1
Introduction

Enzymes are the catalysts of biological processes. Like any other catalyst, an enzyme brings the reaction catalyzed to its equilibrium position more quickly than would occur otherwise; an enzyme cannot bring about a reaction with an unfavorable change in free energy unless that reaction can be coupled to one whose free energy change is more favorable. This situation is not uncommon in biological systems, but the true role of the enzymes involved should not be mistaken.

The activities of enzymes have been recognized for thousands of years; the fermentation of sugar to alcohol by yeast is among the earliest examples of a biotechnological process. However, only recently have the properties of enzymes been understood properly. Indeed, research on enzymes has now entered a new phase with the fusion of ideas from protein chemistry, molecular biophysics, and molecular biology. Full accounts of the chemistry of enzymes, their structure, kinetics, and technological potential can be found in many books and series devoted to these topics [1–5]. This chapter reviews some aspects of the history of enzymes, their nomenclature, their structure, and their relationship to recent developments in molecular biology.

1.1
History

Detailed histories of the study of enzymes can be found in the literature [6], [7].

Early Concepts of Enzymes. The term "enzyme" (literally "in yeast") was coined by Kühne in 1876. Yeast, because of the acknowledged importance of fermentation, was a popular subject of research. A major controversy at that time, associated most memorably with Liebig and Pasteur, was whether or not the process of fermentation was separable from the living cell. No belief in the necessity of vital forces, however, survived the demonstration by Buchner (1897) that alcoholic fermentation could by carried out by a cell-free yeast extract. The existence of extracellular enzymes had, for reasons of experimental accessibility, already been recognized. For example, as early as 1783, Spallanzani had demonstrated that gastric juice could digest meat in vitro, and Schwann (1836) called the active substance pepsin.

Enzymes in Industry. Wolfgang Aehle

ISBN: 3-527-29592-5

Kühne himself appears to have given trypsin its present name, although its existence in the intestine had been suspected since the early 1800s.

Enzymes as Proteins. By the early 1800s, the proteinaceous nature of enzymes had been recognized. Knowledge of the chemistry of proteins drew heavily on the improving techniques and concepts of organic chemistry in the second half of the 1800s; it culminated in the peptide theory of protein structure, usually credited to Fischer und Hofmeister. However, methods that had permitted the separation and synthesis of small peptides were unequal to the task of purifying enzymes. Indeed, there was no consensus that enzymes were proteins. Then, in 1926, Sumner crystallized urease from jack bean meal and announced it to be a simple protein. However, Willstätter argued that enzymes were not proteins but "colloidal carriers" with "active prosthetic groups". However, with the conclusive work by Northrop *et al.*, who isolated a series of crystalline proteolytic enzymes, beginning with pepsin in 1930, the proteinaceous nature of enzymes was established.

The isolation and characterization of intracellular enzymes was naturally more complicated and, once again, significant improvements were necessary in the separation techniques applicable to proteins before, in the late 1940s, any such enzyme became available in reasonable quantities. Because of the large amounts of accessible starting material and the historical importance of fermentation experiments, most of the first pure intracellular enzymes came from yeast and skeletal muscle. However, as purification methods were improved, the number of enzymes obtained in pure form increased tremendously and still continues to grow. Methods of protein purification are so sophisticated today that, with sufficient effort, any desired enzyme can probably be purified completely, even though very small amounts will be obtained if the source is poor.

Primary Structure. After the protein nature of enzymes had been accepted, the way was clear for more precise analysis of their composition and structure. Most amino acids had been identified by the early 20th century. The methods of amino acid analysis then available, such as gravimetric analysis or microbiological assay, were quite accurate but very slow and required large amounts of material. The breakthrough came with the work of Moore and Stein on ion-exchange chromatography of amino acids, which culminated in 1958 in the introduction of the first automated amino acid analyzer [8].

The more complex question – the arrangement of the constituent amino acids in a given protein, generally referred to as its primary structure – was solved in the late 1940s. The determination in 1951 of the amino acid sequence of the β-chain of insulin by Sanger and Tuppy [10] demonstrated for the first time that a given protein does indeed have a unique primary structure. The genetic implications of this were enormous. The introduction by Edman of the phenyl isothiocyanate degradation of proteins stepwise from the N-terminus, in manual form in 1950 and subsequently automated in 1967 [11], provided the principal chemical method for determining the amino acid sequences of proteins. The primary structures of pancreatic ribonuclease [12] and egg-white lysozyme [13] were published in 1963. Both of these

enzymes, simple extracellular proteins, contain about 120 amino acids. The first intracellular enzyme to have its primary structure determined was glyceraldehyde 3-phosphate dehydrogenase [14], which has an amino acid sequence of 330 residues and represents a size (250 – 400 residues) typical of many enzymes. Protein sequencing is increasingly performed by liquid chromatography/mass spectrometry (LC/MS) techniques, and several tools and software packages are now available for protein identification and characterization. The methods of protein sequence analysis are now so well developed that no real practical deterrent exists, other than time or expense, to determination of the amino acid sequence of any polypeptide chain [9].

A more recent fundamental concept called proteome (*prote*in complement to a gen*ome*) will enable researchers to unravel biochemical and physiological mechanisms of complex multivariate diseases at the functional molecular level. A new discipline, *proteomics*, complements physical genome research. Proteomics can be defined as "the qualitative and quantitative comparison of proteomes under different conditions to further unravel biological processes" [15].

Active Site. The fact that enzymes are highly substrate specific and are generally much larger than the substrates on which they act quickly became apparent. The earliest kinetic analyses of enzymatic reactions indicated the formation of transient enzyme – substrate complexes. These observations could be explained easily if the conversion of substrate to product was assumed to occur at a restricted site on an enzyme molecule. This site soon became known as the active center or, as is more common today, the active site.

Particular compounds were found to react with specific amino acid side chains and thus inhibit particular enzymes. This suggested that such side chains might take part in the catalytic mechanisms of these enzymes. An early example was the inhibition of glycolysis or fermentation by iodoacetic acid, which was later recognized as resulting from reaction with a unique cysteine residue of glyceraldehyde 3-phosphate dehydrogenase, which normally carries the substrate in a thioester linkage [16].

Many such group-specific reagents have now been identified as inhibitors of individual enzymes; often they are effective because of the hyper-reactivity of a functionally important side chain in the enzyme's active site. However, a more sophisticated approach to the design of enzyme inhibitors became possible when the reactive group was attached to a substrate; in this way, the specificity of the target enzyme was utilized to achieve selective inhibition of the enzyme [17]. Such active-site-directed inhibitors have acquired major importance not only academically in the study of enzyme mechanisms but also commercially in the search for a rational approach to selective toxicity or chemotherapy.

Three-Dimensional Structure. Chemical studies showed that the active site of an enzyme consists of a constellation of amino acid side chains brought together spatially from different parts of the polypeptide chain. If this three-dimensional structure was disrupted by denaturation, that is, without breaking any covalent bonds, the biological activity of the enzyme was destroyed. In addition, it was found that all

the information required for a protein to fold up spontaneously in solution and reproduce its native shape was contained in its primary structure. This was part of the original "central dogma" of molecular biology.

The X-ray crystallography of proteins [18] demonstrated unequivocally that a given protein has a unique three-dimensional structure. Among the basic design principles was the tendency of hydrophobic amino acid side chains to be associated with the hydrophobic interior of the folded molecule, whereas charged side chains were almost exclusively situated on the hydrophilic exterior or surface. The first high-resolution crystallographic analysis of an enzyme, egg-white lysozyme, confirmed these principles and led to the proposal of a detailed mechanism [19]. The active site was located in a cleft in the structure (Fig. 1), which has subsequently proved to be a common feature of active sites. According to this, the enzymatic reaction takes place in a hydrophobic environment, and the successive chemical events involving substrate and protein side chains are not constrained by the ambient conditions of aqueous solution and neutral pH.

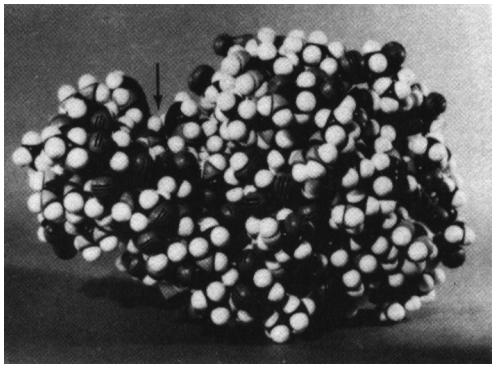

Figure 1. A molecular model of the enzyme lysozyme: the arrow points to the cleft that accepts the polysaccharide substrate (Reproduced by courtesy of J. A. Rupley)

1.2 Enzyme Nomenclature

Strict specificity is a distinguishing feature of enzymes, as opposed to other known catalysts. Enzymes occur in myriad forms and catalyze an enormous range of reactions. By the late 1950 s the number of known enzymes had increased so rapidly that their nomenclature was becoming confused or, worse still, misleading because the same enzyme was often known to different workers by different names; in addition, the name frequently conveyed little or nothing about the nature of the reaction catalyzed.

To bring order to this chaotic situation, an International Commission on Enzymes was established in 1956 under the auspices of the International Union of Biochem-

istry (IUB). Its terms of reference were as follows: "To consider the classification and nomenclature of enzymes and coenzymes, their units of activity and standard methods of assay, together with the symbols used in the description of enzyme kinetics." The Commission's recommendations have formed the basis of enzyme nomenclature since its first report in 1961 [1].

Responsibility for enzyme nomenclature passed to the Nomenclature Committee of IUB in 1977, which has subsequently published several reports, e.g., [20] and supplements, e.g., [21]; it is expected that further supplements will be published from time to time in the *European Journal of Biochemistry*. The growth in scale can be appreciated from the fact that the 1961 *Report of the Enzyme Commission* listed 712 enzymes, whereas the 1992 version of Enzyme Nomenclature listed 3196. The most recent information about changes or additions to enzyme nomenclature is available at http://www.chem.qmw.ac.uk/iubmb/, which also offers an up-to-date version of the *Enzyme Nomenclature* list.

1.2.1 General Principles of Nomenclature

The accepted system for classification and nomenclature of enzymes embodies three general principles.

The first is that enzyme names, especially those ending in *-ase,* should be used only for single enzymes, i.e., single catalytic entities. They should not be applied to systems containing more than one enzyme.

The second general principle is that an enzyme is named and classified according to the reaction it catalyzes. This refers only to the observed chemical change produced by the enzyme, as expressed in the chemical equation. The mechanism of action is ignored, and intermediate cofactors or prosthetic groups are not normally included in the name. Thus, an enzyme cannot be named systematically until the reaction it catalyzes has been identified properly.

The third general principle is that enzymes are named and classified according to the type of reaction catalyzed, which enables Enzyme Commission (E.C.) code numbers to be assigned to enzymes to facilitate subsequent unambiguous identification. For the purpose of systematic nomenclature, all enzymes in a particular class are considered to catalyze reactions that take place in a given direction, although only the reverse direction may have been demonstrated experimentally. However, the recommended name for the enzyme may well be based on the presumed direction of the reaction in vivo.

Thus, a given enzyme often has two names, one systematic and the other recommended or trivial. The latter is generally the name in current usage, shorter and more readily applied. After its systematic name and E.C. code number have identified an enzyme, the recommended name can be used without fear of ambiguity. This practice is now generally followed in the literature.

1.2.2 Classification and Numbering of Enzymes

According to the report of the first Enzyme Commission in 1961, enzymes are divided into six main classes according to the type of reaction catalyzed. They are assigned code numbers, prefixed by E.C., which contain four elements separated by points and have the following meaning:

1. the number first indicates to which of the six classes the enzyme belongs,
2. the second indicates the subclass,
3. the third number indicates the sub-subclass, and
4. the fourth is the serial number of the enzyme in its sub-subclass.

The six classes are distinguished in the following manner:

1. *Oxidoreductases*
   This class encompasses all enzymes that catalyze redox reactions. The recommended name is *dehydrogenase* whenever possible, but *reductase* can also be used. *Oxidase* is used only when O_2 is the acceptor for reduction. The systematic name is formed according to *donor : acceptor oxidoreductase*.
2. *Transferases*
   Transferases catalyze the transfer of a specific group, such as methyl, acyl, amino, glycosyl, or phosphate, from one substance to another. The recommended name is normally *acceptor grouptransferase* or *donor grouptransferase*. The systematic name is formed according to *donor : acceptor grouptransferase*.
3. *Hydrolases*
   Hydrolases catalyze the hydrolytic cleavage of C—O, C—N, C—C, and some other bonds. The recommended name often consists simply of the substrate name with the suffix *-ase*. The systematic name always includes *hydrolase*.
4. *Lyases*
   Lyases catalyze the cleavage of C—C, C—O, C—N, and other bonds by elimination. The recommended name is, for example, *decarboxylase, aldolase, dehydratase* (elimination of CO_2, aldehyde, and water, respectively). The systematic name is formed according to *substrate group-lyase*.
5. *Isomerases*
   Isomerases catalyze geometric or structural rearrangements within a molecule. The different types of isomerism lead to the names *racemase, epimerase, isomerase, tautomerase, mutase,* or *cycloisomerase*.
6. *Ligases*
   Ligases catalyze the joining of two molecules, coupled with the hydrolysis of a pyrophosphate bond in ATP or another nucleoside triphosphate. Until 1983, the recommended name often included *synthetase,* but the current recommendation is that names of the type *X – Y ligase* be used instead, to avoid confusion with the name *synthase* (which is not confined to enzymes of class 6). The systematic name is formed according to *X : Y ligase* (ADP-forming).

A few examples will serve to illustrate how this system works. (The full list can be found in *Enzyme Nomenclature* 1992 [20].)

The enzyme alcohol dehydrogenase (recommended name) catalyzes the reaction

$$\text{Alcohol} + \text{NAD}^+ \rightleftharpoons \text{Aldehyde or Ketone} + \text{NADH} + \text{H}^+$$

The enzyme has been assigned E.C. number 1.1.1.1. It may also be called aldehyde reductase, but its systematic name is alcohol: NAD^+ oxidoreductase.

Similarly, the enzyme hexokinase (recommended name), which catalyzes the reaction

$$\text{ATP} + \text{D-Hexose} \rightleftharpoons \text{ADP} + \text{D-Hexose 6-phosphate}$$

has been given the E.C. number 2.7.1.1. It has such other names as glucokinase and hexokinase type IV, and its systematic name is ATP: D-hexose 6-phosphotransferase.

1.3 Structure of Enzymes

Enzymes are proteins (for an exception, see Section 1.3.4) and, as such, are amenable to structural analysis by the methods of protein chemistry, molecular biology, and molecular biophysics.

1.3.1 Primary Structure

The primary structure of enzymes can be determined by direct chemical methods which, in sensitivity and automation, have reached very high levels of sophistication [9], [22]. However, for many proteins, particularly those with long polypeptide chains, direct sequence analysis would be very time-consuming; others may be available only in very small amounts. In these cases, a more profitable approach is to clone the relevant structural gene and determine its DNA sequence [9], [23], [24]. From this, the amino acid sequence can be inferred. Whenever possible, this sequence should be checked, e.g., for genetic reading frame, against whatever amino acid sequence information is available from direct methods. The recombinant DNA approach is so quick and so powerful, however, that amino acid sequence information about enzymes is growing much more rapidly from this source than from direct chemical analysis [25], [26]. Indeed, the information now available is so large in total that computer data banks are required to store it and make it available for systematic access [27].

1.3.2 Three-Dimensional Structure

The three-dimensional structure of an enzyme can be obtained at high resolution by X-ray crystallography [28] and, for molecules up to ca. 300 amino acids in length, by NMR spectroscopy. By this means, the detailed structures of many enzymes have been determined, and a broad understanding of the principles of protein structure has resulted [29], [30]. Proteins are generally well-ordered; their interiors are well-packed (comparable to other crystalline organic molecules) to produce a hydrophobic core with a dielectric constant similar to that of a hydrocarbon. Proteins vary in the amount of regular secondary structure (α-helix and β-sheet) they contain and can be grouped into four classes according to the combination and packing of these structural features [31]. Although the number of possible combinations of amino acids in a given protein is virtually unlimited, it is estimated that there are not more 1000 different families of folding patterns for protein structures [32].

Despite their close-packed and generally well ordered structure, enzymes are usually not entirely rigid molecules, and some conformational flexibility in solution is widely observed, particularly by NMR spectroscopy [33–37]. These conformational changes may be limited to a molecular "breathing" or flexing of the structure, they may involve various "hinge-bending" motions, or they may extend to more substantial conformational mobility in parts of the polypeptide chain. All such motions, contribute to the mechanisms of enzyme catalysis [2], [38].

As of August 28, 2001, 15 905 3D structures were freely accessible from the website of the Protein Data Bank (http://www.rcsb.org/pdb/) [39]. This site also offers several programs for analysis and visualization of protein (and DNA) 3D structures on numerous computer platforms.

1.3.3 Quaternary Structure, Folding, and Domains

Many enzymes consist of more than one polypeptide chain (or subunit), and these must form an aggregate, usually with relatively simple symmetry, before full (or even any) biological activity is conferred (Table 1). The subunits within an oligomer or multimer are often identical or at least limited to a few different types. Aggregation is generally some form of self-assembly dictated by coherent binding patterns between the subunits, which provide the necessary recognition sites in sorting out the subunits required for assembly [29], [40].

The complexity of this sorting process in a cell becomes evident from the fact that many intracellular enzymes are dimers or tetramers. Increasingly more complicated structures are being recognized and their design principles analyzed. These range from enzymes with simple cyclic symmetry up to those with the most elaborate cubic point group symmetry, e.g., octahedral and icosahedral [29], [40].

The folding of polypeptide chains, along with their aggregation into ordered structures, is a spontaneous process in solution, and this implies that it is exergonic [39]. However, calculation of the time required for a protein to explore all possible struc-

Table 1. Quaternary structures of some typical enzymes

Enzyme	E.C. number *[CAS registry number]*	Source	Number of subunits	Point symmetry	
				Crystallographic symbol	Schönflies symbol
Alcohol dehydrogenase	1.1.1.1 [9031-72-5]	horse liver	2	2	C_2
Glutathione reductase	1.6.4.2 [9001-48-3]	human red blood cells	2	2	C_2
Triose phosphate isomerase	5.3.1.1 [9023-78-3]	chicken muscle	2	2	C_2
Lactate dehydrogenase	1.1.1.27 [9001-60-9]	dogfish muscle	4	222	D_2
Glyceraldehyde 3-phosphate dehydrogenase	1.2.1.12 [9001-50-7]	*Bacillus stearothermophilus*	4	222	D_2
Pyruvate kinase	2.7.1.40 [9001-59-6]	cat muscle	4	222	D_2
Aspartate carbamoyltransferase	2.1.3.2 [9012-49-1]	*Escherichia coli*	6 + 6	32	D_3
Dihydrolipoamide acetyltransferase	2.3.1.12 [9032-29-5]	*Escherichia coli*	24	432	0
Dihydrolipoamide acetyltransferase	2.3.1.12 [9032-29-5]	*Bacillus stearothermophilus*	60	532	Y

tures during the folding process indicates that the search for the "right" structure cannot be entirely random. Thus, even for a small protein such as bovine pancreatic ribonuclease (124 amino acid residues), such a search might take around 10^{95} years, whereas the experimentally determined time in vivo is a few milliseconds. This dramatic discrepancy led to the concept of *kinetic pathways* during folding. Such pathways have been experimentally explored, and intermediates identified for various proteins. The stable structure of a protein in solution is therefore identified as the lowest free energy form of the kinetically accessible structures [29], [30], [40].

A typical enzyme is not an entity completely folded as a whole, as is evident from the growing catalogue of three-dimensional protein structures determined by X-ray crystallography. On the contrary, enzymes frequently consist of apparently autonomous or semiautonomous folding units, called *domains* (Fig. 2). Sometimes, these may be identified as products of limited proteolysis, i.e., regions of the polypeptide chain that can be excised from the chain with retention of their biological properties. Indeed, this has proved in many instances to be a valuable guide to the actual activity contributed by that part of the enzyme. Classical examples of such functional domains can be found in the study of muscle contraction and antibody-antigen recognition [29], [30].

In other cases, domains are not readily released as biologically active entities, and their existence must be inferred from the three-dimensional structure of the

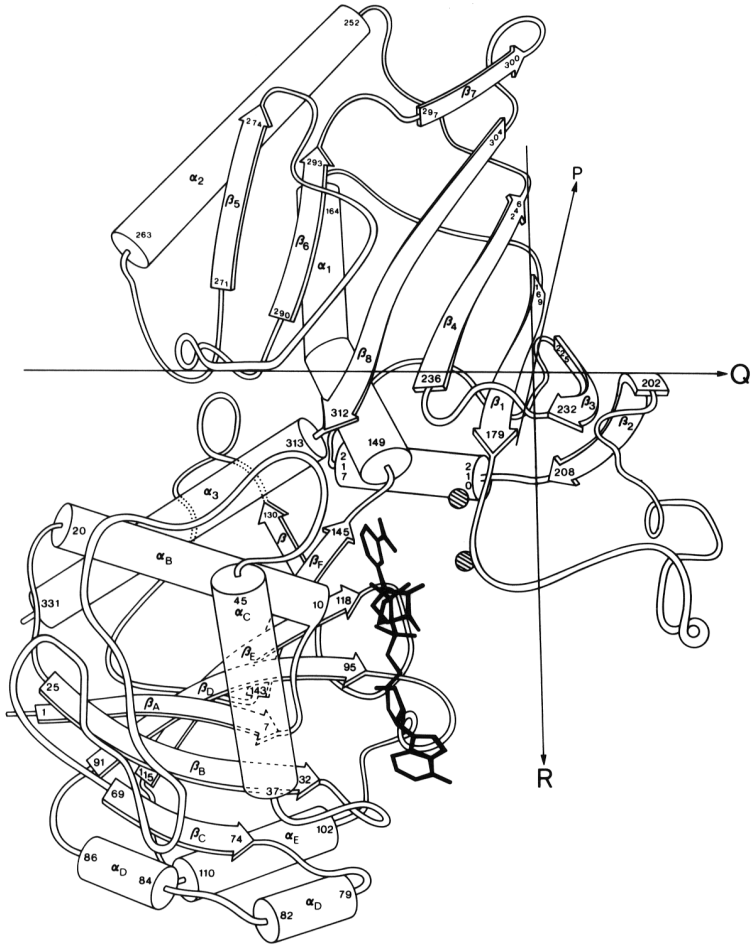

Figure 2. The domains in glyceraldehyde 3-phosphate dehydrogenase from *Bacillus stearothermophilus* [467]
Reproduced with permission.

enzyme. Most globular proteins can in fact be subdivided into such regions, which generally have molecular masses of 20 000 or less [29]. The active site of an enzyme is often located at the interface between two such domains as, for example, in the well-known cleft of lysozyme (Fig. 1) or in glutathione reductase. Other domains appear to represent favored folding patterns in the assembly of proteins, but biological activity associated with them can often be inferred from comparison of the structures of related proteins: a typical example is the NAD-binding domain present in dehydrogenases.

Structural domains may be regions of the polypeptide chain that fold independently of each other. Functional domains, as defined above, do indeed fold independently; and individual subunits of oligomeric enzymes appear to fold before association [29], [30], [40], [41].

1.3.4 The Ribozyme

Enzymes are proteins, but the specific involvement of RNA molecules in certain reactions concerned with RNA processing in vivo is worth noting. Until Cech *et al.* [42] and Altman *et al.* [43] published their observations, it was generally accepted knowledge that the major duties in a biological system, namely, to encode information and to catalyze chemical reactions, are neatly split, one being performed by nucleic acids, the other by proteins. With the discovery of special RNAs which store genetic information and can also catalyze reactions on themselves or on other RNAs, this dogma was destroyed [42], [43]. Over the years, it has become evident, that group I and group II introns, catalyze various transesterifications. In cellular systems these reactions facilitate their excision from pre-RNAs and the ligation of flanking exons (self-splicing). In vitro these intron RNAs perform a variety of reactions in cis (i.e., on the same strand of the RNA genome) and in trans (i.e., on another RNA), such as cleavage and ligation of RNAs, transfer of nucleotides between RNAs, polymerization, and editing-like reactions. These RNAs thus can act as enzymes and are therefore called "ribozymes" [44].

In *Escherichia coli*, tRNA precursors are cleaved by ribonuclease P to generate the correct 5′-ends of the mature tRNA molecules, and the enzyme contains an essential RNA moiety that can function in the absence of the protein. In fact, this RNA moiety fulfills all the criteria of an enzyme [45]. Similarly, the ribosomal RNA of *Tetrahymena thermophila* undergoes self-splicing to perform a highly specific intramolecular catalysis in the removal of an intervening sequence. A truncated version of the intervening sequence, lacking the first 19 nucleotides of the original excised RNA, can then behave as an enzyme in vitro, capable of acting as an RNA polymerase and a sequence-specific ribonuclease under appropriate conditions [46].

The structure of the ribosome's large subunit has since been solved. This largest unique structure established that the ribosome is a ribozyme in which the ribosomal RNA, and not the protein, performs catalytic functions, including the peptidyl transferase reaction that forms the peptide bond [47], [48]. One of the most remarkable findings to emerge from this is that although enzymes composed entirely of protein promote virtually all chemical reactions that occur in living organisms, the protein synthesis reaction that occurs on the ribosome is due to the two-thirds of its mass that is RNA, not the one-third that is protein. In addition to enhancing the understanding of protein synthesis, this work will have significant medical implications, because the ribosome is a major target for antibiotics [49].

Ribozymes also offer an excellent opportunity to compare and contrast the behavior of RNA enzymes with that of protein enzymes. The differences between the RNA and protein enzymes highlight features that are distinct and thus enable a better understanding of each of these classes of biological macromolecules. On the other hand, the features of protein and RNA enzymes that are similar may represent aspects that are fundamental to biological catalysis. Indeed, these studies have suggested that RNA enzymes, like their protein counterparts, can use binding interactions remote from the site of bond transformation to facilitate that transformation

[50]. Beyond this, recent results suggest that RNA enzymes are ideally suited for exploration of the energetic origins of this interconnection between binding and catalysis [51]. This use of binding energy provides a natural connection between rate enhancement and specificity, the two hallmarks of biological catalysis. Finally, ribozymes will not only offer new clues about evolution [52], but also offer the potential for specific inactivation of disease-associated mRNAs or viral RNA genomes that, unlike conventional therapeutics, require no knowledge of the structure or function of proteins that target RNAs encode [53].

1.4 Biosynthesis of Enzymes

Enzymes are synthesized in cells by the normal machinery of protein synthesis. The structure of any given enzyme is encoded by a structural gene, whose DNA base sequence is transcribed into a messenger RNA, and the mRNA is translated from its triplet code into the amino acid sequence of the desired protein by the ribosomes and associated factors [54], [55]. The enzyme then folds spontaneously into its active conformation. Posttranslational modifications may be required to target an enzyme to its ultimate intracellular or extracellular location.

1.4.1 Enzymes and DNA

For many years, the chemical manipulation of DNA lagged behind that of proteins. The chemical complexity and variety of proteins, with up to 20 different naturally occurring amino acids, served to make them more amenable to increasingly sophisticated methods of analysis. On the other hand, DNA, composed of only four different nucleotides, appeared dauntingly large, with few structural features to make it yield to available methodology.

Paradoxically, this very lack of variety in the nature of the constituent nucleotides of DNA has permitted the revolution in genetic engineering, in which the enzymology of DNA [56] has played a prominent part. For example, the discovery and purification of restriction enzymes enabled DNA to be cleaved selectively into defined fragments; phosphatases and ligases permit the fragments to be rejoined selectively; and DNA polymerases allow DNA to be synthesized and sequenced at astonishing speed, all in vitro [23], [54–56].