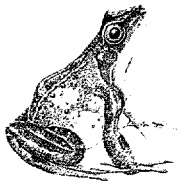


[redacted]

AN INTRODUCTION TO THE USE OF GENE PROBES IN THE BACTERIOLOGICAL ANALYSIS OF WATER

A special report produced for DoE to support Contract ET 9418

DoE 2357-M

DECEMBER 1989

AN INTRODUCTION TO THE USE OF GENE PROBES IN THE BACTERIOLOGICAL ANALYSIS OF WATER

A special report produced for DoE to support Contract ET 9418

Report No: DoE 2357-M

December 1989

Author: Dr P Gale

Contract Manager: G Stanfield

Contract No: ET 9148

DoE reference no: PECD 7/7/302

Contract duration: September 1988 - March 1991

Any enquiries relating to this report should be referred to the Contract Manager at the following address:

Water Research Centre (1989) plc, Henley Road, Medmenham, PO Box 16, Marlow, Buckinghamshire SL7 2HD. Telephone: Henley (0491) 571531

a gene of a single stranded DNA molecule.

SUMMARY

I OBJECTIVES

To provide an introduction to the techniques and theory currently used in developing gene probes to detect and quantify total coliforms or Escherichia coli in water.

II REASONS

This report has been prepared by WRc at the request of DoE to aid in understanding and provide background knowledge to work being carried out at the University of Leicester in connection with DoE contract "Development of Gene Probes For Coliform Bacteria" (PECD 7/7/302).

III RESUME OF CONTENTS

To understand the mechanism by which gene probes may be deployed to detect specific bacteria in a water sample requires some background knowledge of molecular biology. This is introduced in Section 2. In particular, the structure of DNA, which is the genetic material, is described along with how the sequence of bases contains the genetic information. In Section 3, the structure of gene probes and how they recognise and bind to their specific target sites within the bacterial genetic material is discussed. The strategy to develop gene probes for the bacterial analysis of water is outlined in Section 4. To develop a gene probe that is specific for a particular species (eg E. coli) or group (eg total coliforms) of bacteria requires the identification of sequences within the genetic material that are unique to that species or group. Gene probes designed to target these sequences will thus detect that particular species or group of bacteria. Two potential target sites for coliform or E. coli specific gene probes are the lac operon and the 16S rRNA. The advantages and disadvantages of these two target sites are summarised. One disadvantage of using gene probes is the sensitivity of the detection method. In Section 5, a method, called the

polymerase chain reaction, which may overcome the problem is introduced. Much of the research and development of gene probes requires techniques in genetic engineering. Some of these, in particular cloning and sequencing of DNA fragments are presented in the Appendices.

CONTENTS

CONTENTS - continued

SECTION 1 - INTRODUCTION

Current methods of determining the number of bacteria in water involve plating the bacteria onto agar plates and incubating for a couple of days while the individual bacteria multiply to produce colonies large enough to be counted with the naked eye. It is possible to quantify specific groups of bacteria by using selective nutrient media that only allow that type of bacteria to grow. In the water industry, one group of bacteria known as total coliforms are of particular importance in the monitoring of water quality since their detection may indicate faecal pollution and hence the presence of pathogenic agents. In this role as indicators, one member of the total coliform group, Escherichia coli, is of special public health significance.

Because of the need to ensure that water for human consumption does not contain any pathogenic organisms, analysis for coliform bacteria needs to be as rapid as possible. Despite many advances in analytical techniques, the goal of being able to analyse a water sample for total coliforms within a normal working day has always evaded microbiologists. One technique, however, which offers the potential of achieving this goal is currently being developed under contract at the University of Leicester. This technique utilises gene probes, which are produced using the relatively new technology of genetic engineering. The work at Leicester aims to produce two gene probes. One is being deigned to detect the presence of bacteria of the component genera of the total coliforms group, while the other will detect only E. coli.

Different types of bacteria differ in many ways. They have different surface structures allowing their detection by immunological techniques, they have different nutrient requirements, allowing their identification by growth on selective media, and they also have different genetic material. Indeed it is the differences in the genetic material that are directly responsible for the differences in surface structure and nutrient requirements. It is the differences in the genetic material that are targeted by gene probes. Gene probes confirm the presence and identity of a bacterium by detection of a specific portion of its

genetic material. A gene probe can be designed so that it binds to areas of communality in the genetic material of related bacteria (eg total coliforms) or to a unique area of a specific bacterium (eg E. coli). The gene probe can be detected by means of an isotopic label, eg ^{32}P, which is incorporated into the probe itself, and a technique called auto-radiography (Section 3.2).

To understand how a probe is designed and how it discriminates between genetic material from different species it is important to have a basic knowledge of the structure and function of the genetic material. A brief and hopefully simple, description of what the genetic material is and how it works is provided in Section 2. In Section 3, the structure of gene probes and how they recognise and bind to their specific target sequences are described. Section 4 introduces the two target sites currently being investigated for gene probes specific for coliforms and E. coli and presents the basic techniques deployed in the development of gene probes for bacteriological analysis of water. The more fundamental tools and methodology of genetic engineering, for example cloning and gene sequencing, are outlined in the Appendices.

SECTION 2 - BACKGROUND BIOLOGY

2.1 PROTEINS, GENES AND THE GENETIC MATERIAL

Each type of bacterium has its own particular genetic material, which contains the genetic information for the generation, growth and development of that type of bacterium. The genetic information contains all the biological instructions for a bacterium. In particular it stores the coding sequences for the various proteins produced within a bacterium. Proteins are long, linear, chains of amino acids. There are twenty different amino acids that can be incorporated into proteins. Proteins differ from each other only in the number, and sequence of amino acids. Different proteins perform different functions. Some have structural roles, while others, often called enzymes, serve as highly specific catalysts, eg converting sugar into a form of energy that can

be used by the cell. Whether a bacterium can grow on a particular nutrient depends on whether the bacterium possesses an enzyme that can breakdown and utilise that particular nutrient. This will depend on whether the genetic material contains the coding sequence for that enzyme. Thus, differences in nutrient requirements reflect differences at the level of the genetic material. The identity of the genetic material is a substance whose name is abbreviated to DNA.

DNA stands for deoxyribonucleic acid. A DNA molecule has two properties that make it suitable for carrying the genetic information. First, by virtue of its unique structure, DNA can be replicated to give two daughter DNA molecules that are absolutely identical in information content to the first. One of each of these can then be passed on to each daughter bacteria. Clearly, it is essential for a bacterium to pass on an exact copy of the genetic information to its daughter cells when it divides. Second, DNA provides a medium for storing information. A particular protein is coded for by a specific region of the DNA called a **gene**. Each type of protein has its own gene somewhere within the DNA. Thus, for example, the lacZ gene codes for a enzyme that degrades the sugar lactose. The DNA is analogous to a computer disk, with the genes being comparable with individual files. Each file, like each gene, contains a specific sequence of code, that is complete within itself. Files, like genes within a DNA molecule, are allocated to a specific region on the disk where the information for that particular file is stored. Also, genes like files, are kept separate from other genes. The total DNA present in a bacterium contains all the genes for that bacterium and is collectively called the **genome** or the genomic DNA.

2.2 THE STRUCTURE OF THE DNA MOLECULE

The DNA molecule is made of units called nucleotides. Each nucleotide consists of one of four bases, adenine, guanine, cytosine, or thymine, attached to a backbone component. These four nucleotide units are shown in Figure 1a. In Figure 1 the four bases, adenine, guanine, cytosine and thymine are abbreviated to A, G, C and T, respectively. Individual nucleotides are joined together by their backbone components to form a

single-stranded DNA chain (as shown in Figure 1b). It should be noted that the different nucleotides can be joined in any order (Section 2.3). In bacteria, the DNA is not single-stranded, but double-stranded. In double-stranded DNA, the two single strands are linked together by weak bonds (hydrogen bonds) between the bases on opposite strands. The interaction between bases is called base-pairing and is of fundamental importance. There are very precise restrictions controlling which two bases are allowed to base-pair in the DNA double helix. In fact only two base-pair combinations are allowed. These are

adenine - thymine

and

guanine - cytosine

Other base-pair combinations such as adenine-adenine, adenine-cytosine and thymine-cytosine are simply not allowed in double-stranded DNA. Adenine is said to be complementary to thymine, and guanine is complementary to cytosine. Base-pairing between complementary bases is called complementary base-pairing. Diagrammatic representations of complementary base-pairing are shown in Figure 1c. Each adenine is base-paired to thymine by two hydrogen bonds and each quanine is base-paired to cytosine by three hydrogen bonds. The hydrogen bonds are represented as dotted lines. **Complementary base-pairing is the single most important structural property of DNA.** It makes possible all the functions of DNA. It provides a mechanism for the exact replication of DNA molecules prior to cell division and for the transfer of information into the sequence of amino acids in a protein. Furthermore, many techniques in genetic engineering and gene probe technology exploit complementary base-pairing. For example, the actual mechanism by which gene probes detect a particular piece of genetic information, the construction of plasmids and the selection of sequences to amplify in the polymerase chain reaction are all dependent on complementary base-pairing.

DNA molecules may be of any length. The length of a double-stranded DNA molecule is determined by the number of base pairs. It should be noted

that there is no limit on the number of nucleotides joined together in a DNA strand and that some DNA molecules are very long indeed, with up to 100 000 000 base-pairs.

Since DNA molecules may vary in both length and nucleotide sequence, an almost infinite number of different molecules are possible. Indeed, there are ~1.6 x 10^{60} (4^{100}) possible different sequences for a DNA molecule that is only 100 bp long!

2.3 STORAGE OF GENETIC INFORMATION AND THE GENETIC CODE

The backbone of the DNA is the same throughout the whole molecule and serves to provide a framework for storing the sequence of bases. The sequence of bases within a DNA molecule varies. Any sequence can be accommodated and is permissible within a DNA molecule. Two double-stranded DNA molecules with different lengths and base sequences are shown in Figure 1d. The genetic information is stored in the sequence of bases. It is the sequence of bases within a gene that specifies the sequence of amino acids in a protein. The sequence of bases in a gene is decoded into a sequence of amino acids using the genetic code. The genetic code reads the bases in the DNA molecule in groups of three. Each group of three bases is called a codon. Thus, there are 64 (4^3) different codons. Each codon specifies one and only one particular amino acid. There are twenty amino acids. Some amino acids are coded for by more than one different codon. The amino acid leucine, for example, is coded for by six different codons. Of the 64 codons, three do not code for any amino acid, but instead are read as stop signals, indicating the end of the protein. In Figure 2, a portion of single-stranded DNA is shown. Each set of three consecutive bases constitutes a codon and specifies a particular amino acid. The corresponding amino acid sequence coded for is shown above.

2.4 REPLICATION OF DNA MOLECULES

The vital feature of DNA replication is to preserve the sequence of bases, such that the sequences of the two daughter molecules are both identical to the original DNA molecule. In this way the genetic

information and protein coding sequences are preserved and passed on without mistakes to future generations. The double-stranded DNA molecule to be replicated is separated (by breaking the hydrogen bonds between the bases-pairs) into two single-strands, each of which then serves as the template for synthesis of a new strand. Replication of the DNA molecule in a bacterium requires many enzymes in addition to a large supply of the four component nucleotides, which are synthesised in the cell. The major enzyme involved is DNA polymerase, which sequentially adds bases, that are complementary with those in the template strand, onto the end of the new chain. Thus, by way of complementary base-pairing the information for the opposite single-strand can be retrieved and so an identical copy synthesised. Figure 3 shows a DNA molecule in the middle of replication. The original double-stranded molecule has been split at the top end and new complementary strands are being synthesised using the original strands as template. A nucleotide is shown being added to each of the new chains. Note the base is complementary to that on the template. Replication proceeds sequentially towards the bottom end of the molecule until the end. The result is two identical double-stranded DNA molecules. It should be noted that each daughter molecule contains one original strand and one new strand.

DNA polymerase is not able to start the synthesis of a new DNA strand at a template from scratch. Instead, the enzyme requires what is known as an **oligonucleotide primer** from which it can begin synthesis of the new DNA strand. An oligonucleotide is a short, single-stranded chain of about 15-30 nucleotides in length. A primer does exactly what it implies, primes the system for DNA synthesis. The primer for DNA replication is synthesised by an enzyme, called primase. Thus, DNA polymerase begins synthesis of the new strand at a primer.

2.5 TRANSFER OF GENETIC INFORMATION FROM GENE SEQUENCE INTO PROTEIN SEQUENCE

The cellular process which produces a protein molecule with the correct amino acid sequence as directed by the sequence of codons stored in the gene for that protein takes place in two stages. These two stages are

called transcription and translation. The DNA in the gene is not decoded into protein sequence directly. Instead, the information containing the coding sequence for the protein is read into a single-stranded chain of nucleotides called messanger RNA, or mRNA for short. RNA is very similar to DNA, except it is always single-stranded. mRNA is synthesised on the DNA itself by the process of transcription, and requires an enzyme called RNA polymerase. The events of transcription are shown in Figure 4. Each gene contains a special sequence of bases, called the promoter, which is recognised by RNA polymerase as a start site for transcription. RNA polymerase recognises the promoter and binds to it. The promoter tells the RNA polymerase which way along the DNA to begin transcription. The DNA in the gene to be transcribed is separated into single-strand regions and RNA polymerase uses one of these strands as template to direct the synthesis of a mRNA molecule using complementary base pairing. The nucleotide base sequence of the mRNA is thus complementary to that of the gene and contains the information for the coding sequence of the protein. The completed mRNA molecule then dissociates from the DNA at the end of the gene.

The base sequence of the mRNA is converted to the sequence of amino acids in a protein by a process called translation. Translation occurs on subcellular particles called ribosomes. Ribosomes are oval-shaped structures, made of protein and another type of RNA, called ribosomal RNA, or rRNA for short. There are several rRNA molecules in each ribosome, one of which called the 16S rRNA, is of particular interest to workers at Leicester University as a possible target sequence for gene probes (Section 4.1.2). Another type of RNA molecule called transfer RNA (or tRNA) occupies the central role in translation. There are 61 different tRNA molecules each of which recognises a different codon on the mRNA. Each tRNA molecule has a region of three bases, called an the anticodon, which is complementary to a particular codon on the mRNA and is thus able to recognise and bind to that codon by complementary base-pairing. Attached to the other end of each tRNA molecule is the amino acid coded for by the particular codon. Thus the tRNA serves as the decoder, converting the codon in mRNA into the amino acid in

protein. The mRNA molecule feeds into the ribosome and is pulled through. As each codon on the mRNA goes through a special area on the ribosome, so it is presented to the various tRNAs. The particular tRNA with the complementary anticodon is bound and its amino acid added to the growing protein chain. Translation is depicted in Figure 5.

2.6 OPERONS AND THE CONTROL OF GENE EXPRESSION

A gene that is regularly transcribed is said to be transcriptionally active or expressed. Some genes in the bacterial genetic material are transcribed all the time, whereas others are inactive a lot of the time and only transcribed when changes occur in the environmental conditions. For example, if E. coli bacteria are grown in a medium that does not contain the sugar lactose, then they do not produce the specific enzymes to break down lactose and consequently cannot use lactose as a food. The genes that code for the enzymes that degrade lactose are located in what is called the lac operon. An operon is a cluster of genes which are located very close to each other on the DNA. Furthermore, all the genes in an operon are transcribed from the same promoter. Generally, all the genes in a particular operon code for proteins that function together. Thus, the proteins coded for in the lac operon are all involved in utilisation of lactose as a food. The structure of the lac operon is shown in Figure 6. It consists of three genes, lacZ, lacY and lacA, that code for enzymes that allow the bacterium to use lactose as a food source. When lactose is not available, the genes of the lac operon are not transcribed. This ensures the bacterium does not waste resources producing enzymes that have no use. In the presence of lactose, however, the genes of the lac operon are transcribed and the enzymes coded by the three genes are produced. The bacterium is now able to use lactose as a food. The lac operon is also of interest to workers at Leicester as a potential target site for a coliform specific gene probe (Section 4.1.1).

SECTION 3 - WHAT ARE GENE PROBES AND HOW DO THEY DETECT THE PRESENCE OF SPECIFIC DNA SEQUENCES?

Any diagnostic test for a pathogenic micro-organism requires the identification of a characteristic that is unique to that particular micro-organism. This characteristic can then be tested for to confirm the presence or absence of that pathogen. The characteristic detected by a gene probe is a particular sequence of bases, called the target sequence, within the genetic material. Detection of that particular sequence of bases in a water sample is confirmation of the presence of that pathogen.

It is possible to design a gene probe to target and detect a particular DNA sequence, which has been identified as unique to a particular species or group of pathogens. A gene probe confirms the presence or absence of its specific target sequence within genetic material by its ability to bind specifically to that target sequence. Genetic material containing the target sequence will bind the gene probe, while if the target sequence is not present, the gene probe will not bind. A gene probe has to fulfil two functions. First, its design should ensure that it binds specifically to the chosen target sequence and not to any other sequences. Second, it must contain a marker group of some form to provide a method of detecting whether or not it has bound successfully to the bacterial genetic material. This marker group is called a label.

The basic molecular biology of the genetic material of bacteria has been introduced (Section 2) and the scene is set to describe the structure of gene probes and the mechanism by which a gene probe recognises and binds to its particular target sequence in the bacterial genetic material.

3.1 THE STRUCTURE OF GENE PROBES

A gene probe is a single-stranded DNA molecule, which may be of any length longer than about 30 nucleotides. It is the nucleotide base sequence of the gene probe that is of fundamental importance in recognising a base sequence within the bacterial genetic material as the

target site. Gene probes with different base sequences recognise and bind to different target sequences. The process by which a gene probe recognises and binds to its particular target sequence within the genetic material of a bacterium is known as **hybridisation.** A gene probe that has recognised and bound to its particular target site is said to have **hybridised** to its target site. The label to allow detection of the gene probe after it has hybridised to its target site is usually one or more radioactive phosphate groups, which are incorporated into the backbone of the probe molecule, itself. Workers at Leicester University, however, are currently developing non-radioactive labels for gene probes, so that the probes can be used routinely by the water industry without having to take the safety precautions associated with using radioactive material.

3.2 HYBRIDISATION AND AUTO-RADIOGRAPHY

If two single-stranded DNA molecules possess complementary base sequences over part or all of their respective lengths, then these complementary sequences can undergo base-pairing (Section 2.2) and so anneal to form a region of double-stranded DNA, which effectively binds the two molecules together. This process, called hybridisation, is shown in Figure 7, and is similar in concept to the action of "velcro". To hybridise, the complementary sequences on the two DNA molecules must extend for more than approximately 30 consecutive bases.

It is possible, therefore, to design a DNA molecule to hybridise to any desired target sequence of DNA by making its nucleotide sequence complementary to that of the target sequence. This is the principle of designing gene probes that target specific sequences. If the DNA molecule that hybridises to a specific target sequence is radiolabelled then it can serve as a gene probe to detect the presence of that particular target sequence of nucleotides within, for example, the genomic DNA from a bacterium. This is depicted in Figure 7, where the circular DNA represents single-stranded DNA from a bacterium and the radio-labelled, single-stranded DNA molecule is the gene probe. The gene probe will only hybridise to sequences, which are exactly

complementary for more than 30 consecutive bases. The genomic DNA from a bacterium presents thousands of different sequences to the gene probe. If any are complementary to the sequence of the gene probe, then gene probe molecules will recognise these as target sites and hybridise to them. Successful hybridisation and binding of a gene probe, which is labelled with a radioactive isotope, can be visualised using photographic film. High energy rays emitted by the decaying ^{32}P-nuclei create a black mark on the film, which is easily seen. This technique is known as auto-radiography.

A single probe molecule cannot be detected by auto-radiography. Many rays are needed to make a mark on the photographic film large enough to be seen with the naked eye. In practice many copies of the radiolabelled probe must be hybridised in order to be detected. Thus, when the term "gene probe" is used it is taken to mean multiple copies, each radiolabelled and each identical, of the specific DNA sequence to be used. Since each copy of a target sequence is only capable of hybridising to one probe molecule, multiple copies of the target sequence are also required for auto-radiography. As discussed later, this requirement tends to limit the sensitivity of gene probe techniques in quantifying the numbers of bacteria.

SECTION 4 - STRATEGY IN DEVELOPING GENE PROBES FOR THE BACTERIOLOGICAL ANALYSIS OF WATER

The aim of the work is to develop two genes probes, one of which detects all bacteria in the group called total coliforms, while the other is specific for the species E. coli.

The strategy adopted at Leicester University is first to identify suitable target sequences within the bacterial genetic material and then to design gene probes that specifically hybridise to them. Target sequences have to be selected to meet the desired requirements of the probe. Thus, a particular sequence which occurs in all bacteria of a particular group, for example total coliforms, but not in bacteria of

other groups, presents a potential target site for a gene probe which detects the presence of bacteria in that group. Similarly, a sequence which is found only in a particular species of bacterium, for example E. coli, is a suitable target site for a gene probe that detects just that one species of bacterium. Once a gene probe for a particular target sequence is obtained, its specificity has to be evaluated to confirm that it actually does identify the group or particular species of bacteria it is designed to. Thus, a probe designed to detect total coliforms must be tested to ensure that it does indeed identify all bacteria of the coliform genera and furthermore that it does not detect any bacteria outside this group. Similarly, a probe specifically for E. coli must detect only E. coli and no other bacteria. Further development of the gene probe and revaluation may then be necessary.

This section begins by introducing two sequences identified as possible target sites for gene probes in the bacteriological analysis of water. The method adopted to obtain a gene probe that hybridises to a particular target site is then described. Also presented is the procedure by which the number of bacteria in a sample of water may be quantified using gene probes. Finally, some of the advantages and disadvantages of the two target sites currently under investigation are discussed.

4.1 IDENTIFICATION OF SUITABLE TARGET SITES FOR A TOTAL COLIFORM GENE PROBE AND AN E. COLI GENE PROBE

Researchers at Leicester University are currently investigating two sequences as potential target sites for gene probes in the bacteriological analysis of water. The first sequence, which is part of the lac operon (Section 2.6), presents an ideal target site for a gene probe to detect bacteria in the total coliforms group. The second target sequence investigated is actually an RNA molecule, called 16S rRNA (Section 2.5), the base sequence of which varies between bacteria of different species. The 16S rRNA thus offers the potential to provide a target site for gene probes specific for different species of bacteria. Indeed, it is hoped to develop a gene probe that specifically

targets the 16S rRNA of E. coli. Such a probe could serve as an E. coli specific gene probe.

4.1.1 Lac operon

The lac operon (Section 2.6) is a cluster of three genes, located in the genomic DNA. The structure of the lac operon is shown in Figure 6. The three enzymes coded for by the genes in the lac operon are required for the utilisation of lactose as a food source. Indeed bacteria without the lac operon are unable to grow on media with lactose as the only carbon source, because they are unable to use the lactose as a food. In contrast bacteria with the lac operon are able to grow on lactose only media, because they have the enzymes to degrade it. In water microbiology, the group of bacteria called coliforms are defined as those which are able to grow on lactose only media in the presence of bile salts. Thus, coliforms must have the cluster of genes called the lac operon in their genome, while non-coliforms do not. Clearly, sequences within the lac operon fulfil the requirements for a suitable target sequence for a gene probe that specifically detects the presence of a coliform.

4.1.2 16S rRNA

Ribosomes are the sites of protein synthesis. They serve as the factory where a protein is synthesised according to the instructions specified in a mRNA molecule (Section 2. 5). Each ribosome consists of two subunits, which are comprised of many proteins and several RNA molecules. The RNA molecules making up ribosomes are called ribosomal RNA, or rRNA. One of the rRNA molecules in the smaller subunit is called the 16S rRNA. It is named "16S" in reference to its density. The 16S rRNA is of particular interest as a target site for gene probes for bacteriological analysis of water because it contains sequences of bases that are unique to particular species of bacteria. Indeed, the base sequence of 16S rRNA in a species of micro-organism is used for characterising micro-organisms and establishing their evolutionary links.

The 16S rRNA molecules from different species of bacteria are ~1500 nucleotides long. Using RNA sequencing methods, described in Appendix B, workers at Leicester have obtained the exact base sequences of 16S rRNA from different species. Computer aided comparison of the base sequences obtained from different species of bacteria has identified portions of sequence within the 16S rRNA molecules in which the actual sequence of nucleotides varies from species to species. These are called variable regions and present potential target sites for species-specific gene probes, such as one specific for E. coli. Workers at Leicester are also developing 16S rRNA gene probes that are specific for the total coliforms group. This involves identifying regions within the 16S rRNA which are common to all coliforms but not present in non-coliforms.

4.2 OBTAINING A GENE PROBE THAT TARGETS A PARTICULAR SEQUENCE

A gene probe is designed to hybridise to a particular target sequence by making its base sequence complementary to that of the desired target site. The usual source of a gene probe for a specific target sequence is to produce a clone of that target sequence (Appendix C). (Remember that both the clone and the target sequences will be double stranded and of identical base sequence. Thus, on splitting both DNAs into single-stranded DNA, the single-stranded probes will also be complementary to the single-stranded target sites.)

To obtain a gene probe that specifically binds and targets the lac operon (to detect coliforms), workers at Leicester produced a clone of a portion of the lac operon. The segment of DNA coding for the lacZ protein (Figure 6) was isolated and incorporated into a plasmid to produce a recombinant plasmid as described in Appendix C. This plasmid was cultured in suitable host bacteria to produce an indefinite supply of the lacZ gene ie it was cloned. Identical DNA segments containing the lacZ gene were separated from other fragments of the plasmid by agarose gel electrophoresis (Appendix B) and labelled with radioactive ^{32}P. If the cloned lacZ gene fragment is split into single strands then

each single strand will be complementary in base sequence to one of the strands in the lacZ gene of the lac operon in a coliform chromosome. Thus the radio-labelled lacZ gene, after conversion into single-stranded DNA, serves as a gene probe to detect the presence of bacteria with the lac operon, ie total coliforms.

4.3 METHOD FOR DETERMINING THE NUMBER OF BACTERIA IN A WATER SAMPLE BY HYBRIDISATION OF GENE PROBES

The principle of detecting bacteria with gene probes is that if a gene probe binds to the genetic material extracted from a bacterium, then the genetic material must have a particular target sequence within it, which then confirms the identity and presence of a particular type of bacterium. The general technical procedure for extracting the DNA from the bacteria in a water sample such that it can be presented to gene probes for hybridisation is outlined below.

A sample of water is filtered. Bacteria trapped on the filter are then cultured on an agar medium to allow the individual bacteria to multiply. Each bacterium present thus generates one colony, which contains thousands of bacteria each with the same genetic information. To produce enough bacteria in each colony to be detected by autoradiography by gene probes for which there is a single target site in each bacterium requires a 5 hour culture period. The colonies on the filter are then blotted onto a nitrocellose filter, which is the material used for providing a solid support on which hybridisation is able to take place. Blotting is performed by placing the nitro-cellulose filter onto the filter containing the bacterial colonies. The nitro-cellulose filter is then peeled off and some bacteria from each and every colony adhere to it. Thus a replica of the pattern of colonies is obtained on the nitro-cellulose filter. Before hybridisation can take place, the bacteria have to be lysed and their double stranded DNA converted into single-stranded DNA by separation of the two strands. This is achieved by alkali treatment. Nitro-cellulose filters provide an ideal support for hybridisation because single-stranded DNA molecules stick instantly to their surface and are immobilised, which prevents the complementary

single strands of the bacterial genetic material from annealing to reform double-stranded DNA. The filter is then baked at 80 °C to permanently fix the single-stranded DNA to the surface. Furthermore, baking renders the nitro-cellulose filter unable to adsorb single-stranded DNA molecules, such as the probe to its surface. Thus, the only way further single-stranded DNA molecules, namely the radio-labelled probe, can attach to the filter is through complementary base-pairing, ie hybridisation to single strand portions of the bacterial genetic material already fixed to the filter. The filter is now ready for hybridisation. The radiolabelled double-stranded probe DNA molecules are first denatured into single-strand DNA by heat treatment and are then incubated with the filter. The pH, ionic strength, and temperature conditions are chosen to permit the single-stranded gene probe DNA molecules to hybridise to single-stranded regions of complementary nucleotide sequence in the bacterial genetic material fixed to the nitro-cellulose filter. After washing to remove unhybridised probe, the positions of gene probes hybridised to the filter can be detected by autoradiography. The position of each bacterial colony, the genetic material of which the gene probes have bound to, can thus be detected and the number of colonies counted.

4.4 MERITS OF THE TWO TARGET SEQUENCES

Both the lac operon and the 16S rRNA have their advantages and disadvantages as target sites for gene probes in bacteriological analysis of water.

Clearly, if gene probes are to be of use to the water industry in providing a rapid method of counting the number of bacteria in a water sample, they must be able to detect individual bacteria. One of the main limitations of using gene probes is the poor sensitivity, particularly in detecting radiolabelled probe molecules using autoradiography. Multiple copies of the labelled probe molecule would have to be hybridised to each bacterium at the same time to detect it. This means each bacterium would need multiple copies of the target site, since each target site can only hybridise one probe molecule. This is a

significant problem for target sequences, such as the lac operon, for which there is only one copy per bacterium. Thus, only one gene probe molecule is able to hybridise to each bacterium. Workers at Leicester University estimate that between 50 000 and 500 000 bacteria is the minimum number that can be detected using a gene probe with a single target site in each bacterium. Thus to detect individual bacteria requires a culture period of approximately 5 hours, while each bacterium forms a small colony of bacteria. Each colony then has enough copies of the lac operon to allow detection by autoradiography. For a rapid bacteriological detection procedure (ie <6 hours) it is important to reduce or if possible remove this 5 hour culture step.

One method to overcome the sensitivity problem is to chose target sequences, which are present in large numbers in each bacterium. For this reason the 16S rRNA seemed an obvious choice. There are thousands of identical copies of 16S rRNA within the same bacterium. This is because there are several thousand ribosomes in a single bacterium and in each ribosome is one 16S rRNA molecule. Thus, each bacterium has the potential of binding thousands of probe molecules. The main disadvantage of using 16S rRNA as a target site is that 16S rRNAs are present in every species of bacteria. Furthermore, some regions of the rRNA base sequence are the same in a wide range of bacterial species while the base sequence of other regions of the molecule are variable and specific to a particular species. Thus, workers at Leicester are faced with the problem of deciding which part of the rRNA sequence to target the probe. To overcome this, they are currently determining the actual nucleotide sequences of the rRNAs from a variety of bacteria. (Methods of obtaining the actual base sequences of RNA and DNA molecules are described in Appendix B.) By using computers to align and compare the base sequences obtained from the 16S rRNA molecules from different species of bacteria, Leicester University is optimistic in identifying sequences which are specific to a particular strain of bacteria. The potential of these sequences as target sites for species-specific gene probes can then evaluated.

SECTION 5 - POSSIBLE USE OF THE POLYMERASE CHAIN REACTION TO OVERCOME THE SENSITIVITY PROBLEMS OF GENE PROBES IN BACTERIOLOGICAL ANALYSIS OF WATER

As described in Section 3.2, the detection of a gene probe by auto-radiography requires multiple copies of that radio-labelled gene probe to be hybridised. This necessitates multiple copies of the target sequence for each gene probe molecule to hybridise to. Thus, bacteriological analysis using gene probes that target DNA sequences, such as the lac operon, for which only a single copy is present per bacterium, are particularly limited by low sensitivity. Indeed, it is impossible to detect individual bacteria using gene probes for such single copy sequences. As discussed in Section 4.4, Leicester have overcome this by culturing each bacterium into a small colony. Each colony contains many identical bacteria and hence multiple copies of the lac operon. Another possible way to overcome the low sensitivity associated with such single copy target sequences is by using a relatively new technique, called the polymerase chain reaction (PCR), which is described in Appendix D.

5.1 SELECTIVE AMPLIFICATION OF TARGET SITE SEQUENCES IN INDIVIDUAL BACTERIA BY THE POLYMERASE CHAIN REACTION

The polymerase chain reaction is used to selectively amplify a particular sequence of DNA. For example, it is possible to selectively amplify the lac operon sequence from the genetic material of coliforms using the PCR. Other DNA sequences in the genetic material are not amplified. The PCR is capable of producing a selective increase in numbers of a specific DNA sequence, eg lac operon, by a factor of 1 000 000, which would greatly facilitate subsequent detection by hybridisation of a gene probe.

In theory, amplification of lac operon target sites in individual bacteria by the polymerase chain reaction could alleviate the need for the 5 hour culturing step during which individual bacteria form

colonies, which have enough target site DNA to be detected by autoradiography.

The polymerase chain reaction is an important breakthrough in overcoming sensitivity limitations in using gene probes to detect pathogenic microorganisms in infected human beings. The potential of the PCR to detect the presence of DNA from the retrovirus HIV (human immunodeficiency virus), believed to be the causative agent for AIDS, has been assessed by Kwok et al (1987). Development of full blown AIDS may not occur for 4-7 years after initial infection by the virus. Immunological methods to detect the virus have disadvantages. Thus, gene probe techniques are now being developed to test for the presence of viral DNA in the cells. Kwok et al (1987) used primers complementary to conserved regions of the HIV genome to selectively amplify HIV DNA sequences in DNA extracted from cells cultured from patients with AIDS. The method proved more sensitive than previous gene probe techniques.

REFERENCES

KWOK S, MACK D H, MULLIS K B, POIESZ B, EHRLICH G, BLAIR D, FRIEDMAN-KIEN A and SNINSKY J J (1987) Identification of human immunodeficiency virus sequences by using in vitro enzymatic amplification and oligomer cleavage detection. Journal of Virology 61, 1690-1694.

LANE D J, PACE B, OLSEN G J, STAHL D A, SOGIN M L and Pace N R (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences, USA 82, 6955-6959.

SAIKI R K, SCHARF S, FALOONA F, MULLIS K B, HORN G T, ERLICH H A, and ARNHEIM N (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. Science, 230, 1350-1354.

SANGER F S, NICKLEN S and COULSON A R (1977) Proceedings of the National Academy of Sciences, USA 74, 5,463.

APPENDIX A - ENZYMES USED IN GENETIC ENGINEERING AND GENE PROBE TECHNOLOGY

The technology used in the design and production of a gene probe is only possible because of the availability of enzymes which are able to make, break and join pieces of DNA. These enzymes have been purified from a variety of bacteria, viruses and animal cells. Without these enzymes, the manipulations of DNA required for the cloning a fragment of DNA selected for a gene probe could not even be contemplated. Furthermore, subsequent analysis of DNA fragments, for example determination of their base sequences (gene sequencing), would be impossible. Three types of enzymes that perform different functions are described.

A1 ENZYMES WHICH SYNTHESISE DNA

Two enzymes, **DNA polymerase** I and **reverse transcriptase** are frequently used.

The central role of the enzyme DNA polymerase in the replication of chromosomal DNA has been discussed (Section 2.4). This enzyme uses a single-stranded DNA molecule as template to direct the synthesis of a second DNA strand, which is exactly complementary in nucleotide sequence to the template. In addition to the template strand, DNA polymerase requires a fragment of DNA (or RNA) from which it can begin adding nucleotides. This initial strand is called the primer. DNA polymerase then sequentially adds nucleotide units to the primer, observing the rules of base-pairing (Section 2.2) with the nucleotides in the template strand so that an exact complementary strand is synthesised. The result is a double-stranded DNA molecule with one new and one old strand (the template) (Figure 3). DNA polymerase I is extracted from bacterial sources. In E. coli cells, the major function of DNA polymerase I appears to be repairing damaged DNA molecules with the polymerase activity filling in the damaged strands. In addition, to help it with its repair activity, DNA polymerase I also has a DNA digesting activity, which it uses to remove damaged parts of the DNA strand before replacing with new DNA synthesised by the polymerase activity. This digesting

activity is analogous to a dentist drilling out damaged parts of teeth before putting in a filling. This "cut out-and-replace" repair property of DNA polymerase I is used in genetic engineering for incorporating radio-active nucleotides into cloned segments of DNA to produce radio-labelled probes. DNA polymerase is also used in the polymerase chain reaction (Appendix D).

The enzyme reverse transcriptase, so named because it catalyses the reverse of transcription (Section 2.5), is isolated from the nucleoprotein core of retroviruses (RNA tumour viruses). Retroviruses, which have received increasing publicity recently after being linked with causing cancer in animals and AIDS in humans, are small viruses which have a single-strand RNA molecule for a genome. In the retrovirus life-cycle, reverse transcriptase uses the RNA strand of the reterovirus genome as a template and synthesises a complementary DNA strand, which is called complementary DNA (or cDNA). The important feature of reverse transcriptase in genetic engineering is that it specifically copies single-stranded RNA molecules into DNA molecules which can then be cloned by insertion into plasmids. It is also used in sequencing RNA molecules (Appendix B) and workers at Leicester University successfully obtain sequences of bacterial 16S rRNA molecules by this method.

A2 ENZYMES THAT CLEAVE DNA MOLECULES AT SEQUENCE-SPECIFIC SITES

Certain proteins and enzymes are able to recognise specific base sequences within the DNA double helix. Among these are a group of enzymes called the restriction endonucleases or restriction enzymes. Each restriction enzyme recognises a particular target sequence of nucleotides within the DNA double helix, and cleaves both strands at this site. The names of some frequently used restriction endonucleases and their respective target sequences are presented in Figure 8. Depending on the particular restriction enzyme, the sequence recognised is usually between 4 and 6 base-pairs in length. Restriction endonucleases not only differ in the target sequence cleaved but also in how they cut the target sequence. Some, eg PvuII cut the recognition site at the centre of symmetry thus generating blunt-ended termini

(Figure 8d), whereas others eg EcoRI, HindIII, HpaII and KpnI cut their respective target sequences off-centre so generating overhanging ends (see Figure 8a-c). The two overhanging ends can anneal with each other by virtue of their complementarly base sequences and are sometimes known as "sticky" ends. "Sticky ends" produced by restriction endonuclease cleavage provide a method for joining DNA fragments together in cloning (Appendix C).

A large number of restriction endonucleases, each with different target sequences have been purified from bacterial sources and are now available from major suppliers. In 1983, restriction enzymes were available that recognised and cleaved some 91 different target sequences. The name of a particular restriction endonuclease reflects its bacterial source, eg EcoRI is obtained from E. coli and HindIII is isolated from Haemophilus influenza stereotype d.

The number of base-pairs in the recognition sequence determines the frequency of sites within a DNA fragment. In a random stretch of DNA, a four base recognition sequence will occur, on average once in every 256 base pairs, while a six base recognition sequence will be present once every 4096 bases. If a particular DNA molecule does not have a sequence identical to that recognised by a particular restriction endonuclease, it will not be cleaved. Since different restriction endonucleases recognise different target sequences, a particular DNA molecule may not be cleaved by one restriction endonuclease but may, for example, contain two cleavage sites for another restriction endonuclease. A map of a particular DNA molecule showing the location of these cleavage points is known as a restriction map. The restriction map for the plasmid known as pBR322, is shown in Figure 9a. Agarose gel electrophoresis, the main technique used in elucidating the positions of the cleavage sites in restriction maps, is discussed in Appendix B.

The most important property of restriction enzymes is reproducibility. Thus, a particular restriction endonuclease will cleave identical DNA molecules eg a cloned DNA fragment or a plasmid, at exactly the same sites, so producing identical product fragments every time. This is

particularly important for techniques that are used to analyse DNA, such as agarose gel electrophoresis and DNA sequencing. These techniques, which are discussed in Appendix B, require a large number of identical fragments for a successful result. This is illustrated in Figure 10 where five identical copies of a plasmid with three EcoRI restriction sites are treated with EcoRI. Cleavage occurs at each EcoRI site on each plasmid generating the same three fragments, A, B and C. Thus a homogeneous sample of cleavage products, containing five identical copies of the three fragments A, B and C will by obtained. In contrast, treatment of the same five plasmids with the enzyme DNase I, which cleaves DNA fragments anywhere, produces a random selection of completely different fragments. It would be most unlikely that any two fragments would be the same.

Restriction endonucleases are also of prime importance in cloning and the construction of recombinant plasmids (Appendix C).

A3 ENZYMES THAT JOIN THE ENDS OF DNA MOLECULES TOGETHER

The T4 bacteriophage contains an enzyme that covalently bonds the end of one DNA strand to the end of another DNA strand to permanently join the two fragments. T4 ligase can join DNA double helices with both blunt ends or cohesive ("sticky") ends.

APPENDIX B - TECHNIQUES FOR SEPARATING AND ANALYSING FRAGMENTS OF DNA

The analysis of a cloned DNA fragment is normally performed in two stages, which differ in the amount of detail they yield. The first stage of analysis reveals a limited amount of detail over a large length of DNA. It entails mapping the positions of cleavage sites for various restriction endonucleases (Appendix A) within the DNA fragment. A diagram of the DNA fragment showing the relative positions of, and the distances (measured in thousands of base-pairs, kbp) separating the sites cut by various restriction endonucleases is called a restriction map. The distances between restriction sites are determined by cutting the DNA with the various restriction endonucleases and measuring the lengths of the fragments produced. The technique used to measure the lengths of different DNA fragments is called **agarose gel electrophoresis**. The second stage of analysis reveals very detailed information over a relatively small fragment of DNA and involves determination of the exact nucleotide base sequence of a particular DNA fragment. This is known as **DNA sequencing**.

B1 AGAROSE GEL ELECTROPHORESIS FOR SEPARATION OF DNA FRAGMENTS BY LENGTH

Agarose gel electrophoresis is used to separate DNA fragments according to their size, ie length. It can be used to determine the lengths in base pairs of DNA molecules. The agarose gel contains microscopic pores and can thus serve as a molecular sieve. The DNA fragments are negatively charged at neutral pH due to the phosphate groups, and thus migrate towards a positive electrode if placed in an electric field. The principle of agarose gel electrophoresis is that DNA molecules are pulled through the gel towards the positive electrode and because the shorter molecules can move through the pores faster than the longer fragments, a mixture of DNA molecules of a variety of sizes are separated according to their size. DNA molecules of identical sizes, eg the same DNA molecule, will move exactly the same distance and, after staining with ethidium bromide, which specifically binds to nucleic acid, will appear as a sharp band on the gel.

The principle of agarose gel electrophoresis is demonstrated in Figure 11. Here five identical plasmids with three EcoRI sites were digested with EcoRI to produce five copies of three fragments A, B and C as shown in Figure 10. The products of digestion are loaded into one of the troughs on the gel. The three fragments A, B and C are of different sizes and begin migrating towards the positive electrode. The five identical C fragments all move at the same speed and faster than the A and B fragments. Similarly the five B fragments all move at the same speed and faster than the A fragments. The five A fragments all move at the same speed but slower than all the other fragments. Thus, at any time, identical DNA fragments will all be at the same position within the gel and can be collectively seen as a band. The separation of the bands depends on the difference in sizes of the fragments, A, B and C. By comparison with the distance moved by DNA marker molecules of known length, in other troughs in the gel, it is possible to determine the lengths of the unknown DNA molecules. This technique is only possible because of the reproducibility of cleavage by restriction endonucleases. If the plasmids were cleaved by DNase I, which cuts DNA randomly, then individual discrete bands would not be obtained. Instead, there would be one blurred continuous band spanning the length of the gel, resulting from the presence of DNA fragments of all sizes.

B2 DNA SEQUENCING - DETERMINATION OF THE BASE SEQUENCE OF DNA

The principles of DNA sequencing techniques are outlined below for completeness, but the only important fact to be aware of is that there are routine methods for sequencing between 250 and 400 consecutive bases on a cloned DNA fragment.

The DNA sequencing method currently used is called the chain termination method (Sanger et al 1977) and relies on the ability of polyacrylamide gel electrophoresis (PAGE) to resolve nucleic acid molecules differing in length by only one nucleotide in a total length of several hundred nucleotides. (Note that the principle of PAGE is identical to agarose gel electrophoresis described in Appendix B1). It allows between 300 and 400 consecutive nucleotides to be sequenced. The principle is that

synthesis of a DNA strand is terminated at one of the four bases A, T, G or C. The start point of synthesis is determined by using a short oligonucleotide primer which is complementary to a region of DNA adjacent to that to be sequenced. The double-stranded DNA to be sequenced must first be denatured into single-strands, one of which serves as a template for synthesis. The new chain is synthesised using DNA polymerase I. Termination at a specific base is achieved using one of the four dideoxynucleoside triphosphates, ddNTP (ddATP, ddCTP, ddTTP or ddGTP). The structure of these nucleotides is such that once one has been incorporated into a chain then additional nucleotides cannot be added. The concentration of a particular ddNTP is such that on average chains of several hundred nucleotides are synthesised before a ddNTP terminates synthesis. Consider the synthesis mixture to which ddATP has been added. Multiple, identical copies of the fragment to be sequenced and the primer are present. Different chains will be randomly terminated at different A residues, and because there are so many chains being synthesised each and every A will be represented. Thus, determination of the lengths of each synthesised fragment by PAGE gives the position of an A nucleotide. Similar techniques using the other ddNTPs gives the positions of the G, C and T residues. Thus, the complete base sequence can be obtained.

B3 METHODS FOR SEQUENCING RNA MOLECULES

The dideoxy chain termination method can also be applied to sequencing RNA molecules by using reverse transcriptase instead of DNA polymerase I (Lane et al 1985). Indeed Lane et al (1985) used this procedure as a rapid method to sequence 16S ribosomal RNAs in bacteria to assess their evolutionary relationships. It should be noted that numerous identical copies of the RNA segment to be sequenced are required. Each bacterial cell contains thousands of ribosomes and hence thousands of identical 16S rRNA molecules. Thus, 16S rRNA is extracted from a culture of bacteria and then directly sequenced using the dideoxy-chain termination method.

APPENDIX C - CLONING A DNA MOLECULE

C1 WHAT DOES CLONING A DNA MOLECULE MEAN?

The term clone can be applied to both cells and biological molecules. A clone is defined as a large number of cells or molecules all identical with an original ancestral cell or molecule. To produce and maintain a clone, that clone must be capable of undergoing replication. Thus, cells in a clone undergo mitotic divisions producing daughter cells which are identical in structure, function and genetic composition to the parent cell. These in turn divide to produce more identical cells. In genetic engineering, cloning a particular DNA molecule means constructing a system that replicates, perpetuates and maintains that DNA molecule such as to provide an indefinite supply of DNA molecules all with identical base sequences to that single, original DNA molecule.

The important point to remember about a clone of a particular DNA molecule is that the base sequences of all the DNA molecules in that particular clone are identical.

C2 WHY IS IT NECESSARY TO CLONE DNA MOLECULES?

An indefinite supply of multiple, identical copies of a DNA fragment is essential for:

i) analytical techniques, such as agarose gel-electrophoresis and DNA sequencing (Appendix B).

ii) gene probes. Furthermore, the gene probe samples must be pure, ie free from all other DNA sequences.

Thus, a clone of the DNA fragment to be analysed or to serve as a gene probe provides a satisfactory source of the required DNA.

C3 THE NEED FOR CLONING VECTORS IN CLONING A FRAGMENT OF DNA

To clone a DNA molecule, an environment must be chosen that not only promotes its replication but also maintains it. As discussed in Section 2.2, the molecular structure of the DNA double helix confers a mechanism for its own replication. Individual DNA fragments, for example a fragment containing the lac operon from a coliform chromosome, are unable to replicate themselves. A battery of replication enzymes (eg DNA polymerase) are required in addition to an unlimited supply of raw materials (Section 2.4). An obvious environment to supply the enzymes and precursors required for replication is within a bacterial host cell. E. coli are often used. However, incorporating fragments of DNA into E. coli cells is not sufficient to ensure their replication. As a further requirement, each DNA molecule to be replicated must contain a special nucleotide sequence - known as an origin of replication - which the DNA-replicating enzymes of the host bacterium are able to recognise. This is provided by using genetic engineering techniques to insert the DNA fragment to be cloned into a second DNA molecule, which already possesses an origin of replication, recognisable by the DNA-replicating enzymes of the bacterial host cell. This second DNA molecule thus serves as a carrier vehicle for the segment of interest and in genetic engineering jargon is called a cloning vector, or simply a vector.

C4 PLASMIDS AS CLONING VECTORS

Bacterial plasmids are commonly used as cloning vectors. Plasmids are small circular DNA molecules that are separate from the main chromosome of the bacterium. Each plasmid has its own origin of replication and is thus able to be replicated within the bacterium. Plasmids are maintained in the bacterial cell at a certain number (called the copy number) per bacterial chromosome and each time the bacterium undergoes a mitotic cell division and the chromosomal DNA molecule replicates, the plasmid(s) are replicated and **each daughter cell inherits a copy(s) of the plasmid.** Insertion of a foreign DNA fragment, such as the lacZ gene, into a bacterial plasmid produces what is called a recombinant

plasmid. These recombinant plasmids replicate in bacteria just like the original plasmid and so are maintained and provide an indefinite source of the inserted sequence of foreign DNA.

Plasmids typically contains genes for enzymes that confer resistance on the host bacterium to certain antibiotics. A "cartoon" diagram of a plasmid, called pBR322, frequently used as a cloning vector is shown in Figure 9b. The genes, AmpR and TetR, that code for the proteins that confer resistance to the antibiotics ampicillin and tetracycline, respectively, are shown as boxes in the diagram. (It should be realised that this is only to highlight the positions of these genes in the diagram; in the actual plasmid molecule, the double-stranded DNA in these genes is of identical structure to that in the rest of the plasmid.) The positions of the cleavage sites for the restriction endonucleases, EcoRI, HindIII, BamHI, SalI, PstI, and XorII are shown. A map of a piece of DNA indicating the positions of the various restriction endonuclease cleavage sites is called a restriction map. The site labelled Ori is the origin of replication, which is recognised as the start site of replication of the plasmid inside the bacterial cell.

C5 PRINCIPLES OF CLONING A FRAGMENT OF DNA

The plasmid to be used as cloning vector is cut at a selected site, usually in an antibiotic resistance gene, using a restriction endonuclease (Appendix A). In Figure 9b, the plasmid pBR322 is cleaved in the middle of the TetR gene, using the restriction endonuclease, BamHI. The ends of the linear plasmid are joined to the ends of the DNA fragment to be cloned, in this case a fragment containing the lacZ gene from the lac operon (Figure 9c), using the enzyme T4 ligase (Appendix A), so producing a circular, recombinant plasmid. The recombinant plasmid (Figure 9d), which contains the lacZ gene as an insert, is then incorporated into a single host bacterium of the appropriate strain, often E. coli, by a process called transformation. If this single bacterium is placed on an agar medium, which provides nutrients for growth, then the bacterium will divide into two. Each

daughter cell will also divide into into two and so on. With each cell division both the bacterial chromosome and the recombinant DNA plasmid (including the inserted fragment with the lacZ gene) are replicated. Each daughter cell inherits a copy of the chromosome and a copy of the recombinant plasmid. The end result is a colony of bacteria each with a copy of the lacZ gene. This fragment of DNA, containing the lacZ gene has thus been cloned. By plating some of the bacteria onto new agar plates, it is possible to maintain an indefinite supply of the bacteria and hence the cloned DNA fragment. If the plasmid DNA is extracted from all the bacteria in a colony then multiple, identical copies of cloned DNA fragment containing the lacZ fragment are obtained. These can be radio-labelled and used as gene probes (for total coliforms).

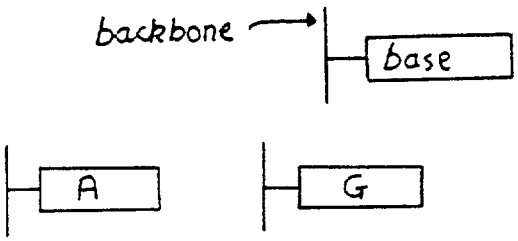
APPENDIX D - THE POLYMERASE CHAIN REACTION (PCR)

The PCR simulates replication of a fragment of DNA in a cell-free environment such as a test tube. In contrast to cloning, incorporation of the DNA fragment into plasmid vectors using restriction endonucleases and the culturing of bacteria are not required. Replication is performed by the enzyme DNA polymerase, hence the word "polymerase" in the name PCR. The PCR technique involves repeated rounds of replication, usually between 20 and 40. With each round the number of copies of the selected DNA fragment is theoretically doubled (in practice, however, each step is not 100% efficient) and so the number of copies increases expontially, hence the term "chain reaction". After 20 rounds of replication, a 1 048 580 fold (2^{20}) increase in the number of copies of the single, particular fragment would be expected assuming 100% efficiency. Saiki et al (1985) estimated a 200 000 fold amplification of a 110 bp fragment from the globin genes after 20 rounds. This is consistent with an overall efficiency of approximately 85%.

DNA polymerase, alone, cannot initiate replication of a double-stranded DNA fragment. As discussed in Section 2.4, the double-stranded DNA first has to be separated into single-stranded DNA to expose the bases for complementary base-pairing. Second, an oligonucleotide primer has to be available to provide a starting point for the DNA polymerase to begin complentary strand synthesis. In a chromosome or a plasmid this occurs at a specific site called the origin of replication, which is recognised by the replication machinery of the cell. In the PCR technique, the double stranded DNA fragments are separated into their component single-strands by heating the sample at the beginning of each round of replication. Multiple copies of two oligonucleotide sequences (~20 nucleotide in length) are added to the reaction mixture. The sequences of these two oligonucleotides are specially designed to be complementary to the regions of DNA sequence that flank the region of DNA to be amplified. These primers thus anneal specifically to the DNA sequences at the end of the sequence to be amplified, where they serve as primers for complementary strand synthesis by DNA polymerase. Thus,

the oligonucleotide primer, by virtue of its particular sequence, directs amplification of a particular fragment of DNA, in preference to all the other DNA sequences within the sample. Once a double-stranded DNA fragment has been replicated into two daughter fragments, each daughter fragment can itself serve as template and be replicated in the next round. Thus, with each round of replication, the number of templates for the next round doubles.

FIG 1 :- THE STRUCTURE OF DNA



Generalised structure of a nucleotide

Fig 1a :- The four nucleotides found in DNA

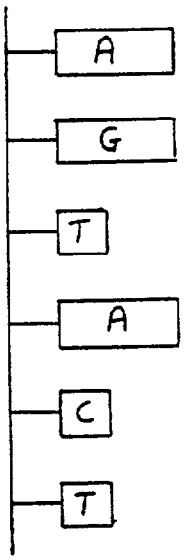
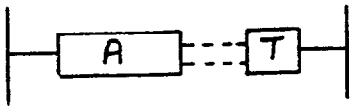
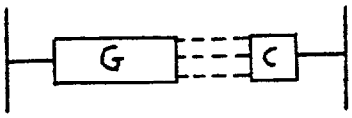


Fig 1b :- Single-stranded DNA molecule

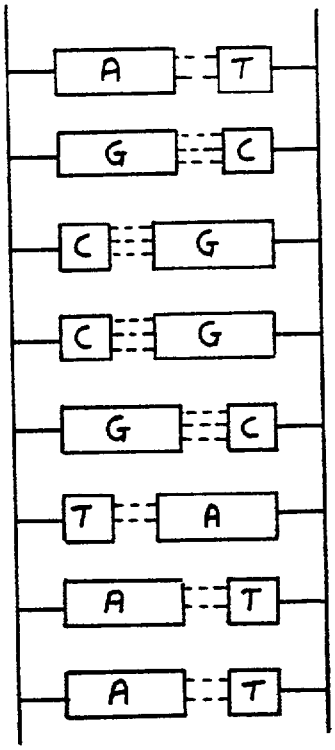


Adenine - Thymine base-pair



Guanine - Cytosine base-pair

Fig 1c :- Complementary base-pairing



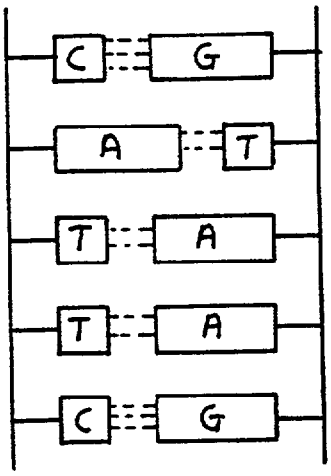
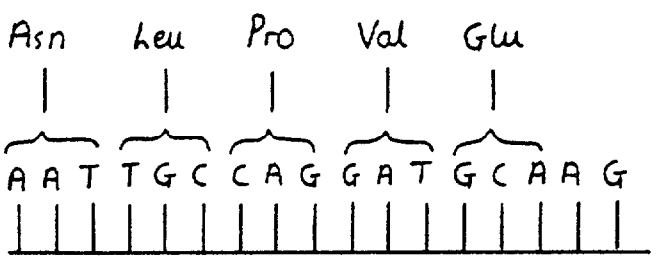


Fig 1d :- Two different double-stranded DNA molecules, one 8 bp long, the other 5 bp long. DNA molecules differ in length and base sequence

FIG 2 :- THE GENETIC CODE

Bases on the DNA are read in groups of three, called codons

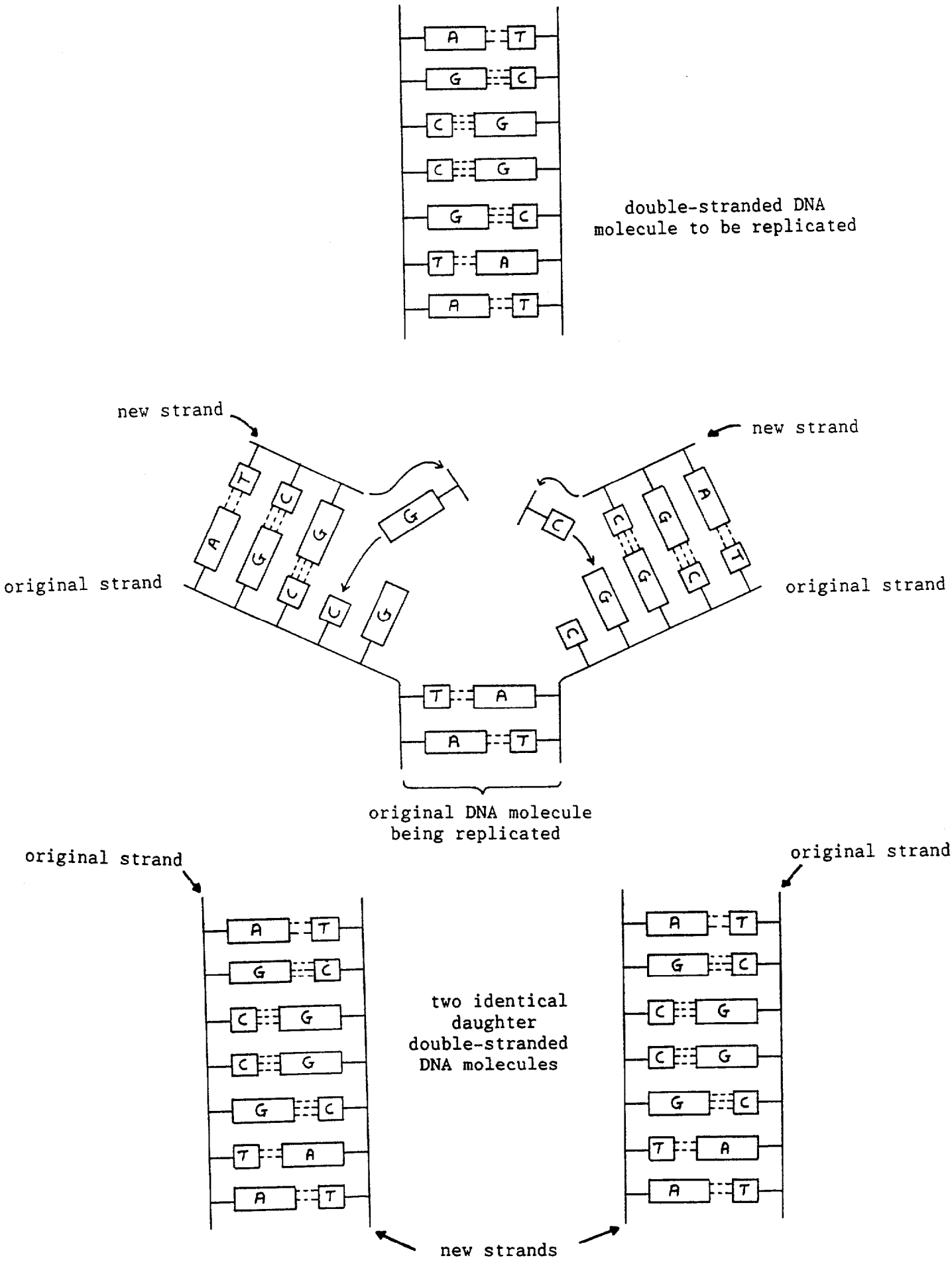


amino acid sequence of protein

portion of single-stranded DNA in a gene

(Asn, Leu, Pro, Val and Glu are abbreviations for five of the twenty different amino acids.)

FIG 3 :- REPLICATION OF A DNA MOLECULE



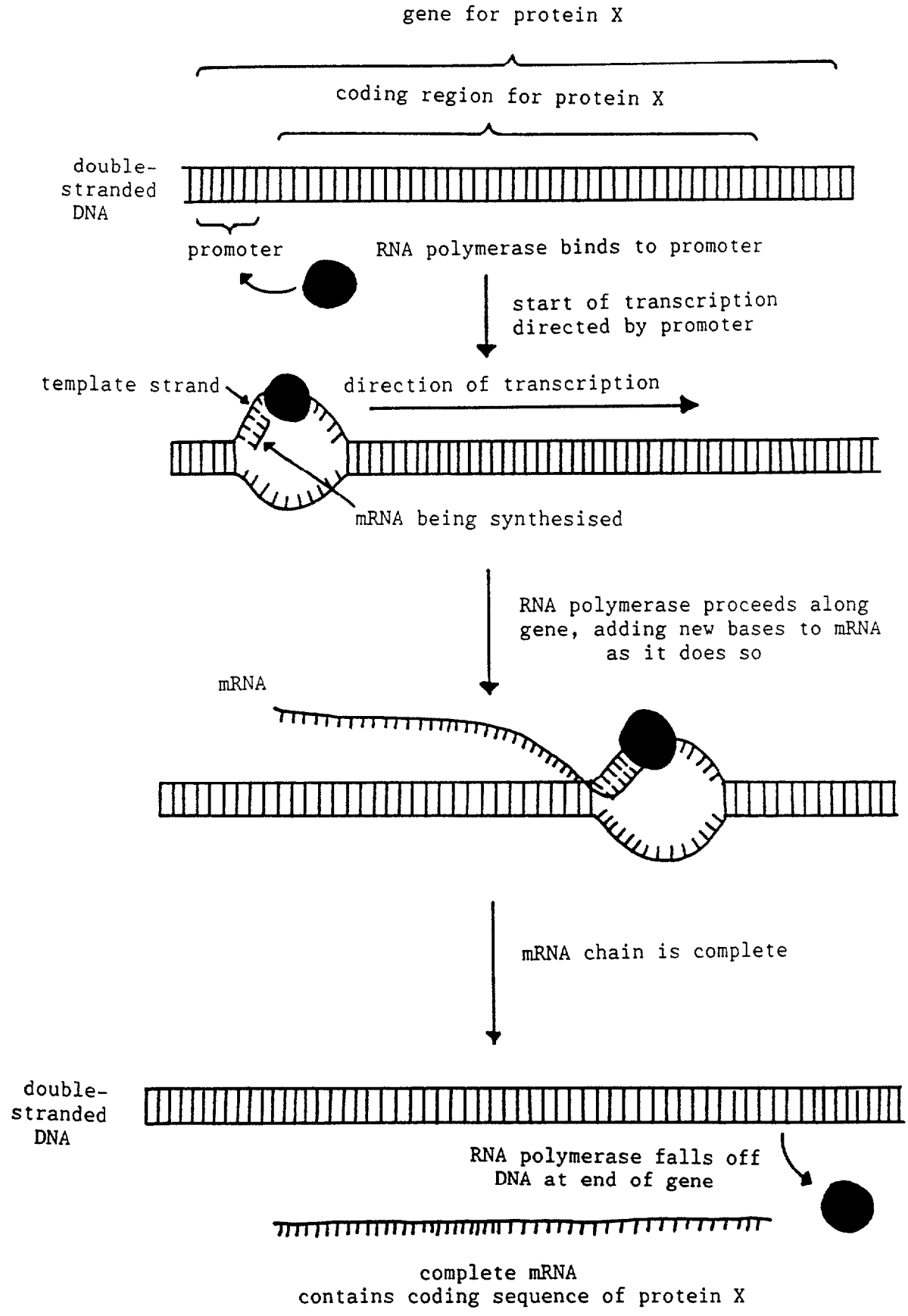
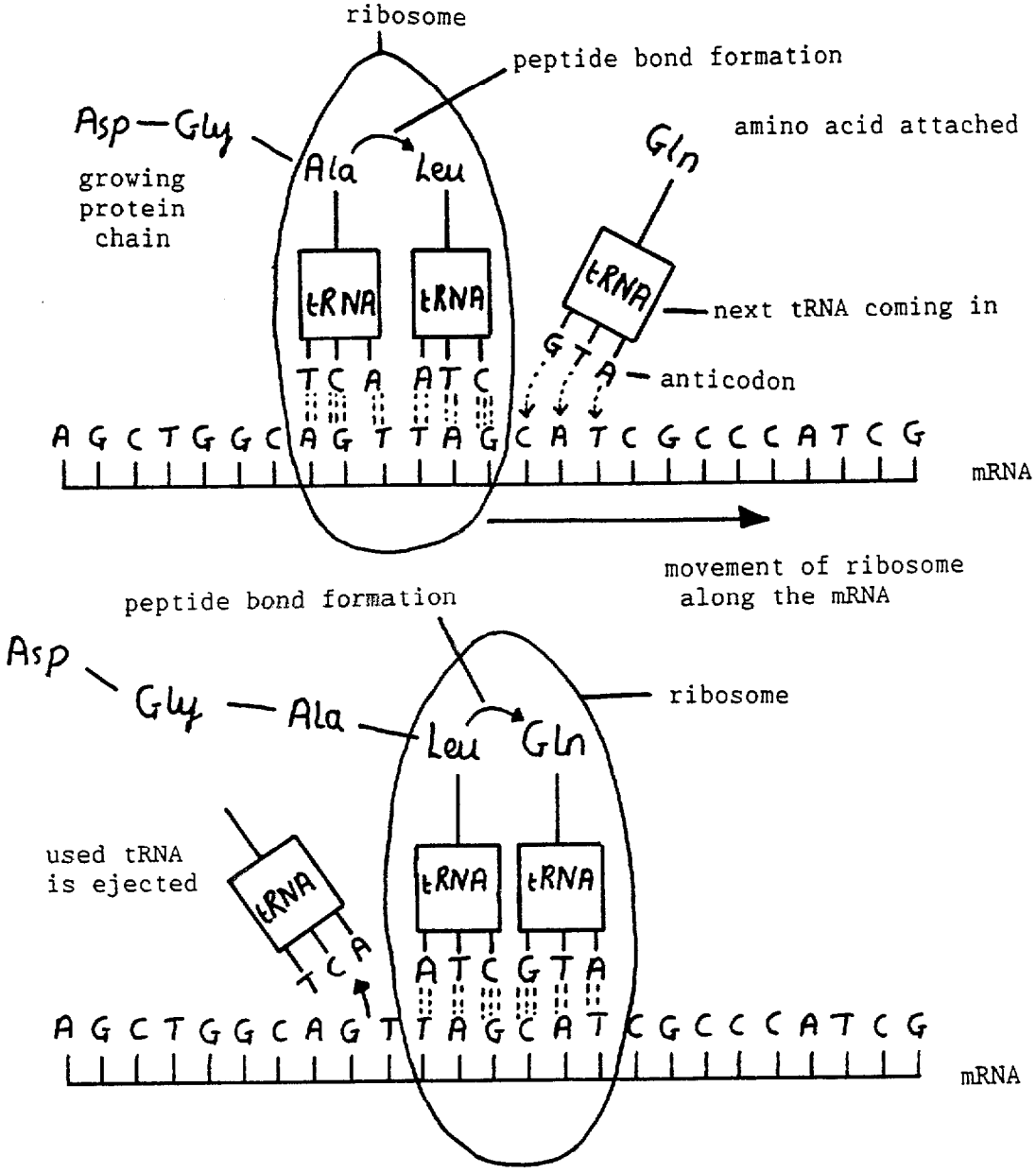

gene for protein X
coding region for protein X
double-stranded DNA
promoter
RNA polymerase binds to promoter
start of transcription directed by promoter
template strand
direction of transcription
mRNA being synthesised
RNA polymerase proceeds along gene, adding new bases to mRNA as it does so
mRNA
mRNA chain is complete
double-stranded DNA
RNA polymerase falls off DNA at end of gene
complete mRNA contains coding sequence of protein X

FIG 5 :- TRANSLATION



Asp, Gly, Ala, Leu and Gln represent different amino acids

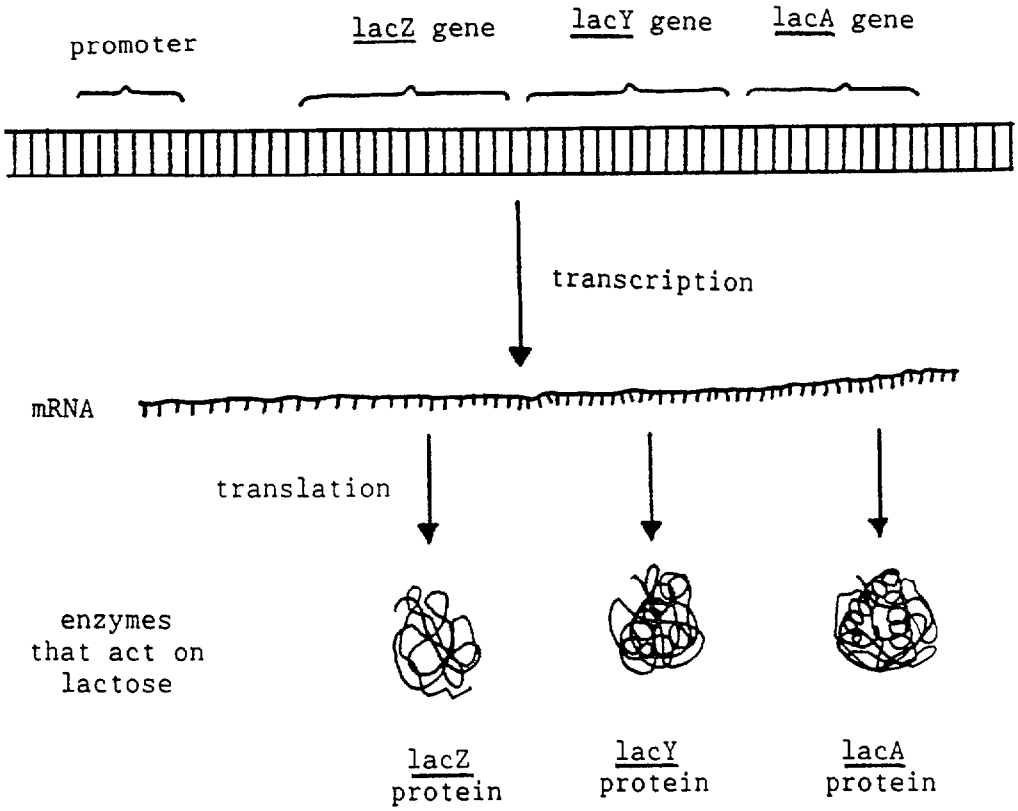
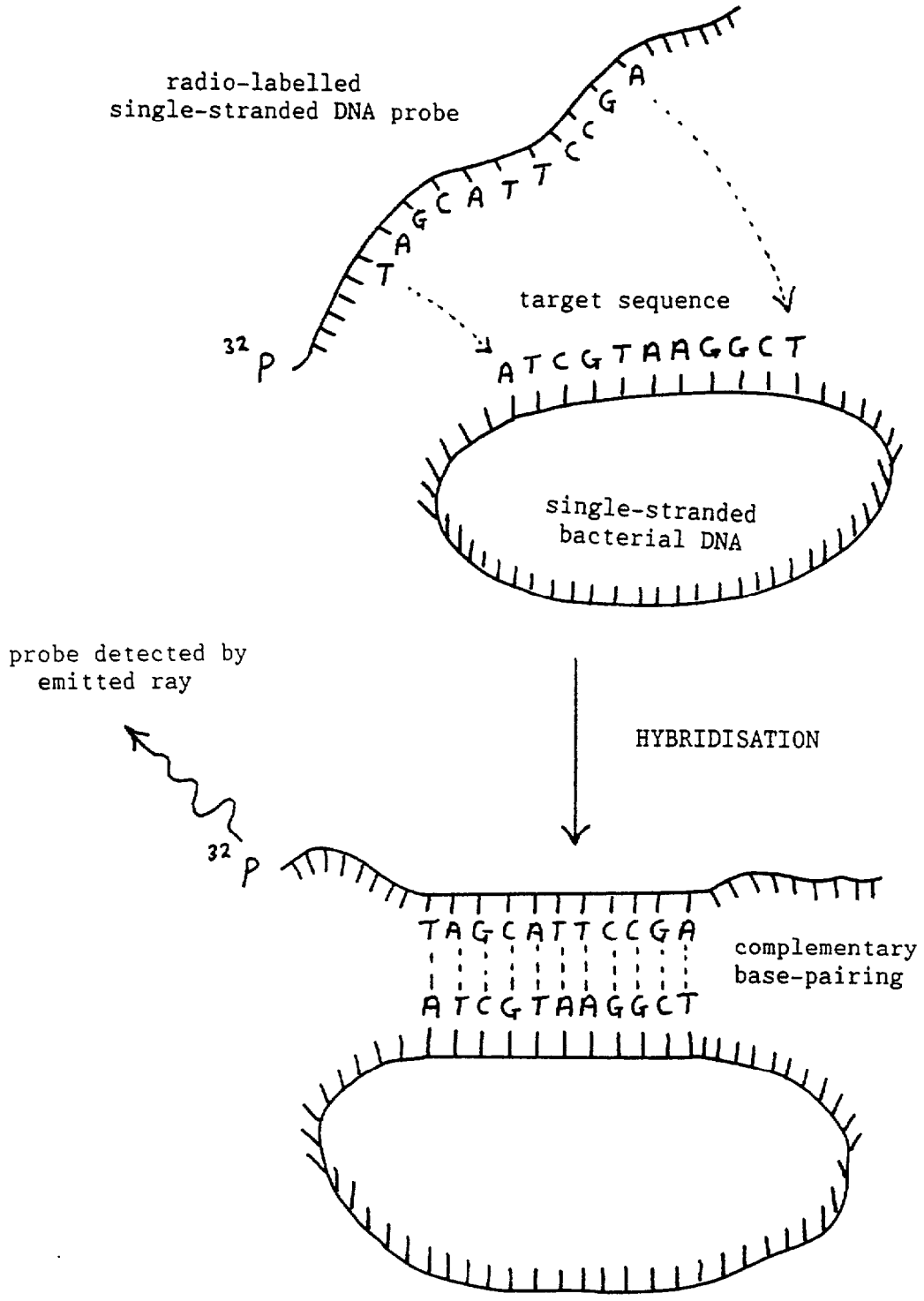

promoter
lacZ gene
lacY gene
lacA gene
transcription
mRNA
translation
enzymes that act on lactose
lacZ protein
lacY protein
lacA protein

FIG 7 :- GENE PROBE HYBRIDISING TO TARGET SEQUENCE IN BACTERIAL GENETIC MATERIAL BY COMPLEMENTARY BASE-PAIRING

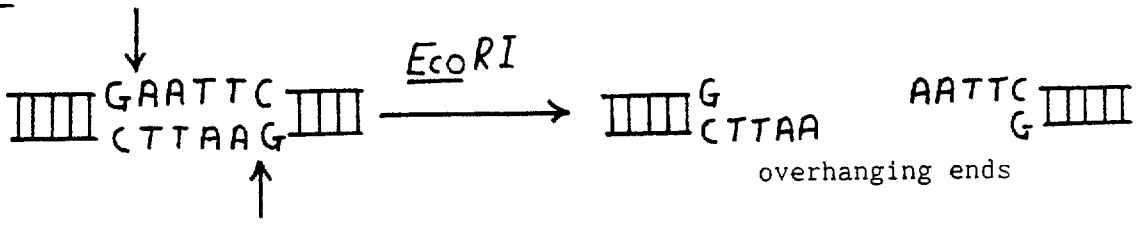


probe hybridised to target sequence in bacterial genetic material

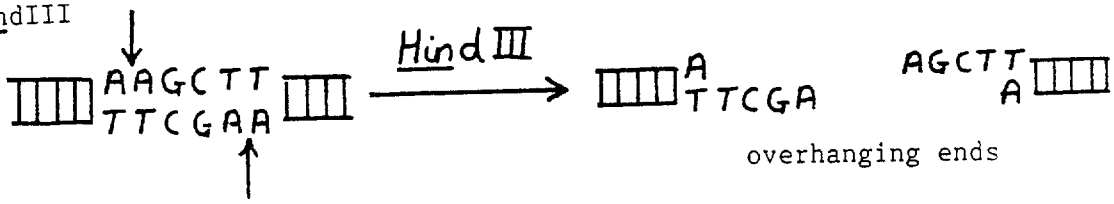
FIG 8 :- EXAMPLES OF COMMONLY USED RESTRICTION ENDONUCLEASES AND HOW THEY CUT THEIR SPECIFIC TARGET SEQUENCES

staggered cuts

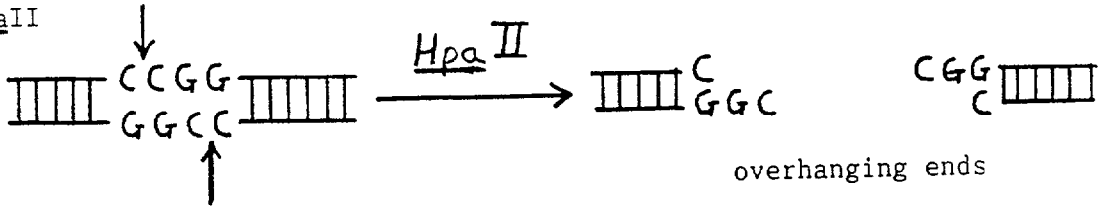
a) EcoRI



b) HindIII

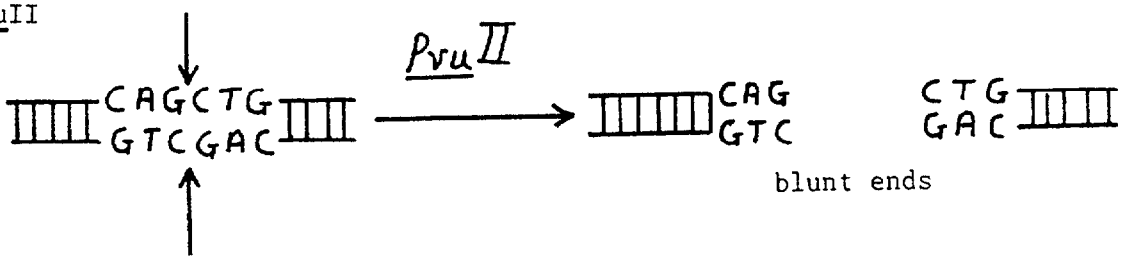


c) HpaII



central cut

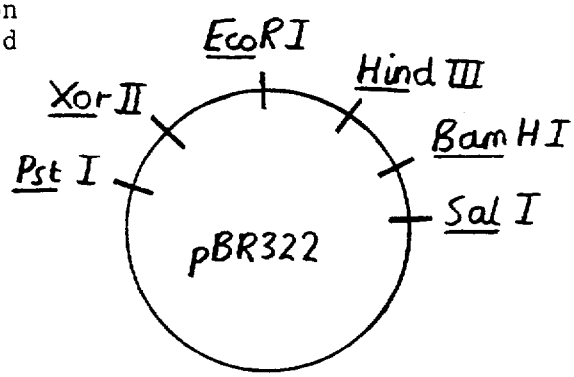
d) PvuII



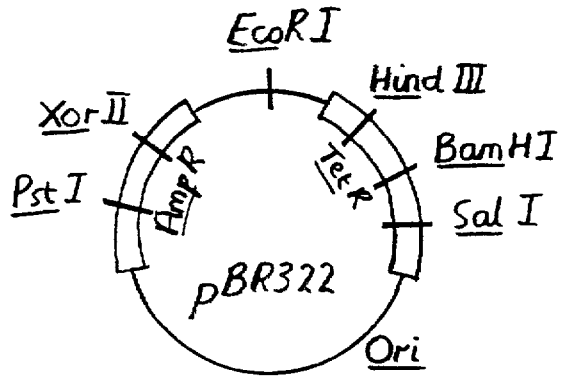
arrows show where each strand is cut

FIG 9 :- STRUCTURE OF A PLASMID AND DEMONSTRATION OF CONSTRUCTION OF A RECOMBINANT PLASMID USED FOR CLONING

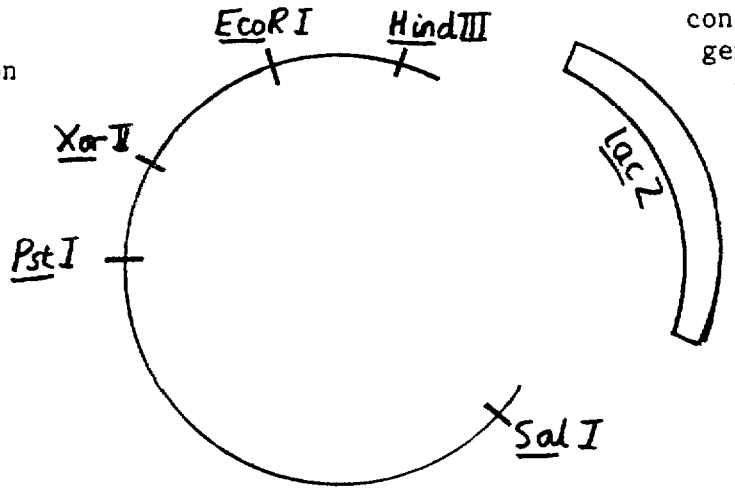
a) Restriction Map of plasmid pBR322



b) Location of antibiotic resistance genes and origin of replication



c) plasmid cut with restriction endonuclease BamHI



fragment of DNA containing lacZ gene from lac operon

d) lacZ gene is inserted to give a recombinant plasmid

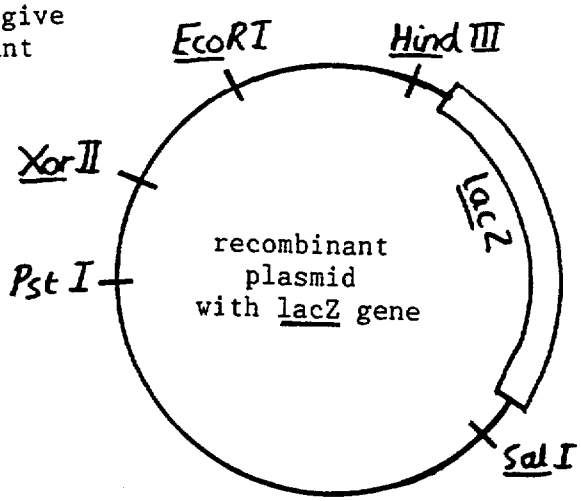


FIG 10 :- DIGESTION OF IDENTICAL PLASMIDS WITH RESTRICTION ENDONUCLEASE PRODUCES IDENTICAL COPIES OF PRODUCT FRAGMENTS, WHEREAS AN ENZYME THAT CUTS DNA ANYWHERE PRODUCES A RANDOM MIXTURE OF PRODUCT FRAGMENTS

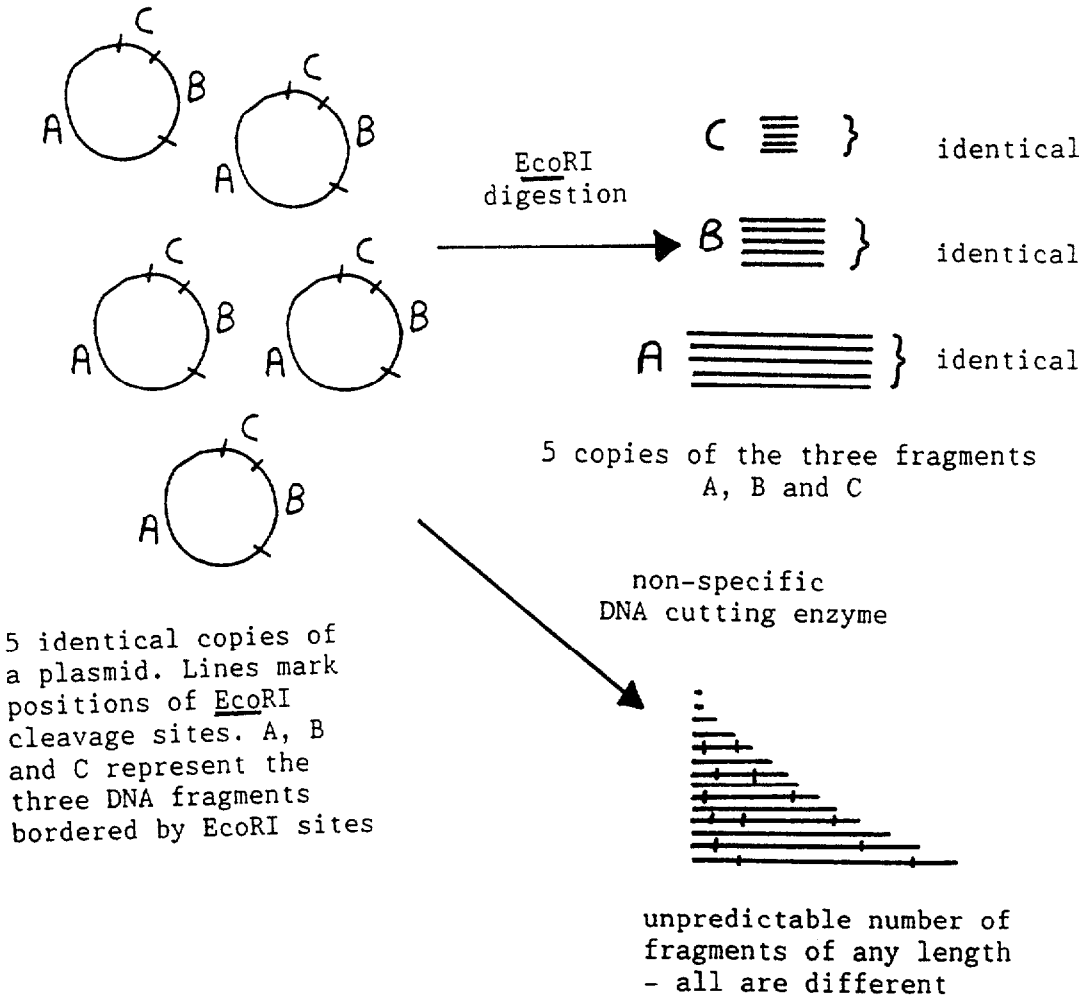


FIG 11 :- SEPARATION OF DNA MOLECULES OF DIFFERENT LENGTHS BY AGAROSE GEL ELECTROPHORESIS