# THE STRUCTURE OF PROTEINS IN RELATION TO BIOLOGICAL PROBLEMS[1]

MAX BERGMANN

*Laboratories of The Rockefeller Institute for Medical Research, New York, New York*

*Received February 12, 1938*

There are in protein chemistry two fundamental theories which have been accepted for many years. One is the well-known peptide theory of Emil Fischer (17) and of Hofmeister (19). According to this theory the molecule of a protein is supposed to be constructed of a large number of amino acids which are combined one by one through peptide bonds, thus forming a long chain containing many CO—NH groups or peptide bonds. We know about twenty-five amino acids to be constituents of proteins. The various proteins are supposed to differ with regard to the length of the peptide chain and the relative content and sequence of the individual amino acid residues. Fischer has emphasized that there is an almost infinite number of possible sequences of amino acid residues. For example, it was calculated that thirty amino acid residues, among them eighteen of different nature, can give rise to the existence of $1.28 \times 10^{27}$ different proteins (18).

The second general theory concerns the enzymic digestion of proteins. Proteins have been supposed to be degraded, step by step, into peptones, polypeptides, dipeptides, and eventually amino acids. It is assumed that each step is performed by a special enzyme or group of enzymes (22, 16). The first step, which consists in the degradation of the original high molecular proteins, is attributed to enzymes called proteinases, such as pepsin or trypsin. None of these proteinases was found capable of digesting simple peptides,—either natural or synthetic peptides. On the other hand, it was observed that the enzymes that attack simple peptides could not digest high molecular proteins. There was considerable discussion (20, 21) about the question as to how to reconcile the peptide theory with the fact that proteins are digested by enzymes which do not attack the peptide bonds of all the known low molecular polypeptides. Until very recently it has been assumed that proteinases are restricted to high molecular

[1] The text of this paper, which is based principally on results obtained in this laboratory, is presented as delivered, by invitation, at the Seventh National Organic Chemistry Symposium of the American Chemical Society, December 29, 1937, at Richmond, Virginia.

substrates. On the other hand, it has repeatedly been suggested that proteins may contain large numbers of linkages that differ from peptide bonds and may be the points at which pepsin, trypsin, and other proteinases attack the protein molecule. None of these hypotheses has found sufficient experimental support.

And, indeed, no experimental support could have been expected. If proteinases are in reality capable of attacking only high molecular proteins or certain special linkages found in high molecular proteins, it is obviously futile to expect to obtain any experimental evidence on the specificity of proteinases by the use of simple synthetic peptides. However, if we reëxamine our knowledge regarding proteins, there arises some doubt whether the original premise of our argument is correct. The disparity in molecular weight is not the sole difference between proteins and peptides. Another significant difference lies in the fact that proteins contain practically no free α-amino or free α-carboxyl groups, while peptides do contain such groups. The presence of these ionizable groups should influence decisively the polar character of the peptide molecule and of its peptide bonds, thus affecting the enzymic digestibility. Consequently we synthesized peptide-like substances of low molecular weight, which had neither free α-carboxyl nor α-amino groups and conformed to the general scheme

$$R'\cdot CO—NH\cdot CHR''\cdot CO—NHR'''$$

To our satisfaction, certain of these protein models were found to be easily digested by crystalline trypsin (8), others by crystalline chymotrypsin (6) and heterotrypsin (6, 12), and still others by the intracellular proteinases papain (13, 14, 15), cathepsin, and bromelin (7). For each of the representative proteinases, with the exception of pepsin, we were thus able to synthesize numerous substrates; a few of them are reported in table 1.

Since the structure of these substrates may be varied in many ways, there is no longer any obstacle to a detailed investigation of the specificity of each of these enzymes. It was found that each proteinase has its individual specificity and therefore its individual substrates. These synthetic substrates are the long-needed tools for a characterization of the specific nature and the exact quantity of proteinases contained in crude biological preparations and for an investigation of the homogeneity and the kinetics of purified enzymes. As an example of the usefulness of the synthetic substrates I should like to mention the discovery that commercial pancreatin contains large quantities of a hitherto unknown tryptic enzyme which has been designated heterotrypsin.

As another result of the study of synthetic protein models it must be mentioned that the digestive action of trypsin, pepsin, and other proteinases does not stop with the formation of peptones but may, in part, proceed

to the formation of simple peptides or even amino acids. How far the digestion of a protein proceeds is a question not of the molecular size but of the structure of the intermediate products. The presence or absence of α-amino or α-carboxyl groups and the nature and sequence of side chains direct the course and limit the extent of the digestion. Very remarkable indeed is the highly differentiated specificity of our various digestive proteinases as illustrated by trypsin, chymotrypsin, and heterotrypsin. It is generally assumed that the task of digestion is a complete breakdown of food proteins into amino acids. It appears, however, as though the digestive tract takes special precautions in order to conduct the digestion along certain structural lines.

Table 1 also reports the fact that all the proteinases are specifically adapted to the splitting of peptide bonds. Therefore there is no doubt that the digestive action of these enzymes on proteins is performed at the peptide bonds. We no longer have to fear that proteins contain large

TABLE 1

*Substrates of proteinases*

| ENZYME | SUBSTRATE |
|---|---|
| All enzymes | R′·CO—NH·CHR″·CO—NHR‴ |
| Trypsin | Benzoyl-*l*-arginine amide |
| Chymotrypsin | Benzoyl-*l*-tyrosylglycine amide |
| Heterotrypsin | Benzoylglycyl-*l*-lysine amide |
| Papain | Benzoylglycine amide |
| Cathepsin | Carbobenzoxy-*l*-leucylglycylglycine |
| Bromelin | Carbobenzoxyglycyl-*l*-glutamylglycine amide |

numbers of linkages of unknown nature, but can without reservation consider the peptide bonds to form the essential links connecting the amino acid residues inside a protein molecule.

Through our experiments with artificial substrates we have killed two birds with one stone: we have arrived at a new concept of enzymatic proteolysis and, simultaneously, we have strengthened the general validity of the peptide theory—or, at least, of one claim of the peptide theory. We can now be certain that proteins have the structure of peptide chains. However, within the peptide scheme of proteins there are still a number of variables, such as (*1*) the length of the peptide chain expressed in the total number of amino acid residues contained in one molecule of a protein, and (*2*) the ratio and the sequence of the individual amino acid residues contained in that protein molecule. By the variation of the total number, the ratio, and the sequence of the amino acid residues, there is possible, theoretically, an infinite number of proteins. Actually, it was found that only a

limited number of these variations are realized in nature (9, 10, 11). This important fact is demonstrated in tables 2 and 3. In table 2 there is reported the analysis of several constituents of cattle blood globin. In table 3 there are found the numbers of amino acid residues per molecule in four proteins of rather different physiological significance, such as the albumin of the chicken egg, the globin and the fibrin of cattle blood, and

TABLE 2

*The number of amino acid residues per molecule of cattle blood globin*

| 1<br>AMINO ACID RESIDUE | (2)<br>WEIGHT | (3)<br>RESIDUE WEIGHT | (4)<br>GRAM-EQUIVALENTS PER 100 GRAMS OF GLOBIN | (5)<br>RECIPROCAL FRACTIONAL VALUE | (6)<br>RATIO (NUMBER OF INDIVIDUAL RESIDUES PER MOLECULE) |
|---|---|---|---|---|---|
| | *per cent* | | | | |
| Lysine | 7.01 | 128 | 0.0546 | 16 | 36 |
| Histidine | 6.54 | 137 | 0.0478 | 18 | 32 |
| Aspartic acid | 5.53 | 115 | 0.0479 | 18 | 32 |
| Glutamic acid | 3.07 | 129 | 0.0239 | 36 | 16 |
| Tyrosine | 2.97 | 163 | 0.0182 | 48 | 12 |
| Proline | 1.77 | 97 | 0.0182 | 48 | 12 |
| Arginine | 2.78 | 156 | 0.0181 | 48 | 12 |
| Cysteine | 0.47 | 103 | 0.0046 | 192* | 3* |
| Average | | 115.5 | 0.865 | 1 | 576 |

* Calculated as cysteine.

The percentage found for an individual amino acid residue (column 2), divided by the weight of this residue (column 3), gives the number of gram-equivalents of this residue found per 100 grams of globin (column 4).

The average weight of all amino acid residues in globin is 115.5. Therefore, 100 grams of globin contains 100/115.5 = 0.865 gram-equivalent of an average amino acid residue.

Lysine comprises 0.0546/0.865 = 1/16 of all the constituent residues of globin; histidine comprises 0.0478/0.865 = 1/18; etc. (column 5).

The ratios in column 6 are obtained directly from column 4.

By multiplying a value of column 6 by the corresponding value in column 5, one obtains the total number of amino acid residues contained in one molecule of globin. Globin thus contains 36 × 16 = 576 residues of an average weight of 115.5. Globin therefore has a molecular weight of 576 × 115.5 = 66,520. This is a minimum value. The correct value may be a whole number multiple thereof.

the fibroin of the silkworm fiber. Time does not permit a description of the newer and relatively simple methods which enable us to perform the estimation of quite a number of amino acids with a high degree of precision (11, 3, 1, 2).

When we analyze a protein as, for example, cattle hemoglobin, our analytical methods enable us to find that one molecule of this protein

contains thirty-six lysine units, and that these units are one-sixteenth of the total number of units contained in one molecule of cattle hemoglobin. When the number of individual units ($N_i$) and the reciprocal fractional value ($F_i$) are multiplied by each other, we find the total number of amino acid residues in one protein molecule, as shown in equation 1 of table 4. The values reported in table 3 for the amino acid content of four proteins

TABLE 3

*The number of amino acid residues (units) per molecule of various proteins*

| AMINO ACID | CATTLE GLOBIN | CATTLE FIBRIN | CHICKEN EGG ALBUMIN | SILK FIBROIN |
|---|---|---|---|---|
| All amino acids | $2^6 \times 3^2$ | $2^6 \times 3^2$ | $2^5 \times 3^2$ | $2^5 \times 3^4$ |
| Arginine | $2^2 \times 3^1$ | $2^5 \times 3^0$ | $2^2 \times 3^1$ | $2^2 \times 3^1$ |
| Lysine | $2^2 \times 3^2$ | $2^4 \times 3^1$ | $2^2 \times 3^1$ | $2^2 \times 3^0$ |
| Histidine | $2^5 \times 3^0$ | $2^2 \times 3^1$ | $2^2 \times 3^0$ | $2^0 \times 3^0$ |
| Aspartic acid | $2^5 \times 3^0$ | $2^5 \times 3^0$ | $2^4 \times 3^0$ | |
| Glutamic acid | $2^4 \times 3^0$ | $2^3 \times 3^2$ | $2^2 \times 3^2$ | |
| Glycine | | | | $2^4 \times 3^4$ |
| Alanine | | | | $2^3 \times 3^4$ |
| Tyrosine | $2^2 \times 3^1$ | | $2^3 \times 3^0$ | $2^1 \times 3^4$ |
| Proline | $2^2 \times 3^1$ | $2^5 \times 3^0$ | | |
| Tryptophane | | $2^1 \times 3^2$ | | |
| Cysteine | $2^0 \times 3^1$ | $2^0 \times 3^2$ | $2^2 \times 3^0$ | |
| Methionine | | $2^2 \times 3^1$ | $2^2 \times 3^1$ | |
| Molecular weight | 66,520 | 69,300 | 35,700 | 217,700 |

TABLE 4

*General formulas regarding the number of units per molecule of the proteins of table 3*

(1) $N_i \times F_i = N_t$

(2) $N_t = 2^m \times 3^n$, where $m$ and $n$ are positive whole numbers

(3) $N_i = 2^{m'} \times 3^{n'}$ where $m'$, $m''$, $n'$, and $n''$ are either zero or positive whole numbers

(4) $F_i = 2^{m''} \times 3^{n''}$

(5) $m = m' + m''$

(6) $n = n' + n''$

reveal the surprising fact that $N_i$ and $F_i$, and therefore $N_t$, can be expressed by powers of 2 and 3, as shown in the general equations 2, 3, and 4 of table 4. Everyone who is familiar with the history of protein chemistry may feel somewhat amazed on being confronted with a simple stoichiometry of the protein molecule and with numerical rules such as are reported in table 4. Perhaps it would not seem to be superfluous to point out that these rules are not mere hypothetical conceptions but actual experimental results.

The experiments on which these results are based cover, at the present time, only a restricted number of proteins. However, the numerical rules observed are of so pronounced a uniformity that it seems natural to generalize their validity beyond the four proteins which have been discussed. There already exist indications that similar numerical rules hold good for many other natural proteins. As such proteins there may be mentioned collagen (analyzed by Niemann and Stein in this laboratory), elastin (analyzed by Miller and Stein in Dr. Hans T. Clarke's laboratory), and insulin (du Vigneaud). Furthermore, the finding of Svedberg and his collaborators that the particle sizes of many proteins exhibit approximate numerical regularities has been interpreted as pointing to a common plan for the building up of the protein molecules (23).

This common plan has its basis in our equation 2 (table 4). Equation 2 expresses the fact that the molecules of various proteins fall into classes containing a definite number of amino acid units: for example, chicken egg albumin falls into the class with 288 units, cattle blood globin and fibrin into the class with 576 units, and silk fibroin into the class with 2592 units.

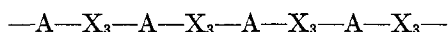
If a protein consists exclusively or almost exclusively of amino acid units, then an elementary calculation permits the transformation of the total number of units into the molecular weight of the protein. The chemical analysis reveals not only the total number of units but also their average weight. By multiplying the total number of units by their average weight we get a fairly accurate value of the molecular weight. As in the case of simpler compounds, a molecular weight obtained by chemical analysis is a minimum value and the true value may be a whole number multiple thereof. The molecular weights obtained for the four proteins considered in table 3 are, respectively: 35,700; 66,520; 69,300; and 217,700. Here we find, indeed, a regularity similar to that observed by Svedberg by means of the particle size method. It is apparent, however, that the regularity of the molecular weights is, and can only be, an approximate one. It is quite instructive to compare fibroin with egg albumin. The average residue weight of fibroin is unusually low, since it consists for the most part of small amino acid units; therefore its molecular weight is only six times that of egg albumin, although fibroin has nine times the number of units of egg albumin.

In order to understand the meaning of a molecular weight of a protein just as in the case of a simpler compound, we have to know the kind and the number of its constituents and the structural pattern in which they are arranged. In the case of many natural proteins the number of units is expressed by the numerical rules stated in equations 2 to 4. These rules must have a basis in the structural pattern of the protein,—that is, in the sequence of the amino acid units. These units cannot be distributed at

random. Their sequence must rather be such that it includes by implication the numerical rules found experimentally. There is only one structural principle that fulfills this requirement: it may be illustrated by the example of silk fibroin, since of all the well-known high molecular proteins fibroin has the simplest pattern. Exactly one-half of the constituents of fibroin are glycine residues. This fact may be expressed in the conclusion that every second amino acid residue in the peptide chain of fibroin must be a glycine residue (G):

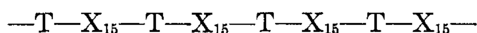
$$—G—X—G—X—G—X—G—X—$$

Every fourth residue in the chain is an alanine residue (A):



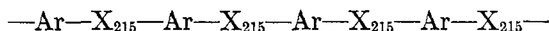
$$—A—X_3—A—X_3—A—X_3—A—X_3—$$

Every sixteenth residue is a tyrosine residue (T):



$$—T—X_{15}—T—X_{15}—T—X_{15}—T—X_{15}—$$

And every two hundred and sixteenth is an arginine residue (Ar):



$$—Ar—X_{215}—Ar—X_{215}—Ar—X_{215}—Ar—X_{215}—$$

When the above four schemes are combined into one, we obtain the following structure of a segment of the silk fibroin molecule representing 432 amino acid residues or one-sixth of the whole molecule: —G—A—G—T—G—A—G—Ar—G—A—G—X—G—A—G—X—[G—A—G—T—G—A—G—X—G—A—G—X—G—A—G—X$]_{12}$—G—A—G—T—G—A—G—X—G—A—G—X—G—A—G—Ar—[G—A—G—T—G—A—G—X—G—A—G—X—G—A—G—X$]_{13}$—. As a general rule, the amino acid residues of a protein molecule may be arranged in such a way that each individual residue repeats itself throughout the protein molecule at constant intervals, i.e., with a regularly recurring frequency. The frequencies are different, in general, with respect to various kinds of amino acid residues of the same molecule. The protein molecule thus contains a number of different, superimposed frequencies. This principle of the superimposed frequencies confers a relatively simple structural pattern upon the giant protein molecule. Living organisms therefore do not achieve the synthesis of the immense number of proteins provided by the peptide theory in its original conception, but seem to synthesize only those proteins that exhibit these simple numerical rules and the pattern of the superimposed frequencies. How is such a limitation in the number and types of naturally occurring proteins to be explained? It seems to me to have its origin in the mechanism of the biological synthesis of protein molecules. As long as we expected to meet in nature every kind of protein provided by the peptide theory in its general, unrestricted form, we had no

indication of the means by which the variety of natural proteins is produced. The discovery of the quantitative rules governing the protein molecule makes the biological synthesis of an individual protein molecule appear as a process which involves a specificity that is both highly delicate and extremely complex. Such specificity phenomena are a clear indication of the operation of an enzymatic process which directs all the steps involved in the synthesis of the special pattern of each individual protein. We have approached this problem experimentally in the belief that in living cells proteinases should direct the synthesis of proteins and should exhibit some kind of specificity hitherto unknown.

It has been supposed for quite a while that proteolytic enzymes may be involved in protein synthesis. A number of investigators—as, for example, Wasteneys and Borsook (26), Taylor (24), and Voegtlin and his collaborators (25)—have studied the changes in substrate concentration, pH, and activation which might be necessary in order to divert the action of those enzymes from protein splitting to protein synthesis. All these investigators had to perform their experiments with very complex mixtures of protein digestion products, the very structure of which was unknown.

Under such experimental circumstances a large number of chemical and physical conditions influence a multitude of different peptide bonds, and the analytical findings are often a summation of many conflicting, partial results. In order to obtain clear-cut results, it is necessary to simplify the experimental conditions fundamentally. Here we were able to use to advantage our finding that the enzymes which attack genuine proteins also act on very simple substrates, provided that these substrates meet the specificity requirements of the enzyme involved.

Our experiments were performed with the best known intracellular proteinase, the papain of the so-called melon tree. It was readily established (4, 5) that such enzymes, when activated, are capable of performing four different types of reaction (table 5).

(*1*) The hydrolytic effect of papain, which may best be illustrated by the splitting of benzoylglycine amide into benzoylglycine and ammonia, is well known.

(*2*) The phenyl derivative of the benzoylglycine amide just mentioned is not at all split by papain; on the contrary, it is synthesized from benzoylglycine and aniline when papain is present. Such syntheses, in many cases, reach 100 per cent and proceed with remarkable speed. When the synthesis of benzoylleucine anilide from benzoyl-*l*-leucine and aniline is performed in a 2 to 3 per cent solution, after 5 minutes the mixture is already solidified by the crystallization of the anilide synthesized. As another example the synthesis of benzoylleucylleucine anilide from benzoylleucine and leucine anilide may be mentioned. Here the enzymic action combines two natural amino acids with each other through a genuine peptide bond.

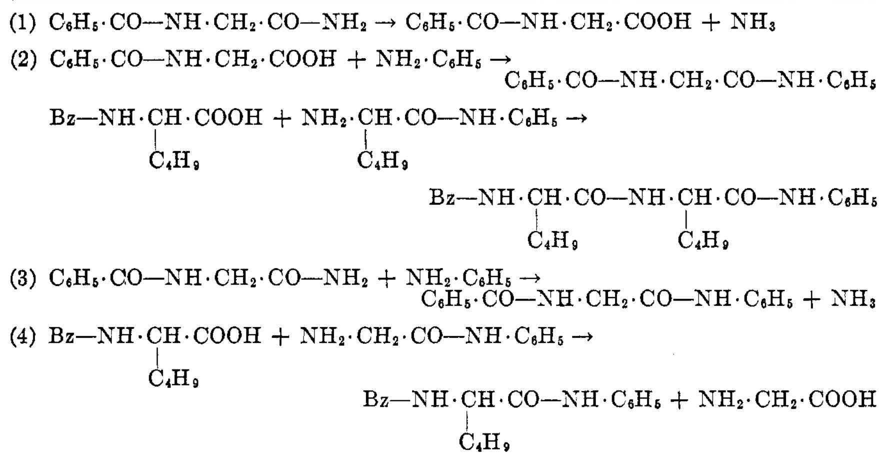
(*3*) When benzoylglycine amide is treated with a diluted solution of aniline in the presence of papain, the aniline replaces the amide group with the formation of benzoylglycine anilide.

(*4*) On the other hand, when a mixture of benzoylleucine and glycine anilide is treated with papain, the benzoylleucine replaces the glycine residue with the formation of benzoylleucine anilide.

Reactions of types 3 and 4 deal with the replacement of one group by another.[2] Such replacement reactions have hitherto not been considered in connection with proteolytic enzymes. And yet they may in many cases complicate the situation when we perform an enzymic digestion of a protein.

TABLE 5

*Reaction types catalyzed by papain*



(1) $C_6H_5\cdot CO{-}NH\cdot CH_2\cdot CO{-}NH_2 \rightarrow C_6H_5\cdot CO{-}NH\cdot CH_2\cdot COOH + NH_3$

(2) $C_6H_5\cdot CO{-}NH\cdot CH_2\cdot COOH + NH_2\cdot C_6H_5 \rightarrow C_6H_5\cdot CO{-}NH\cdot CH_2\cdot CO{-}NH\cdot C_6H_5$

$Bz{-}NH\cdot \underset{C_4H_9}{\underset{|}{CH}}\cdot COOH + NH_2\cdot \underset{C_4H_9}{\underset{|}{CH}}\cdot CO{-}NH\cdot C_6H_5 \rightarrow Bz{-}NH\cdot \underset{C_4H_9}{\underset{|}{CH}}\cdot CO{-}NH\cdot \underset{C_4H_9}{\underset{|}{CH}}\cdot CO{-}NH\cdot C_6H_5$

(3) $C_6H_5\cdot CO{-}NH\cdot CH_2\cdot CO{-}NH_2 + NH_2\cdot C_6H_5 \rightarrow C_6H_5\cdot CO{-}NH\cdot CH_2\cdot CO{-}NH\cdot C_6H_5 + NH_3$

(4) $Bz{-}NH\cdot \underset{C_4H_9}{\underset{|}{CH}}\cdot COOH + NH_2\cdot CH_2\cdot CO{-}NH\cdot C_6H_5 \rightarrow Bz{-}NH\cdot \underset{C_4H_9}{\underset{|}{CH}}\cdot CO{-}NH\cdot C_6H_5 + NH_2\cdot CH_2\cdot COOH$

With respect to the problem of the biological synthesis of proteins, our finding that the four types of enzymic reactions described above occur under identical conditions of pH, activation, etc., seems to be of some significance. No longer can it be claimed to be a general rule that hydrolysis occurs at one pH and synthesis at another.

Which one of the four types of reaction discussed above may occur in a living cell at a given moment is determined in the first place by specificity phenomena, that is, by the specificity of the enzyme involved and by the structure of the substrates available. The mutual interdependence between the substrate structure and the specific action of an enzyme may be

[2] It is at present not certain whether replacements of type 4 proceed directly or through some more complex mechanism.

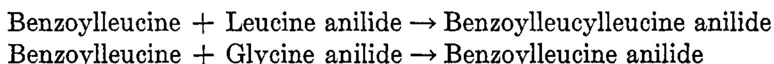
illustrated by the fate of several comparative series of substrates when in contact with an enzyme.

*Series A (papain)*

| Benzoylleucylglycyl\|glycine | Hydrolyzed |
|---|---|
| Benzoylleucyl\|leucylglycine | Hydrolyzed |

Series A deals with two molecules of identical chain length and identical general type. They differ only with respect to the nature of one aliphatic amino acid residue, and yet this small difference effects the shift of the point of hydrolysis from one peptide bond to another.

*Series B (papain)*



Benzoylleucine + Leucine anilide → Benzoylleucylleucine anilide
Benzoylleucine + Glycine anilide → Benzoylleucine anilide

Here we compare two reaction systems of the same type. They differ only in that the first system contains a leucine residue, whereas the second system contains a glycine residue. This difference causes the two substrates to react differently under the influence of papain. The first system undergoes a synthesis, the second system a replacement reaction.

*Series C (papain)*

| Benzoylleucyl\|leucylglycine | Hydrolysis |
|---|---|
| Benzoylleucine + Leucine anilide | Synthesis |

The two systems considered in series C contain the same amino acid residues as constituents participating in the enzymic reaction. The two systems differ, however, with respect to substituents outside the sphere of enzymic action. As a result, the two systems react under the influence of papain in opposite directions.

*Series D (chymotrypsin)*

| Benzoyltyrosine amide | No hydrolysis |
|---|---|
| Benzoyltyrosylglycine | No hydrolysis |
| Benzoyltyrosyl\|glycine amide | Quick hydrolysis |
| Benzoyltyrosylglycylglycine amide | Very slow hydrolysis |

Series D deals with substrates for chymotrypsin. Exactly as in series C, we find remote sections of the molecule influencing the splitting of a given peptide bond.

The various reactions just discussed demonstrate the pronounced specificity of proteinases and, on the other hand, the rather complex nature of

this specificity. By the term "enzymatic specificity" we have usually designated the fact that a relatively slight modification of the substrate may inhibit the enzymatic action. The specificity of proteinases, however, is such that structural alterations of the substrate very often do not inhibit the action of the enzyme, but shift the point of enzymatic attack or alter the type of enzymic reaction. A proteinase is capable of producing a variety of reactions and of reacting on a variety of substrates. Nevertheless, on each substrate it performs in general only one unequivocal reaction. This combination, in a single enzyme, of versatility and unequivocal action may be of the greatest significance in the biological synthesis of individual proteins.

Before we discuss the possible course of such a biological synthesis, there should be mentioned another aspect of the specificity phenomenon, having its basis in the complex nature of the substrate. It lies in the fact, which was mentioned before, that the enzymatic behavior of a peptide bond is frequently influenced by distant sections of the substrate molecule. A peptide bond between two individual amino acid residues is not an independent entity endowed with specific properties of its own. The energy content and the specific nature of a peptide bond are the resultants of the interaction of many sections of the entire substrate molecule. Such an interaction of the various sections of a molecule is common to all kinds of compounds. However, its consequences are very impressive and extremely significant in the synthesis of protein molecules, since here we have to deal with the action of enzymes possessing a highly sensitive specificity and with a reaction sequence consisting of at least hundreds of single steps, each of which offers a great variety of modifications.

Thus, for example, during the synthesis of a fibroin molecule several thousands of peptide bonds are formed. We may visualize this synthesis as consisting of a large number of single steps. A few of them are represented in table 6. Each of these steps means the attachment of an amino acid residue or a peptide residue to the rudimentary fibroin molecule. In the interest of simplicity only the attachment of single amino acid residues is considered in table 6. In each single step of the synthesis the enzyme makes a precise selection from the variety of available amino acids, attaching in one step of the synthesis glycine, in the next step alanine, then glycine, tyrosine, glycine, alanine, and so on. Each step of the synthesis changes the size and the structure of the rudimentary fibroin molecule, and simultaneously the nature of the amino acid to be attached varies also from step to step. The synthesizing enzyme has to act on a different substrate in each single step; by its action in one step it has used up the reaction product of the foregoing step and has synthesized the substrate of the next step.

Thus, according to this picture, the highly organized pattern of a protein molecule is the result of a reaction sequence, consisting of numerous single yet interdependent steps. It is the capability of the proteinases to perform long sequences of reactions in an unequivocal way that is instrumental in producing the unequivocal pattern of an individual protein. Neither an enzyme specifically restricted to a single substrate nor an enzyme acting on various substrates in an unspecific manner could produce the unique pattern of an individual protein. We can explain the formation of proteins only by postulating the presence in living organisms of enzymes capable of acting on a multitude of substrates and having the property of acting on each of these substrates in a sharply defined manner. The specificity of an individual enzyme predetermines the molecular pattern of the protein synthesized by this enzyme. The numerical rules governing a protein molecule have their basis in the specificity of the enzyme involved. Here

TABLE 6

*Synthesis of a segment of the fibroin molecule*

The symbol *R* refers to the rudimentary fibroin molecule at a given stage of the synthesis. The symbols G, A, and T refer respectively to the residues of glycine, alanine, and tyrosine.

| | |
|---|---|
| R + glycine | → R·glycine |
| R·glycine + alanine | → R·G·alanine |
| R·G·alanine + glycine | → R·G·A·glycine |
| R·G·A·glycine + tyrosine | → R·G·A·G·tyrosine |
| R·G·A·G·tyrosine + glycine | → R·G·A·G·T·glycine |
| R·G·A·G·T·glycine + alanine | → R·G·A·G·T·G·alanine |
| R·G·A·G·T·G·alanine + glycine | → R·G·A·G·T·G·A·glycine |
| R·G·A·G·T·G·A·glycine + arginine | → R·G·A·G·T·G·A·G·arginine |

we arrive for the first time at a physicochemical concept of the predetermination which is an inherent attribute of many phenomena of life.

The question of whether hereditary phenomena are connected with and explained by a transmission of individual proteins has frequently been discussed. On the basis of the conclusions which we have reached, it seems that the essential substances transmitted from one generation of cells to the next must be enzymes, and that they have to be enzymes gifted with the capability of synthesizing individual proteins by predetermined sequences of specificity reactions. There is already considerable evidence that proteinases themselves are proteins, or contain proteins as essential molecular constituents. Therefore the proteinases owe their existence to the preëxistence of other proteinases. There is, in life, a practically endless sequence of sequence reactions, in which one proteinase synthesizes the next by a predetermined reaction, and so forth. The sequence breaks

off whenever a proteinase has synthesized a protein that does not possess enzymatic properties.

Will we ever be able to copy *in vitro* the synthesis of natural proteins? I do not know whether we may succeed sooner or later (probably later) in synthesizing proteins with the aid of proteinases and without the coöperation of living cells. However, I am doubtful how much a synthesis of this kind would add to our understanding of protein chemistry and life phenomena. At the present moment it would seem to be of still greater significance that we have available methods to study the composition and the transformations of proteins and the specificity of proteinases with the same precision as in the case of simpler substances and simpler phenomena. Thus we may hope to extract much new information about the numerous physiological and pathological processes that are dependent upon the formation, the presence, or the transformation of proteins.

## REFERENCES


(1) Bergmann, M.: J. Biol. Chem. **110**, 471 (1935).
(2) Bergmann, M.: J. Biol. Chem. **122**, 569 (1938).
(3) Bergmann, M., and Fox, S. W.: J. Biol. Chem. **109**, 317 (1935).
(4) Bergmann, M., and Fraenkel-Conrat, H.: J. Biol. Chem. **119**, 707 (1937).
(5) Bergmann, M., and Fraenkel-Conrat, H.: J. Biol. Chem. **124**, 1 (1938).
(6) Bergmann, M., and Fruton, J. S.: J. Biol. Chem. **118**, 405 (1937).
(7) Bergmann, M., Fruton, J. S., and Fraenkel-Conrat, H.: J. Biol. Chem. **119**, 35 (1937).
(8) Bergmann, M., Fruton, J. S., and Pollok, H.: Science **85**, 410 (1937).
(9) Bergmann, M., and Niemann, C.: J. Biol. Chem. **115**, 77 (1936).
(10) Bergmann, M., and Niemann, C.: J. Biol. Chem. **118**, 301 (1937).
(11) Bergmann, M., and Niemann, C.: J. Biol. Chem. **122**, 577 (1938).
(12) Bergmann, M., and Ross, W. F.: J. Am. Chem. Soc. **58**, 1503 (1936).
(13) Bergmann, M., Zervas, L., and Fruton, J. S.: J. Biol. Chem. **111**, 225 (1935).
(14) Bergmann, M., Zervas, L., and Fruton, J. S.: J. Biol. Chem. **115**, 593 (1936).
(15) Bergmann, M., Zervas, L., and Ross, W. F.: J. Biol. Chem. **111**, 245 (1935).
(16) Bertho, A., and Grassmann, W.: Biochemisches Praktikum, p. 99. Walter de Gruyter and Co., Berlin (1936).
(17) Fischer, E.: Ber. **39**, 530 (1906).
(18) Fischer, E.: Untersuchungen über Aminosäuren, Polypeptide und Proteine, Vol. II, p. 41. Julius Springer, Berlin (1923).
(19) Hofmeister, F.: Naturw. Rundschau **17**, No. 42/43 (1902).
(20) Linderström-Lang, K.: Ergeb. Physiol. exptl. Pharmakol. **35**, 415 (1933).
(21) Oppenheimer, C.: Die Fermente und ihre Wirkungen, 5th edition, Supplement, p. 611. W. Junk, The Hague (1936).
(22) Reference 21, p. 618.
(23) Svedberg, T.: Ind. Eng. Chem., Anal. Ed. **10**, 113 (1938).
(24) Taylor, A. E.: J. Biol. Chem. **3**, 87 (1907).
(25) Voegtlin, C., Maver, M. E., and Johnson, J. M.: J. Pharmacol. **48**, 241 (1932).
(26) Wasteneys, H., and Borsook, H.: Physiol. Rev. **10**, 110 (1930).