Tertiary Level Biology

Plant Molecular Biology

Donald Grierson
Simon Covey

Blackie
USA: Chapman & Hall, New York

TERTIARY LEVEL BIOLOGY

Plant Molecular Biology

DONALD GRIERSON, B.Sc., Ph.D.

Reader in Plant Physiology
School of Agriculture
University of Nottingham

SIMON N. COVEY, B.Sc., Ph.D.

Senior Scientific Officer
John Innes Institute
Norwich

Blackie

Glasgow and London
Distributed in the USA by
Chapman and Hall
New York

Blackie & Son Limited
Bishopbriggs, Glasgow G64 2NZ

Furnival House, 14–18 High Holborn, London WC1V 6BX

Distributed in the USA by
Chapman and Hall
in association with Methuen, Inc.
733 Third Avenue, New York, N.Y. 10017

First published 1984

British Library Cataloguing in Publication Data

Grierson, D.
Plant molecular biology.—(Tertiary level biology)
1. Plant physiology 2. Molecular biology
I. Title II. Covey, Simon N. III. Series
581.8'8 QK711.2

ISBN 0–216–91631–2
ISBN 0–216–91632–1 Pbk

For the USA, International Standard Book Numbers are
0–412–00651–0
0–412–00661–8 (pbk)

Filmset by Thomson Press (India) Limited, New Delhi
and printed in Great Britain by
McCorquodale (Scotland) Ltd., Glasgow.

Preface

The development of molecular biology has had a tremendous impact on the biological sciences over the last 30 years. Not only has it proved intellectually exciting and provided a fascinating insight into how cells work, but recent advances have been so successful that they are already forming the basis of industrial processes. Most of the early progress was made by research workers who concentrated on animals and micro-organisms, but plant molecular biology has now developed sufficiently for it to be treated as a discipline in its own right.

Plant cells contain DNA in nuclei, plastids and mitochondria, and so offer the unique challenge of studying the interaction of three separate genetic systems in a single organism. Furthermore, since plant development involves co-ordinated gene expression in response to internal and external signals, plant molecular biology can provide a fundamental insight into how development is regulated. It is also of value in breeding programmes, in understanding the interactions between plants and micro-organisms, and may suggest new ways of manipulating plant growth, development and productivity.

Despite the fact that there is considerable interest in the topic, reports of much of the work that has taken place on plants over the last few years have been confined, largely, to scientific journals and highly-priced conference proceedings. This has tended to maintain plant molecular biology as a specialist field and has hindered the wider appreciation of the subject. This volume arose from the conviction that there is a need for a short book that highlights important recent developments in this very exciting area, yet contains the background information useful to students and non-specialists. We have dealt with topics that we feel should form the 'core' of any course on plant molecular biology. We have assumed a basic knowledge of molecular genetics, biochemistry and plant and cell biology,

but have included an outline of some important methods used in the study of plant nucleic acids, together with a discussion of the prospects for plant genetic engineering. We believe that this rather limited interpretation of the scope of plant molecular biology is justified on the grounds that it is intellectually the most exciting and it is in these areas where most progress is being made. We have kept references in the text to a minimum, citing recent important papers and review articles, but access to the extensive and rapidly developing literature on the subject is provided in the detailed bibliography at the end of the book.

We are indebted to many colleagues for their invaluable advice and assistance during the preparation of the manuscript. In particular we thank Dr J. A. Downie, Dr J. Burgess, Dr N. J. Brewin, Dr C. J. Leaver, Dr J. L. Firmin, Mr T. Collins, Mr P. G. Scott and Ms K. Stallard.

Donald Grierson
Simon Covey

Contents

CHAPTER ONE

GENE CLONING, IDENTIFICATION AND SEQUENCING

Recombinant DNA technology utilizes a few simple but powerful methods which allow researchers to identify, purify and determine the structure and regulation of genes and their products. It also provides the means for transferring genes from one organism to another and, perhaps, for designing new genes. Unfortunately, this area of research has its own jargon and abbreviations which make it difficult for the uninitiated to make head or tail of what is going on. This chapter has been written with this in mind and a general outline of the procedures is given in order to help the reader interpret the experiments discussed in the rest of the book. Detailed accounts of experimental methods may be found in the excellent reviews of the subject cited at the end of the book.

1.1 Restriction enzymes

Restriction enzymes are essential for the characterization of DNA molecules and for gene cloning and sequencing because they generate small defined fragments of DNA which are easy to manipulate and study. Over 200 restriction endonucleases have been purified from bacteria. Three classes of restriction enzymes are recognized but only Type II enzymes are extensively used for genetic manipulation. The names of the enzymes are derived from the names of the bacteria from which they are purified. For example, *Eco* R1 comes from *Escherichia coli* and *Hae* III comes from *Haemophilus aegyptus*. In nature these enzymes are probably involved in the recognition and destruction of foreign DNA sequences which may enter bacterial cells. They recognize and cleave double-stranded DNA molecules at or near specific base sequences.

Restriction enzyme recognition sites generally consist of 4–6 base pairs with twofold symmetry. Cleavage can occur in the recognition sequence

Table 1.1 Restriction enzyme recognition and cleavage sites

Restriction enzyme	Recognition site	Ends created	
Eco R1	5′ GAATTC 3′ 3′ CTTAAG 5′	G CTTAA 5′	5′ AATTC G
Sma 1	5′ CCCGGG 3′ 3′ GGGCCC 5′	CCC GGG 5′	5′ GGG CCC
Hind 111	5′ AAGCTT 3′ 3′ TTCGAA 5′	A TTCGAA 5′	5′ AGCTT A
Hind 11	5′ GTPyPuAC 3′ 3′ CAPuPyTG 5′	GTPy CAPu 5′	5′ PuAC PyTG
Pst 1	5′ CTGCAG 3′ 3′ GACGTC 5′	CTGCA G 5′	5′ G ACGTC
Hinf 1	5′ GANTC 3′ 3′ CTNAG 5′	G CTNA 5′	5′ ANTC G
Sta N1	5′ NNNNNNNNNNGATGC 3′ 3′ NNNNNNNNNNCTACG 5′	N NNNNN 5′	5′ NNNNNNNNNGATGC NNNNNCTACG

itself or some distance away (Table 1.1). The DNA molecule can either be cut symmetrically to produce 'blunt ends', as with *Sma* 1, or asymmetrically to generate staggered cuts with 3′ or 5′ projections. Restriction enzymes such as *Pst* 1 and *Eco* R1, which make staggered cuts within the recognition sequence, generate cohesive or 'sticky' ends, which can base-pair with each other.

The frequency and distribution of restriction enzyme sites within a DNA molecule vary greatly with the sequence being studied. For example, in the repeat unit of the 25S and 18S rRNA genes in wheat (about 9000 base pairs long) there is only one *Eco* R1 site, whereas in the spacer region between the 25S and 18S genes (see Chapter 2) there are at least 17 *Hha* 1 sites. On the other hand, cauliflower mosaic virus DNA (the Strasbourg isolate, 8024 base pairs, see Chapter 9) has only one *Hha* 1 site but seven *Eco* R1 recognition sequences.

Restriction enzymes are generally unable to cut at their recognition sites if specific cytosine (C) or adenine (A) residues are methylated (Table 1.2). This can sometimes complicate the job of restriction enzyme mapping but it

Table 1.2 Effect of methylation on restriction enzyme site recognitions

Restriction enzyme	Recognition sequence	Methylated sequence cleaved	Methylated sequence not recognized
Eco R1	↓ GAATTC	—	GAA*TTC
Eco R11	↓CC A GG T	—	CC* A GG T
Sau 3A1	↓GATC	GA*TC	GATC*
Taq 1	↓ TCGA	TC*GA	TCGA*

can be advantageous when studying gene regulation. There is evidence that specific methylation of C residues may be involved in rendering genes transcriptionally inactive (see Chapter 3) and restriction enzymes provide a very effective means of detecting changes in methylation patterns related to gene expression.

1.2 Analysis of restriction enzyme digests

Restriction endonuclease fragments are generally fractionated by electrophoresis in agarose gels, to separate the molecules on the basis of molecular weight. The lengths of the fragments are measured by comparing their mobility with that of marker DNAs of defined length included in the gel. Only microgram quantities of DNA are needed for the analysis. The fragments can be visualized in the gel by staining with ethidium bromide and observing the orange fluorescence of the dye under ultraviolet light. If a complex genome is cut with restriction enzymes, many fragments of different lengths are generated and these produce a smear through the gel. However, repeated sequences are often present in the genome and these give rise to many copies of a particular fragment which can be seen as a distinct band. In contrast, restriction enzyme digestion of simple genomes, such as those from chloroplasts or plasmids, generally produces a relatively small number of unique fragments.

Digestion of DNA regions containing blocks of repeating sequences can sometimes lead to the production of 'ladders' when the digest is frac-

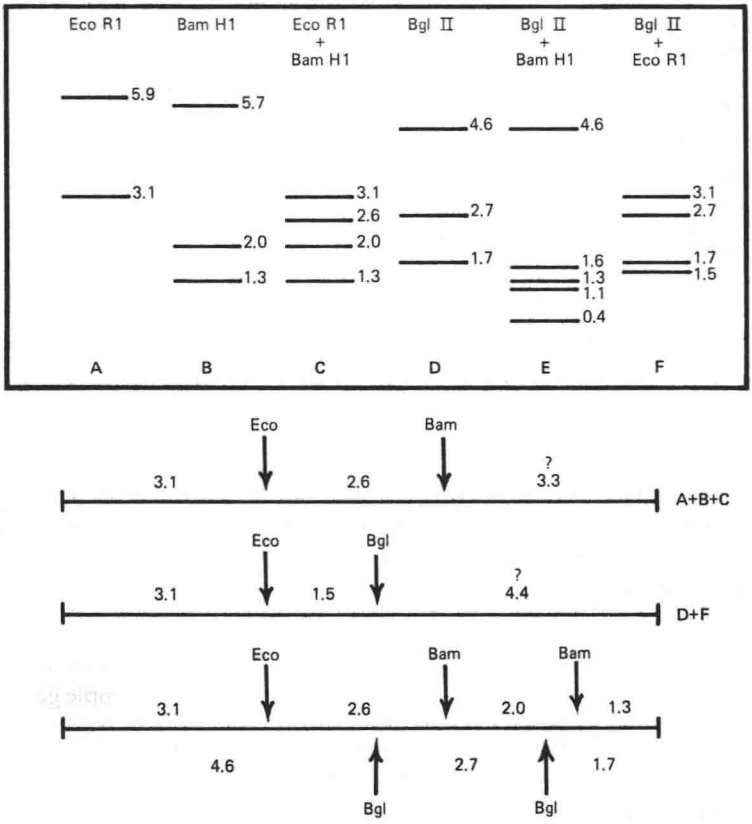

Figure 1.1 Mapping of restriction enzyme sites in a DNA segment by single and double digestions. The DNA molecule in this example is cleaved with *Eco* R1, *Bam* H1 and *Bgl* 11 in separate reactions. Further samples of the DNA are then digested with mixtures of two restriction enzymes. The fragments are separated by electrophoresis in agarose gels and visualized by their fluorescence under ultraviolet light in the presence of ethidium bromide. The molecular weights of the individual fragments, shown $\times 10^{-6}$, are calculated from the mobility of known marker DNAs fractionated in the same gel. Partial maps can easily be made from the data. For example, the relative positions of the *Eco* R1 site and one of the two *Bam* H1 sites can be deduced from the information in A, B and C. Similarly, one of the *Bgl* 11 sites can be positioned relative to the *Eco* R1 site from the information in D and F. The problem is to locate the second *Bgl* 11 and *Bam* H1 sites to the right of the diagram. The solution is found by ordering the overlapping fragments shown in B, D and E. The order is confirmed by the information in F.

tionated by gel electrophoresis. There are two ways in which this can come about. Incomplete digestion of a block of repeating sequences can lead to the production of multimers of the repeating unit; in this situation, incubation with more restriction enzyme or for a longer time results in complete digestion to the basic repeating sequence. Frequently, however, the repeating sequences are not perfect and some restriction sites are altered, by base changes, so that the enzyme no longer recognizes or cuts the DNA at a particular site. This also produces ladders but the patterns are unaffected by prolonged digestion.

A physical map of a particular DNA region can be obtained by locating internal restriction enzyme sites and determining the order of these, as illustrated in Figure 1.1. Restriction enzyme maps of large DNA molecules are constructed by first cutting them into smaller fragments and working out the order of internal restriction sites in each fragment. Digestion of the DNA with different enzymes produces a new series of overlapping fragments which allows the complete map to be assembled. Once the physical location of specific fragments is determined, they can be cut out of the DNA molecule and their properties and function investigated. For example, DNA–RNA hybridization can be used to locate regions coding for rRNA or mRNA (Chapter 2), and it can be used in coupled transcription–translation systems to locate protein-coding regions (Chapter 4) or for cloning or DNA sequencing.

1.3 Mapping of specific nucleotide sequences to individual restriction enzyme fragments

After electrophoresis, DNA fragments can be transferred directly from the gel and fixed to a sheet of nitrocellulose paper, DBM-paper (diazobenzyloxymethyl-paper), or some other suitable support, preserving the original banding pattern of the DNA. This procedure, known as 'Southern blotting' (named after the man who invented it) involves a high-pH denaturation step so that the DNA is attached to the paper in a single-stranded state. This allows the immobilized DNA to form hybrids when the paper is soaked in solutions of radioactive complementary nucleic acid probes. Large numbers of DNA fragments can be transferred to a sheet of paper and challenged in this way to locate regions coding for specific nucleotide sequences. An analogous procedure, used to transfer RNA molecules from gels to nitrocellulose or DBM-paper, known as 'Northern blotting' enables the detection and study of RNA sequences which have homology with specific DNA probes. DNA probes are generally made

radioactive *in vitro* by 'nick translation'. In this procedure unlabelled double-stranded DNA of interest is nicked with deoxyribonuclease I, to introduce single-stranded breaks in the polynucleotide chain. DNA polymerase I is then used to incorporate ^{32}P-or ^{35}S-labelled deoxyribonucleotides into the DNA at the site of the nicks. DNA probes of extremely high specific radioactivity can be generated by this means since, starting from one nick, the DNA polymerase enzyme catalyses the introduction of many radioactive nucleotides. RNA molecules are also used as probes and can be labelled *in vitro* by several methods, including the addition of radioactive ribonucleotides to the 5′ end with the enzyme polynucleotide kinase.

1.4 Gene cloning

Gene cloning provides a means of purifying and propagating specific DNA segments. Large DNA molecules are first dissected with restriction enzymes to produce specific fragments. These are then inserted into a cloning vector, which is capable of being replicated in *E. coli* or some other suitable host, by recombination *in vitro*. The chimaeric molecule, containing the vector DNA and inserted foreign DNA, is introduced into bacterial cells where it multiplies. Many copies of the inserted DNA can subsequently be recovered by purifying the hybrid vector from the cells. Cloning is generally carried out with a complex mixture of foreign DNA sequences. Thousands of recombinant DNA molecules are generated and these are replicated individually in bacterial cells to produce a 'library' of different DNA clones. The cloned DNA is not necessarily expressed in the bacteria. Special screening methods have to be devised in order to identify clones of particular interest. Cloning can be carried out in other bacteria and yeast but the systems based on *E. coli* are the most highly developed. Broadly speaking there are three types of cloning vectors for use in *E. coli* cells: plasmids, bacteriophage lambda and cosmids, discussed below.

Plasmids

Plasmids are extra-chromosomal, double-stranded, circular DNA molecules found in prokaryotic and eukaryotic cells. They carry genes for DNA replication and segregation and also frequently contain other genes for heavy metal tolerance, antibiotic resistance or the ability to metabolize exotic organic compounds. In *Rhizobium*, which forms root nodules in association with leguminous plants (Chapter 7), the genes for nitrogen

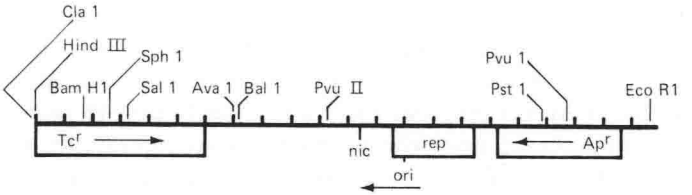

Figure 1.2 Physical map of plasmid pBR322 showing restriction enzyme sites. The map is actually circular but is shown in a linear form with the *Hind* III site at the start of the Tc^r (tetracycline resistance) gene on the left. Ap^r is the ampicillin resistance gene; *ori* is the origin of replication. The direction of transcription of particular genes is shown by the arrows. Modified from Thompson (1982).

fixation are carried on a plasmid. In *Agrobacterium*, which causes plant cancers, the tumour genes are carried on a plasmid and can be transferred to the plant genome and stably inherited (Chapter 8). Plasmids are also found in plant cell organelles (Chapter 5). Some plasmids are extremely large but small plasmids are more suitable for genetic manipulation experiments and naturally-occurring plasmids have been modified for this purpose. They are designated by 'p' for plasmid, followed by the initials of the research worker, laboratory or organism they came from, plus a strain number. One engineered plasmid which is very frequently used for cloning in *E. coli* is pBR322 (Figure 1.2). The entire plasmid DNA consists of 4362 base pairs and the complete sequence is known. It carries genes for resistance to the antibiotics tetracycline and ampicillin and has a number of single restriction enzyme sites into which foreign DNA can be inserted. Some of the restriction sites are in the antibiotic-resistance genes (Figure 1.2) and cloning into these sites inactivates these genes, a phenomenon which is useful during selection for recombinants. Plasmids are introduced into *E. coli* cells by transformation. The bacterial cells are rendered 'competent' by washing them in ice-cold $MgCl_2$ and incubating them in $CaCl_2$ overnight. Subsequently, approximately 20% of the surviving cells take up added DNA. Transformation efficiency is greatest with circular plasmids containing only small inserts of passenger DNA. For this reason vectors other than plasmids are often favoured for cloning long stretches of DNA.

Bacteriophage lambda

The lambda genome consists of a double-stranded linear DNA molecule of approximately 50 kb (50 000 base pairs), which is packaged into the

bacteriophage particle. The 5′ ends of the genome have 12-base projections which are complementary and form 'sticky ends'. This allows the bacteriophage DNA to form a circle after infection of *E. coli*. Only about half of the genes are essential for bacteriophage growth and plaque formation and consequently the remaining genes can be removed and replaced with foreign DNA. A number of restriction enzyme sites are available for doing this. The recombinant DNA molecules can be mixed with cell extracts containing bacteriophage proteins and packaged with high efficiency into phage particles *in vitro*, providing the total length of the DNA is between 78–105% of the wild type. DNA molecules outside this size range will not package and this automatically discriminates against non-recombinant DNAs and selects for recombinant molecules. Because of the limitations imposed by the packaging requirement, lambda vectors cannot be used to clone more than 18–21 000 base pairs of foreign DNA. Introduction of the recombinant DNA into *E. coli* cells (transduction) occurs by the normal bacteriophage infection mechanism and is very efficient.

Cosmids

These are hybrid cloning vehicles which combine some of the advantages of plasmids and lambda phage. Cosmid vectors contain the cohesive (*cos*) sites, or 'sticky ends' of lambda and plasmid DNA. They can be packaged into lambda bacteriophage particles *in vitro*. However, most of the lambda DNA is removed, allowing for up to 52 000 base pairs of foreign DNA to be inserted. The packaged DNA can be introduced very efficiently into *E. coli* cells by transduction. Once inside the cells the cosmid replicates as a plasmid.

Recombination in vitro

An outline of the procedure for introducing a piece of foreign DNA into a cloning vector is shown in Figure 1.3. The vector and the DNA to be cloned are cut with the same restriction enzyme to produce sticky ends. If a

Figure 1.3 Insertion of a DNA restriction enzyme fragment into a cloning vector. The vector and DNA containing the gene to be cloned are digested with the same restriction enzyme, producing identical, overlapping, sticky ends. The two DNA preparations are annealed so that the sticky ends can base-pair with one another. (Note that the association of the vector and the DNA fragment to be cloned is only one of several possible combinations). The hybrid molecules are then covalently linked (ligated) by adding T4 DNA ligase and ATP. The region of the plasmid DNA marked 'Res' represents an antibiotic-resistance gene.