chapter 9

ON THE ACCURACY
OF
PROTEIN SYNTHESIS

Except for the relatively minor influence of environ-
mental factors, the phenotypic potentialities of an individual are de-
termined by the genetic information in his chromosomes. In this
book we have taken as a working hypothesis an even more specific
proposition, namely, that the phenotypic picture presented by an or-
ganism is the summation of the effects, physical and catalytic, pro-
duced by the complement of protein molecules characterizing the
species in question. That is to say, we assume that the genetic in-
formation available to an organism is first translated into protein
structure, each gene determining specific details of a corresponding
protein molecule, either alone or in collaboration with other genes.
For example, we may think of the structure and physiological prop-
erties of hemoglobin as being determined by a single hemoglobin
gene or "cistron" (see Chapter 4) or perhaps by several cooperating
genes having to do with each of the peptide chains of hemoglobin
and with the folding and cross attachments between them. The
biosynthesis and arrangement of nonprotein substances such as car-

bohydrates and fats, and the distribution and integration of these secondary products into characteristic cellular systems, might then be considered to be the business of the proteins as agents of the genotype, acting as enzymes, hormones, and structural subunits of morphology.

Since we know that even a single gene mutation, causing a very limited change in the structure of a single protein species (e.g. the sickle-cell hemoglobin case discussed in the previous chapter), can induce a marked variation in phenotypic behavior, it becomes of prime importance in these considerations to examine the accuracy of the mechanisms by which proteins are synthesized and to try to appraise the degree to which errors in these mechanisms may occur. If a change in the structure of a protein involving a single amino acid residue can cause a marked change in function, we may justifiably be rather concerned by the influence of random heterogeneities in structure caused by biosynthetic errors. When we speak of the "control" of protein synthesis by genes, we would like to have some idea of the limits within which this control is exercised.

The highly developed techniques now available for the physicochemical study of proteins have made it possible to detect inhomogeneities in protein preparations caused by differences as small as the presence or absence of a single amide nitrogen group in a fraction of a population of molecules in solution. The apparent inhomogeneity of "pure" proteins may become even more marked as time goes on. In spite of such elegant tools as ion exchange chromatography, countercurrent distribution, and refined electrophoresis machines, we must recognize that the criteria of homogeneity are only relative and that heterogeneities not observable by presently available methods may, tomorrow, be detected on the basis of new, more discriminating procedures.

Since proteins are very complicated organic chemicals, they may naturally exist in a number of isomeric forms. We can conveniently classify such variations under two major headings, using terms suggested by D. Steinberg and E. Mihalyi in their recent review on protein chemistry.[1] Variations in the structure of a protein may be attributed to *sequential* or to *configurational* isomerism. The former classification refers to differences in amino acid sequence between individual molecules and, for convenience, is defined to include other aspects of structure which involve covalent bonds such as disulfide bridges, phosphate ester linkages, and amide nitrogen groups. These are the stable, black-or-white parameters of structure, not modified by the ordinary methods of handling proteins during purification and storage. Configurational isomerism, on the other hand, refers to differences in the mode of coiling of peptide chains or to the location, frequency, and stability of noncovalent bonds. Such isomerism is to be expected in aqueous solutions since the bonds responsible for the secondary and tertiary structure of proteins are strongly affected by the acidity, polarity, and temperature of the environment.

Colvin, Smith and Cook, in their review[2] written in 1954, have presented a list of examples from the literature in which inhomogeneity has been demonstrated in protein preparations, presumably of a high degree of purity. Thus, to quote only a few examples, lysozyme, ribonuclease, and ovalbumin showed reversible boundary spreading upon electrophoretic analysis, human gamma globulin appeared heterogeneous both by electrophoretic and by ultracentrifugal criteria, and insulin could be resolved into two components by countercurrent distribution techniques. These authors have chosen to interpret the experimental results in terms of a "microheterogeneity" in structure and suggest that "it seems more correct to describe a native protein, not in terms of a finite number of definite chemical entities, but as a population of closely related individuals which may differ either discretely or continuously in a number of properties." They have suggested that, should "microheterogeneity" be observed for all native proteins, the cellular mechanisms for the synthesis of proteins need not be specific and rigid, and that there might exist a broad spectrum of individual protein "subspecies" within any single "species," differing in enzymic, hormonal, or physicochemical properties. This conclusion was certainly an understandable one on the basis of the information available in 1954. It is now apparent, however, that many of the examples which supported the concept of "microheterogeneity" could be included in the list quoted only because of the inadequacy of the available knowledge about the chemistry of these proteins. In the case of beef pancreatic ribonuclease, for example, we can separate on proper chromatographic columns two major and two very minor components. A similar family of ribonucleases may be shown to be present in sheep pancreas.[3] These four components appear to be present normally in pancreas tissue since they may be separated both from purified, crystalline starting materials and from crude extracts of pancreas glands. Electrometric titrations on the two major components isolated from beef pancreas have suggested that they differ by a single carboxyl group (or conversely by a single amide nitrogen group). Rather than assume a broad microheterogeneity in the sense of a spectrum of related materials, we might equally well conclude that "ribonuclease" is *not* a

THE MOLECULAR BASIS OF EVOLUTION

ON THE ACCURACY OF PROTEIN SYNTHESIS

statistical population of related proteins but rather a limited group of well-defined chemical entities.

The complete absence of sequential isomerism in any sample of protein can only be established by the *quantitative* recovery of *all* the fragments on which the sequence reconstruction is based. In practice the optimal situation has never been reached, and, for the proteins and polypeptides that have been studied in detail hitherto, sequential purity can only be inferred from the fact that aberrant sequences of amino acid residues, not fitting into the final reconstruction, have not been observed. In the studies of Sanger and his colleagues on the structure of insulin, for example, an enormous array of peptide fragments was examined, and none of these was found to be incompatible with the sequences of the two chains as they finally emerged. Such studies give strong presumptive evidence for the sequential purity of the starting material. In other work, such as that of Shepherd and his collaborators on the structure of ACTH,[4] and in the careful chromatographic separation of the enzymatically produced fragments of ribonuclease by C. H. W. Hirs, S. Moore, W. H. Stein and L. Bailey (Chapter 5), careful balance sheets of recoveries were kept, and for many portions of the over-all sequences recoveries were quantitative within the experimental error of determination. In the case of ACTH none of the fragments was isolated in less than 70 per cent yield, and total recovery, on the average, was 93 per cent of the total starting material. In those fragments recovered in yields less than quantitative it was shown, by paper chromatography, by terminal amino acid analysis, and by the presence of integral molar ratios of constituent amino acids, that sequential purity was extremely likely.

These examples illustrate two of the ways in which we may appraise the homogeneity of a protein; first, by an examination of the internal consistency between the sequences of a large number of small peptide fragments in terms of the final reconstructed sequence of the polypeptide chain representing the common denominator and, second, by a consideration of the completeness of recovery of these fragments. Both methods are as good as the accuracy of methods of detection or analysis available for peptides and give a lower limit for the degree of purity. Except for special sorts of inhomogeneity which we shall discuss more fully later, most proteins or polypeptides that have been examined will probably appear to be at least 90 per cent or more pure by these criteria. The last 5 to 10 per cent (or 3 to 5 per cent when analysis is done by careful ion exchange chromatography) remains an unknown quantity and might conceivably obscure the presence of small amounts of physically similar but structurally different protein material.

Another more worrisome factor in this regard has to do with the extent to which closely related proteins might be removed from the major fraction during isolation procedures. The purification of a protein such as hemoglobin presents no problem of this sort, since the starting material, washed red cells, contains only insignificant amounts of other proteins and the yield of hemoglobin can be made nearly quantitative by careful experimental manipulation. Most enzymes and hormones, however, must be concentrated many hundreds or even thousands of times from their initial *in vivo* state of purity. Modern purification procedures are tremendously discriminating. Under circumstances in which two forms of the same protein, differing only by a carboxylic acid group, may be separated, it is not unlikely that very minor variations in charge or polarity within a family may result in the complete separation of related molecules.

For these reasons we cannot be categorical about the absence of microheterogeneity, even in the sense in which this term was used by Colvin, Smith and Cook. There is, however, no evidence which *requires* that we take the "broad spectrum" point of view, and it would seem unnecessary at the present time to think of protein biosynthesis as an inaccurate or arbitrary process.

A more optimistic alternative may be given as an explanation for the bald fact that a well-defined sequential isomerism *does* occur in certain proteins. Isomerism, as observed for the β-lactoglobulins, hemoglobins, serum haptoglobin, etc., may be thought of as the reflection of heterozygosity in the corresponding genetic material. In those instances in which adequate genetic analyses have been carried out, the occurrence of more than one form of a single protein has been attributable to the presence of sets of allelic genes. In most cases only two forms are observed. We do not find, for example, more than one abnormal hemoglobin in any one individual (except for varying amounts of the fetal form which many believe to be extremely similar or identical with a portion of adult homoglobin). On the other hand, multiple forms of certain other proteins have been observed in some instances. Proteins in the β-globulin fraction of the serum of cattle, for example, may be divided, by the sensitive starch gel electrophoresis method of O. Smithies,[5] into four or five well-separated subcomponents. It is important to emphasize, however, that these are *separable* and that what is obtained is not a smear of overlapping substances but rather a well-defined and reproducible pattern. By some techniques such as gradual salting-out

precipitation, even highly purified human hemoglobin from normal individuals appears to be subdivisible into several components as observed by Roche and his colleagues.[6] Although most workers in the field consider this phenomenon to be due to artifacts introduced by interactions with salts and buffer molecules, it is still quite possible that the effect is a real one, one that is simply not demonstrable by the usual electrophoretic analysis employed for the study of the hemoglobin series.

We have discussed, in a previous chapter, the concept of genetic fine structure which suggests that each "gene" may be composed of a large number of subgenic chemical units, each contributing a small piece of information to the protein biosynthetic process. The occurrence of many abnormal forms of a protein in addition to the normal one cannot be excluded if this concept is generally applicable. Thus, the portion of the genetic material that has to do with a particular protein molecule might involve more than one "cistron." The synthesis of separate chains, or even of parts of the same chain of a protein, might be controlled by different genetic functional units. If this were the case, and if the mutations were not lethal ones and still permitted the synthesis of a functionally adequate protein, we can easily see that a multiplicity of closely related proteins could result through the cooperation of the several cistrons involved. The hemoglobins that have so far been separated and studied have differed in electric charge. Since electrophoresis would not distinguish between two hemoglobins differing, let us say, by the substitution of valine for isoleucine or of serine for threonine, other techniques for fractionation will be required to test this possibility.

The Significance of Amino Acid Analogue Incorporation

A large number of structural analogues of amino acids have been synthesized (Figure 89) and tested for their utilizability in protein biosynthesis. These include the methionine analogues, selenomethionine and ethionine, the phenylalanine analogues, o- and p-fluoro phenylalanine and β-2-thienylalanine, and the tryptophane analogue, 7-azatryptophane. They may be synthesized in radioactive form and thus furnish a powerful and sensitive tool for testing whether the mechanisms of protein biosynthesis are absolutely precise and specific or whether alternative structures may be formed by replacement of natural amino acid residues with man-made substitutes.

The results of such tests are clear-cut. Abnormal amino acids

can be used. Cowie and Cohen,[7] for example, have grown a methionine-requiring mutant of E. coli in a medium completely free of methionine but containing selenomethionine instead. The cells were able to synthesize certain enzymes in a relatively normal manner in spite of the unusual nutritional circumstance, and the proteins of the daughter cells were of the selenomethionine variety. In another similar study, M. Gross and H. Tarver have shown that the proteins of *Tetrahymena pyriformis* can incorporate C^{14}-labeled ethionine. The incorporation represents true peptide bond formation since it was found that ethionine-containing peptides could be isolated from partial acid hydrolysates. An interesting experiment on analogue incorporation has been carried out by D. Steinberg and M. Vaughan, who studied the *in vitro* uptake of tritium-labeled o-fluorophenylalanine (see Figure 89) into the proteins of the minced hen's oviduct.[8] Pure lysozyme was then isolated from the tissue and digested with trypsin and chymotrypsin. The digest was subjected to fingerprinting as described in Chapter 7, and the peptides thus separated were analyzed for the presence of aromatic amino acids. A small proportion of the peptides that normally contained phenylalanine were found to contain the radioactive analogue in place of the natural amino acid.

Although analogues may be incorporated into proteins, the compromise with normalcy does not seem to be a happy one, for most of them also cause marked inhibition of growth. We can say nothing at the moment about the mechanism of this inhibition. Studies are now in progress in several laboratories to determine the efficiency of incorporation of analogues as a function of their degree of dissimilarity from the natural amino acid they mimic.

Nature appears to have been extremely clever in her choice of standard amino acids and has managed to choose some twenty which differ sufficiently to preclude mistakes in recognition. Valine and isoleucine residues are extremely similar from the point of view of three-dimensional structure, and it would not be too surprising to find an occasional lapse in the precision of protein assembly at points of protein structure involving one or the other of these amino acids. To my knowledge, however (within the limits of analytical accuracy mentioned earlier in this chapter), valine-isoleucine interchange has not been observed, except in samples of the same protein or polypeptide isolated from two different species or from an individual who is "heterozygous" for the material in question. On the other hand, D. Cowie and his colleagues have recently shown that a considerable amount of the methionine in E. coli proteins may be replaced by
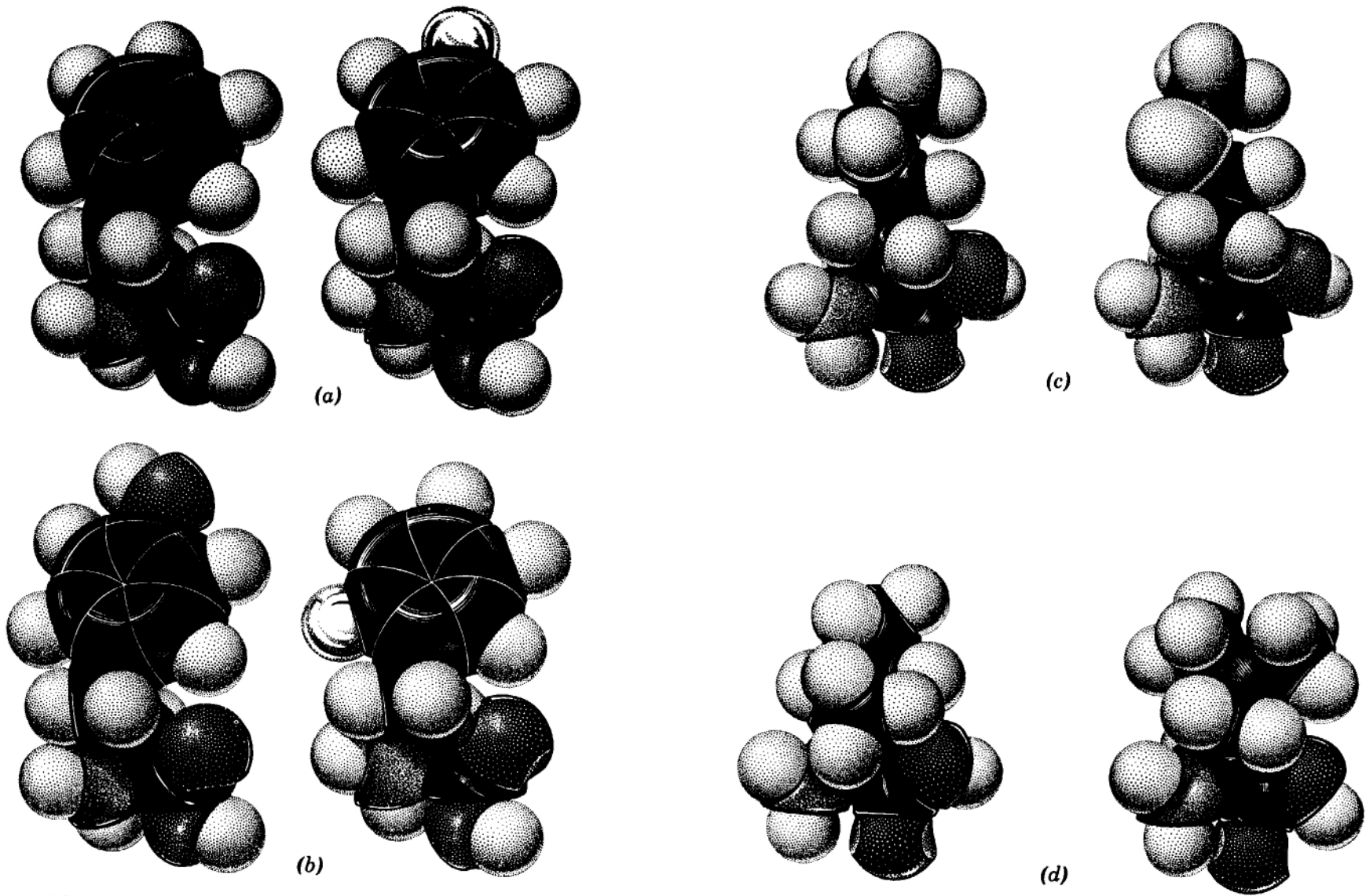
Figure 89. The molecular structure of some amino acids and amino acid analogues. (a) Phenylalanine and p-fluorophenylalanine, (b) tyrosine and o-fluorophenylalanine, (c) norleucine and methionine, (d) isoleucine and leucine. Although norleucine is a chemical isomer of leucine and isoleucine, its molecular shape is much more similar to that of the sulfur-containing amino acid methionine than to that of isoleucine or leucine.

norleucine,[9] an amino acid which is not normally found in proteins but which exhibits a remarkable similarity in molecular appearance (Figure 89) to methionine.

Protein synthesis is not an absolutely precise process. Some amino acid analogues *can* substitute for natural amino acids. Nevertheless, the weight of evidence indicates that no mistakes are detectable under normal circumstances and that the protein assemble mechanism must have built into it an extraordinary capacity for structural discrimination.

REFERENCES

1. D. Steinberg and E. Mihalyi, *Ann. Rev. Biochem.*, **26**, 373 (1957).
2. J. R. Colvin, D. B. Smith, and W. H. Cook, *Chem. Revs.*, **54**, 687 (1954).
3. S. Åqvist and C. B. Anfinsen, *J. Biol. Chem.*, **234**, No. 5 (1959); C. B. Anfinsen, S. Åqvist, J. Cooke, and B. Jönsson, *J. Biol. Chem.*, **234**, No. 5 (1959).
4. R. G. Shepherd, S. D. Wilson, K. S. Howard, P. H. Bell, D. S. Davies, S. B. Davis, E. A. Eigner, and N. E. Shakespeare, *J. Am. Chem. Soc.*, **78**, 5067 (1956).
5. O. Smithies, *Biochem. J.*, **61**, 629 (1955).
6. J. Roche, Y. Derrien, and M. Roques, *Bull. soc. chim. biol.*, **35**, 933 (1953).
7. D. B. Cowie and G. N. Cohen, *Biochim. et Biophys. Acta*, **26**, 252 (1957).
8. I am grateful to Dr. Daniel Steinberg and Dr. Martha Vaughan of the National Heart Institute, Bethesda, Md., for information on these studies prior to publication.
9. Personal communication from Dr. D. B. Cowie of the Carnegie Institution, Department of Terrestrial Magnetism, Washington, D. C.