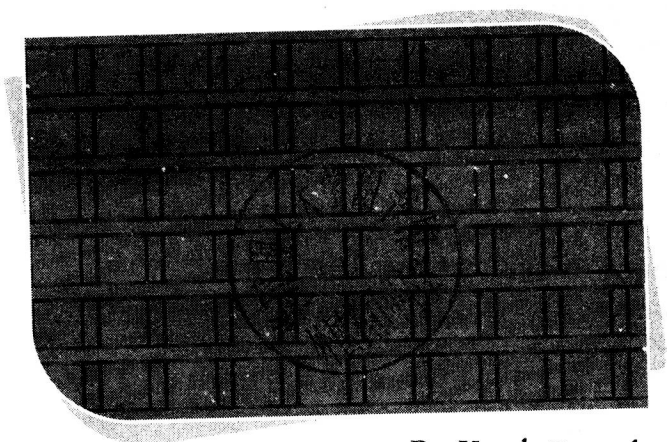
MOLECULAR Biotechnology

Vatsala Piramal

MOLECULAR Biotechnology



By: Vatsala Piramal

dominant

Dominant Publishers and Distributors
NEW DELHI-110002

Molecular Biotechnology

By: Vatsala Piramal

Exclusive rights by
Dominant Publishers And Distributors
for manufacture and export.

ISBN 81-7888-264-7

dominant

Dominant Publishers and Distributors

Editorial Office :
116–A, South Anarkali, Delhi - 110051.
Ph. 22415687, 30926471.

Sales & Marketing :
4378/4-B, Murari Lal Street, Ansari Road,
Daryaganj, New Delhi - 110002.
Ph. 23281685, Fax. 91-11-23270680.

Production:
199/5, C.B. Marg, Moujpur, Delhi - 110053.
Ph. 22913460.
e-mail: dominantbooks@post.com

Published by. A.S. Saini for Dominant Publishers And Distributors.
Printed at Orient Offset, Delhi - 110053.

Preface

Over a century of work has gone into the current framework of molecular biotechnology, but at its core, our understanding can be distilled down to what has become known as 'the central dogma' of molecular biotechnology: DNA is copied into genetic messages, which are then translated into proteins that go on to perform the underlying biochemistry of an organism.

The effort to understand inheritance has been an important journey for science. It has required a broad range of disciplines, from chemistry to microbiology to zoology, as well as the development of many new and exciting technologies. This mixing of disciplines, combined with overcoming many engineering difficulties, eventually resulted in the emergence of the field of biotechnology. An exciting enterprise is bringing molecular biotechnology to the forefront of society and causing us to reshape the way we see the world.

This book of Molecular Biotechnology offers an in-depth overview of contemporary biotechnology. It aims at providing the student with a profound fundamental basis, as well as offering him an integrated study package, compiled from a wide scope of scientific disciplines, which together underscore the practical realisations of application-oriented molecular biotechnology.

Completely revised and updated this book covers both the underlying scientific principles and the wide-ranging industrial, agricultural, pharmaceutical and biomedical applications of recombinant DNA technology. Ideally suited as a text, this book is also an excellent reference for health professionals, scientists, engineers or attorneys interested in biotechnology.

Contents

Chapter 1

Introduction

What is 'Biotechnology'?

Biotechnology generally refers to the use of microorganisms to produce certain chemical compounds. Long before the term 'biotechnology' was coined for the process of using living organisms to produce improved commodities, people were utilising living microorganisms to produce valuable products.

Since then 'biotechnology' has rapidly progressed and expanded. In the mid-forties, scale-up and commercial production of antibiotics such as penicillin occurred. The techniques used were:

(a) isolation of an organism producing the chemical of interest using screening/selection procedures and

(b) improvement of production yields via mutagenesis of the organism or optimisation of media and fermentation conditions. This type of 'antique' biotechnology is limited to chemicals produced in nature. It is also limited by its trial-and-error approach and requires a lengthy timeframe (years or even decades) for yield improvement.

About two decades ago, biotechnology became much more of a science (rather than an art). Regions of DNA (called genes) were found to contain information that would lead to synthesis of specific proteins (which are strings of amino acids). Each of these proteins have their own identity and function, many catalyse (facilitate) chemical reactions and others are structural components of entities in cells. If one now is able to express a natural gene in simple bacteria

such as *Escherichia coli* (*E. coli*), a bacterium living in intestines that has become the model organism for much of biotechnology, one can have this bacterium make a lot of the protein coded for by the gene, regardless its source. The techniques used for this development include:

(a) isolation of the gene coding for a protein of interest,

(b) cloning of this gene into an appropriate production host and

(c) improving expression by using better promoters, tighter regulation, etc., together these techniques are known as recombinant DNA techniques.

The commercial implications are that a large number of proteins, existing only in tiny quantities in nature, can now be produced in large amount if needed. Also, the yields of (bio)chemicals to be produced can be increased much faster than was possible with classical fermentation. These modern biotechnology techniques started with the expression of human genes such as that coding for insulin, but have since been extended to mammalian, microbial and plant genes. Also, the spectrum of 'bioreactors' (organisms used for production) recently has been broadened to include a variety of animals and plants. Perceived needs and marketability, the researchers' imagination, ethics and governmental regulations essentially are the major factors in setting the stage and boundaries for developments in biotechnology.

About a decade ago, 'protein engineering' became possible as an offshoot of the recombinant DNA technology. Protein engineering differs from 'classical' biotechnology in that it is concerned with producing new (man-made) proteins, which have been modified or improved in some way. The techniques involved in protein engineering are more complicated than before and involve

(a) various types of mutagenesis (to cause changes in specific locations or regions of a gene to produce a new gene product),

(b) expression of the new gene to form a stable protein,

(c) characterisation of the structure and function of the protein produced and

(d) selection of new locations or regions to modify as a result of this characterisation.

It has become possible to transform (genetically modify) plants and animals that are important for food production. 'Transgenic' animals and plants, including cows, sheep, tomatoes, tobacco, potato and cotton have now been obtained. Genes introduced may make the organism more resistant to disease, may influence the rate of fruit ripening or may increase productivity. As this approach leads to release of genetically altered organisms into the environment, this part of biotechnology is quite strictly regulated at government levels. Recent advances in this area of modern biotechnology are numerous.

Below is an overview of recombinant DNA based biotechnology:

1953	Double helix structure of DNA was first described by Watson and Crick.
1973	Cohen and Boyer developed genetic engineering techniques to 'cut and paste' DNA and to amplify the new DNA in bacteria.
1977	The first human protein (somatostatin) was produced in a bacterium (E. coli).
1982	The first recombinant protein (human insulin) appeared on the market.
1983	Polymerase Chain Reaction (PCR) technique conceived.
1990	Launch of the Human Genome Project (HGP), an international effort to sequence the human genome.
1995	The first genome sequence of an organism (*Haemophilus influenzae*) was determined.
2000	A first draft of the human genome sequence was completed.

Fundamentals of molecular biotechnology

Contemporary molecular biology is concerned principally with understanding the mechanisms responsible for transmission and expression of the genetic information that ultimately governs cell

structure and function. All cells share a number of basic properties and this underlying unity of cell biology is particularly apparent at the molecular level. Such unity has allowed scientists to choose simple organisms (such as bacteria) as models for many fundamental experiments, with the expectation that similar molecular mechanisms are operative in organisms as diverse as *E. coli* and humans. Numerous experiments have established the validity of this assumption and it is now clear that the molecular biology of cells provides a unifying theme to understanding diverse aspects of cell behaviour.

Initial advances in molecular biology were made by taking advantage of the rapid growth and readily manipulable genetics of simple bacteria, such as *E. coli* and their viruses. More recently, not only the fundamental principles but also many of the experimental approaches first developed in prokaryotes have been successfully applied to eukaryotic cells. The development of recombinant DNA has had a tremendous impact, allowing individual eukaryotic genes to be isolated and characterised in detail. Current advances in recombinant DNA technology have made even the determination of the complete sequence of the human genome a feasible project.

The molecular biotechnology revolution

In 1885, a scientist named Roux demonstrated embryonic chick cells could be kept alive outside an animal's body. For the next hundred years, advances in cell tissue culture have provided fascinating glimpses into many different areas such as biological clocks and cancer therapy. Monoclonal antibodies are new tools to detect and localise specific biological molecules. In principle, monoclonal antibodies can be made against any macromolecule and used to locate, purify or even potentially destroy a molecule as for example with anticancer drugs.

Molecular biology is useful in many fields. DNA technology is utilised in solving crimes. It also allows researchers to produce banks of DNA, RNA and proteins, while mapping the human genome. Tracers are used to synthesise specific DNA or RNA probes, essential for localising sequences involved in genetic disorders. With genetic engineering, new proteins are synthesised.

They can be introduced into plants or animal genomes, producing a new type of disease resistant plants, capable of living in inhospitable environments (i.e. temperature and water extremes).

When introduced into bacteria, these proteins have also produced new antibiotics and useful drugs. Techniques of cloning generate large quantities of pure human proteins, which are used to treat diseases like diabetes. In the future, a resource bank for rare human proteins or other molecules is a possibility. For instance,

DNA sequences which are modified to correct a mutation, to increase the production of a specific protein or to produce a new type of protein can be stored. This technique will probably be playing a key role in gene therapy.

Chapter 2

DNA, RNA and Protein Synthesis

Structure and function of DNA

DNA molecules are incredibly long, but also very thin. One DNA molecule from the chromosome of a mammal may be about 1 m long when unravelled. However, it has to fit in a nucleus of some 5-6 orders of magnitude smaller and is folded up in chromosomes in a highly organised manner. DNA is a linear polymer that is composed of four different building blocks, the nucleotides. It is in the sequence of the nucleotides in the polymers where the genetic information carried by chromosomes is located. Each nucleotide is composed of three parts:

(1) a nitrogenous base known as purine (adenine (A) and guanine (G)) or pyrimidine (cytosine (C) and thymine (T)),

(2) a sugar, deoxyribose and

(3) a phosphate group.

The nitrogenous base determines the identity of the nucleotide and individual nucleotides are often referred to by their base (A, C, G or T). One DNA strand can be up to several hundred million nucleotides in length. T can form a hydrogen bond with A and C with G. Two DNA strands wind together in an antiparallel fashion in a double helix.

Inside the cell, the DNA acts like an 'instruction manual': in its sequence, it provides all the information needed to function, but the actual work of translating the information into a medium which can be used directly by the cell is done by RNA, (ribonucleic acid.

The structural difference RNA and DNA is that RNA contains a -OH group both at the 2' and 3' position of the ribose ring, whereas DNA (which stands, in fact, for deoxy-RNA) lacks such a hydroxy group at the 2' position of the ribose. The same bases can be attached to the ribose group in RNA as occurring in DNA, with the exception that in RNA thymine does not occur and is replaced by uracil, which has an H-group instead of a methyl group at the C-5 position of the pyrimidine. The RNA has three functions:

(a) it serves as the messenger that tells the cell (the ribosomes) what protein to make (messenger RNA, mRNA),

(b) it serves as part of the structure of the ribosome, the protein/RNA complex that synthesises proteins according to the information presented by the mRNA (ribosomal RNA, rRNA) and

(c) it functions to bring amino acids (the constituents of the proteins) to the ribosome when a specific amino acid 'is called for' by the information on the mRNA to be put in into the protein that is being synthesised, this RNA is called transfer RNA (tRNA).

An important point of emphasis should be that all vegetative cells of one organism contain the same genetic information. Upon division, each daughter cell obtains an 'exact' copy of the DNA of the parent. However, the specific genes that are expressed at specific times may be very different between different tissues. These differences in gene expression allow for the regulation of development of the organism and for the development of different tissues. For the most part, DNA-binding proteins (encoded by the DNA) play an important role in the regulation of expression of genes encoded on the DNA.

RNAs

The messenger RNA (mRNA) serves as an intermediate between DNA and protein. Parts of the DNA are 'transcribed' into mRNA, a single-stranded molecule. Transcription starts at a specific site on the DNA called a promoter. Each gene or operon has its own promoter(s). Transcription ends at a terminator sequence on the

DNA. The transcripts (mRNA molecules) usually are 300-50,000 nucleotides long and contain the information to make protein. In eukaryotes (organisms with cells containing a nucleus, in fact, any higher organism) generally the transcripts needs to be processed before they can serve as a blueprint for a protein. The processing involves the removal of intervening sequences (introns) in the gene. The introns may be anywhere between 50 and 10,000 nucleotides in length. The coding regions of the mRNA are called exons. There may be up to 100 introns in a single gene. The introns are spliced out by small ribonucleoprotein particles (consisting of RNA and protein), which appear to pull the two ends of the intron together. However, there are also introns that splice out without the need of a protein: the RNA sequence itself appears to contain sufficient information to know where to splice out the intron. After processing, the information in the mRNA can be used to be 'translated' into a protein of specific sequence. However, in prokaryotes introns are rare and mRNA generally does not get processed before translation.

Ribosomal RNAs (rRNAs) are essential components of an important part of the protein synthesis machinery: the ribosomes. In addition to rRNA, there are some 70 different proteins in a ribosome. There are hundreds of copies of rRNA genes per genome, thus making the production of lots of rRNA possible. There are four different rRNAs, each with a different size. Each ribosome contains one molecule of each of the four rRNA types. In prokaryotes, ribosomes bind to the mRNA close to the translation start site. This ribosome-binding site is referred to as the Shine-Dalgarno sequence or as the ribosome recognition element. In eukaryotes, ribosomes bind at the 5' end of the mRNA and scan down the mRNA until they encounter a suitable start codon.

Transfer RNA (tRNA) carries amino acids to the ribosomes, to enable the ribosomes to put this amino acid on the protein that is being synthesised as an elongating chain of amino acid residues, using the information on the mRNA to 'know' which amino acid should be put on next. For each kind of amino acid, there is a specific tRNA that will recognise the amino acid and transport it to the protein that is being synthesised and tag it on to the protein once the information on the mRNA calls for it.

All tRNAs have the same general shape, sort of resembling a clover leaf. Parts of the molecule fold back in characteristic loops, which are held in shape by nucleotide pairing between different areas of the molecule. There are two parts of the tRNA that are of particular importance: the aminoacyl attachment site and the anticodon. The aminoacyl attachment site is the site at which the amino acid is attached to the tRNA molecule. Each type of tRNA specifically binds only one type of amino acid. The anticodon (three bases) of the tRNA base-pairs with the appropriate mRNA codon at the mRNA-ribosome complex. This temporarily binds the tRNA to the mRNA, allowing the amino acid carried by the tRNA to be incorporated into the polypeptide in its proper place. Thus, the sequence of the codon (three bases) in the mRNA dictates the amino acid to be put in the protein at a specific site. The 'dictionary' of codons coding for amino acids is called the genetic code. The three codons for which there is no matching tRNA (UAA, UGA and UAG) serve as 'stop-translation' signals at which the ribosome falls off.

Protein synthesis, processing and regulation

Transcription and RNA processing are followed by translation, the synthesis of proteins as directed by mRNA templates. Proteins are the active players in most cell processes, implementing the myriad tasks that are directed by the information encoded in genomic DNA. Protein synthesis is thus the final stage of gene expression. However, the translation of mRNA is only the first step in the formation of a functional protein. The polypeptide chain must then fold into the appropriate three-dimensional conformation and, frequently, undergo various processing steps before being converted to its active form. These processing steps, particularly in eukaryotes, are intimately related to the sorting and transport of different proteins to their appropriate destinations within the cell.

Although the expression of most genes is regulated primarily at the level of transcription, gene expression can also be controlled at the level of translation and this control is an important element of gene regulation in both prokaryotic and eukaryotic cells. Of even broader significance, however, are the mechanisms that control the activities

of proteins within cells. Once synthesised, most proteins can be regulated in response to extracellular signals by either covalent modifications or by association with other molecules. In addition, the levels of proteins within cells can be controlled by differential rates of protein degradation. These multiple controls of both the amounts and activities of intracellular proteins ultimately regulate all aspects of cell behaviour.

Translation of mRNA

Proteins are synthesised from mRNA templates by a process that has been highly conserved throughout evolution. All mRNAs are read in the 5′ to 3′ direction and polypeptide chains are synthesised from the amino to the carboxy terminus. Each amino acid is specified by three bases (a codon) in the mRNA, according to a nearly universal genetic code. The basic mechanics of protein synthesis are also the same in all cells: Translation is carried out on ribosomes, with tRNAs serving as adaptors between the mRNA template and the amino acids being incorporated into protein. Protein synthesis thus involves interactions between three types of RNA molecules (mRNA templates, tRNAs and rRNAs), as well as various proteins that are required for translation.

Transfer RNAs

During translation, each of the 20 amino acids must be aligned with their corresponding codons on the mRNA template. All cells contain a variety of tRNAs that serve as adaptors for this process. As might be expected, given their common function in protein synthesis, different tRNAs share similar overall structures. However, they also possess unique identifying sequences that allow the correct amino acid to be attached and aligned with the appropriate codon in mRNA.

Transfer RNAs are approximately 70 to 80 nucleotides long and have characteristic cloverleaf structures that result from complementary base pairing between different regions of the molecule.