Organization for Protein Biosynthesis

Alan Wiseman

Organization for Protein Biosynthesis

Alan Wiseman
B.Sc. A.R.C.S. Ph.D. M.I.Biol. A.R.I.C.
Scientific Staff of Medical Research Council
Metabolic Reactions Research Unit,
Department of Biochemistry,
Imperial College of Science and Technology,
University of London

Blackwell
Scientific Publications
Oxford

First published 1965

Printed in Great Britain by
Spottiswoode, Ballantyne and Co Ltd
London and Colchester
and bound by
Webb, Son and Co Ltd
London

Preface

A large number of proteins are synthesized by the living cell for a variety of functions. Many are enzymic catalysts of chemical reactions, others form part of the intracellular structures seen under the microscope. Recent investigations have confirmed that certain proteins, at least, consist of a unique sequence of amino acids. This book sets out to describe in a readily understandable manner how this genetically-determined sequence is specified and constructed, so forming the molecular species necessary to perform the functions of the cell.

Research on this fundamental topic has been performed on a large number of organisms and a variety of plant and animal tissues. Especially popular have been studies upon the intestinal bacterial species *Escherichia coli*, immature red blood corpuscles (reticulocytes), pea-seedlings, ascites tumour cells, rat liver, and yeast. In this connection the author has had considerable personal experience of the last two materials in this list, and in fact all the fundamental discoveries concerning the so-called classical pathway of protein biosynthesis (see on, Part A) seem to apply in their essentials to all these systems. Mention is made in the text of the particular organism used to obtain the results, and significant differences are pointed out where appropriate.

This book is intended both as an introduction to the field for students of biochemistry, microbiology, physiology and chemistry, and as a reference book and entry to the literature for research workers in related fields. Review references are provided wherever expedient and the other references are often as recent as possible except when historical aspects are under review. Sections on Recent Trends and Future Developments are included at the ends of chapters, and these are usually reserved for the frontiers of the subject. The student reader is advised to read the Summary near the end of Chapters 2 to 11 before starting upon the chapter itself so that the fundamentals of the subject may be clearly understood.

Many thanks are due to J.A.W. for preparing the diagrams and assisting with the reading of the proofs, and to M.M.T. for typing the manuscript.

Contents

Preface

Introduction

1 The Molecular Structure of Proteins and Nucleic Acids
I PROTEINS
(*a*) *The primary structure of proteins*, 3. (*b*) *Structural studies of insulin*, 7. (*c*) *Other proteins*, 11. (*d*) *The nature of enzymes*, 12.

II NUCLEIC ACIDS
(*a*) *Primary structure*, 13. (*b*) *Comparative account of the structures of DNA and RNA*, 15.

A Biochemical Mechanisms in the 'Classical Scheme' of Protein Biosynthesis

2 ATP and Activation of Amino Acids 23
(*a*) *Formation of ATP*, 24. (*b*) *Function of ATP*, 27. (*c*) *Mechanism and measurement of activation of amino acids*, 29. (*d*) *The amino acid adenylates*, 32: isolation, 32; synthetic routes, 33; high reactivity, 34. (*e*) *Amino acid activating enzymes*, 35: mode of action, 35; isolation and purification, 36; specificity, 37. (*f*) *Evidence for an acceptor of activated amino acid* (Transfer RNA), 38.

SUMMARY TO CHAPTER 2, 39

RECENT DEVELOPMENTS, 39

FUTURE TRENDS, 40

3 Transfer Ribonucleic Acid—its nature and role 41
(*a*) *Isolation and separation of amino acid-specific species of transfer RNA*, 41: molecular weight, 42; amino acid-specific transfer RNA species, 43. (*b*) *Structure of transfer RNA*, 45: methods of sequential degradation, 46; specific degradation of transfer RNA, 47; the presence of abnormal bases, 49. (*c*) *Specificity of transfer RNA*, 50: species specificity, 51. (*d*) *Formation and properties of transfer RNA–AA compounds*, 52: the chemical properties of transfer RNA–AA, 52. (*e*) *Transfer of AA from transfer RNA–AA to protein*, 53: studies *in vivo*, 54; certain additives are required, 55; transfer enzymes, 55; after transfer, 56; inhibitors of transfer, 56; certain well-known proteins, 57.

SUMMARY TO CHAPTER 3, 58

RECENT DEVELOPMENTS, 58
X-ray diffraction data

FUTURE TRENDS, 60
Site of cellular origin of transfer RNA

4 Ribosomes—*in vivo* and *in vitro*; their structure and function 62
(*a*) *Studies on living cells*, 62. (*b*) *Methods of isolation of ribosomes*, 65: disruption of the cell, 65; ultracentrifugation, 65. (*c*) *Structural studies on ribosomes*, 66: chemical composition, 66; physical properties, 67; aggregates of ribosomes, 69. (*d*) *Role of ribosomes in cell-free preparations*, 70: stepwise synthesis and unequal labelling, 73; nascent proteins, 75.

SUMMARY TO CHAPTER 4, 76

RECENT DEVELOPMENTS, 77
Origin of ribosomal RNA

FUTURE TRENDS, 78

B Specificity and Controlling Mechanisms of Enzyme Biosynthesis (classical scheme)

5 Directing Function of DNA 83
(*a*) *Structural studies on DNA*, 83. (*b*) *Replication of DNA*, 85. (*c*) *DNA—the genetic determinant*, 86: bacterial transformation, 87; studies on bacteriophage, 88; transduction, 88; bacterial-mating (conjugation), 89. (*d*) *The gene concept*, 89: mapping of the positions of genes, 94. (*e*) *Synthesis of messenger RNA*, 95.

SUMMARY TO CHAPTER 5, 96

RECENT DEVELOPMENTS, 97
The replication of DNA

FUTURE TRENDS

6 Occurrence and Function of Messenger RNA (M.RNA) 99
(*a*) *Discovery of messenger RNA in bacteria, infected by phage*, 99. (*b*) *Occurrence of messenger RNA in normal bacteria*, 101. (*c*) *Widespread occurrence of messenger RNA; proof of role*, 104. (*d*) *Template role of messenger RNA*, 107: the adaptor hypothesis, 108. (*e*) *Completion of proteins*, 110.

SUMMARY TO CHAPTER 6, 111

RECENT DEVELOPMENTS, 112
Messenger RNA in animal cells, 112; actinomycin D, 114.

FUTURE TRENDS, 114

7 Elucidation of Triplets of Nucleotides constituting the Genetic Coding Sequences (CODONS) for each Amino Acid 116
(*a*) *Replacement data from naturally-occurring amino acid substitutions*, 117. (*b*) *Data from mutations in tobacco mosaic virus (TMV)*, 118. (*c*) *Synthetic polynucleotides—directed incorporation of amino acids in cell-free systems from* E. coli, 120. (*d*) *Synthetic polynucleotides—other features*, 123.

SUMMARY TO CHAPTER 7, 124

RECENT DEVELOPMENTS, 125

FUTURE TRENDS, 127

C Regulation of Enzyme Biosynthesis

8 Induced Synthesis of Enzymes in Micro-organisms 131
(*a*) *Adaptation to environment*, 131. (*b*) *Sequential and gratuitous induction*, 133: kinetics of induction, 134. (*c*) *Deadaptation*, 135. (*d*) *Active transport of sugars by permease systems*, 136. (*e*) *Repression of enzyme synthesis, end-product and catabolite*, 138: end-product repression, 138; catabolite repression, 139. (*f*) *Mode of action of inducers and repressors (end-product)—genetic regulation*, 140.

SUMMARY TO CHAPTER 8, 143

RECENT DEVELOPMENTS, 143

FUTURE TRENDS, 144

9 Other Regulatory Systems for Protein Biosynthesis 146
(*a*) *Induced synthesis of enzymes in animal tissues*, 146. (*b*) *Production of antibodies—an induced synthesis*, 147. (*c*) *End-product inhibition of the action of enzymes*, 149. (*d*) *Regulation of the synthesis of ribonucleic acid (RNA)—in relation to protein biosynthesis*, 151. (*e*) *Regulation reflected in cell-free systems*, 152. (*f*) *Differentiation of animal cells—repression of genes* 152.

SUMMARY TO CHAPTER 9, 153

RECENT DEVELOPMENTS, 154

FUTURE TRENDS, 154

D Other Pathways of Protein Biosynthesis

10 Protein Synthesis in Cellular Organelles—other than cytoplasmic ribosomes 159
(*a*) *In the nucleus of the cell*, 159. (*b*) *In the mitochondrion of the cell*, 162. (*c*) *In bacterial cytoplasmic membranes—Role of Lipids*, 165.

SUMMARY TO CHAPTER 10, 166

RECENT DEVELOPMENTS, 166

FUTURE TRENDS, 167

11 A Role for Activated Peptides? 168
(*a*) *Distribution, Isolation and properties*, 168. (*b*) *Degree of activation*, 171. (*c*) *Structures of anhydrides from yeast*, 173. (*d*) *Presence of 'Esters' in yeast extracts*, 176. (*e*) *Changes in activated peptides during metabolic shifts*, 177. (*f*) *Isolation of enzyme systems producing activated peptides*, 179, (*g*) *Role of activated peptides*, 180.

SUMMARY TO CHAPTER 11, 182

RECENT DEVELOPMENTS, 182
Protein turnover, 183

FUTURE TRENDS, 183

12 Conclusions and Comments 185

References 193

Indexes 205

Introduction

Chapter One
The Molecular Structure of Proteins and Nucleic Acids

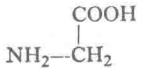
I Proteins

(*a*) *The Primary Structure of Proteins*

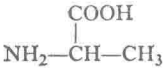
Protein molecules are derived from the 'head to tail' condensation of many units of each of about twenty amino acids (see fig. 1). Every free amino acid used for the building of proteins has an amino and a carboxyl grouping attached to the same carbon atom. Further, often complex, groups are attached to this carbon atom so forming the twenty recognized varieties of amino acid. If one considers a linear arrangement of amino acids, each amino grouping is seen to be adjacent to a carboxyl grouping in its neighbour. Loss of the atoms of one molecule of water per amino acid with the simultaneous linking by peptide-bond formation of the remaining amino acid residues then gives rise to the protein molecule (see fig. 2). This idealized scheme for the formation of protein obviously allows protein molecules to vary in length and to have in general a free amino group at one end of the chain-like sequence and a free carboxyl group at the other end. These features of the structure of protein molecules will be discussed more fully later, for their synthesis is now known to be a stepwise process rather than the simultaneous condensation of several amino acids. The protein chain is lengthened by one amino acid at a time, where the latest amino acid to arrive at the site of protein synthesis joins on to the end of the growing protein chain. The latter has its free carboxyl grouping in reactive form due to combination with transfer ribonucleic acid (see chapter 3). Because the amino acids are carboxyl-activated, chain-growth proceeds from what will be the free amino (N-terminal) end of the finished protein chain. Thus the amino acid required for the free amino end of the protein chain is the first to be laid down on the ribonucleic acid building site (template action of messenger ribonucleic acid on ribosome support; chapter 6).

Protein chains can consist of well over a hundred amino acid residues, and in fact some protein molecules contain more than one protein

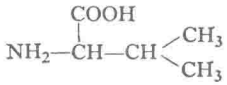
A. 1. Glycine (Gly)



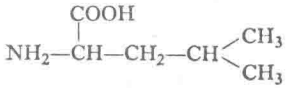
2. Alanine (Ala)



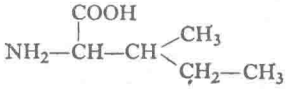
3. Valine (Val)



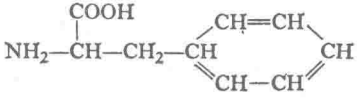
4. Leucine (Leu)



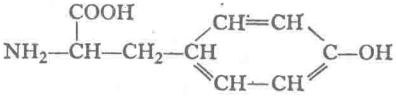
5. Iso-Leucine (Ileu)



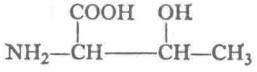
B. 6. Phenylalanine (Phe)



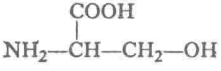
7. Tyrosine (Tyr)



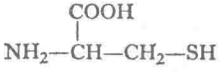
C. 8. Threonine (The)



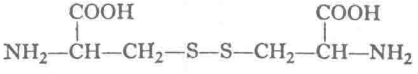
9. Serine (Ser)



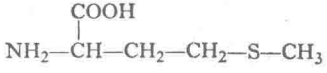
10. Cysteine (CySH)



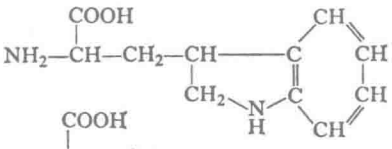
11. Cystine (Cys)



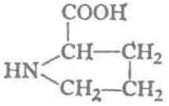
12. Methionine (Met)



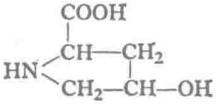
D. 13. Tryptophan (Try)



14. Proline (Pro)

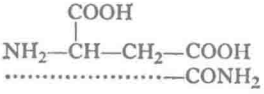


15. Hydroxyproline (HOPro)

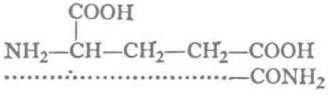


E. *Acidic Amino Acids and their Amides*

16. Aspartic acid (Asp)
Asparagine (Asp NH_2)

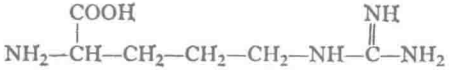


17. Glutamic acid (Glu)
Glutamine (Glu NH_2)

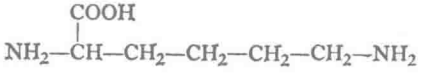


F. *Basic Amino Acids*

18. Arginine (Arg)



19. Lysine (Lys)



20. Histidine (His)

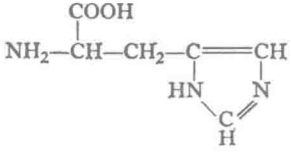


FIG. 1. Amino acids usually found in proteins—neutral amino acids categories A–D

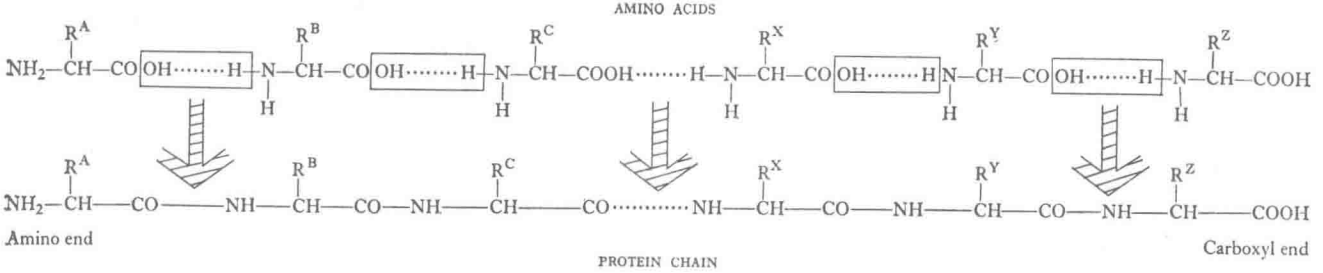


FIG. 2. Primary structure of proteins—formation from amino acids

chain. Haemoglobin (see on) the oxygen-carrying substance of blood contains the protein called globin, which consists of four such chains, two of one amino acid sequence, and two of another. It was precisely the fact that proteins often have high molecular weights that was responsible for their early recognition, and naming as colloids; they being non-dialysable and viscous substances which displayed opalescence in solution. The explanation for most of these properties can be readily formulated simply on the basis of the high molecular weight. Thus the molecules of protein are too large to pass through the fine molecular mesh of dialysis membranes, while their high viscosity can be considered, in simplified terms, to be due to the entangling of their long chains, giving rise to a dragging friction.

(*b*) *Structural Studies of Insulin**

The characteristic activity, enzymic or hormonal, of certain proteins is known to be directly controlled by the linear arrangement of their constituent amino acids in each of their protein chains. Although the particular proportions of the various amino acids in many proteins have been known for some time, the complete sequence of amino acids in a protein molecule was reported for the first time, as recently as 1955, by Sanger and co-authors [1]. They elucidated the structure of insulin, the well-known hormone popularly associated with the metabolism of carbohydrate, which can be conveniently obtained in reasonable quantity from ox pancreas. The amino acid composition of insulin had been extensively investigated by previous workers. In general, the determination of the gross amino acid composition of a protein involves the breakage of the peptide bonds joining the amino acid residues by a hydrolytic procedure with hot mineral acid; often 6N-hydrochloric acid for sixteen hours at 100°C. Almost all the amino acids of a fairly pure protein can be recovered in good yield after this treatment, and a satisfactory correction can be made for the losses due to the destruction of the more labile amino acids. In addition to this, one amino acid called tryptophan (fig. 13) is destroyed. However, it can be recovered, unlike many amino acids, after the breakdown of protein with hot alkaline solutions. Insulin, in fact, was found to be free of tryptophan, so this complicating factor was absent.

Physical studies on purified insulin gave some promise that this protein, unlike many others, was of relatively small size, a feature which would enormously facilitate the subsequent investigation. An

* Recently synthesized by chemical methods.

account of these techniques can be found in *The Proteins* [2]. The presence of the amino acid cystine in relatively large amounts in insulin allows the possibility that each half of this amino acid is in a distinct protein chain. Reference to fig. 3 shows how these protein chains are joined together by the two sulphur atoms constituting the disulphide bond. This bond must be broken before the discrete protein chains in the intact hormone can be separated and examined. This was accomplished using a specific oxidizing agent, performic acid, thus separating the insulin molecule into its two protein chains without causing any appreciable secondary damage. Each cystine residue is converted into two cysteic acid residues by this treatment, one remaining in each of the two chains (see fig. 3).

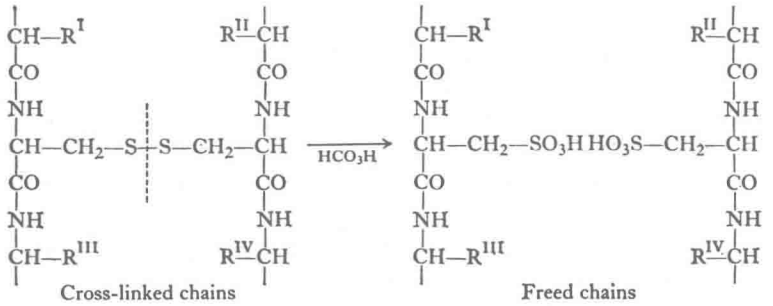


FIG. 3. Cleavage of disulphide crosslinks between protein chains with performic acid (HCO_3H)

These investigators (Sanger *et al.* [1], among others) have spent many years in developing methods for determining the sequence of amino acids in protein chains. These methods involve the fragmentation of the long chains into very small portions containing, in some cases, only two amino acid residues joined together as a dipeptide. This can be achieved using mineral acids under carefully controlled conditions, milder than those which cause complete breakdown to amino acids. Somewhat larger units can be isolated from proteins by the careful use of a variety of proteolytic* enzymes. The peptides obtained are separated, and the sequence of the amino acids contained in them can be elucidated by a variety of techniques. One of these involves the reaction of the terminal amino acid, bearing the free amino group at one end of the peptide chain, with 2:4-dinitro 1-fluorobenzene.

* Protein-splitting.

The reactive fluorine atom displaces a hydrogen atom from the terminal amino group with the formation of the dinitrophenyl derivative of the protein. The dinitrophenyl derivative of this N-terminal amino acid is liberated intact after hydrolysis of the modified protein with acid: these derivatives of amino acids are bright yellow compounds, which facilitates their separation and identification by, for example, paper chromatography [3]. Identification of the N-terminal amino acid of a dipeptide automatically identifies the sequence of the two amino acids it contains, as the overall amino acid composition can be easily determined. The amino acid sequences of many small peptides can be elucidated in this way until fragments of the whole protein chain have been examined, many of which overlap. The information on overlapping peptide sequences was fitted together for insulin, and was found to correspond to a unique unequivocal structure. This amazingly successful piece of work earned a Nobel Prize for F. Sanger, and was in fact the first demonstration that a unique covalent structure existed for any protein. Up to this time many research workers considered that all protein preparations, even when of highest purity and physical homogeneity, consisted of a closely related mixture of molecular species showing a range of similar amino acid sequences. This may be true in some cases, but is certainly not so for the insulin from various species of animals. The outstanding conclusion of biological interest from this work was that the existence of a unique primary structure for a protein must mean that the information for making it is just as specific. Thus a precise genetic code existed which in some way allowed the passage of information from the nuclear genes to the protein-synthesizing site in the cytoplasm of the cell. This code, held in the structure of messenger ribonucleic acid, is discussed in chapter 7. Ox insulin was found to contain two relatively short protein chains, one containing twenty-one amino acid residues and the other thirty, joined together by two disulphide bonds. Subsequent investigations [4] showed that the insulins from pig and sheep pancreas were very similar to ox insulin, except that one trio of amino acids in the shorter of the two protein chains was different in each type (fig. 4). Whether this difference has arisen because of some fine distinction between the metabolism of these species of animals is not clear. Such differences may arise accidentally, where the two structures may be of equal survival value to the species. Contrary to the action of some enzymes (see on), the complete structural integrity of insulin seems necessary to maintain its hormonal action (Young [5]).