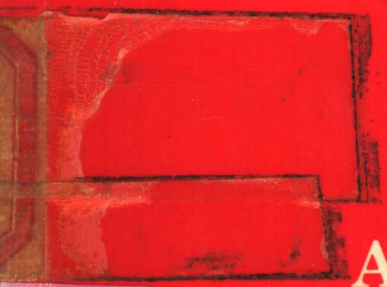
Genetic Engineering

3

edited by

Robert Williamson



Academic Press

A Subsidiary of Harcourt Brace Jovanovich, Publishers

London New York

Paris San Diego San Francisco Sao Paulo

Sydney Tokyo Toronto

Genetic Engineering 3

Edited by

Robert Williamson

Professor of Biochemistry,
St Mary's Hospital Medical School,
University of London

ACADEMIC PRESS · 1982

A Subsidiary of Harcourt Brace Jovanovich, Publishers

London · New York · Paris · San Diego
San Francisco · São Paulo · Sydney · Tokyo · Toronto

ACADEMIC PRESS INC. (LONDON) LTD
24/28 Oval Road,
London NW1

United States Edition published by
ACADEMIC PRESS INC.
111 Fifth Avenue,
New York, New York 10003

British Library Cataloguing in Publication Data

Genetic engineering. — Vol. 3
1. Genetic engineering
575.1 QH442

ISBN 0—12—270303—0

LCCCN 80—41976

Printed in Great Britain at the Alden Press
Oxford London and Northampton

Preface

Like many genetic engineers, I have recently been receiving the attention of various venture capital companies, international drug houses and Members of Parliament. I will not discuss which of these approaches are most welcome, but it did cause me to consider the speed of advance in genetic engineering, and the implications of this rapid growth. There were few who anticipated it — only five years ago, most scientists thought applications would come at the end of the century, yet we see products such as insulin and interferon already available for clinical testing.

In Europe in general and Britain in particular, this explosive growth in our own field has coincided with a general industrial depression and a marked reduction in funding for biomedical research. The brain drain from Britain is a serious matter, for we are losing the best of our younger scientists, on whom we would rely to train the next generation of molecular biologists. These volumes have come from British labs (mostly because I happen to be based in London, and my contacts and friends are here), and I feel that the quality of the contributions also shows that our current research is of a high standard. It has been based on a scientific and social philosophy of service to both the scientific and the wider community. I hope that the entry of big money into the field will not distort this philosophy, nor in the end destroy the ethics of the scientific community that has given birth to the new genetics.

This volume continues the policy of its predecessors. There are three major articles, from Russell Thompson on plasmid vectors, Bill Brammar on phage vectors and Peter Rigby on the expression of cloned genes in eukaryotic cells using vectors based on viruses and similar systems. In addition, Kay Davies has prepared a list of all recombinants containing eukaryotic genes, as of October 1981. We hope to update this occasionally, although I expect that computerized lists available on line via satellites will eventually supplant it.

Those who have read the first two volumes of this series seem to enjoy them and use them — I hope the same will be true for this

effort. When Volume 4 joins the series, the set will represent a complete description of the state of the art of genetic engineering at this time, suitable for those learning about the field, entering it for the first time, or working in it actively.

London, 29 November 1981 *Bob Williamson*

Contents

Vectors based on bacteriophage lambda

W.J. Brammar

Expression of cloned genes in eukaryotic cells using vector systems derived from viral replicons

P.W.J. Rigby

A comprehensive list of cloned eukaryotic genes

K. E. Davies

Plasmid and phage M13 cloning vectors

RUSSELL THOMPSON

Institute of Virology, University of Glasgow, Glasgow, UK

I Introduction

The essence of DNA cloning is the joining of a stretch of DNA of interest to a vector molecule which serves to propagate that DNA segment in bacteria. Vectors for *Escherichia coli* are derived from the natural phages and plasmids of this organism. Plasmid vectors have figured in all of the early achievements of recombinant DNA technology, the cloning and synthesis in *E.coli* of insulin and other hormones, interferon and animal virus antigens. To a large extent this reflects the wide range and versatility of the plasmids available for DNA cloning, a range that is continually expanding. The purpose of this chapter is to review the plasmid vectors currently available for cloning DNA in *E.coli* cells and to provide sufficient background material to enable the reader to follow the constant improvements in vector systems. Considerable effort has been directed at making *E.coli* cells an efficient source of the protein encoded by the cloned segment; vectors designed to allow expression of cloned genes will be discussed in detail in a separate chapter (see Carey, this series, Vol. 4) and they will be dealt with fairly lightly here. Plasmids which expand the cloning range to host cells other than *E. coli* will be discussed and finally, the special advantages of using the single-stranded DNA phages, such as M13, as cloning vectors will be described.

The application of DNA cloning techniques to the study of plasmids themselves has led to a considerable increase in our understanding of several fundamental plasmid properties such as replication, partition and copy number control. To provide a background for the more detailed discussion of plasmid vectors we will first turn to these more general topics.

II Bacterial plasmids

A Plasmid genes

Plasmids are extrachromosomal, self-replicating and stably inherited nucleic acid molecules. All plasmids so far isolated from bacteria have been molecules of double stranded circular DNA. Their stable inheritance suggests that plasmids may code for functions involved in their *replication* and *segregation* into daughter cells at cell division. Indeed genes involved in these two processes would seem to be the sole requirement for a piece of DNA to exist in the plasmid state and cryptic plasmids, which have no detectable phenotypic effect on their host cells, have been found (Kretschmer *et al.*, 1975). The vast majority of plasmids however carry many more genes than the

minimum required for maintenance within a bacterial cell. Some of these genes confer on the host cell properties of medical and economic importance such as antibiotic or heavy metal resistance, virulence, toxin production and the ability to degrade exotic organic compounds. An extensive discussion of these properties, as well as much other background information can be found in the books of Falkow (1975) and Broda (1979).

Naturally occurring plasmids have been modified by *in vivo* and *in vitro* genetic manipulations to improve their usefulness as vectors for DNA cloning. The replication and transfer properties of plasmids are central to these improvements.

B Plasmid replication in *E.coli*

Three aspects of plasmid replication which relate to their use as cloning vectors are:

(a) the number of plasmid copies per chromosome
(b) the size of the region essential for replication and partition
(c) the phenomenon of plasmid incompatibility

1 *Copy number*

Plasmids are maintained at characteristic copy numbers relative to the host chromosome. There is a continuous spectrum of copy numbers but it is convenient to define two groups: the low copy number plasmids present at a level of 1—5 copies per chromosome and the multicopy plasmids present at 15 or more copies per chromosome. Plasmid mutants which affect the copy number have been isolated from both classes showing the involvment of plasmid functions in copy number control. Gustafsson and Nordström (1978) isolated temperature sensitive and amber copy mutants of the low copy number drug resistance plasmid R1—19. The mutants have an elevated copy number under non-permissive conditions and exhibit an increase in their level of ampicillin resistance corresponding to the increased gene dosage of the β-lactamase gene on the plasmid. A second type of copy mutant, isolated by Nordström's group is a thermosensitive runaway replication mutant. Uhlin *et al.* (1979) have made derivatives of these mutants into useful cloning vectors which after a temperature shift replicate rapidly such that within a short period the plasmid DNA represents 75% of the total DNA. This runaway replication is lethal to the cell.

Copy mutants have also been isolated from the derivatives of the multicopy plasmid ColE1, a replicon from which several useful vectors have been derived. Gelfand *et al.* (1978) isolated a copy

mutant which was present at the level of 30% of total intracellular DNA compared to 5% for the parental plasmid. A spontaneous deletion derivative of the mutant was still maintained at a level of 30% of the total DNA and since the plasmid DNA was smaller the copy number of the deletion mutant must have increased to maintain the same plasmid DNA level. This observation led Gelfand *et al.* (1978) to suggest that copy number is regulated by a plasmid specific factor that represses replication. The characteristic copy number is determined perhaps by the affinity of a repressor for its binding site, as had been suggested by others (Pritchard *et al.*, 1969; Cabello *et al.*, 1976). In the absence of this repressor, copy number would increase until host-encoded functions required for replication, such as DNA polymerase, became limiting. Direct evidence for a ColE1 coded replication repressor such as the isolation of an amber suppressible copy mutant has so far not been obtained. But consistent with the idea of negative control of ColE1 copy number is the finding by Shepherd *et al.* (1979) that the ColE1 elevated copy number mutant is recessive and falls when a wild-type ColE1 plasmid is present in the same cell. They have mapped a 2 kb region spanning the replication origin and presumably coding for the replication repressor which can suppress the DNA overproducer phenotype of the copy mutant in *cis* or in *trans*.

Johnson and Willetts (1980) have reported a bacterial strain which can stably accommodate 39% of its total DNA as plasmid. Thus the limiting level of plasmid DNA in a viable cell may be 30—40% of the total DNA. The manipulation of copy number allows manipulation of the dosage of cloned genes; the effect of this on the expression of the gene products will be discussed in a later section.

A finding of relevance to those interested in growing plasmid-containing cultures on an industrial scale is that copy number can fall during nutrient-limited growth. Jones *et al.* (1980) grew a ColE1-containing strain in a chemostat under conditions of glucose or phosphate limitation and found that during 80 generations the plasmid content of the cells fell five-fold. The ColE1 plasmid contained a transposon Tn*1* insertion coding for ampicillin resistance. Subculture in media containing ampicillin could reverse the drop in copy number showing that the decrease was not due to selection of mutants with a lower copy number but rather was a phenotypic change in response to nutrient limitation.

2 *The minimal replicon*

Small size is a desirable feature in a cloning vector. It maximizes the ratio of passenger to vector DNA and simplifies restriction digest

patterns such that mapping the cloned segment and isolating fragments for sequencing are easier. These considerations lead to the question of what is the minimum component of a plasmid that can direct its own replication. Many cloning vectors have been derived from the small multicopy plasmids ColE1, pMB1 and P15A, which share the properties of continued replication in the absence of protein synthesis and dependence on DNA polymerase I (for reviews of plasmid replication see Kolter and Helinski (1979) and Staudenbauer (1978)). A 580 bp fragment from the replication origin region of pMB1 contains all of the genetic information necessary for replication as a plasmid in *E.coli* cells (Backman *et al.*, 1978). The transition point from primer RNA to DNA has been mapped for the closely related plasmid ColE1 (Bird and Tomizawa, 1978). It is 13 bp from one end of the 580 bp fragment showing that no information is necessary downstream from the origin. At the other end of this fragment is a region which is transcribed *in vivo* to yield a 100-nucleotide transcript. Backman *et al.* (1978) have proposed a nomadic primer model in which this transcript is processed and migrates to the replication origin where it can act as a primer for DNA synthesis. However, Oka *et al.* (1979) have isolated ColE1 derivatives which lack the nomadic primer region so that the source of the primer RNA remains obscure. The 580 bp fragment, which has been completely sequenced, contains no obvious sequence which might code for a polypeptide. This, together with experiments which indicate that no ColE1 encoded protein is needed for replication (Donoghue and Sharpe, 1978; Kahn and Helinski, 1978) suggests that the plasmid supplies a stretch of DNA which is recognized as a replication origin and that host enzymes are solely responsible for the replication reactions.

Such a simple picture is not the case for all plasmids. The antibiotic resistance plasmid R6K is 38 kb in size and a 2 kb segment from it is stably maintained in *E.coli* at the same copy number as the parental plasmid. The 2 kb fragment contains a gene *pir* coding for the π protein, which is essential for R6K replication, and an *ori* region which functions as an origin of replication (Kolter *et al.*, 1978). Kolter and Helinski (1979) have proposed a model in which the π protein has a dual role firstly as a positive element regulating the frequency of initiation of replication of the R6K origin and secondly as a negative element regulating its own synthesis. The π protein is proposed to regulate its own synthesis by binding to the nucleotide sequence repeats in the operator region of the *pir* gene and repressing transcription.

Replication alone is not enough to ensure stable inheritance of plasmid DNA molecules within a growing bacterial population. The

plasmid molecules must be segregated accurately into the daughter cells at cell division. To accomplish this, plasmids have a stretch of DNA which is functionally equivalent to the centromere of eukaryotic chromosomes. This insight has come from recent experiments of Meacock and Cohen (1980). They have identified a locus, designated *par* for partition, that is required for stable plasmid maintenance. The *par* locus of pSC101 lies in a 270 bp segment adjacent to the replication origin but is not directly associated with plasmid replication functions. Partition defective plasmids which lack a *par* locus can be maintained in a population by continuous selection, for instance for expression of a plasmid coded drug resistance gene. On removing the selection, however, the plasmid is slowly lost from the growing population and the rate of segregation of plasmid-free cells is proportional to the plasmid copy number.

Meacock and Cohen found that in the course of the DNA manipulations used to convert the naturally occurring plasmid P15A into the cloning vector plasmid pACYC184 the *par* locus of P15A had been deleted. As a result of this pACYC184 is unstable and is lost from cells cultured for long periods in non-selective medium. Cloning of the DNA fragment containing the *par* locus of pSC101 into pACYC184 can restore plasmid stability; the pSC101 *par* locus can function actively to segregate the unrelated plasmid pACYC184. This stabilization of a par^- plasmid only works in the *cis* configuration, that is a *par* locus is only active in segregation of the DNA molecule to which it is physically linked. The *par* locus is presumably a DNA site which interacts with cellular components to accomplish partitioning of plasmid DNA molecules during cell division.

In seeking to reduce the size of vector plasmids, DNA regions which were non-essential for replication and which did not contain selectable markers have been removed. As a consequence of this many cloning vectors probably are par^-. Jones *et al.* (1980) found that pBR322 and pMB9-containing cells give rise to plasmid-free segregants after about 30 generations of growth in a nutrient limited chemostat. This rate of loss is likely to be due to two factors, the lack of accurate partitioning and the drop in plasmid copy number in conditions of nutrient limitation. While the absence of a *par* locus from vector plasmids will not affect their growth on a laboratory scale, particularly if care is taken that the inoculum is 100% plasmid-carrying, it may lead to problems on an industrial scale. The problem can be overcome by simply cloning the 270 bp *par* fragment from pSC101 into the vector. Alternatively a continuous selection could be applied by incorporating an essential host gene, say a cell wall gene, into the vector and using a host that was deleted for that gene.

3 Plasmid incompatibility

It is possible to isolate cells containing any number of different plasmid types provided that the plasmids are from different incompatibility groups. Two plasmids which cannot be stably maintained in the same cell are said to be incompatible; they are members of the same incompatibility group. Naturally occurring plasmids have been found to fall into a large number of incompatibility groups (see Appendix B in Bukhari *et al.*, 1977). The many cloning vectors have been derived from a small number of parental plasmids. Of these ColE1 and pMB1 fall into the same incompatibility group so that experiments to examine the interaction of the products of genes cloned on different derivatives of these two plasmids are not practical. A third small multicopy plasmid P15A which is the progenitor of several vector plasmids is, however, compatible with ColE1 and pMB1 (Chang and Cohen, 1978). Likewise the plasmids pSC101, F and RP4 all fall into different incompatibility groups such that vectors derived from one of these plasmids are stable in cells containing plasmids derived from any of the others.

C Transfer and mobilization

Plasmid DNA molecules range from 2 kb to over 200 kb in size; this range is similar to that of organelle and viral genomes. Those larger than about 30 kb often carry a set of genes which mediate conjugal transfer of the plasmid DNA to other bacterial cells. The best studied and archetypal example of these conjugative plasmids is the F factor of *E.coli*, although many other plasmid transfer systems have been described. Interbacterial DNA transfer by conjugation is a complex process requiring the products of at least 20 transfer genes (for review see Clark and Warren, 1979; Willetts and Skurray, 1980). Plasmids that are too small to code for complete transfer systems can often be transferred if a conjugative plasmid is present in the same cell. This process of transfer of a small, non-conjugative plasmid by a coresident, large, conjugative one is termed mobilization.

Plasmid mobilization has been extensively studied using the small plasmid ColE1. It has been shown that mobilization requires both a specific site on the ColE1 DNA and ColE1-specified diffusible gene products (Warren *et al*, 1978). About one-third of the ColE1 genome or 2 kb of DNA is necessary for mobility and mobilization deficient mutants of ColE1 have been grouped into three complementation groups (Dougan *et al.*, 1978; Inselburg and Ware, 1979). Most cloning vectors derived from ColE1 or the related plasmid pMB1 have lost the DNA region which codes for the mobility proteins.

The proteins can, however, be supplied in *trans* by a compatible plasmid such as ColK. The mobility proteins probably act at a site designated *nic* (see Clark and Warren, 1979, for a discussion of this point). The ColE1 *nic* site has been sequenced (Bastia, 1978) and this sequence is conserved in pBR322 (Sutcliffe, 1978a). Several vectors have had the *nic* site deleted during their construction (see below). They cannot be mobilized and the only possible route of conjugal transmission for such plasmids is if they physically become part of a conjugative plasmid by recombination to form a fused or cointegrate plasmid. The use of recombination-deficient host strains such as *recA* strains removes this possibility, so that vectors deleted for *nic* in a *recA* host are considered more biologically "contained" than nic^+ vectors.

III General purpose amplifiable vectors

To be of use as a cloning vector a plasmid must have a unique site for one or more restriction enzymes at which insertion of DNA does not interfere with plasmid replication functions. New plasmids formed by inserting DNA fragments at these restriction sites must be capable of reintroduction into bacterial cells and cells inheriting them should be easily identifiable. This latter point usually means that insertion of new DNA at a particular restriction site must leave at least one selective marker on the plasmid intact. The first DNA cloning experiments to be carried out (Cohen *et al.*, 1973) used pSC101, a plasmid isolated from *Salmonella* (Cohen and Chang, 1977). This plasmid contains a single *Eco*RI site in a position such that cloning of DNA into this site does not affect either replication or the only marker selective for the presence of the plasmid, a gene coding for resistance to tetracycline. Since then, the trend has been to develop plasmids of minimal size that carry two or three selective markers. Often the unique cloning sites are within one or other of the selective markers such that insertion of DNA at the site inactivates the particular marker and allows ready identification of plasmids carrying DNA inserts. The early cloning vectors are described in the review by Collins (1977). More recent reviews are those by Brammar (1979), Sherratt (1979), Bolivar and Backman (1980), Bernard and Helsinki (1980), Kahn *et al.* (1979) and Timmis (1981).

A Choice of vector

Often the sole purpose of cloning a DNA fragment is to allow isolation of large quantities of the DNA in pure form. Plasmids derived from

ColE1, pMB1 or P15A are particularly useful for this purpose for several reasons. They are multicopy plasmids maintained in cells at levels of 10 or more copies per chromosome equivalent. The copy number can be amplified to as much as 1500 by treatment of the culture with inhibitors of protein synthesis such as chloramphenicol or spectinomycin (Clewell, 1972; Chang and Cohen, 1978). This allows the isolation of the plasmid DNA with yields in excess of 1 mg/litre of cells. A further 2—3-fold amplification may be achieved by addition of high concentrations of uridine (Norgard *et al.*, 1979). The P15A based vectors are however much less amplifiable than the others (Chang and Cohen, 1978). As discussed above, only a small segment of these plasmids is necessary for replication; regions outside of this can therefore be deleted during rearrangement of the cloned segment.

Table 1 lists the most useful of the multicopy vectors and physical maps are presented in Fig. 1. Most of the plasmids shown in Fig. 1 contain the tetracycline resistance gene originally found on pSC101 and the maps have been aligned using the common *Hin*dIII site in this gene. The plasmid pKC7 was constructed by replacing the small *Hin*dIII—*Bam*Hl fragment of pBR322 with a fragment containing the kanamycin resistance gene of transposon Tn*5* (Rao and Rogers, 1979). The position of the *Hin*dIII site is therefore unchanged. The only plasmid lacking the *Hin*dIII site, pACYC177, has been arbitrarily linearized from one end of the Ap^r gene.

In general it is desirable to be able to recover the cloned fragment free of vector DNA by restriction enzyme cleavage. Choice of vector is to some extent influenced by the enzyme(s) used to generate the desired fragment (but see the section on regeneration of restriction sites). If the fragment to be cloned encodes a function which can be selected, then all of the vectors with unique sites for the appropriate enzyme would be equally useful. This is rarely the case and vectors may be preferred which have restriction sites positioned such that insertion of a DNA fragment inactivates a particular gene. For example, cloning *Eco*RI fragments into pBR322 does not give a detectable change in plasmid phenotype. Thus, to distinguish within a transformed cell population between cells carrying the vector alone and cells carrying vector plus insert, properties such as the size of the plasmid DNA or the potential to hybridize to a suitable probe must be examined. In contrast, cloning into the *Eco*RI site of pBR328 inactivates the chloramphenicol resistance gene, and screening for chloramphenicol sensitive transformants thus identifies a population which is greatly enriched for plasmids carrying inserts (there will be a background of Cm^s clones with no insert which arise by aberrant recircularization of the vector).