INTRODUCTION TO PRACTICAL MOLECULAR BIOLOGY

Philippa D. Darbre

INTRODUCTION TO PRACTICAL MOLECULAR BIOLOGY

Philippa D. Darbre

Imperial Cancer Research Fund, London, UK

A Wiley–Interscience Publication

WILEY

Chichester New York Brisbane Toronto Singapore

Library of Congress Cataloging-in-Publication Data

Darbre, Philippa D.
Introduction to practical molecular biology.

'A Wiley–Interscience publication.'
Includes index.
1. Molecular biology – Technique. I. Title.
QH506.D36 1988 574.8'8 88-10700
ISBN 0 471 91965 9 (pbk.)

British Library Cataloguing in Publication Data

Darbre, Philippa D.
Introduction to practical molecular biology
1. Molecular biology
I. Title
574.8'8

ISBN 0 471 91965 9

Typeset by MHL Typesetting Ltd, Coventry
Printed and bound in Great Britain

Preface

Molecular biology has become a very powerful tool in many fields of biological, biochemical and medical science. It is now used very widely both in fundamental research and for diagnosis of disease. This methodology has enabled major advances to be made in medical research resulting in the better diagnosis and understanding of blood disorders, immunological deficiencies, cystic fibrosis and cancer. It has helped research into animal diseases and there is now also wide application in botanical fields. More and more scientists are discovering a need to use these techniques. Yet for many people who are not established molecular biologists, it is an awesome task to know where to start. It is for just such people that this book has been written.

This book was developed from a collection of laboratory protocols, which have been used successfully in the laboratory over the past five years. Simple protocols for the preparation of DNA and RNA are described, followed by a step-by-step procedure for DNA analysis by Southern blotting and RNA analysis by Northern blotting. Each chapter begins with a simple explanation of the principles involved. This is followed by a plan of how to fit the successive stages into a weekly schedule, a list of equipment and materials required, and finally a simple-to-follow protocol. Here lies the way for scientists wishing to take up molecular biology.

This volume is not intended to provide a major laboratory manual for the established and highly experienced molecular biologist but is an introduction to the field. It concentrates on simple protocols giving advice on the problems which might be encountered. It is assumed that genes to be used have already been cloned. Procedures are described for the amplification of cloned DNA and the preparation of probes required. No attempt is made to discuss the more complex strategies for gene cloning or engineering. For excellent comprehensive manuals, where answers to more detailed

questions will be found, the reader is referred to the books by Maniatis *et al.** and Ausubel *et al.***

The production of reliable protocols for experiments in any scientific discipline poses the problem of acknowledgements. All relevant references quoted in the text are to original procedures but it is inevitable that methods are continually being modified and improved. In this latter respect, I am indebted to my many colleagues at the laboratories of the Imperial Cancer Research Fund, London, for their advice and support.

Finally, I want to express by deep gratitude to my family who gave encouragement when it was most needed – to my husband, my daughter and my parents without whose support this book might never have been completed.

Philippa D. Darbre

* Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.

** Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) *Current Protocols in Molecular Biology*, Green Publishing Associates and Wiley–Interscience, New York, USA.

Contents

Chapter 1

Introduction

Twenty years ago, DNA was one of the most difficult biochemical macromolecules to analyse and it appeared to be a daunting task to try to understand the function of such an enormously long nucleotide structure. However, today the situation has entirely changed and DNA has become the easiest macromolecule of the cell to study. This rapid development of modern molecular biology and the so-called recombinant DNA technology has resulted from a unique combination of new discoveries (new enzymes and nucleic acid hybridisation) with old techniques (microbial genetics) and these will be described in simple terms in this book.

DNA can be easily purified from a cell and once isolated it is much more stable than many other macromolecules. It can be cut very precisely and reproducibly with restriction enzymes, enabling excision of specific pieces of DNA, which can then be obtained in essentially unlimited quantities (DNA cloning). Hybridisation techniques and rapid sequencing methods for cloned DNA have enabled determination of the structure and organisation of large parts of the genome of many organisms. It is now within the realms of possibility to obtain a sequence for the entire structure of the human genome.

RNA can also be easily purified, although being less stable than DNA, it has to be handled with greater care. With the use of cloned DNA, both qualitative and quantitative studies of any specific RNA can be made. This approach has provided a powerful tool in the study of eukaryotic gene expression.

The early years following production of the first recombinant DNA molecules did not open up a whole new era of research but rather were filled with fears about the safety of such powerful technology. At the Asilomar conference in 1975, the landmark guidelines for the use of recombinant DNA were finally drawn up. Fortunately today most forms of such research are no longer subject to any preventative form of regulation.

NUCLEIC ACID STRUCTURE

A general introduction to molecular biology must inevitably begin with a brief outline of nucleic acid structure. Both deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are chainlike macromolecules made up of strings of monomeric units called nucleotides. The monomeric units of DNA are called deoxyribonucleotides, those of RNA are ribonucleotides. Each nucleotide is composed of three components: (1) a nitrogenous base, which is either a pyrimidine or purine derivative; (2) a pentose sugar; and (3) a molecule of phosphoric acid. DNA and RNA differ both in the sugar present and in their bases. The sugar 2-deoxy-D-ribose is present in DNA, whereas D-ribose is present in RNA. The four bases present in DNA are adenine and guanine (purines), cytosine and thymine (pyrimidines). RNA has the bases adenine, guanine and cytosine but uracil instead of thymine. The structure of these five bases is given in Fig. 1.1 and the general structure of the nucleotides in Fig. 1.2.

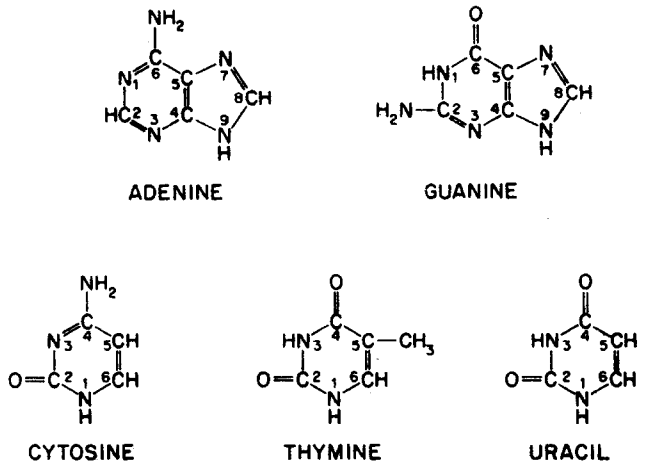


Fig. 1.1 Bases present in nucleic acids. There are two purines – adenine and guanine, and three pyrimidines – cytosine, thymine and uracil

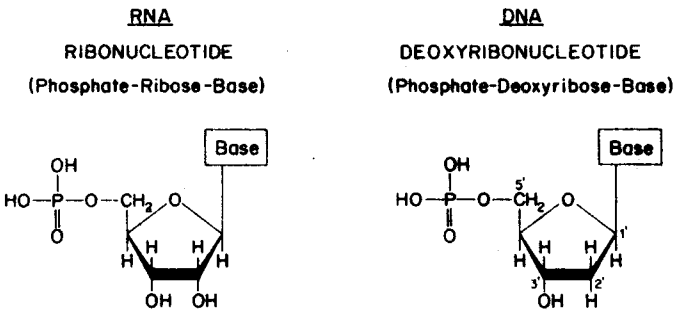


Fig. 1.2 The general structure of nucleotides in DNA and RNA

DNA consists, then, of covalently linked chains of deoxyribonucleotides and RNA of chains of ribonucleotides. The linkage is by phosphodiester bridges between the 5′-hydroxyl group of the phosphate on one nucleotide and the 3′-hydroxyl group of the sugar molecule of the next nucleotide. The backbone of both DNA and RNA is thus composed of alternating phosphate and pentose groups, with the bases acting as distinct side-chains.

In 1953, Watson and Crick made the landmark discovery of the double-helical structure of DNA. In this structure, two polynucleotide chains are wound into a helix such that the backbone strands are of alternating phosphate and sugar residues, leaving the bases projecting perpendicularly into the centre axis. The bases of the two chains are then held together very specifically by hydrogen bonding such that adenine can only pair with thymine, and cytosine only with guanine. This structure was based on X-ray data from Franklin and Wilkins, together with Chargaff's observations that in DNA the ratio of adenine to thymine and of cytosine to guanine was always very close to 1.0.

Thus, in DNA there are two chains coiled around a common axis and held together by base pairing between adenine and thymine, and between cytosine and guanine. This produces two chains which are not identical but, because of base pairing, are precise complements of each other. In addition, the chains do not run in the same direction with respect to their internucleotide linkages but rather are antiparallel. That is, if two adjacent deoxyribonucleotides T and C in the same chain are linked 5′-3′, the complementary deoxyribonucleotides A and G in the other chain will be linked 3′-5′. This structure is outlined in Fig. 1.3.

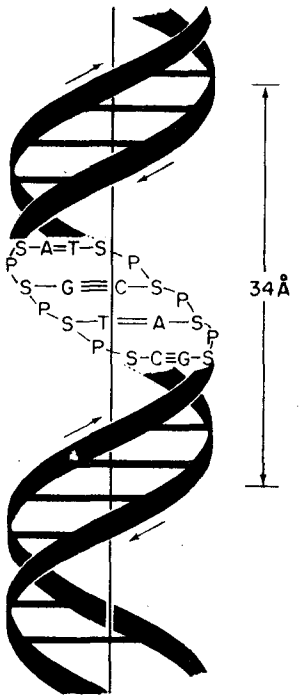


Fig. 1.3 Double-helical structure of DNA. S = sugar; P = phosphate; A = adenine; C = cytosine; G = guanine; T = thymine

For RNA, there are three major types in cells: messenger RNA (mRNA); transfer RNA (tRNA); and ribosomal RNA (rRNA). The most abundant species is rRNA, making up about 80% of the total; next is tRNA, making up 15%; mRNA is present in the smallest amounts – 5% or less. Messenger RNA is synthesised in the nucleus during the process of transcription, in which the sequence of bases in one strand of DNA is enzymatically transcribed into the form of a single strand of mRNA with complementary base sequence. mRNA is thus composed of only the four bases – adenine, cytosine, guanine and uracil. After transcription, the RNA is processed and passed out into the cytoplasm, where it then serves at the ribosomes as the template for sequential ordering of amino acids during protein synthesis. Although mRNA molecules make up only a very small fraction of the total RNA of the cell, there is a wide variety of them, differing greatly in molecular weight and in base sequence.

tRNAs are relatively small molecules that act as the carriers of specific amino acids during protein synthesis. They have similar molecular weights in the range of 23 000 to 30 000 (sedimentation coefficient of about 4S). These RNAs characteristically contain a rather large number of minor or 'odd' bases, and there are one or more tRNAs for each amino acid found in proteins.

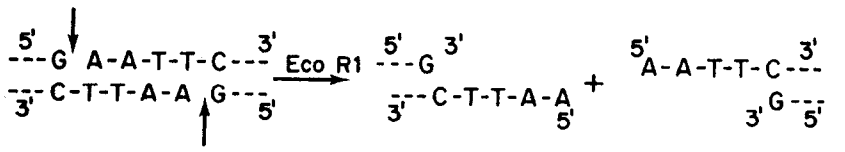
rRNA is the most abundant RNA and is found in the ribosomes of eukaryotic cells in three forms, with sedimentation coefficients of 28S, 18S and 5S.

RESTRICTION ENZYMES

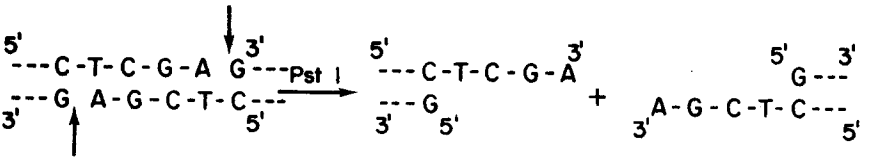
Many bacteria make enzymes called restriction endonucleases, which protect them by degrading any invading foreign DNA molecules. Each enzyme recognises a specific sequence within double-stranded DNA, which is typically four to six nucleotides in length with a two-fold axis of symmetry. The corresponding sequences in the genome of the bacterium itself are protected by methylation at an adenine or cytosine residue, but any foreign DNA molecule is immediately recognised and cut to pieces.
Many restriction endonucleases have now been purified from prokaryotes and are commercially available. The very specific cleavage of DNA by these enzymes and the availability of such a wide variety of enzymes cutting at different nucleotide sequences have been very important in the development of molecular biology.

Each restriction enzyme will cut any length of DNA double helix into a series of fragments, known as restriction fragments. However, many of these enzymes do not cut DNA exactly at the axis of dyad symmetry and thus the cutting can be broadly divided into three types:

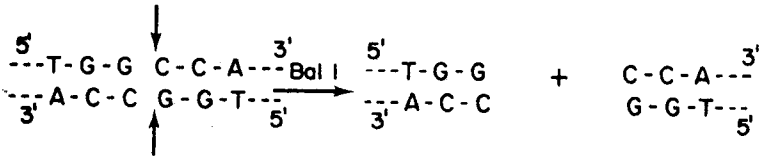
1. Enzymes which cut DNA in a staggered fashion producing restriction fragments with protruding cohesive 5′ termini.



2. Enzymes which cut DNA in a staggered fashion but in the opposite way to produce restriction fragments with protruding cohesive 3′ termini.



3. Lastly, there are some enzymes which cleave exactly at the axis of dyad symmetry to produce blunt-ended fragments.



In general, different restriction enzymes recognise different sequences, but there are some examples of enzymes isolated from different sources that cleave within the same target sequences. These are known as isoschizomers.

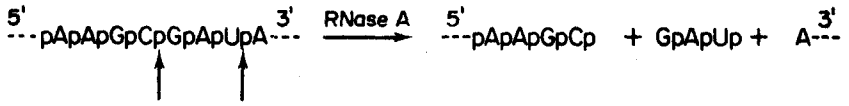
Each enzyme has a set of optimal reaction conditions, nowadays given on the information sheet supplied by the manufacturer. The major variables are the temperature and salt composition, although some enzymes are affected by methylation of the DNA, in particular of cytosine residues.

OTHER ENZYMES

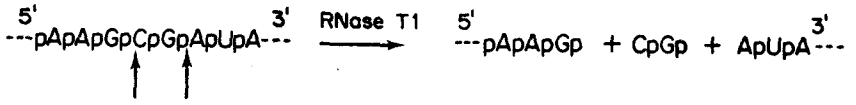
Many enzymes, apart from restriction endonucleases, are used widely in molecular biology and the actions of those referred to in this book are summarised below.

Ribonucleases A and T_1

(Used in preparation of DNA, Chapter 2.)
Ribonuclease A (bovine pancreas) is an endoribonuclease which attacks pyrimidine nucleotides at the 3′-phosphate group and cleaves the 5′-phosphate linkage to the adjacent nucleotide.

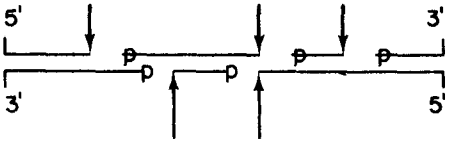


Ribonuclease T_1 (*Aspergillus oryzae*) is an endoribonuclease which attacks RNA in a similar fashion to ribonuclease A but at the 3′-phosphate group of guanine-containing nucleotides.



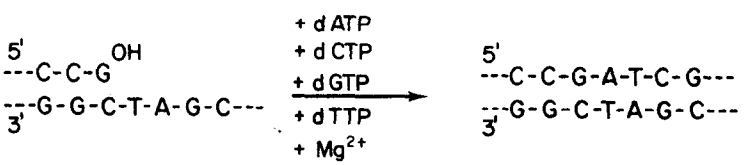
Deoxyribonuclease I (DNase I)

(Used in nick translation of DNA, Chapters 3 and 5.)
This enzyme is an endonuclease that hydrolyses double- or single-stranded DNA to a complex mixture of polynucleotides with 5′-phosphate termini. In the presence of magnesium ions, it attacks each strand of DNA independently giving a random distribution of cleavage.



E. coli DNA polymerase I

(Used in nick translation of DNA, Chapters 3 and 5.)
This enzyme carries three separate enzyme activities. It carries a 5′ → 3′ polymerase activity, adding nucleotide residues to the 3′-hydroxyl terminus that is created when one strand of a double-stranded DNA molecule is damaged, for example:



However, it also carries a 5′ → 3′ exonuclease activity and a 3′ → 5′ exonuclease activity. This makes it extremely useful in the nick translation reaction since it can both remove nucleotides from the 5′ side of the nick and add them to the 3′ side. By replacing the pre-existing nucleotides with radiolabelled (^{32}P) nucleotides, it is possible to prepare ^{32}P-labelled DNA of specific activities of 10^8 cpm/μg.

Alkaline phosphatase

(Used in labelling DNA molecular weight markers, Chapter 3.)
This enzyme can be prepared from either bacteria or calf intestine; it catalyses the removal of 5′-phosphate residues from fragments of DNA or RNA of any length, either single- or double-stranded.

$$5'_{p}\,DNA \xrightarrow[\text{phosphatase}]{\text{Alkaline}} 5'_{OH}\,DNA + p$$

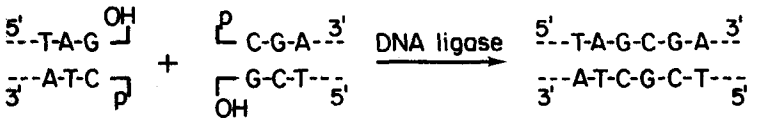
T4 polynucleotide kinase

(Used in labelling DNA molecular weight markers, Chapter 3.)
This enzyme is prepared from T4-infected *E. coli* and catalyses the transfer of the γ-phosphate of ATP to a free 5′-hydroxyl terminus in DNA or RNA, either single- or double-stranded. Since it reverses the action of alkaline phosphatase, it is possible to radiolabel any piece of nucleic acid by removing phosphate with alkaline phosphatase and adding radiolabelled phosphate with polynucleotide kinase.

$$5'_{OH}\,DNA + \gamma\text{-}^{32}P\text{-ATP} \xrightarrow[+Mg^{2+}]{\text{Polynucleotide kinase}} 5'_{32p}\,DNA + ADP$$

T4 DNA ligase

(Used in joining DNA molecules while cloning, Chapter 6.)
This enzyme is prepared from T4-infected *E. coli* and catalyses the formation of a phosphodiester bond between adjacent 3′-OH and 5′-P termini in double-stranded DNA. It can thus be used for joining together DNA molecules that are either blunt-ended or have cohesive termini sufficiently compatible to allow base pairing to one another.



DENATURATION OF DNA

The double-helical form of DNA unwinds into a single-stranded disordered structure called random coils when subjected to: (1) extremes of pH; (2) increased temperature; (3) decrease in dielectric constant by alcohols, ketones, etc.; and (4) exposure to urea or amides. During this denaturation process, no covalent bonds in the backbone structure are broken.

Two types of forces maintain the double-helical structure of DNA: (1) hydrogen-bonding between base pairs and (2) hydrophobic interactions between successive stacked bases. When either or both sets of forces are interrupted, the double helix denatures into a random coil.

HYBRIDISATION OF DNA

In 1961 it was discovered that complementary single strands of DNA will re-form double helices in a process called DNA renaturation or hybridisation. In fact, under the appropriate conditions, hybridisation can occur between any two single-stranded nucleic acid chains (DNA : DNA, RNA : RNA or RNA : DNA) provided that they have a complementary nucleotide sequence.

Use of this technique of hybridisation is now very wide. By using cloned, single-stranded, radiolabelled DNA (commonly referred to as a DNA probe), one can determine the number and structure of any particular gene in the DNA of a cell and also the type and quantity of transcribed RNA. Early studies involved the use of reactions in solution, called solution hybridisation, but now the DNA or RNA under test is usually immobilised on a nitrocellulose or nylon filter, the so-called hybridisation of Southern (DNA) and Northern (RNA) blots.

The technique of solution hybridisation was based on the kinetics of nucleic acid hybridisation. Since the rate of hybridisation is limited by the rate at which two complementary nucleic acid chains happen to collide, the concentration of specific nucleotide sequences can be measured by the rate at which they hybridise to a radiolabelled cloned DNA of complementary sequence. This is such a stringent test that sequences present as a single copy in the genome can be detected. Such studies can also be used with RNA to determine whether cloned DNA sequences are transcribed into RNA, and if so how many copies of RNA are made in a cell. Use of more elaborate hybridisation procedures (the S_1 mapping technique) enables identification of the exact region of the cloned DNA that hybridises to RNA and thereby definition of start and stop sites for RNA transcription.

Nowadays, the DNA or RNA under test is usually immobilised on a filter and hybridisation to a single-stranded radiolabelled DNA probe is carried out on the filter itself so that the end result can be visualised by autoradiography. If the DNA or RNA fragments under test are initially separated by size on gel electrophoresis before transfer to a filter, the sizes of fragments containing nucleic acid sequences complementary

to the cloned DNA probe can be determined. For RNA, this enables an instant estimation of the size of the transcribed RNA. If DNA is cut into a series of restriction fragments using different restriction enzymes before separation on the gel, a restriction map of the gene of interest can be built up. In this map, the relationship of each cutting site to its neighbours can be estimated (Fig. 1.4). Since such maps reflect the arrangement of selected nucleotide sequences, they are used for comparing normal with mutant or variant genes and also for studying homology between genes.

One of the major medical applications of restriction maps is in the diagnosis of genetic diseases. Sickle-cell anaemia results from a mutation that changes a glutamic acid residue to a valine residue at position 6 in the β-globin chain of haemoglobin. The mutation involves a single base change in the DNA. This disease was one of the first genetic diseases to be diagnosed directly at the gene level by restriction enzyme analysis of the DNA. The restriction enzyme MstII generates from normal DNA a 1.1 kilobase β-globin gene fragment, but in sickle-cell DNA this is replaced by a 1.3 kilobase fragment.

In many higher plant and animal DNAs, a significant fraction of the cytosine residues exist in a modified form in which a methyl group is attached to the 5 carbon atom of the pyrimidine ring (5-methylcytosine). It is now clear that these methyl groups are added after the DNA chains are synthesised and may play a key role in the control of DNA transcription. Analysis of restriction fragments produced by enzymes sensitive to DNA methylation (e.g. HpaII and MspI cut at the same site in DNA but differ in their sensitivity to cytosine methylation) can be useful in determining the role and position of DNA methylation in gene expression.

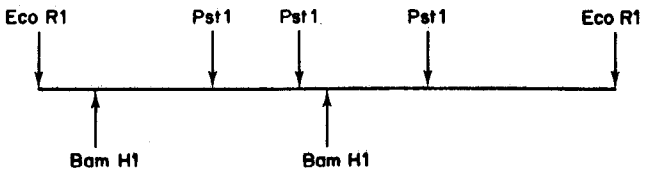


Fig. 1.4 A simple diagrammatic representation of a restriction map of a DNA fragment

DNA CLONING

In cloning procedures, a specific DNA fragment is integrated into a rapidly replicating genetic element (plasmid or bacteriophage) so that it can be amplified in bacteria or yeast cells.

There are two key features in the cloning of DNA: (1) the use of the cloning vector and (2) the combined use of restriction enzymes with DNA ligase.

1. Plasmids are small, circular, double-stranded DNA molecules that occur naturally in both bacteria and yeasts, where they replicate as independent units. Their features essential to DNA cloning are their small size (enabling simple purification) and the fact that they carry genes (which can be used for selection, such as resistance to antibiotics).

2. The combined use of restriction enzymes and DNA ligase has made it possible to graft and excise fragments of any DNA into these self-replicating elements. The staggered cleavage by restriction enzymes producing cohesive ends enables complementary base pairing between any two DNA fragments cut with the same enzyme. A circular DNA molecule, such as a plasmid, that is cut at a single site so it has cohesive ends, will tend to re-form a circle by annealing (base-pairing) of its cohesive ends. However, if a second piece of DNA cut with the same enzyme is mixed in, the two DNAs can join together. Once the ends are base-paired, they can be sealed with the enzyme DNA ligase (Fig. 1.5).

Thus, for DNA cloning, the plasmid DNA is cut with a restriction enzyme giving compatible cohesive or blunt ends to the DNA fragment to be cloned. The two DNA molecules are then joined (ligated) together using the enzyme DNA ligase. The hybrid plasmid is reintroduced into bacteria and those cells which have taken up the DNA are selected by the antibiotic resistance conferred on them by the plasmid. As these bacteria multiply, the plasmid also replicates to produce enormous numbers of copies of the original DNA fragment. The plasmid can be easily purified by virtue of its small size compared to the host cell DNA, and the copies of the original DNA fragment can then be recovered from the plasmid by excision with a second treatment with the same restriction enzyme.

The DNA to be cloned can be obtained either from cleavage of the entire genome of a cell with a specific restriction enzyme (genomic cloning) or from DNA synthesised from purified cellular mRNA using the enzyme reverse transcriptase (cDNA cloning). Cloned cDNA will contain mainly sequences coding for protein, but cloned genomic fragments may also contain DNA sequences 5′ or 3′ to the gene of interest or intervening sequences.

The ability to clone DNA and the ease with which DNA can now be analysed by hybridisation techniques and sequencing have had innumerable scientific and commercial applications. Cloned DNA can be used to search for DNA sequences or transcribed RNA in any cell, and comparisons can be made between normal and variant or abnormal situations. For example, in 1982 a cancer gene from human bladder cells was isolated and cloned in *E. coli*. The base sequence of this cancer gene was found to differ from its normal counterpart by only a single base change, leading to a single amino acid change in the protein product.