Genetic Engineering

4

edited by

Robert Williamson

Academic Press

A Subsidiary of Harcourt Brace Jovanovich, Publishers

Genetic Engineering 4

Edited by

Robert Williamson

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

ACADEMIC PRESS · 1983

A Subsidiary of Harcourt Brace Jovanovich, Publishers

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

Contributors

R. K. Craig *Courtauld Institute of Biochemistry, The Middlesex Hospital Medical School, Mortimer Street, London W1P 7PN, UK*

R. Everett *MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, UK*

L. Hall *Courtauld Institute of Biochemistry, The Middlesex Hospital Medical School, Mortimer Street, London W1P 7PN, UK*

T. Harris *Celltech Limited, 244—250 Bath Road, Slough SL2 4DY, UK*

R. F. Lathe *Transgenè S. A., 11 rue Humann, 67000 Strasbourg, France*

J. P. Lecocq *Transgenè S. A., 11 rue Humann, 67000 Strasbourg, France*

Preface

In the two years since the first volume of this series was conceived, a remarkable change has taken place, one of which few of us foresaw. From a research technique mastered by a few hundred scientists, recombinant DNA is now a tool of virtually every contemporary laboratory in any field of biology or biochemistry: a routine method for protein sequencing, the way to determine gene number, domain evolution, chromosome localisation or family inheritance. During the next five years, it will probably become the standard technique for antenatal diagnosis, for crop improvement and for the industrial production of many chemicals, not only pharmaceuticals but also heavy chemicals and foodstuffs.

We see an absurd claim from time to time in the financial pages of the press: "the genetic engineering bubble has burst". I honestly wonder what some journalists spend their time doing! Having hyped the biotechnology field for the sake of "exciting copy", and succeeded in overestimating the potential in the *short run* of even this limitless field, they now burst their own bubble and pretend that the field is going down. The truth is that there is an enormous potential in biotechnology for research and application, but this will come to fruition over the next decade, not in the next three months. The payoff for the community, in any case, will not be measured in dollars or pounds, but in the contributions that genetic engineering makes to pure and applied clinical, agricultural, biological and industrial science.

My first Preface two years ago expressed the hope that the initial four volumes of "Genetic Engineering" would combine to give a complete primer of recombinant DNA technology. Thanks to the contributors, I think that the series comes close to achieving this object. We have covered phage, plasmid and virus cloning systems and discussed yeast, bacteria and animal cells, and looked at clinical and industrial applications as well as many fundamental uses to which genetic engineering can be put. To the extent that we have failed, the fault is mine as editor. It would have been good to have had articles about plant systems, and the industrial and clinical applications have come about faster than even I, an incurable

optomist, thought possible. I understand that Peter Rigby, who is taking over as editor after this volume, means to correct some of these weaknesses as quickly as possible.

In this volume, Len Hall and Roger Craig discuss the most recent results obtained studying polypeptide hormones, and succeed once again in introducing a great deal of up-to-the-minute data, as have all the contributors. Rick Lathe, Dick Everett and Jean-Pierre Lecocq provide a detailed review of chemical techniques for site-directed mutagenesis and manipulation of restriction sites, an article which I feel will be a fine manual for anyone starting work on the chemical side of genetic engineering. Finally, Tim Harris has written an up-to-date summary of the way in which industry is seeking to obtain the expression of cloned genes at a high level, extremely timely at the moment when both human insulin and interferon synthesised using cloned genes are undergoing clinical trials.

I found it a pleasure to edit these volumes, largely because of the commitment of the authors and the help I have received from the staff of Academic Press. I hope that the book has been useful. I commented in the first preface that the only justification for a series such as this is the fact that people find it of value for teaching and for research, and if it meets that need, I will be pleased.

London, November 1982 *Bob Williamson*

Contents

Recombinant DNA technology: application to the characterisation and expression of polypeptide hormones

R. K. Craig and L. Hall

Expression of eukaryotic genes in *E. coli*

T. J. R. Harris

DNA engineering: the use of enzymes, chemicals and oligonucleotides to restructure DNA sequences *in vitro*

R. F. LATHE and J.P. LECOCQ

Transgène S.A., 11 rue Humann, 67000 Strasbourg, France

R. EVERETT

Laboratoire de Génétique Moléculaire du CNRS, 11 rue Humann, 67085 Strasbourg, France. Present address: MRC Unit of Virology, Church Street, Glasgow, UK

GENETIC ENGINEERING
ISBN 0 12 270304 9

I Introduction

It is only during the past few years that scientists have been able to manipulate DNA in any precise manner. DNA engineering is a new technology, using methods and techniques often quite dissimilar from those of classical biochemistry. The enormous amplification of a replicating DNA segment occurring after introduction into a host cell has changed the character of DNA analysis, such that primary manipulations are now performed upon minute quantities of material. Indeed, some experimental protocols involve manipulation of a population comprising only several thousand molecules.

In the present review we discuss the advantages and limitations of certain central techniques of DNA manipulation, and we aim primarily at the clarification of technical and theoretical details. It will be noted that DNA engineering is used by two quite different schools. The fundamental scientist is more concerned with altering a DNA sequence in order to determine its function, whereas the applied scientist may be interested in manipulating genes coding for products of medical or commercial importance. In practice, both groups share the core of common techniques which form the subject of this review.

In the first sections we discuss the tools used for DNA engineering: the restriction enzymes, ligases, polymerases and nucleases which

make the manipulation of DNA possible. We stress their properties and have indicated how they have been or could be used in a sometimes quite complex manner to achieve a desired result. We go on to describe the uses of small synthetic oligonucleotides which have met increasing application in this field for they permit manipulations which involve successive transfers of DNA segments from one genetic location to another. The section on methylation of DNA illustrates how unexpected problems may be encountered during routine manipulation, and some ways of circumventing them. Finally we describe the application of these methods to the mutagenesis of DNA sequences, in particular to the alteration of a sequence lying at some distance from landmarks such as restriction enzyme sites. This approach has led in recent years to a considerable increase in our understanding of gene structure and function.

Throughout we have avoided the temptation to include a historical survey of the literature, nor do we describe experiments in the same detail as would be found in a laboratory manual. For these reasons certain older techniques have been omitted as they have been superseded by more efficient procedures. The principles, applications and limitations of techniques are presented in the hope that the reader may be able to apply these methods to any experimental problem in this field.

A Abbreviations and definitions

All nucleotide sequences are presented $5' \rightarrow 3'$ except where otherwise specified. Positions of cleavage within recognition sequences are represented by /, e.g. G/AATTC. Where the recognition sequence and cleavage site are symmetrical, only one strand is presented. ds = double strand; ss = single strand; N = any nucleotide; dNTP = any of the 4 deoxyribonucleoside triphosphates; bp = base pair(s): kb = kilobase pair(s), EGTA = ethylene glycol-bis (β-amino ethyl ether)-N, N′-tetra acetic acid; PPi = inorganic pyrophosphate; NAD = nicotinamide adenine mononucleotide; NMN = nicotinamide mononucleotide.

II Breaking and joining DNA

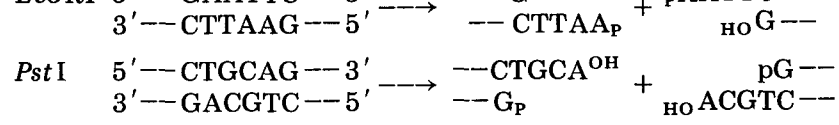
In this section we discuss the basic principles and materials used to manipulate DNA. These are the restriction enzymes and ligases which allow site-specific cleavage of DNA and *in vitro* ligation.

A Type II restriction endonucleases

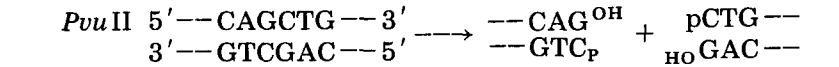
This topic has been reviewed earlier in this Series (Malcolm, 1981) and elsewhere (Modrich, 1979; Wells *et al.*, 1981) and we will restrict our discussion accordingly. In contrast to type I and type III restriction endonucleases those of class II cleave ds DNA at specific sites (for a discussion of types I and III, see Yuan, 1981). Type II restriction endonucleases recognise a sequence of bases in DNA, usually 4 to 6 in length, and cleave ds DNA at a defined position in both strands at or near the recognition sequence. Roberts (1981) presents a detailed listing of restriction enzymes, their recognition sequences and exact cleavage sites; we will mention a few examples from this list to illustrate specific points. Type III enzymes could potentially be used for specific cleavage of DNA since, unlike the type I enzymes which generally cleave at a considerable distance from their recognition sites, the cleavage sites of type III enzymes fall within a restricted number of base pairs in the proximity of the recognition site.

We will nevertheless concentrate our discussion on restriction enzymes of type II: in the body of this article we will use restriction enzyme to refer exclusively to those of the second class. The actual position of cleavage in each strand relative to the recognition sequence varies from one enzyme to the next, and the enzymes can be subclassified on this basis.

Enzymes generating only one type of terminus: enzymes of this class recognise a simple, usually rotationally symmetrical, sequence of bases and cleave within this sequence. Examples are *Eco*RI and *Pst*I which cleave as below:



The protrusions generated by cleavage with one such enzyme are self-complementary and can associate spontaneously by base-pairing under appropriate conditions ("cohesive end" cleavage). In contrast, *Pvu*II cuts both strands at the same position (blunt or flush cleavage).



The distinction between cohesive and blunt cleavage is of paramount importance to DNA engineering.

Enzymes generating non-identical termini: this subclass of restriction enzymes includes those which recognise and cleave at a group of related sequences rather than at a single defined sequence. One such enzyme is *Ava*I which recognises the sequence CYCGRG where Y represents pyrimidine (T or C) and R represents purine (A or G). Thus the sequences CCCGGG, CTCGGG, CCCGAG and CTCGAG are all cleaved by this enzyme. When an asymmetric site is cleaved by *Ava*I the two ends, although complementary, are not identical.

$$\begin{array}{l} 5'\text{--CTCGGG--}3' \\ 3'\text{--GAGCCC--}5' \end{array} \xrightarrow{AvaI} \begin{array}{l} \text{--C}^{OH} \\ \text{--GAGCC}_P \end{array} + \begin{array}{l} {}_P\text{TCGGG} \\ {}_{HO}\text{C--} \end{array}$$

A further example, *Bgl*I, cleaves DNA at the sequence GCCNNNN/NGGC. In other cases, the recognition sequence is not symmetrical and cleavage occurs at some distance to one side of the site recognised. For example, *Hga*I cleaves as shown below:

$$\begin{array}{l} 5'\text{--GACGCNNNNN NNNNNNN--}3' \\ 3'\text{--CTGCGNNNNNNNNNN NN--}5' \end{array}$$

In these cases the exact termini produced by cleavage differ between one site and the next, and this is of particular importance when DNA fragments are to be joined end-to-end (see later sections).

B DNA ligases

The basic strategies of DNA manipulation involve enzymatic cleavage of DNA at defined sites followed by end-to-end joining of the various fragments produced. This second step is normally accomplished *in vitro* through the action of a DNA ligase (for a review see Engler and Richardson, 1982) though *in vivo* ligation after transformation occurs at a detectable frequency (Chang and Cohen, 1977).

Ligation consists of the joining of an extremity carrying a 5′-phosphate group to an extremity possessing a 3′-hydroxyl, thus reconstituting a normal internucleotide phosphodiester bond. Although DNA ligase is ubiquitous in living cells, only two enzymes are commonly used to carry out this reaction *in vitro*.

T4 DNA ligase. This is the enzyme that has been most widely used for the joining of DNA fragments. The relevant gene from bacteriophage T4 has been cloned in *E. coli* (Wilson and Murray, 1979) and the elevated production of ligase from this strain (Murray *et al.*, 1979) contributes to a reduction in the levels of contaminating

activities in the purified enzyme. During the ligation reaction ATP is hydrolysed to AMP and PPi (Weiss and Richardson, 1968), while the phosphodiester bond generated derives from the 5′-phosphoryl group of the DNA substrate.

Bearing in mind the requirement for a 5′-phosphate and 3′-hydroxyl, the preferred substrate for T4 DNA ligase is a nicked ds DNA molecule. The enzyme is thus able to join cohesive termini generated by restriction enzyme cleavage, for base-pairing of the ss protrusions generates a substrate with staggered nicks in the two strands. The low temperature optimum of this joining reaction (4—15 °C, Ferretti and Sgaramella, 1981a) presumably reflects in part the stabilisation of the base-paired form.

The T4 enzyme, quite unlike other DNA ligases described in the literature, has the ability to link DNA molecules possessing "blunt" termini (Sgaramella *et al.*, 1970), although this "flush-end" reaction is some two orders of magnitude less efficient than "cohesive-end" joining (Sugino *et al.*, 1977). In addition, the T4 enzyme can carry out additional reactions such as the sealing of gaps in duplex DNA (Nilsson and Magnusson, 1982) and this low substrate specificity can lead to the generation of unexpected recombinant molecules.

It has been suggested that T4 RNA ligase is involved in the blunt-end joining activity associated with the DNA ligase (Sugino *et al.*, 1977), although the DNA ligase gene cloned in *E. coli* was subsequently shown to specify an enzyme with blunt-end joining activity (Murray *et al.*, 1979), demonstrating that this activity is an intrinsic property of the enzyme.

E. coli DNA ligase. E. coli ligase catalyses the repair of ss nicks in ds DNA but cannot catalyse end-to-end joining of DNA molecules with flush termini. The activity of the *E. coli* ligase, in contrast to the T4 enzyme, is substantially stimulated by the presence of NH_4^+ (Modrich and Lehman, 1973). In addition, the ligation reaction occurs concomitantly with the cleavage of NAD^+ to AMP+NMN rather than consuming ATP (Olivera and Lehman, 1967). The higher substrate specificity of the *E. coli* enzyme is useful in circumstances where nicks are to be sealed but where end to end joining of different molecules is to be avoided (Anderson, 1981). However, reports that the blunt-end joining activity of the T4 enzyme can be inhibited without reducing the cohesive-end joining activity (Ferretti and Sgaramella, 1981b) may lead to the increased use of the T4 enzyme in such circumstances. Strains overproducing the *E. coli* ligase have also been constructed (Panasenko *et al.*, 1978).

III Joining of fragments generated by restriction enzyme cleavage

A Basic DNA manipulations

The most central techniques of DNA manipulation are site-specific cleavage of a DNA molecule and subsequent rejoining of different DNA fragments to generate a recombinant DNA molecule. Such manipulations generally involve the specific cleavage of a DNA preparation (the "donor") with a restriction enzyme, mixing with a correspondingly cleaved vector molecule capable of autonomous replication, subsequent random end-to-end ligation *in vitro* followed by the transfer (transformation or transfection) of circular recombinant molecules into a suitable host. It must be stressed that, in many vector systems, circular molecules are necessary prerequisites for replication, and linear molecules are not normally recovered as clones.

The element of randomness during ligation reactions is particularly noteworthy, for much of the work involved during genetic manipulation lies in determining which of the "clones" produced contains the desired DNA fragments inserted in the correct position and orientation. There are, luckily, a number of strategies for ensuring that the correct hybrid molecules will predominate.

1 DNA concentration

In most ligation reactions the joining of extremity A to extremity B will be in competition with A-A and B-B fusions. Equally, in a 2 component mixture of donor (D) and vector (V), the proportion and yield of viable circular V + D molecules will depend on the absolute concentrations of both elements. Much mathematical analysis revolves about the apparent "j" value (Dugaiczyk *et al.*, 1975) of a molecule, which reflects the concentration of one end of a molecule in the immediate vicinity of the other. The shorter a DNA molecule, the more likely is it to circularise at any given concentration, and the fewer molecules which therefore participate in intermolecular ligation. To offset this problem, the concentration of the second DNA molecule is usually increased to compete this process, with the proviso that the probability of formation of linear oligomers is also increased. A more sophisticated strategy involves the ligation of only a fraction of the ends in the population, followed by dilution and religation to allow circularisation of joint molecules. Note that very small DNA fragments ($\leqslant$ 250 bp) are invariably cloned with high

efficiency for their recircularisation is impaired on steric grounds (Shore *et al.*, 1981).

2 *"Sabotage" strategies*

The cloning of DNA fragments from a biological source *de novo* is performed relatively infrequently, and routine work more often consists of the transferring of a DNA fragment, defined by flanking restriction sites, from one location to another. Parental DNA molecules can often be eliminated by "sabotage" strategies. The simplest strategy involves the identification of a restriction recognition sequence in a parental or otherwise unwanted molecule but not in the molecule required. For instance, the fusion of restriction termini may eliminate the recognition sequences for both enzymes involved. Cleavage of the products of ligation with an appropriate enzyme prior to transformation will enrich the desired recombinant.

A more general procedure is to precleave the parental molecule containing the inserted target fragment with an enzyme lacking a recognition site within the target. Dephosphorylation (see below) of these termini and excision of the target fragment prior to religation very effectively eliminates the parental form.

3 *Cloning between dissimilar sites*

In cases where a fragment to be inserted into a circular vector molecule is flanked by dissimilar restriction termini, self-circularisation of the fragment is prohibited. Equally, the vector molecule can be cleaved by 2 different restriction enzymes at adjacent sites and the central fragment removed by physical techniques, thus ensuring incorporation of the target fragment. Kurtz and Nicodemus (1981) used this technique in conjunction with plasmid pBR322 to enhance the generation of recombinants having acquired an exogenous *Eco*RI-*Sal*I fragment; greater than 95% of the viable molecules recovered had incorporated a target DNA fragment.

4 *Dephosphorylation*

A vector DNA molecule cleaved at a single site with a restriction enzyme will normally be efficiently recircularised by DNA ligase. Thus, hybrid circular molecules having incorporated an exogenous DNA fragment will be in a minority. One successful solution has been to dephosphorylate the vector with either bacterial alkaline phosphatase or calf intestinal phosphatase. Since DNA joining

mediated by DNA ligase requires a 5′ phosphoryl group, abortive recircularisation is prevented (Ullrich *et al.*, 1977); when both 5′ termini have been dephosphorylated, ligation can proceed in neither strand. However, at a junction between a donor and a vector molecule the donor is able to provide a single phosphoryl group suitable for ligation in one strand alone, and the residual nick is usually repaired after transformation by enzymes present *in vivo*. In certain cases it has been found advantageous to dephosphorylate the target DNA rather than the vector DNA (Ish-Horowicz and Burke, 1980) in order to prevent the cloning of more than one insert fragment per vector.

One note of caution: phosphatases are difficult to remove effectively and it is helpful to chelate their Zn^{2+} cofactor with EGTA and denature with phenol prior to subsequent manipulation.

B Homologous joining of restriction termini

We have discussed how the efficiency of end-to-end joining of DNA fragments is stimulated markedly by the formation of a substrate with staggered nicks in the two DNA strands. If the cohesive ends generated by cleavage do not match, then ligation is severely impaired. A fundamental technique of genetic manipulation takes advantage of the fact that certain enzymes generate identical cohesive ends even though they recognise different base sequences in DNA (Table 1). Although fusion of an *Eco*RI terminus with an identical terminus always regenerates the recognition sequence for *Eco*RI, recognition sequences are seldom regenerated when joining termini generated by dissimilar restriction enzymes.

An extension to the above technique is suggested by the finding that certain enzymes cleave at a distance from their recognition sequence. For example *Bbv*I cleaves as below:

5′-GCAGCNNNNNNNN↓
3′-CGTCGNNNNNNNNNNNN↑

Depending on the DNA sequence at the cleavage site, then cohesive ends may be generated compatible with the termini produced by a number of other enzymes. However, after site fusion by any method the recognition site remains intact (but not at both sides of the junction).

C Ordered ligation

Several restriction enzymes cleave ds DNA to give termini which are nonidentical (Section II.A). In such cases, ligase treatment of the