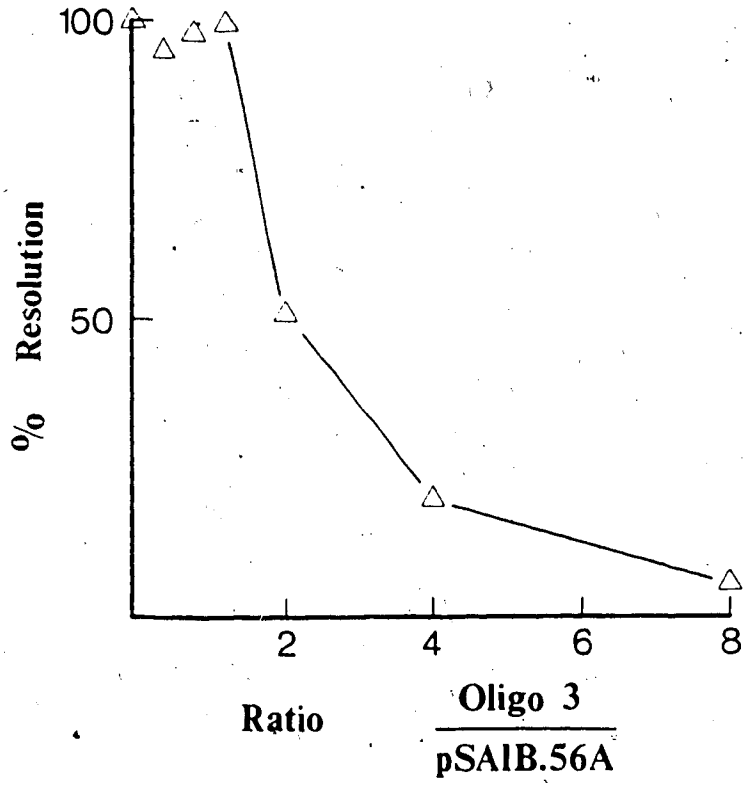


Figure A2.1. Inhibition of T7 endonuclease I with single-stranded DNA. Oligo 3 refers to the synthetic oligonucleotide 3 described in Chapter III. The ratio of the DNAs is based on weight. Resolution implies the specific cleavage of the cruciform junction in pSAlB.56A, converting the circular plasmid to the linear form with nicked hairpin termini.

Appendix 3

Cruciformation in Vaccinia virus telomere sequences. Vaccinia telomeres were cloned originally in yeast (DeLange, A. M., Futcher, B., Morgan, A. R. and McFadden, G. (1984) *Gene* **27**, 13-21) and the vaccinia telomere concatameric junction was subsequently cloned into pUC13 (DeLange, A. M., Reddy, M., Scraba, D., Upton, C. and McFadden, G. (1986) *J. Virol.* **59**, 249-259). One such bacterial plasmid, pVCB.5, bore a 242 base pair inverted repeat of the telomere sequences (see Figure V.2). Unlike plasmid pSAlB.56A, pVCB.5 was isolated from bacteria with the viral sequences predominantly in the lineform configuration (~90%). Plasmid pVCB.5 was partially relaxed with eukaryotic topoisomerase I and analyzed in the two-dimensional agarose gel electrophoresis system, described in Chapter II, in order to determine the energy of formation of the extruded cruciforms. Cruciformation in pVCB.5 had been verified in the original publication by electron microscopy and other means (DeLange *et al.*, 1986) and its behaviour in the 2-D gels clearly demonstrated a structural transition capable of relaxing highly supercoiled topoisomers. The computed energy of formation related to the transition was 58 kcal/mol.

There are 12 extra-helical bases in vaccinia telomeres, as opposed to just 8 in SFV telomeres. This difference did not seem sufficient to account for the large difference in stabilization energies. A rough estimate of the incremental difference that one extra-helical base makes to the energy calculations would be about 2 kcal/mol (based on the data in Table II.1, for SFV telomere sequences). Therefore, a ΔG_f of 48 to 50 kcal/mol might have been more reasonable. In fact, what likely contributed to the higher ΔG_f value was a pre-transitional event that adsorbed an estimated 4 to 6 superturns. This structural

transition was not the previously described vector-related transition because gel migration shifts at 20°C were significantly greater and it occurred at the higher estimated supercoiling energy of 45 kcal/mol. It may represent a structural transition within the long A+T stretch in the vaccinia telomere sequence. In any event, the true ΔG_f for the vaccinia cruciform is probably much closer to 45 kcal/mol, but even at this value and in light of the additional structural transition, cruciformation seems highly unlikely *in vivo* unless the cruciform structure is stabilized by something other than superhelical torsion.