THE MECHANISM OF SITE-SPECIFIC RECOMBINATION ENCODED BY Tn3

A thesis submitted for the Degree of

Doctor of Philosophy at the

University of Glasgow

by

Paul J. Dyson

Institute of Genetics

Church Street

Glasgow

Scotland

January 1984

The research reported in this thesis is my own original work except where otherwise stated, and has not been submitted for any other degree.

CONTENTS

ABBREVIATIONS

(i) Chemicals.

Ac	acetate
APS	ammonium persulphate
ATP	adenosine triphosphate
DNA	deoxyribonucleic acid
DS	Denhardts solution
DTT	dithiothreitol
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphoshate
dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
dH_2O	distilled water
EDTA	ethylenediaminetetra-acetic acid (disodium salt)
EtBr	ethidium bromide
EtOH	ethanol
oc	covalently closed relaxed DNA
PVP	polyvinylpyrrolidine
RNA	ribonucleic acid
sc	supercoiled DNA
SDS	sodium dodecylsulphate
TEMED	N,N,N',N'-tetramethyl ethylenediamine
Tris	tris(hydroxymethyl)aminomethane

(ii) Antibiotics

Ap	Ampicillin
Cm	Chloramphenicol
Str	Streptomycin(chromosomal resistance)

Tc	Tetracycline
Tp	Trimethoprim

(iii) Phenotype

X^r	resistance to X
X^s	sensitivity to X
Ori	origin of replication
res	resolution site
Rec^+	recombination proficient

(iv) Measurements

mA	milliamps
bp	basepair
kb	kilobasepair
^oC	degrees centigrade
cps	counts per second
cpm	counts per minute
hr	hour
g	centrifugal force equal to gravitational acceleration
g	gramme
mg	milligramme
ug	microgramme
ng	nanogramme
l	litre
ml	millilitre
ul	microlitre
M	Molar(moles per litre)
mM	millimolar

uM	micromolar
cm	centimetre
mm	millimetre
min	minute
pH	acidity [negative $\log_{10}$(Molar concentration H^+ ions)]
sec	second
V	volts
w/v	weight per volume
v/v	volume per volume
W	watts

(v) Miscellaneous.

Fig.	figure
LH	left-hand
RH	right-hand
log	logarithm
Tn	transposon
wt	wild-type

Acknowledgements

I must thank all members of the Genetics Department, Glasgow University, both past and present, for invaluable help and moral support. They are too numerous to mention, but a special thanks must go to the media ladies who kept my bacteria well-fed and my glass-ware clean.

I am indebted to Linda Taylor and Russell Thompson of the MRC Virology Unit, Glasgow who taught me sequencing techniques - even if my first attempt was merely to confirm a little of Gregor Sutcliffe's work.

Finally acknowledgements to two people who have particularly inspired me in the last 3+ years. Avril provided both a project to work on and a very special kind of support. Professor David Sherratt, my supervisor, offered advice, enthusiasm and all kinds of encouragement. Above all, he opened my eyes to an intuitive insight into the ways of DNA within the organism.

My benefactors, the MRC, provided support for 3 years for which I am grateful.

SUMMARY

The bacterial ampicillin resistance transposon Tn3 encodes a site-specific recombination system which mediates the second step in its transposition pathway: the resolution of cointegrates (Arthur and Sherratt (1979) Mol.gen.Genet. **175**, 267-274). *In vitro* analysis has shown that this recombination system is simple: dependent on the transposon encoded resolvase protein acting at *res*, a site required *in cis* (Reed (1981) Cell **25**, 713-719). Resolvase binds at *res* at three sites (Grindley *et al* (1982) Cell **30**, 713-719).

res was defined to a region of 139bp by *in vivo* analysis, which confirms that the extent of DNA bound by resolvase *in vitro* constitutes *res*. Truncated *res* regions have been analysed for resolvase binding and function. The results confirm the importance of a sequence determinant for resolvase binding, suggest non-cooperative binding at each of the three sites, but cooperativity for binding at both half-sites of the individual binding sites.

In terms of function a truncated *res* region was shown to have secondary site activity in that it could recombine with *res*-wt but not efficiently with itself. The difference in primary and secondary *res* sites is the presence of resolvase binding site III in the former. This binding site can be reintroduced at a different position to a truncated *res* to partially reconstitute primary site activity.

The normal cointegrate substrate contains directly repeated copies of *res*. Inversion substrates, containing two inverted *res* regions, were analysed for recombination. Inversion is approximately 50 times less efficient than resolution. This rate is comparable to

that of resolution of a substrate containing directly repeated secondary *res*-regions.

These results, taken in context with evidence from other workers on both Tn3 encoded site-specific recombination and bacteriophage lambda site-specific recombination, suggest that the mechanism by which two *res* regions are brought together prior to recombination is determined by the higher-order structure formed by resolvase and *res*.

CHAPTER 1

Introduction

1.1 Genetic recombination: mechanistic and evolutionary distinctions.

Genetic recombination involves the reassortment of DNA strands between two participating duplexes which may be derived from the same molecule (intramolecular recombination), or two different molecules (intermolecular recombination). Within prokaryotes genetic recombination can be divided into three types on the basis of the extent of DNA sequence homology between each duplex which is required to promote recombination: generalised recombination, site-specific recombination, and transposition.

Generalised recombination depends on there being extensive regions of homology between the participating duplexes so that recombination intermediates can be recognised and acted upon by the recombination enzyme machinery: the stability of intermediates is proportional to the extent of homology and results from base pairing between complementary single-stranded regions of DNA derived from each participating duplex. Generalised recombination can act on any regions of sequence homology and is processive, depending on the function of a number of unlinked gene products (in Escherichia coli most notably RecA).

While there is a requirement for some sequence homology between the participating duplexes for site-specific recombination, the region of homology is much smaller and appears not to be involved so much in intermediate base pairing but for firstly binding of a recombination protein and secondly sequence-specific strand cleavage and ligation mediated by that protein giving rise to reciprocal strand exchange. Recombination intermediates are much more transitory than for generalised recombination suggesting a concerted recombination

mechanism mediated by the recombination protein. This recombination protein is often encoded by a gene adjacent to the recombination site.

The third type of recombination event allows one DNA sequence to be inserted into another without reliance on sequence homology. Transposable elements are discrete units of DNA which appear to encode the recombination machinery required for transposition themselves. Each insertion event involves the joining of exactly the same nucleotides at the end of the element to a target DNA molecule broken by staggered nicks. The strand cleavage-ligation events are assumed to be mediated by the transposon-encoded transposase protein which acts at the elements' termini, typically inverted repeats, and presumably introduces a staggered break at the target sequence for which there is little or no sequence homology. In this sense transpositional recombination is site-specific in terms of one participating duplex but not for the other. Current models of transposition postulate DNA breakage, single-strand exchange, and rejoining, followed by replication of the transposable element. A key difference between these models and site-specific recombination as defined above is that the initial strand exchange is not thought to be reciprocal in transposition. Non-reciprocal exchange leaves a free 3' end that serves as a primer for strand elongation and hence replication of the element as a consequence of transposition. Transposable elements also promote other types of DNA rearrangements such as deletions, inversions and transposition of the intervening sequences between the two copies of an element, and these are mechanistically related to transposition.

In an evolutionary context, by shuffling genes recombination

allows favourable and unfavourable mutations to be separated and tested as individual units in new assortments. Just as it provides a means to spread favourable alleles, it can eliminate an unfavourable allele without the loss of other genes with which this allele may have been associated in the past. Any mechanism that produces changes in DNA must have occasionally contributed to evolution by creating adaptive gene combinations that have survived. The respective roles in evolution of the three types of recombination are considered different however.

With due regard to the pitfalls of semantics a 'conventional' genome is regarded as a generally static arrangement of sequences. This concept allows the construction of genetic maps which identify the loci at which known genes reside, which in turn implies that other sequences remain at constant positions. The stability of genetic organisation is indicated by the retention of linkage relationships even after speciation; for example, a very similar genetic map describes the different bacterial species E.coli and Salmonella typhimurium. Within the framework of a static genome the reassortment of alleles can most easily be explained by the mechanism of generalised recombination, generally accepted as the recombinational work-horse of evolution. Generalised recombination can also utilise close regions of homology to either duplicate or delete intervening sequences via unequal exchange. Moreover gene conversion is promoted by generalised recombination.

A diploid eukaryote can successfully reassort alleles via generalised recombination between two homologous chromosomes. This process is a key feature of meiosis and underlines the importance of

sex in allelic reassortment in populations as a whole. Within prokaryotes, in the absence of sex, extrachromosomal elements are responsible for mediating genetic exchanges. Plasmids allow conjugation of bacteria, while phages are released by infection. Both may occasionally transfer host genes along with the autonomous replicon which then provide a substrate for generalised recombination between the recipient chromosome and incoming episome.

Transposition provides a powerful means by which sequences may be translocated between non-homologous regions of the same chromosome, and between chromosome and plasmid. Concurrent with translocation is the opportunity for duplication of essential sequences which can then provide material for further evolution by mutation and rearrangement without loss of original function. Moreover transposable elements can provoke cellular systems by functioning as portable regions of homology: two copies of an element at different locations (even on different chromosomes) may provide sites for generalised recombination. Such exchanges could result in deletions, insertions, inversions or translocations of DNA.

That transposition is indirectly important to prokaryote evolution is underlined by the role of transposable elements in the evolution of plasmids. Comparisons of the genomic organisation of related plasmids show that frequently they differ from one another by insertion of transposable elements and by replicon fusions arising from transposition or generalised recombination between two homologous elements. The integration of the F plasmid of E.coli into the bacterial chromosome often occurs by a host-mediated generalised recombination between an element in the plasmid and a homologous

element in the bacterial DNA.

Whilst the role of transposition *per se* in evolution is potentially a major one there remains the doubt as to whether the process makes significant contribution to progressive evolution, or if the transposition events observed within the laboratory are simply constituting a cyclical change in the numbers and locations of elements. But the potential flexibility accorded to the genome by the presence of transposable elements within it is enormous. That flexibility extends enormously the range of permutations and combinations of alleles that can be tried and tested over the course of evolution. Perhaps a clearer definition of the evolutionary roles of generalised recombination and transposition comes in context with the distinction between an effect and a function. In the highly specialised case of meiotic crossing over in eukaryotes an apparent function of generalised recombination is the reassortment of alleles. In *E.coli*, which may be representative of many prokaryotes, the case is probably the same although an associated DNA repair function is shared by the recombination machinery. It is not nearly as plausible to suggest that the function of transposition is to reassort alleles. Instead such reassortment is an effect of the ability of each discrete transposable element to transpose and spread itself duplicatively within both individual and population.

Indeed an argument has been expounded which maintains that transposable elements need have no phenotypic or evolutionary function to maintain themselves within a particular host. This is the 'selfish DNA' hypothesis which suggests that 'natural selection operating within genomes will inevitably result in the appearance of DNAs with

no phenotypic expression whose only "function" is survival within genomes' (Doolittle and Sapienza 1980). An alternative to this view is that transposable elements might provide beneficial phenotypic effects at two levels: the level of individual selection, which involves direct effects of the element on the fitness of its host; and at the level of the population, which involves the capacity of such elements to generate and disseminate novel gene combinations (Campbell 1981).

There is now evidence to suggest this latter view may be applicable to elements found within E.coli. Edlin et al (1975, 1977) and Lin et al (1977) have reported that lysogens of lambda, Mu, P1 and P2 temperate phages are all favoured over their isogenic (but non-lysogenic) counterparts in energy-limited aerobic chemostats. This is the case for Tn10 too although there is a dependence on the initial frequency of this element, and the advantage it confers to its host appears related to mutation arising from transposition (Chau and Cox 1983). Tn5 can also confer a selective advantage upon its host and in this case transposition appears not to be involved. Either the Tn5 encoded transposase or repressor seem to be necessary for this effect (Biel and Hartl 1983).

The distinction between progressive evolutionary changes and changes of a more cyclical nature is more defined when comparing generalised and site-specific recombination. The acquisition of lambda as a lysogen by an E.coli strain depends on site-specific recombination. But instead of speaking of this process as evolution, it is generally regarded that the transition from phage to prophage and back again defines the life cycle of the virus. The requirement for both sites involved in site-specific to be defined places

constraints on the ubiquity of this process which are not necessarily influential on transposition. These constraints are why site-specific recombination has been adopted for precise biological processes since the fidelity of these processes is thereby maintained. In several instances site-specific recombination is utilised in the dissemination of phages and transposons: in this respect it has a contributory feature to evolutionary change.

1.2 A catalogue of site-specific recombination systems.

Three types of recombination event can be sponsored by site-specific recombination depending on the relative positions of both participating sites: intermolecular fusion, intramolecular inversion and intramolecular deletion (resolution) as depicted in Fig. 1.1. In fact for the systems so far analysed there are representatives for each type of event (Table 1.1). Moreover the choice of recombination event is crucial to the biological function of each system, and despite similarities between these recombination systems, they have evolved diverse biological roles. The relative efficiences of these systems may reflect both the activities and amount of the recombination enzymes and the nature and position of the normal recombination site. Because the energy states of the reaction substrates and products ought to be very similar(see below), the presence of a functional site-specific recombination system should generate an equilibrium mixture of similar quantities of reactants and products. Though this appears to be the case for the inversion in the 'flip-flop' recombination systems, the other systems have evolved ways of apparently shifting this equilibrium.

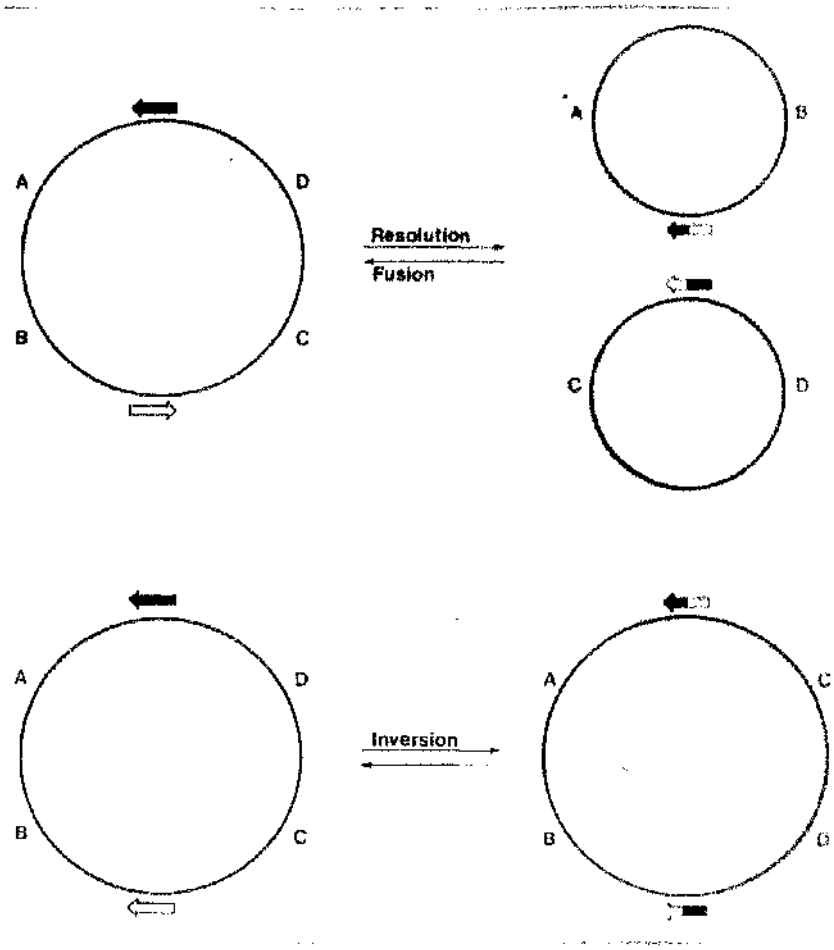


Figure 1.1 **Potential recombination events sponsored by site-specific recombination.**

⇨ Recombination site

Letters describe the fate of intervening DNA segments subsequent to recombination.

Table 1.1 **Site-specific recombination systems.**

ORGANISM	SUBSTRATE	RECOMBINATION PROTEIN(S) INVOLVED	EVENT SPONSORED	BIOLOGICAL FUNCTION	REFERENCE
Lambda	attP x attB	Int, IHF	Fusion	Integration and Excision of lambda	reviewed by Weisberg and Landy 1983
Lambda	attL x attR	Int, Xis, IHF	Resolution		
P1	loxP x loxP	Cre	Resolution	Stability of P1	Austin et al 1981
P1	loxP x loxB	Cre	Fusion	Occasional integration and Excision of P1	Hoess et al 1982
P1	loxL x loxR	Cre	Resolution		
P1	cixL x cixR	Cin	Inversion	Determines host specificity of P1	Iida et al 1982
Mu	IRL x IRR	Gin	Inversion	Determines host specificity of Mu	van de Putte et al 1980
S.typhimurium chromosome	IRL x IRR	Hin	Inversion	Antigenic phase variation	Simon et al 1980
E.coli chromosome	IRL x IRR	Pin	Inversion	Unknown	Plasterk et al 1983
Class II transposons	res x res	Resolvase	Resolution	Resolution of cointegrates: second step in transposition	Arthur and Sherratt 1979
Yeast 2u plasmid	FloxL x LloxR	FLP	Inversion	Unknown	Broach 1982

Bacteriophage lambda integration and excision is the archetype of site-specific recombination involving conservative breakage and reunion of DNA strands as first proposed by Campbell in 1962 (reviewed by Weisberg and Landy 1983). Integration of lambda *in vivo* requires lambda Int protein and the products of two host encoded genes, *him*A and *him*D, which in combination form intergration host factor (IHF). It normally occurs between *att*P, a region of about 240bp on lambda, and *att*B, a region of some 25bp on the bacterial chromosome, generating two hybrid sites *att*L and *att*R. Though *att*P and *att*B are non-identical, they share a 15bp homologous 'core' sequence within which recombination occurs. Binding of Int and IHF occurs on both arms of *att*P surrounding the core, whereas Int and IHF bind solely to the core of *att*B. The reverse recombination reaction, between *att*L and *att*R, which leads to lambda excision *in vivo*, requires Int protein, IHF and a second lambda protein specified by the *xis* gene. *In vitro*, Int promoted recombination occurs on negatively supercoiled DNA molecules containing both *att*P and *att*B, with the loss of two superhelical turns (Nash and Pollock 1983), no net DNA synthesis and no need for an exogenous energy source such as ATP. If lambda excision were simply a reversal of integration, then establishment of stable lysogeny would be difficult or impossible due to the equilibrium of reactants and products. But a measure of control on stable lysogeny arises because the asymmetric lambda recombination sites (*att*P and *att*B) generate different integration and excision substrates, which in turn allow the additional Xis protein to be required for excision.

Bacteriophage P1 exists as a plasmid in its lysogenic state and encodes a site-specific recombination enzyme, Cre, which rapidly

converts any dimeric plasmid molecules (generated during replication or homologous recombination) to monomers, by intramolecular recombination at a specific site, loxP (Austin et al 1981). It is postulated that such recombination may be important for stable maintenance of P1 as a plasmid, since its copy number control is so stringent that P1 replication is inhibited once there are two P1 replication origins in a cell. A single P1 dimer cannot be partitioned to both daughter cells. Conversion of P1 dimers to monomers is reversible but it would appear that intramolecular resolution between two loxP sites is more efficient than the intermolecular fusion recation. A model to explain this observation is presented in Chapter 8.

Occasional intermolecular recombination between loxP and loxB results in integration of P1 within the E.coli chromosome. Whereas loxP consists of a 13bp inverted repeat (binding sites for Cre; K.Abremski pers.comm.) surrounding an 8bp 'core' region (containing the crossover site), loxB can be considered an imperfect loxP consisting of a 10bp inverted repeat sharing 8bp homology with loxP, surrounding a 9bp non-identical core. The hybrid sites, loxL and loxR, are also imperfect loxP sites which in turn constrains intramolecular excisive recombination of P1 (Hoess et al 1982).

The invertible DNA regions tabulated have different functions and genetic organisation, but the actual inversion systemsare closely related in both the recombination sites and the amino acid sequence of the recombinases. Pin, Gin, Hin and Cin are more than 60% homologous at the amino acid level. To take just one example, the Hin protein can invert a 1kb region of the S.typhimurium chromosome by intramolecular

recombination across a 14bp inverted repeat to determine which of two flagellar types is synthesised (Simon et al 1980). To date the nature of the sequences which bind Hin is unknown. Inversion for these systems occurs relatively rarely; there would be little sense in rapid switching from one host range or flagellar type to another. This in turn appears to be a function of low-level expression of the recombinase genes (Plasterk et al 1983).

Transposons of the Tn3 family specify site-specific recombination systems that act on directly repeated copies of the recombination site res. Such molecules are generated in the first step of intermolecular transposition and are rapidly resolved to the normal transposition end products by the site-specific recombination (Fig. 1.2). The related transposons Tn1, Tn3 and Tn1000 share interchangeable resolvase and res sites, and are the subject of this thesis. A subgroup of ClassII elements containing Tn501 (encoding resistance to mercuric ions), Tn1721 (encoding resistance to tetracycline) and Tn21 (encoding resistance to mercuric ions, sulphonamides and streptomycin) share interchangeable site-specific recombination functions between themselves but not with Tn3 (Diver et al 1983).

There is one categorised example of a eukaryotic site-specific recombination system. Most strains of Saccharomyces cerevisiae harbour several copies of a plasmid termed the 2 micron circle DNA. This plasmid contains two 599bp precise inverted repeats across which a site-specific inversion occurs in vivo, promoted by the plasmid encoded function called FLP (Broach 1982). A functional recombination site is contained within a 63bp region which contains three binding sites for FLP, arranged as two direct repeats and one inverted repeat

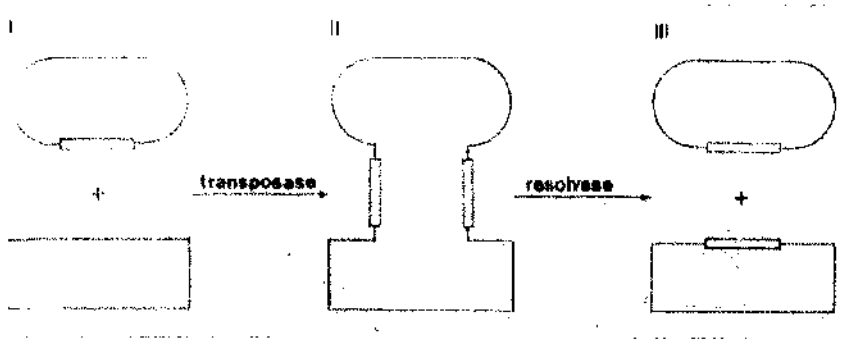


Figure 1.2 **Transposition pathway of Tn3 (from Arthur and Sherratt 1979).**

transposon sequences

I: donor replicon containing the transposon, plus the recipient replicon.

II: cointegrate replicon formed by transposition.

III: normal end products of transposition: donor and recipient replicon each containing one copy of the transposon.

about an 8bp core (Vetter et al 1983). There is in fact remarkable homology between FLP binding sites and Cre binding sites of P1; in fact FLP is reported to bind at loxP but not recombine two of these sites (B.Andrews pers.comm.). Thus there may have been strong conservation of this recombination system during evolution from prokaryote to yeast.

1.3 Site-specific recombinases are topoisomerases.

The topological state of DNA, and the enzymes that control this state, play a crucial role in determining the function of DNA in cells. In eukaryotes, DNA supercoiling is an essential step in forming the DNA-histone complexes of chromatin, and thus in folding and organizing the chromosome. In prokaryotic cells, DNA supercoiling results in torsional strain, and the DNA can be considered to be in an energetically activated state. This activation is an important element in the processes of DNA replication, transcription, and genetic recombination.

DNA supercoiling is controlled by a class of enzymes called topoisomerases. Some of these, originally named DNA relaxing enzymes, carry out the relaxation of DNA. Another group, the DNA gyrases, carry out the reverse reaction of converting relaxed closed-circular DNA to a superhelical form, in a reaction coupled to the hydrolysis of ATP. The level of supercoiling in bacteria is presumably set by the balance of DNA supercoiling and DNA relaxing activities.

Topoisomerases can also catalyze the interconversion of other topological isomers of DNA. Various enzymes of this class carry out the formation of knotted structures in single-stranded DNA or the

formation of knots and catenanes in duplex DNA. Of these reactions, the separation of catenated circles may be particularly significant biologically, because catenanes are produced in some processes of replication and genetic recombination, and the products must be separated to allow segregation of the DNA.

All topological interconversions of DNA require the transient breakage and rejoining of DNA strands. Since the integrity of the DNA is jeopardised during breakage and reunion, these alterations must be carefully and efficiently executed. Topoisomerases break and rejoin DNA in a concerted fashion and catalyse the interconversion of various toopological isomers of DNA by passing a DNA segment through an enzyme-bridged interruption of a second DNA segment (reviewed by Gellert 1981). Type 1 topoisomerases transiently break one DNA strand, and type 2 enzymes coordinately break both strands of the helix. Topoisomerases remain bound to the broken ends during strand passage; one of the bonds is covalent and stores the energy of the broken phosphodiester bond, allowing resealing of the break in the absence of an external energy source. This mechanism ensures that topological alterations are conservative and controlled.

Incubation of a supercoiled plasmid containing the core sequences of lambda att with Int, or a supercoiled plasmid containing directly repeated res sites with resolvase, results in relaxation of these plasmids (Kikuchi and Nash 1979; Craig and Nash 1983; Symington 1982; Krasnow and Cozzarelli 1983). The set of relaxed topoisomerases are related by a difference of linking number of one, showing that both Int and resolvase are type 1 topoisomerases. Accompanying this topoisomerase activity, it has been shown that both proteins form

covalent linkage to the DNA at their specific transient break sites. When Int-dependent breakage occurs, Int becomes linked to the strand that is 5' to the breakage site through a 3' phosphate bond; the other product DNA strand has a new 5' OH terminus. In this regard, Int is unlike all previously studied prokaryotic topoisomerases, which link protein to a 5' DNA phosphate, but is similar to several eukaryotic topoisomerases eg. rat-liver nicking closing enzyme (Champoux 1977) and the HeLa topoisomerase 1 (Edwards et al 1982). Topoisomerase activity of resolvase is associated with covalent attachment of resolvase through a 5' phosphate bond leaving free 3' hydroxyl groups (Reed and Grindley 1981).

The identification of topoisomerase activity for these two site-specific recombination proteins suggests they have at least two roles after synapsis has occured. Firstly they provide the catalytic sites for strand breakage and reunion, and secondly they provide the specificity for strand exchange. The breakage reactions are confined to specific sequences within the recombination sites. This sequence-specific action defines where strand exchange will occur and assures that corresponding positions in the parental DNAs are broken so that rejoining, and hence recombination, occurs conservatively at fixed internucleotide junctions without nucleotide loss. The mechanism by which strand exchange is accomplished by strand breakage and rejoining of DNA is unknown although several models have been presented; one is based on the hypothesis that DNA synapsis occurs via a four-stranded helix and that strand exchange involves cleavage of DNA at the edges of the structure followed by rotation of the parental DNA (Kikuchi and Nash 1979), whilst another predicts that protein-protein interactions

may direct the movement of broken strands (Sherratt *et al* 1981).

Both the observed topoisomerase 1 activity of resolvase and the production of breaks at the recombination site occur only when substrates are used which contain two directly repeated *res* sites. Thus it appears that the formation of the synapse stimulates a protein conformational change which, in turn, activates the strand breakage activity of resolvase. Whereas Mg^{2+} ions are required for *in vitro* resolution, in the absence of Mg^{2+} ions synapse, a resolvase-*res* complex, is formed which contains DNA molecules that have been nicked at *res*, but recombination is inhibited (Reed and Grindley 1981). Similarly, in the absence of Mg^{2+} ions topoisomerase 1 interacts with a superhelical DNA to form transient covalent complexes in which the enzyme is linked to the 5' phosphoryl group at the site of the nick, but there is no change in linking number (Liu and Wang 1979). Addition of Mg^{2+} ions leads to both a change in linking number and dissociation of the protein-DNA complex. Moreover, whilst restriction enzymes require Mg^{2+} ions for catalysis of cleavage, thay can bind to their recognition sites without Mg^{2+} ions (Halford and Johnson 1980). Kinetic analysis suggests that Mg^{2+} ions enhance the rate of dissociation of the *Eco*RI enzyme from DNA (Paoletti *et al* 1971). Though it may appear paradoxical that a co-factor required for catalysis should weaken the binding of an enzyme to its substrate, this is readily accomodated by one theory for enzyme catalysis. The destabilisation of the enzyme-DNA complex by Mg^{2+} ions reduces the activation barrier between enzyme-substrate and enzyme-product complexes and thus facilitates catalysis.

For convenience it is possible to dissect site-specific recombination into three processes. (1) the process by which the recombination protein locates and binds to the recombination site. (2) the process of juxtaposition of both sites involved in recombination prior to strand exchange, and (3) the events of strand cleavage, passage and ligation which produce the recombinant products. The remainder of this thesis addresses the first two of these steps.

CHAPTER 2

Materials and Methods

2.1 Organisms.

All bacterial strains listed are derivatives of *Escherichia coli* K-12.

Name	Relevant markers	Source
C600	*thr leu thi*.	F.Cannon
AB2463	*arg his leu pro thr lac gal ara rpsL*(=Str) *recA*-13.	A.J.Clark
△M15	[*lac pro*]$_{\triangle}$ *thi*Ø 80d*lacZ* M15 *ara strA recA*.	U.Ruther
N100	*galK pro recA*.	K.McKenney
N4830	lambda CI^{857} containing strain. It is deleted for all lambda genes from *cro* to the right of the *att* site and from *gal*(71.3%) to *int*(57.8%).	R.Reed
LS415	N4830 containing pLS213: *tnpR* overproducing strain.	L.Symington
K37	*gal* Str^{r}.	H.I.Miller
K5175	K37 *himA* due to a Tn10 insertion linked to a lesion.	H.I.Miller
K5428	K37 *himD*63.	H.I.Miller
K5482	K37 *himD*63 *himA*42.	H.I.Miller

2.2 Plasmids and transposons.

Plasmids other than those whose construction is described in this thesis are listed in Table 2.1 and the transposons in Table 2.2.

Table 2.1 **Plasmids**

Name	Description	Phenotype	Size(kb)	Source or Reference
pACYC184	vector derived from p15A	Cm^rTc^r	4.0	Chang and Cohen 1978
pAA33	pACYC184::Tn3	$Cm^rTc^rAp^r$	9.0	Arthur 1981
pAA231	pACYC184::Tn103	Cm^r	8.1	Arthur and Sherratt 1979
pLS208	pACYC184 + Tn3 TaqI scramble	Cm^rAp^r	7.2	L.Symington
pPAK100	pAA231 BamHI deletion	Cm^r	7.2	Kitts 1982
pJD100	pACYC184::Tn103 + pMB9::Tn103 fusion	Cm^rTc^{r*}	17.4	D.Sherratt
pJD101	" opposite orientation	Cm^rTc^{r*}	17.4	D.Sherratt
pBR322	vector derived from pMB8	Ap^rTc^r	4.36	Bolivar et al 1977
pPAK329	pBR322 + res-RI	Ap^rTc^r	4.64	Kitts 1982
pPAK330	" opposite orientation	Ap^rTc^r	4.64	Kitts 1982
pUC8	vector derived from pBR322	Ap^r	2.7	Vieira and Messing 1982
pLS213	tnpR overproducing plasmid derived from pBR322	Ap^r	5.2	Symington 1982
RR12	pBR322::Tn1000 + EcoRI linker inserted at HaeIII site in tnpR gene	Ap^rTc^r	10.06	Kitts et al 1982
pK01	derivative of pBR322 containing galK	Ap^r	3.9	McKenney et al 1980
pPAK322	pBR322 + Tn3 96bp TaqI fragment	Ap^r	4.46	Kitts 1982
pPAK333 to pPAK337	derivatives of pBR322 containing 96bp TaqI fragment + Sau3A res-wt fragment in different orientations	Ap^r	4.81	Kitts 1982
pDS4153	ColK::Tn1 HaeII53	Ap^r	8.1	Tacon et al 1981
RSF1365	pMB8::Tn365	Ap^r	5.6	Heffron et al 1977
R388::Tn3	R388(naturally occuring) + Tn3 insert	Tp^rAp^r	38	Arthur 1981
pMB9::Tn103	Insertion of Tn103 into pMB9	Tc^r	9.5	L.Symington

* Tc^r is derived from the tnpA promoter and hence repressible by resolvase.

Table 2.2 **Transposons**

Name	Derivation	Phenotype	Reference
Tn1	naturally occuring on RP4	A^+R^+*res*$^+Ap^r$	Hedges and Jacob 1974
Tn3	naturally occuring on R1	A^+R^+*res*$^+Ap^r$	Heffron *et al* 1975
Tn1000	naturally occuring on F	A^+R^+*res*$^+$	Guyer 1978
Tn103	spontaneous deletion of Tn1	A^+R^-*res*$^+Ap^s$	Heffron *et al* 1977
Tn365	*in vitro* deletion of Tn3	A^-R^+*res*$^+Ap^r$	Heffron *et al* 1977
on pDS4153	*Hae*II deletion of Tn1	A^-R^+*res*$^+Ap^r$	D.Sherratt pers.comm.
on RR12	*Eco*RI linker insertion on Tn1000	A^+R^-*res*$^+$	Kitts *et al* 1982

A = *tnp*A R = *tnp*R

2.3 Chemicals and enzymes.

Chemical	Source
General chemicals and organic compounds	BDH, Hopkins and Williams, Koch-Light Laboratories, May and Baker.
Media	Difco, Oxoid.
Biochemicals	Sigma.
Antibiotics	Sigma.
Agarose	Sigma, BRL.
Restriction enzymes and DNA modifying enzymes	BRL.
X-Gal(5-Bromo-4chloro-3indolyl-βD-galactoside)	BRL.
Sephadex G-50, G-75, G-150	Pharmacia Fine Chemicals.
CM-Sepharose	Pharmacia Fine Chemicals.
Pd/Pt wire	Agar acids.
Piperidine	Fluka A.G.
Dimethyl Sulphide	Aldrich Chemical Co. Ltd.
Hydrazine	Kodak.
Nitrocellulose filters	Schleicher and Schull.
Bioflour	New England Nuclear.
Radiochemicals	Amersham.
SDS	Serva.
Urea	Schwarz/Mann Inc.

2.4 Basic Media.

L-broth: 10g tryptone, 5g yeast extract, 5g NaCl, 1g glucose; made up to 1 litre with distilled water, adjusted to pH7.0 with NaOH.

Nutrient agar: 25g Oxoid No.2 nutrient broth, 12.5g agar; made up to 1 litre in distilled water.

L-agar: As L-broth but without glucose and solidified by adding 10g/l of oxoid No.1 agar prior to autoclaving.

Iso-sensitest agar: 23.4g isosensitest broth, 12.5g agar; made up to 1 litre in distilled water.

MacConkey galactose agar: 40g MacConkey agar made up in 1 litre distilled water, pH7.1 at 25°C, to which 10g galactose were added.

Davis-Mingioli Salts(X4): 28g K_2HPO_4, 8g KH_2PO_4, 4g $(NH_4)_2SO_4$, 1g Na_3 citrate, 0.4g $MgSO_4.7H_2O$; made up to 1 litre with distilled water. For minimal agar 100ml were added to 300ml molten 2%, plus the appropriate supplements as necessary.

Supplements: Where required growth supplements were added at the following concentrations: glucose 2mg/ml, amino acids at 40ug/ml, thymine at 20ug/ml, thyamine (B1) at 2ug/ml, casamino acids 5mg/ml, plus 40ug/ml tryptophan when required.

2.5 Sterilization.

All growth media were sterilzed by autoclaving at 120°C for 15min; supplements, gelatin solution and buffer solutions at 108°C for 10min, and $CaCl_2$ at 114°C for 10min.

2.6 Buffers

E: 0.04M Tris, 0.02M NaAc, 0.001M EDTA, pH8.2 with acetic acid; usually kept as 10X concentrate.

TBE(X10): 108g Tris, 55g Boric acid, 9.3g EDTA; made up to 1 litre with distilled water. The pH should be about 8.3.

SRB: 15.5g Tris, 72.05g glycine, 0.1% SDS; made up to 5 litres with distilled water.

SDS-PAGE lower buffer(X4): 1.5M Tris-HCl, 0.4% SDS, pH8.8.

SDS-PAGE upper buffer(X4): 0.5M Tris-HCl, 0.4% SDS, pH6.8.

TE: 10mM Tris-HCl, 1mM EDTA, pH7.5.

ESB: 10% Ficoll (w/v), 0.5% SDS (w/v), 0.06% bromophenol blue (w/v), 0.06% orange G (w/v), made up with buffer E.

SCFSB: 2.5% Ficoll (w/v), 1.25% SDS (w/v), 0,015% bromophenol blue (w/v), 0.015% orange G (w/v); made up with buffer E.

PFSB: 10% glycerol (v/v), 0.01% bromophenol blue (w/v), 5% mercaptoethanol (v/v), 3%SDS (w/v), 0.625M Tris, pH8.0.

Elution buffer: 500mM NH_4Ac, 1mM EDTA, 0.1% SDS (w/v), 10ug/ml tRNA.

Lytic mix: 2% Triton X-100, 0.05M Tris pH8.0, 0.06M EDTA pH8.0; made up with distilled water.

Low salt restriction buffer-LSRB(X10): 100mM Tris-HCl, 100mM $MgCl_2$, 10mM DTT, pH7.4. Used for restrictions with HaeII, HaeIII, HpaII.

Medium salt restriction buffer-MSRB(X10): 100mM Tris-HCl, 100mM $MgCl_2$, 10mM DTT, 50mM NaCl, pH7.4. Used for restrictions with AccI, AluI, BamHI, ClaI, HincII, HindIII, PstI, PvuII, Sau3A.

High salt restriction buffer-HSRB(X10): 500mM Tris-HCl, 100mM $MgCl_2$, 1M NaCl, pH7.4. Used for restriction with EcoRI, SalI.

TaqI restriction buffer(X10): 100mM Tris-HCl, 60mM $MgCl_2$, 60mM 2-mercaptoethanol, 1M NaCl.

Core buffer-supplied by BRL(X10): 500mM Tris-HCl, 100mM $MgCl_2$, 500mM NaCl. USed for restrictions with HaeII, HaeIII, HindIII, PstI.

NTB(X10): 500mM Tris-Hcl, 50mM $MgCl_2$, 100mM mercaptoethanol, pH7.9.

Ligation buffer(X10): 0.67M Tris-HCl pH7.6, 0.1M $MgCl_2$, 10.0mM EDTA pH9.0, 0.1M DTT. Made up to a final volume of 192ul by addition of 132ul 1M Tris-HCl pH7.6, 20ul 1M $MgCl_2$, 20ul 100mM EDTA pH9.0, and 20ul 1M DTT.

Buffered ATP(X10): 0.67M Tris-HCl pH7.6, 4mM ATP.

Denhardt's solution (DS): 0.2mg/ml BSA, 0.2mg/ml Ficoll-400, 0.2mg/ml PVP.

Phage buffer: 7g Na_2HPO_4, 3g KH_2PO_4, 5g NaCl, 0.25g $MgSO_4.7H_2O$, 15mg $CaCl_2.2H_2O$, 1ml 1% gelatin solution made up to 1 litre in dH_2O.

20XSSC: 3.0M NaCl, 1mM EDTA; 1mM mercaptoethanol.

TEM: 20mM Tris, pH7.5, 1mM EDTA, 1mM mercaptoethanol.

2.7 Growth conditions.

Liquid cultures were grown in L-broth at 37°C with aeration unless otherwise stated. 'Stationary phase cultures' were small liquid cultures, usually 2ml, grown without shaking at 37°C overnight.

Plates contained 25ml agar with the appropriate supplements and were incubated at 37°C for 16-18hr unless otherwise stated. All dilutions prior to plating were made in phage buffer.

Bacterial strains were stored on Nutrient agar slopes at room

temperature or in 50% L-broth, 40% glycerol at -20°C. Inocula from the slopes or glycerol cultures were streaked out onto selective plates of the appropriate types and after growth at 37°C the plates were stored at 4°C for up to 1month before fresh streaks were made. These plates provided a source of inocula for the 2.5ml stationary phase cultures.

2.8 Antibiotic selections.

Throughout this work the standard antibiotic concentrations given below were used in both liquid and plate selections, unless otherwise stated.

Antibiotic	Encoded by	Selective conc.	Stock soln.
Ampicillin(Ap)	transposon	50ug/ml	5mg/ml in water
Chloramphenicol(Cm)	plasmid	25ug/ml	2.5mg/ml in EtOH
Streptomycin(Str)	chromosome	50ug/ml	5mg/ml in water
Tetracycline(Tc)	plasmid	10ug/ml	1mg/ml in 0.1M NaOH-made fresh
Trimethoprim(Tp)	plasmid	50ug/ml	5mg/ml in 50% EtOH

The appropriate volume of antibiotic stock solution was mixed into the medium, agar solutions having been pre-cooled to 50°C. The medium employed was dictated by the required selections as follows:

Tp - trimethoprim acts by inhibiting thymine synthesis and this selection is not effective in media containing exogenous thymine (Stacey and Simpson 1965). Therefore Tp selections were on minimal agar lacking thymine, or Isosensitest agar.

Ap, Cm, Km, Tc, Str - these selections were on Nutrient agar, L-agar, or in L-broth except when used in conjunction with Tp.

Indicators: X-Gal (5-Bromo-4chloro-3indolyl-B-Dgalactoside) was used at 20ug/ml in L-agar. A stock solution in methanol was made up freshly.

This indicator was used in conjunction with cloning DNA fragments into the poly-linker of pUC8. Ligated DNA was used to transform ΔM15 and transformants were plated onto Ap, X-Gal plates.

2.9 Transformation with plasmid DNA.

A fresh overnight culture of the recipient strain in L-broth was diluted 1:40 in 20ml of L-broth. The culture was grown at 37°C, with shaking, for about 90min, or until there were about 2 x 10^8 cells/ml. The time required to reach this stage is strain dependent, recA$^+$ strains grow much faster than recA$^-$ derivatives. Cells were pelleted at 12,000g for 5min, resuspended in 10ml cold 50mM $CaCl_2$ and left in an ice/water bath for 10min. The cells were pelleted again, resuspended in 1ml of cold 50mM $CaCl_2$ and left in an ice/water for between 1 and 5hr. The competence for DNA uptake by recA$^+$ cells increases for up to 24hr on ice at this stage (Dagart and Ehrlich 1979). 0.2ml of the cell suspension were mixed with up to 50ul of DNA solution (DNA was usually diluted in dH_2O - 0.1ug DNA is ample for each transformation). The DNA/cell mixture was mixed well then incubated in an ice/water bath for 20min; at 37°C for 7min then ice/water again for 30min. 1.0 ml of fresh L-broth was added to each tube and incubated at 37°C, shaking, for 30-120min to allow expression of antibiotic resistance. All antibiotic selections required expression times of at least 90min, except ampicillin, which needs less than 30min for expression. 100-500ul aliquots were plated onto

selective agar and incubated overnight at 37°C.

2.10 Purification of plasmid DNA

(i) Cleared lysates.

Cleared lysates were prepared from 100ml cultures. Cells were harvested by centrifugation at 12,000g, 4°C for 10min. The cell pellet was resuspended in 3.3ml cold 25% sucrose/0.25M Tris-HCl, pH8.0. 0.67ml of fresh lysozyme solution, made up at 10mg/ml in 0.25M Tris-HCl, pH8.0, were added and the cell suspension swirled frequently, on ice, for 10min. 1.3ml of 0.25M EDTA, pH8.0 were added and swirled again, while on ice, for 5min. 5.3ml of lytic mix were added gently to the cell suspension, the lysate was swirled gently to ensure thorough mixing. The cells were judged to have lysed when the mixture was clear and viscous; this usually took about 5min. The lysate was centrifuged at 43,000g, 4°C for 20min to pellet the chromosomal DNA and membranous material. The supernatant, containing mainly plasmid DNA, was carefully decanted - this is referred to as the cleared lysate. The cleared lysate was purified further by phenol/isopropanol extraction or by CsCl/EtBr equilibrium centrifugation.

(ii) Phenol/isopropanol extraction.

The cleared lysate was mixed with an equal volume of freshly distilled phenol, saturated with 1M Tris-HCl, pH8.0. The phases were resolved by centrifugation at 5,000g for 5min. The upper aqueous layer was phenol extracted twice more. Residual phenol was removed by ether extraction. An equal volume of diethyl ether was added, mixed well,

the upper phase (ether) was discarded and the process repeated until the lower (aqueous) phase was clear. Residual ether was removed by blowing compressed air over the surface of the solution. 0.1volume of 3M NaAc was added, then 0.54volume of isopropanol. This was mixed well and left at room temperature for 45min to precipitate nucleic acids. The solution was centrifuged at 12,000g, 15°C for 25min; the pellet obtained was washed with 70% ethanol/H_2O, then dried. The dried pellet was resuspended in 500µl TE and stored in air-tight plastic tubes at 4°C.

(iii) CsCl/EtBr equilibrium centrifugation.

For DNA of greater purity, cleared lysates were run through CsCl/EtBr gradients. For each tube 5.0g CsCl, 4.83ml cleared lysate, 0.33ml EtBr(3mg/ml), 0.1ml 0.2M Na_2HPO_4 were mixed and placed in Beckman quickseal tubes. The density of the mixture was checked (it should be about 1.58/cm^3), the tubes were filled with liquid paraffin and heat sealed. They were cetrifuged in a Ti50 or Ti70 rotor in a Beckman L-8 ultracentrifuge at 48Kopm for 16hr at 15°C. Two bands were formed within the gradient, the lower, denser one contains covalently closed DNA. This band was removed by inserting a syringe needle through the side of the tube and slowly drawing off the band laterally; this is enabled by previously piercing the top of the tube with a hot needle. EtBr was removed from the DNA by repeated extraction with a butan-1-ol. The solution was diluted with three volumes of water, then precipitated by addition of two volumes of ethanol at -20°C for one hour. The DNA was collected by centrifugation at 32,000g, 4°C for 25min. The pellet was washed with 70% ethanol/H_2O,

dried, then resuspended in 500ul TE.

(iv) Birnboim-Doly DNA purification.

The two previous methods each had their problems. Phenol extracted DNA could not be guaranteed to restrict; 'caesium DNA' takes two days preparation and is an expensive procedure for the isolation of, for example, putative recombinant plasmids prior to their restriction. The Birnboim-Doly method was designed for rapid purification of small quantities of plasmid DNA. I scaled up the procedure and introduced an additional purification step in order to purify equivalent amounts of plasmid DNA. Plasmid DNA purified was 14kb or less. The purified plasmid DNA is suitable for restriction, subsequent ligation, end labelling and sequencing.

The basic principle of this method is alkaline denaturation of linear DNA molecules. Cells are lysed with NaOH/SDS; chromosomal DNA is denatured, but not covalently closed DNA. On neutralisation the mass of chromosomal DNA renatures and aggregates, proteins complex with SDS; when centrifuged the supernatant contains only covalently closed DNA and low molecular weight RNA (Birnboim and Doly 1979).

25ml overnight cultures were harvested by centrifugation at 12,000g, 4°C for 5min. The supernatant was discarded, and any residual supernatant withdrawn using a pasteur pipette connected to a vacuum pump. The cell pellet was resuspended in 1ml 50mM glucose, 25mM Tris-HCl, pH8.0, 10mM EDTA, 4mg/ml lysozyme added on the day, and then incubated at 22°C for 5min. 2ml of a fresh solution of 0.2M NaOH/1% SDS were added, gently mixed, and left on ice for 5min. To neutralize, 1.5ml precooled 5M KAc, pH4.8 (3M KAc pH'd with acetic acid) were

added, mixed gently and left on ice for 5min. The crude lysate was centrifuged at 12,000g, 4°C for 5min, then the supernatant removed (approx. 4.5ml). An equal volume of a 1:1 phenol/chloroform mix was added and mixed. The phases were resolved by centrifugation at 5,000g for 5min. The upper aqueous layer was removed and to it was added 2 volumes of 95% ethanol/H_2O. This was mixed and left at room temperature for 2min. The solution was centrifuged at 32,000g, 4°C for 25min. The pellet was resuspended in 8ml 70% ethanol/H_2O and this solution centrifuged at 32,000g,4°C for 15min. The pellet, containing RNA and closed circular DNA was dried and then resuspended in 200ul TE and 5ul RNase solution (RNase 1mg/ml in TE). This solution was left at room temperature for 25min. In this time Sephadex G-50 'mini-columns' were prepared. These consist of one blue Eppendorf tip plugged with siliconized glass wool into which was added 1ml of a suspension of Sephadex G-50 equilibrated in TE. This tip is placed in another blue tip which in turn is placed in a 1.5ml Eppendorf tube. The Sephadex is packed in the column by centrifugation at 100g for 2min, and washed by addition of 1ml TE and centrifugation at 100g for 2min. The prepared column is then placed in a 500ul Eppendorf tube.

The solution of DNA in TE was then placed on the top of the column, and centrifuged at 100g for 2min. Recovery of DNA is approximately 100%, and it is now ready for restriction.

(v) Single colony cleared lysates (referred to as SCCL)

This technique provides a quick method for analyzing the total DNA content of a particular clone. The single colonies (isolated from transformation) were 'patched' onto selective plates. Clumps of cells

were resuspended in 100ul of SCFSB. These suspensions were left at 42°C for 15min. The lysates were spun in an Eppendorf or Sarstedt microfuge for 15min; 50ul of supernatant was loaded directly onto agarose gels for visualization of the DNA content.

2.11 DNA Electrophoresis through gels.

Vertical gel kits held two glass plates measuring 16cm x 16cm, and separated by perspex spacers of 3mm (1mm for acrylamide gels). The gel liquid was poured between the two plates; 3mm teflon combs (1mm thick perspex for acrylamide gels) with 10 or 15 teeth were pushed into the top displacing the gel material to produce sample pockets, when removed from the set gel. Electrophoresis was from top to bottom with the cathode at the top.

'Mini-gels' were poured in horizontal gel kits measuring 10cm x 10cm; the gel liquid was poured to a depth of 3mm in an area 10cm x 8cm allowing the electrodes to be uncovered at each end. An eight tooth 1mm comb was placed in the still molten gel and removed when set. The cathode was at the pocket end of the gel.

(i) Agarose gels

Sigma agarose (type II, medium EEO) and buffer E were used for electrophoresis unless otherwise stated. Agarose gels for SCCL, for checking DNA concentrations and for mini-gels were made up at 0.9%(w/v); for restriction gels 1.2% agarose was usually used. The agarose/buffer E mixture was heated to 100°C until completely molten, then cooled to 52°C for gel pouring. DNA samples were mixed with 5ul FSB and loaded with a 50ul Finn pipette. The gel was run with buffer

at anode and cathode at between 25 and 100V until the blue dye reached the bottom, unless otherwise stated. The relative separation between bands increases when gels are run at low voltage for longer times, but bands tend to be more diffuse. Restriction gels were usually run at room temperature at 80V for 5hr; other gels were run at room temperature at either 70V for 6hr, or overnight at 25V. The gels were stained in gel running buffer containing 0.5ug/ml EtBr for 30min. Gels were viewed and photographed using a 35mm SLR camera (Ilford HP5 film) with a red filter.

(ii) Low melting point agarose gels.

These gels were used when isolating higher forms of plasmid DNA, or particular fragments from a restriction digest which were to be used in cloning.

The agarose (1%) was dissolved in TE buffer at 65°C and was cooled to 37°C. The molten agarose was poured into a mini-gel kit and cooled. After the gel had set, the comb was removed and the gel was immersed in TE buffer. The samples (10-20ul) were loaded and the gel was run at 5V/cm at 4°C for 2-4hr. The gel was stained and photographed as above.

A band could be excised and melted at 65°C. This was diluted in 1 x TBE and used for transformation or ligation.

(iii) Polyacrylamide gels.

Optimal resolution of small DNA fragments was achieved by analyzing restriction digests of plasmid DNA on polyacrylamide gels. 10%(w/v) acrylamide, made up in TBE, was used to analyze small

fragments of 10-300bp; 5%(w/v) acrylamide was used to assay for larger fragments, 80-500bp; and 3.5%(w/v) used for fragment sizes between 100-1000bp. Acrylamide solutions (kept as 20% acrylamide, 1% bis-acrylamide stock) were polymerized by the addition of 0.36ml 10%(v/v) TEMED and 0.72ml 10%(w/v) APS in 60ml total volume. Gels were run with TBE buffer at anode and cathode at 30mA at room temperature until the yellow dye reached the bottom of the gel. The gel was removed from the running apparatus, separated from the glass plates, then stained and photographed as previously described.

(iv) Sequencing polyacrylamide gels.

High resolution sequencing gels resolve DNA strands of n and n+1 nucleotides over a range of one to several hundred base pairs. The common sequencing gel techniques are versions of methods described by Peacock and Dingman (1967) and Maniatis et al (1975). For the experiments described in this thesis only 8% sequencing gels were employed. 8% gels contained 7.6%(w/v) deionised acrylamide, 0.4%(w/v) bis-acrylamide, 50%(w/v) urea (8.3M), 100mM Tris-borate, pH8.3, 2mM EDTA, 0.07%(w/v) APS, and 0.7%(w/v) TEMED as catalyst. This solution was injected between two clean, siliconized glass plates, measuring 200 x 400mm, separated by 0.3mm plasticard spacers. The gel material was displaced from the top of the gel by a 0.3mm plasticard comb to form the sample pockets. The gel was installed in the vertical running apparatus, buffer reservoirs filled with TBE. The comb was gently eased from the top of the gel and sample pockets were immediately flushed out with buffer to remove unpolymerized acrylamide, which slides into the pockets from the top of the gel. Gels were prerun at

40W(1600V) for 30min, then disconnected prior to loading the samples. Samples, between 2-3ul, were loaded with a glass capillary tube evenly across the pocket. The power supply was reconnected and the gel run at 40W until the bromophenol dye marker had nearly reached the bottom. The gel was disconnected and duplicate samples were loaded. When the bromophenol blue of the second samples had nearly reached the bottom the gel was disconnected again. Running two duplicated samples this way gives better resolution of bands over a wide range. One glass plate (the notched one) was lifted from the gel, a sheet of clingfilm was spread over the gel, smoothed down and secured with tape on the underside of the remaining gel-plate. This was placed in an aluminium light-tight box, covered by a pre-flashed sheet of X-ray film (Kodak X-omat S1) juxtaposed with an intensifying screen (Du Pont, Cronex lightning plus). A second glass plate was placed on top, and the aluminium lid then firmly clamped. Exposure at -70^{o}C varied in length according to the radioactivity of samples between 48 and 120hr.

(v) Interpretation of gel data.

The distance migrated is related to the size of a molecule of a given conformation; small molecules run the furthest. The distance migrated by a given DNA molecule is dependent on its conformation; supercoils generally run fastest and open circles slowest, with linear molecules between the two. Above a certain size, all linears run at the same rate, therefore large plasmid linears run faster than corresponding supercoils. Large linear fragments of chromosomal DNA appear as a single band.

The size of linear restriction fragments was estimated from

graphs of the $\log_{10}$ molecular size plotted against the distance travelled, according to the formula:

$$\log M = C \times 1/D \qquad \text{(Helling } \underline{\text{et}} \text{ } \underline{\text{al}} \text{ 1974)}$$

M = molecular size in bp D = distance migrated C = a constant

Molecular weight standards were obtained by restriction of λ cI857Sam7 DNA, or pBR322 DNA (Philippsen et al 1978, Haggerty and Schleif 1976, Sutcliffe 1978).

2.12 Elution of DNA fragments from gels.

(i) Electroelution/ DNA precipitation.

Fragments of DNA were separated by electrophoresis through acrylamide or agarose gels. The stained gel was placed on the long-wave transilluminator, the relevant bands were excised with a scalpel and placed in a dialysis sac with 250ul of TBE or E buffer. The dialysis sac was subjected to electrophoresis (in TBE or E buffer) at 50V for 2hr. The current was reversed for 2min to remove any DNA from the sides of the dialysis sac. The contents of the sac were removed with a Finn pipette leaving the gel material, and placed in a 1.5ml Eppendorf tube. The DNA was precipitated by the addition of 10ul yeast tRNA (1mg/ml), 0.1vol 3M NaAc plus 2vol 95% EtOH/H_2O, mixed and placed in dry ice/methanol for 20min, then centrifuged in an Eppendorf microfuge for 10min. The pellet was resuspended in 250ul 0.3M NaAc and reprecipitated by addition of 750ul 95% EtOH/H_2O, mixed and placed in dry ice/methanol for 20min, and centrifuged for 10min. The pellet was washed by addition and resuspension in 25ul dH_2O. The DNA was

repricipitated by addition of 1ml 95% EtOH/H_2O, mixed and placed in dry ice/methanol for 20min, and centrifuged for 10min. The pellet was dried in a rotary freeze-drier, and resuspended in TE, ready for use to restrict, ligate etc..

(ii) Crush-soak-precipitate method.

This method was used exclusively for extracting labelled DNA fragments from acrylamide gel slices. The DNA-containing band was mashed with a siliconized glass rod in a 1.5ml Eppendorf tube, to which 400ul of elution buffer was added. The mixture was thoroughly vortexed and placed in an orbital shaker at 42°C overnight. Acrylamide was filtered out by passing the solution through a blue Eppendorf tip plugged with siliconized glass wool. The DNA from the filtered solution was precipitated by addition of 2vol 95% EtOH/H_2O, mixed and placed in dry ice/methanol for 20min. The DNA was reprecipitated, washed and dried as described above.

2.13 DNA manipulations *in vitro*.

(i) Restriction of plasmid DNA.

Restriction of plasmid DNA was performed in 1.5ml Eppendorf tubes. The final reaction volume contained:

0.5 - 1.0ug plasmid DNA

0.1vol 10x appropriate restriction buffer

1 - 10 units enzyme

gelatin/water (at 0.1mg/ml) to make up volume

Complete digestion was usually achieved in 3hr at 37°C (65°C for Taq1). Digests were analyzed by electrophoresis through agarose or acrylamide gels.

(ii) Ligation of restriction fragments.

Endonucleases used to digest plasmid DNA were destroyed by heating to 65°C for 5min. Ligations were prepared by addition of 5-15ul of each restricted DNA, plus 5ul ligation buffer, plus 5ul buffered ATP, plus 1ul ligase (prepared by E. Nimmo, this laboratory) for 'sticky ends', or 2-4ul for blunt ends, and finally made up to 50ul with dH_2O. Final concentrations were:

67mM Tris-HCl, pH7.6

10mM $MgCl_2$

10mM DTT

1mM EDTA

0.4mM ATP

0.1 unit ligase/ug DNA

Ligation mixtures were incubated at 14°C for 16-18hr. These were then diluted in dH_2O for use in transformation.

(iii) DNA polymerase end-labelling.

For DNA sequencing and footprinting reactions, and for preparation of radioactive probes for Southern hybridisation, DNA fragments with labelled termini are required. Restriction enzymes which produce a staggered cleavage with a 3' recessed hydroxyl group provide a substrate for DNA polymerase I. Only one labelled [α-^{32}P]

triphosphate is required to be polymerized adjacent to the recessed 3'-hydroxyl group for end-labelling. The labelled triphosphate is chosen in reference to the sequence cleaved by the particular restriction enzyme. Restricted plasmid DNA was precipitated, reprecipitated, washed and dried as described in section 2.11. The following were added:

1-5ul (depending on activity) [α-^{32}P]XTP (10mCi/ml when new)

5ul NTE(x10)

0.5ul *E.coli* DNA polymerase I large fragment (1 unit)

40-45ul dH_2O

The reaction was incubated at 16°C for one hour.

(iv) Chemical sequencing reactions.

Chemical sequencing involves three consecutive steps: modification of the base, removal of the modified base from its sugar, and strand scission at that sugar (Maxam and Gilbert 1977; 1980). A DNA restriction fragment with one end-labelled terminus is treated such that each strand should be cleaved only once, generating fragments with one common end and the other variable in length. An array of fragments is produced, representative of cleavage at each position held by a particular base. By performing four separate reactions each using a chemical which cleaves only one of the four bases, separating the bands on a sequence gel, then analyzing the ladder of bands produced, it is possible to read the sequence from an autoradiogram of the gel.

(a) Limited DNA cleavages at guanines.

To a 1.5ml Eppendorf tube on ice, 200ul DMS buffer (50mM Na cacodylate pH8.0, 10mM $MgCl_2$, 1mM EDTA), 5ul [^{32}P]labelled DNA and 1ul DMS were added and mixed. The mix was incubated at 20°C for between 5 and 10min, and then 50ul of DMS stop (1.5M NaAc pH7.0, 1.0M mercaptoethanol, 100ug/ml tRNA) was added, mixed on ice, and then 750ul 95%EtOH/H_2O added and mixed. The mixture was chilled on dry ice/methanol for 10min and centrifuged in a microfuge for 5min. The pellet was resuspended in 250ul 0.3M NaAc, and 750ul 95%EtOH/H_2O added and mixed. The mixture was chilled in dry ice/methanol for 10min and centrifuged in a microfuge for 5min. The pellet was rinsed with 1ml 95% EtOH/H_2O, centrifuged and dried in a centrifugal vacuum pump. This was now ready for the strand scission reaction.

(b) Limited DNA cleavage at purines.

10ul [^{32}P]labelled DNA was dried down in a 1.5ml Eppendorf tube. The DNA was resuspended in 15ul fresh DPU mix (66% formic acid (v/v), 2% diphenylamine (w/v), 1mM EDTA) and incubated at 25°C for 5-10min. The reaction was stopped by addition of 45ul dH_2O. The mixture was ether extracted three times by addition of 500ul diethyl ether, mixing and extraction of the upper phase. The solution was then lyophilised; the dried DNA was now ready for the strand scission reaction.

(c) Limited DNA cleavage at pyrimidines.

On ice 10ul dH_2O was added to 10ul [^{32}P]DNA, and then 30ul hydrazine was added and mixed. The mixture was incubated at 20°C for 10+/-5min; and the reaction stopped by addition of 200ul HZ stop (0.3M

NaAC, 0.1mM EDTA, 25ug/ml tRNA). 750ul 95%EtOH/H_2O was added, mixed and then the mixture chilled in dry ice/methanol for 10min, and centrifuged in a microfuge for 5min. The pellet was reprecipitated, rinsed and dried as described above (a) in preparation for the strand scission reaction.

(d) Limited DNA cleavage at cytosines.

This procedure is similar to (c) except initially 5ul [^{32}]]DNA was combined with 15ul 5M NaCl, followed by addition of 30ul hydrazine.

(e) Strand scission reactions.

Each DNA sample was redissolved in 100ul 1.0M piperidine and incubated at 90^{o}C for 30min. The reaction was stopped and DNA precipitated by addition of 100ul dH_2O, 20ul 3M NaAc, 10ul tRNA(1mg/ml), 750ul 95%EtOH/H_2O; this was mixed and placed in dry ice/methanol for 15min, and then centrifuged in a microfuge for 10min. The pellet was lyophilised and then resuspended in 10ul dH_2O, frozen and lyophilised. This was repeated and finally the DNA was resuspended in 6ul formamide-containing marker dyes (80%(v/v) deionised formamide, 0.1%(w/v) bromophenol blue(w/v), 0.1%(w/v) xylene cyanol, 50mM Tris-borate pH8.3, 1mM EDTA). The sample was heated to 90^{o}C for 1min to separate DNA strands, then quickly chilled. 2-3ul of each sample were loaded per gel pocket.

(v) DNA footprinting

This technique combines the chemical sequencing reactions and the technique of DNase protected fragment isolation (Galas and Schmitz 1978). Fragments of an end-labelled, double stranded DNA segment, partially degraded by DNase in the presence and absence of the binding protein, are visualized by electrophoresis and autoradiography alongside the base-specific reaction products of the Maxam-Gilbert sequencing method. It is then possible to see the protective 'footprint' of the binding protein on the DNA sequence, in this case resolvase bound to the res region.

The enzyme DNaseI works under most conditions, whereas resolvase has specified reaction conditions. For this reason all DNaseI digestions, plus or minus resolvase, were performed using conditions optimal for in vitro resolution. The concentration of DNaseI required to give a ladder of bands representing restriction fragments betwen 200-300bp was 80ng (Symington 1982).

To the standard in vitro resolution conditions (20mM Tris-HCl, pH8.0, 10mM $MgCl_2$,1mM DTT, 50mM NaCl) were added 0.4ug of lambda DNA as competitor for DNaseI, and about 30ng of labelled DNA fragment, and 0-0.6ug resolvase in a total volume of 50ul. Reactions were incubated for 30min at 30°C before addition of 80ng DNaseI. The DNaseI digests were performed for 1min,then stopped by the addition of 12.5ul of 3M NH_4Ac/0.25M EDTA. DNA was precipitated by the addition of 10ul tRNA(1mg/ml) and 2vol 95%EtOH/H_2O, placed in dry ice/methanol for 15min, then centrifuged in a microfuge. The DNA was resuspended, precipitated and washed as described above (2.11). The pellet was lyophilised and resuspended in 6ul formamide containing dyes. This

solution was heated to 90°C to separate the DNA strands, then quick-chilled on ice prior to gel loading. 2-3ul samples were loaded per pocket.

(vi) Southern transfer and hybridisation.

After staining and photography, the agarose gel was exposed to UV light for 10min, and the DNA was then denatured by gently shaking the gel in 1.5M NaCl, 0.5M NaOH for 40min. The gel was washed briefly with four changes of distilled water, and was then neutralized by shaking in 1.5M NaCl, 0.5M Tris-HCl pH7.5.

The fragments were then transferred from the gel to a nitrocellulose filter, according to the method of Southern (1975). The filter was placed on top of the gel and 10xSSC was drawn through the gel into a stack of absorbent paper. After blotting for 16hr, the filter was washed in 2xSSC for 30min, and was then baked at 80°C for 2hr in a vacuum oven.

The filter was sealed in a plastic bag with 20ml 6xSSC, 1xDS, 0.5% SDS and 100ug/ml denatured salmon sperm DNA, and was pre-hybridised overnight at 65°C. A hybridisation mixture was prepared from : labelled DNA (approx 10^4cpm), 125ul 20%SDS, 500ug salmon sperm DNA, plus a volume of dH_2O to give a final volume of 3.40ml. This mixture was boiled for 6min, cooled to 70°C and 1.5ml 20xSSC, 0.25ml 20xDS were added. The pre-hybridisation mix was removed from the bag containing the filter, and the hybridisation mix was added. The filter and probe were shaken gently overnight at 68°C in order to allow hybridisation to occur. The filter was then removed and was washed as follows:

15min at 65°C in 10ml 1xDS, 6xSSC + 0.5%SDS

1hr at 65°C in 100ml 1xDS, 6xSSC + 0.5%SDS

3x 40min at 65°C in 100ml 2xSSC, 0.5%SDS.

The filter was dried at 37°C and placed on a glass plate which was then wrapped in cling-film. A sheet of X-omat film was pre-flashed and placed on top of the filter, and a tungsten intensifying screen was laid over the film, sandwiched with a second glass plate. After exposure at -70°C the autoradiograph was developed.

2.14 Purification of resolvase.

Tn3 resolvase was purified using the method developed by Reed, 1981b. The strain LS415 used as the source of resolvase contains the plasmid pLS213, which was constructed so that the tnpR gene was uncoupled from its promoter, and hooked up to the bacteriophage lambda leftward promoter, P_L (Symington 1982). A UAA stop codon before the ribosome binding site of the tnpR gene allows the translation of a fully functional tnpR gene product, and avoids production of a truncated N gene - tnpR gene fusion product. P_L is a high expression promoter and when unrepressed leads to uncontrolled transcription interfering with plasmid maintenance functions, probably replication (Bernard and Helinski 1979). P_L is usually controlled by the lambda repressor protein CI and a temperature sensitive CI repressor has been isolated, CI^{857}. This allows repression at 30°C, but at 42°C uncontrolled transcription from P_L (Franklin 1971; Bernard and

Helinski 1979). LS415 contains lambda CI^{857}, so that a shift from 30°C to 42°C allows overproduction of resolvase.

(i) Checking resolvase overproduction/ SDS-polyacrylamide gel electrophoresis (for proteins).

To check that resolvase was being produced in large amounts it was necessary to assay the total cellular proteins present. LS415 was grown in 4ml L-broth at 30°C to $5x10^8$ cells/ml with constant shaking. 2ml of this culture were transferred to 43°C for 1hr to induce transcription from P_L. The samples of induced and non-induced cultures were centrifuged at 12,000g for 2min to pellet the cells. The cells were resuspended in 100ul of PFSB and boiled for 2min. The total protein content was analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE).

'Stacking' gels were used, containing 12.5%(w/v) acrylamide in the lower separating gel and 4.5%(w/v) acrylamide in the upper stacking gel. Gels were prepared from a stock solution containing 30%(w/v) acrylamide and 0.8%(w/v) N,N'-bis-methylene acrylamide. 0.03%(w/v) APS and 0.025%(v/v) TEMED were added prior to pouring the gels. The gel apparatus consisted of two glass plates of the same dimensions as used for DNA gels, separated by 1.5mm perspex spacers. Buffer and stock acrylamide were mixed, then freshly made APS and TEMED were added and the gel poured. 0.1% SDS solution was sprayed over the top of the lower gel to ensure an even surface. When the lower gel had set the SDS solution was poured off. The constituents of the top gel were mixed and poured over the lower gel. A 12 space perspex comb was inserted into the top of the gel and left for a

further 20min to set. Top and bottom tanks of the vertical gel apparatus were filled with SRB. The comb was removed and sample pockets were flushed out with buffer to remove any unpolymerized acrylamide. Electrophoresis was carried out at 30-40mA until the blue marker dye reached the bottom of the separating gel. Gels were stained for 1hr, with slight agitation, in 0.2% coomassie blue in fix (50% methanol. 40% water and 10% acetic acid). The stain was poured off and then the gel was washed several times in fix until destained. The gel was placed in 10% acetic acid for the final wash to enlarge the gel to its original size, then photographed. For photography, Panatomic-X film was used in a 35mm camera with an orange or yellow lens filter.

The molecular weight marker proteins used to calibrate gels were purchased from Bio-Rad, and contained the following proteins at 0.1ug/ul:

Phosphorylase B	93,000 Mdal
BSA	68,000 Mdal
Ovalbumin	43,000 Mdal
Carbonic anhydrase	30,000 Mdal
Soybean lectin inhibitor	21,000 Mdal
Lysozyme	14,300 Mdal

(ii) Resolvase purification (Reed 1981b).

Fresh overnight cultures of LS415 in L-broth were diluted 1:20 into two litres of L-broth. Cultures were grown to about $5x10^8$ cells/ml at 30°C. Flasks were transferred to 43°C to induce transcription of P_L. Culture flasks containing at least 500ml of liquid would normally take 20-30min to heat to 43°C; an hour after

this point the cells were harvested.

Cells were harvested by spinning at 9,820g for 5min. Pellets were resuspended in 20ml of 100mM NaCl, 20mM Tris pH7.5. The cells were again pelleted, then resuspended in 10ml of 10% sucrose/TEM buffer. The cell suspension was sonicated in four 30sec bursts using a Dawe Soniprobe type 7532B. The crude lysate was centrifuged at 12,000g for 10min to remove unlysed cells and cell debris. The supernatant was carefully decanted and to it added polymin P (10% w/v in TEM), dropwise, to a final concentration of 0.5%. This was left on ice for 10min to allow polymin P to aggregate with DNA/protein complexes; these were then precipitated by centrifugation at 12,000g for 10min. The supernatant was discarded, and the pellet was resuspendedin 0.3M NaCl/TEM and left on ice for 10min. At this concentration of NaCl most proteins dissociate from DNA, but most of the resolvase remains bound. The DNA/protein/poymin P complex was collected by centrifugation as before then resuspended in 10ml of 1M NaCl/TEM and left on ice for 10min to remove resolvase from the DNA/polymin P complex. The DNA/polymin P pellet was removed by centrifugation.

The resulting supernatant was loaded directly onto a 100x2.5cm Sephadex G-75 (medium grade) gel filtration column, equilibrated with 1M NaCl/TEM. Fractions were collected using an LKB 7000 Ultrorac fraction collector and scanned using an LKB 8300 Uvicord II, set at 280nm during collection. Fractions were analysed further by SDS-PAGE to determine those containing resolvase. Fractions containing resolvase were pooled and dialyzed three times against 7M urea/TEM. The pooled, dialyzed protein was then applied to a 10ml CM-Sepharose cation exchange column, equilibrated with 7M urea/TEM. Afer loading,

one bed volume of running buffer (7M urea/TEM) was passed through the column. A salt gradient of 0-200mM NaCl was applied to the column and collected as before. Fractions were assayed by optical density and gel electrophoresis. Two main peaks were indicated by OD readings and these were assayed by SDS-PAGE to determine which contained resolvase.

Resolvasecontainingfractions were pooled and dialyzed against 7M urea/1M NaCl/TEM, then twice against 1M NaCl/TEM, then further against 1M NaCl/TEM/ 50% glycerol (v/v). Resolvase was then stored at -20°C; samples stored at -20°C in 50% glycerol retained activity for more than 12 months.

2.15 Galactokinase assay.

The galactose operon (*gal*) of *E. coli* consists of three genes: epimerase(*galE*), transferase(*galT*) and kinase(*galK*). These genes are normally expressed from a polycistronic mRNA in the order E,T,K (Michaelis and Starlinger 1967). The promoter distal gene, *galK*, was used to constuct the plasmid pKO-1 (McKenney *et al* 1981) which was then manipulated to fuse transcriptional regulatory signals from Tn3 at positions 5' to *galK*. Galactokinase catalyses the reaction:

galactose + ATP -----> galactose-1-phosphate.

This reaction can be monitored with a simple, sensitive and linear assay (from about 5 to 200,000 molecules/cell), derived from Wilson *et al* 1966, Adhya and Miller 1979.

Cells containing plasmid weregrown in L-broth to an OD_{650} of between 0.2 and 0.35. 1ml of culture was transferred to an Eppendorf tube and 40ul of lysis buffer (100mM EDTA, 100mM DTT, 50mM Tris pH8.0)

and 5 drops of toluene were added. The samples were vortexed vigorously for 1min, then placed at 37°C , shaking to facilitate the evaporation of the toluene for between 30min and 2hr. At this point the cell lysate was ready to be assayed. Separate reaction tubes were made up by addition of 20ul mix1 (5mM DTT, 16mM NaF), 50ul mix2 (8mM $MgCl_2$, 200mM Tris-HCl, pH7.9, 3.2mM ATP), 10ul diluted ^{14}C-galactose.

Amersham D-(1-^{14}C) galactose at 49.4 mCi/mmol, No. CFA 435, was used. The labelled galactose was diluted to a final specific activity of 4.5 x 10^6dpm/umole, and this was filtered twice through 2.3cm DE81 Whatman filters.

Reactions were started by addition of 20ul cell lysate to reaction tubes (plus a blank for which 20ul dH_2O was added) and these were then incubated for 15-30min at 32°C. 25ul samples from the incubated reactions were spotted onto 2.3cm DE81 Whatman filters. The filters were washed twice in 1l dH_2O for 10min, gently shaking and then dried in a 65°C oven. Two additional filters were spotted with 25ul of reaction mix chosen at random, and these filters were not washed.

Dried filters were placed in disposable scintillation vials and 5ml of Biofluor added. Samples were counted using an SL30 liquid scintillation spectrometer.

Galactokinase units were calculated from:

$$\frac{(\text{cpm} \times 2 - \text{blank}) \times 2500}{\text{average unwashed filters} \times \text{incubation time} \times OD_{650}} = \text{galactokinase units}$$

CHAPTER 3

Defining the extent of *res*: the site required *in cis* for Tn3 site-specific recombination.

INTRODUCTION

The pioneering work of Gill et al (1978) and Heffron et al (1978) suggested that Tn3 might encode a site-specific recombination system. This was substantiated by Arthur and Sherratt (1979) who showed that the 19,000 molecular weight transposon-specified protein as detected in mini-cells (Dougan et al 1979) could act in trans to resolve a plasmid containing two directly repeated copies of a Tn3 deleted around its BamHI site. They concluded that the components of the site-specific recombination system were the genestraddlingthe BamHI site, whose diffusable product acted at a site within a 540bp region to the left of this gene. This gene product was also thought to be involved in the repression of a second gene involved in transposition (Heffron et al 1978), but the precise nature of this bifunctionality was unclear. Indeed other components of the site-specific recombination system could not be ruled out. For instance, both host factors and other transposon encoded functions could be implicated.

The analysis of the nucleotide sequence, in conjunction with the phenotypes of several deletion mutants used in early studies and those of newly created insertion mutations, showed the likely gene organization of Tn3 (Heffron et al 1979). This gene organization (Fig. 3.1) correlated with the known polypeptides produced by Tn3 (Gill et al 1979, Dougan et al 1979). A 23bp sequence within the amino terminus of the transposase gene which shares strong sequence homology with part of the 38bp inverted repeat was suggested by Heffron to be a component of res, the cis-acting site of recombination.

Heffron et al (1979) also showed that a streptomycin gene placed

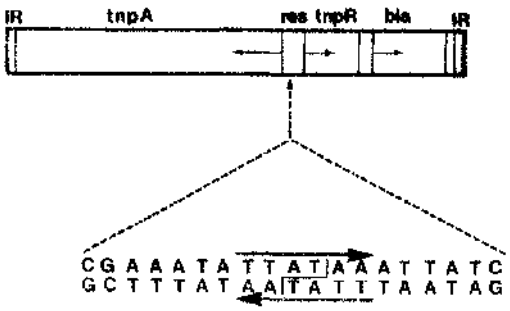


Figure 3.1 **Gene organization and cross-over sequence of Tn3.**

Tn3 encodes three gene products: transposase(tnpA), resolvase(tnpR) and β-lactamase(bla). Arrows indicate the direction of transcription for these genes.

The three sites required *in cis* for transposition are indicated: the terminal 38bp inverted repeats (IR) and the REsolution Site, res, containing the divergent promoters for tnpA and tnpR. The cross-over sequence is shown: resolvase recognizes the 6bp palindrome, arrowed, and cleavage products containing a 2bp staggered break have been isolated (Reed and Grindley 1981).

under the control of the tnpA promoter was repressible by resolvase. Moreover Chou et al (1979a,b) demonstrated with fusions between tnpR and lacZ, where expression of β-galactosidase is under tnpR control, that resolvase, supplied in trans, represses expression from the tnpR promoter. So it appeared that resolvase could bind to either one or multiple operator sites within the intercistronic region between tnpA and tnpR to act as repressor of tnpA expression and to autoregulate its own expression.

The work of Kitts et al (1982) showed that the tnpR gene was the only transposon gene encoding a trans-acting protein involved in Tn3 site-specific recombination, and that this protein was interchangeable with that of the related transposon Tn1000. Moreover Reed (1981b) defined the cross-over sequence of recombination by analyzing the product of recombination of a plasmid containing one 890bp internal segment of Tn1000 and one 360bp internal segment of Tn3. These segments contained the intercistronic region where the Tn1000 and Tn3 sequences are 77% homologous. So when the hybrid res regions in the resolution products were sequenced, the point of the cross-over could be determined. The results localized the cross-over site to an A+T rich sequence 19bp long and exactly homologous between Tn3 and Tn1000 (Fig. 3.1), at nucleotide 3100 within Tn3, and overlapping the putative promoter of tnpA (Heffron et al 1979).

So a functional res was found to be located within a 360bp Sau3A fragment extending from nucleotide 2909 to 3269 of Tn3 containing the entire intercistronic region (Fig. 3.2; pullout at front of thesis). Kostriken et al (1981) independently derived the same cross-over sequence in Tn3, but implicated a second sequence to the left of this

cross-over sequence as being essential for recombination. A deletion from nucleotide 2790 to 3052 was judged res⁻, and this deleted the aforementioned 23bp sequence which shares homology with the terminal 38bp inverted repeat of Tn3.

In this laboratory we were interested in further defining the extent of res. The general dearth of restriction sites within the intercistronic region of Tn3 made this task difficult. However a good candidate for analysis was the 282bp EcoRI* fragment extending from nucleotide 2933 to 3215 (Fig. 3.2). The analysis of the resolution activity of this fragment, referred to as res-RI, by Kitts (1982) showed that this fragment contained all the sequences required for res.

RESULTS

3.1 Derivation and analysis of the Tn1 HincII-BamHI segment for resolution activity.

Tn3 is one of several closely related transposons. The transposon Tn1 specifies three polypeptides which are of identical function with those specified by Tn3; the transposase, resolvase and β lactamase (Kitts et al 1982b). Moreover the interchangeability of transposase and resolvase shows that the sites required in cis: the terminal inverted repeats and res, are natural variants. Using this variation, it was possible to delineate res further.

A single base-pair transversion at Tn3 coordinate 3074 changes a hexanucleotide sequence TTTAAC to GTTAAC found in Tn1, which is recognized and cleaved by HincII, The plasmid pPAK100 (Kitts 1982) was

constructed by deleting the 1kb BamHI fragment from pAA231 (Arthur and Sherratt 1979) and contains all those Tn1 sequences to the left of the BamHI site in tnpR. pJD118 was constructed by deleting the two smallest HincII fragments from pPAK100 (Fig. 3.3), and contains Tn1 sequences between the HincII and BamHI sites: Tn3 nucleotide 3076 to 3576 (Fig. 3.2). This Tn1 segment I shall refer to as res-HB.

The most rigorous test for functionality of res-HB was to determine if two copies of res-HB in direct repeat constituted a resolution substrate; the results referred to in subsequent chapters indicate that a substrate containing a 'trimmed' res region and a 'wild-type' region in direct repeat does not provide an ultimate test for functionality of the trimmed region. A simple way to derive a plasmid containing directly repeated res-HB's was to isolate dimeric forms of pJD118.

pJD118 was introduced into the Rec^+ strain C600. In Rec^+ strains an equilibrium exists betwen plasmid monomers and higher forms (Bedbrook and Ausubel 1976). After subculturing the strain C600 containing pJD118 for 75 generations, plasmid DNA was prepared. An aliquot of DNA was run into a low melting point agarose gel, and the dimer band, running just above open circles, was excised. The agarose/DNA was melted at 65°C, diluted with TE and used to transform the recA$^-$ strain AB2463 to Cm^r. Twenty-four independent transformants were analyzed by the SCCL method, revealing two clones containing pure dimers of pJD118.

The test strain was constructed by introducing the compatible plasmid pDS4153 containing a tnpA$^-$ tnpR$^+$ Tn1. Ap^r transformants were patched out and subsequently analyzed by SCCL (Fig 3.4). Complete

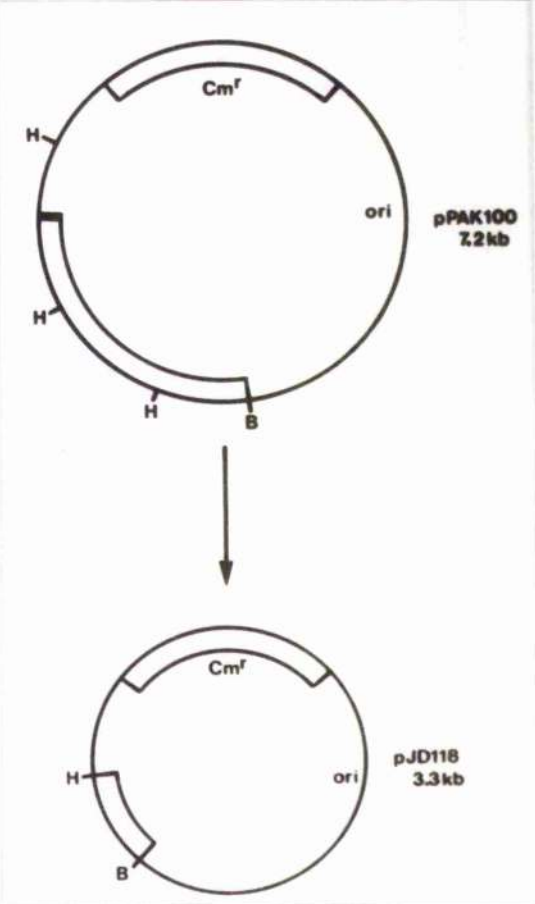


Figure 3.3 **Construction of pJD118, containing res-HB.** (see text for details)

H = HincII site

B = BamHI site

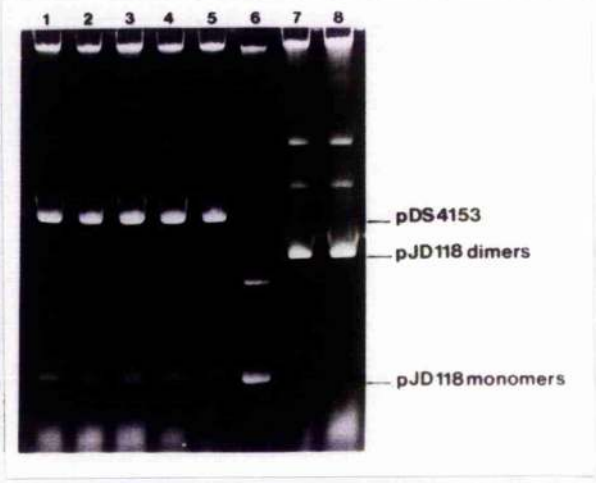


Figure 3.4 **In vivo analysis of res-HB function.**

0.8% agarose gel

Lanes 1 to 5: conversion of pJD118 dimers to monomers by complementation with the tnpR$^+$ plasmid pDS4153.

Lane 6: pJD118 monomers/dimers.

Lanes 7 and 8: pJD118 dimers.

resolution of dimers to monomers was observed in the 35 generations required to grow up patches of cells. From this data the Tn1 HincII to BamHI segment was judged to contain all the sequences required for res.

DISCUSSION

The results of these experiments help to delineate a fully functional res: by inference from both the EcoRI* fragment data and the data above it is located on a 139bp segment derived from the intercistronic region of Tn3. This data contradicts those of Kostriken et al (1981) who implicated sequences between nucleotide 2790 to 3052, a portion of the amino-terminus of the tnpA gene, as being important for res; this region is totally outwith the 139bp segment. Moreover deletions in the res region of Tn1000 (up to the equivalent Tn3 coordinate 3037) confirm that no sequences in the tnpA gene are required for res (Grindley et al 1982).

The substrate requirements for Tn3 site-specific recombination in vivo are therefore a DNA molecule containing two directly repeated copies of this 139bp res^+ segment. Other requirements and components of the system became apparent with the development of an in vitro site-specific recombination system. This has been achieved primarily due to the facility of expression vectors to allow the overproduction and purification of resolvase, and was first achieved by Reed (1981a) for Tn1000. He found that the components of the system are a supercoiled DNA substrate containing two res sites in direct repeat, purified resolvase, Mg^{2+} and a suitable medium-salt buffer. As was found for the cell-free system for lambda site-specific recombination

(Nash 1975; Gottesman and Gottesman 1975; Nash et al 1981), Tn1000 site-specific recombination requires no DNA synthesis, nor an energy source such as ATP. The products of the reaction are catenated supercoiled circular molecules, demonstrating that the reaction is reciprocal, i.e. all strands broken during the recombination event are rejoined to produce continuous DNA helices. Subsequently it has been demonstrated that Tn3 site-specific recombination is achieved with exactly the same reaction requirements (Symington 1982, Krasnow and Cozzarelli 1983).

In the absence of Mg^{2+} it was found that recombination intermediates could be 'trapped' (Reed and Grindley 1981). The requirements for generation of these intermediates were a plasmid containing directly repeated res regions, incubated with resolvase, which yielded cleaved resolvase-res complexes. Cleavage was induced by protease treatment. Analysis of these intermediates showed that resolvase made a single or double-stranded cleavage in its DNA substrate at the recombination site. The protein was covalently attached to the 5' ends of the cleaved DNA strands and generated free 3' hydroxyl ends. The DNA was cut within the region previously identified as containing the cross-over point (Fig. 3.1) at the palindromic sequence 5'-TTATAA-3' to generate 3' extensions of two bases.

The specific binding of resolvase to res was analyzed by the nuclease footprinting technique (Galas and Schmitz 1978). For this technique the susceptibilty of an end-labelled duplex DNA fragment to a non-specific nuclease is examined in the presence and absence of a DNA binding protein. Regions of the DNA that are protected from

nuclease attack are inferred to contain specific binding sites for the protein under investigation. Using this method, resolvase binding sites were analyzed for Tn3 (Symington 1982) and for Tn1000 (Grindley et al 1982).

It was demonstrated that resolvase binds to three DNA sites contained within the tnpA-tnpR intercistronic region of the transposons Tn1000 and Tn3 (the protected regions are illustrated by red bars in Fig. 3.2). The protected sites do not overlap into either tnpA or tnpR coding sequences, although site III extends right up to the start of tnpR. Each protected site spans 30 to 40 bp and the total protected region is about 124bp with only a few unprotected bases separating sites I and II. The protected sequences span the region from Tn3 coordinate 3085 to 3209; this is in close agreement with the data derived from the results of this chapter as to the extent of res as defined in vivo. Hereafter a res region containing all the Tn3 sequences between coordinate 3076 and 3215 will be termed res-wt.

Both the in vivo and in vitro results suggest that Tn3 site-specific recombination is a relatively uncomplex reaction, primarily dependent on the interaction of a small region of DNA, res, and the tnpR gene product. This interaction can also satisfactorily explain the second function of resolvase as repressor of tnpA and tnpR gene expression. The promoters of tnpA and tnpR are both located in protected regions; by binding to res resolvase effectively obscures both promoters from access for RNA polymerase.

As previously mentioned, there are few restriction sites in Tn3 res (Fig. 3.2). However the presence of internal TaqI sites presented us with an opportunity to analyse further recombinant res regions.

CHAPTER 4

Analysis of the structure and function of truncated *res* regions containing the 96bp *Taq*I fragment.

INTRODUCTION

The construction of the Tn3 resolvase over-producing plasmid, pLS213 (Symington 1982), depended on cloning partial TaqI digests of a plasmid containing Tn3. In this way a partial restriction fragment containing the 66bp and 150bp TaqI fragments was cloned. These fragments contain the tnpR structural gene isolated from its promoter which is located on the 96bp TaqI fragment derived from the intercistronic res region. However, in the course of isolating a suitable recombinant plasmid, one chimeric plasmid, pLS208, was isolated which contained the 96bp TaqI fragment but not the two fragments which are adjacent to it in Tn3 (Kitts 1982).

The plasmid pLS208 was investigated further (Kitts 1982). An *in vivo* resolution assay, similar to that described in Chapter 3, of pLS208 dimers was conducted. After 35 generations no conversion of dimers to monomers was observed, and after 100 generations only a small number of monomers were obtained which might have arisen by low level resolution. Therefore the res region contained in pLS208, termed res-96, was judged to not contain sufficient sequences to allow efficient site-specific recombination.

Another substrate was constructed containing res-96 and res-wt in direct repeat (Kitts 1982). A resolution asssay *in vivo* revealed fairly proficient recombination, providing the first evidence of unequal partners in Tn3 site-specific recombination, reminiscent of the phage lambda integration system. In this system an extensive 250bp phage att site recombines with the much smaller 23bp bacterial att site (Hsu *et al* 1980; Mizuuchi and Mizuuchi 1980; Mizuuchi *et al*

1981). I shall refer to the unequal partners as a primary site (res-wt) and secondary site (truncated site).

By the nature of the construction of pLS208, the sequences adjacent to the 96bp TaqI fragment were unknown, and restriction mapping could only provide ambiguous results as to the plasmid's structure (P.Kitts pers. comm.). To test whether the adjacent sequences were important for secondary site activity of res-96, the 96bp TaqI fragment was cloned into a different environment: the ClaI site of pBR322.

Recombinant plasmids were constructed so that the 96bp TaqI fragment cloned in either orientation was in direct repeat with the 360bp Sau3A fragment containing res-wt (Kitts 1982). These plasmids were tested as substrates for resolution in vivo, but no products of recombination were detected. Hence it was concluded that in either orientation, hereafter termed res-322A and res-322B, the 96bp TaqI fragment inserted into the ClaI site of pBR3222 could not act as asecondary res site. Two explanations were provided for this observation: (i) the sequences immediately adjacent to the 96bp TaqI fragment in pLS208 are important in the observed resolution activity of res-96, and that the sequences around the ClaI site of pBR322 cannot substitute for these sequences, (ii) the strong transcription through the ClaI site of pBR322 (Stueber and Bujard 1981) may interfere with binding of resolvase to the 96bp TaqI fragment.

4.1 Structural analysis of res-96.

As stated above pLS208 was derived from cloning a 'scramble' of TaqI fragments from a plasmid containing Tn3. Restriction mapping of pLS208 using enzymes recognizing tetranucleotide sequences suggested at least four possible arrangements of TaqI fragments adjacent to the 96bp fragment (P.Kitts pers. comm.; Fig. 4.1). An additional complication was to assign the position or positions of at least one copy of a 6bp TaqI fragment from Tn3, which was known to be present in pLS208 as a result of the cloning. There was also the possibility that during the cloning a fragment or fragments may have been partially deleted by exonuclease activity and then become incorporated as part of res-96. I decided to proceed with the structural analysis by identifying which HpaII fragment from pLS208 contained the 96bp TaqI fragment, whilst assuming at least one of the four suggested structures was correct and taking into account the possible presence of one 6bp fragment. On this basis the correct HpaII fragment would be between 187 and 208bp long (Fig. 4.1).

A HpaII digest of pLS208 revealed that the bands in question ran in a cluster of fragments between 190 and 210bp, three of which were derived from the parent plasmid pACYC184. I planned to test each unique fragment of pLS208 in the 187 to 208bp range by hybridising them with a restriction fragment containing the 96bp TaqI fragment, using the Southern transfer method (Southern 1975). To do this it was necessary to further separate the bands between 190 and 210bp so that no contaminating fragments from pACYC184 were present. The plasmid

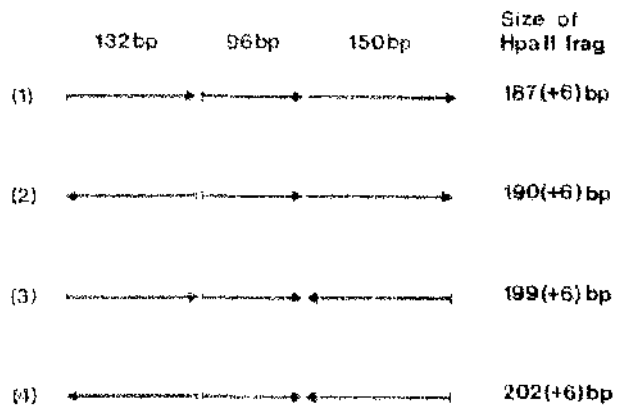


Figure 4.1 **The four possible orientations of TaqI fragments relative to the 96bp fragment derived from res-wt.**

The size of corresponding HpaII fragments containing res-96 are shown, possibly including a copy of the 6bp TaqI fragment from Tn3.

pPAK309 (Kitts 1982) is a derivative of pLS208 containing a wild-type res in direct repeat with res-96. A 1.69kb PstI fragment was isolated from this plasmid, containing most of the scrambled TaqI fragments and only 6bp of pACYC184 sequence. This PstI fragment was used as the source of pLS208 specific HpaII fragments.

The 1.69kb PstI fragment was excised from a 1% agarose gel and the DNA electroeluted and precipitated. This DNA was then digested with HpaII, and then radioactively labelled by repairing the 5' extensions this enzyme generates with DNA polymerase I large fragment, and α-^{32}P-dCTP. The labelled fragments were separated on an acrylamide gel, and three bands were excised: one approximately 190bp, another approximately 210bp and a third 242bp HpaII fragment derived from the Ap^r gene of pLS208 (and hence Tn3) to serve as a control for the hybridisation reaction.

EcoRI-SalI digests of pPAK332 (pBR322 plus the 96bp TaqI fragment cloned at the ClaI site; Kitts 1982) and pBR322 were run on a 1% agarose gel and then transferred to a nitrocellulose filter. The filters were hybridised to the three HpaII fragment probes and then autoradiographed. As a positive control, the 242bp HpaII fagment hybridised to the large EcoRI-SalI fragments of both pPAK322 and pBR322 which both contain the β-lactamase gene derived from Tn3, whilst not hybridising to the smaller EcoRI-SalI fragments of both plasmids as a negative control (Fig. 4.2). Only the 190bp HpaII fragment hybridised to the smaller EcoRI-SalI fragment of pPAK322 which contains the 96bp TaqI fragment.

By identifying the relevant HpaII fragment, this provided the basis to go on and determine the sequences adjacent to the 96bp TaqI

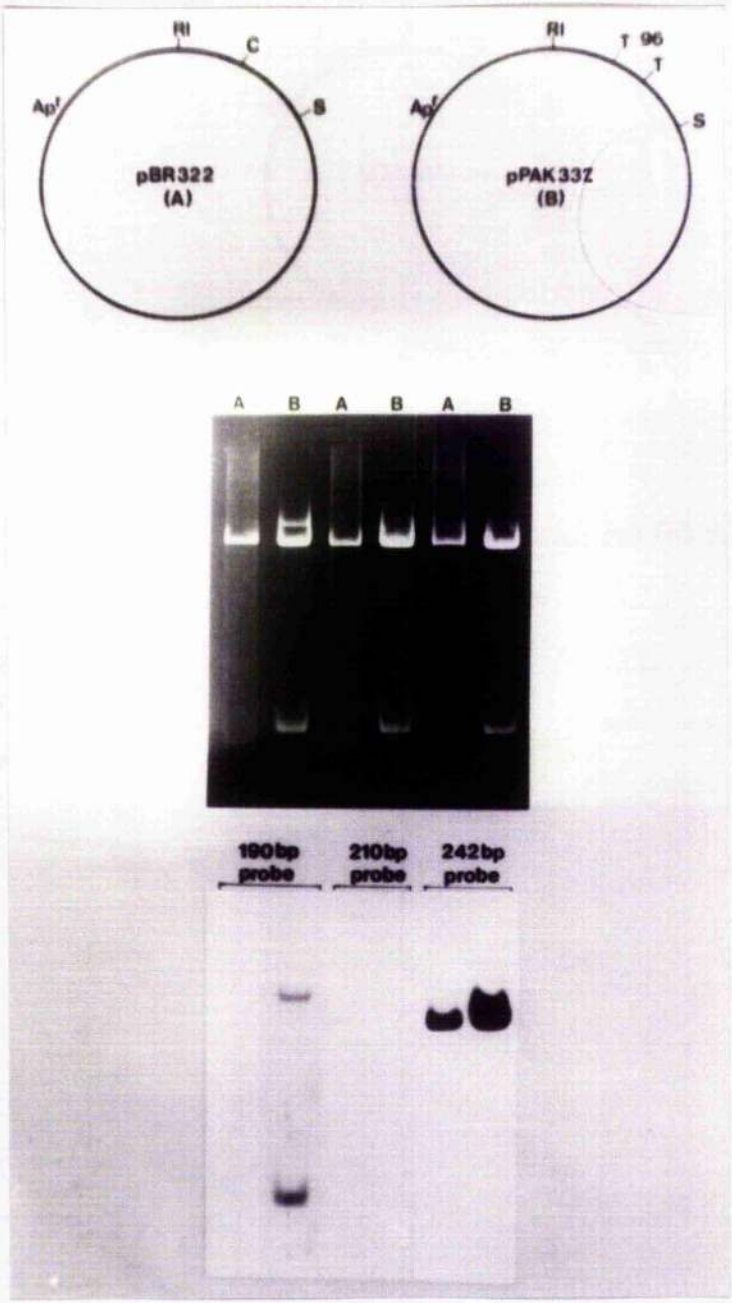


Figure 4.2 **Southern hybridisation analysis to identify the (TaqI 96bp)$^+$ HpaII fragment.**

Top is a 1% agarose gel: A lanes = pBR322 EcoRI/SalI digests

B lanes = pPAK332 EcoRI/SalI digests.

Bottom is an autoradiograph of the different HpaII probes hybridised to the transferred DNA of the top gel.

RI = EcoRI S = SalI

C = ClaI T = TaqI

fragment using the Maxam-Gilbert sequencing technique. As there are no restriction sites within the 96bp fragment it was expected that a digestion of the labelled HpaII fragment would yield at least two fragments each with one labelled terminus; the larger of these, after separation by electrophoresis and elution from the gel material could be sequenced. However the number of steps involved to derive this fragment meant a consequent reduction of DNA fragment concentration at each step, and often incomplete enzyme reactions, resulting in insufficient labelled fragment for Maxam-Gilbert sequencing.

A simple alternative to the process of successive fragment isolation from gels was to digest pLS208 with both Sau3A and HpaII, end-label the total digest with α-^{32}P-dCTP, and then separate the fragments on an acrylamide gel. To do this it was necessary to determine the size of the large fragment resulting from digestion of the 190bp HpaII fragment with Sau3A. The end-labelled HpaII fragment was isolated as described above, cut with Sau3A, and run on a 400mm long acrylamide gel together with a labelled HpaII digest of pBR322 as a size marker. This showed the relevant HpaII-Sau3A fragment was approximately 140bp long. A total labelled HpaII-Sau3A digest of pLS208 revealed that this 140bp fragment ran almost inseparably with another fragment which would contaminate sequence analysis, so an alternative strategy was adopted.

The 1.69kb PstI fragment was cloned into the PstI site of pUC8. The recombinant plasmid, pJD130, was digested with HpaII and Sau3A and run on a gel with HpaII-Sau3A digests of both parent plasmids, pLS208 and pUC8, and a HpaII digest of pBR322 as a size marker. This revealed the 140bp HpaII-Sau3A fragment running uniquely, and this was the

source for initial sequencing.

Maxam-Gilbert sequencing of this fragment showed that the 96bp TaqI fragment was bounded on one side by the 132bp fragment in the orientation predicted in configurations 2 and 4, Fig. 4.1. The other boundary fragment was eqivocal though, as it was represented by the dozen or so bases at the top of the sequencing gel where resolution was poor. The prediction based purely on the relative position of the Sau3A site to the TaqI sites was that configuration 2, Fig. 4.1 would be correct. If this was so it was likely that sequencing a 160bp HpaII-AluI fragment would provide confirmation. A HpaII-AluI digest of pJD130 showed that this 160bp fragment ran inseparably with another fragment (Fig. 4.5). But assuming configuration 2 was correct the 160bp would not be cut with HaeIII, and a HpaII-AluI-HaeIII digest separated the fragments sufficiently well for isolation and subsequent sequencing of the correct fragment (Fig. 4.3). Configuration 2, without the 6bp Taq fragment or any deletion, constituted the structure of res-96.

4.2 Comparison of sequences of res-wt, res-96 and res-322A/B.

A conclusion from the data presented in Chapter 3 is that a functional res site consists of three resolvase binding sites. Moreover each binding site is comprised of similar sequences, suggesting sequence specificity for binding. Hence the functionality of res-96, res-322A and res-322B may be determined by those sequences which have been substituted compared to res-wt in positions where resolvase is known to bind. As the TaqI sites are located in site I and site III, the sequences respectively upstream and downstream from

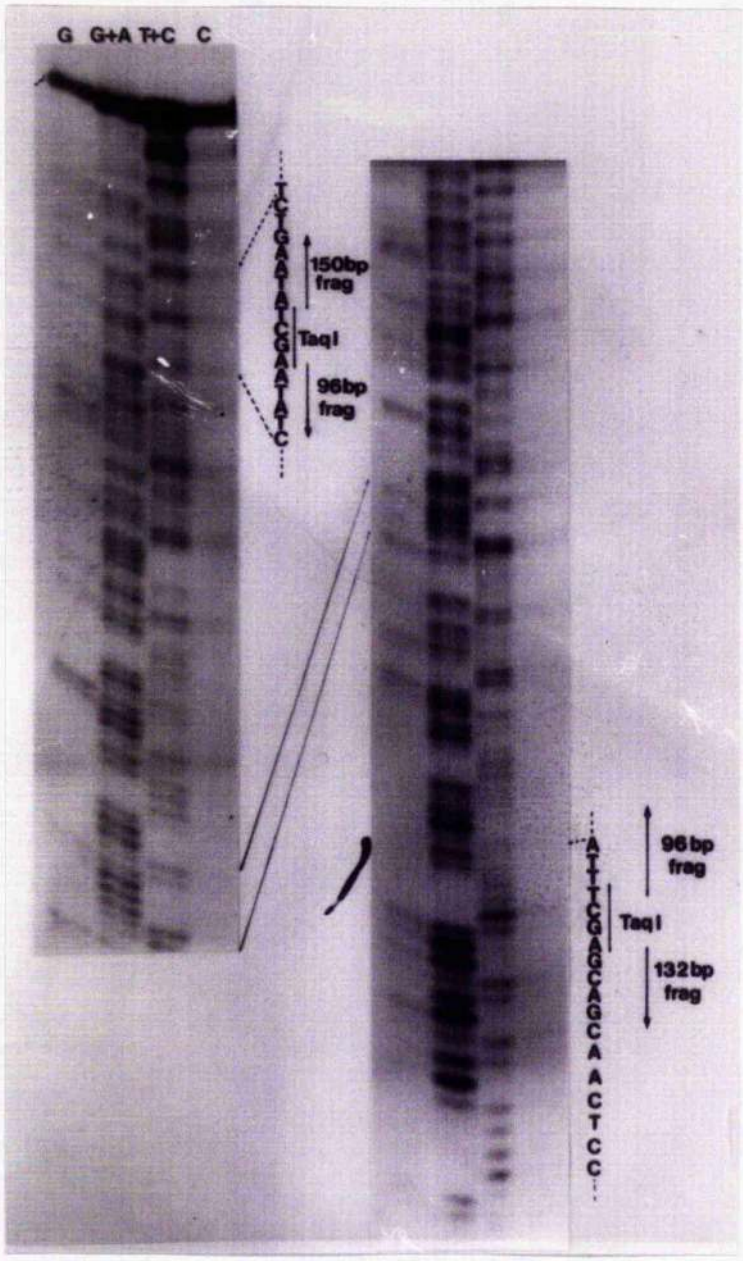


Figure 4.3 **Autoradiograph of an 8% sequencing gel of the 160bp HpaII-AluI fragment derived from pJD130.**

Samples were run on the gel for 3hr(right) and 1.5hr(left): the overlap is indicated. The TaqI junctions are illustrated.

Site IL

consensus	a (a/t) g T G T C Y g (a/t) T A (a/t) (a/t) T (c/t) A (a/t) A
res-wt	A A C C G T T C G A A A T A T T A T A
res-96	T G C T G C T C G A A A T A T T A T A
res-322A	A G C T T A T C G A A A T A T T A T A
res-322B	T T A T C A T C G A A A T A T T A T A

Site IIIR

consensus	a (a/t) g T G T C Y g (a/t) T A (a/t) (a/t) T (c/t) A (t/a) A
res-wt	A A A T G T A C C T T A A A T C G A A
res-96	C T C T G A T C T G A A T A T C G A A
res-322A	T T G A C A G C T T A T C A T C G A A
res-322B	C G C A T T A A A G C T T A T C G A A

Figure 4.4 **Sequence comparisons of site IL and site IIIR.**

The consensus sequence is derived from comparison of all six half-sites of both Tn3 and Tn1000 (Grindley et al 1982).

res-322A and res-322B sequences outside of the 96bp TaqI fragment are derived from the pBR322 sequence (Sutcliffe 1978).

these sites were compared (Fig. 4.4). Comparison of the left-hand half site I (site IL) of res-96 and res-wt shows one transversion at the 5' end,plus three transitions, including one C to T to consensus. Possibly significant also was that the triplet which is represented as a highly conserved TGT in the consensus, and as CGT in res-wt, is represented as TGC in res-96. res-322A site IL shows least homology with res-wt: 5 transversions and one transition are present in the 7bp substitution. res-322B site IL has 4 out of 7bp substitutions: two transitions, including a C to T to consensus, and two transversions.

Comparison of the right-hand half site III (site IIIR) revealed that res-96 contains nine out of thirteen substitutions: two transitions and seven transversions; res-322A contains eleven substitutions: one transition and ten transversions; and res-322B contains nine substitutions: one transition and eight transversions. This data suggested that the basis for functionality of res-96 as a secondary site probably did not reside in its site III equivalent, but might have involved sequences in the left-hand site I. In this case the sequence substitutions in res-96 might allow resolvase binding, whereas those of res-322B, and more notably of res-322A, which shares with res-96 reasonably good homology to res-wt in this position, would presumably prevent resolvase binding.

4.3 *In vitro* resolution assays to test res-322A and res-322B functionality.

The idea that res-322A/B functionality, even as asecondary site, was absent due to the transcriptional activity around the ClaI site of pBR322 could be tested by assaying for resolution *in vitro*. In this

case there will be no transriptional activity whatsoever.

In cloning the TaqI 96bp fragment into the ClaI site it was impossible to determine the orientation of insertion due to the absence of internal restriction sites in the TaqI fragment. The colonies containing the recombinant plasmids fell into two groups based on different levels of resistance to tetracycline. The two levels of Tc^r observed were thought to reflect insertion in either orientation of the TaqI 96bp fragment (Kitts 1982) which contains the divergent promoters for tnpA and tnpR (A.Lamond pers. comm.). Constructing plasmids derived from these with res-wt in both orientations ensured that two of the four recombinant plasmids contained respectively res-322A and res-wt, and res-322B and res-wt in direct repeat (Kitts 1982). The four recombinants: pPAK333, pPAK334, pPAK335 and pPAK336 were tested for resolution in vitro. After incubation with resolvase one sample from each assay was cut with HindIII which cuts once within each molecule - this procedure ensures that the products of resolution can be visualized on a gel, as the normal in vitro products are catenanes which run adjacent to the supercoils (Reed 1981a; Symington 1982). Control tracks of linearised and supercoiled starting plasmid were run on gels with the assayed samples (Fig. 4.5).

No resolution was observed in vitro for plasmids pPAK333-6, suggesting that the absence of resolution observed in vivo is due to the specific sequences around the ClaI site being unable to substitute for the wild type sequences for resolvase binding, rather than any interference from transcriptional activity.

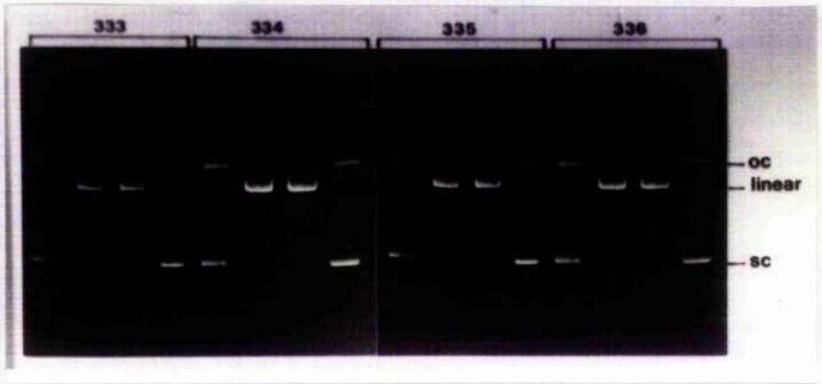


Figure 4.5 **In vitro resolution analysis of plasmids containing res-322A and res-322B.**

1% agarose gels

pPAK333 to 336 were treated in four ways: each plasmid is represented by four lanes which show from left to right:

1. Plasmid incubated with resolvase.
2. Plasmid incubated with resolvase and then HindIII.
3. Plasmid incubated with HindIII.
4. Plasmid incubated alone.

Any resolution products would run below the linear band of lane 2 for each plasmid.

4.4 DNA footprinting reactions.

To aid further analysis and manipulation of res-96 I cloned the 190bp HpaII fragment containing the 96bp TaqI fragment into the poly-linker of pUC8. A HpaII digest of pJD130 was run on an acrylamide gel. The 190bp fragment was excised and eluted and ligated with AccI linearised pUC8. Ap^r transformants of M15 which gave white colonies on X-gal plates were screened for recombinant plasmids by SCCL. Plasmid DNA was prepared from colonies containing putative recombinant plasmids and analyzed by digestion with TaqI (Fig. 4.6). In this way recombinants were isolated containing the 190bp HpaII fragment inserted in the AccI site of pUC8 in both orientations: pJD131 and pJD133.

The advantages of constructing recombinants in the poly-linker are manifold. Originally 'designed' to aid sequencing by the di-deoxy method (Vieira and Messing 1982), the poly-linker also offers an easy way to isolate restriction fragments for end-labelling and footprinting. Moreover the copy number of pUC8 is very high which allows preparation of sufficient plasmid DNA by the Birnboim-Doly method for footprinting and sequencing reactions.

pJD133 was used as the source of res-96 DNA. The plasmid was cut with HindIII and BamHI, and the HindIII ends of the total digest were partially repaired with DNA polymerase I large fragment and α-^{32}P-dATP. The labelled fragments were separated on an acrylamide gel and the 210bp labelled HindIII-BamHI fragment excised and eluted in preparation for footprinting reactions. The plasmid pPAK329 (Kitts 1982) was used to obtain res-wt DNA, by cutting with HindIII and PvuII, repairing the 5' extensions of the HindIII generated ends with

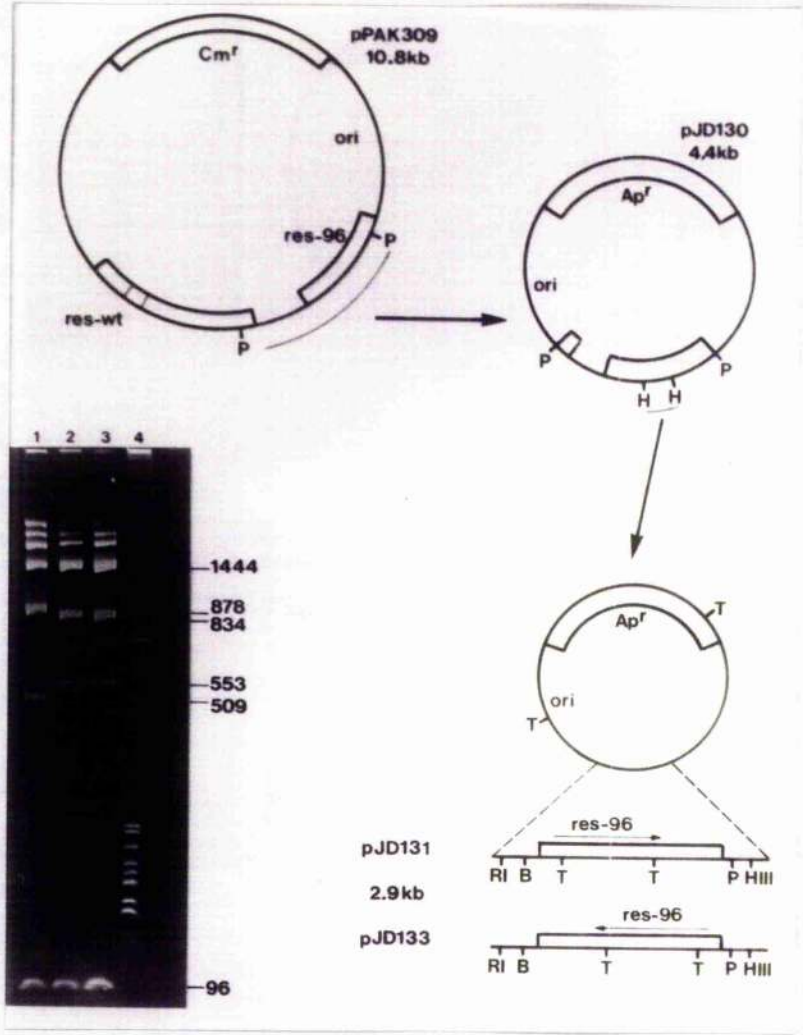


Figure 4.6 **Derivation of pJD131/3.**

P = PstI H = HpaII T = TaqI

RI = EcoRI B = BamHI HIII = HindIII

The 1.69kb PstI fragment containing res-96 was cloned into pUC8 to create pJD130. A 190bp HpaII fragment was isolated from digested pJD130 and cloned in either orientation into AccI linearised pUC8.

Clones were checked by restriction analysis on a 3.5% acrylamide gel:

Lane 1 = pJD131 TaqI digest

Lane 2 = pJD133 TaqI digest

Lane 3 = pBR322 HpaII digest

The sizes in bp of TaqI fragments are indicated.

DNA polymerase I large fragment and α-^{32}P-dATP. The end-labelled 269bp HindIII-PvuII fragment was isolated from an acrylamide gel.

Fragments containing res-322A and res-322B were isolated from pPAK331 and pPAK332 (Kitts 1982) respectively. Each plasmid was digested with HpaII and HaeIII. The 5' extensions of the HpaII ends were repaired with DNA polymerase I and α-^{32}P-dCTP. Total end-labelled digests were run onto a 400mm long preparative gel and the 274bp fragments containing the truncated res sites excised and eluted from the acrylamide.

For each fragment aliquots were measured for separate reactions: a pyrimidine sequencing reaction, a minus resolvase DNase I cleavage reaction, and plus resolvase (between 0.1ug and 0.6ug) DNase I cleavage reactions. The DNase I digestions were performed in conditions optimal for in vitro resolution, using 80ng DNase I per reaction as determined by Symington (1982). The products of each reaction were run on 8% sequencing gels which were subsequently autoradiographed (Figs. 4.7 to 4.9).

A summary of gel data showing the binding patterns of resolvase for the four different res regions is presented (Fig. 4.10). Protection from DNase I digestion reveals the three binding sites for resolvase on res-wt as demonstrated by Grindley et al (1982) and Symington (1982). Binding to all three sites is observed by addition of 0.2ug resolvase; lowering the resolvase concentration results in incomplete protection of all three sites. Within the protected regions certain phospodiester bonds remain susceptible to DNase I or are cleaved with even enhanced efficiency; the enhanced cleavage within site II with 0.1ug resolvase may indicate a slightly higher binding

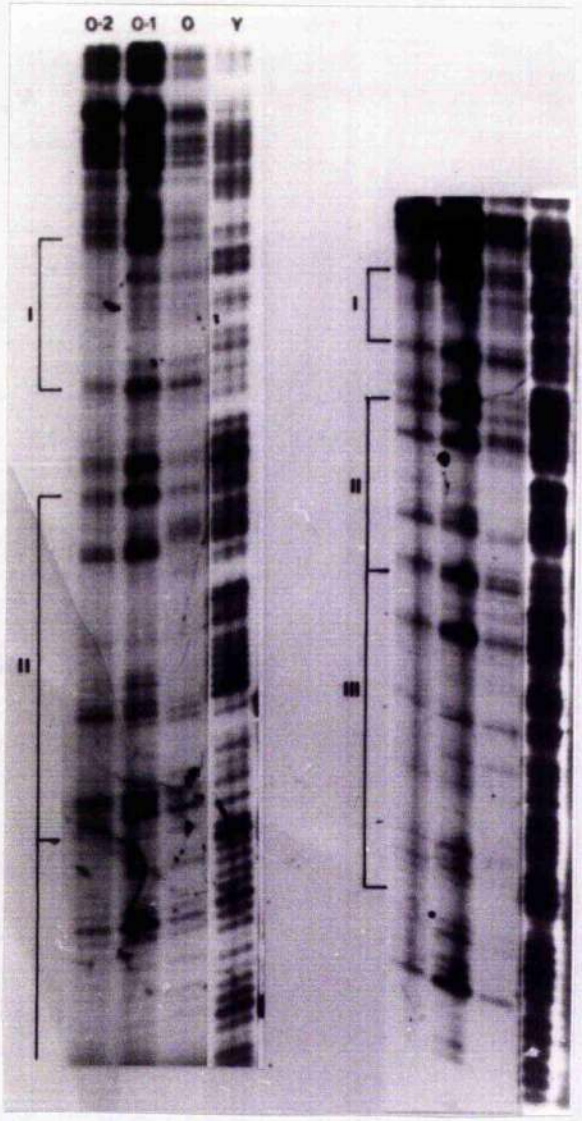


Figure 4.7 **Determination of resolvase binding sites of res-wt on the 3' labelled HindIII-PvuII fragment.**

8% sequencing gel

0.2 = +0.2ug resolvase

0.1 = +0.1ug resolvase

0 = no resolvase

Y = TC sequence track

The right hand samples were run for 3hr, the left for 1.5hr. The extent of each binding site is indicated. The extent of site III distal to site II is better represented by protection studies of the other strand (Symington 1982).

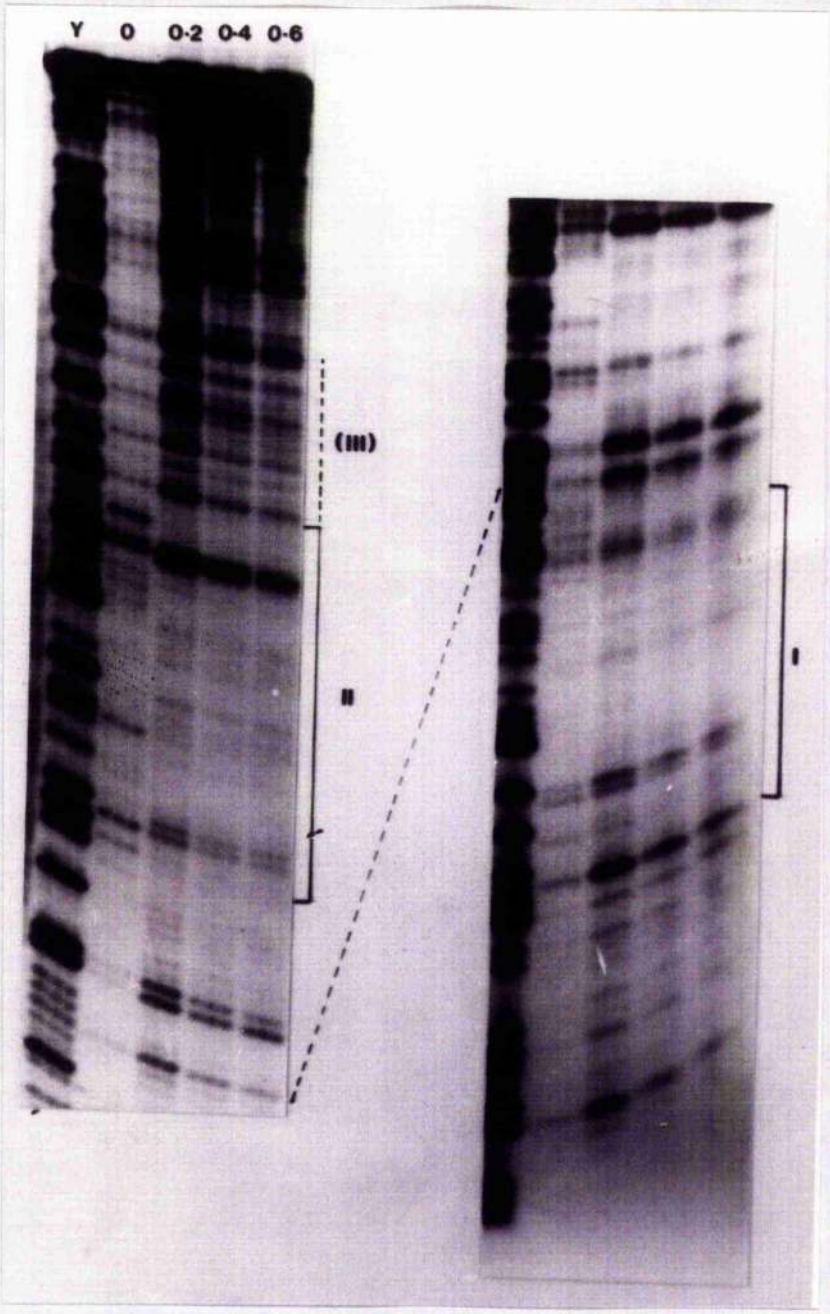


Figure 4.8 **Determination of resolvase binding sites of res-96 on the 3' labelled HindIII-BamHI fragment.**

8% sequencing gel

Y = TC sequence track 0 = no resolvase 0.2 = +0.2ug resolvase

0.4 = +0.4ug resolvase 0.6 = +0.6ug resolvase

The right hand samples were run for 3hr, the left for 1.5hr. The extent of each binding site is indicated. The position of site III is shown by the dotted line.

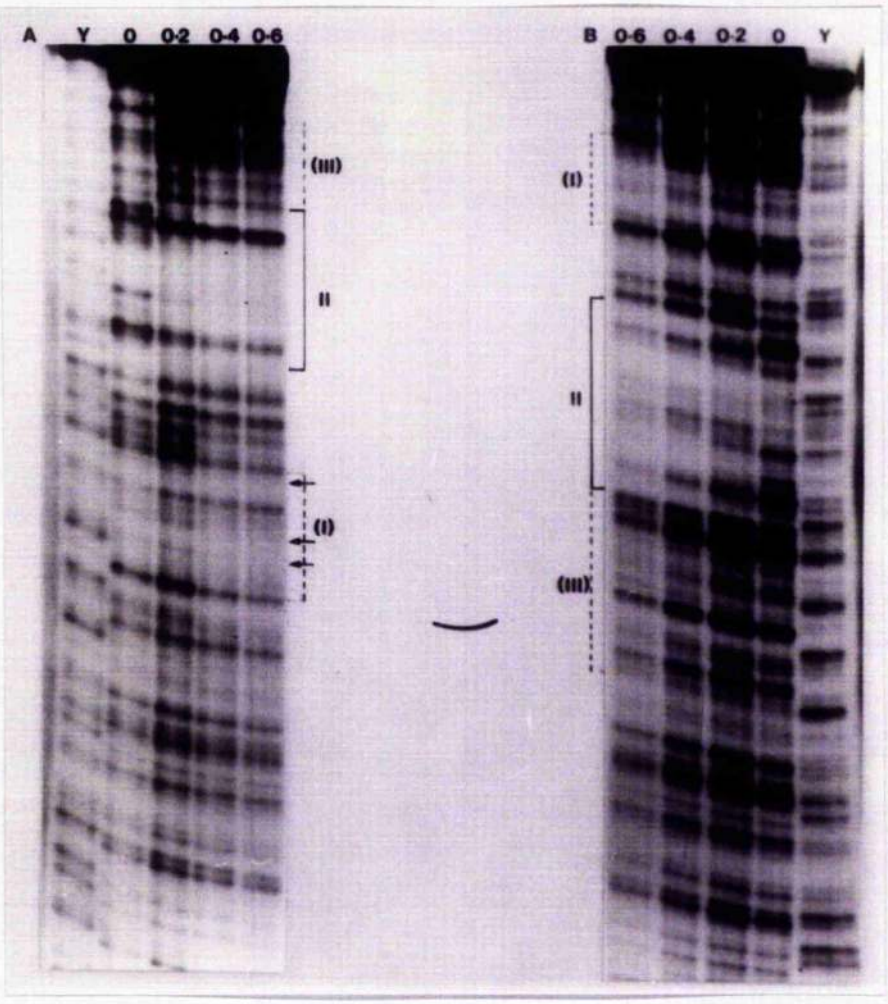


Figure 4.9 **Determination of resolvase binding sites of res-322A and res-322B.**

A=8% sequencing gel of 3' labelled HpaII-HaeIII res-322A fragment.

B=8% sequencing gel of 3' labelled HpaII-HaeIII res-322B fragment.

Y = TC sequence track 0 = no resolvase 0.2 = +0.2ug resolvase

0.4 = +0.4ug resolvase 0.6 = +0.6ug resolvase

Protection of site II in each gel is indicated. Positions of sites I and III are indicated by dotted lines. Arrows point to bands showing reduced intensity in site I of res-322A which may indicate protection at high resolvase concentration.

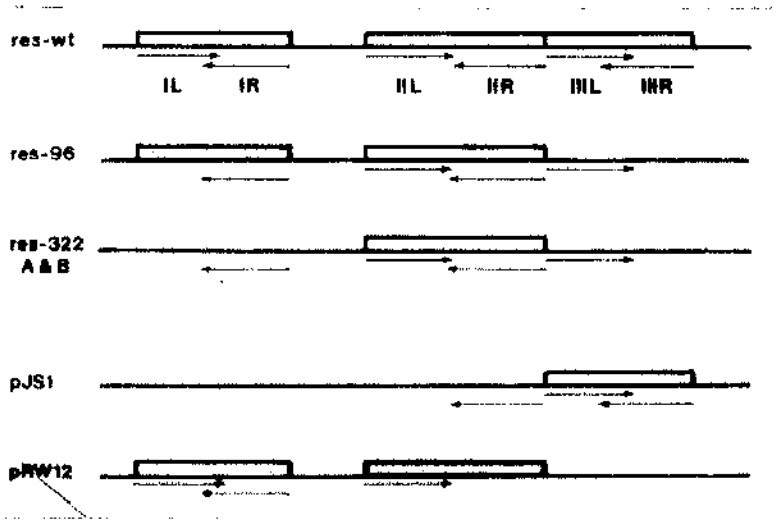


Figure 4.10 **Footprint patterns of resolvase bound to res sites.**

There are three resolvase binding sites for res-wt; each site comprising of two half-sites, arrowed. Open boxes represent resolvase binding with affinities indistinguishable from affinities for wild-type binding sites. Stipled boxes represent observable binding with lower affinities. Unprotected half-sites are indicated.

pJS1 and pRW12 contain truncated Tn1000 res-regions (Grindley et al 1982); see discussion.

affinity for this site than for sites I and III.

res-96 shows protection by resolvase at positions equivalent to site II and site I of res-wt, but not at site III. Full protection of site II is observed with 0.2ug resolvase, whereas site I is only fully protected with 0.4ug resolvase. This indicates that the sequence substitutions in site I left of res-96 confer a reduced binding affinity for resolvase. Site III, although containing a complete wild-type half site (right-hand) does not bind resolvase, suggesting a requirement of two half-sites for resolvase binding. That site II is protected at lower resolvase concentrations than site I, and in the absence of site III binding, indicates non-cooperative binding of resolvase at adjacent sites.

For res-322A, resolvase binding is represented by protection of sequences corresponding to site II with 0.2ug resolvase. Site III is unprotected at all resolvase concentrations, but the slight loss in intensity of bands (arrowed in Fig. 4.9) in site I with 0.6ug resolvase may indicate that binding with very low affinity can occur. This is not apparent for site I of res-322B, where only site II is protected.

DISCUSSION

The difference in resolvase binding sites for res-96 and res-322A/B suggests the formers functionality as a secondary res site is due to its ability to bind resolvase at site I. As the crossover site is located within site I, it is not surprising that resolvase binding at this site is important for recombination. It is also possible to derive conclusions about the nature of resolvase binding from the

observations of non-cooperative binding at adjacent sites and the absence of binding at half sites. As the 'state of the art' as regards studying interactions with DNA for resolvase is behind that of some other prokaryotic DNA binding proteins, it is possible in some cases to only derive conclusions by analogy, and in as much some points below are made speculatively.

For the best studied protein-DNA interaction: the Cro-repressor of bacteriophage lambda binds to any one of six operator sites, three techniques have been used to study the interaction. Firstly DNA footprinting defines the extent of the binding site (Johnson et al 1979). This is more apparent in conjunction with defined operator mutants located within the binding site: base pair substitutions which have differing effects on binding affinity suggest roles for specific interactions of each base pair with the protein. Furthermore an embellishment of this technique using chemical probes gives further insight. Appropriately end-labelled DNA fragments can be methylated with dimethyl sulphate (DMS) either in the presence or absence of the protein. DMS methylates the N^7 of guanine and the N^3 of adenine in double-stranded DNA; these are exposed, respectively, in the major and minor grooves (Maxam and Gilbert 1980). Cro protects guanines from DMS methylation, but enhances adenine methylation, indicating that it contacts DNA in the major but not minor grooves (Humayon et al 1977; Johnson et al 1978). In related experiments, ethyl nitrosourea was used to determine which phosphate groups of the operator are contacted by the protein (Ptashne et al 1980); these are the phosphate groups which are ethylated and hinder phosphate binding (Siebenlist and Gilbert 1979).

The three dimensional structure of Cro has been determined by crystallography (Anderson et al 1981). It includes three β-strands and three α-helices. The third of these helices (α_3) protrudes from the surface of the protein and is an obvious candidate for an interaction with DNA. To date, no structure has been determined of a sequence-specific DNA binding protein complexed with its target DNA sequence. Structural models for DNA-Cro recognition therefore rely primarily on inspection of the uncomplexed protein as seen in the respective crystal structures. This is complemented by the nature of the binding site which would place symmetrical sequences in successive major grooves of a right-handed B-DNA double helix, and that Cro itself is dimeric in solution (Takeda et al 1977). Hence a dyad axis of symmetry of the protein would coincide with a dyad axis of the DNA site.

For the dimeric protein, the 34Å spacing between the twofold related α_3 helices, together with their angle of tilt, strongly suggests that these α-helices bind within successive major grooves of right handed B-DNA as illustrated (Fig. 4.11; Anderson et al 1981; Matthews et al 1983; Ohlendorf et al 1982). It is presumed that the flexible carboxy-terminal residues of Cro participate in DNA binding by lying along the minor groove. This model, based on stereochemical data, is consistent with the chemical protection and modification studies of the DNA, and is supported by recent nuclear magnetic resonance (NMR) studies (Arndt et al 1983): the third technique to shed light on the interaction.

In order to obtain an insight into the sequence-specific recognition, model building and energy refinement were used to develop a detailed model for the complexes between Cro and DNA (Ohlendorf et

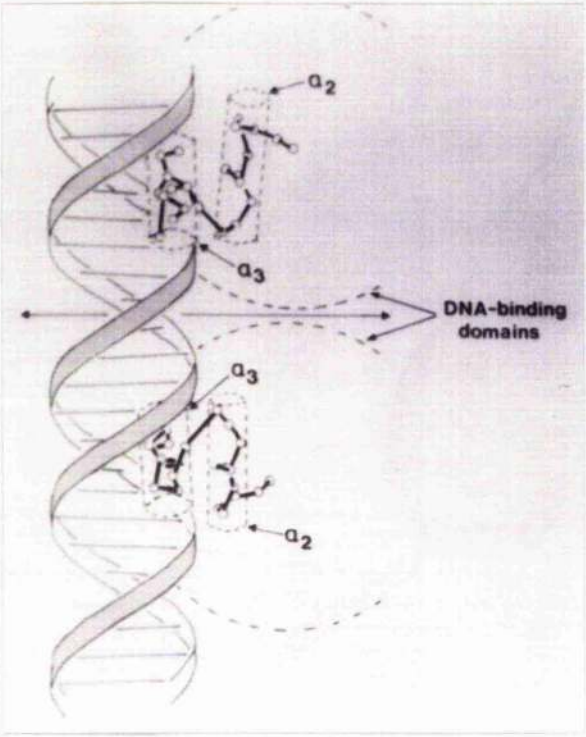


Figure 4.11 Scheme of the nature of the interaction presumed to occur between the bihelical fold of a DNA binding protein and right-handed B-form DNA after Takeda et al 1983.
Represented is a side view with the two-fold axis of symmetry (arrowed) extending from left to right.

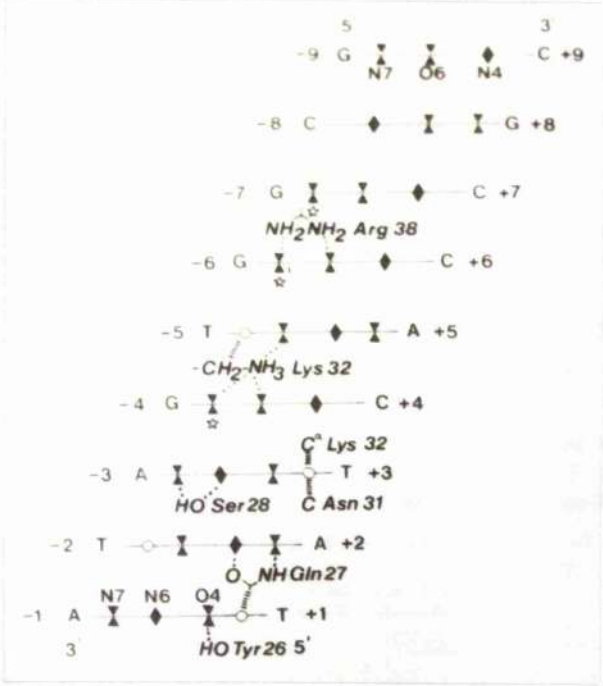


Figure 4.12 Schematic representation of the presumed sequence-specific interactions between Cro and parts of the base-pairs exposed within the major groove of the DNA of one half binding site (after Ohlendorf et al 1982).

⧗ = hydrogen bond acceptor

♦ = hydrogen bond donor

✱ = guanine N7 which is protected from methylation when Cro is bound

○ = methyl group of thymine

····· = presumed hydrogen bonds

||| = apparent van der Waals contacts

al 1982). This model is consistent with the known affinities of Cro for its sixbindingsites on lambda DNA and for mutant sites too. The model suggests that the recognition of a specific base on the DNA is due largely to a complementary network of hydrogen bonds between amino acid side chains of the protein and DNA base-pair atoms exposed within the grooves of the DNA. The hydrogen bond network that is presumed to exist between Cro and its tightest known binding site, O_R3, is shown schematically in Fig. 4.12 (Ohlendorf et al 1982). In the figure, the successive base pairs are imagined to be seen edge-on, with all the possible hydrogen bond acceptor and donor atoms indicated. Atoms that do not hydrogen bond to the protein are presumed to remain hydrogen bonded to solvent. One striking feature of the model is the multiple hydrogen bonding of the amino acid side chains. Such bi- and multidentate interactions provide a clear rationale for enhancing the specificity of DNA-protein recognition. Also, as has been shown for Lac repressor, hydrophobic interactions can play an important part in recognition (Caruthers 1980). But while protein-DNA interactions of the sort illustrated are presumed to be responsible for recognition of a specific base sequence, it is understood that the overall energy of the interaction of the complex comes primarily from interactions with the DNA that do not depend on the base sequence. In the Cro-DNA model there are many such contacts, including about 10 potential ionic interactions with the phosphate backbone (Ohlendorf et al 1982).

Is such a model specific for Cro, or can it be applied to other prokaryotic DNA binding proteins? Two other crystallised DNA binding proteins: catabolite gene activator protein (CAP) and lambda repressor both contain a substructure consisting of two consecutive alpha

helices that is virtually identical with that of Cro (Steitz et al 1982; Pabo and Lewis 1982). Structural and amino acid comparisons suggest that this bihelical fold occurs in a number of proteins that regulate gene expression, and is an intrinsic part of the DNA-protein recognition event (Takeda et al 1983). Comparison of the different sites of action of the regulatory proteins have shown that they all present a two-fold rotational symmetry and that most of them contain a consensus sequence $\frac{TGTGT}{ACACA}\ N_{6-10}\ \frac{ACACA}{TGTGT}$ or $\frac{CAC}{GTG}\ N_{5-10}\ \frac{TGT}{ACA}$ (Gicquel-Sanzey and Cossart 1982).

These findings and the observation that the regulatory proteins are multimeric, suggest a general type of DNA-regulatory protein interaction.

(i)Symmetry. Twofold symmetrical recognition sites on the DNA are recognized by oligomeric proteins with at least one twofold axis of symmetry, the dyad axis of the protein coinciding with the dyad axis of the DNA site.

(ii) DNA conformation. The DNA maintains an essentially standard right-handed B-form double-helical conformation, possibly with some bending of the DNA. Neither intercalation between base-pairs nor localized unwinding or conformational rearrangement is necessary for sequence-specific recognition.

(iii) α-Helical fold. Many gene regulatory proteins appear to contain a two helical fold derived from a common evolutionary precursor.

(iv) Helix in groove. The above two helical unit binds to DNA with the first α-helix making predominantly sequence independent interactions and the second α-helix occupying, or partly occupying,

the major groove and making sequence-specific interactions.

(v) Specificty. Specificity of recognition derives in large part from multi-dentate hydrogen bonds between amino acid side chains and the parts of the base pairs exposed within the major groove of the DNA.

Does the resolvase-DNA interaction conform with this model? The DNA binding domain has been assigned to the C-terminal 40 amino acids of Tn1000 resolvase (N.Grindley pers.comm.). Mutations of the tnpR gene which encode resolvase that can bind but not recombine map in the N-terminal half of the gene, and resolvase has been cleaved to release the C-terminal 40 amino acid peptide which binds to res sites. A two-helical fold motif has been located within this region of the protein. This example of a binding domain near the C-terminus is unusual; except for CAP, the binding domains of other prokaryotic regulatory proteins are located in the N-terminal part of the proteins studied.

Another point to consider is symmetry of the DNA recognition sites. Each binding site for resolvase exhibits dyad symmetry, and loss of one half site prevents resolvase binding to the other. Moreover the observation of non-cooperative binding between sites suggests each of the three binding sites may be considered uniquely. However, in disagreement with recognition sites for other proteins which extend for between 15 and 20bp, resolvase binding sites extend between 32bp (site III) and 40bp (site II). Simply this represents a duplication of the extent of the binding site, and in evolutionary terms that is what can be invoked. Subsequent to a duplication of a symmetrical binding site, one half (the inner halves) of each site has

presumably diverged to generate the sequence diversity present at the 3' ends of each present half-site. This would suggest that the conserving force, ie. the requirement for symmetrical binding of the two-helical fold structure to the inner halves, was negated. The evolved structure would then consist of a tetrameric protein with two symmetrical two-helical fold structures of two monomers binding within the major grooves separated by two major grooves or three turns of the helix. This could be stabilised by 'minor' binding within the unbound major grooves: this may be represented by the conserved TA^{a}_{t} sequence and the general AT richness in the 3' half of each half-site. The two monomers containing two unbound helical fold structures could participate in the mechanism to bring two res sites into alignment prior to recombination (see Chapter 8). This speculative scheme is consistent with the evolution of a bifunctional regulatory-recombinase protein from a more simple regulatory protein, but there is no evidence of the oligomeric structure of resolvase at res; it may bind simply as a dimer to each site. The purified preparations contain predominantly dimeric resolvase with some tetramers and a few monomers (Symington 1982). Resolvase may well have co-ordinately evolved with its binding sites, so that dimers could bind to each complete binding site.

The 5' half of each half-site of res contains a highly conserved TGT triplet (except for site IL: CTG) which agrees with the consensus sequence for binding of a single two-helical fold structure. This TGT triplet and the sequences immediately adjacent are presumed to constitute the specific recognition sequence where amino acid side chain-base pair bonding can occur for resolvase-res complexes. In all

cases the TGT trinucleotide is in the same position relative to the ends of the protected region. In addition in several instances the TGT is followed after one base in the 3' direction on the same strand by a short region (two phosphodiester bonds) that remains susceptible to DNase I action. In all cases DNase I sensitivity is also observed on the complementary strand around the 3'A residue of the ACA sequence that is base paired with the TGT (Grindley et al 1982). A Tn3 mutation, cis_{10}, which is a G to A transition in the TGT sequence of the left half of site II, and reduces the autoregulatory effect of resolvase (Chou et al 1979a), implicates site II in the regulation of tnpR expression and, more specifically, underlines the importance of the TGT trinucleotide in resolvase binding.

Any sequence changes within this region can affect binding affinity of resolvase for the site by changing the available sites for hydrogen bonding. It is postulated that if hydrogen bonding is a major determinant in the specificity of protein-recognition, then the loss of successive hydrogen bonds might correspond to successive reductions in the affinity of binding by a factor of seven (Ohlendorf et al 1982). The base substitutions present in res-96 site IL could constitute a slight change in available hydrogen bonding positions, in turn conferring reduced binding affinity for resolvase, whilst those base substitutions of res-322A/B site IL are presumed to result in a significant loss of available hydrogen bonding positions conferring a greater reduction in binding affinity. These sequences are compared with a consensus for the region, together with those from site IIIR where no binding was observed for res-96 and res-322A/B (Fig. 4.13). Only res-96 site IL contains a TGC triplet comparable to the TGT

trinucleotide suggested as the major determinant for resolvase binding.

Footprinting of truncated Tn1000 res regions has been carried out (Grindley et al 1982; Fig. 4.10). A ClaI restriction site in the centre of site II was used to generate truncated regions spanning (a) site I and site IIL present in pRW12, and (b) site IIR and site III present in pJS1. Footprinting of the res region of pJS1 showed protection of site III at low resolvase concentration but no protection of site IIR even at high resolvase concentration, consistent with no half site binding. Protection experiments of the res region of pRW12 showed that the half site (site IIL) was protected by resolvase although less efficiently than the intact siteII of a res-wt and only at higher resolvase concentrations than required for complete protection of site I of either pRW12 or res-wt. The authors also observed that the protected region extended a few bases into the non-Tn1000 DNA and was accompanied by a very pronounced enhancement of DNaseI cleavage 13 bases from the end of Tn1000 sequences. Despite this resolvase mediated effect in non-Tn1000 sequences, the authors suggest the protection constitutes only half site binding. Reinterpretation of their data however suggests full site binding similar to that of res-96.

Immediately adjacent to the pronounced DNaseI cleavage site, and away from the Tn1000 sequences is a 3-4 base region resistant to DNaseI even in the unprotected control reaction. It is therefore impossible to determine resolvase protection at this point. Analysis of the sequence shows this region contains a TGC trinucleotide which could allow recognition by resolvase. Moreover comparison of this

triplet and the sequences immediately adjacent to it with res-96 site IL shows 5 out of 9 base-pair homology:

consensus	T G T C Y N N T A
res-96 site IL	T G C T C G A A A
pRW12 site IIR	T G C T C A T C C

Together this suggests that resolvase might well bind, albeit with reduced affinity, to a region analogous to a complete site II in pRW12.

Interestingly the 40 amino acid C-terminal peptide obtained by cleavage of resolvase appears to bind to res DNA with greater affinity than the native protein (N.Grindley pers.comm.). Binding is observed at half sites, with differing affinities for each half site. Hence lack of observable binding in vitro of native resolvase to half sites would be a result of low binding affinity presumably arising due to a combination of the size of the native protein and its oligomerization; the truncated peptide has lost the domain invoked for oligomerisation. A half site is not a sufficiently strong anchor for resolvase dimers or tetramers. Binding to whole sites I, II and III with equal affinity implies the sum of the affinity for two constituent half sites is near equal for each site.

In conclusion, resolvase can bind to sites II and III independently of the others in vitro; this may well be the case for site I. Each site consists of two half sites in dyad symmetry together spanning a region between 30 and 40bp, and effective binding is dependent on simultaneous binding to each constituent half site. The major sequence determinant for resolvase recognition is a TGT triplet located at the 5' end of each half site, although in certain cases a C

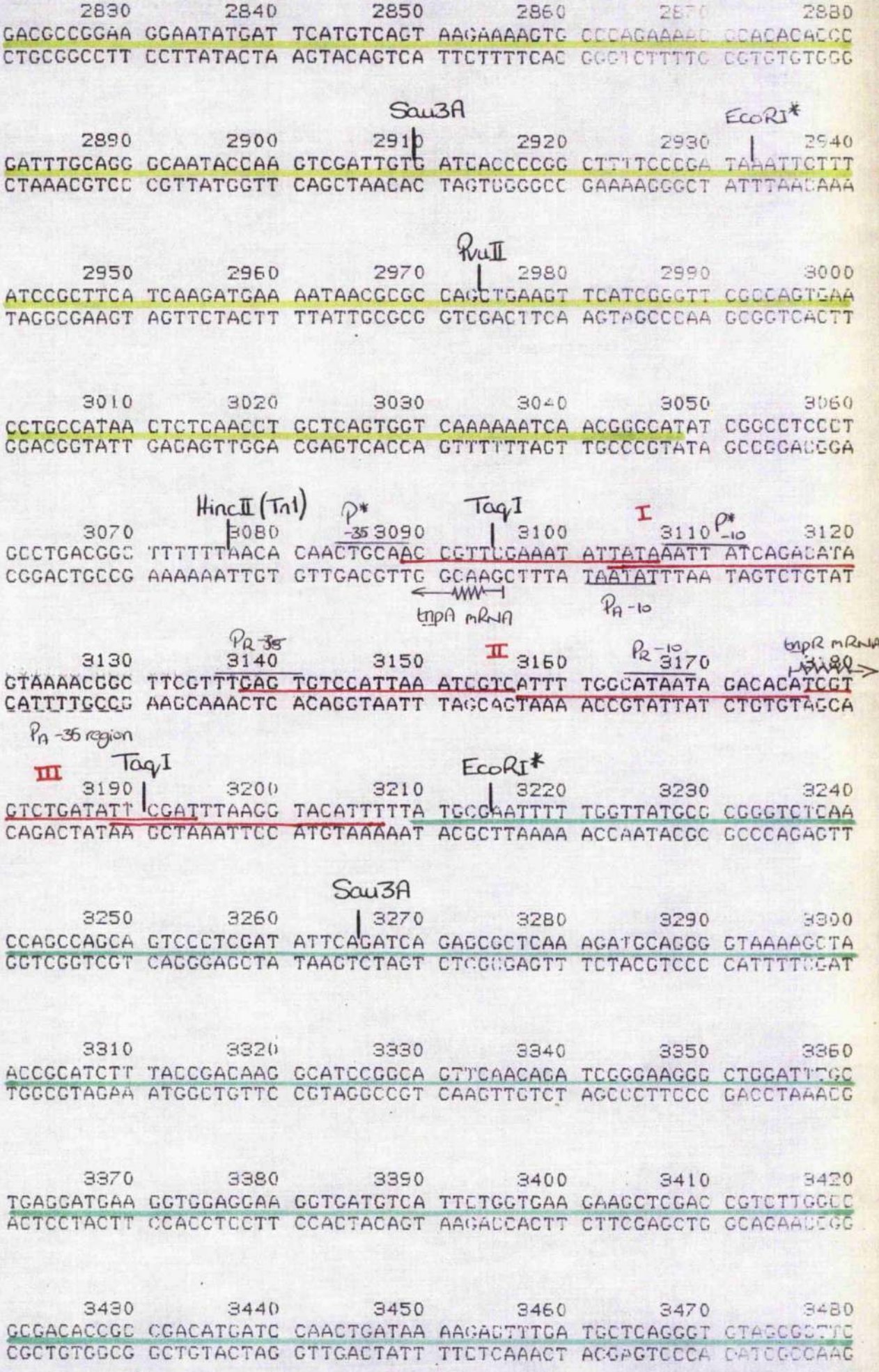
can substitute for one of the T bases. On the whole there appear to be fairly equal binding affinities for each of the three sites despite the sequence variance present. Moreover binding affinities are not grossly affected by the difference of the size of the binding sites: the 5'-TGT...ACA conserved sequences related by dyad symmetry are sparated by 22bp, 28bp and 19bp in sites I, II and III respectively. These separations alter both the linear distance and the angle of rotation between the two TGT trinucleotides of each site. This implies that binding at each site is accompanied by either flexibility of the double helix, or of resolvase itself.

Although there are three binding sites only site I acts as a recombination site. This implies that the interaction between resolvase and site I is qualitatively different from the interaction with the other sites. Only site I has the central palindromic sequence 5'-TTATAA which contains the cleavage site; this sequence may itself be important. In addition it is possible that the correct alignment of resolvase as determined by the precise separation of the TGT sequences, plays a role in activating the recombinational potential of the protein. However small differences in that separation which conserve the internal palindrome are reported to make little difference to the recombinational activity, whereas substitution of a single base in the palindrome result in complete loss of activity, although resolvase does bind to the site (N.Grindley pers.comm.).

CHAPTER 5

Do *invitro* binding studies reflect the nature of resolvase-*res* interactions *in vivo*?

Figure 3.2


2830 2840 2850 2860 2870 2880
GACGCCGGAA GGAATATGAT TCATGTCAGT AAGAAAAGTG CCCAGAAAAC CCACACACCC
CTGCGGCCTT CCTTATACTA AGTACAGTCA TTCTTTTCAC GGGTCTTTTG CGTGTGTGGG
Sau3A
EcoRI*
2890 2900 2910 2920 2930 2940
GATTTGCAGG GCAATACCAA GTCGATTGTG ATCACCCCGG CTTTTCCCGA TAAATTCTTT
CTAAACGTCC CGTTATGGTT CAGCTAACAC TAGTGGGGCC GAAAAGGGCT ATTTAAGAAA
PvuII
2950 2960 2970 2980 2990 3000
ATCCGCTTCA TCAAGATGAA AATAACGCGC CAGCTGAAGT TCATCGGGTT CGCCAGTGAA
TAGGCGAAGT AGTTCTACTT TTATTGCGCG GTCGACTTCA AGTAGCCCAA GCGGTCACTT
3010 3020 3030 3040 3050 3060
CCTGCCATAA CTCTCAACCT GCTCAGTGGT CAAAAAATCA ACGGGCATAT CGGCCTCCCT
GGACGGTATT GAGAGTTGGA CGAGTCACCA GTTTTTTAGT TGCCCGTATA GCCGGAGGGA
HincII (Tn1)
P*
-35
TaqI
I
P*
-10
3070 3080 3090 3100 3110 3120
GCCTGACGGC TTTTTTAACA CAACTGCAAC CGTTCGAAAT ATTATAAATT ATCAGACATA
CGGACTGCCG AAAAAATTGT GTTGACGTTG GCAAGCTTTA TAATATTTAA TAGTCTGTAT
tnpA mRNA
PA -10
PR -35
II
PR -10
tnpR mRNA
3130 3140 3150 3160 3170 3180
GTAAAACGGC TTCGTTTCAG TGTCCATTAA ATCGTCATTT TGGCATAATA GACACATCGT
CATTTTGCCG AAGCAAACTC ACAGGTAATT TAGCAGTAAA ACCGTATTAT CTGTGTAGCA
PA -35 region
III
TaqI
EcoRI*
3190 3200 3210 3220 3230 3240
GTCTGATATT CGATTTAAGG TACATTTTTA TGCGAATTTT TGGTTATGCG CGGGTCTCAA
CAGACTATAA GCTAAATTCC ATGTAAAAAT ACGCTTAAAA ACCAATACGC GCCCAGAGTT
Sau3A
3250 3260 3270 3280 3290 3300
CCAGCCAGCA GTCCCTCGAT ATTCAGATCA GAGCGCTCAA AGATGCAGGG GTAAAAGCTA
GGTCGGTCGT CAGGGAGCTA TAAGTCTAGT CTCGCGAGTT TCTACGTCCC CATTTTCGAT
3310 3320 3330 3340 3350 3360
ACCGCATCTT TACCGACAAG GCATCCGGCA GTTCAACAGA TCGGGAAGGG CTGGATTTGC
TGGCGTAGAA ATGGCTGTTC CGTAGGCCGT CAAGTTGTCT AGCCCTTCCC GACCTAAACG
3370 3380 3390 3400 3410 3420
TGAGGATGAA GGTGGAGGAA GGTGATGTCA TTCTGGTGAA GAAGCTCGAC CGTCTTGCCC
ACTCCTACTT CCACCTCCTT CCACTACAGT AAGACCACTT CTTCGAGCTG GCAGAACGGG
3430 3440 3450 3460 3470 3480
GCGACACCGC CGACATGATC CAACTGATAA AAGAGTTTGA TGCTCAGGGT CTAGCGGTTC
CGCTGTGGCG GCTGTACTAG GTTGACTATT TTCTCAAACT ACGAGTCCCA GATCGCCAAG
tnpA gene
resolvase binding
tnpR gene

INTRODUCTION

The correlation of the extent of res-wt defined in vivo (Chapter 3) and the extent defined in vitro by protection experiments with resolvase suggests the latter provides a good insight into the nature of the protein-DNA interaction. This is despite the nature of the res-wt DNA in the in vitro experiments: a non-supercoiled restriction fragment, whereas a requirement for the in vitro recombination reaction is a supercoiled substrate as found in vivo. Hence resolvase binding, but not recombination, can occur on relaxed substrates. With truncated sites a supercoiled res, which has unprotected half sites as determined in vitro, may provide alternative interactions for resolvase and the site in vivo. As stated in Chapter 4, the in vitro observations of lack of binding of resolvase to half sites is a result of reduced binding affinity to these sites; conditions in vivo may indeed enable binding to half-sites (a) by allowing different localised concentrations of resolvase to DNA, (b) optimising the local pH and ionic conditions for binding, and (c) providing a superhelical substrate.

The possibility to test such alternative interactions arises from the bifunctional nature of resolvase, being both a site-specific recombinase and a repressor of both tnpA and tnpR expression. By comparing the location of the divergent promoters of these genes with the binding sites of resolvase for each res region, it is possible to assess whether binding will shield these promoters from access by RNA polymerase. If the binding patterns observed are a representation of the in vivo interactions, then the amount of expression of these genes

will be correlated to the degree of shielding of each promoter as predicted from the protection experiments.

A criterion for such experiments is that the precise position of both promoters is known. Originally the location of both promoters was surmised by comparison with other known promoter sequences. It was noted by Pribnow (1975a,b) and Schaller et al (1975), after comparing only a few prokaryotic promoter sequences, that each contained a 7bp sequence, generally homologous to the sequence 5'-TATAATG, which was centred about 10bp upstream of the mRNA startpoint. Sequence similarity has also been noted among promoters in a region about 35bp preceding the mRNA starpoint (Takanami et al 1976; Seeburg et al 1977). Sequences at this -35 region are generally homologous to the sequence 5'-TTGACA. On this basis the tnpA promoter was initially located around Tn3 coordinates 3100 (-10) and 3120 (-35); and the tnpR promoter around 3170 (-10) and 3140 (-35) (Heffron et al 1979). A mutation which was isolated and gave increased expression of tnpA led to the suggestion that the -10 for tnpA might be located at position 3070 (Casadaban et al 1980); however no corresponding -35 sequence was invoked.

The criteria adopted to recently compile a list of sixty E.coli promoter DNA sequences are; (i) the 5' terminal nucleotides of the transcript be determined, or (ii) one or more promoter mutations have been sequenced (Hawley and McClure 1983). Sequencing the 5' terminus of the tnpA mRNA of Tn3 (A. Lamond pers. comm.) showed the -10 sequence around coordinate 3100 as proposed by Heffron et al (1979). At this position is a 5'-TATAAT sequence which is a perfect match to the consensus -10 sequence. The spacing between -10 and -35 regions is

critical for promoter function (Stefano and Gralla 1982; Mandecki and Reznikoff 1982) and spacing of 17bp gives strongest promoter function, although promoters with spacings of 15 and 20bp have been reported to retain partial function (Hawley and McClure 1983). Using these parameters, ie. searching for a -35 candidate sequence between 15 and 20bp upstream of the -10, it is impossible to find a good sequence match with the consensus -35: 5'-TTGACA, or in fact any 'known' -35.

The tnpR mRNA sequence (A.Lamond pers.comm.) suggests a -10 at coordinate 3170 as proposed. The -10 sequence present is 5'-CATAAAT which is exactly homologous to the -10's of the chloramphenicol resistance gene, cat, lambda P_o, and the supB-E gene. Using the 15 to 20bp spacing parameter no good candidate for a -35 is present. Extending the scope of the spacing parameter to 22bp the sequence 5'-TTGAGT is found. This is exactly homologous to three other -35's: lambda P_o (as for the -10), R100 RNAII, and P_{ori-L}.

The 5' termini of both tnpA and tnpR mRNA's of Tn1000 have been sequenced (Reed et al 1982). The tnpA -10 is the same as for Tn3, and the -35 bears no real resemblance to consensus. The published promoter signals for tnpR are at completely different positions than for Tn3, but whilst these represent real promotional signals (P^*) the position of the Tn1000 tnpR promoter has been corrected to the equivalent position as that for Tn3 (N.Grindley pers.comm.).

So whilst the positions of -35 sequences for both genes are under dispute, the locations of the -10 sequences are established. The tnpA -10 sequence lies within resolvase binding site I, and the tnpR -10 within site II (Fig. 5.1). This provided the basis for the experiments described below.

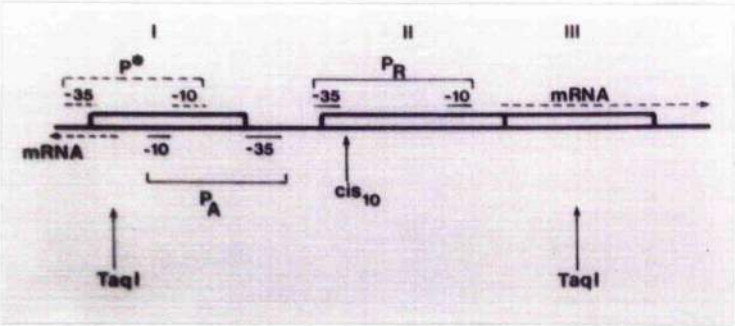


Figure 5.1 **Schematic representation of the positions of P_A and P_R relative to resolvase binding sites.**

P_A andP_R -10, -35 and mRNA start positions are indicated, as are the TaqI restriction sites. The position of P^* is shown, and the point of the cis10 mutation - see discussion.

Previously both the strength of tnpA and tnpR promoters and the effect of repression by resolvase has been assayed by the technique of gene fusion. Both genes were fused to the lacZ gene and the amount of β-galactosidase produced was assayed with and without resolvase being supplied in trans (Chou et al 1979a,b). Gene fusions present several limitations to the study of promoters though: perhaps greatest of these are the polar effects (transcriptional, translational, or both) that fusions may introduce within the operon under study. These effects can be quite variable, depending on the the position at which the fusion occurs. Apparently fusion of the same gene at even slightly different locations in the operon can result in substantially different degrees of expression (Mercereau-Puijalon and Kourilsky 1979; Merril et al 1978).

A system that circumvents these problems uses the E.coli galactokinase gene to study regulatory signals (McKenney et al 1981); it was devloped to satisfy the following requirements.

(i) Fusions should be to a gene with a readily assayable product. The assay should be simple, sensitive, and linear over a wide range.

(ii) Expression of the gene function should be proportional to the amount of transcription, thereby measuring the overall efficiency of the promoter. This requires that the translation efficiency of the mRNA remain nearly constant, irrespective of the regulatory region to which it is fused. To achieve this, all translation except that of the assayable function must be uncoupled from the transcription unit to avoid differential effects on the expression of that gene. In addition, the translation efficiency of the mRNA must be relatively immune to changes in the upstream 5' terminal RNA structures resulting

from the fusion.

(iii) The gene function should be a readily selectable genetic marker; those with both positive and negative selections are the most flexible.

(iv) A second selectable marker other than that used for characterisation of the regulatory site should be part of the vector.

(v) The existence of several unique cloning sites allows diversity, simplicity and precision in fusing small DNA fragments that contain the regulatory site of interest to the gene.

(vi) Transcriptional expression of the gene must depend solely on the inserted regulatory site. Transcription originating elsewhere on the vector must not traverse the gene.

(vii) The exact DNA sequence should be known between the cloning site and the gene. In addition, extensive restriction information of the gene is extremely useful.

The plasmid pKO1 was designed to satisfy these requirements. The galK gene, including 168bp of the region normally preceding the galK coding sequence, was inserted into a pBR322 derivative from which was deleted the entire Tc region of the plasmid. Several unique restriction sites were constructed to allow precise cloning of a variety of DNA fragments upstream of galK. Any of these sites can be used to insert DNA fragments with either defined or potential transcriptional regulatory signals. When fused in the proper orientation the inserted regulatory signal will control galK expession. The galK gene is not expressed by the starting vector, pKO1, and this plasmid cannot complement an $E^+T^+K^-$ host. Thus, there

is little if any transcription of the galK coding region without insertion of a promoter site in the proper orientation at one of the cloning sites.

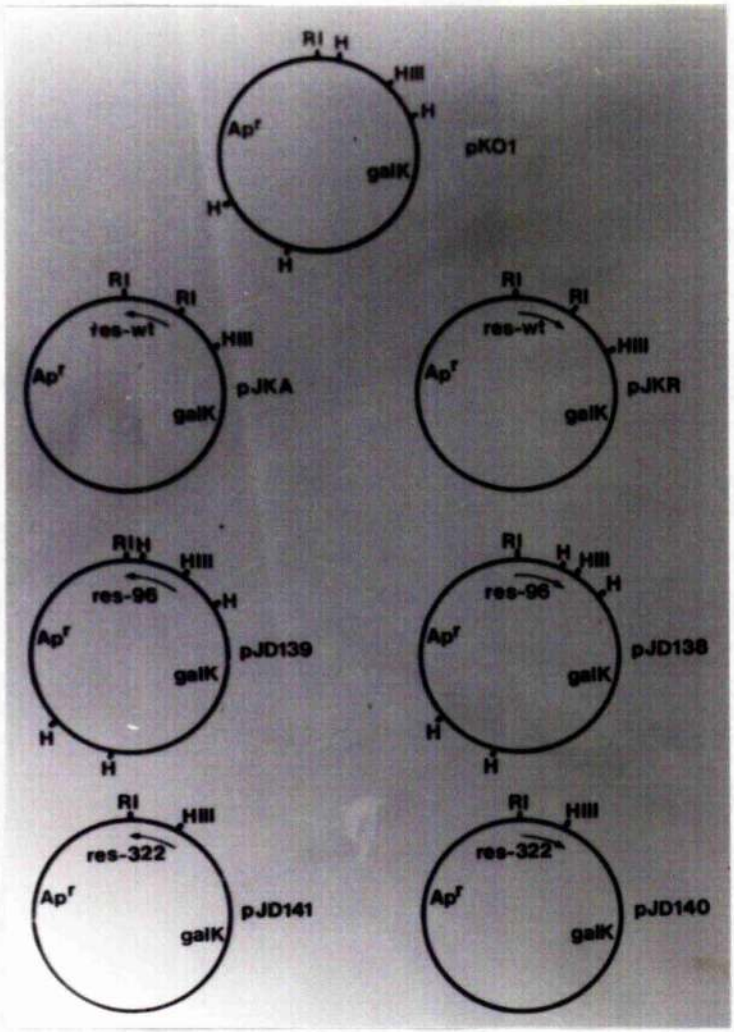
To minimize the potential effects of upstream translation, translation stop codons were constructed in all three reading frames beyond the cloning sites, located at 174, 121 and 74bp before the galK coding sequences. These stop codons uncouple from the galK transcript all translation initiating in any DNA insert. Thus, galK has a relatively uniform untranslated leader region, ensuring constant efficiency for galK translation.

RESULTS

5.1 Construction of vectors.

Plasmids containing the tnpA promoter (P_{A-wt}) and tnpR promoter (P_{R-wt}) of res-wt controlling galK were constructed by an undergraduate project student, Patrick Kelly. He cloned the EcoRI fragment containing res-wt from pPAK330 into the EcoRI site of pKO1, and proceeded to isolate recombinant plasmids containing the insert in either orientation. Ligated DNA was used to transform N100; the transformation mix was plated on MacConkey galactose plates. Recombinant clones showed up as red colonies.

A vector containing P_{A-96} controlling galK was constructed by replacing the 295bp EcoRI-HindIII fragment of pKO1 with a 210bp EcoRI-HindIII fragment from pJD131. Similarly for the P_{R-96} vector the 295bp EcoRI-HindIII fragment was replaced with a 210bp EcoRI-HindIII fragment from pJD133. Promoter plus recombinant clones were



facing page 104

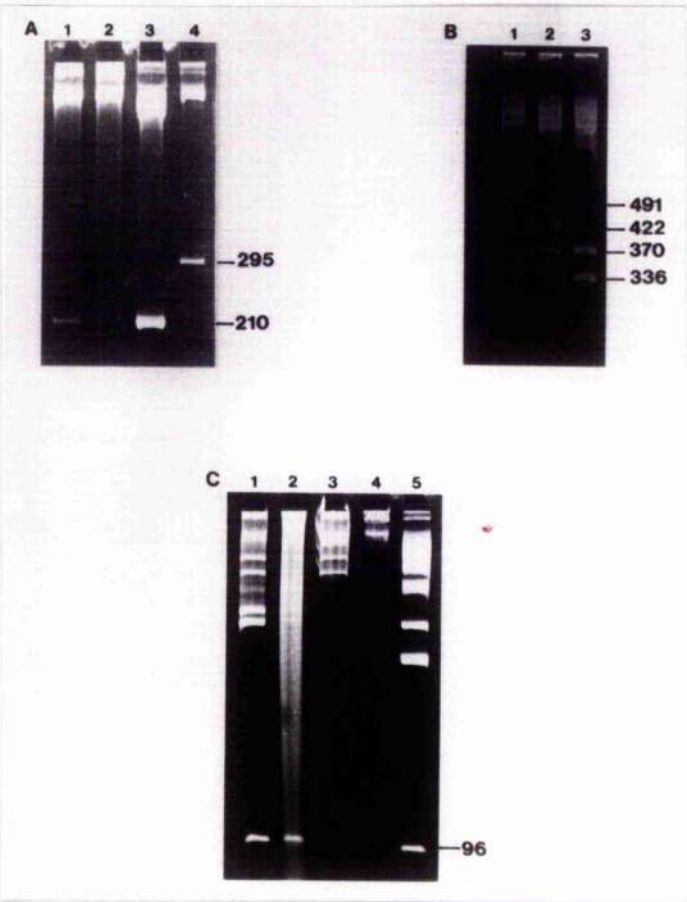


Figure 5.2 **Structures of pKO1 derived promoter vectors.**
RI = EcoRI H = HaeII
HIII = HindIII

Structures were confirmed by restriction analysis (sizes in bp):
Gel A: 5% acrylamide
lane 1 = pJD138 HindIII/EcoRI digest
lane 2 = pJD139 HindIII/EcoRI digest
lane 3 = pJD133 HindIII/EcoRI digest
lane 4 = pKO1 HindIII/EcoRI digest
Gel B: 5% acrylamide
lane 1 = pKO1 HaeII digest
lane 2 = pJD139 HaeII digest
lane 3 = pJD138 HaeII digest
Gel C: 5% acrylamide
lane 1 = pJD140 TaqI digest
lane 2 = pJD142 TaqI digest
lane 3 = pKO1 TaqI digest
lane 4 = pPAK331 TaqI digest
lane 5 = pPAK332 TaqI digest

The orientation of the 96bp fragment in each plasmid is inferred from data presented in Chapter 4.

selected by their ability to show up as red colonies on MacConkey galactose plates.

To avoid complicated constructions only the P_{A-322A} and P_{R-322B} promoters were linked up to galK. In both cases vector construction proceeded by replacing the 295bp EcoRI-HindIII fragment of pKO1: for pJD141 containing P_{A-322}, the replacement EcoRI-HindIII 125bp fragment was derived from pPAK331, and for pJD140 containing P_{R-322B}, it was derived from pPAK332. In both cases promoter plus recombinant clones were detected as red colonies on MacConkey galactose plates.

All recombinant plasmids were checked by restriction mapping prior to further strain constuction and the galactokinase assays (Fig. 5.2).

5.2 Galactokinase assays.

A set of strains was constructed to assay for galactokinase expression. For negative controls, N100 alone and N100 containing pKO1 were used. For the other strains galactokinase expression was measured for cells containing (i) the test vector alone, (ii) the test vector plus the plasmid pACYC184, tnpA$^-$ tnpR$^-$, (iii) the test vector plus the plasmid pAA33 (pACYC184::Tn3) tnpA$^+$ tnpR$^+$, and (iv) the test vector plus the plasmid pAA231 (pACYC184::Tn103) tnpA$^+$ tnpR$^-$.

The choice to use pACYC184 and derivatives thereof as complementing plasmids was made because they are of similar copy number to that of the pKO1 derivatives. The levels of repression mediated by resolvase may indeed be dependent on the overall cellular concentration of resolvase. For instance, complementation with low copy number tnpR$^+$ plasmids reduces the level of repression (P.Kelly

pers.comm.). But for the experiments reported here, for each test strain a similar steady state of resolvase compared to available binding sites is assumed. There is also an effect of differing strength promoters on the copy number of plasmid vectors in which the promoter resides (Wong et al 1982; Stueber and Bujard 1982). This manifests itself as an inverse correlation between promoter strength and plasmid levels in the galK expression vector system (Lamond and Travers 1983). For the experiments reported here, no correction for altered plasmid copy number has been made. However analysis of the results and overall conclusions obtained from these experiments shows that they would not be qualitatively altered by a plasmid copy number correction factor.

The results of the galactokinase assays are presented in Table 5.1, the figures being derived from three separate assays per strain; they each gave consistent results. Table 5.2 summarises the results of the GalK activities of each promoter with and without resolvase present, and expresses the level of repression as a ratio of unrepressed activity divided by repressed activity. For P_{A-wt} the repressed activity is about one third of unrepressed. Remarkably both unrepressed P_{A-96} and P_{A-322} show between one-tenth and one fifth of the activity of P_{A-wt}. The repression ratios for these three promoters are quite different.

P_{R-wt} has about one third the activity of P_{A-wt}, and is repressed by a factor of four. P_{R-322} and P_{R-96} show one half to four-fifths the activity of P_{R-wt}, and have similar repression ratios of about2.5.

Table 5.1 **Galactokinase activities of promoters plus and minus resolvase.**

Galactokinase units are expressed as nanomoles of galactose phosphorylated per minute per ml of cells at $OD_{650}=1.0$.

Strain / Vector	N100	N100+ pACYC184	N100+ pAA231	N100+ pAA33
-	2			
pKO1	10			
pJKA $P_{A\text{-}wt}$	160	160	168	50
pJKR $P_{R\text{-}wt}$	51	48	55	12
pJD139 $P_{A\text{-}96}$	19	21	21	10
pJD138 $P_{R\text{-}96}$	39	40	42	18
pJD141 $P_{A\text{-}322}$	30	29	36	29
pJD140 $P_{R\text{-}322}$	30	33	34	13

facing page 107

Table 5.2 **Summary of galactokinase activities.**
Galactokinase units expressed as above. The repression ratio is obtained by dividing unrepressed activity by repressed activity.

	unrepressed	repressed	repression ratio
P_{A-wt}	163	50	3.2
P_{A-96}	20	10	2.0
P_{A-322}	32	29	1.1
P_{R-wt}	51	12	4.3
P_{R-96}	40	18	2.2
P_{R-322}	32	13	2.5

DISCUSSION

The set of experiments described above were designed to see if an *in vivo* analysis of resolvase-*res* interactions could complement the *in vitro* data derived from DNA footprinting. Ultimately they have raised important questions as to the nature of the sequences involved in efficient promoter activity which must be addressed prior to discussion of the effects of repression.

This issue is raised primarily from the observation that compared to P_{A-wt}, the GalK activities of P_{A-96} and P_{A-322} are both much reduced compared to the wild type level. Simple explanations such as the presence of alternative same-sense promoters in the *Eco*RI fragment of pJKA but not in the *res* regions cloned in pJD138 and pJD141 are ruled out; no sequences corresponding to such promoter signals can be found. Moreover there is no evidence to suggest the presence of a CAP binding site upstream to P_{A-wt}, and deleted in P_{A-96} and P_{A-322}, to which CAP might bind and stimulate promoter activity. Promoter sequences of 'opposite sense' which might interfere with P_{A-96} and P_{A-322} activity have also been searched for; again there are no candidates in either of the constructs. A more realistic explanation would therefore involve the sequences substituted downstream of the -10 region; for this an overall picture of the events involved in effective promotion is required. It should be noted that the different GalK activities are not due to the loss of ribosome binding sites and translation start signals in the cloned *res* regions of pJD138 and pJD141. The GalK expression system is designed to make expression independent of the translation, ie. the translation of *galK* is

independent of upstream translation signals.

For convenience, the overall promoter-polymerase interaction can be divided into four sequential steps. These processes, or mechanistic steps, are (1) promoter location; (2) promoter recognition and formation of the 'closed'-promoter polymerase complex; (3) formation of the 'open'-promoter polymerase complex, and (4) mRNA initiation.

Promoter location is thought to involve the typical processes now invoked for DNA binding proteins that can act *in trans* (Berg *et al* 1982). A simple process of three-dimensional diffusion cannot account for the speed of promoter location. Instead, following non-specific binding to DNA, the polymerase is thought to 'slide' along the DNA or 'hop' between superhelical domains, essentially in a one-dimensional diffusion process (see Chapter 8). Promoter recognition and formation of the 'closed complex' is a sequence-dependent event. Much as described for Cro in Chapter 4, polymerase must recognise and bind to specific base-pair sequences - presumably in the -10 and -35 regions; the specificity being afforded by potential bi- and multi-dentate hydrogen bonding between amino acid side chains and acceptor/donor sites along the major and minor grooves of the DNA double helix. 'Open complex' formation is a far less defined process, and involves an isomerization of the 'closed complex' in which a limited number of base pairs are opened. Mutations in the sequences around -10 can affect this isomerisation but not initial binding to any extent, implying the -10 region to be crucial for 'melting-out'. In fact the open complex extends from approximately -9 to +3. Evidence suggests that the 'strength' of a promoter is dependent on steps (2) and (3), and that either can be limiting (Hawley *et al* 1982).

Following open complex formation the system moves into the intiation phase, which appears to be complex, and even its definition has not been agreed upon. The open complex interacts with nucleoside triphophates to carry out de novo RNA chain initiation. Mutational evidence suggests some sequence specificity to this process and indicates that the RNA polymerase, at least in vitro, undergoes a promoter sequence-specific recycling reaction for the first polymerisation steps (Reznikoff et al 1982).

The differences between GalK activities of P_{A-wt} and those of P_{A-96} and P_{A-322} must be assessed in terms of steps (2), (3) and (4), as for each of these steps there is some sequence dependence. The question arises as to which step is limited in terms of the sequence changes which extend from +3 downstream in P_{A-96} and P_{A-322}. Step (2) involves recognition of -35 and -10 sequences (von Hippel et al 1982), which are conserved in all three promoters. That is at least those sequences corresponding to the -35 position are conserved; as discussed before there is no good candidate for a -35 sequence in this region. For this reason it is quite possible that P_A recognition involves other sequences. One distinct possibility is cooperative binding between polymerase bound to P_R and P_A; here again there should be little difference for the three P_A-promoters since P_R -10 and -35 sequences are conserved. Another possibility is that sequences involved in P_A location, by polymerase are found downstream of the -10, but there is no precedent for this and such sequences are found upstream of 'stable' promoters such as that of $tRNA^{Tyr}$ (Travers et al 1983).

In considering step (2) a pertinent comparison might be made with

the promoter of the uvrA gene. A region between the ssb and uvrA genes carries divergent promoters for these genes, and LexA binds around the uvrA -35 to partially repress transcription of this gene. By cloning restriction fragments from this region into pKO1, it has been shown that the uvrA promoter is efficient even when divorced from the ssb promoter sequences and its own -35 sequence (Backendorf et al 1983). A feature of the sequences around -10 are two direct repeats of the sequence 5'-TTGTGT separated by 13bp, the second at +2 - these are thought to be involved in polymerase recognition. As yet there is no data to show the effect of deleting either of these repeats on promoter activity. A search for similar repeats around P_{A-wt} shows no candidates.

On the whole it appears that sequence changes of +3 onwards would not be significant in giving altered GalK activities due to differences in promoter recognition and formation of closed complexes. Step (3), the isomerisation of closed to open complex, is sequence dependent to the extent that for unstable promoters the 'melting out' of DNA is dependent on general AT richness between positions -6 to +3 (A. Travers pers. comm.). The comparison of these sequences for the three promoters is:

	-6 +3
P_{A-wt}	5'-TATTTCGAACGG
P_{A-96}	5'-TATTTCGAGCAG
P_{A-322}	5'-TATTTCGATAAG

The A to G transversion at position +3 of P_{A-96} could easily affect open complex formation and hence initiation. Less easy to see

is an equivalent affect of the A to T transversion at +13 of P_{A-322} which preserves the AT richness.

As suggested above, a sequence dependent recycling process occurs for the first polymerisation steps in vitro. A summary of this process is that several different products are made from each different promoter system with significant production of 2 to 8 mers as well as elongated products; the relative abundance of various products differs for different promoters. So at approximately position +8 (differing for each promoter) there is a choice between release of the product oligonucleotide followed by reinitiation at the +1 site versus proceeding to the next polymersisation step. Mutations in the lac-controlling element sequence at position +5 and +10 give altered expression, and these are assigned to alter initiation and/or recycling efficiencies (Reznikoff et al 1982).

In conclusion, the best explanation for the different activities of P_{A-wt}, P_{A-96} and P_{A-322} lies in the sequence changes from +3 downstream. These changes are likely to result in reduced initiation and elongation efficiences for P_{A-96} and P_{A-322}. The lack of comparable studies on the 3' sequences of other promoters makes this conclusion speculative at this time. An experiment which might help to discern if open-complex formation is less efficient for P_{A-96} and P_{A-322} would be footprinting of RNA polymerase bound to these promoters: this should reveal the extent of open-complex formation for each promoter.

It is noticeable that activities of P_{R-96} and P_{R-322} are both down on wild-type. This less dramatic effect is not readily attributable to the above process of polymerase-promoter interaction,

as all sequences from -80 to +18 are conserved in the constructs. Moreover there are no anti-sense promoters in pJD139 and pJD140 which might interfere with P_R. However, the same-sense promoter, P^*, at coordinate 3105 (-10), initially reported as being the Tn1000 P_R (Reed et al 1982) is deleted for its -35 sequence in pJD139 and pJD140. This could explain reduction in GalK activity for these plasmids.

Whilst this second promoter has not been shown to bind RNA polymerase (Wishart et al 1983) compared to weak binding to P_A and P_R, it does express galK weakly when divorced from P_R (N.Grindley pers.comm.). Therefore it would be reasonable to suggest that some GalK activity of P_{R-wt} is in fact derived from this second promoter.

Whilst P_{A-wt} is of reasonable strength, P_{R-wt} is a weak promoter. Both display sub-optimal sequences at -35, indicative of weak binding of RNA polymerase as observed in vitro (Wishart et al 1983). The -10 sequence of P_{A-wt} fits with consensus, which together with the effect of +3 downstream base substitutions suggests initiation at P_{A-wt} is good. The spacing between -35 and -10 of P_{R-wt} is sub-optimal, and even with the effect of an upstream, albeit weak, same-sense promoter, the promoter is quite weak, indicative of poor initiation.

The effect of resolvase binding is to reduce both P_{A-wt} and P_{R-wt} activities by similar amounts. The amount of repression in fact is not large. The repression ratio for P_{A-wt} resulting from resolvase binding at site I is 3.3. Using the lacZ fusion system, Chou et al (1979b) obtained a repression ratio at 30°C (the temperature at which the GalK assays were carried out) of 5.2 for P_A. For P_{R-wt} the repression ratio

is 4.25: larger possibly because resolvase binding at both sites II and III may prevent access by RNA polymerase. The repression ratio for P_R obtained by Chou et al (1979a) was 5.51. These figures can be put in perspective if one considers the difference in expression of the lactose operon between its induced and repressed states is 10^3-fold.

So why should resolvase mediated repression be poor? One reason would be a smaller binding affinity of resolvase for any of its binding sites than for instance lac repressor for its operator. Secondly RNA polymerase may compete successfully for binding at the res region; this competition may be enhanced if an adjacent sequence specificity for polymerase allowed it to 'nudge' its way into the region, displacing resolvase. A third effect concerns the cellular concentration of resolvase compared to available res regions. Probably all proteins that have a great affinity for a specific sequence also possess a low affinity for any DNA sequence. A large number of low affinity sites will compete just as well for a repressor as a small number of high-affinity sites; this implies that the binding of resolvase to the res region will be very sensitive to both the total DNA concentration and the total resolvase concentration in the cell, and at low resolvase concentration it could be impossible to establish tight repression.

The effect of introducing the sequence substitutions that reduce binding affinity of resolvase to the three sites is to lower the repression ratios. The sequence changes in site IL of res-96 result in a reduction of the repression ratio, which would correlate with the observed reduction in binding affinity observed in vitro. The substitutions present in site IL of res-322A, which resulted in no

resolvase binding at site I *in vivo*, confer a repression ratio near unity. This would suggest *in vivo* there is no binding at *res*-322A site I, despite complete conservation of the site IR half-site.

The sequence changes in site IIIR of *res*-96 and *res*-322B resulted in complete loss of binding *in vitro* to site III. The *in vivo* repression data shows that this correlates with a slight reduction in repression of P_R. As site II overlaps with the -35 and -10 regions of this promoter it, not site III, is probably the most significant binding site in regard to P_R repression. In conjunction with this, the *cis*10 mutation, which is a G to A transition at position 3142, gives a repression ratio of 2.36 compared to 5.51 of wild-type (Chou *et al* 1979a). This transition changes the TGT binding determinant of site IIL to TAT. It would be interesting to analyse RNA polymerase-resolvase binding competition for this mutation, and to know the resolvase protection pattern *in vitro*.

As the same extent of site IIIL half-site is present in *res*-96 and *res*-322B, it is impossible to say if the repression observed results from resolvase binding just at site II, or at the site IIIL half-site aswell. If the prediction was true that a half-site alone does not constitute a resolvase binding site then further substitution of sequences in site IIIR would not change the repression ratios observed for P_{R-96} or P_{R-322}.

CHAPTER 6

Further analysis of the nature of DNA sequences required for *res* function.

INTRODUCTION

As reported in Chapter 4, a substrate containing two res-96 sites in direct repeat was not observed to recombine in vivo, whereas res-96 x res-wt recombines reasonably well. Several other substrates were constructed and tested for their ability to recombine in vivo (Kitts et al 1983); the results are summarised in Figure 6.1. res-LH and res-RH are both products of a res-96 x res-wt resolution, and from the results of Chapter 4 the structures of each res site in terms of resolvase binding sites is known. res-96 contains a reduced affinity binding site I, wild-type site II and no site III for resolvase binding in vitro. res-LH contains wild-type sites I and II, and no site III. res-RH contains the same reduced binding affinity binding site I as res-96, and wild-type sites II and III.

The truncated Tn1000 res site contained in pRW12, referred to in Chapter 4, was also tested in vivo for recombination (Grindley et al 1982). This shows similar properties to those of res-96 although containing wild-type site I and reduced affinity binding site II.

The conclusion from this data is that transposon-encoded site-specific recombination can involve unequal partners. That is, a primary site, res-wt, can recombine with a secondary site such as res-96, but two secondary sites cannot recombine. The basic difference between primary and secondary sites is the presence of site III in the former and its absence in the latter.

The efficiency of recombination involving a primary and secondary site of the ones tested is reduced compared to recombination involving two res-wt sites. Is this reduced efficiency derived from loss of site

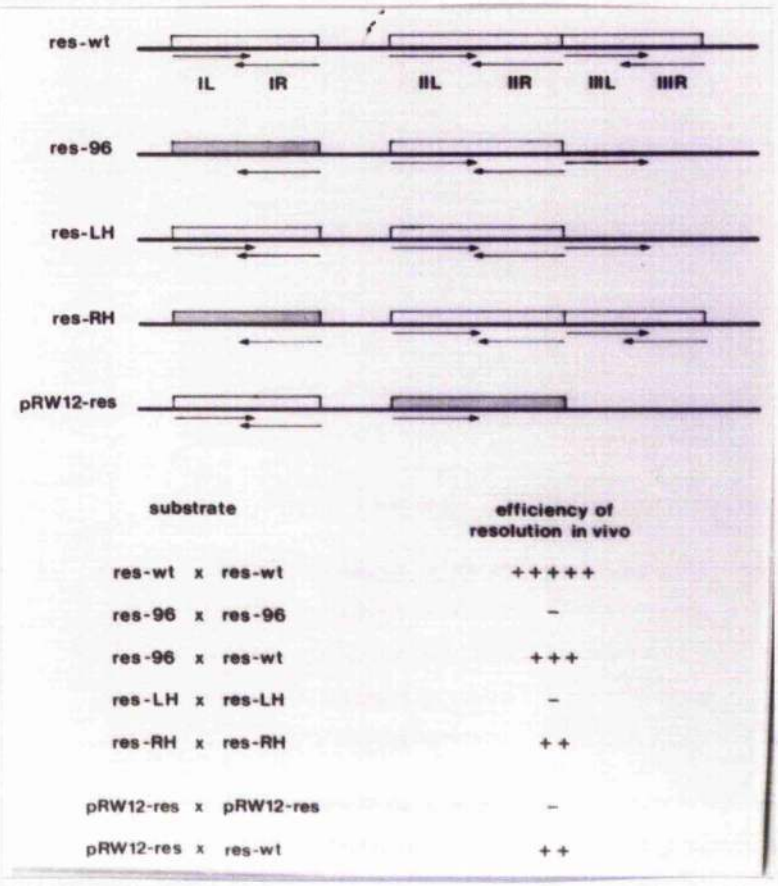


Figure 6.1 **Structures and efficiencies of various recombinant *res* regions.**

Resolvase binding sites are indicated by boxes; shading denotes reduced affinity binding. Arrows indicate the positions of half-sites.

Resolution efficiencies are derived from Kitts *et al* 1983 for Tn3 *res*, and Grindley *et al* 1982 for Tn1000 *res*.

III *per se* in the secondary site, or is it a reflection of the reduced binding affinity of resident binding sites in the secondary site? Moreover what properties does site III confer on a *res* site to give it the characteristics of being a primary site in recombination?

RESULTS

6.1 The efficiency of primary site x secodary site recombination.

Further confirmation of the effect of the reduced binding affinity of resolvase for site I of *res*-96 could be gained by comparing resolution efficiencies of similar substrate molecules differing only in the site I in question. To this end four substrates were constructed: pairs of similar molecules containing *res*-96 x *res*-wt to compare with *res*-wt x *res*-wt, and *res*-LH x *res*-wt to compare with *res*-wt x *res*-wt.

For the first pair the plasmid pJD118, containing the *Hin*cII-*Bam*HI *res*-wt site, was used to clone in adjacent *res* regions. pJD120 was constructed by inserting a 358bp *Sau*3A fragment containing *res*-wt from Tn3. The fragment was isolated from an acrylamide gel fragment, and ligated with *Bam*HI linearised pJD118. Recombinants were enriched for by cutting the ligated DNA with *Bam*HI before using it to transform with; from the known sequence it was possible to predict that the *Bam*HI-*Sau*3A junctions of the recombinants would not now contain a *Bam*HI recognition sequence. Transformants were screened by SCCL to check plasmid sizes, and likely candidates were further checked by restriction mapping. The two recombinants containing the *Sau*3A fragment in either orientation, were isolated in this manner. The

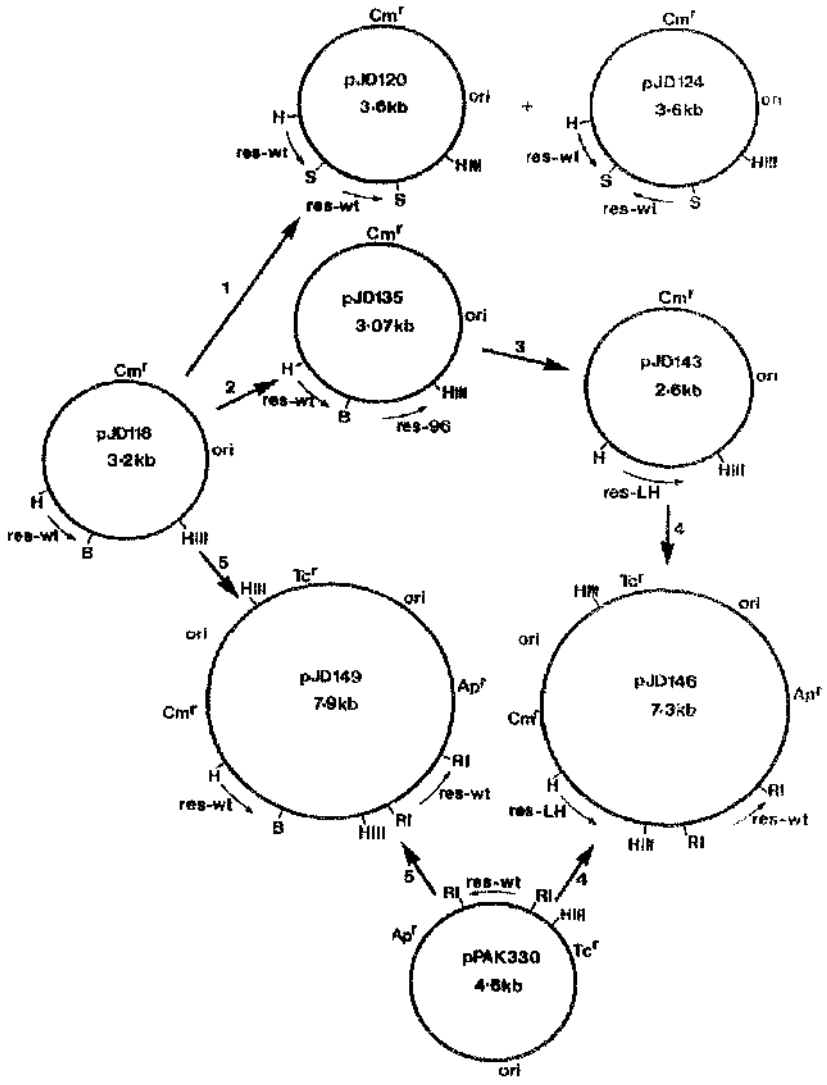


Figure 6.2 **Structure and derivation of substrate plasmids.**

H = HincII S = Sau3A HIII = HindIII

B = BamHI RI = EcoRI

Steps in construction were:

1. Cloning the 358bp Sau3A fragment into the BamHI site of pJD118.

2. Replacing the 300bp BamHI-HindIII fragment of pJD118 with a similar 120bp fragment from pJD131.

3. Resolution of pJD135.

4. Fusion of HindIII linearised pJD143 and pPAK330.

5. Fusion of HindIII linearised pJD118 and pPAK330.

3.6kb plasmid pJD120 contains two copies of res-wt in direct repeat separated by 650bp (Fig. 6.2). pJD124 contains the Sau3A fragment in the opposite orientation and was used to monitor resolvase mediated inversion (see Chapter 7).

To construct the res-wt x res-96 substrate, pJD118 was cut with both HindIII and BamHI, and ligated with pJD133 cut in a similar way. Recombinants where the 300bp HindIII-BamHI fragment was replaced by a 120bp fragment from pJD131 were screened for size by SCCL, and the structure confirmed by restriction mapping. The 3.07kb plasmid pJD135 contains directly repeated copies of res-wt and res-96 separated by 510bp (Fig. 6.2).

The product of resolution of pJD135 contains a single res region: res-LH. This plasmid, pJD143, was cut with HindIII and ligated with HindIII linearised pPAK330, containing res-wt. $Tc^rAp^rCm^r$ transformants were selected for and subsequently screened by SCCL for recombinant plasmids. The phenomenon of palindromic lethality directs the orientation of the plasmid fusion. The structure of the recombinant plasmid, pJD146, was confirmed by restriction mapping. It is a 7.3kb plasmid carrying res-wt and res-LH in direct repeat, separated by 385bp (Fig. 6.2).

In a similar way a fusion of pJD118 and pPAK330 was constructed. This 7.9kb plasmid, pJD149, contains direct repeats of res-wt separated by 1kb (Fig. 6.2).

These pairs of plasmids were tested for resolution efficiency both *in vivo* and *in vitro*. For the *in vivo* experiments resolvase was supplied *in trans* by (a) the low copy number plasmid R388::Tn3, and (b) the high copy number plasmid pDS4153 in AB2463. The test stains

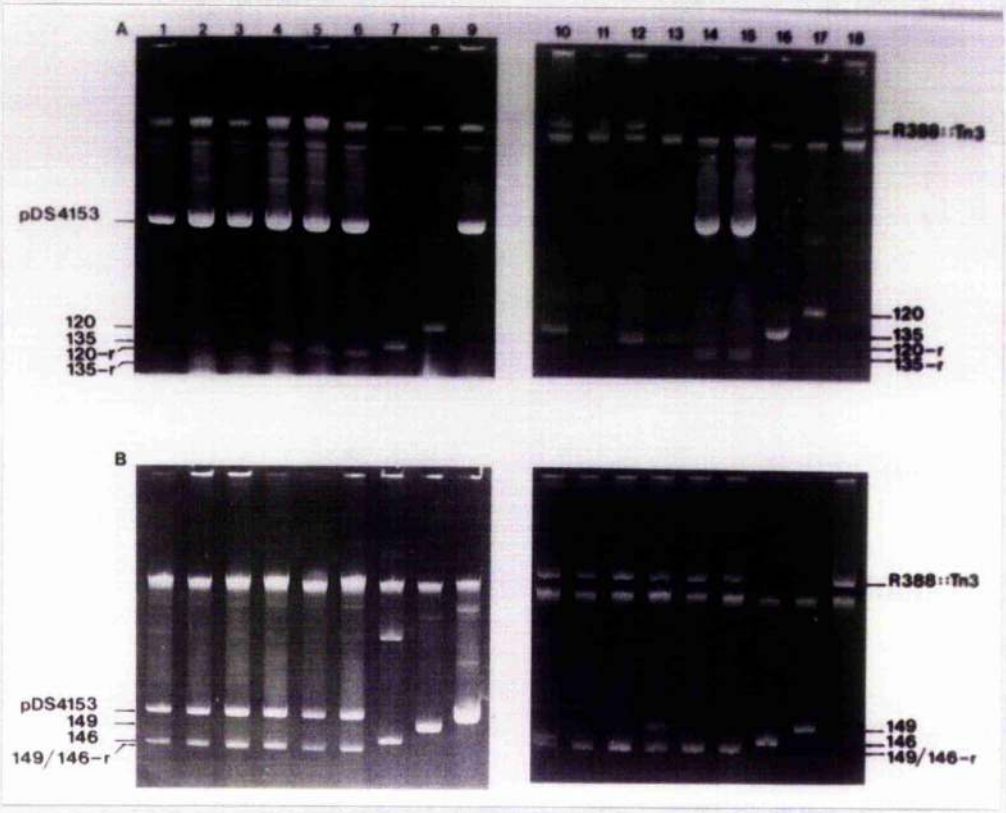
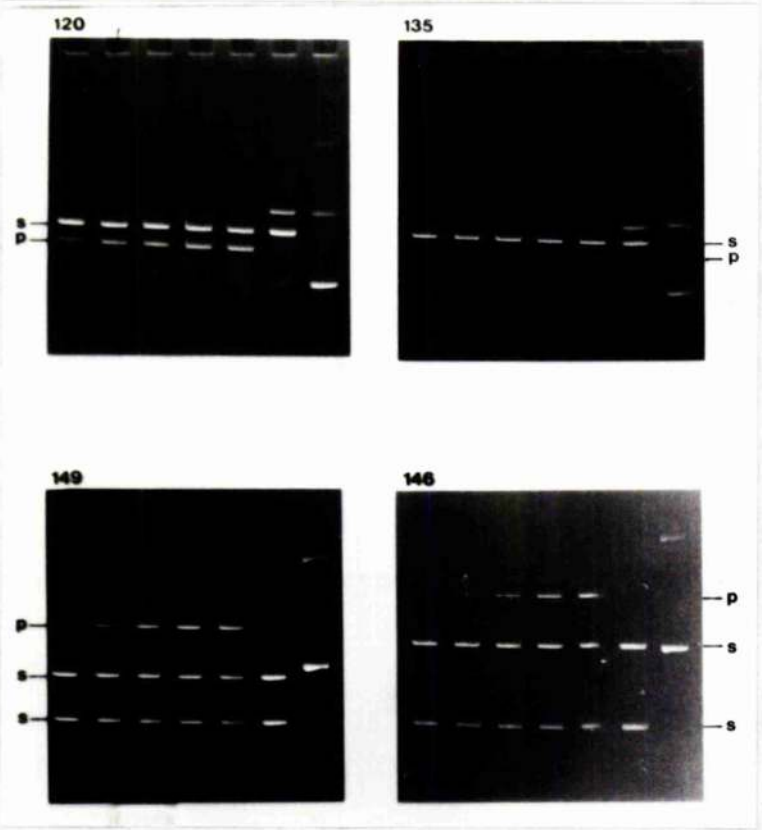


Figure 6.3 **In vivo resolution assays to compare secondary site activity.**
0.8% agarose gels
Substrates labelled as, eg. 135 = pJD135, 135-r = pJD135 resolution product.

	Gel A	Gel B
lanes 1to3	pJD135+pDS4153, gen.35	pJD146+pDS4153, gen.35
lanes 4to6	pJD120+pDS4153, gen.35	pJD149+pDS4153, gen.35
lane 7	pJD135	pJD146
lane 8	pJD120	pJD149
lane 9	pDS4153	pDS4153
lane 10	pJD135+R388::Tn3,gen.25	pJD146+R388::Tn3,gen.25
lane 11	pJD135+R388::Tn3,gen.35	pJD146+R388::Tn3,gen.35
lane 12	pJD120+R388::Tn3,gen.25	pJD146+R388::Tn3,gen.35
lane 13	pJD120+R388::Tn3,gen.35	pJD149+R388::Tn3,gen.25
lanes 14&15	pJD135+pDS4153, gen.45	pJD149+R388::Tn3,gen.35
lane 16	pJD135	pJD146
lane 17	pJD120	pJD149
lane 18	R388::Tn3	R388::Tn3

were constructed in each case by introducing the substrate into the tnpR$^+$ background by transformation. Plasmids were subsequently visualised by SCCL (Fig. 6.3). Complete resolution was observed for res-wt x res-wt and res-wt x res-96 within 35 generations when complemented for resolvase by either the low or high copy number plasmid. Within 25 generations more pJD149 was converted than pJD146 when complemented by R388::Tn3. The res-wt x res-96 substrate, pJD135, shows approximately 40% resolution after 35 generations when complemented by pDS4153. So the observed resolution efficiencies *in vivo* are res-wt x res-wt > res-wt x res-LH > res-wt x res-96.

The effects of competition between resolved and unresolved plasmids within bacterial populations might in fact bias the observed efficiencies of recombination for these substrates. So *in vitro* analysis may give a better idea of the relative efficiencies of recombination for each substrate. For the *in vitro* reactions, identical concentrations of DNA, resolvase and buffers were used. A time course of reactions were carried out. To visualize the difference between reacted (supercoiled catenanes) and unreacted supercoiled molecules, each reaction was subsequently digested with a restriction enzyme before gel electrophoresis and staining with EtBr. The pJD120/pJD135 pair were cut with HindIII, and the pJD146/pJD149 pair were cut with ClaI (Fig. 6.4). The relative amounts of DNA in each band were assessed by densitometry, and these figures normalised to correct for the size differences of each linear fragment. To do this a linear ratio for size of fragment and EtBr incorporated was assumed. The results were plotted on graphs: percentage of resolved product against time (Fig. 6.4).



facing page 124

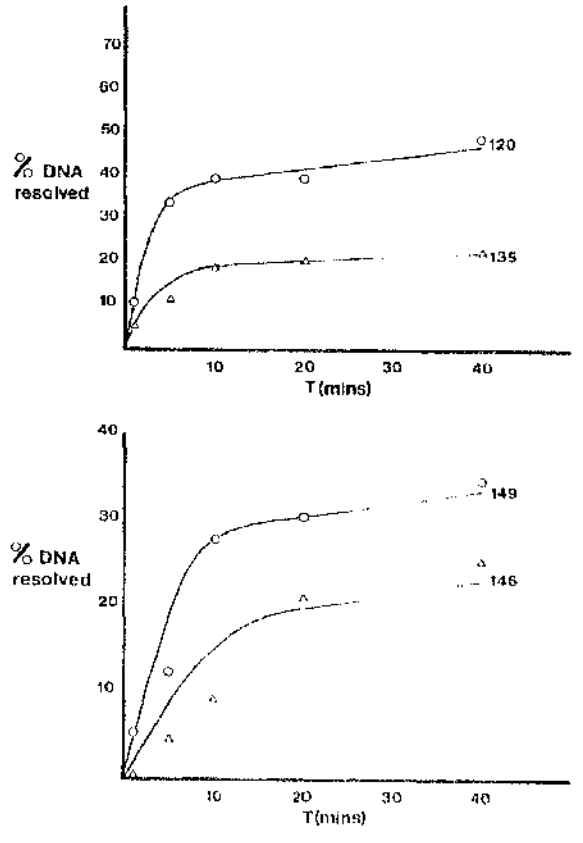


Figure 6.4 **<u>In vitro</u> resolution assays to compare secondary site activity.**

1% agarose gels

s = resolution substrate band 120 = pJD120 149 = pJD149

p = resolution product band 135 = pJD135 146 = pJD146

For each gel, lanes represent from left to right substrate DNA incubated with resolvase, heat-shocked and incubated with a restriction enzyme: 1. + resolvase for 1min; 2. + resolvase for 5min; 3. + resolvase for 10min; 4. + resolvase for 20min; 5. + resolvase for 40min; 6. -resolvase +restriction enzyme; 7. -resolvase -restriction enzyme.

Graphs plot the percentage DNA resolved against time for each substrate.

From this data it is concluded that *in vitro*, the efficiency of resolution of *res*-LH x *res*-wt is approximately 70% that of *res*-wt x *res*-wt. The efficiency of resolution of *res*-96 x *res*-wt is 50% that of *res*-wt x *res*-wt.

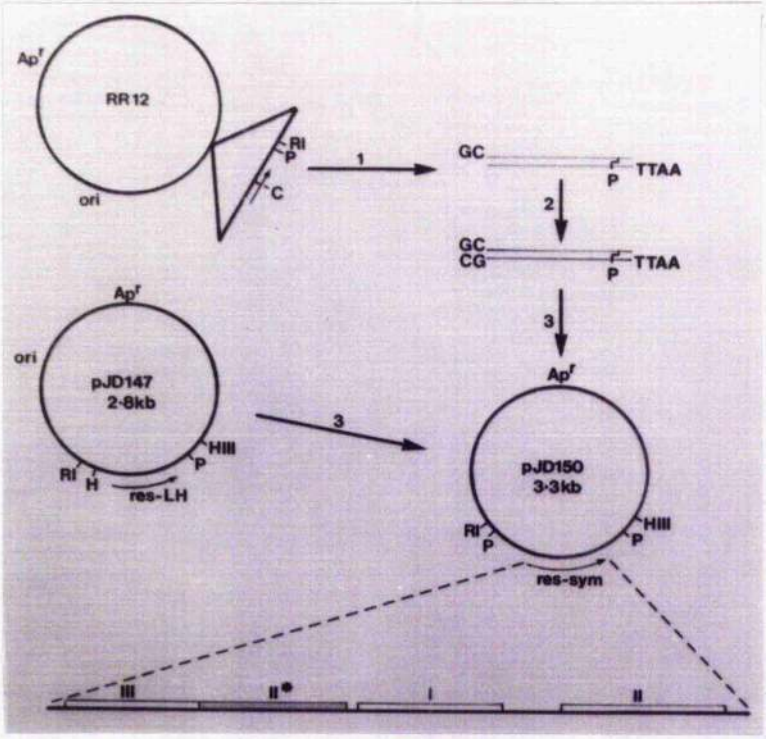
6.2 The readdition of site III to *res*-LH: a symmetrical *res*.

It is apparent from the above data that site III is important for primary site function, while *res*-LH 'functions well' as a secondary site without site III. What is it about site III that makes it so crucial? The effect could simply be a result of the quantity of resolvase molecules required at the *res* region; only, for example, when *res* is bound by three times the number of resolvase molecules required to bind at one site does the whole complex become 'primed' to initiate the mechanisms that bring about recombination. Moreover the actual arrangement of individual binding sites relative to each other may be an important aspect. For *res*-wt the arrangement of sites is assymetrical with respect to site I, the crossover site. This arrangement might facilitate those mechanisms involved in recombination. Resolvase, like many other proteins, forms multimers (Symington 1982); the 'communication' between resolvase bound at each site may be essential. This would arise from multimeric quaternary conformations involving resolvase-resolvase interactions which thereby change the local spatial arrangement of the *res* DNA. This would predict that a disruption of the organisation of resolvase binding sites, whilst conserving the sequences themselves and the potential for resolvase to bind to that DNA, may grossly affect the recombinational potential of the complex.

To test out some of these ideas a Tn1000-Tn1 hybrid region containing all three binding sites, but in the order III-I-II was constructed. The choice of this particular configuration was made due to the relative ease of its construction compared to that of other alternative arrangements. This is largely due to the shortage of restriction sites within *res*; a more systematic approach to reorganising *res* might make use of DNA synthesising facilities.

A pUC8 derivative containing *res*-LH was constructed by replacing the 12bp *Hinc*II-*Hind*III fragment within the polylinker with the 193bp *Hinc*II-*Hind*III fragment from pJD143. White colonies on XGal plates were screened by SCCL to check their size, and recombinants checked by restriction to confirm the presence of the 193bp fragment. The recombinant plasmid was called pJD147.

The plasmid RR12 (Reed 1981) is a pBR322::Tn1000 derivative. It contains an *Eco*RI linker inserted at a *Hae*III site within the *tnp*R gene (at the equivalent Tn3 coordinate 3551), and the *Cla*I site in the centre of site II (Tn3 coordinate 3152). RR12 DNA was digested with *Eco*RI and *Cla*I, and the *Cla*I ends of the total digest were repaired with DNA polymerase large fragment and dCTP and dGTP. Meanwhile pJD147 was cut with *Eco*RI, *Hinc*II and *Bam*HI (which cuts within the 15bp *Eco*RI-*Hinc*II fragment). The two restricted plasmid DNA's were then ligated and subsequently restricted again with *Bam*HI and *Hinc*II to enrich for the correct recombinants prior to transformation. Transformants were analyzed by SCCL to check plasmid sizes. In this way the recombinant plasmid, pJD150, was obtained and subsequently analysed by restriction (Fig. 6.5). Both the predicted sizes of restriction fragments which are observed and the loss of the *Hinc*II



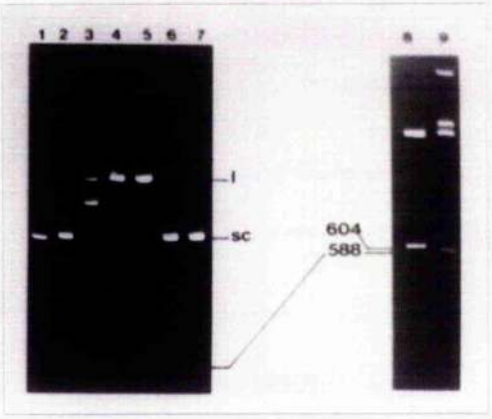


Figure 6.5 **Construction of res-sym.**
H = HincII C = ClaI P = PstI
RI = EcoRI HIII = HindIII
1. RR12 was digested with ClaI and EcoRI.
2. The ClaI 'sticky ends' were repaired with Klenow fragment, dCTP and dGTP.
3. pJD147 was digested with EcoRI and HindIII, and recombinants were selected which contained the repaired ClaI-EcoRI fragment from RR12.

Restriction analysis confirmed the structure of pJD150: lanes 1 to 7: 1% agarose gel lanes 8 to 9: 3.5% acrylamide gel
(fragment sizes in bp)
lane 1=pJD150
lane 2=pJD150: ClaI digest
lane 3=pJD150: PstI digest
lane 4=pJD150: EcoRI digest
lane 5=pJD150: HindIII digest
lane 6=pJD150: BamHI digest
lane 7=pJD150: HincII digest
lane 8=pJD150: HindIII/EcoRI digest
lane 9=pJD150: PstI digest

restriction site in pJD150 suggest that the ClaI 'sticky' ends were faithfully filled in, and subsequently blunt-end ligated to the HincII site of pJD147. The structure of the reorganised res-site (res-sym) is presented in Figure 6.5. It is noted that site III is now inverted in respect to sites I and II; however the symmetrical feature of the binding site would predict that the nature of resolvase-res interactions at site III would be similar in either orientation. Moreover one half-site of site II (IIR) is present adjacent to site III. From data in the previous two chapters it is expected that resolvase would only bind at this position if the sequences to the right of IIR can substitute for IIL; in specific sequences substituting for the conserved TGT trinucleotide and adjacent sequences. There are in fact two candidates for a substituted IIL. The conserved sequences of the binding determinant and adjacent sequences are compared to these candidate sequences:

consensus	5'-TGTCYNNTA
candidate (1)	5'-TGTTCGATT
candidate (2)	5'-TGTGTTCGA

This presents the distinct possibility of resolvase binding at this site, termed II*; however it has not yet been possible to test this by DNA footprinting. It should be noted that footprint analysis with the truncated C-terminal fragment of resolvase shows that the affinity of this smallest fragment is smallest for the IIR half-site than for any of the other half-sites (Abdel-Meguid et al in Press). A combination of this low affinity to IIR and the sub-optimal sequences for binding in the substituted IIL might confer an overall reduced

affinity for site II*.

6.3 Construction of substrates to test res-sym function.

It was necessary to construct three substrates to test res-sym function: (1) res-sym x res-wt, (2) res-sym x res-sym, and (3) res-sym x res-LH.

(1) res-sym x res-wt.

This substrate was constructed by digesting both pPAK329 and pJD150 with PstI and ligating the restricted DNA. The ligated DNA was used to transform AB2463 to Tc^r and colonies were then tested for Ap^S; Tc^rAp^S colonies were subsequently analyzed by SCCL. Recombinants were isolated containing the PstI fragment containing res-sym inserted at the PstI site of the β-lactamase gene in both orientations, as shown by restriction analysis. In terms of orientation of res sites, the orientation of res-sym is referred to compared to res-wt, so that pJD151 contains res-sym in order of sites III-II*-I-II relative to res-wt: I-II-III. pJD152 contains res-sym in inverted orientation to res-wt (Fig. 6.6).

(2) res-sym x res-sym.

Initially the 588bp PstI fragment carrying res-sym was cloned into the PstI site of pBR322. This plasmid, pJD153, was cut with EcoRI and ligated with similarly cut pJD150. Ap^rTc^r transformants were selected and analyzed by SCCL. Two classes of recombinants were identified by restriction analysis. pJD154 contains directly repeated res-sym's; pJD155 contains res-sym's in inverted orientation (Fig. 6.6).

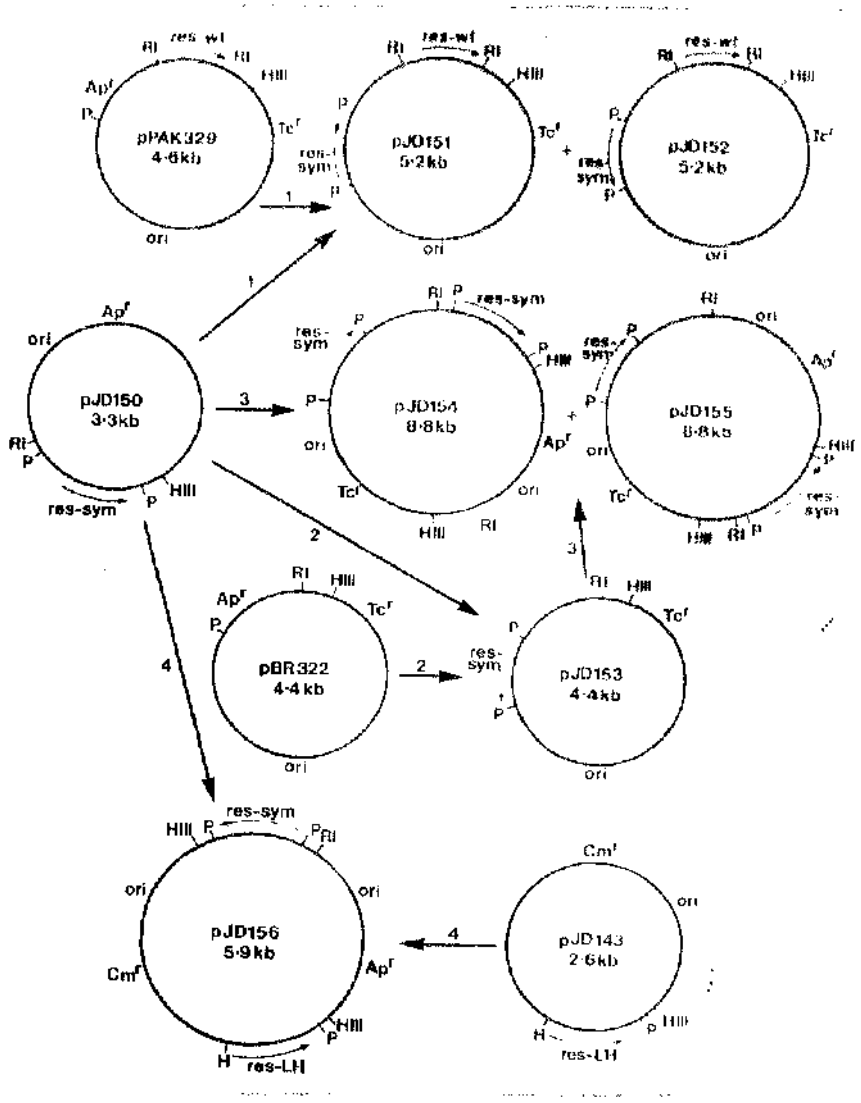


Figure 6.6 **Construction of substrates to test res-sym function.**

1. A 588bp PstI fragment from pJD150 was cloned in either orientation into pPAK329.

2. The same fragment was cloned into pBR322.

3. EcoRI cut linears of pJD153 and pJD150 were ligated together.

4. HindIII cut linears of pJD143 and pJD150 were ligated together.

(3) res-sym x res-LH.

For this construction pJD150 was cut with HindIII and ligated with pJD143 linearised with HindIII. Ap^rCm^r transformants were isolated. Palindromic lethality allows only one orientation of recombinant plasmid to be isolated: pJD156 containing res-sym and res-LH in direct repeat (Fig. 6.6).

6.4 Testing res-sym function in vivo and in vitro.

Initially plasmids pJD151, pJD152, pJD154, pJD155 and pJD156 were introduced into the strain AB2463 containing the $tnpR^+$ plasmid pDS4153 or pAA33. These test strains were analyzed by SCCL after 35 generations growth. For the three substrates containing res sites in direct repeat resolution was observed, but with differing efficiencies. For those substrates containing res sites in inverted orientation no deletion was observed; these substrates were analyzed for resolvase-mediated inversion as reported in Chapter 7.

Within 35 generations pJD151 (res-wt x res-sym) was observed to resolve completely; pJD154 (res-sym x res-sym) gave 25% resolution; pJD156 (res-sym x res-LH) gave 20% resolution (Fig. 6.7).

These substrates were subsequently tested for resolution in vitro. pJD151, pJD154 and pJD156 were incubated with resolvase for 40min, then digested with HindIII to separate resolved catenates from unreacted supercoils. Approximately 50% of pJD151 recombined in vitro, in marked contrast to the complete absence of recombination products for pJD154 and pJD156 (Fig. 6.7).

The lack of observable recombination in vitro for res-sym x res-sym and res-sym x res-LH substrates could be explained if the in vivo

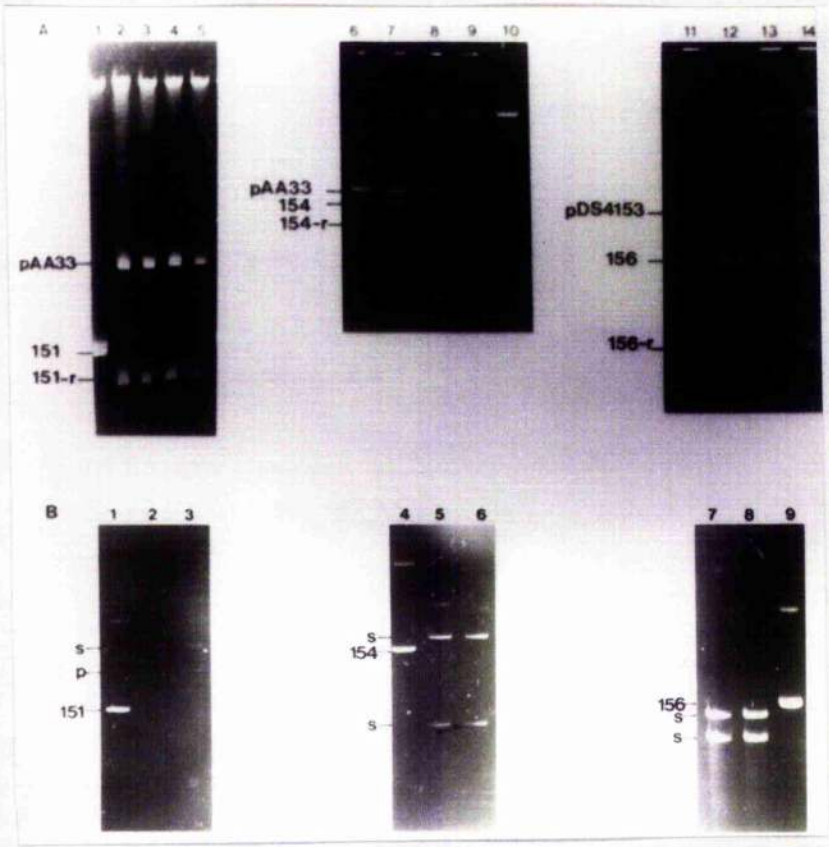


Figure 6.7 **In vivo and in vitro resolution assays to test res-sym function.**
A: 0.8% agarose gels
Substrate plasmids identified as, eg. 151 = pJD151,
151-r = in vivo resolution product.
lane 1 = pJD151 lanes 2 to 5 = pJD151+pAA33
lanes 6 to 8 = pJD154+pAA33 lane 9 = pJD154
lane 10 = pAA33 lane 11 = pDS4153
lane 12 = pJD156
lanes 13 & 14 = pJD156+pDS4153

B: 1% agarose gels
s = substrate bands p = in vitro product bands
lane 1 = pJD151 lane 2 = pJD151: HindIII digest
lane 3 = pJD151 + resolvase: HindIII digest
lane 4 = pJD154 lane 5 = pJD154: HindIII digest
lane 6 = pJD154 + resolvase: HindIII digest
lane 7 = pJD156 + resolvase: HindIII digest
lane 8 = pJD156: HindIII digest
lane 9 = pJD156

level of recombination was sufficiently low that the *in vitro* efficiency would result in recombination being relatively undetectable. The rate of recombination *in vivo* of the *res*-wt x *res*-96 substrate is similar to that of both pJD154 and pJD156, and the former could easily be detected *in vitro* with production of approximately 25% resolved product after 40min incubation with resolvase. So by comparison, *in vitro* resolution of pJD154 and pJD156 should be detectable.

An alternative explanation for this observation is that for those substrates where *res*-sym acts as a primary site, another component, probably host-encoded, is required for recombination. In the case of *res*-wt x *res*-sym, the wild-type site would be conceived as dominant over *res*-sym in terms of primary site function, so that there would be no requirement for this extra component.

To test this idea an obvious step was to assay these substrates for recombination in a variety of strains containing host mutations until the relevant strain which does not allow resolution was found. In the limited time available a short cut to testing recombination in a plethora of strains was to look for a precedent of host-encoded functions required for site-specific recombination. In fact there is only one precedent as yet: lambda integrative and excisive site-specific recombination reqires IHF, integration host factor, which is comprised of two subunits, one encoded by the host *him*A gene, the other by the host *him*D gene (alternatively called *hip*; reviewed by Weisberg and Landy 1983).

An isogenic set of strains were initially used to test recombination: K37 wild-type, K5175 *him*A$^-$ due to a Tn10 insertion

linked to a lesion in himA, K5428 himD63 (a base-pair substitution), K5482 himD63 himA42 (a base-pair substitution; Miller 1981; Hoyt et al 1982). The tnpR$^+$ plasmid pDS4153, plus the substrate plasmid pJD156 were introduced into these strains. Transformants were analyzed by SCCL after 35 generations (Fig. 6.8). The res-sym x res-LH substrate resolves to give approximately 20% product in K37 and in K5428. A slightly reduced proportion of product is observed in K5482, whilst in K5175 no product was produced. This result suggested a role for HimA in res-sym recombination. The difference in recombination levels observed in K5175 and K5482 could be due to the point mutation in himA in the latter being leaky; this leakiness has been observed in lambda studies with himA point mutations (P. Kitts pers. comm.).

HimA has a role not only in integrative and excisive recombination, but also in regulation of lambda genes. Efficient translation of CII mRNA in vivo requires the HimA protein (Hoyt et al 1982). Moreover IHF binds specifically to lambda DNA in a region between the t_{RI} termination site and the P_{RE} transcription start site, although little is known about the mechanism of IHF action on repressor synthesis (Wulff and Rosenberg 1983). So it was important to ascertain if resolvase synthesis was affected by the Tn10 insertion in the himA gene of KS175. To test this a res-wt x res-wt plasmid, pJD120, was tested for resolution (Fig. 6.8). Efficient resolution was observed in each strain.

The level of resolution of a res-wt x res-96 substrate was shown to be directly related to the copy number of the tnpR$^+$ complementing plasmid and hence the resolvase concentration within the cell (Kitts 1982). If resolvase synthesis in a himA$^-$ strain was reduced, but in

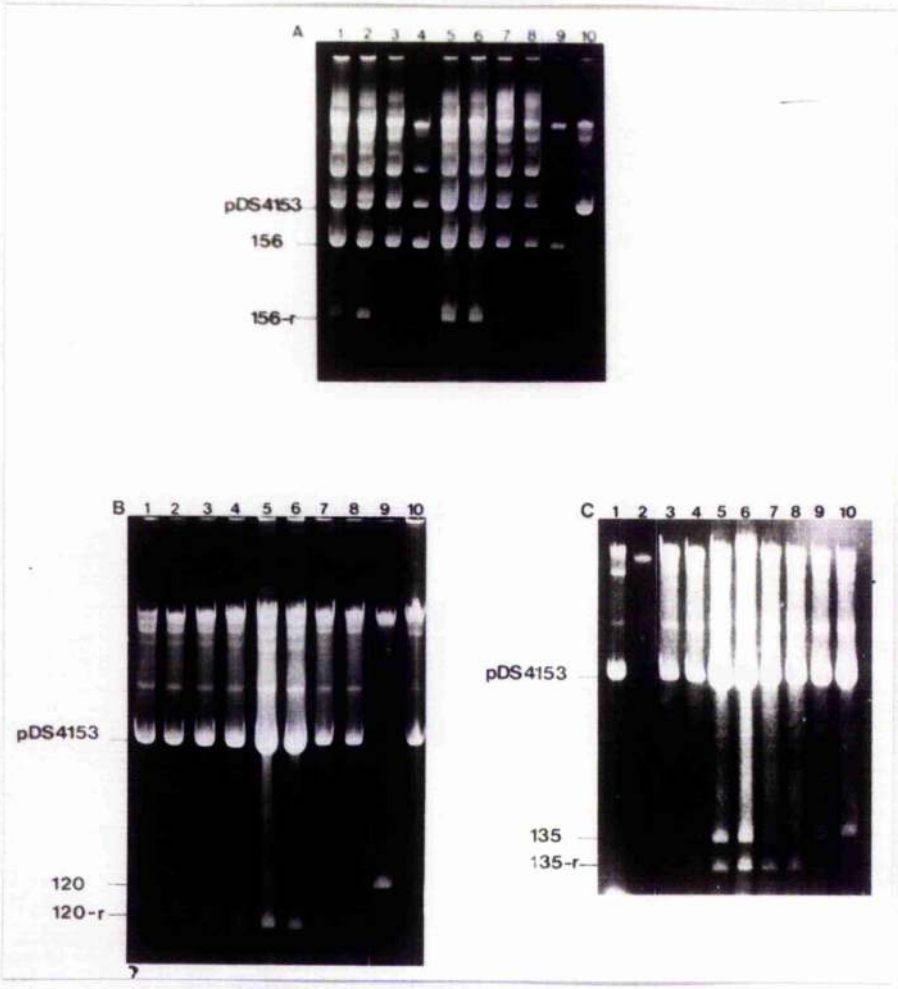


Figure 6.8 **Testing <u>res</u> function in <u>him</u>A and <u>him</u>D strains.**
0.8% agarose gels SCCL after 35 generations

	Gel A	Gel B
lanes 1&2:	pJD156+pDS4153 in K37	pJD120+pDS4153 in K37
lanes 3&4:	pJD156+pDS4153 in K5175	pJD120+pDS4153 in K5175
lanes 5&6:	pJD156+pDS4153 in K5428	pJD120+pDS4153 in K5428
lanes 7&8:	pJD156+pDS4153 in K5482	pJD120+pDS4153 in K5482
lane 9:	pJD156	pJD120
lane 10:	pDS4153	pDS4153

	Gel C
lane 1:	pDS4153
lane 2:	pJD135
lanes 3&4:	pJD135+pDS4153 in K5482
lanes 5&6:	pJD135+pDS4153 in K5428
lanes 7&8:	pJD135+pDS4153 in K5175
lanes 9&10:	pJD135+pDS4153 in K37

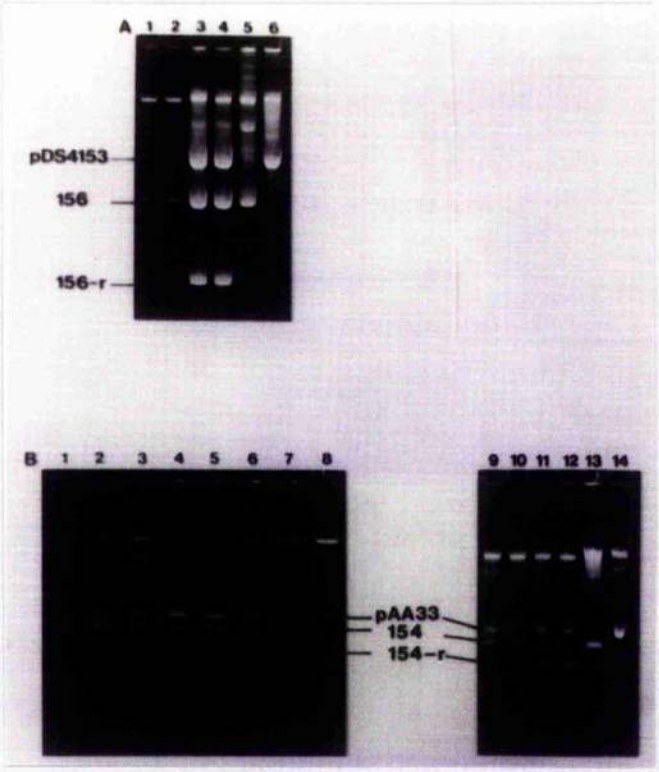


Figure 6.9 **Testing res-sym function in a himA background.**
0.8% agarose gels

Gel A:

lanes 1&2:	pJD156+pDS4153 in K5175 after 85 generations
lanes 3&4:	pJD156+pDS4153 in K37 after 85 generations
lane 5:	pJD156 in K37 after 85 generations
lane 6:	pDS4153

Gel B:

lanes 1to3:	pJD154+pAA33 in K5175 after 35 generations
lanes 4to6:	pJD154+pAA33 in K37 after 35 generations
lane 7:	pJD154 in K37 after 35 generations
lane 8:	pAA33
lanes 9&10:	pJD154+pAA33 in K5175 after 85 generations
lanes 11&12:	pJD154+pAA33 in K37 after 85 generations
lane 13:	pJD154 in K37 after 85 generations
lane 14:	pAA33

sufficient amount to allow pJD120 resolution but not pJD156 resolution, then the level of resolution of a *res*-wt x *res*-96 substrate might be reduced compared to that in the wild-type strain. This was tested by introducing pJD135 into each strain to observe resolution. Resolution was observed in each strain, but the ratio of product to substrate was slightly different in the K5175 strain - if anything greater than in the wild-type K37, K5428 or K5482 (Fig. 6.8). This observation suggests that resolvase synthesis is not limiting in a *him*A$^-$ strain.

pJD156 was subcultured for 85 generations in the *tnp*R$^+$ *him*A$^-$ strain and no resolution was observed. Moreover the *res*-sym x *res*-sym substrate was tested for recombination in both K37 containing pAA33 and K5175 containing pAA33; no resolution was observed (Fig.6.9) after 35 generations. A small amount of product was observed after 85 generations subculture.

DISCUSSION

The bulk of points raised in this Chapter are discussed in detail in Chapter 8.

The differences in resolution efficiencies for *res*wt x *res*96 and *res*-wt x *res*-LH (approximately 20%) can be attributed to the reduced binding affinity of resolvase for *res*-96 site I. This difference in resolution efficiency correlates well with both the reduced binding affinity observed in the DNA footprintig experiments, and the decrease in resolvase mediated repression discussed in Chapter 5. The resolution efficiency of *res*-wt x *res*-LH is 70% that of an equivalent *res*-wt x *res*-wt substrate. This reduction in efficiency can be

attributed to the absence of site III in res-LH. Without site III there are at least two ways by which the recombination efficiency may be reduced:

(i) The formation of synapse ultimately depends on juxtaposition of res sites mediated by resolvase. For res-LH there are just 66% of the resolvase binding sites present in res-wt. If the juxtaposition mechanism depends on resolvase bound at one res region recognising resolvase bound to the other region, or simply the unbound res DNA of the other region, then res-LH will offer in either case only 66% of the 'attractive features' that a res-wt might offer.

(ii) The structure of a resolvase-res-LH complex may be 'non-receptive'. If resolvase molecules bound at adjacent sites form multimeric quaternary structures, altering the local DNA conformation, then the overall structure formed this way at res-LH may be less recognizable by the resolvase-DNA complex formed at the primary res-wt region than a quaternary structure formed at a similar res-wt region.

Addition of site III to res-LH changes the region from having secondary site function to primary site function; this is shown by the ability of a res-sym x res-LH substrate to recombine. Moreover the primary site function of res-sym is observed despite site III being at a different position to that in res-wt.

The function of res-sym is dependent on its orientation. Thus for a res-sym x res-sym substrate, resolution is observed only when the two regions are in direct repeat. A feature of an exactly symmetrical recombination region would be that inversion and deletion reactions would occur at similar frequencies, that is there would be no

intrinsic orientation to the region (Sherratt *et al* 1981). Palindromic lethality, however, probably prevents such symmetrical regions existing *in vivo*. The orientation of *res*-sym relative to other *res* sites maintains the orientation assumed by a similarly positioned *res*-wt, so that for directly repeated *res*-LH x *res*-sym the order of resolvase binding sites is *res*-LH: site I - site II, followed by *res*-sym: site III - site II* - site I - site II.

CHAPTER 7

Site-specific inversions mediated by resolvase.

INTRODUCTION

Intramolecular site-specific recombination between two sites will result in deletions or inversions depending on the orientation of the recombination sites relative to each other. For both phage lambda excision and Tn3 resolution the 'normal' substrate consists of direct repeats of the site. There are a number of closely related systems in prokaryotes where site-specific recombination acts on inverted sites to generate inversions of the intervening DNA. Two systems have been well studied: (1) inversion of the G-region of bacteriophage Mu (Kamp et al 1978; Van de Putte et al 1980) and the C-region of the related phage P1 (Chow and Bukhari 1976) and (2) phase variation in Salmonella typhimurium (Silverman and Simon 1980). The gin (Mu), cin (P1) and hin (S.typhimurium) gene products can complement each other in all combinations so far tested (Kutsukake and Iino 1980; Kamp and Kahmann 1981), suggesting a close evolutionary relationship between these inversion systems. More recently an invertible region in the E.coli chromosome has been analyzed. A function which can complement gin mutations was called pin, and it catalyses inversion of a 1600bp P-region (Plasterk et al 1983).

Chiang and Clowes (1980) reported that a molecule containing two copies of Tn2660 (a Tn3-like transposon) in inverted repeat did not show any evidence of resolvase-mediated inversion between the two inverted res regions. A prediction was made that resolvase mediated inversion would lead to establishment of equilibrium between two forms of such a molecule: one containing an internal region inverted relative to the other molecule (Arthur and Sherratt 1979). Moreover

the sequences of Hin and resolvase have been shown to contain 33% homology at the amino acid level (Simon et al 1980), suggesting that resolvase might be capable of promoting inversions.

RESULTS

7.1 Inversion substrates containing two res-wt's.

A substrate molecule containing two copies of Tn103 (tnpA$^+$tnpR$^-$ res$^+$Tn1) in inverted orientation was constructed in this laboratory by D.Sherratt (Fig. 7.1). Initially this substrate, pJD100A, was introduced into a rec$^-$ strain and complemented in trans for resolvase by the low copy number plasmid R388::Tn3. Plasmid DNA was isolated from this test strain and restricted to identify any products of inversion. No inversion could be detected (D. Sherratt pers. comm.).

To test whether inversion could be detected in the presence of a high copy number tnpR$^+$ plasmid, a strain was constructed containing pJD100A and pDS4153 in AB2463. This strain was subcultured for 100 generations, plasmid DNA isolated and restricted with HindIII which cuts asymmetrically within the plasmid. The presence of two new restriction fragments indicated that a proportion of pJD100A had inverted. Approximately 15% of pJD100A DNA was present in the inverted form. The strain was further subcultured and analyzed after 200 generations for inversion. Approximately 60:40 proportions of non-inverted and inverted forms were present. This DNA was used to transform AB2463 to Cm^{r}Ap^r and then the DNA was isolated from individual colonies. In this way the inverted form of pJD100A was isolated; termed pJD100B (Fig. 7.1).

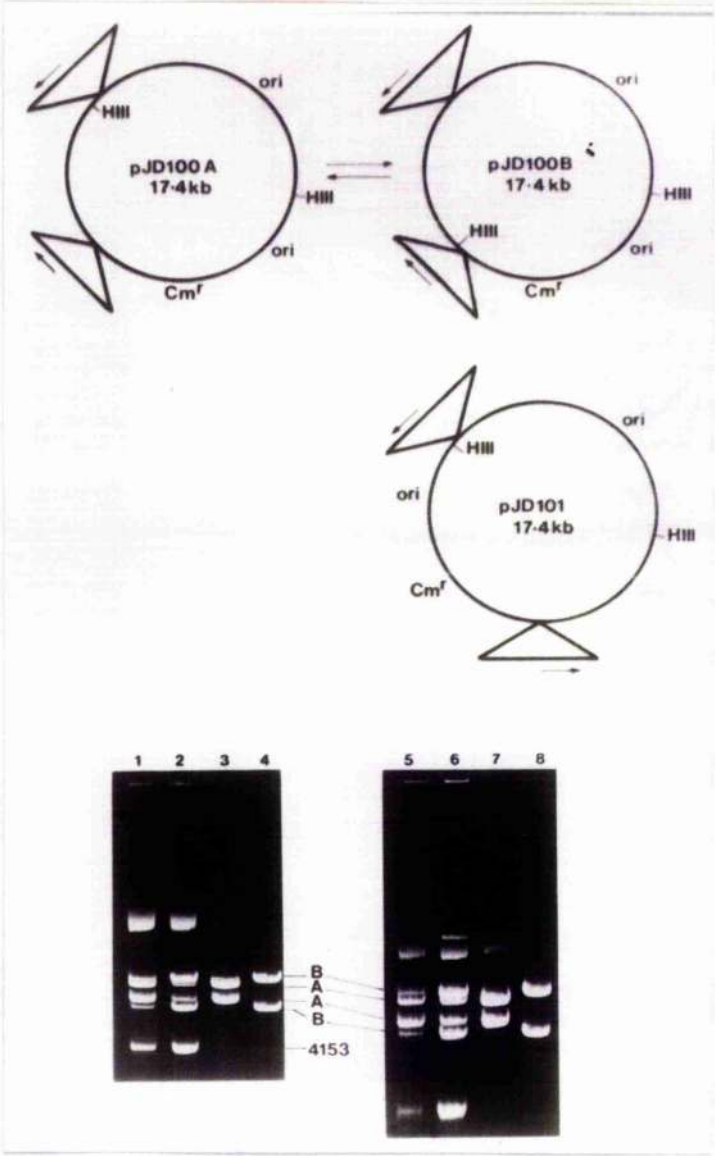


Figure 7.1 **Inversion analysis of pJD100.**

Inversion was monitored by restricting plasmid DNA with asymmetrically cutting HindIII (HIII): all lanes of the gels (1% agarose) show DNA digested with HindIII.

lane 1: pJD100A+pDS4153 gen.95 lane 2: pJD100B+pDS4153 gen.95

lane 3: pJD100A gen.95 lane 4: pJD100B gen.95

lane 5: pJD100A+pDS4153 gen.155 lane 6: pJD100B+pDS4153 gen.155

lane 7: pJD100A gen.155 lane 8: pJD100B gen.155

Subsequently inversion of both plasmid configurations was monitored when complemented by pDS4153. Approximately 15% of both forms had inverted within 95 generations subculture, and 30% within 155 generations (Fig. 7.1). Equilibrium between both forms was reached after 250-300 generations independent of the configuration of the starting plasmid.

A sister molecule, pJD101, containing both copies of Tn103 in direct repeat (Fig. 7.1) resolves totally within 25 generations (the shortest time to allow analysis of the DNA on a gel) when complemented by pDS4153. The time for complete resolution to occur will be dependent on the amount of available resolvase within the cell, which itself is dependent on the copy number of the complementing plasmid. It is reported that after infection with a lambda phage derivative carrying directly repeated *res* sites, over 90% resolution occurs after 40min (2-3 generations) exposure to a *tnpR*$^+$ cytoplasm (Muster *et al* 1983). In this situation pMB8::Tn3 was used to complement for resolvase, representing an initial approximate copy number ratio of 20:1 for *tnpR*$^+$ plasmid to substrate. In Chapter 6 it was reported that complete resolution of a high copy number substrate, when complemented by a low copy number *tnpR*$^+$ plasmid, was observed between 25 and 35 generations after the latter plasmid was introduced into the test strain. So it is probable that complete resolution of pJD101, when complemented by pDS4153, occurs within 10 generations.

From these figures estimates for both inversion rate and resolution rate can be made (see Appendix at the end of this Chapter for derivation). Assuming 10 generations for complete resolution of pJD101, then an approximate rate for resolution is 0.1 per cell per

generation. For inversion, for either pJD100A or pJD100B, the rate is 0.001 per cell per generation. Resolvase mediated inversion, therefore, would appear to be 100 times less efficient than resolution.

At this time it was reported that resolvase mediated inversion was in fact as efficient as resolution (Heffron et al 1981). This conclusion was derived from the observation that equilibrium for an invertible substrate could be established within approximately 50 generations. In order to reconcile this observation with the above result it was necessary to compare experimental procedures. The principal differences appeared to lie in the nature of the substrates used. Whereas pJD100A/B contain two copies of the major part of Tn1, Heffron et al used a substrate containing res sites divorced from the main body of the transposon: one res from Tn3, the second from Tn1000. Moreover this substrate contained res regions separated by only 1kb, in comparison to the 5kb separating the res sites of pJD100A/B.

To test if these differences could affect inversion rates, a second set of substrates were analyzed. The construction of pJD124A is described in Section 6.1. This plasmid contains the HincII-BamHI res-wt fragment and the 358bp Sau3A res-wt fragment in the opposite orientation, separated by 630bp (Fig. 7.2). In order to isolate the 'flipped' form of this plasmid it was introduced into AB2463 containing the high copy number tnpR$^+$ plasmid RSF1365. This strain was subcultured for 100 generations, plasmid DNA isolated and restricted with HaeII to look for inversion. Approximate equilibrium had been established in this period, and pJD124B was isolated by transforming AB2463 to Cm^r with this DNA, and analyzing the plasmid DNA content of

individual colonies.

pJD124A and pJD124B were reintroduced into AB2463 containing RSF1365. Plasmid DNA was isolated from these strains after 45, 75, 100, and 135 generations, and restricted with HaeII. Inversion of pJD124A occurs reasonably rapidly, reaching a 50:50 proportion with pJD124B at generation 100. However, two features of these invertible substrates are noticeable. Firstly, the 50:50 ratio of both forms of pJD124 present at generation 100 shifts towards pJD124B with subsequent subculturing. Secondly, even after 135 generations hardly any (<10%) of pJD124B has inverted to give pJD124A (Fig. 7.2).

Possible explanations for these observations are (a) the two Tn1/3 hybrid res sites which result from inversion of pJD124A are less functional in terms of inversion than one Tn1 res and one Tn3 res; (b) for some reason pJD124A is less competitive in a population, or within a cell, than pJD124B. The first explanation was tested by replacing the 358bp Sau3A fragment of pJD124A, derived from Tn3, with a similar fragment derived from Tn1. The resulting plasmid, pJD142A, and its flipped form pJD142B, were monitored for inversion. As observed before for the pJD124 substrates, pJD142A inversion proceeds more rapidly than for pJD142B (results not shown).

Plasmid competition within a bacterial population was examined by mixing equal volumes of stationary phase cultures of AB2463 containing on the one hand pJD124A, and on the other pJD124B. This mixed culture was subcultured and plasmid DNA isolated after 75 generations. This DNA, when restricted, showed that equal amounts of both plasmid were present within the population (Fig. 7.2), suggesting no competitive effects within the population. However this experiment does not

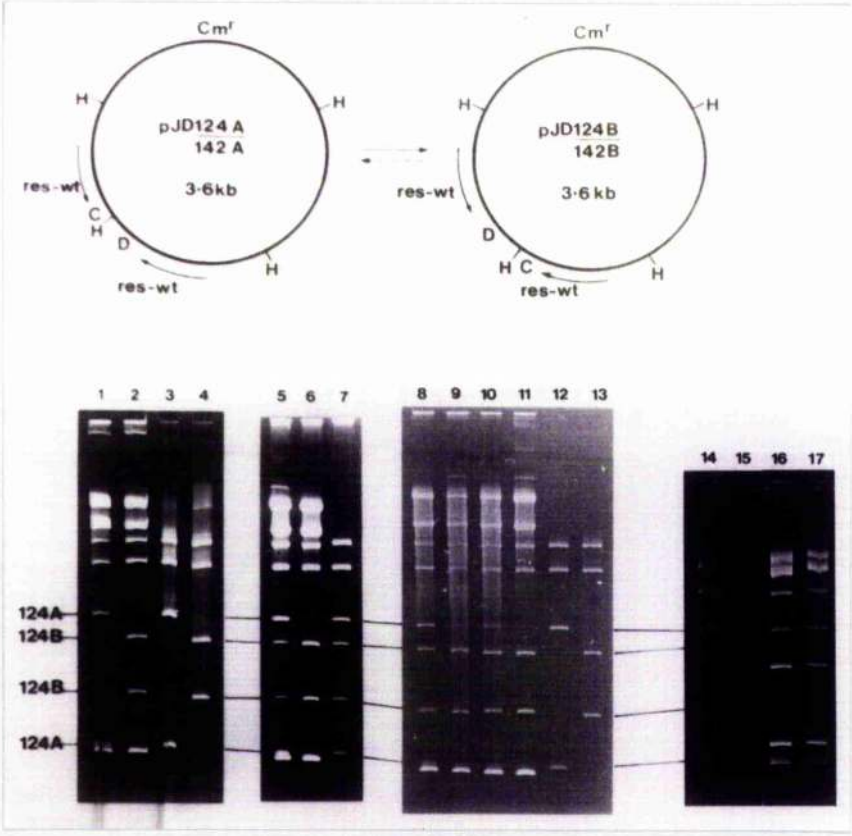


Figure 7.2 **Inversion analysis of pJD124.**

Inversion was monitored by restricting plasmid DNA with HaeII (H) which distinguishes the orientation of the intervening segment C/D. All lanes of the gels (3.5% acrylamide) show DNA previously incubated with HaeII.

lane1: pJD124A+RSF1365 gen.45 lane2: pJD124B+RSF1365 gen.45

lane3: pJD124A gen.45 lane4: pJD124B gen.45

lane5: pJD124A+RSF1365 gen.75 lane6: pJD124B+RSF1365 gen.75

lane7: pJD124A+pJD124B gen.75 lane8: pJD124A+RSF1365 gen.100

lane9: pJD124B+RSF1365 gen.100 lane10:pJD124A+RSF1365 gen.135

lane11:pJD124B+RSF1365 gen.135 lane12:pJD124A gen.135

lane13:pJD124B gen.135 lane14:pJD124A gen.525

lane15:pJD124B gen.525 lane16:pJD124A+pJZ110 gen.525

lane17:pJD124A+pJZ110 gen.525

preclude plasmid competition within individual cells giving rise to the biased distributions of inverted forms.

If it is assumed that the inversion rate of pJD124B to pJD124A is negligible, then an approximate inversion rate of pJD124A can be derived on the basis that complete inversion occurs within 300 generations. This would give a rate of 0.003 per cell per generation, which in turn would imply that inversion is approximately 30 times less efficient than resolution.

The effect of a bias in inversion rate could also be invoked for a separate construct designed to test resolvase complementation *in cis* on inversion. The plasmid pAA33* is a pACYC184 derivative containing a Tn3 insertion. This plasmid was digested with *Eco*RI and ligated together with pPAK329 digested with *Eco*RI. $Cm^{s}Tc^{r}Ap^{r}$ transformants were isolated, and their plasmids screened for size and then checked by restriction. In this way a substrate containing a *res*-wt (Tn3) and an *Eco*RI *res*-wt fragment in opposite orientation was isolated (Fig. 7.3). This plasmid, pJD134, was subcultured in AB2463 and DNA isolated after successive generations. The same residual amount of inverted plasmid was present after 250 generations as was present at 50 generations (Fig. 7.3). Perhaps plasmid competition effects can explain this observation too; in this case pJD134A would be more competitive than its flipped form.

The possibility of Hin-mediated inversion on a substrate containing two inverted *res*-wt's was tested. The plasmid pJZ110 is a pBR322 derivative carrying the entire phase variation system cloned from the *S.typhimurium* genome (Zieg *et al* 1978). This plasmid was introduced, together with pJD124A, into AB2463. Even after 525

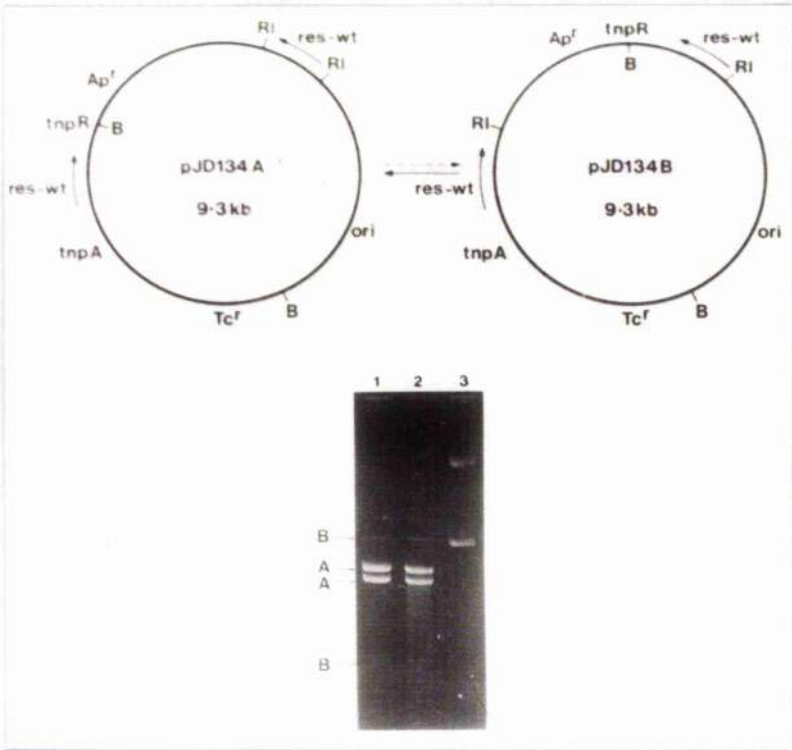


Figure 7.3 **Inversion analysis of pJD134.**

RI=EcoRI B=BamHI

Inversion was monitored by restriction with BamHI which cuts asymmetrically within the plasmid.

1.2% agarose gel

lane1: pJD134 gen.50, BamHI digestion

lane2: pJD134 gen.250, BamHI digestion

lane3: pJD134

generations there was no evidence of pJD124B form DNA (Fig. 7.2), indicating a complete absence of interchangeability between resolvase and Hin. Hin is unable to resolve a substrate containing direct repeats of res-wt (results not shown).

7.2 Inversion substrates containing secondary res sites, and res-sym.

To construct a substrate containing res-wt and res-96 in inverted orientation, pJD118 was cut with both HindIII and BamHI, and ligated with pJD131 cut in a similar way. Recombinants were screened by size and then by restriction mapping. pJD136 is a 3.07kb plasmid containing the HincII-BamHI res-wt and res-96 in opposite orientation, separated by 638bp (Fig. 7.4).

pJD145, containing inverted res-wt and res-LH, was constructed by ligating HindIII linearised pJD143 and pPAK329. AB2463 was transformed to $Tc^rAp^rCm^r$; transformants were screened and checked by restriction. Palindromic lethality predicts the relative orientation in which the original linears will be found in the recombinants (Fig. 7.4).

AB2463 containing pJD136 and the high copy number $tnpR^+$ plasmid RSF1365 was subcultured for 120 generations. AB2463 containing pJD145 and pDS4153 was subcultured for 135 generations. No inversion was evident in this period for either substrate (Fig. 7.4).

The constructions of pJD152, a substrate containing inverted res-sym and res-wt, and pJD155, containing two copies of res-sym in inverted orientation, are described in Section 6.3. Both plasmids were isolated after 50 generations subculture in AB2463 containing pAA33 and restricted to check for any inversion products (Fig. 7.5). No inversion was evident for either substrate, in marked contrast to the

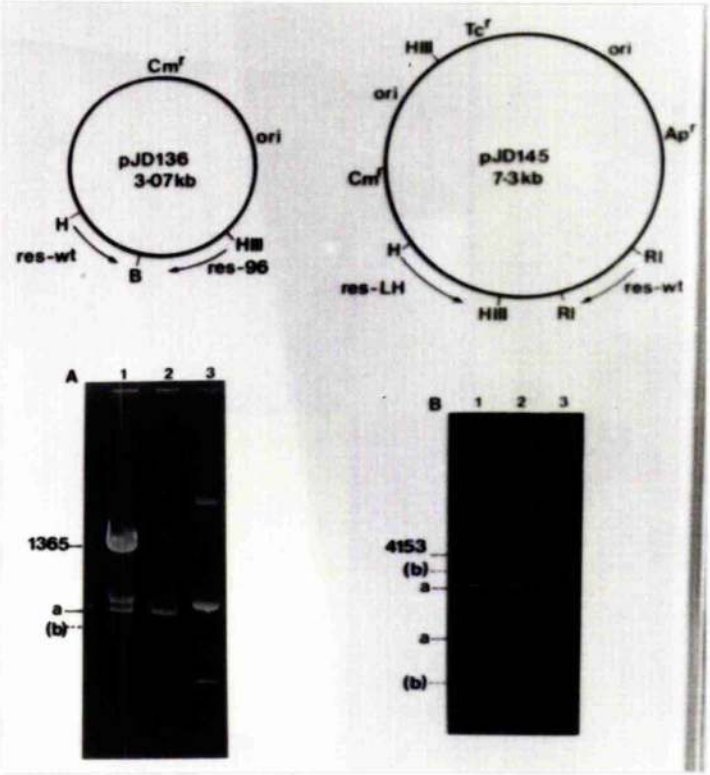


Figure 7.4 **Inversion analysis of pJD136 and pJD145.**

H=HincII B=BamHI H=HindIII RI=EcoRI

1% agarose gels

a=position of initial orientation plasmid bands

b=proposed position of flipped plasmid bands

A: lane1: pJD136+RSF1365 gen.120, HindIII/BamHI digest

lane2: pJD136 gen.120, HindIII/BamHI digest

lane3: pJD136+RSF1365 gen.120

B: lane1: pJD145+pDS4153 gen.135, HindIII digest

lane2: pJD145 gen.135, HindIII digest

lane3: pJD145+pDS4153 gen.135

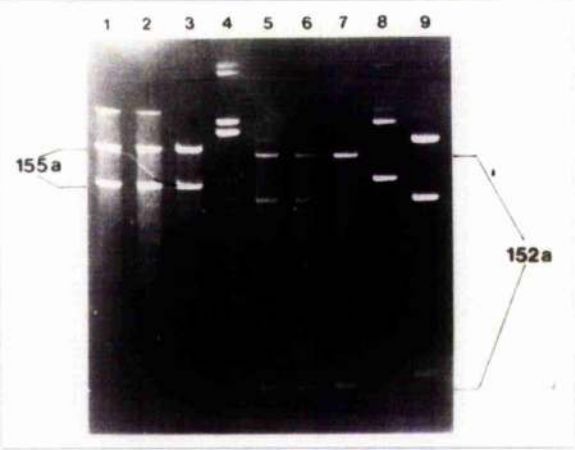


Figure 7.5 **Inversion analysis of pJD152 and pJD155.**

After 50 generations subculture pJD155 inversion was monitored by EcoRI restriction; pJD152 inversion by PstI restriction, and DNA's run on a 1.2% agarose gel.

lanes 1&2: pJD155+pAA33, EcoRI digest lane 3: pJD155, EcoRI digest

lane 4: pJD155+pAA33 lanes 5&6: pJD152+pAA33, PstI digest

lane 7: pJD152, PstI digest lane 8: pJD152+pAA33

lane 9: pAA33, PstI digest

rapid resolution for pJD151 (*res*-sym x *res*-wt in direct repeat), and reasonably efficient resolution for pJD154 (*res*-sym x *res*-sym in direct repeat).

DISCUSSION

There have been numerous conflicting reports (some unpublished) as to the efficiency of resolvase mediated inversion compared to resolution. In agreement with the results of Heffron *et al*(1980), equilibrium can be established *in vivo* between two forms of an invertible substrate containing Tn1000 *res*-wt's within 50 generations (N.Grindley pers. comm.). This would imply a minimum inversion rate of 0.007 per cell per generation, approximately 20 times less efficient than the corresponding resolution reaction.

It is reported that inversion cannot be detected *in vitro* (Reed and Grindley 1981; Symington 1982). However, using more sensitive silver staining techniques to assay for inverted product, it has been shown that inversion is 50 times less efficient than the corresponding resolution reaction *in vitro* (N.Cozzarelli pers. comm.).

The general conclusion is that inversion between two copies of *res*-wt is less efficient than resolution, but several factors appear to influence the former which do not noticeably affect the latter reaction. Firstly the distance between adjacent sites would appear to affect the inversion rate by an inverse proportion.

Secondly inversion does not necessarily appear reciprocal *in vivo*. This can be explained by assuming inversion is reciprocal, but that one form of the substrate is less competitive within a cell when

plasmid incompatibility effects come into play. A second, and conventionally less tenable idea, is that in fact inversion can be non-reciprocal. This idea has little credence if the absolute requirements for resolvase mediated site-specific recombination are only as described in Chapter 3, that is, two copies of res-wt on a supercoiled plasmid, resolvase and Mg^{2+} ions. However, as reported in Section 6.4, in certain circumstances the resolution reaction can be driven only in the presence of a host-factor(s). Clearly if some sequences adjacent to res, which were previously described as unnecessary for res function in terms of the resolution reaction, but which may promote inversion, and these are deleted in either one of the participating res sites in an inversion substrate, then there is a distinct possibility that inversion may be non-reciprocal due to the nature of the participating res sites.

Whilst resolution of a res-wt x res-LH substrate was shown to be 75% as efficient as resolution of res-wt x res-wt, the former substrate, with inverted res sites, does not appear to promote inversion. This can be explained if inversion was occuring (conceivably at 75% efficiency of res-wt x res-wt, ie. at approximately 0.0015 per cell per generation which would result in equilibrium at 250 generations), but the inverted product is less competitive than pJD145 and hence undetectable. Alternatively, the process of juxtaposition of res-sites prior to strand exchange is mechanistically different for inversion than for resolution (see Chapter 8). If this process is far slower for inverted sites than for directly repeated sites, then resolvase-DNA complexes may form at res-LH well before the juxtaposition mechanism can bring the two res-

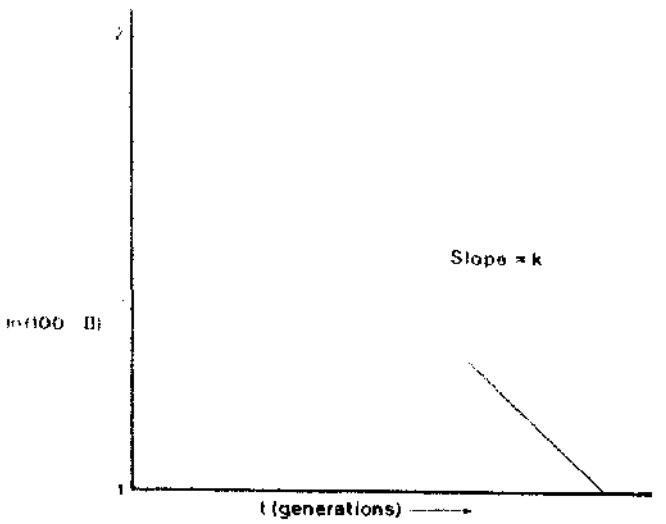
regions together. The quaternary structure formed at res-LH may then be unrecognizable to that formed at res-wt; efficient juxtaposition may not occur thereby not allowing recombination.

The absence of Hin-resolvase interchangeability is presumably at least partly due to their respective affinities for different DNA binding sites. Whilst the nature of Hin binding sites has not been established, there are no sequences similar to the resolvase consensus binding site in or around the inverted repeats of the phase variation system (sequence reference: Zieg and Simon 1980). Recent results, showing the amino acid sequence of the α-helical fold binding domain of resolvase (Abdel-Meguid et al in Press), would in fact predict that shared binding sites are extremely unlikely; if the amino acid sequence of Hin and resolvase are compared then the region of least homology coincides with the proposed binding domains of the proteins. Whereas Hin, Pin and Gin share >50% amino acid homology at this region, resolvase shares <25% homology.

7.3 Appendix.

In order to compare rates of resolution and inversion, equations were derived to represent (a) resolution, where substrate A is converted to product B in time (t); this assumes the reaction is irreversible which is unlikely, but in fact the reverse reaction: intermolecular fusion, is rare enough to be undetectable (Arthur 1981, Krasnow and Cozzarelli 1983). (b) inversion, where substrate A is converted to substrate B and vice versa in time (t), the end-point being when 50:50 equilibrium of A and B is first established. I am indebted to Dr. D.K. Summers for help with this problem.

A.



B.

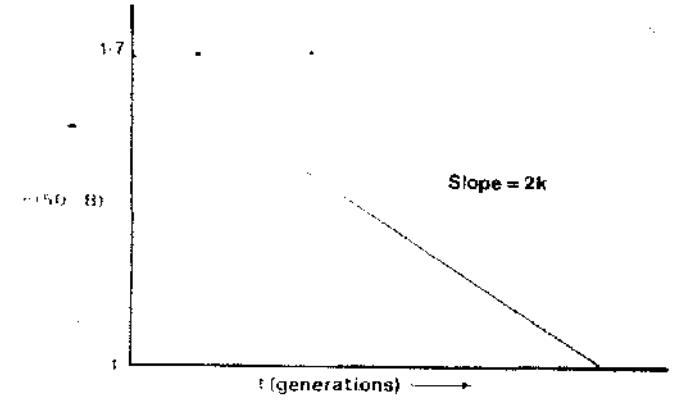


Figure 7.6 **Graphs to estimate rates of resolution and inversion.**

A. Resolution

B. Inversion

See text for details.

facing page 155

(a) Resolution.

$$A \longrightarrow B, \quad B=(100-A)$$

$$\frac{dB}{dt} = k(A)$$

$$= k(100-B)$$

$$\Rightarrow \quad dt = \frac{dB}{k(100-B)}$$

$$\Rightarrow \quad kt = -\ln(100-B) + c$$

This equation can be represented on a graph (Fig. 7.6), plotting ln(100-B) against t, where the slope (k) is a measure of the rate of resolution.

(b) Inversion.

$$A \rightleftharpoons B$$

$$\frac{dB}{dt} = k'(A) - k'(B)$$

$$= k'(100-B) - k'(B)$$

$$\Rightarrow \quad dt = \frac{dB}{(k'(100-B) - k'(B))}$$

$$= \frac{dB}{(100k' - k'B - k'B)}$$

$$= \frac{dB}{(100k' - 2k'B)}$$

$$\Rightarrow \quad 2k't = -\ln(50 - B) + c$$

This equation can be represented on graph (Fig. 7.6), plotting ln(50-B) against t, where the slope (k') is a measure of the rate of inversion.

CHAPTER 8

Discussion

8.1 Location of target sites by DNA binding proteins: sliding and tracking.

The observations and results obtained in the course of the experimental work presented in this thesis address mainly the second of three processes involved in site-specific recombination. That is, for convenience, it is possible to dissect site-specific recombination into (1) the process by which the recombination protein locates and binds to the recombination site, (2) the process of juxtaposition of both sites involved in recombination prior to strand exchange, and (3) the events of strand cleavage, passage and ligation which produce the recombinant products. This third process, as exhibited classically by topoisomerases, is discussed in Chapter 1.

The location and binding of resolvase to any of its three binding sites is a process which is shared by, for example, all site-specific genome regulatory proteins such as polymerases, repressors and activators involved in transcription. When considering *E.coli*, the number of copies per cell of any particular site-specific genome regulatory protein is much smaller than the number of base pairs of DNA (for example, approximately 10^1 *lac* repressor molecules, and approximately 5×10^3 RNA polymerase molecules compared to approximately 10^7 base pairs). Therefore, in specific binding interactions the target sites are each in competition with 10^7 potential (overlapping) non-specific binding sites. It is clear that there are two types of interaction between these regulatory proteins and DNA. Firstly they show affinity for non-specific sites largely of an 'electrostatic' nature; that is, binding involves charge-charge

interactions between DNA phosphates and the basic residues of the protein (Record et al 1976). Specific binding involves the interaction of a matrix of DNA hydrogen bond donors and acceptors, located in the grooves of the double-helix, with the sterically complementary acceptors and donors of the protein binding site, as discussed in Chapter 4.

Clearly the combination of relatively slow, three-dimensional diffusion of a regulatory protein, coupled with the potentially competitive role of the many non-specific sites, should combine to make straightforward specific site location a very slow process. Yet a paradox exists that a regulatory protein seems to be able to locate and bind to a specific DNA target site much more rapidly than permitted by a simple three-dimensional diffusion reaction. This was first observed by measuring the in vitro kinetics of the binding of E.coli lac repressor to lac operator sites (Riggs et al 1970; Berg et al 1981; Winter and von Hippel 1981; Winter et al 1981). To account for this observation two mechanisms have been proposed for the facilitated transfer of the regulatory protein once bound non-specifically to DNA.

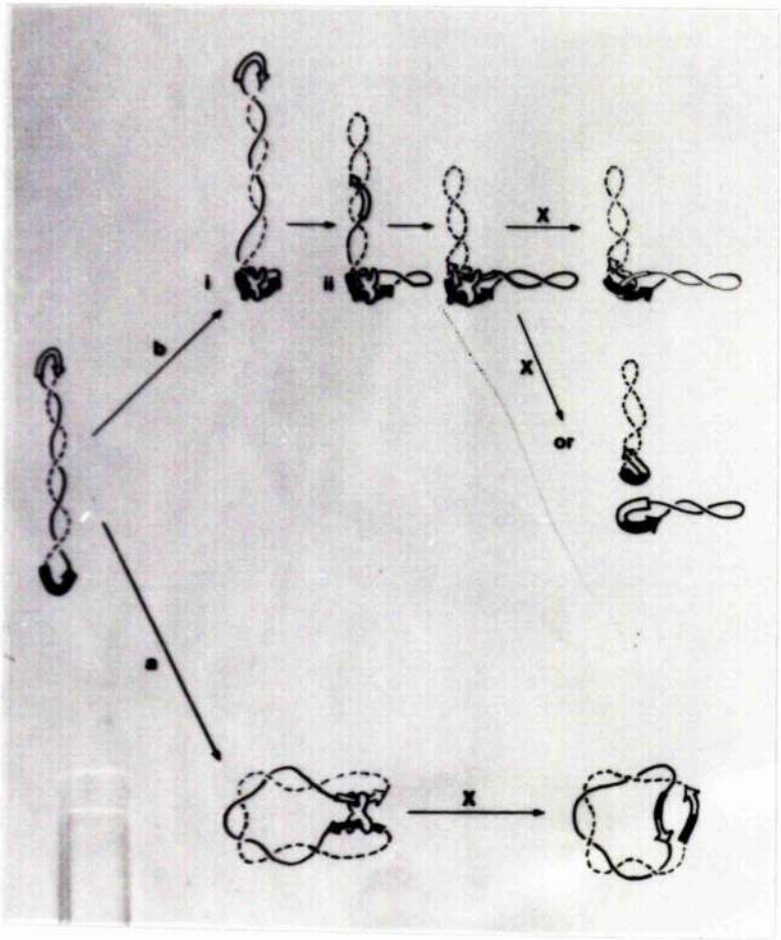
One is called 'sliding' and is defined as the diffusion of the regulatory protein, while non-specifically bound, in a one-dimensional random walk along the DNA molecule. Sliding depends on the electrostatic attraction between the protein and DNA but results in no net ion displacement: the monovalent cation displaced from in front of the protein as it moves along the DNA molecule is simply replaced by one binding to the DNA behind the protein. Thus, there is no thermodynamic barrier to repressor sliding along the DNA, since the

relaxation of the ion atmosphere is fast relative to the repressor sliding rate.

The other proposed facilitated transfer mechanism is called 'intersegment transfer', and involves the rapid and direct transfer of the regulatory protein from one segment of a DNA molecule to another as a consequence of the relative diffusion of these segments within the 'domain' of the molecule. That is, the protein can become transiently bound between two DNA segments and when the segments diffuse apart one of the DNA protein contacts will break and, if the two binding interactions are equally tight, the repressor will have a 50% chance of being transferred to a new and distant site on the DNA molecule. This process may be very fast because it circumvents the large activation barrier involved in the dissociation of the protein into solution.

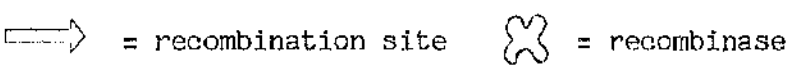
Several experiments (reviewed by Berg et al 1982) have shown that proteins can translocate along DNA by sliding, and that this facilitated transport mechanism is used by proteins in locating specific sites under in vitro conditions on otherwise 'naked' DNA at a rate corresponding to the scanning of approximately 10^3 base pairs in one second. It is proposed that both facilitated transfer mechanisms play a significant role in DNA target location in vivo (Winter et al 1981).

The second process involved in site-specific recombination: the juxtaposition of both recombination sites, shares common features to the first. That is, it involves the location by a resolvase-res complex of either another similar complex, or simply a second 'naked' res site. An important prediction can be made to distinguish between



facing page 160

Figure 8.1 **Catenation resulting from collision or tracking.**

 = recombination site = recombinase

(a) Collision: cross-over (X) results in formation of complex catenanes as a consequence of the intertwining between sites in the substrate.

(b) Tracking: cross-over can potentially generate two distinct outcomes. The following processes are involved:

i. specific binding of the recombinase to a site, and non-specific binding to adjacent DNA.

ii. tracking by the elecrostatic attraction between the recombinase and non-specific DNA. This process segregates intertwining into the two different domains.

Juxtaposition: simply for descriptive purposes the sites are drawn antiparallel. This arrangement would not favour recombination-see Fig. 8.2.

Crossover can result in either simple catenanes containing one interlock, or free circles. This is dependent on the nature of the strand-exchange mechanism.

whether this process relies on passive three-dimensional diffusion (collision), or facilitated diffusion in one dimension. Assuming the commonly held view that superhelical DNA adopts a plectonemic configuration in solution then, as depicted in Fig. 8.1, different products of *in vitro* resolution will result from (a) a passive three-dimensional diffusion, or (b) a one-dimensional facilitated diffusion process to juxtapose the recombination sites. Analysis of the catenated products of *in vitro* resolution show that they are interlinked once which is predicted if the juxtaposition of *res* sites is via one-dimensional diffusion (Krasnow and Cozzarelli 1983). The presence of one interlock is assigned to a feature of the strand-exchange mechanism directed by resolvase (Krasnow *et al* 1983).

This second one-dimensional diffusion process, which is proposed to be formally equivalent to the first, involving electrostatic attraction of a protein subunit binding domain to non-specific DNA, has been termed 'tracking' (Sherratt *et al* 1983; Krasnow and Cozzarelli 1983; Kitts *et al* 1983). The model explains two additional results: intermolecular recombination is proscribed since by tracking resolvase confines its search for a second site to a single dimension and determines whether sites are joined by a continuous piece of DNA. Secondly, tracking would juxtapose inverted *res* sites in an anti-parallel alignment, a configuration thought to be unfavourable for recombination (Fig. 8.2), hence explaining the relative efficiencies of resolution compared to inversion.

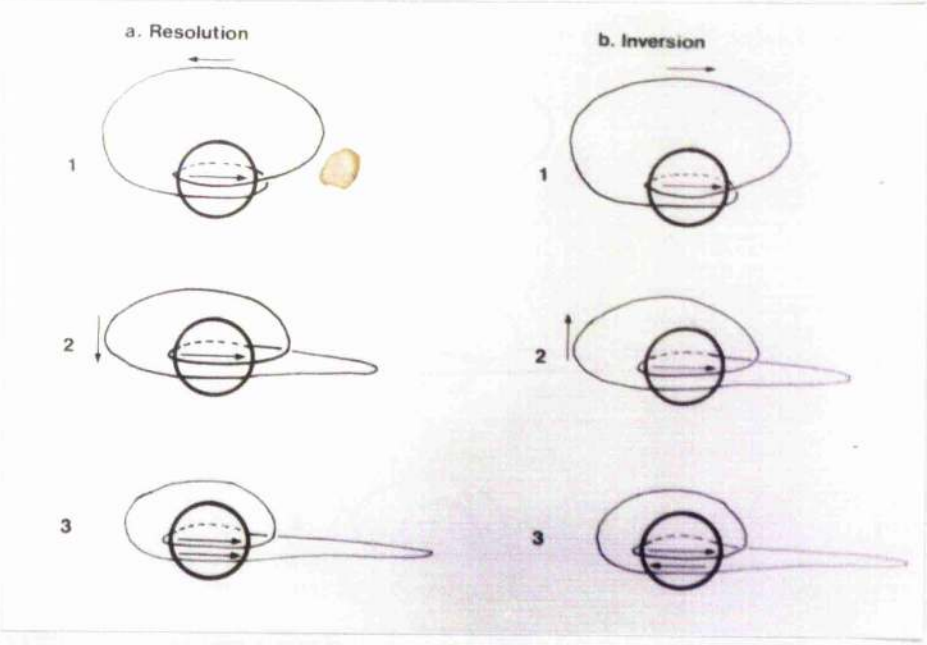


Figure 8.2 **Tracking on resolution and inversion substrates.**

The wrapping of DNA around the recombination protein dictates how recombination sites are juxtaposed.

(a) Direct repeats of the site are brought together in a parallel manner(3) prior to recombination.

(b) Inverted repeats of the site are brought together in an anti-parallel alignment(3) - a configuration unfavourable for recombination.

8.2 Tracking versus collision: influence of higher order structures at the recombination site.

The substrate specificities of site-specific recombination systems are different. Lambda integrative recombination is an intermolecular fusion reaction, and excisive recombination is an intramolecular resolution reaction across direct repeats of the two hybrid sites attL and attR. The recombination systems discussed in Chapter 7 such as phase-variation and G-loop flipping use intramolecular recombination across inverted sites. As discussed above, the efficiency of each type of reaction will depend on the mode adopted by a recombination protein-site complex to juxtapose the second site. Moreover as each type of recombination system has evolved, the structure of the recombination protein-site complex will have become more suited to the mode of juxtaposition most appropriate.

It was stated rather simply above that if the mode adopted is random collision (which would allow inversion and intermolecular fusion) then the efficiency of resolution might be reduced compared to that promoted using a tracking mode. However, it is important to define random collision in the *in vivo* context in order to illustrate how this mode can be efficiently optimised. Whereas tracking is initiated by the trapping of a small loop of DNA between each binding site, involving non-specific binding to DNA adjacent to the bound site (Fig. 8.1), collision might initially utilise intersegment transfer to enable the protein-site complex to become non-specifically bound to a distant region on either the same molecule or a different molecule, and location of the second site could then utilise further intersegment transfers coupled to a sliding mechanism capable of

scanning small regions. Clearly there are random features to this mode: the process is dependent on random DNA diffusion, but it can be efficiently optimised if the protein-site complex offers, for example, two exposed non-specific DNA binding sites to facilitate intersegment transfer. Presumably quite subtle differences in the structure of the recombination protein-site complex can determine the choice of mode: if scanning is confined to an orientation dependent one-dimensional process, or an orientation independent three-dimensional mode.

An example of how the higher order structure of complexes at a recombination site can determine the mode of juxtaposition is provided by lambda site-specific recombination. The properties of various lambda site-specific recombination reactions are summarised in Table 8.1 (Craig and Nash 1983). Clearly integration and Xis-promoted excision share a collision mode of juxtaposition whereas Xis-independent excision has the features of utilising the tracking mode. From electron microscopy analysis (Better et al 1982) and topology studies (Pollock and Nash 1983) it is proposed for attP, which contains a multitude of specific binding sites for both Int and IHF, that these proteins associate to form a complex around which attP would be wrapped. Figure 8.3 shows a diagram of a nuleosome-like structure formed by the wrapping of attP around a complex of Int and IHF. But whereas attP has binding sites on either arm relative to the core sequence, for the hybrid sites attL and attR only one arm, the P arm, contains specific recognition sites for Int and IHF, and therefore complexes at these sites cannot assume complete nucleosome-like structures (Fig. 8.3).

Craig and Nash (1983) propose that complexes of Int and IHF and

Table 8.1 **Properties of bacteriophage lambda site-specific recombination reactions.**

For all three reactions a superhelical substrate was incubated with Int, IHF and where indicated Xis *in vitro*.

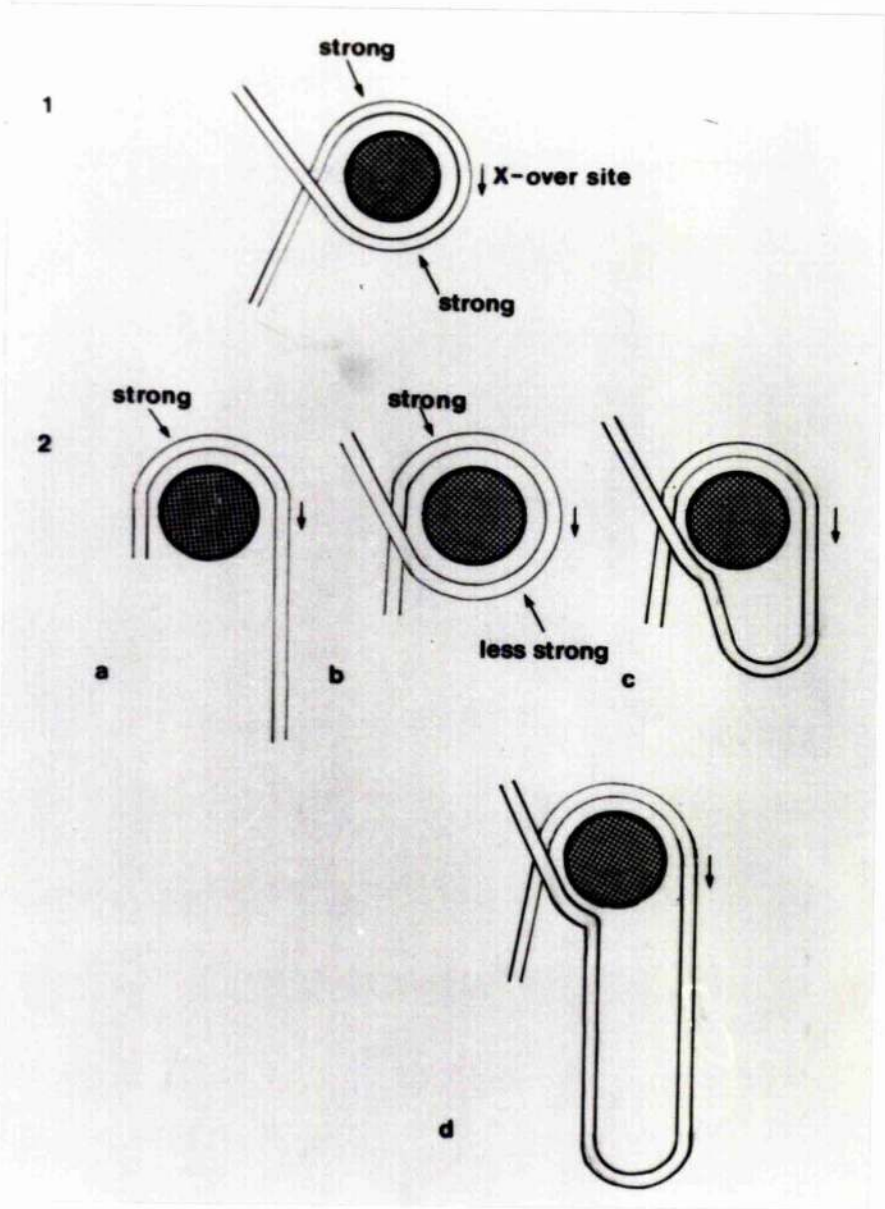
	Efficient recombination			Catenation after resolution of supercoiled DNA
	Fusion	Inversion	Resolution	
Integration (attPxattB)	Yes	Yes	Yes	Complex
Xis-promoted Excision (attLxattR)	Yes	Yes	Yes	Complex
Xis-independent Excision (attLxattR)	No	No	Yes	Free circles

prophage att sites derive their potential for tracking from their asymmetric nucleosome-like structure as diagrammed in Fig. 8.3. The position of the P or P' arm with respect to the complex is likely to be fixed because of the strong specific interactions. Because the nonspecific interactions between the B or B' arm and the protein complex are less strong, this DNA can translocate with respect to the protein-att site complex. This would effectively be one-dimensional diffusion of the protein-att site complex along the adjacent DNA.

The properties of Xis-promoted and Xis-independent excision are different although both reactions utilise the same att sites. It is proposed that Xis activates complexes of Int, IHF and prophage att sites so that collision between att sites provides effective precursors to recombination and, furthermore, that Xis alters the higher order structure of complexes of Int, IHF and prophage att sites so that tracking does not occur (Craig and Nash 1983). Xis may strengthen the non-specific interactions of protein and the B arm, effectively stabilising the asymmetric nucleosome, or weaken those interactions sufficiently so that tracking does not occur. The differences in the Xis-promoted and Xis-independent pathways illustrates how the proteins involved can influence the higher order structure of complexes of att sites and recombination proteins and thereby determine the pathway of recombination.

8.3 Higher order structures at res may determine the mode of juxtaposition.

An interpretation of the results presented in Chapters 6 and 7 can be made at the molecular level in terms of the higher-order



facing page 167

Figure 8.3 **Nucleosome and semi-nucleosome structures at the recombination site.**

1. Strong interactions between protein and the DNA of both arms of the site surrounding the crossover site dictate that a full nucleosome structure may be formed.

2. If binding interactions between one arm of the site and the protein are substantially weaker than for the other arm, then the DNA can potentially translocate with respect to the semi-nucleosome allowing a one-dimensional diffusion of the protein-site complex along adjacent DNA. This sliding pathway could bring into juxtaposition the second recombination site.

structures of resolvase-res complexes. The higher order structure at res-wt can be envisaged as being an asymmetric nucleosome analogous to prophage att-Int-IHF complexes. The arrangement of binding sites for resolvase is in fact asymmetric in respect to the cross-over site; this would facilitate asymmetric wrapping of the DNA around resolvase, in turn favouring the tracking mode for juxtaposition of sites.

Whereas a resolvase-res-wt complex can very effectively juxtapose a second res-wt or a secondary site such as res-LH, the latter cannot. The absence of site III would grossly affect the higher order structure of a resolvase-res complex to the extent that the essential asymmetry of the proposed nucleosome-like structure would be absent. Hence a tracking mode of juxtaposition might be impossible if emanating from res-LH. Is there any evidence that a collision mode might be adopted by res-LH? In fact a small amount of resolution was observed *in vivo*, increasing after subculture, for a res-LH x res-LH substrate (Kitts 1982), consistent with an inefficient juxtaposition mechanism.

The collision mode would also be implicated to bring about the albeit inefficient inversion of res-wt x res-wt. The increased rate of inversion when both sites are relatively close together is explicable if random collision juxtaposes inverted sites. If the rate of inversion is between 0.005 and 0.001 per cell per generation, then a resolution reaction similarly dependent on collision at the same efficiency would go to completion within 200 to 1000 generations, which is consistent with the low level resolution for res-LH x res-LH. Clearly resolvase-res complexes are ill equipped to make use of collision as an efficient juxtaposition mode, and perhaps this is a

reflection of a low potential for intersegment transfer.

Further evidence of a role of higher order structures at res determining the mode of juxtaposition is provided by studies on res-sym function. The in vivo evidence that res-sym is a primary site, although 'recessive' to res-wt, and that resolution is more efficient than inversion for a res-sym x res-sym substrate, suggests that the tracking mode of juxtaposition is utilised by res-sym. Furthermore the HimA protein is required to drive recombination. Just as HimA, together with HimD in the form of IHF, is required in the formation of complexes with Int at lambda att sites, it appears essential to the formation of potential tracking complexes together with resolvase at res-sym. The nature of resolvase binding sites at res-sym is such that an asymmetric nucleosome structure in respect to the crossover site is not feasible for a resolvase-res-sym complex, which will in turn proscribe tracking emanating from this site. Instead, the evidence presented suggests that in some way HimA interaction with the complex can enable the formation of a suitably asymmetric structure that can potentiate tracking.

In the absence of HimA, a low level rate of resolution for res-sym x res-sym occurs as evident from the appearance of resolution products after 85 generations subculture; this could be the product of a collision juxtaposition mechanism. In the same period no resolution products for a res-sym x res-LH substrate could be detected. For a res-wt x res-LH inversion substrate no products of recombination could be detected, whereas a similar resolution substrate recombines efficently, and res-wt x res-wt inversion is easily detectable for an inefficient reaction. This suggests that 'efficient' collision must

involve similar res sites, be they in direct or inverted repeat. So either the higher order structures or naked res sites offer better targets for a collision mechanism if they share identity with the other participating resolvase-res complex. If this is the case then it might be expected that for collision dependent res-LH x res-LH resolution and inversion substrates, resolution and inversion may proceed at equivalent rates.

Best targets for tracking may also be identical sites. The efficiency of the tracking mechanism may enable the juxtaposition to occur before the second site itself is occupied by resolvase, or resolvase may be displaced at the second site by the action of tracking. A plausible reason for the slight reduction of res-wt x res-LH resolution, compared to that involving two res-wt's, might be that the higher order structure at res-LH cannot be displaced or recognised by tracking emanating from res-wt.

The absolute requirement for HimA of resolution reactions involving the res-sym primary site raises the tantalising question of the role, if any, of HimA at res-wt. The in vitro reqirements for resolution suggest absolutely no role at all. However, an obvious difference between lambda and Tn3 specified site-specific recombination is the in vitro reqirement for the latter of a supercoiled substrate. For every combination of att site substrate and protein requirement tested, recombination is observed with both relaxed and supercoiled substrates although for the latter recombination proceeds more efficiently. The available evidence suggests that neither initial site location by Int and IHF (+/- Xis), nor strand exchange are limited by using a relaxed substrate.

Recombination using relaxed substrates is most efficient at low ionic strength: conditions which would promote nucleosome formation in the absence of supercoiling by facilitating protein-protein interactions (Pollock and Nash 1983). So it remains to be tested if Tn3 specified site-specific recombination can proceed on relaxed substrates *in vitro* at low ionic strength in the presence of purified HimA.

The role of IHF in lambda site-specific recombination is unclear apart from being invoked in the part of promoting the formation of higher order structures. The requirement for IHF does not appear absolute. In a *hip* mutant *att*P x *att*P and *att*P x *att*L is reduced about 100-fold (Enquist *et al* 1979), whereas excisive recombination is less severely affected, especially when additional Int is provided in *trans* from an *int*-constitutive phage (Enquist *et al* 1979a; Miller and Friedman 1980), or both *int* and *xis* genes are derepressed (Gottesman and Ambremski 1980). Moreover two *int* mutants that partially relieve the host factor requirement for integrative and excisive recombination have been found: *int*-h3 (Miller *et al* 1980) and *xin* (reviewed by Weisberg and Landy 1983). The Int-h3 protein promotes a low level of integrative recombination *in vitro* in the absence of IHF. Presumably in the instances described other proteins, especially Int, can compensate for the absence of IHF in higher order structure formation at *att*.

8.4 The role of juxtaposition mechanisms for other site-specific recombination sytems.

The simple relationship between the ability to produce free or singly interlocked circles from recombination of a resolution

substrate and the inability to recombine inversion substrates or intermolecularly efficiently suggests a tracking mode for juxtaposition. However it is not inconceivable that a protein-site complex may be proficient for both tracking and collision. This is one interpretation of how the bacteriophage P1 encoded site-specific recombination system can function; this involves the recombinase Cre interacting with the recombination sites, loxP. The in vitro products of recombination on a resolution substrate are 50% free circles and 50% singly interlocked catenates (Abremski et al 1983; K.Abremski pers. comm.). Moreover Cre can efficiently recombine inversion substrates and fuse molecules by intermolecular recombination. For this system there is absolutely no requirement for supercoiling; substrates can be either relaxed or linearised. This may be a consequence of the ability of Cre to avidly wrap DNA around itself at loxP; Cre is a quite large recombination protein of 35K and possibly the number of DNA binding domains present gives it a great degree of flexibility to determine the higher order structure at loxP.

Due to problems of protein purification, in vitro inversion systems specified by S.typhimurium and Mu have not been established. However, in vivo analysis shows that low level deletion can occur when one of the recombination sites is reorientated with respect to the other (Scott and Simon 1982; Plasterk et al 1983). Moreover separate molecules, each carrying one site of the phase variation system are fused by Hin (Scott and Simon 1982). These results would suggest that the family of inversion recombinases when complexed to sites juxtapose adjacent sites by collision. It will be interesting to see the binding patterns for these proteins to their recombination sites; for such

similar proteins as resolvase to act by collision rather than tracking a prediction is that the protein binding sites would be arranged symmetrically around the cross-over site.

8.5 Future experiments.

An experiment being conducted in this laboratory is testing the prediction that tracking will juxtapose adjacent *res* sites prior to recombination. Using a substrate molecule containing one *res*-wt and several *res*-96 sites in direct repeat means that after the initial *res*-wt x *res*-96 recombination event, successive recombination events are proscribed since no suitably asymmetric *res* sites will be present on the molecule to initiate tracking again. The products of recombination will be analysed to see if *res*-wt always recombines with the adjacent *res*-96, and if there is any directionality to tracking emanating from *res*-wt. Both *in vitro* and *in vivo* analysis of recombination of this substrate might indicate how tracking copes with a naked and protein coated DNA molecule.

Further analysis of the nature of resolvase-*res* complexes is essential if we are to learn more about the importance of higher order structures. Such analysis is technically difficult, but one possible approach is an adaptation of DNA footprinting using chemical probes to supercoiled substrates. For example, a supercoiled plasmid with one copy of *res* can be incubated with and without resolvase. Subsequent addition of DMS would methylate the N^7 of guanine and the N^3 of adenine of unprotected bases. The plasmid DNA would then be suitably restricted, end-labelled, and the appropriate fragment chemically sequenced using a G+A reaction. Methylation protection patterns for

resolvase-*res*-wt structures and resolvase-*res*-LH structures could be obtained this way, and may be informative as to the comparative nature of higher-order structures at these different *res* sites.

The extent of DNA wrapping around resolvase at these different *res* sites can also be tested. A superhelical plasmid containing a copy of *res* can be relaxed by eukaryotic topoisomerase I in the presence and absence of resolvase. It might be possible to see a difference in the average linking number of the distribution of the relaxed topoisomers due to the binding of resolvase. A comparison of this difference using plasmids containing different *res* sites may complement the chemical probe-superhelical substrate experiments.

The experiments reported using *res*-sym have raised the question as to the part played by HimA on higher order structures at *res*. Critical confirmation awaits *in vitro* resolution analysis with the addition of purified HimA (and possibly HimD). Otherwise both standard footprinting techniques and analysis of higher order structures upon addition of HimA might shed light on this role, if not at *res*-wt, then at *res*-sym.

The experiments described in previous Chapters have helped to shed light on how coevolution of the resolvase protein and the recombination site *res* have optimised a particular mode of juxtaposition of recombination sites. This mode will very efficiently help to recombine directly repeated recombination sites; this is the arrangement of recombination sites in the cointegrate transposition intermediate which is consequently rapidly resolved to generate the

end-products of Tn3 transposition. As a consequence of the adoption of the tracking mode of juxtaposition other site-specific recombination events, such as intramolecular inversion and intermolecular fusion, are proscribed.

REFERENCES

Abdel-Meguid,S.S., Grindley,N.D.F., Templeton,N.S., Steitz,T.A. (1984)
Cleavage of the site-specific recombination protein, γδ resolvase: the smaller of two fragments binds DNA specifically.
Cell, in Press.

Abremski,K., Hoess,R. & Sternberg,N. (1983)
Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination.
Cell **32**, 1301-1311.

Adhya,S. & Miller,W. (1979)
Modulation of the two promoters of the galactose operon of *Escherichia coli*.
Nature **279**, 492-494.

Anderson,W.F., Ohlendorf,D.H., Takeda,Y. & Matthews,B.W. (1981)
Structure of the Cro repressor from bacteriophage lambda and its interaction with DNA.
Nature **290**, 754-758.

Arndt,K.T., Boschelli,F., Cook,J., Takeda,Y., Tecza,E. & Lu,P. (1983)
Lambda phage Cro repressor interaction with DNA.
J.Biol.Chem. **258**, 4177-4183.

Arthur,A.K. (1981)
Molecular genetics of the transposition process.
PhD thesis, University of Sussex.

Arthur,A.K. & Sherratt,D.J. (1979)
Dissection of the transposition process: a transposon-encoded site-specific recombination system.
Mol.Gen.Genet. **175**, 267-274.

Austin,S., Ziese,M. & Sternberg,N (1981)
A novel role for site-specific recombination in maintenance of bacterial replicons.
Cell **25**, 729-736.

Bachmann,B.J. (1972)
Pedigrees of some mutant strains of *Escherichia coli* K12.
Bacteriol.Rev. **36**, 525-557.

Backendorf,C., Brandsma,J.A., Kartasova,T. & van de Putte,P. (1983)
In vivo regulation of the *uvrA* gene: role of the '-10' and '-35' promoter region.
Nucl.Acids Res. **11**, 5795-5810.

Bedbrook,J.R. & Ausubel,F.M. (1976)
Recombination between bacterial plasmids leading to the formation of plasmid multimers.
Cell **9**, 707-716.

Berg,O.G., Winter,R.B. & von Hippel,P.H. (1981)
Diffusion-driven mechanisms of protein translocation on nucleic acids. 1 Models and theory.
Biochem. **20**, 6926-6948.

Berg,O.G., Winter,R.B. & von Hippel,P.H. (1982)
How do genome-regulatory proteins locate their DNA target sites?
Trends Biochem.Sci. **7**, 52-55.

Bernard,H.W. & Helinski,D.R. (1979)
Use of the lambda phage promoter to promote gene expression in hybrid plasmid cloning vehicles.
Meths.Enzymol. **68**, 482-492.

Better,M., Lu,C., Williams,R. & Echols,H. (1982)
Site-specific DNA condensation and pairing mediated by the Int protein of bacteriophage lambda.
Proc.Natl.Acad.Sci.USA **79**, 5837-5841.

Biel,S.W. & Hartl,D.L. (1983)
Evolution of transposons: natural selection for Tn5 in Escherichia coli K12.
Genetics **103**, 581-592.

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

Bolivar,F., Rodriguez,R.L., Greene,P.J., Betlach,M.C., Heyneker,H.L. & Boyer,H.W. (1977)
Construction and characterization of new cloning vehicles II. A multipurpose cloning system.
Gene **2**, 95-113.

Broach,J.R. (1982)
The yeast 2u circle
Cell **28**, 203-204.

Campbell,A. (1962)
Episomes.
Adv.Genet. **11**, 101-145.

Campbell,A. (1981)
Some general questions about movable elements and their implications.
Cold Spring Harbor Symp.Quant.Biol. **45**, 1-9.

Caruthers,M.H. (1980)
Deciphering the protein-DNA recognition code.
Acc.Chem.Res. **13**, 155-160.

Casadaban,M.J., Chou,J., Lemaux,P., Tu,C.-P.D. & Cohen S.N. (1981)
Tn3: transposition and control.
Cold Spring Harbor Symp.Quant.Biol. **45**, 269-273.

Champoux,J. (1977)
Strand breakage by the DNA untwisting enzyme results in covalent attachment of the enzyme to DNA.
Proc.Natl.Acad.Sci.USA **73**, 3488-3491.

Chang,A.C.Y. & Cohen,S.N. (1978)
Construction and characterization of amplifiable multicopy cloning vehicles derived from the P15A cryptic miniplasmid.
J.Bacteriol. **134**, 1141-1156.

Chau,L. & Cox,E.C. (1983)
Competition between high and low mutating strains of Escherichia coli.
Evolution **37**, 125-134.

Chiang,S. & Clowes,R.C. (1980)
Intramolecular transposition and inversion in plasmid R6K.
J.Bacteriol. **142**, 668-682.

Chou,J.,Casadaban,M.J.,Lemaux,P.G. & Cohen,S.N. (1979a)
Identification and characterisation of a self-regulated repressor of translocation of the Tn3 element.
Proc.Natl.Acad.Sci.USA. **76**, 4020-4024.

Chou,J.,Lemaux,P.G.,Casadaban,M.J. & Cohen,S.N. (1979b)
Transposition protein of Tn3: identification and characterization of an essential repressor-controlled gene product.
Nature, **282**, 801-806.

Chow,L.T. & Bukhari,A. (1976)
The invertible DNA segments of coliphages Mu and PI are identical.
Virology **74**, 242-248.

Collins,J. (1981)
Instability of palindromic DNA in Escherichia coli.
Cold Spring Harbor Symp.Quant.Biol. **45**, 409-416.

Craig,N.L. & Nash,H.A. (1983)
The mechanism of phage lambda site-specific recombination: collision versus sliding in att site juxtaposition.
in UCLA Symposia on Molecular and Cellular Biology, Eds. Cozzarelli,N. and Fox,C.F. (In Press).

Craig,N.L. & Nash,H.A. (1983)
The mechanism of bacteriophage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase.
Cell 35, 795-803.

Dagart,M. & Ehrlich,S.D. (1979)
Prolonged incubation in calcium chloride improves the competence of E.coli cells.
Gene 6, 23-28.

Datta,N. & Hedges,R.W. (1972)
Trimethoprim resistance coferred by W plasmids in Enterobacteriacae.
J.Gen.Microbiol. 72, 349-355.

Diver,W.P., Grinsted,J., Fritzinger,D.C., Brown,N.L., Altenbucher,J., Rogowsky,P. & Schmitt,R. (1983)
DNA sequences of and complementation by the tnpR genes of Tn21, Tn501 and Tn1721.
Mol.Gen.Genet. 191, 189-193.

Doolittle,F.W. & Sapienza,C. (1980)
Selfish genes, the phenotype paradigm and genome evolution.
Nature 284, 601-603.

Dougan,G.,Saul,M.,Twigg,A.,Gill,R. & Sherratt,D.J. (1979)
Polypeptides expressed in Escherichia coli K12 minicells by the transposition elements Tn1 and Tn3.
J.Bacteriol. 138, 48-54.

Edlin,G., Lin,L. & Bitner,R. (1977)
Reproductive fitness of P1, P2 and Mu lysogens.
J.Virol. 21, 560-564.

Edlin,G., Lin,L. & Kudrna,R. (1975)
Lambda lysogens of E.coli reproduce more rapidly than non-lysogens.
Nature 255, 735-737.

Edwards,K. Halligan,B., Davis,J., Nivera,N. & Liu,L. (1982)
Recognition sites of eukaryotic DNA topoisomerase I: DNA nucleotide sequencing analysis of Topo I cleavage sites on SV40 DNA.
Nucl.Acids Res. 10, 2565-2576.

Enquist,L.W., Nash,H. & Weisberg,R.A. (1979b)
Strand exchange in site-specific recombination.
Proc.Natl.Acad.Sci.USA 76, 1363-1367.

Enquist,L.W., Kikuchi,A. & Weisberg,R.A. (1979a)
The role of lambda integrase in integration and excision.
Cold Spring Harbor Symp.Quant.Biol. 43, 1115-1120.

Franklin,N.C. (1971)
The N operon of lambda: extent and regulation as observed in fusions to the tryptophan operon of E.coli.
In, The bacteriophage lambda. (Ed. A.D.Harshey).
Cold Spring Harbor Lab., Cold Spring Harbor, New York. pp. 621-637.

Galas,D.J. & Schmitz,A. (1978)
DNase footprinting: a simple method for the detection of protein-DNA binding specificity.
Nucl.Acids Res. 5, 3157-3170.

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

Gicquel-Sanzey,B. & Cossart,P. (1982)
Homologies between prokaryotic DNA-binding regulatory proteins and betweeen their sites of action.
EMBO Journal 1, 591-595.

Gill,R.,Heffron,F.,Dougan,G. & Falkow,S. (1978)
Analysis of sequences transposed by complementation of two classes of transposition deficient mutants of Tn3.
J.Bacteriol. 136, 742-756.

Gill,R.E.,Heffron,F. & Falkow,S. (1979)
Identification of the protein encoded by the transposable element Tn3 which is required for its transposition.
Nature 282, 797-801.

Gillen,J.R.,Karu,A.E.,Nagaishi,H. & Clark,C.J. (1977)
Characterization of the deoxyribonuclease determined by lambda reverse as exonuclease of Escherichia coli.
J.Mol.Biol. 113, 27-41.

Gottesman,S. & Abremski,K. (1980)
The role of himA and xis in lambda site-specific recombination.
J.Mol.Biol. 138, 503-512.

Gottesman,S. & Gottesman,M. (1975)
Excision of prophage λ in a cell-free system.
Proc.Natl.Acad.Sci.USA 72, 2188-2192.

Grindley,N.D.F.,Lauth,M.R.,Wells,R.G.,Wityk,R.J.,Salvo,J.J. & Reed,R.R. (1982)
Transposon mediated site-specific recombination: identification of three binding sites for resolvase at the res sites of γδ and Tn3.
Cell 30, 19-27.

Guyer,M.S. (1978)
The γδ sequence of F is an insertion sequence.
J.Mol.Biol. 126, 347-365.

Haggerty,D.M. & Schleif,R.F. (1976)
Location in bacteriophage lambda DNA of cleavage sites of the site-specific endonuclease from Bacillus amyloliquefaciens H.
J.Virol. **18**, 659-663.

Halford,S.E. & Johnson,N.P. (1980)
The EcoRI restriction endonuclease with bacteriophage lambda DNA: equilibrium binding studies.
Biochem.J. **191**, 593-604.

Hawley,D.K., Malan,T.P., Mulligan,M.E. & McClure,W.R. (1982)
Intermediates on the pathway to open-complex formation.
in Promoters: structure and function. Eds. Rodriguez, R.L. & Chamberlin,M.J., Praeger Publishers, New York, pp54-68.

Hawley,D.K. & McClure,W.R. (1983)
Compilation and analysis of Escherichia coli promoter DNA sequences.
Nucl.Acids Res. **8**,2237-2255.

Hedges,R.W. & Jacob,A.E. (1974)
Transposition of ampicillin resistance from RP4 to other replicons.
Mol.gen.Genet. **132**, 31-40.

Heffron,F., Bedinger,P., Champoux,J.J. & Falkow,S. (1977)
Deletions affecting the transposition of an antibiotic resistance gene.
Proc.Natl.Acad.Sci.USA **74**, 702-706.

Heffron,F., Kostriken,R., Morita,C. & Parker,R. (1981)
Tn3 encodes a site-specific recombination system: identification of essential sequences, genes, and the actual site of recombination.
Cold Spring Harbor Symp.Quant.Biol. **45**, 259-268.

Heffron,F.,McCarthy,B.J.,Ohtsubo,H. & Ohtsubo,E. (1979)
DNA sequence analysis of the transposon Tn3: three genes and three sites involved in transposition of Tn3.
Cell **18**, 1153-1163.

Heffron,F.,So,M. & McCarthey,B.J. (1978)
In vitro mutagenesis of a circular DNA molecule by using synthetic restriction sites.
Proc.Natl.Acad.Sci.USA. **74**, 702-706.

Helling,R.B.,Goodman,H.M. & Boyer,H.W. (1974)
Analysis of R.EcoRI fragments of DNA from lambdoid bacteriophage and other viruses by agarose-gel electrophoresis.
J.Virol. **14**, 1235-1244.

Hoess,R.H., Ziese,M. & Sternberg,N. (1982)
P1 site-specific recombination: nucleotide sequence of the recombining sites.
Proc.Natl.Acad.Sci.USA 79,3398-3402.

Hoyt,M.A., Knight,D.M., Das,A., Miller,H.I. & Echols,H. (1982)
Control of phage lambda development by stability and synthesis of CII protein: role of the viral CIII and host hflA, himA and himD genes.
Cell 31, 565-573.

Hsu,P.-L.,Ross,W. & Landy,A. (1980)
The λ phage att site: functional limits and interaction with Int protein.
Nature 285, 85-91.

Humayon,Z., Kleid,D. & Ptashne,M. (1977)
Sites of contact between lambda operators and lambda repressor.
Nucl.Acids Res. 4, 1595-1607.

Iida,S., Meyer,J., Kennedy,K.C. & Arber,W. (1982)
A site-specific conservative recombination system carried by bacteriophage P1. Mapping the recombinase gene cin and the cross-over sites cix for the inversion.
EMBO Journal 1, 1445-1453.

Johnson,A.,Meyer,B.J. & Ptashne,M. (1979)
Mechanism of action of the Cro protein of bacteriophage λ .
Proc.Natl.Acad.Sci.USA 75, 1783-1787.

Johnson,A.,Meyer,B.J. & Ptashne,M. (1979)
Interactions between DNA-bound repressors govern regulation by the λ phage repressor.
Proc.Natl.Acad.Sci.USA 76, 5061-5065.

Kamp,D. & Kahman,R. (1981)
The relationship of two invertible segments in bacteriophage Mu and Salmonella typhimurium DNA.
Mol.Gen.Genet. 189, 564-566.

Kamp,D., Kahman,R., Zipser,D., Broker,T.R. & Chow,L.T. (1978)
Inversion of the G DNA segment of phage Mu controls phage infectivity.
Nature 271, 577-580.

Kikuchi,Y. & Nash,H. (1979)
Nicking-closing activity associated with bacteriophage lambda int gene product.
Proc.Natl.Acad.Sci.USA 76, 3760-3764.

Kitts,P.A.K. (1982)
The transposition mechanism of Tn3.
Ph.D. thesis, University of Glasgow.

Kitts,P.,Symington,L.,Burke,M.,Reed,R. & Sherratt,D. (1982a)
Transposon-encoded site-specific recombination.
Proc.Natl.Acad.Sci.USA. 79, 46-50.

Kitts,P., Lamond,A. & Sherratt,D.J. (1982b)
Inter-replicon transposition of Tn1/3 occurs in two sequential genetically seperable steps.
Nature 295, 626-628.

Kostriken,R.,Morita,C. & Heffron,F. (1981)
Transposon Tn3 encodes a site-specific recombination system: identification of essential sequences, genes, and actual site of recombination.
Proc.Natl.Acad.Sci.USA. 78, 4041-4045.

Krasnow,M.A. & Cozzarelli,N.R. (1983)
Site-specific relaxation and recombination by the Tn3 resolvase: recognition of the DNA path between oriented res sites.
Cell 32, 1313-1324.

Krasnow,M.A., Stasiak,A., Spengler,S.J., Dean,F., Koller,T. & Cozzarelli,N.R. (1983)
Determination of the absolute handedness of knots and catenanes of DNA.
Nature 304, 559-560.

Kutsukake,K. & Iino,T. (1980)
Inversions of specific DNA segments in flagellar phase variation of Salmonella and inversion systems of bacteriophages P1 and Mu.
Proc.Natl.Acad.Sci.USA 77, 7338-7341.

Laban,A. & Cohen,A. (1981)
Interplasmidic and intraplasmidic recombination in Escherichia coli K12.
Mol.Gen.Genet. 184, 200-207.

Lamond,A.I. & Travers,A.A. (1983)
Requirement for an upstream element for optimal transcription of a bacterial tRNA gene.
Nature 305, 248-250.

Lin,L., Bitner,R. & Edlin,G. (1977)
Increased reproductive fitness of Escherichia coli lambda lysogens.
J.Virol. 21, 554-559.

Liu,L.F. & Wang,J.C. (1979)
Interaction between DNA and Escherichia coli DNA topisomerase I.
J.Biol.Chem. 254, 11082-11088.

Mandecki,W. & Reznikoff,W.S. (1982)
A lac promoter with a changed distance between -10 and -35 regions.
Nucl.Acids Res. **10**, 903-912.

Maniatis,T.,Jeffrey,A. & van de Sande,H. (1975)
Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis.
Biochem. **14**, 3787-3794.

Matthews,B.W., Ohlendorf,D.H., Anderson,W.F., Fisher,R.G. & Takeda,Y. (1983)
How does Cro repressor recognize its DNA target sites?
Trends Biochem.Sci. **8**, 25-29.

Maxam,A.M. & Gilbert,W. (1977)
A new method for sequencing DNA.
Proc.Natl.Acad.Sci.USA. **74**, 560-564.

Maxam,A.M. & Gilbert,W. (1980)
Sequencing end-labelled DNA with base-specific chemical cleavages.
Meths.Enzymol. **65**, 499-560.

McKenney,K., Shimatake,H., Court,D., Schmeissner,U., Brady,C. & Rosenberg,M. (1981)
A system to study promoter and terminator signals recognized by Escherichia coli.
In, Gene amplification and Analysis Vol.II. (Eds. J.G.Chirikjian & T.S.Papas).
Elsevier/North Holland, Amsterdam. pp.383-417.

Mercereau-Puijalon,O. & Kourilsky,P. (1976)
Escape synthesis of β-Galactosidase under the control of bacteriophage lambda.
J.Mol.Biol. **108**, 733-751.

Michaelis,G. & Starlinger,D. (1967)
Sequential appearance of the galactose enzymes in E.coli.
Mol.Gen.Genet. **100**, 210-215.

Merril,C., Gottesman,M., Court,D. & Adhya,S. (1978)
Discoordinate expression of the Escherichia coli gal operon after prophage lambda induction.
J.Mol.Biol. **118**, 241-245.

Miller,H.I. (1981)
Multilevel regulation of bacteriophage lambda lysogeny by the E.coli himA gene.
Cell **25**, 269-276.

Miller,H.I. & Friedman,D.I. (1977)
An E.coli gene product required for lambda site-specific recombination.
Cell **20**, 711-719.

Miller,H.I., Mozola,M.M. & Friedman,D.I. (1980)
int-h: an *int* mutation of phage lambda that enhances site-specific recombination in *Escherichia coli*.
Cell 20, 721-729.

Miller,J.H. (1972)
Experiments in molecular genetics.
Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Mizuuchi,K., Weisberg,R., Enquist,L., Mizuuchi,M.,Buraczynska,M., Foeller,C.,Hsu,P.-L.,Ross,W. & Landy,A. (1981)
Structure and function of the phage λ *att* site: size, Int-binding sites, and location of the crossover point.
Cold Spring Harbor Symp. Quant. Biol. 45, 429-437.

Mizuuchi,M. & Mizuuchi,K. (1980)
Integrative recombination of bacteriophage λ : extent of the DNA sequence involved in attachment site function.
Proc.Natl.Acad.Sci.USA 77, 3220-3224.

Muster,C.J., MacHattie,L.A. & Shapiro,J.A. (1983)
pλCM system: observations on the roles of transposable elements in formation and breakdown of plasmids derived from bacteriophage lambda replicons.
J.Bacteriol. 153, 976-990.

Nash,H.A. (1975)
Integrative recombination of bacteriophage lambda DNA *in vitro*.
Proc.Natl.Acad.Sci.USA 72, 1072-1076.

Nash,H.A. & Pollock,T.J. (1983)
Site-specific recombination of bacteriophage lambda: the change in topological linking number associated with exchange of DNA strands.
J.Mol.Biol. 170, 19-38.

Nash,H.A.,Mizuuchi,K.,Enquist,L.W. & Weisberg,R.A. (1981)
Strand exchange in lambda integrative recombination: genetics, biochemistry, and models.
Cold Spring Harbor Symp. Quant. Biol. 45, 417-428.

Ohlendorf,D.H., Anderson,W.F., Fisher,R.G., Takeda,Y. & Matthews, B.W. (1982)
The molecular basis of DNA-protein recognition inferred from the structure of Cro repressor.
Nature 298, 718-723.

Pabo,C.O. & Lewis,M. (1982)
The operator-binding domain of lambda repressor: structure and DNA recognition.
Nature 298, 443-447.

Paoletti,C., Le Pecq,J.-B. & Lehman,I.R. (1971)
The use of ethidium bromide-circular DNA complexes for the flourometric analysis of breakage and joining of DNA.
J.Mol.Biol. **55**, 75-100.

Peacock,A.C. & Dingman,C.W. (1967)
Resolution of multiple ribo-nucleic acid species by polyacrylamide gel electrophoresis.
Biochem. **6**, 1818-1827.

Philippsen,P.,Kramer,R.A. & Davies,R.W. (1978)
Cloning of yeast ribosomal DNA repeat unit in SstI and HindIII lambda vectors using genetic and physical size selections.
J.Mol.Biol. **123**, 371-386.

Plasterk,R.H.A., Brinkman,A. & Van de Putte,P. (1983)
DNA inversions in the chromosome of E.coli and in bacteriophage Mu; relationship to other site-specific recombination systems.
Proc.Natl.Acad.Sci.USA **80**, 5355-5358.

Plasterk,R.H.A., Ilmer,T.A.M. & van de Putte,P. (1983)
Site-specific recombination by Gin of bacteriophage Mu: inversions and deletions.
Virology **127**, 24-36.

Pollock,T. & Nash,H. (1983)
Knotting of DNA caused by a genetic rearrangement: evidence for a nucleosome-like structure in site-specific recombination of bacteriophage lambda.
J.Mol.Biol. **170**, 1-18.

Pribnow,D. (1975a)
Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter.
Proc.Natl.Acad.Sci.USA **72**, 784-789.

Pribnow,D. (1975b)
Bacteriphage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites.
J.Mol.Biol. **99**, 419-443.

Ptashne,M.,Jeffrey,A.,Johnson,A.D.,Maurer,R.,Meyer,B.J.,Pabo,C.O. Roberts,T.M. & Sauer,R.T. (1980)
How the λ repressor and Cro work.
Cell **19**, 1-11.

Record,M.T.Jr., Lohman,T.M. & deHaseth,P.L. (1970)
Ion effects on ligand-nucleic acid interactions.
J.Mol.Biol. **107**, 145-158.

Reed,R.R. (1981a)
Transposon-mediated site-specific recombination: a defined *in vitro* system.
Cell **25**,713-719.

Reed,R.R. (1981b)
Resolution of cointegrates between transposons γδ and Tn3 defines the recombination site.
Proc.Natl.Acad.Sci.USA. **78**, 3428-3432.

Reed,R.R. & Grindley,N.D.F. (1981)
Transposon mediated site-specific recombination: DNA cleavage and protein-DNA linkage at the recombination site.
Cell **25**, 721-729.

Reed,R.R., Shibuya,G.I. & Steitz,J.A. (1982)
Nucleotide sequence of γδ resolvase gene and demonstration that its gene product acts as a repressor of transcription.
Nature **300**, 381-383.

Reznikoff,W.S., Maquat,L.E., Munson, L.M., Johnson, R.C. & Mandecki,W. (1982)
The *lac* promoter: analysis of structural signals for transcription, initiation and identification of a new sequence-specific event.
in Promoters: structure and function. Eds Rodriguez,R.L. & Chamberlin,M.J., Praeger Publishers, New York, pp80-95.

Riggs,A.D., Suzuki,H. & Bourgeois,S. (1970)
lac repressor-operator interaction.
J.Mol.Biol. **48**, 67-83.

Ruther,U.,Koenen,M.,Otto,K. & Muller-Hill,B. (1981)
pUR222, a vector for cloning and rapid chemical sequencing of DNA.
Nucl.Acids Res. **9**, 4087-4098.

Schaller,H., Gray,C. & Hermann,K. (1975)
Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd.
Proc.Natl.Acad.Sci.USA **72**, 737-741.

Scott,T.N. & Simon,M.I. (1982)
Genetic analysis of the mechanism of the *Salmonella* phase variation site-specific recombination system.
Mol.Gen.Genet. **188**, 313-321.

Seeburg,P.H., Nusslein,C. and Schaller,H. (1977)
Interaction of RNA polymerase with promoters from bacteriophage fd.
Eur.J.Biochem. **74**, 107-113.

Sherratt,D.J., Arthur,A.K. & Dyson,P.J. (1981)
Site-specific recombination.
Nature **294**, 608-610.

Sherratt,D.J., Arthur,A., Bishop,R., Dyson,P., Kitts,P. & Symington,L. (1983)
Genetic transposition in bacteria.
in Genetic Rearrangement, The Fifth John Innes Symposium, Eds. Chater,K.F., Cullis,C.A., Hopwood,D.A., Johnston,A.W.B. & Woolhouse,H.W.. Croon Helm, London, pp59-74.

Siebenlist,U. & Gilbert,W. (1980)
Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7.
Proc.Natl.Acad.Sci.USA **77**, 122-126.

Silverman,M. & Simon,M. (1980)
Phase variation: genetic analysis of switching mutants.
Cell **19**, 845-854.

Simon,M., Zieg,J., Silverman,M., Mandel,G. & Doolittle,R. (1980)
Phase variation: Evolution of a controlling element.
Science **209**, 1370-1374.

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

Stacey,K.A. & Simpson,E. (1965)
Improved method for the isolation of thymine requiring mutants of Escherichia coli.
J.Bacteriol. **90**, 554-555.

Steitz,T.A., Ohlendorf,D.H., McKay,D.B., Anderson,W.F. & Matthews,B.W. (1982)
Structural similarity in the DNA-binding domains of catabolite gene activator and Cro repressor proteins.
Proc.Natl.Acad.Sci.USA **79**, 3097-3100.

Stephano,J.E. & Gralla,J.D. (1982)
Spacer mutations in the lac p^s promoter.
Proc.Natl.Acad.Sci.USA **79**, 1069-1072.

Stueber,D. & Bujard,H. (1981)
Organization of transcriptional signals in plasmids pBR322 and pACYC184.
Proc.Natl.Acad.Sci.USA **78**, 167-171.

Stueber,D. & Bujard,H. (1982)
Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid specified genes.
EMBO Journal **1**, 1399-1404.

Sutcliffe,J.G. (1978)
pBR322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs long.
Nucl.Acids Res. **5**, 2721-2728.

Symington,L.S. (1982)
Transposon-encoded site-specific recombination.
Ph.D. thesis, University of Glasgow.

Tacon,W., Bhamra,B., Sunar,B. & Sherratt,D.J. (1981)
Structure and function of plasmid ColK.
Plasmid **6**, 358-359.

Takanami,M., Sugimoto,K. & Okamoto,T. (1976)
Sequence of promoter for coat protein of bacteriophage fd.
Nature **260**, 297-302.

Takeda,Y., Folkmanis,A. & Echols,H. (1977)
Cro regulatory protein specified by bacteriophage lambda: structure, DNA-binding and repression of RNA synthesis.
J.Biol.Chem. **252**, 6177-6183.

Takeda,Y., Ohlendorf,D.H., Anderson,W.F. & Matthews,B.W. (1983)
DNA-binding proteins.
Science **221**, 1020-1026.

Travers,A.A., Lamond,A.I., Mace,H.A.F. & Berman,M.L. (1983)
RNA polymerase interactions with the upstream region of the *E.coli tyrT* promoter.
Cell **35**, 265-273.

Van de Putte,P., Cramer,S. & Giphart-Gassler,M. (1980)
Invertible DNA determines host specificity of bacteriophage Mu.
Nature **286**, 218-222.

Vetter,D., Andrews,B.J., Roberts-Beatty,L. & Sadowski,P.D. (1983)
Site-specific recombination of yeast 2 micron DNA *in vitro*.
Proc.Natl.Acad.Sci.USA **80**, 7284-7288.

Vieira,J. & Messing,J. (1982)
The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers.
Gene **19**, 259-268.

von Hippel,P.H., Bear,D.G., Winter,R.B. & Berg,O.G. (1982)
Molecular aspects of promoter function: an overview.
in Promoters: structure and function. Eds. Rodriguez,R.L. & Chamberlin,M.J., Praeger Publishers, New York, pp3-33.

Weisberg,R.A. & Landy,A. (1983)
Site-specific recombination in phage lambda.
in Lambda II Eds. Hendrix,R.W., Roberts,J.W., Stahl,F.W. & Weisberg,R.A., Cold Spring Harbor Laboratory, pp211-250.

Wilson,D.B. & Hogness,D.S. (1966)
Galactokinase and uridine diphosphogalactose 4-epimerase from *Escherichia coli*.
Methods in Enzymology **8**, 229-240.

Winter,R.B. & von Hippel,P.H. (1981)
Diffusion-driven mechanisms of protein translocation on nucleic acids. 2 The *E.coli lac* repressor-operator interaction: equilibrium measurements.
Biochem. **20**, 6948-6960.

Winter,R.B., Berg,O.G. & von Hippel,P.H. (1981)
Diffusion-driven mechanisms of protein translocation on nucleic acids. 3 The *E.coli lac* repressor-operator interaction: kinetic measurements and conclusions.
Biochem. **20**, 6961-6977.

Wishart,W.L., Machida,C., Ohtsubo,H. & Ohtsubo,E. (1983)
Escherichia coli RNA polymerase binding sites and transcription initiation sites in the transposon Tn3.
Gene **24**, 99-113.

Wong,E.M., Meusing,M.A. & Polisky,B. (1982)
Temperature sensitive copy number mutants of ColE1 are located in an untranslated region of the plasmid genome.
Proc.Natl.Acad.Sci.USA **79**, 3570-3574.

Wulff,D.L. & Rosenberg,M. (1983)
Establishment of repressor synthesis.
in Lambda II, Eds. Hendrix,R.W., Roberts,J.W., Stahl,F.W. & Weisberg,R.A., Cold Spring Harbor Laboratory, pp53-73.

Zieg,J., Hilmen,M. & Simon,M. (1978)
Regulation of gene expression by site-specific inversion.
Cell **15**, 237-244.

Zieg,J. & Simon,M. (1980)
Analysis of the nucleotide sequence of an invertible controlling element.
Proc.Natl.Acad.Sci.USA **77**, 4196-4200.