GENE REARRANGEMENT

GENE REARRANGEMENT

Edited by

B.D.Hames

Department of Biochemistry, University of Leeds, Leeds LS2 9JT, UK

and

D.M.Glover

Cancer Research Campaign, Eukaryotic Molecular Genetics Research Group, Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, UK

IRL PRESS
at
OXFORD UNIVERSITY PRESS
Oxford New York Tokyo

Oxford University Press
Walton Street, Oxford OX2 6DP

Oxford is a trade mark of Oxford University Press

Published in the United States
by Oxford University Press, New York

British Library Cataloguing in Publication Data
Gene rearrangement.
1. Genes. Recombination
I. Hames, B.D. (B. David) II. Glover, David M. III.
574.87322
ISBN 0-19-963050-X
ISBN 0-19-963051-8 pbk

Library of Congress Cataloging-in-Publication Data
Gene rearrangement/edited by B.D. Hames and D.M. Glo.
(Frontiers in molecular biology)
Includes bibliographical references.
1. Genetic recombination. 2. Gene expression. 3. Gene amplification. I. Hames, B.D. II. Glover, David M. III. Series.
QH443.G44 1989 574.87'322—dc20 89-20021
ISBN 0-19-963050-X
ISBN 0-19-963051-8 (pbk.)

Previously announced as:

ISBN 1 85221 141 5
ISBN 1 85221 142 3 (Pbk)

Typeset and printed by Information Press Ltd, Oxford, England.

Preface

DNA sequence rearrangements are widespread in both prokaryotes and eukaryotes and are the subject of intense current research effort. Indeed the topic scope is now so large that it is impossible to review in any depth in a single book the recent discoveries which have been made in the field as a whole. An earlier volume in this series was devoted solely to the reorganization and expression of genes of the immune system. The present book extends this interest into several other related areas. Lex Van der Ploeg's chapter describes the molecular basis of antigenic variation in trypanosomes, covering DNA recombinational mechanisms in differential VSG gene expression in detail as well as other important expression events including discontinuous transcription and *trans*-splicing. In both the trypanosome and the immune system, DNA rearrangements are geared to providing a wide range of qualitative structural changes in the expressed gene product. In other circumstances it is the relative amount of a gene product which is the limiting factor, where the need for a particular product cannot be met by the maximal expression of the existing DNA sequences. Here DNA amplification is a solution to the problem. The chapter by George Stark, Michelle Debatisse, Geoffrey Wahl, and David Glover considers both developmentally programmed DNA amplification, particularly in dipteran flies where most progress has been made, and other increasingly important cases of amplification which occur as more isolated events such as the amplification of drug resistance genes and oncogenes in mammalian cells and the amplification of genes for resistance to toxic agents in whole organisms. Finally, every practising molecular biologist is all too aware that deletions, amplifications, inversions, insertions of foreign DNA, and translocations also occur in bacteria. We feel sure therefore that the chapter by Doug Berg, which examines the diverse mechanisms whereby these DNA rearrangements occur in prokaryotes, will be fascinating to a large cross-section of readers and not only to the specialists in this field. We thank each of the authors for their hard work and enthusiasm throughout this project and hope that their efforts will be as well-received by the readers as they were by the editors.

David Hames
David Glover

Contributors

Douglas E.Berg
Department of Microbiology and Immunology and Department of Genetics, Box 8093, Washington Medical School, St Louis, MO 63110, USA.

Michelle Debatisse
Unité de Génétique Somatique, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France.

David M.Glover
Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, UK.

George R.Stark
Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK.

Lex H.T.Van der Ploeg
Department of Genetics and Development, College of Physicians and Surgeons, Columbia University, 701 West 168th Street, New York, NY 10032, USA.

Geoffrey M.Wahl
The Salk Institute for Biological Studies, San Diego, CA 92138, USA.

Contents

1

Genomic rearrangements in prokaryotes

Douglas E.Berg

1. Introduction

Deletions, duplications and further gene amplifications, inversions, insertions of foreign DNAs, and translocations of resident DNA sequences to new sites have all been well documented in prokaryotes. These various DNA sequence rearrangements arise by diverse mechanisms: classical homologous recombination between repeated sequences; site-specific recombination; the movement of transposable elements; and several 'illegitimate recombination' processes. Certain rearrangements are as rare as point mutations whereas others are much more frequent, often developmentally regulated and, at the extreme, induced in each cell in a population by specific environmental or physiological signals. DNA rearrangements are of great biological interest. Some result from errors in DNA replication or repair or the movement of transposable elements, and provide models for frequent types of germ line and somatic mutations in humans. Some speed bacterial evolution, for example by facilitating the flow of genes among unrelated species and by generating duplications in which DNA sequences can diverge without selection and acquire new functions. Some contribute to pathogenicity or are part of a developmental program of prokaryotic cell differentiation. Some chromosomal inversions that were anticipated have not been found; such apparently 'forbidden' inversions may give insights into forces that determine the overall organization of the bacterial genome. Finally, some foreign DNAs seem prone to rearrange when cloned in *Escherichia coli*, an outcome of general significance to modern, recombinant DNA-based molecular biology.

2. Homologous recombination and genome organization

Repeated DNA segments in the size range of 1–10 kb collectively constitute several per cent of *E.coli* chromosomal DNA, and include

various insertion sequence (IS) transposable elements, ribosomal DNA genes, and fragments of defective or cryptic prophages. Each of these repeated DNAs is normally present at less than 10 copies per genome (1 – 3). Repeated DNAs that are matched for at least 20 – 30 bp constitute

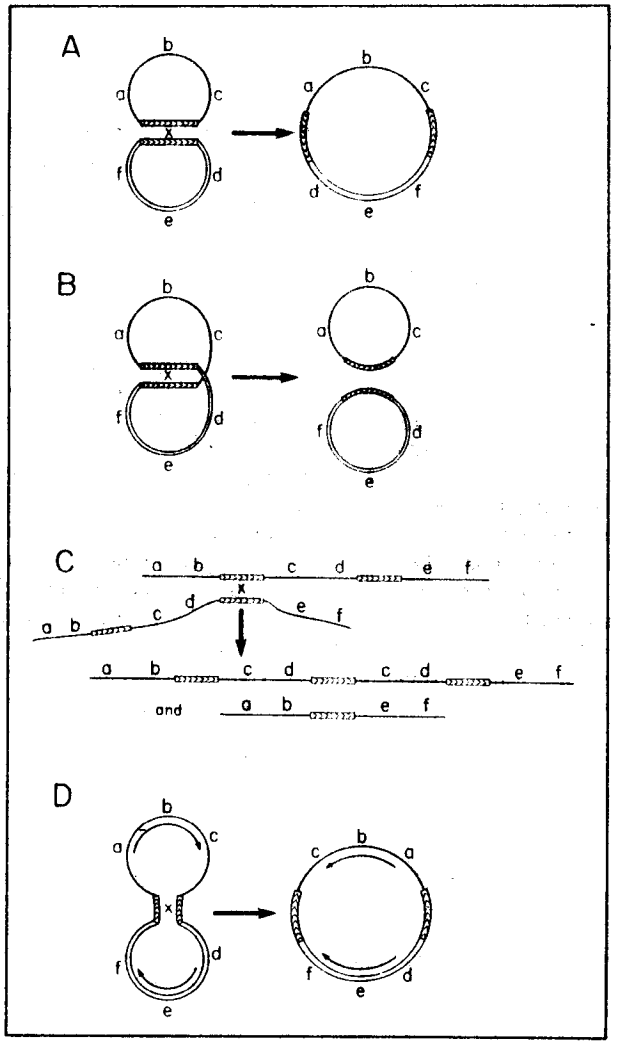

Figure 1. Consequences of recombination between homologous (repeated) sequences (⟿). (A) Intermolecular recombination; (B) Intramolecular recombination involving direct repeats; (C) intermolecular recombination between direct repeats ('unequal' crossing-over); (D) intramolecular recombination involving inverted repeats.

substrates for classical homologous crossing-over mediated by the protein products of *recA* and of several other recombination genes (4,5). *Figure 1* summarizes how the positions and arrangement of homologous sequences determines the type of rearrangements formed by crossing-over between them.

2.1 Intermolecular recombination

A single reciprocal cross-over between two circular DNA molecules generates a single large molecule, a co-integrate containing both parental DNAs (*Figure 1A*). This is illustrated by the formation of Hfr strains of *E.coli*: the conjugative F factor plasmid present in F^+ strains contains copies of the insertion sequences IS*2* and IS*3*; these elements are also present at multiple sites in the bacterial chromosome. Crossing-over between IS*2* or IS*3* in F and in the chromosome generates Hfr strains (co-integrates) at frequencies of about 10^{-3} – 10^{-2} (6,7). Many different Hfr strains have been found, differing in the site of F insertion and in the orientation of F in the chromosome, which reflects in part the many chromosomal positions of IS*2* or IS*3*. (Additional Hfr strains arise by transposition, as detailed in Section 4.2.) Equivalent intermolecular recombination, but involving λ and pBR322-derived plasmid DNAs, has been used to select specific clones from recombinant DNA libraries as an alternative to screening methods such as plaque hybridization (8), and also to estimate the length of homology needed for recombination (9).

Recombination can affect the arrangement of sequences in multicopy plasmids such as pBR322 (10) and λ*dv* (an autonomously replicating deletion derivative of phage λ; 11, 12). Crossing-over between monomers as in *Figure 1A* generates dimeric forms of the same plasmid; higher oligomeric forms can arise by further recombinational interactions (13,14). The number of plasmid replication origins per cell tends to remain constant, so that plasmid oligomerization decreases the number of separate plasmid DNA molecules and when not checked can foster the segregation of plasmid-free cells (13).

The first λ*dv* plasmid was isolated in the late 1960s, before the ability of plasmids to exist stably in a dimeric form was recognized. Electron microscope measurements had indicated that the contour length of λ*dv* DNA was longer than expected from its known λ gene content. A reasonable interpretation was that some foreign DNA segments had been acquired during λ*dv* formation (11). Subsequent tests showed, however, that the predominant form in the *recA*$^+$ strain was a dimer, and that monomers generated by homologous recombination could be stably maintained in *recA*$^-$ cells (13,15). A selection that permitted new λ*dv* plasmids to be isolated easily was developed and used to study λ*dv* structure. In each case the new plasmids were found to lack foreign sequences (12,16).

Some pBR322-derived recombinant plasmids are found preferentially as monomers, whereas others are found preferentially as dimers in *recA*$^+$

cells. The predominant form is affected by the size of the monomer unit. In general, monomeric plasmids predominate when the size is less than about 5 kb, whereas dimers predominate when the size is in the range of 8 – 15 kb (17). The size distribution may reflect selection because the expression of certain genes is affected by the total size of the vector plasmid (C.M.Berg, personal communication).

2.2 Recombination between direct repeats

A single reciprocal cross-over between direct repeats within one DNA molecule generates two complementary smaller circular DNAs, each with a copy of the repeated sequence (*Figure 1B*). This is the major cause of reversion of Hfr strains of *E.coli* to F^+ strains. Crossing-over involving other repeated DNAs can generate autonomous F′ plasmids in which chromosomal DNAs that had been adjacent to the inserted F remain linked to the autonomous F factor. The chromosomes in these F′ strains carry a complementary deletion (18,19). Equivalent deletions elsewhere in the chromosome must also arise by recombination, but they will not be detected if the product lacks a replication origin but contains genes needed for bacterial viability.

2.3 Unequal intermolecular crossing-over

A single cross-over between different copies of a directly repeated segment on different DNA molecules, or as drawn here (*Figure 1C*), in the two arms of a replication fork (unequal crossing-over), can lead to chromosomes with complementary duplications and deletions of the interstitial segment (the immediate product, upon completion of replication, is a dimeric molecule: a second cross-over can regenerate twin monomeric DNAs, one with the tandem duplication, and the other with the deletion). Further unequal crossing-over coupled with selection (e.g. for enhanced expression of a gene within the duplication) can lead to extensive amplification of the segment. Such amplified DNAs are inherently unstable; their extended homology makes them much more prone to become haploid by recombination than to arise in the first place (20).

2.4 Recombination between inverted repeats: 'permitted' and 'forbidden' chromosomal inversions

A single cross-over between inverted repeats reverses the orientation of the interstitial segment, but does not change the size or content of the complete DNA molecule (*Figure 1D*). Small inversions in the *E.coli* chromosome were easily detected in early studies by screening for derivatives of an Hfr strain which transferred chromosomal markers near its tail-end with extraordinarily high frequency (18). In contrast, a specific inversion of about one-fifth of the chromosome was not found after selecting recombination between a pair of mutant *lac* operons that were

inverted and in different chromosomal locations (21,22). Inversions were also rare among illegitimate recombinants obtained by selection for fusion of a mutant promoterless *his* operon to other transcription units (23).

Recent systematic searches for inversions as products of homologous recombination have reinforced the view that many of them are in some way 'forbidden' (24–26). In these experiments transposons were used to place different mutant alleles of *lac*, *his*, or *tet* genes in opposite orientations at different chromosomal locations. Any Lac$^+$ (or His$^+$ or Tet^R) recombinant formed by a single reciprocal cross-over would carry a large inversion. The desired inversion derivatives were distinguished from non-inversion strains (that may have resulted from gene conversion) by testing for altered linkage of markers spanning the cross-over site (e.g. c and d in *Figure 1D*) (25), or by testing for an altered pattern of Hfr marker transfer in cross-streaking plate tests (24,26).

The distribution of 'permitted' and 'forbidden' inversions is generally correlated with the pattern of chromosome replication, which proceeds bidirectionally from an origin site to a termination region halfway around the chromosome. Most anticipated inversions with one endpoint in each replication arm (inversions of the origin relative to the terminus) were recovered, whereas many inversions entirely within one replication arm were not (25,26). Some inversions isolated on minimal medium proved to be lethal on rich medium. Others were so deleterious that they could be detected only in transient mating assays by an inverted pattern of Hfr transfer from young cultures, and were overgrown by fast-growing derivatives which had undergone an additional inversion and regained the standard gene arrangement (26).

A variety of models have been invoked to explain why certain inversions seem to be 'forbidden'. Some of them will be mentioned here.

(i) Fitness or viability might depend on the relative dosage of certain key genes (25). Chromosomes are replicated continuously in rapidly growing cells, often with reinitiation of a new replication cycle before the previous cycle is complete, so that during rapid growth origin-linked genes tend to be 2- to 4-fold more abundant than terminus-linked genes. The lethality of inversions within one replication arm could reflect a deleterious effect of changes in the dosage of certain key genes or gene products.

(ii) Fitness or viability might depend on the relative time of expression of certain key genes in the cell cycle. Transcription promoters that contain GATC (Dam methylation) sites are generally more active when hemi-methylated (just after passage of the replication fork) than when fully methylated (27,28). The lethality of certain inversions might thus reflect a change in the relative timing of expression of specific Dam-regulated genes.

(iii) The bacterial chromosome might be folded in a highly ordered structure by the binding of specific chromosomal loci on some sort of scaffold in a defined order. Inversions that disrupted the

arrangement of chromosomal loci on this scaffold might be lethal (26).

(iv) Sequences at certain sites might be 'excluded' from the kind of single reciprocal cross-over that efficiently generates inversions by a mechanism that reflects the structure or replication pattern of the folded chromosome (25).

(v) Finally, the failure to recover some inversions and the rich-medium sensitivity and impairment of growth caused by others might reflect a transient cessation of replication each time RNA polymerase collides with the DNA replication machinery (29). In support of this model, the orientation of most very frequently transcribed genes matches that of replication fork movement whereas the orientation of most other genes is random with respect to replication. For example, 98 of 106 genes that specify components needed for protein synthesis are aligned with the direction of replication. The eight exceptions are small isolated genes which, because of their size, are probably often free of RNA polymerase (29). Many of the inversions that were sought but not recovered are also in accord with this model, although some inversions that apparently have been recovered (26) ought to have been lethal according to the polymerase collision model. At this point it seems that these apparent exceptions might reflect additional undetected rearrangements that compensated for the initial inversion.

2.5 Homologous recombination and phase variation: *Neisseria gonorrhoeae* pilin genes

Low levels of heterogeneity for certain surface structures in single clones are widespread in prokaryotes. Such heterogeneity may help them cope with variable and often hostile environments, for example immune responses in the animal or human host, or invasion of a population with virulent phage. The pili of *N.gonorrhoeae* provide the best known case of phase variation due to homologous recombination. Pili help bacteria attach to host mucosal surfaces; they are highly antigenic and in *Neisseria* are extremely variable in protein structure (30). Variants with new antigenically distinct pilins generally constitute 0.1 – 1% of the cells in young cultures grown *in vitro* without immune selection, and often predominate among re-isolates from infected humans.

Detailed analyses at the DNA level have shown that pilin variability is due to a recombinational scrambling of DNA sequences contributed by members of a large and divergent '*pil*' (pilin) gene family. Typical strains contain one or two copies of an expressed *pil* gene and multiple copies of divergent, but silent, and often truncated *pil* genes. Variant subclones arise by replacement of part of an expressed *pil* gene with the corresponding segment from a silent and partially homologous *pil* gene (30; *Figure 2*).

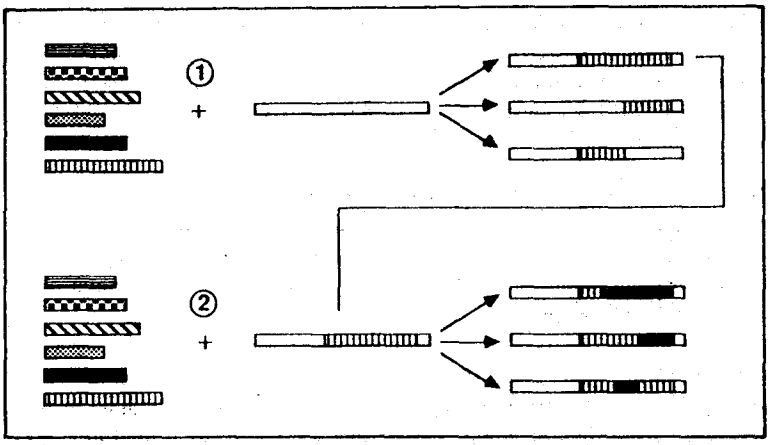

Figure 2. Pilin phase variation in *Neisseria gonorrhoeae*. Serial recombination events in regions of homology between divergent pilin genes (indicated by open and variously filled boxes) at silent loci and at the expressed locus generate a great variety of chimeric expressed pilin genes. Two successive recombination events, 1 and 2, are shown. (Reprinted from ref. 30, with permission.)

Initial studies suggested that allele replacement was not due to reciprocal recombination because in the variant strains the apparent donor locus did not contain sequences removed from the expressed locus; rather, sequences from the expressed locus were simply lost from the cell lineage. This seemed compatible with models of site-specific or transpositional recombination (31). *Neisseria* tend to lyse during growth and are also highly transformable, and it has been concluded that most *pil* gene variability is due to classical recombination between transforming DNA and endogenous chromosomal sequences, rather than between two loci in the same intact chromosome (32,33).

(i) Switching of pilin types is *recA*-dependent, unlike most site-specific recombination.

(ii) Reconstruction experiments using pilin genes marked with inserts of an antibiotic resistance gene showed that silent (incomplete) as well as expressed loci can undergo replacement.

(iii) This variation is blocked by use of mutants that, although recombination proficient, are defective in DNA uptake during transformation, or by the addition of DNase to donor cells prior to mixing them with recipients (30,32,33).

3. Site-specific recombination

Examples of site-specific recombination reactions in prokaryotes include phage λ integration and excision, resolution of transpositional

co-integrates, circularization of linear DNAs, monomerization of plasmid dimers, DNA inversion causing phase variation, and rearrangements implicated in cell differentiation. Each of these events entails protein-mediated DNA breakage at precisely defined sites and rejoining of broken ends without DNA synthesis.

3.1 Integration and excision of phage λ

The integration of λ into the *E.coli* chromosome and λ prophage excision provide well understood examples of how site-specific recombination can be sensitively regulated and, in turn, can affect key steps in a developmental pathway.

3.1.1 General features and an overview of λ

Integration and excision (***Figure 3***) constitute important stages in both lysogenic and lytic λ development (for reviews see refs 34 – 36). Integration results from a single reciprocal cross-over between a 240 bp 'attachment' (*att*) site in the phage (designated PP′ in ***Figure 3***) and a 25 bp *att* site in the bacterium (BB′ in ***Figure 3***). It is mediated by the λ-encoded integrase protein (Int) plus 'integration host factor' (IHF), which is composed of two host proteins. Int and IHF bind cooperatively at several places within PP′ and Int also binds BB′. Int protein makes a 7 bp staggered cut in a 15 bp segment that is identical in both *att* sites, twists the DNAs, and rejoins free 3′ and 5′ ends, thereby achieving integration. λ prophage excision also involves a single reciprocal exchange between the hybrid BP′ and PB′ sites. This is mediated by Int and IHF plus the λ-encoded excisionase (Xis) protein. At low Xis concentrations excision is facilitated by the host-encoded Fis protein. The complex containing Xis, as well as Int and IHF, is specific for the hybrid BP′ and PB′ sites formed by integration; it cannot recombine the PP′ and BB′ sites formed by excision.

There are many different phages related to λ including 21, φ80 and 434 of *E.coli* and P22 of *Salmonella typhimurium* (37 – 39). They resemble λ in the relative locations of genes and sites for functions such as integration, prophage immunity, DNA replication, and DNA packaging, but many of the corresponding segments differ in sequence and in the specificities of the proteins they encode. The lambdoid phages are distinct from P1 (which is normally maintained in lysogens as a non-integrated plasmid; 40), the mutator phage Mu (41; Section 4), and P2 (42), which belongs to another family of temperate phages.

The various lambdoid phages were initially distinguished from one another by immunity. A specific lambdoid phage will not grow on or lysogenize cells that already carry a prophage of the same immunity type (homo-immune), but each can grow on or lysogenize cells that carry a prophage of another immunity specificity (hetero-immune) (43), and can recombine with phage λ in regions of homology.

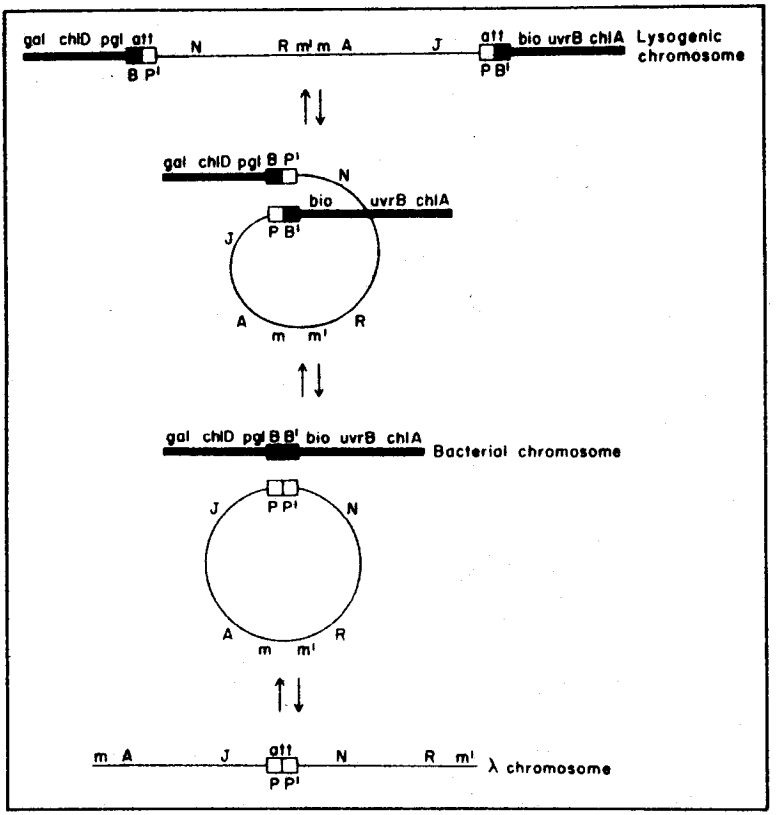

Figure 3. Integration and excision of phage λ. A, J, N, and R are phage genes. PP′ and BB′ are the phage and bacterial attachment sites. m and m′ (often designated *cos* sites) are the ends of the λ DNA in virus particles. *gal*, *chlD*, *pglB*, *bio*, *uvrB*, and *chlA* are bacterial genes. λ DNA as found in phage particles is shown at the bottom. (Reprinted from ref. 36, with permission.)

Integration reactions are also highly phage-specific. For example, the Int protein of λ does not act on φ80, nor, conversely, does Int of φ80 act on λ. Each lambdoid phage also integrates preferentially at a specific site: the *att* site for λ is between *gal* and *bio*; the *att* site for φ80 is one-tenth of the bacterial genome away, near *trp*; the *att* site in *S.typhimurium* for P22 is at a third locus, near *pro*. Phage 434 integrates at the site used by λ although these two phages are hetero-immune, which underscores both the homology between the *int* genes of λ and 434, and the separate determination of integration and immunity specificities (38,39).

Recombination among different lambdoid phages is quite common. Interspersed with dissimilar or divergent sequences are homologous DNA