Reprinted for private circulation from
Raymond E. Zirkle, ed.: A SYMPOSIUM ON MOLECULAR BIOLOGY
University of Chicago Press

CHAPTER 5

Protein Structure and Biological Activity

Christian B. Anfinsen

WE HAVE only recently reached that stage of methodological and conceptual sophistication where a consideration of the relationships between the structure of protein molecules and their biological activity has become possible. Nevertheless, the literature has already been adequately saturated with discussions of the subject, and, for that reason, the following remarks have been restricted to the presentation of a thin background of experimental data, together with a few conclusions and questions which might not be completely obvious to those who are not directly engaged in protein research or in allied fields.

The fundamental work of Sanger and his collaborators (1, 2, 3), followed by similar studies by a number of other groups of investigators (e.g., 4–12), has now given us a body of information on the covalent structure of proteins which permits us to proceed with the business of the systematic degradation of protein structure in parallel with studies of the effects of such modification on the functional capacity of these molecules. A number of investigations have demonstrated that biologically active proteins may be considerably modified with only minimal loss of function. Two examples of such work have been chosen here for more detailed chemical consideration.

The structure of the adrenocorticotrophic hormone (ACTH), prepared from the pituitary glands of two different species, has been completely elucidated in several laboratories (7, 8, 9), and recently the structures of allied melanocyte-stimulating hormones (MSH) have also been worked out (10, 11). The various structures derived from pig pituitary glands are listed in Figure 1. They are arranged in such a way as to make clear the overlapping portions of the amino acid sequences. The figure also includes an indication of some of the sites at which these polypeptides are cleaved by the various proteolytic enzymes that were used during the course of structural study. Although the full story is not

yet available, it appears that at least the eleven C-terminally located residues of ACTH may be removed by digestion with pepsin without loss of hormonal activity (7). We may thus conclude that no more, and perhaps less, than the first twenty-eight residues are sufficient. On the other hand, removal of one or two residues from the N-terminal end of the molecule by leucine aminopeptidase digestion leads to complete inactivation. Recent studies by White and Gross (13) have shown that bovine fibrinolysin, which splits corticotropin following residues 8 and 15 also causes complete loss of ACTH activity. Measurements of MSH activity (which one might expect to disappear during such hydrolysis) were, unfortunately, not reported. The melanocyte-stimulating activity of ACTH is, however, not lost following the more limited modification

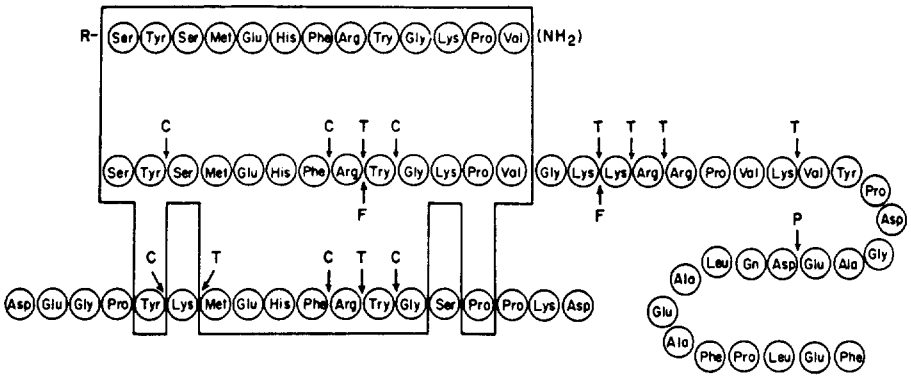

FIG. 1.—The structures of some porcine pituitary hormones. *Upper formula*, α-MSH (10); *middle*, ACTH; *lower*, β-MSH (10, 11); *C*, chymotrypsin; *T*, trypsin; *F*, fibrinolysin; *P*, pepsin. N-terminal groups at left.

at the N-terminal end of the chain brought about by leucine aminopeptidase (9).

Two different polypeptides having melanocyte-stimulating properties but no ACTH activity have been isolated from pig pituitary glands. The first, termed "β-MSH," contains eighteen amino acid residues and includes a sequence of seven residues which is found also in ACTH (10, 11). The second, α-MSH, contains only thirteen residues, and the N-terminal and C-terminal ends are apparently masked by an acyl group and an amide group, respectively (10). The sequence of this polypeptide is completely analogous to the N-terminal portion of porcine ACTH except for the modifications at the ends of the chain just mentioned.

Although it is perhaps superfluous to discuss here the activity interrelations summarized in Figure 1, since the matter has been considered in detail by Li *et al.* (11) and by Harris and Roos (14) and Harris and Lerner (10), the implications of these findings in connection with the mechanism of polypeptide

biosynthesis are worthy of note. The heptapeptide sequence which occurs in all three of the molecules shown is suggestive of a common intermediate in the synthesis of all three. It might be argued that the same "template" is involved in the formation of β-ACTH and α-MSH, since the structure of the latter is completely contained in the former, excepting the terminal modifications, although, of course, the existence of different "templates" having similar information is equally probable. In the case of β-MSH, the former proposition becomes less tenable because the heptapeptide sequence is surrounded by residues not present in the other two cases. The recent studies of Ramachandran and Winnick (14) on pituitary extracts, indicating a very heterogeneous and large pool of assorted peptides, would be compatible with a situation in which peptide fragments are first synthesized *de novo* or produced through controlled degradation of some preformed protein material and then subsequently assembled in a specific way to yield the various biologically active polypeptide materials which characterized the anterior lobe and the pars intermedia of the pituitary gland. It may be argued that these polypeptides are small and that their synthesis may not parallel that of larger protein molecules. Nevertheless, the structural data concerned here are of considerable interest in connection with the hypothesis that protein biosynthesis involves discrete, kinetically distinguishable (15) intermediates whose specific alignment and conjugation is brought about by genetically controlled assembly mechanisms.

A second example which may be considered in a preliminary sort of way is the enzyme, bovine pancreatic ribonuclease (4–6). The structural elucidation of this protein has not yet been completed, but sufficient is already known to permit us to make a few remarks about its unique characteristics. A schematic summary of our present picture of the chemical formula of ribonuclease is shown in Figure 2. It must be emphasized that this scheme has all the deficiencies of a two-dimensional projection and that the spatial relationships between the various parts of the molecule, as it exists in solution, must be arrived at through a careful analysis of the physical studies which are now only in their beginning stage. Ribonuclease is a single chain of 124 amino acid residues, cross-linked through 4 disulphide bridges and an undetermined number of non-covalent bonds of varying strengths. It has a fairly long N-terminal "tail," which, according to the results of Richards (16), may be almost completely removed by a limited exposure to the proteolytic enzyme, subtilisin, without loss of catalytic activity in the macromolecular portion. It also appears that the C-terminal valine residue, and possibly the preceding serine and alanine residues as well, may be removed by carboxypeptidase action without impairment of activity (17).

The converse of these findings is obtained when one exposes ribonuclease to very limited digestion with pepsin (18). From such a digestion mixture one may isolate a derivative of the native molecule which lacks only the C-terminal

tetrapeptide, Asp-Ala-Ser-Val, as evidenced by the presence in this derivative of C-terminal phenylalanine (19) and by the absence of indications of other cleavages in the polypeptide chain. This derivative is completely *inactive*. Concomitant with the covalent change, there occurs a marked change in the ultraviolet absorption of the tyrosine moieties of the protein, of a sort consistent with a process of hydrogen-bond rupture in some tertiary structural feature of the molecule involving the hydroxyl groups of some of the tyrosine rings and electronegative groups elsewhere in the molecule (19). The simultaneity of the changes in activity and spectral properties and of the rate of appearance of

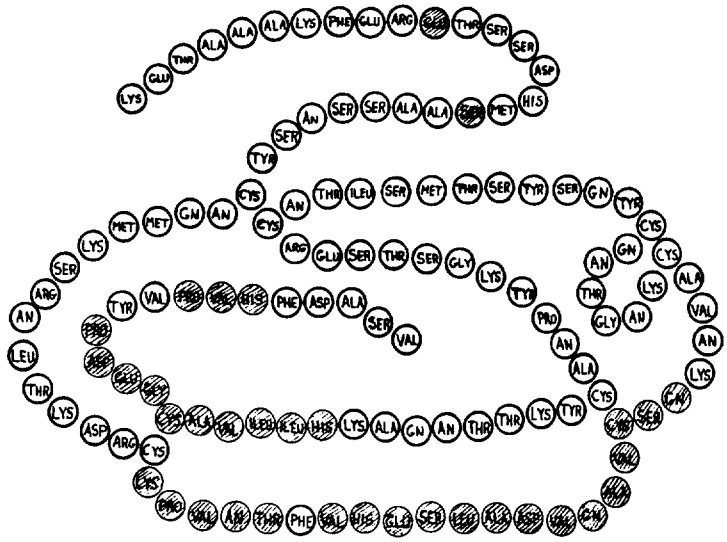

FIG. 2.—Schematic diagram of the structure of bovine pancreatic ribonuclease. *An*, asparagine; *Gn*, glutamine. The sequences of the residues that are crosshatched are not definitely determined at the present time. This figure has been constructed from the combined studies of a number of individuals at the Rockefeller Institute (including Drs. C. H. W. Hirs, S. Moore, W. Stein, D. Spackman, and L. Bailey) and at the National Heart Institute (including R. Redfield, J. Cooke, A. Ryle, and C. B. Anfinsen).

the free tetrapeptide is shown in Figures 3 and 4. This change, induced by pepsin, may be correlated with a more or less direct attack on the "active center" of the molecule and suggests that this catalytically active constellation of amino acid residues and three-dimensional structural parameters is associated with the C-terminal portion of the protein. A similar conclusion may be reached from the recent degradative studies of Rogers and Kalnitsky (20).

It must be kept in mind, however, that cross-linking through disulfide bridges may bring into close spatial relation two areas of the molecule quite widely separated in terms of amino acid sequence. The locations of the disulfide bridges are indicated in Figure 2 (21, 22). The half-cystine residue located at position 40 in the chain is bonded to the half-cystine residue at residue 113, that at 58 to 96, and so forth. Thus C-terminally located residues are

drawn into close proximity to portions of the sequence well separated from this end of the chain in a linear sense (model building, although clearly of enormous difficulty in the case of a polypeptide of this size, appears to be an unfortunate necessity for obtaining a better understanding of the three-dimensional aspects of function in this enzyme).

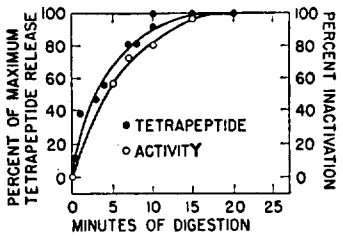

FIG. 3.—Inactivation of ribonuclease and appearance of the tetrapeptide, Asp-Ala-Ser-Val, during pepsin digestion at pH 1.8.

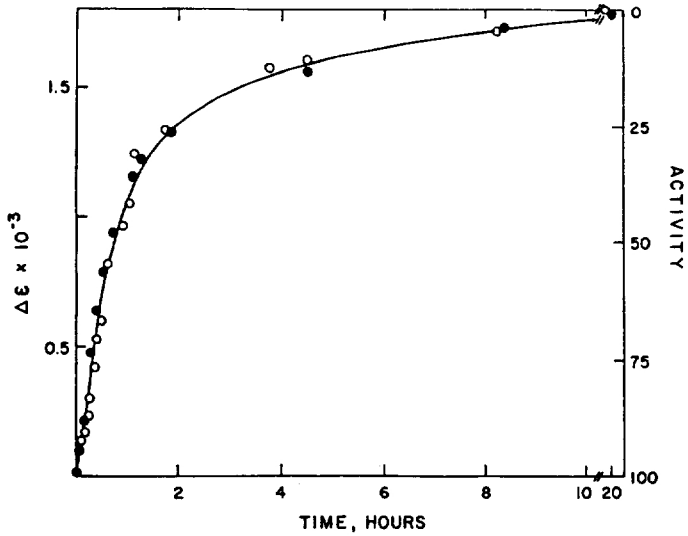

FIG. 4.—Relation between spectral and activity changes during pepsin digestion of ribonuclease. *Left ordinate*, change in extinction at 285 mμ (*open circles*); *right ordinate*, per cent of native activity (*solid circles*).

Recent studies on the stepwise reductive cleavage of the SS bridges of ribonuclease using thioglycollate (23) have been encouraging in connection with the likelihood of obtaining a really small fragment of the enzyme that still possesses enzymatic activity. The experiments, summarized in Figure 5, indicate that certainly one, and perhaps two, of the SS bridges may be so ruptured with only partial inactivation. Furthermore, it can be shown that the activity

which remains is not due to remnants of unattacked native enzyme, since electrophoretic analysis indicates the complete absence of the unchanged protein. Although based on very preliminary data, the conclusion can be drawn that one bridge which is ***not*** ruptured by the limited reduction process and which, therefore, is implicated in the active core of the molecule is the disulfide bond joining half-cystines 3 and 7 (numbering from the N-terminal end of the chain), which are located at residues 58 and 96. This bridge appears to be intact in the active products of partial reduction.

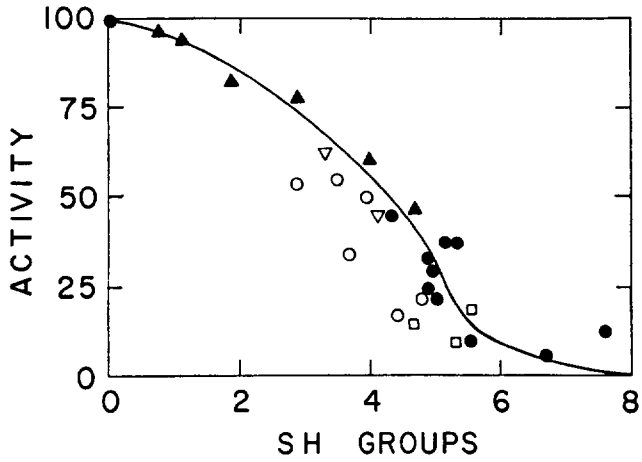

FIG. 5.—Changes in ribonuclease activity during reductive cleavage of disulfide bridges; *solid triangles*, reduction with thioglycollate in absence of urea; *solid circles*, in presence of 8 *M* urea. Open points represent results of reoxidation experiments; open squares, oxidation with gaseous oxygen of fully reduced (8 SH groups/mole), inactive ribonuclease; other open points, oxidation of partially active intermediates. (Recent experiments indicate that the reduction of disulfide bridges in ribonuclease in the absence of urea does not proceed beyond two SH groups per mole and that the higher levels of reduction indicated by three of the solid triangles in the figure may be due to side reactions. The shape of the curve is not changed, however.) For details see reference 23.

These studies on the pituitary gland hormones and on ribonuclease have been chosen as examples from a number of investigations of "permissible modifications" (pepsin, papain, TMV, lysozyme, etc.). They illustrate some generalities that may well be quite applicable to many biologically active proteins and that make it necessary for the protein chemist to expand his horizons and his research plans considerably. A few of the questions raised are the following:

1. If, through the millions of years of mutation and selection, nature has chosen and preserved certain unique molecular structures which appear to our naïve mind's-eye to contain superfluous parts, are we not led to believe that a number of the structural features in proteins have to do with yet undiscovered aspects of cellular engineering and metabolism?

2. Is it, therefore, not likely that, in analogy to the process of natural selec-

tion at the morphological level, there exists a "natural selection of molecules," with a number of different structural aspects within the same molecule each to be considered in the evaluation of a given molecular species in terms of its relative efficiency and suitability in the total cellular economy?

3. If relatively small portions of catalytically active polypetides are sufficient for the job of catalysis, may one properly speculate about the possibilities of pre-Cambrian "organisms," so simple and so well supplied with environmental food supplies that the advantages of macromolecullariness were not yet an overriding consideration?

4. May we begin now to think of certain mutants, such as the temperature-dependent ones described by Hartmann (24) for some of the enzymes involved in histidine biosynthesis in *Salmonella*, as containing modified genetic loci which control the biosynthesis of parts of proteins intermediate in functional importance but perhaps not absolutely obligatory as structural components? The relatively low frequency of occurrence of "temperature mutants" in a mixed mutant population may be a hint in this direction (see, for example, articles by Benzer [25]).

Questions such as these, as well as the whole problem of the significance of species variations in protein structure and the genetic basis for such variations, will clearly occupy investigators in many allied fields for a great number of years. It does not seem too overoptimistic, however, to feel that the present degree of interest in the molecular basis of biological function will increase even further and that partial answers to at least some of these problems may be obtained through experimental study.

BIBLIOGRAPHY

1. SANGER, F., and THOMPSON, E. O. P. Biochem. J., **53**:353, 356, 1953.
2. SANGER, F., and TUPPY, H. Biochem. J., **49**:481, 1951.
3. RYLE, A. P., SANGER, F., SMITH, L. F., and KITAI, R. Biochem. J., **60**:541, 1955.
4. ANFINSEN, C. B., and REDFIELD, R. Advances Protein Chem., Vol. **11**. New York: Academic Press, Inc., 1956.
5. HIRS, C. H. W., STEIN, W. H., and MOORE, S. J. Biol. Chem., **221**:151, 1956.
6. HIRS, C. H. W., MOORE, S., and STEIN, W. H. J. Biol. Chem., **219**:623, 1956.
7. LI, C. H., GESCHWIND, I., COLE, R. D., RAACKE, I., HARRIS, J., and DIXON, J. Nature, **176**:687, 1955.
8. HOWARD, K. S., SHEPHERD, R. G., EIGNER, E. A., DAVIES, D. S., and BELL, P. H. J. Am. Soc., **77**:3419, 1955.
9. WHITE, W. F., and LANDMANN, W. A. J. Am. Chem. Soc., **77**:1711, 1955.
10. HARRIS, J. I., and ROOS, P. Nature, **178**:90, 1956.
    HARRIS, J. I., and LERNER, A. B. Nature, **179**:1346, 1957.
11. GESCHWIND, I., and LI, C. H. J. Am. Chem. Soc., **79**:615, 1957.
12. HILL, R. L., and SMITH, E. L. Biochim. et biophys. acta, **19**:376, 1956.
13. WHITE, W. F., and GROSS, A. M. J. Am. Chem. Soc., **79**:1141, 1957.
14. RAMACHANDRAN, L. K., and WINNICK, T. Biochim. et biophys. acta. **23**:533, 1957.
15. STEINBERG, D., VAUGHAN, M., and ANFINSEN, C. B. Science, **124**:389, 1956.

16. RICHARDS, F. M. Compt. rend. trav. lab. Carlsberger, Sér. chim., **29**:329, 1955.
17. ANFINSEN, C. B., and RICHARDS, F. M. Unpublished data.
KALNITSKY, G., and ROGERS, W. I. Biochim. et biophys. acta, **20**:378, 1956.
18. ANFINSEN, C. B. J. Biol. Chem., **221**:405, 1956.
19. SELA, M., and ANFINSEN, C. B. Biochim. et biophys. acta, **24**:229, 1957.
20. ROGERS, W. I., and KALNITSKY, G. Biochim. et biophys. acta, **23**:525, 1957.
21. RYLE, A., and ANFINSEN, C. B. Biochim. et biophys. acta, **24**:633, 1957.
22. SPACKMAN, D. H., MOORE, S., and STEIN, W. H. Fed. Proc., **16**:196, 1957.
23. SELA, M., WHITE, F., and ANFINSEN, C. B. Science, **125**:691, 1957.
24. HARTMANN, P. Pub. Carnegie Inst. Washington No. 612, 1956.
25. BENZER, S. In: W. D. MCELROY and B. GLASS (eds.), The chemical basis of heredity, p. 70. Baltimore: Johns Hopkins University Press, 1957.