MICROBIAL GENETICS APPLIED TO BIOTECHNOLOGY

PRINCIPLES AND TECHNIQUES OF GENE TRANSFER AND MANIPULATION

VENETIA A SAUNDERS AND JON R SAUNDERS

MICROBIAL GENETICS APPLIED TO BIOTECHNOLOGY

PRINCIPLES AND TECHNIQUES OF GENE TRANSFER AND MANIPULATION

VENETIA A. SAUNDERS AND JON R. SAUNDERS

CROOM HELM
London & Sydney

Croom Helm Ltd, Provident House, Burrell Row,
Beckenham, Kent BR3 1AT
Croom Helm Australia, 44-50 Waterloo Road,
North Ryde, 2113, New South Wales

British Library Cataloguing in Publication Data

Saunders, Venetia A.
Microbial genetics applied to biotechnology.
1. Microbial genetic engineering
2. Biotechnology
I. Title II. Saunders, Jon R.
660'.62 TP248.6

ISBN 0-7099-2365-1
ISBN 0-7099-4435-7 Pbk

Typeset in Times Roman by Leaper & Gard Ltd, Bristol, England
Printed and bound in Great Britain

Preface

This book describes techniques of microbial genetics and how they may be applied to biotechnology. The text is concerned largely with the application of these techniques to microbial technology. We have therefore utilised illustrative material that is given in our own courses in applied microbiology. The book assumes in the reader a basic knowledge of microbial genetics and industrial microbiology. We hope it will prove useful to undergraduates, postgraduates and others taking courses in applied microbiology.

We would like to thank various colleagues, including John Carter, Julian Davies, Gordon Dougan, David Hopwood, Gwyn Humphreys, Alan McCarthy, David O'Connor, Tony Hart, Steve Oliver, Roger Pickup, Hilary Richards, Bob Rowlands, David Sherratt, Peter Strike, Richard Sykes and Liz Wellington, all of whom provided information at various stages during the writing of this book. Many thanks are also due to Linda Marsh for patiently typing the many drafts of the manuscript.

Contents

1

Introduction

Natural genetic variation has always been exploited by man to improve the properties of microbial strains. Spontaneous mutations that arise in microbial populations and that have properties advantageous to man have been gradually selected over centuries of use. However, it is only since the development of modern genetic techniques that more rational approaches have been possible. Such newer technologies have permitted the tailoring of microorganisms, plant or animal cells to manufacture specific products of commercial or social benefit and to manage the environment.

It is the intention of this book to demonstrate how fundamental principles of microbial genetics have been applied to biotechnology. Microorganisms have a wide range of physiological capabilities not encountered in higher organisms. Bacteria and fungi can also normally be grown in culture under controlled conditions. Furthermore, microorganisms possess relatively simple genetic organisations and are more amenable to manipulation than higher organisms. Genetics is a discipline *par excellence* for breaking down and analysing complex biological problems and for using the information so obtained for the rational manipulation of biological processes. Fundamental studies in genetics have provided both direct and indirect benefits to biotechnology by permitting a better understanding of the biology of microorganisms. This has generated an infrastructure of knowledge, particularly for bacteria, such as *Escherichia coli*, and fungi, such as *Saccharomyces cerevisiae*, that has, in turn, promoted the development of modern manipulative techniques. During recent years much emphasis has been placed on the role of *in vitro* recombinant DNA technology (see Chapters 3, 4 and 5) in stimulating a resurgence of biotechnology. Although it is true that *in vitro* techniques do provide a radical new ingredient in the exploitation of biological systems, it should not be forgotten that many commercially important applications of microbial genetics involve the use of traditional *in vivo* gene manipulations (see Chapters 6, 7, 8 and 9). Moreover, the development of much *in vitro* genetic manipula-

tion technology has been strictly dependent on 'conventional' *in vivo* techniques (Chapters 2 and 4).

The ability to exploit microbial genetics has depended to a great extent on gene transfer and other evolutionary mechanisms found in natural populations of microorganisms. Crucial to the development of genetic technologies has been an understanding of the biology of plasmid and virus genomes. Such extrachromosomal elements, which are widespread in both bacteria and fungi, provide important vehicles for the natural exchange of genes both within and between species (see Chapter 2). These elements are relatively small, autonomously replicating, nucleic acid molecules, and are particularly amenable to manipulation by both *in vivo* and *in vitro* genetic techniques. Furthermore, many extrachromosomes are present in host cells at copy numbers greater than chromosomes. This provides a means of amplifying genes that the element happens to carry. Plasmids may be classified into incompatibility (Inc) groups based on the ability of pairs of plasmids to coexist in the same cell. Those plasmids that fail to coexist belong to the same Inc group. Incompatible plasmids share homologous systems for controlling replication and hence copy number (see Chapter 5). The diversity of Inc groups in different bacteria and fungi has permitted the construction of a variety of useful genetic vectors, of both broad and narrow host range, for gene manipulation (see Chapters 2, 3 and 5).

This book describes *in vivo* and *in vitro* genetic manipulation methodologies and considers examples of their application to biotechnology. Genetics can be used for the deliberate enhancement of existing attributes of microorganisms (see Chapter 6) and for their manipulation to provide novel properties or products (see, for example, Chapters 6 and 7). Examples of the application of the principles of microbial genetics to veterinary and human medicine (Chapter 7), to agriculture (Chapter 8) and to the management of the environment (Chapter 9) are also considered.

2

In Vivo Genetic Manipulation

Genes can be shuffled within and between microbial species by a variety of mechanisms. Such mechanisms lead to the formation of new genotypes by bringing together and reassorting genes from different organisms. In bacteria new combinations of genes may be generated by using one of the three natural processes of gene transfer, namely **transformation** (section 2.2), **conjugation** (section 2.3) or **transduction** (section 2.4). In fungi, genetic exchange can be effected through the agency of the **sexual** (section 2.5) or **parasexual** (section 2.6) cycle. Transformation systems are also available for gene transfer in yeasts and certain filamentous fungi. **Protoplast fusion** (section 2.7) provides a further route for combining groups of genes from different strains and modifying the genetic constitution of microorganisms. These various processes for manipulating genomes enable the formation of innumerable genetic combinations, in turn producing variability within microbial populations.

Transposable genetic elements (section 2.1) provide another source of variability among microorganisms. These elements can insert into and excise from a variety of replicons and promote a number of genome rearrangements. Transposable elements can cause mutation when they interrupt the coding sequence of a gene. Various properties of these elements make them useful for manipulating genomes.

This chapter describes mechanisms for gene manipulation *in vivo*.

2.1 TRANSPOSABLE GENETIC ELEMENTS

2.1.1 Properties of transposable elements

Transposable genetic elements are segments of DNA that are capable of inserting as discrete nonpermuted DNA sequences at various sites within a genome. These elements have been found in the genomes of a variety of microorganisms, including bacteria, fungi and bacteriophages, as well as in

the genomes of higher organisms. This section is primarily concerned with the properties of prokaryotic transposable elements, some of which have been studied in detail and have been used widely in genetic manipulation. However, transposable elements of fungi, for example *Ty* elements, have some similar properties.

The smallest prokaryotic transposable elements are **insertion sequences (IS)** or IS-like elements. These encode determinants involved in promoting/regulating transposition. Larger genetic entities (originally termed **transposons (Tn)**)[1] encode accessory determinants (for example, antibiotic resistance, lactose fermentation), in addition to transposition functions. Transposable elements can be divided into a number of classes according to genetic organisation and transposition mechanism (see Table 2.1). Class I includes IS-like modules and composite (compound) elements which are formed from them. Composite transposons comprise a central DNA segment flanked on either side by a copy of an IS (Figure 2.1). Available evidence indicates that the information required for transposition of composite transposons is encoded by the IS constituents. Furthermore, the ISs are capable of transposing independently. Class II (complex transposons) comprises Tn*3* and its relatives and Class III transposing bacteriophages, for example Mu. Certain elements cannot be categorised into these classes. Such 'unclassified' elements include Tn*7* and Tn*916*. The properties of transposable elements have been extensively documented; see for example, Calos and Miller (1980), Cohen and Shapiro (1980), Bennett (1985), Campbell (1981), Kleckner (1981), Shapiro (1983), Cullum (1985) and Grindley and Reed (1985).

Transposition (and associated events discussed in section 2.1.3) in *E. coli* is independent of *recA*-mediated homologous recombination, implying reliance on transposon-encoded functions and/or other host recombination processes. Extensive DNA sequence homology between the element and the site of insertion is not required for transposition. Integrity of the ends of the transposable element is crucial for transposition. Normally the ends are inverted repeats (IRs) of one another. Phage Mu (and D108), Tn *7* and Tn*554* are exceptions in that they do not have true terminal repeats. In the case of IS*1* the minimal terminal sequence required for transposition is 21 to 25 bp. A specific sequence at base pairs 13 to 23, found at each extremity of IS*1*, appears to be an essential site. It has been suggested (Gamas *et al.*, 1985) that a host DNA binding protein, such as integration host factor (IHF) or a similar protein, could be involved in the transposition reaction by binding to such sites. For Tn*10* the outer 13 to 27 bp of the terminal IS*10* sequence, at each end, are essential for transposition, and sequences at 27 to 70 bp are important in the process (Way and Kleckner,

[1] The terms transposable element and transposon are now often used synonymously.

Table 2.1: Selected prokaryotic transposable elements and their properties

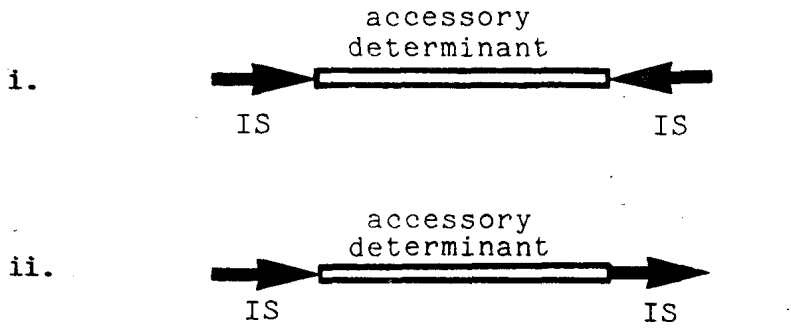
	Element	Size (kb)	Source	Accessory determinants*
Class IA (Individual IS modules)	IS*1*	0.768	*E. coli* and other entero-bacterial genomes, plasmid R100	none
	IS*2*	1.327	*E. coli* K12 genome, plasmid F	none
	IS*21*	2.1	*Pseudomonas aeruginosa* plasmid R68.45	none
Class IB (composite Tn elements)	Tn*5*	5.7 (IR.IS*50*)	*Klebsiella* plasmid JR67	Km
	Tn*9*	2.5 (DR.IS*1*)	plasmid R100	Cm
	Tn*10*	9.3 (IR.IS*10*)	plasmid R100	Tc
	Tn*903* (Tn*601*)	3.1 (IR.IS*903*)	plasmid R6	Km
Class II (Tn*3* family)	Tn*3*	4.957	plasmid R1	Ap
	Tn *1000* (γδ)	5.8	plasmid F	unknown
	Tn*501*	8.2	*P. aeruginosa* plasmid VS1	Hg
	Tn*551*	5.3	*Staphylococcus aureus* plasmid pI258	Em
	Tn*917*	5.3	*Streptococcus faecalis* plasmid pAD2	Em
Class III (transposing bacteriophages)	Mu	37.0	*E. coli* K12 genome	phage life cycle
Unclassified	Tn*7*	13.5	plasmid R483	Tp, Sm

* Ap, Cm, Em, Hg, Km, Sm, Tc, Tp: resistance to ampicillin, chloramphenicol, erythromycin, mercuric ions, kanamycin (and other aminoglycosides including G418), streptomycin, tetracycline and trimethoprim respectively. IR, inverted repeat; DR, direct repeat

1984). Transposons of the Tn*3* family (such as Tn*1/3*, Tn*21* and Tn*1721*) can transpose when only a single end (either right or left IR) is present (Arthur *et al.*, 1984; Avila *et al.*, 1984; Motsch and Schmitt, 1984). Such one-ended transposition occurs at low frequency and results in the formation of cointegrates when the appropriate transposase enzyme (see section 2.1.2.a) is present in the cell.

Transposable elements can transpose either within the same DNA molecule (**intramolecular transposition**), or from one DNA molecule to another (**intermolecular transposition**). Insertion of the element into the target DNA is generally accompanied by duplication of host DNA sequences at the target site (Tn*554* being a notable exception). The size of

Figure 2.1: Structures of composite (compound) transposons. (i) Insertion sequence (IS) in inverted repeat; (ii) IS in direct repeat



the duplication (typically 3 to 13 base pairs) is characteristic of a particular element. Furthermore, the degree of insertion specificity depends upon the element. Some transposons (for example Tn*5*, Tn*10*) show preference for particular target sites; whereas others (for example Tn*1*, Tn*3*) exhibit regional specificity, inserting efficiently at a number of different sites within preferred regions of a target molecule. Generally, AT-rich regions (where denaturability is greater) are preferred sites for integration of Tn*3*. Furthermore insertions of Tn *3* have been found near sequences resembling its ends.

Members of the Tn*3* family exhibit **transposition immunity** whereby the presence of a Tn on a plasmid can inhibit transposition of a second copy of that Tn on to that plasmid, but not on to another Tn-free plasmid residing in the same cell (Robinson *et al.*, 1977). There is, however, no barrier to introducing a second Tn by homologous recombination. Transposition immunity thus limits the number of copies of a Tn that may transpose on to a replicon. Transposition of the Tn*3* family is also regulated by resolvase acting at the level of transcription (see section 2.1.2.a). Different modes of regulation are employed by different transposons. Transposition of composite Tns, such as Tn*10*, can be regulated by copies of their component IS elsewhere in the cell. Regulation of Tn*10* transposition may involve RNA species encoded by IS*10*, which forms the terminal repeats of Tn*10*. One of the two copies of IS*10*, IS*10*-right (IS*10*-R), encodes a long open reading frame (ORF) specifying a function (presumptive transposase) essential for transposition. An inwardly directed promoter, pIN, which is located just upstream of this ORF, is responsible for its expression. A second outwardly directed promoter, pOUT, is located just inside the start of the ORF. The region of overlap includes the ATG translation start codon for the ORF. It is proposed that the transcript obtained from pOUT pairs with the transcript for the putative transposase from pIN, in turn inhibiting translation

of the transposase gene (Simons and Kleckner, 1983; Way and Kleckner, 1984; Kleckner, 1986).

The frequency of transposition is generally within the range 10^{-4} to 10^{-7}. However, transposition frequencies may be influenced by environmental stimuli (for example, temperature or conditions of stress for the host). The transposition of most transposable elements probably depends upon host replication functions. In *E. coli*, mutations in genes such as *polA* (for DNA polymerase I) and *gyrB* (for DNA gyrase, B subunit) have resulted in decreased transposition frequencies for a number of elements (Isberg and Syvanen, 1982; Syvanen *et al.*, 1982). By contrast, strains defective in the *dam* gene product (for DNA methylation) show increased frequencies of transposition (see section 4.12).

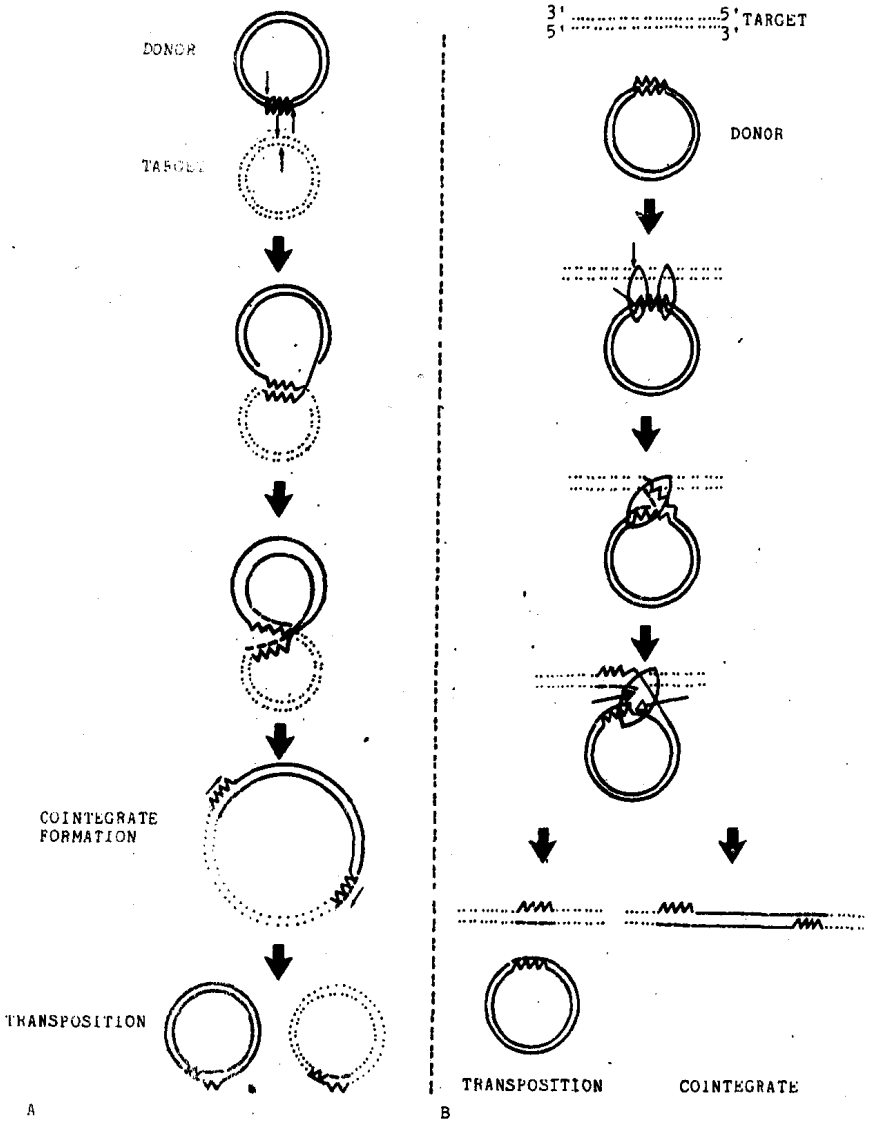
2.1.2 Transposition models

There are two general classes of transposition models, **replicative** and **conservative** (non-replicative). The underlying features of replicative models are (i) the precise joining of transposon sequences to target DNA at the insertion site (break/join process), and (ii) replication of the transposing element. The outcome of intermolecular replicative transposition is that one copy of the element remains at the original locus of the donor DNA molecule while a second copy integrates at the target site:

Replicon 1 : : Tn + Replicon 2 → Replicon 1 : : Tn + Replicon 2 : : Tn

However, transposable elements appear to differ in the manner in which they resolve transposition intermediates formed between donor and target molecules. In some cases, for example Tn*3* and Tn*1000* ($\gamma\delta$), resolution of intermediates (to give final separation of donor and target DNA each carrying a copy of the transposon) can occur after completion of element replication, through a site-specific recombination event involving a site internal to the element. In other cases, where no internal resolution system appears to operate, separation probably occurs as an integral part of the break/join and replication processes themselves, or does not occur at all.

There are various molecular models for replicative transposition (see, for example, Arthur and Sherratt, 1979; Shapiro, 1979; Bukhari, 1981; Galas and Chandler, 1981; Harshey *et al.*, 1982; Craigie and Mizuuchi, 1985), all of which account for break/join and replication processes, but differ in the order and details of these events. Such models can be divided into two classes, **symmetric** and **asymmetric**, depending upon whether transposition is initiated at both or only one end of the transposon. According to the Shapiro model (Figure 2.2A), an archetype for numerous other models, intermolecular transposition is initiated by separation of



transposon sequences from adjacent donor DNA sequences by a single-stranded nick at the boundaries of the transposon. Both ends of the element become ligated to the target DNA. Replication of the element results in the formation of a cointegrate structure comprising two directly repeated copies of the transposon, one at each junction of donor and target molecules. Resolution of the cointegrate occurs by recombination between the two copies of the element. This may be a site-specific recombination (as

Figure 2.2: Models for transposition. (A) The Shapiro model. Transposition is initiated by single-stranded nicks (↓) at the ends of the transposon on the donor molecule. Nicks also occur at the target site on the recipient molecule and the ends of the Tn become ligated to the recipient DNA. Donor and recipient molecules are fused, but four nicks are present in the structure. Replication of the element results in a cointegrate as an obligate intermediate. Resolution of the cointegrate is effected by recombination between the directly repeated copies of the Tn. (B) Harshey and Bukhari model. A protein-mediated association is brought about between the donor and target DNA molecules. The target site undergoes a double-stranded cleavage. Cleavage also occurs at one end of the Tn in one strand that becomes attached to the target DNA. (Cleavage sites are indicated by arrows, ↓.) Replication of the element (roll-in replication) occurs. Where the 3′ end of the newly synthesised strand (indicated by the arrow ⟶ at the left) is recognised for ligation, a simple (direct) transposition results. Where a nick is made in the parental strand (indicated by the arrow ⟶ at the right), and is ligated to the free end of the target strand, insertion of the entire donor molecule (carrying the Tn) into the target site results, generating a cointegrate structure. ∧∧, Transposon; ---, new DNA synthesis; O, protein

can occur in Tn*3* and Tn*1000*), or may be a *recA*-promoted homologous recombination event. This symmetric model invokes cointegrates as obligatory intermediates in the transposition process.

Asymmetric replicative models invoking branching pathways to explain transposition and cointegrate formation have also been proposed (Galas and Chandler, 1981; Harshey and Bukhari, 1981; Harshey *et al.*, 1982). A salient feature of these models is that replication starts from one end of the element that is ligated to the target DNA (roll-in-type replication) while the other end remains unligated until the final stage in the transposition process (see Figure 2.2B). The end product may be either a cointegrate or a simple (direct) transposition of the element, depending upon the precise nature of the break/join processes at the target site when replication is terminated. In 1% to 5% of Tn*3* transposition events a cointegrate is not involved as intermediate. Accordingly, direct transposition of Tn*3* might proceed by mechanisms such as these (Bennett *et al.*, 1983). One-ended transposition might also occur by such mechanisms (Arthur *et al.*, 1984; Avila *et al.*, 1984). For composite transposons, transposition may involve read-through replication, in which replication begins at the outside end of a flanking IS and proceeds through the central region of the transposon to the outside end of the second copy of the IS, such that the composite transposon can function as a unit (Galas and Chandler, 1981).

Although some transposition events appear to be replicative, others apparently do not involve replication of the entire transposon. In such cases the transposon may be excised from the donor site without replication. A model for such conservative mechanisms (Berg and Berg, 1983)

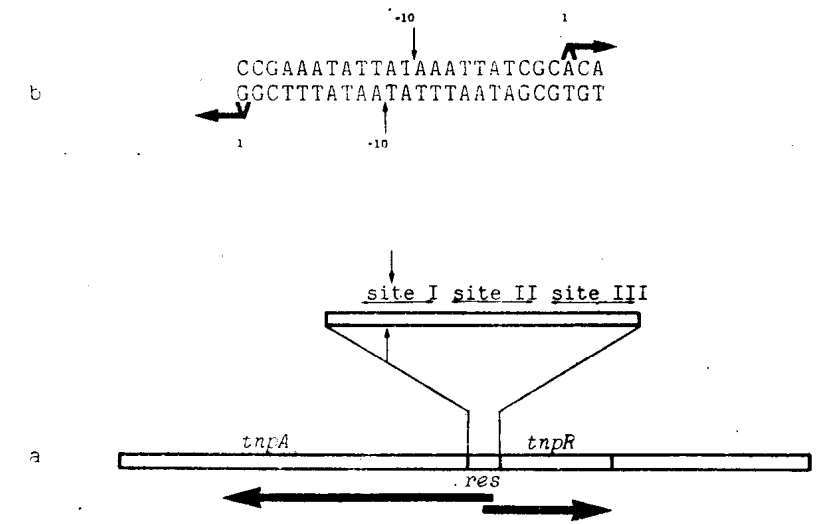
proposes that the transposable element is excised from the donor DNA by double-stranded cleavages at both ends of the element and is inserted into a target molecule. There is no DNA synthesis, except for that required to fill the gaps in the target DNA. The remainder of the donor replicon is lost from the host cell. A second copy of the donor molecule that did not participate in transposition is over-replicated to compensate for the loss of the first copy.

The significance of replicative and conservative modes of transposition would appear to vary from element to element. Some elements, such as IS*1*, might transpose in either mode: utilising the replicative mode where cointegrates are involved, or the conservative mode for direct transposition.

*(a) Structure of Tn*1000 (γδ)

The genetic map of Tn*1000* is given in Figure 2.3. The ***tnpA*** gene product, transposase, is required for the generation of cointegrate structures. The

Figure 2.3: Genetic map and sequenced promoter regions of Tn*1000* (γδ). (a) The element is drawn with the γ end towards the right. Transcription of *tnpA* and *tnpR* (⟶) is divergent from the *tnpA-tnpR* intercistronic region. The intercistronic region is enlarged to show DNA sites protected by resolvase binding. Site I contains cleavage sites (↓) for resolvase. These cleavage sites are within the −10 regions for *tnpA* and *tnpR*. *tnpA* gene product is transposase; *tnpR* gene product is resolvase; *res*, resolution site in the intercistronic space. (b) The −10 regions of the *tnpA* and *tnpR* promoters overlap. ⟵⟶, nucleotides at which *tnpA* and *tnpR* transcription is initiated (adapted from Reed *et al.*, 1982)



tnpR gene product, resolvase, mediates site-specific recombination at *res*, a site located in the intercistronic region between *tnpA* and *tnpR* genes (Reed, 1981). In addition, resolvase appears to regulate the frequency of transposition by repressing transcription of both the *tnpA* gene and its own gene. There are three sites (sites I, II and III) in the *tnpA-tnpR* intercistronic region that bind resolvase. Site I contains the recombinational crossover point and promoters for the divergent transcription of *tnpA* and *tnpR* genes. It is likely that resolvase regulates expression of transposition functions and resolves transposition intermediates by acting at these sites (Grindley *et al.*, 1982; Reed *et al.*, 1982; Grindley and Reed, 1985).

2.1.3 Transposon-mediated genome rearrangements

Transposable genetic elements can promote a number of genome rearrangements, including the inversion, deletion and duplication of DNA. Such rearrangements have a natural role in modulating gene expression and in promoting evolution. Transposable elements have been introduced into various Gram-negative and some Gram-positive organisms, including species with poorly developed genetic systems, and are used for a variety of genetic manipulations and analyses. Examples of transposon-mediated genome rearrangements in prokaryotes are described in Figure 2.4. These genetic rearrangements can arise independently of *recA* function, presumably as a consequence of intra- or intermolecular transpositions. Alternatively, combinations of transposition and homologous recombination events may be involved in which the transposon serves as a region of portable homology. By providing portable regions of homology, transposons can be used to mobilise the chromosome and nonconjugative plasmids (section 2.3), to fuse unrelated replicons and to mediate gene duplications. Furthermore, transposons can provide a source of mobile restriction sites to facilitate gene cloning.

Transposons can cause mutations when they insert into genes (section 4.7). Many of these mutations are strongly polar with respect to expression of those genes in the operon that lie distal to the insertion site. The mechanics of such polar effects are not completely clear. Transcription termination mechanisms that have been implicated in other types of polarity in prokaryotic operons (see Adhya and Gottesman, 1978) may be involved. Strong polarity may result at least in part from the presence of multiple translational stop codons within a transposable element and the absence of nearby translational reinitiation signals.

There are a number of cases where insertion of an IS or transposon (for example IS*2* or Tn*10*) can activate expression of adjacent genes. This is presumably due to the presence of promoter sequences at the ends of the element. In some instances the juxtaposition of determinants within/at the