Chapter 3

Introduction to Molecular Biology

> Everything should be as simple as possible, but not simpler.
> *Albert Einstein*

Biotechnology research seeks to develop applications of molecular biology. Many sources use analogies to recipe books or blueprints to explain the role of DNA and genes in molecular biology. Ultimately, these analogies obscure the importance of topics such as regulation of gene expression, which is of fundamental importance in understanding molecular biology. When applying one's knowledge of biotechnology fundamentals, most metaphors fail. It is only by understanding molecular biology and biotechnology applications that one can appreciate the applications and limitations of techniques used in molecular biology.

This chapter presents a brief, metaphor-free, introduction to molecular biology. Subsequent chapters describe the tools, techniques, and applications of biotechnology and provide greater details on the potential and limitations of molecular biology.

INFORMATION FLOW IN MOLECULAR BIOLOGY

In order to understand the basis of most biotechnology applications, it is necessary to first understand the process by which information in genes leads to the formation of structural and functional proteins.

Proteins serve structural and functional roles that give individual cells—and by extension whole organisms—specific structures and functional characteristics. When many people think of proteins,

Information Flow in Molecular Biology

Genetic information is contained in DNA and leads to the formation of proteins through an intermediary called mRNA.

they think of foods such as meat and beans. While animal muscle and plant seeds are excellent sources of dietary protein, proteins play a central role in all cell types and perform functional and structural roles (see Table 3.1). Examples of structural proteins include keratin, which makes skin waterproof, and myosin, which interacts with other proteins in muscles to make them flex.

DNA contains information that describes the construction of proteins. The process of protein synthesis is as follows:

1. DNA contains the information to produce proteins.
2. Information encoded in DNA is *transcribed* into a molecule called messenger RNA (mRNA)—effectively a "working copy" of the DNA sequence of a given gene.
3. mRNA is *translated* into proteins by the protein synthesis machinery, the composition of the resulting protein corresponding to the original DNA instructions.

This basic mechanism is conserved in all life forms, from bacteria to humans. The implication of this common process that converts information in DNA into functional proteins is that similar techniques can be used to investigate and manipulate all biological systems. Furthermore, it is possible to make human therapeutic pro-

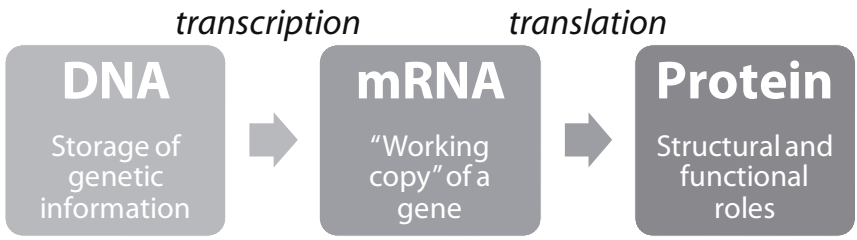

Figure 3.1 *Simplified model of information flow in molecular biology*

teins, for example, in organisms as distantly related as bacteria.

Understanding the roles of DNA, RNA, and protein and their relationships to each other is essential to understanding molecular biology. While there are some specific exceptions (e.g., retroviruses and prions) to the order and direction of information flow shown in Figure 3.1, these examples still fit within the general framework, and the majority of biological systems use the framework as presented.

DNA: STORING AND RELAYING INFORMATION

Deoxyribonucleic acid (DNA) is the primary source of genetic information in cells. Humans, plants, animals, and bacteria all contain DNA. DNA is physically passed from generation to generation, bestowing certain traits of parents to their children. The reason why children have physical characteristics from each of their parents—a child may have their mother's eye color and father's hair color—is

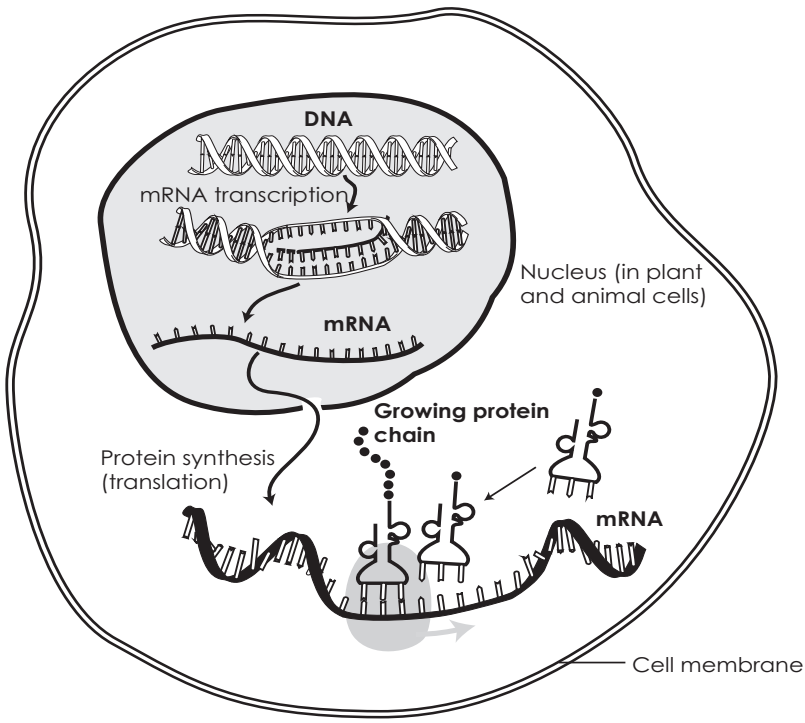

Figure 3.2 *General scheme of gene expression*
Modified from National Human Genome Research Institute

because they received half their DNA from each parent.

Each of our cells (with a few exceptions like red blood cells, eggs, and sperm) contain all the DNA required to code our genetic features. Individual regions of DNA that confer traits are called genes. Information in genes is relayed to the protein synthesis machinery within cells where it dictates the production of proteins. The word "genome" refers to all the DNA in an organism. The human genome contains approximately 30,000 genes arrayed on 46 long stretches of DNA called chromosomes.

DNA is essentially composed of two intertwined strands that form a double helix. The two strands of DNA are said to be complementary because the sequence of one strand indicates the sequence of the opposite strand, like a photograph and its negative. Each strand is physically composed of four different chemical units called nucleotides, the sequence of which encodes the genetic information. These four chemical units, adenine, cytosine, guanine, and thymine, are often abbreviated as A, C, G, and T, respectively. Just as the English language can be expressed in twenty-six letters, the genetic code is expressed in these four chemical units. A DNA "sequence" refers to the specific order of A's, C's, G's, and T's in a stretch of DNA.

There are two essential elements of genes: coding and regulatory elements. The coding elements of genes are first transcribed as mRNA, which is then translated into protein. The chemical sequence of A's, C's, G's, and T's in the coding region of a gene determines the composition and structure of the resulting protein and, by extension, its function. Regulatory elements affect the rate at which genes are transcribed and translated, and may be interspersed within the coding sequence or outside of it. Regulatory elements also control the cell types within which specific genes are activated, and the timing and magnitude of gene expression. Gene regulation thereby allows individual proteins to be expressed only in certain cells at specific times and at specific rates.

Proper regulation of gene expression—the production of gene products—is essential. Under- or over-expression of genes can have deleterious effects. For example, many forms of cancer are caused by mis-regulation of gene expression that results in uncontrolled cell division. A potential solution for diseases resulting from low expres-

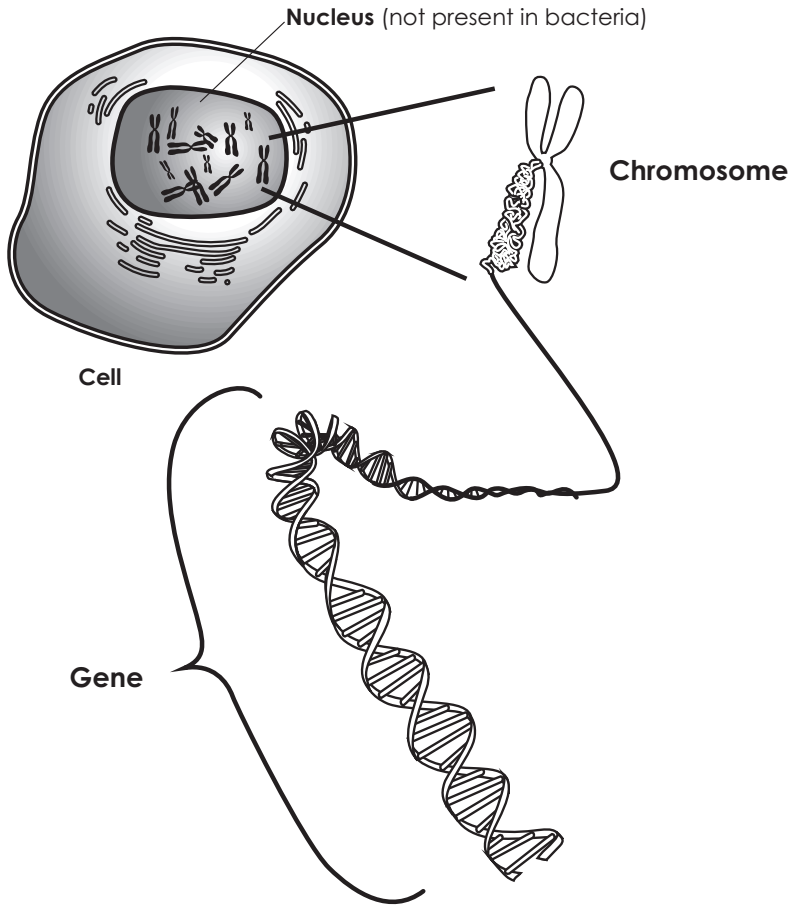

Figure 3.3 *DNA: Chromosomes and genes*
Modified from National Human Genome Research Institute

sion of genes is to use gene therapy to introduce affected genes or regulatory elements to spur additional production. One of the challenges of gene therapy is developing methods to regulate the expression of genes that are introduced into cells and ensure that they are not over-expressed. A solution for diseases caused by over-expressed genes is RNA interference. This procedure prevents translation of mRNA, inhibiting protein production. RNA interference is described in further detail in Chapter 6.

Box

Human chromosomes and genetic trait inheritance

The human genome is composed of chromosomes. We get 23 chromosomes from our mother and 23 chromosomes from our father, constituting 23 pairs. While 22 of the 23 chromosome pairs are similar in both men and women, the 23rd pair is quite different and determines the sex of an individual. For the 23rd pair of chromosomes, women have two X-chromosomes while men have one X- and one Y-chromosome. Because X-chromosomes contain more DNA than Y-chromosomes, they are physically larger than Y-chromosomes. Having too many or too few chromosomes can affect gene regulation and cause diseases. Down's Syndrome, for example, occurs in individuals with three copies of chromosome 21.

The roles of X- and Y-chromosomes are important in understanding sex-linked diseases. Women do not have Y-chromosomes, so diseases that are caused by defective genes on the Y-chromosome can only occur in men. Additionally, men only have one X-chromosome, so mutations in genes on the X-chromosome are more likely to affect males, because the second X-chromosome in women can sometimes compensate for mutations on the first. Color blindness, caused by a mutation on the X-chromosome, is more common in men than women for this reason.

mRNA: THE MESSENGER

Messenger RNA (mRNA) is used to relay information from genes in DNA to the protein synthesis machinery. An additional feature of mRNA is that it can be destroyed once sufficient protein is produced, permitting an extra level of control of gene expression. RNA is also present in forms other than mRNA, some of which are described later in this chapter.

It is possible to affect expression of genes by targeting their mRNA with antisense RNA or DNA—nucleic acids which can bind the mRNA. Unlike DNA, which is usually double-stranded, mRNA is single stranded. Nucleic acids (DNA or RNA) containing a sequence that can bind to a given mRNA will prevent translation by the protein synthesis machinery, inhibiting gene expression. The Flavr Savr tomato, a tomato engineered to have a long shelf life, was

produced by introducing antisense RNA corresponding to mRNA for an enzyme involved in fruit spoilage. Inhibiting expression of this gene delays spoilage. In 1998 the FDA approved Isis Pharmaceuticals' Vitravene, the first antisense drug, to treat cytomegalovirus-induced retinitis.

TRANSLATION: MAKING PROTEINS

Just as DNA and RNA are composed of linked nucleotides, proteins are comprised of chains of amino acid units. When mRNA is translated to produce a protein, the protein-synthesis machinery "reads" the nucleotides three at a time, assembling amino acid chains that correspond to the mRNA sequence. The basic elements of the protein-synthesis machinery are tRNA, a form of RNA that *transfers* amino acids to the protein-synthesis machinery in a way that enables them to be linked together, and ribosomes, which help form the chemical bonds that attach amino acids in a protein chain.

The three-nucleotide sequence elements on mRNA that code for individual amino acids are called codons. These are matched by

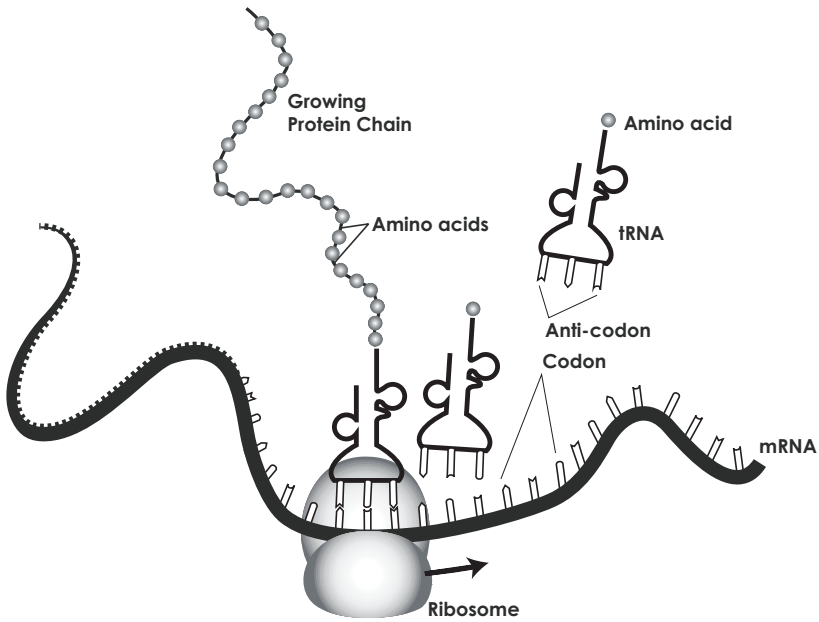

Figure 3.4 *Protein translation*
Modified from National Human Genome Research Institute

anti-codons on tRNA to ensure that the appropriate amino acid is aligned with a given mRNA sequence. The 64 possible combinations of A, C, G, and T at each codon code for only 20 different amino acids. This redundancy in the genetic code, permitting multiple codons to specify common amino acids, is considered a form of protection against DNA mutations and has applications in identifying foreign DNA from sources such as viruses which may use different "dialects" of the genetic code.

The chemical characteristics of amino acids in a protein cause it to fold into a defined 3-dimensional structure. That determines the protein's function. Because the DNA sequence of a gene dictates the sequence of amino acids in a protein, and the sequence of these acids in a protein determines its structure, one can deduce a protein sequence, and potentially its structure and function, from the gene sequence encoding it.

PROTEINS AND ENZYMES

Proteins, the workhorses of cells, are responsible for the majority of structural features and functional characteristics in cells. Enzymes are proteins that perform functional roles as part of the cellular process. Different types of cells get their characteristics by expressing a specific array of genes, resulting in production of a complement of proteins that give each cell type its unique characteristics. Pancreatic cells, for example, produce the protein insulin to regulate blood sugar levels; neurons produce neurotransmitters essential for brain function; and hemoglobin is made in blood cells, enabling them to carry oxygen. Examples of enzymes include proteases that break down proteins or enable digestion of food, and polymerases that assemble DNA and RNA. Some genes are expressed only in certain cell types whereas others are widely expressed. Examples of widely-expressed genes include those encoding proteins and enzymes involved in general cellular activities such as DNA replication, mRNA translation, protein synthesis, energy production and maintenance of structural integrity.

Production of inappropriate proteins in cell types and mis-regulation of protein expression are at the root of many diseases. As

Table 3.1 *Examples of protein and enzyme functions*

Enzyme	Function
Amylase	Breaks down starches and other complex carbohydrates into basic sugars
Cellulase	Breaks down cellulose, found in the cell walls of plants
Lipase	Breaks down fats
Protease	Breaks down proteins
Protein	**Function**
Collagen	Main protein in connective tissue; structural roles in skin, cartilage, teeth, bone, and other tissues
Keratin	Makes skin waterproof and contributes to strength and flexibility
Myosin	Muscle contraction

mentioned above, many cancers result from mis-regulation of gene expression that causes uncontrolled cell division.

Molecular biologists can transfer genes from humans and other animals into bacteria, yeast, and other organisms to confer the ability to produce specific proteins that may be extracted for therapeutic use. For example, Genentech produced its first drug by introducing the gene for human insulin into bacteria and extracted the resulting protein to produce a treatment for human diabetes. Genes can also be transferred from one organism to another to confer new attributes. Pesticide-resistant crops have been produced by incorporating naturally-occuring pesticidal proteins into plants. Bacteria have also been modified to perform roles such as decomposing oil spills by adding genes encoding proteins with the ability to break down components of oil. Additional examples are described in Chapter 6.

Other forms of RNA

Traditional molecular biology held that the primary role of RNA in cells was largely limited to housekeeping functions such as transferring information from DNA to the protein synthesis machinery (mRNA), transporting amino acids to be assembled into proteins (tRNA), and translating mRNA into protein (rRNA).

Sidney Altman and Thomas Cech shared the 1989 Nobel Prize

Table 3.2 Selected *RNA types*

RNA type	Function
mRNA	Messenger RNA. Contains a working copy of a gene sequence and is read by the protein synthesis machinery to produce proteins.
tRNA	Transfer RNA. Transfers amino acids to the protein synthesis machinery to produce proteins.
rRNA	Ribosomal RNA. Part of the protein synthesis machinery. Also useful for determining evolutionary similarity between organisms.
aRNA	Antisense RNA. Used for gene regulation.
siRNA	Small Interfering RNA. Used for gene regulation.
snRNA	Small Nuclear RNA. Used to edit mRNA, regulate gene expression, and maintain chromosome tips (telomeres).

in Chemistry for their discovery of catalytic properties of RNA. The ability to catalyze (increase the rate of) biochemical reactions had previously been thought to only exist in proteins. Altman and Cech found a role for RNA in the splicing of mRNAs, ultimately making it possible for a single gene to give rise to several different proteins. The significance of Altman and Cech's discovery was expanded more than a decade after they received the Nobel Prize. Following sequencing of the human genome it was discovered that the human genome contained only a fraction of the genes previously thought necessary to produce the complete set of proteins comprising human biology. The ability of this small set of genes to produce the full complement of human proteins could largely be explained through mRNA splicing.

More recently myriad forms of RNA have been discovered, and diverse roles for RNA have also been elucidated (see Table 3.2). These discoveries indicate that controlling cellular activities is more complex than previously thought, suggesting that there are also more opportunities to influence cellular activities.

THE BIG PICTURE

Genes interact with the environment and with each other to confer traits. While the presence or absence of a gene can potentially confer a given trait, environmental factors also play a role. Our physical characteristics are a combination of genetic and environmental factors. A child with a hypothetical *tallness* gene, for instance, would not necessarily grow taller than a child without the gene; the child with the *tallness* gene would also require adequate nutrition to fuel the extra growth (and the effect of the *tallness* gene may be limited or enhanced by the action of other genes). Rather than thinking of genes as determinants of physical characteristics, they should be regarded as potentials or predispositions for characteristics.

The ability to modify characteristics of cells is similarly limited by biological and physical constraints. Since some cells are rapidly replaced, induced changes will be quickly lost. Other cells are dormant, precluding their potential to express modifications.

Furthermore, biology is complicated. In fields such as industrial chemistry or engineering, applications are developed from well-characterized principles. With biotechnology on the leading edge of molecular biology research, it can be difficult or impossible to foretell the outcomes of manipulations and they can have unforeseen consequences. Because it is not possible to fully predict the outcome of these procedures, scientists must perform experiments, take observations, refine theories, and finally develop functional applications. This is why biotechnology research is so complex, time consuming, and fraught with unforeseen setbacks and disappointments.