

# Pepsinogen and Pepsin

ROGER M. HERRIOTT

From the Department of Biochemistry, The Johns Hopkins University School of Hygiene and Public Health, Baltimore

**ABSTRACT** Evidence relating to the structure and properties of swine pepsinogen and pepsin has been reviewed and used to suggest a tentative two dimensional picture of the skeleton of these two proteins. When pepsinogen, a folded single peptide chain, is converted to pepsin, there is a profound change in the physical and chemical properties of the protein. In an as yet unknown manner, except that it is initiated by a peptic cleavage of the protein chain, a single enzymic site is formed. This site is made up, quite probably, of the secondary carboxyl group of glutamic acid or of aspartic acid and a tyrosine phenol group in close proximity so that they can form hydrogen or hydrophobic bonds with the substrate in some unique manner that permits hydrolysis to occur at an accelerated rate.

## INTRODUCTION

Volumes have been written about the action of enzymes, but little is known which explains their catalytic action in chemical terms. This is particularly true of the hydrolytic enzymes which have no prosthetic groups or coenzymes. It is generally assumed that in these enzymes some tertiary structure (1) of the protein developed, perhaps, by the folding of the long peptide chain, is responsible for the enzymic property.

Today, largely as a result of the influence of Northrop's work, many highly purified enzymes, and in some instances, their precursors, are available for structural studies, and a number of laboratories are using a variety of means to uncover the explanation for enzyme catalysis. The work on ribonuclease (2-4), chymotrypsin and trypsin (5-10), papain (11, 12), acetylcholinesterase (13), and fumerase (14) is leading the advance, yet in no case has the structure which binds the substrate been well established.

A solution to these problems comes slowly, for unless one has the good fortune to be able to digest away most of the protein without loss of enzymic function, as was found for papain (11), the approaches are indirect and the conclusions tentative. These indirect approaches are of many different types, not all of which have been applied to pepsin. A total amino acid sequence

may be determined and the position of any cross-bonding, as has been obtained for insulin (15) and ribonuclease (2, 3), and will probably be completed for others before long. Useful and important for the total picture as this is, the amino acid sequence of the enzyme is not particularly helpful to an understanding of the site of substrate interaction with the enzyme. Denatured enzymes have the same amino acid sequence as the native protein, but no enzymic activity, so it is presumed that tertiary folding of the protein chain is needed to bring particular amino acid side groups into close proximity for catalytic action on small substrate structures, such as dipeptides. This spatial proximity of the two or more critical side groups of the enzyme would be expected to be disturbed by denaturation of the protein since this is almost invariably accompanied by shape changes.

A second approach to an understanding of the critical structure of an enzyme is made by determining the relationship of chemically reactive side groups of the protein to its enzymic function, using a variety of chemical modifications with relatively specific chemical reagents. The inactivation of esterases (some of which are also proteases) by diisopropyl fluorophosphate (DFP) (16) is considered a model of this approach, for in this case one mole of DFP inactivates one mole of enzyme.

The effect of pH on the enzymic activity, and the determination of the heat of ionization of titrable structures essential for activity also provide useful information. Another approach which in recent years has been utilized with moderate success is the digestion of the active proteins or their precursors proteases or peptidases under varied conditions while following the catalytic function of the degraded parts. Occasionally, the structure around the "active site" is not disrupted, though extensive cleavage of other parts of the molecule is observed (11). In the precursor, the active site may be protected, allowing other less essential components of the molecule to be removed (17). There is no predicting the course of such a study, but the variety of proteases and conditions that can be employed and the experience thus far, indicate its potential. The use of low molecular weight synthetic substrates and their isomers that inhibit the enzyme provides a still different approach (18). Information from such studies permits certain restrictions to be placed on the components of the active site.

It is the purpose of this paper to review all available information on pepsin and pepsinogen, and using it, to suggest a structure that might show these properties.

The seed for this work was sown many years ago when, having isolated crystalline pepsin (19), Northrop initiated a program to study its chemical structure.

Excellent reviews on protein structure (20, 21) and enzymes (22-24) are recommended for those wishing additional information. Two of the reviews

(23, 24) are concerned with the active sites of enzymes, and there is one devoted to pepsin (25).

## PEPSINOGEN

### *Stability to Alkalinity and Heat*

Swine pepsinogen, a protein of 43,000 molecular weight, was isolated in crystalline form from stomach mucosae (26, 27). Its stability in neutral and slightly alkaline solution, as well as its catalytic inactivity, distinguishes it from the enzyme pepsin. Denaturation of this precursor at temperatures above 55°C and pH 7, or by raising the pH to 11 at room temperature, is reversible in the absence of salts. The alkali denaturation is accompanied by the removal of two protons (27) and the  $pK$  of these two groups must be in the neighborhood of 9.4. This is close to the  $pK$  of  $\epsilon$ -amino groups of lysine and phenolic groups of tyrosine in proteins.

When pepsinogen is converted to pepsin at acidities below pH 5, the region of protein pH stability or instability shifts 2 pH units toward the acid side, and simultaneously, the protein becomes extremely labile to alkalinities that did not harm the precursor. The changes in protein structure occurring during this conversion are not well understood. The reaction is catalyzed by pepsin and results in splitting off about 20 per cent of the molecule. The liberated fragments carry most of the basic amino acids of the precursor. There is probably a drastic change in fundamental structure for in addition to the pH instability change, the reversibility of denaturation is largely lost on conversion to pepsin.

### *The Nature of the Protein Structure*

End-group analyses of pepsinogen (28, 29) have revealed only a single  $\alpha$ -amino group belonging to leucine and one terminal carboxylic amino acid of alanine (30, 31). These findings gave rise to the tentative hypothesis that the basic backbone of pepsinogen is a single unbranched peptide chain. Pepsinogen is probably not composed of two or more peptide chains held together by disulfide bridges, for in addition to the end-group analyses, Kern (32) found that after reduction of all three cystine disulfide bridges, the protein yielded the same osmotic pressure as the native untreated protein. Had there been two or more chains held together by disulfide bonds, the osmotic pressure should have risen proportionally on reduction. If a second chain was small enough to diffuse through the membrane it would not raise the osmotic pressure, but no nitrogenous products were observed outside the membrane. This evidence supports the single peptide chain concept for the structure of pepsinogen.

### *The Shape of the Molecule*

A solution of straight or uncoiled molecules of 43,000 molecular weight would be expected to have a high intrinsic viscosity. Since direct measurements of native pepsinogen showed it to have a low intrinsic viscosity, it follows that the peptide chain is not a rigid rod nor a random flexible chain and may be expected to be coiled or folded in some specific manner.

The forces tending to hold the folds of proteins together in some specific fashion include hydrogen and hydrophobic bonds, electrostatic and disulfide bonds (33). The last are the strongest, but there are only three per molecule of pepsinogen, whereas the hydrogen and hydrophobic bonds are more numerous, but weak. Kern (32) observed that reduction of the equivalent of one disulfide bond of pepsinogen with mercaptoethanol produced only a small change in the viscosity of the protein solution and no loss of capacity to form pepsin. This result was also obtained by Bovey and Yanari (34) using sulfite as the reducing agent. Soon after the equivalent of a second —S—S— bond had been reduced there was a sharp rise in the viscosity and a loss of potential enzymic function. The equivalent of a third bond could not be reduced under conditions of temperature and pH which left control pepsinogen stable, suggesting that the third disulfide is either less accessible to the reducing agent or it is a stronger bond by virtue of surrounding structures. Although reaction mixtures were analyzed in this work by Kern, the nature of the results suggests that the three disulfide bridges reduce at such different rates that it amounts to a specific reduction with little overlap (see Fig. 1). Since the three groups reduce differently and the reduction of each produces a different effect, it has been tentatively concluded that the three bridges are located in significantly different positions in the molecule. The first bridge may link only a short segment and be remote from the catalytic site or point of substrate attachment, since its reduction affects neither the viscosity nor peptic function. The second bridge must link large and vital segments of the molecule, for its reduction is soon followed by an increase in viscosity and a loss in catalytic function. It is important to note that these changes *followed* reduction and because the loss of function had a high temperature coefficient, it seems probable that the viscosity and functional changes resulted from denaturation that followed reduction of the second disulfide bond. The third disulfide bridge is different from the other two as shown by its resistance to reduction under conditions used for the first two. An attempt was made to indicate the three disulfide bonds in the diagram in Fig. 3 so that they would reflect the properties noted above. It must be stressed, however, that the true position of the disulfide bridges is not known.

## CONVERSION OF PEPSINOGEN TO PEPSIN

Conversion of pepsinogen takes place in solutions which are more acidic than pH 5. This reaction is autocatalytic; *i.e.*, one of the products (pepsin) catalyzes the conversion. Proof of this conclusion was obtained by observing an

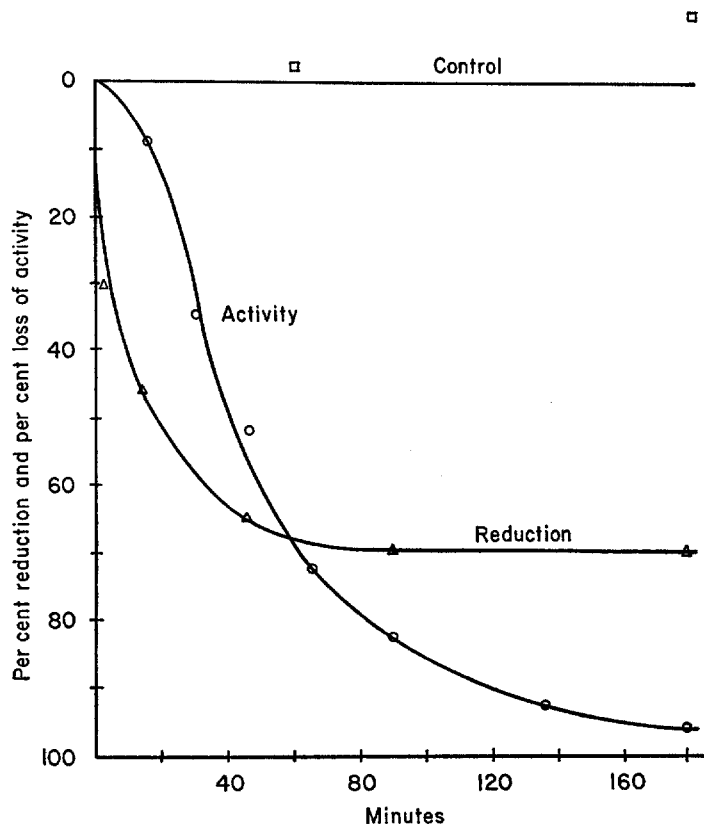


FIGURE 1. Reduction of pepsinogen by mercaptoethanol (Kern).

increased rate on addition of pepsin to the mixture (35). One of the polypeptides liberated during conversion inhibits peptic action at pH 5.5 (35).

Seven to nine (28) peptide bonds are hydrolyzed during the formation of pepsin, but if the single chain concept is correct, it is probable that cleavage of only one of these bonds is required to release the active enzyme and that the other six to eight bonds may reflect other susceptible linkages, unrelated to the release of activity. It would appear likely that at least one peptide bond must be split during conversion, for pepsinogen is not active and pepsin which catalyzes the conversion cleaves only peptide bonds.

*The Position of Pepsin and the Inhibitor in Pepsinogen*

Studies on the amino acid sequence at the *N*-terminal end of pepsin, (28, 31, 36–38), pepsinogen (28), and the inhibitor isolated from activated pepsinogen (28) were found to be different (see Table I) so that neither the enzyme nor the inhibitor can occupy the *N*-terminal position of the precursor. Carboxypeptidase acting specifically on the C-terminal amino acids liberated the same three amino acids, alanine, one of the leucines, and then valine from both pepsin and pepsinogen (30). This means that pepsin probably occupies the carboxyl terminal position of pepsinogen. Williamson and Passmann (39) reported that both the carboxypeptidase and lithium aluminum hydride methods of degrading the carboxyl end of protein chains agreed in showing that two moles of alanine were liberated from pepsin and these were followed

TABLE I  
NITROGEN TERMINAL AMINO ACID SEQUENCE

| Material   | <i>N</i> -Terminal amino acid | Second amino acid in chain |
|------------|-------------------------------|----------------------------|
| Pepsinogen | Leucine                       | Isoleucine or leucine      |
| Inhibitor  | Leucine                       | Glutamic acid              |
| Pepsin     | Isoleucine                    | Glycine                    |

by glutamic acid. These workers conclude that pepsin consists of a branched chain or two chains. Since the earlier study (30) showed only one mole of alanine per mole protein despite extended exposure to carboxypeptidase and no trace of glutamic acid was liberated, it is tentatively suggested that the pepsin preparations were different.

With pepsin occupying the C-terminal position of pepsinogen, the *N*-terminal  $\alpha$ -amino group of isoleucine in pepsin must be one component of a peptide bond in pepsinogen that is split during the autocatalytic conversion. Another of the peptide bonds hydrolyzed during formation of the enzyme must contain the leucyl *N*-terminal amino acid of the inhibitor. The five remaining bonds that are broken have been assigned to the "miscellaneous peptides" which have an average molecular weight of 1,000 and which are found in the neutral fraction (28) of the TCA-soluble portion of a pepsinogen activation mixture.

*Amino Acid Composition*

The amino acid composition of the precursor, pepsin, inhibitor, and miscellaneous peptides, has been recorded elsewhere (28). New values for some of

the amino acids have been found recently (40). Although the subject will be dealt with in detail later, it is worth noting here that on conversion to pepsin there is a disproportionate loss of basic amino acids. This, coupled with the finding that in pepsinogen the  $\epsilon$ -amino groups of the twelve lysines are free (28), may have special significance for the greater stability of pepsinogen than pepsin in weakly alkaline solutions.

#### PEPSIN

Pepsin is a single chain peptide (28, 36) with very little helical coiling, as judged by rotatory dispersion measurements (41, 42). It is folded in such a way that a relatively globular particle is obtained with an axial ratio of 3.0 (43), assuming it to be an unhydrated prolate ellipsoid. The bonds holding the folds of this protein together are not known to be significantly different from those in pepsinogen, but the difference in electrostatic charges noted above may be of importance. Urea concentrations up to four molar have little effect on inactivation of pepsin (44, 45), so that hydrogen bonding may not play the main role in stabilizing the protein. Perlmann (42) has noted that pepsin contains nearly double the usual percentage of amino acids with hydrophobic side chains. Just how much stability is contributed to the pepsin molecule by hydrophobic bonding is not clear, but it should be greater than in most proteins.

The ionizable groups of pepsin consist of thirty-six carboxyls (46, 47), one phosphodiester, which is not essential (48), one  $\epsilon$ -amino group of lysine (40), one  $\alpha$ -amino group of isoleucine (28), between nine and twelve (49) of the 16–18 tyrosine phenol groups, and perhaps the imidazole of the one histidine, two guanido of arginine, and four or six tryptophane indoles (28, 40). The isoelectric point, if one exists for pepsin, must be below pH 1, for at this acidity the protein had not become cationic (50, 51). Some of the physical properties of pepsin are listed in Table II.

#### *Alkaline Denaturation*

Despite Northrop's clear demonstration of the reversibility of alkali denaturation of pepsin (53), the yield was so low that from many points of view the process may be regarded as not readily reversible.

The kinetics of the inactivation have been carefully analyzed by a number of investigators with one finding that the rate of inactivation varied inversely as the fifth power of the hydrogen ion concentration (54), while others (55, 56) found lower values. Casey and Laidler (57), working at pH 4.83 and 50 to 60°C, report that the rate is a function of the pepsin concentration as well as the hydrogen ion concentration. As the concentration of protein decreased,

the exponential dependence on the hydrogen ion (or hydroxyl ion) increased. In view of enormous disparity in cationic *vs.* anionic charges in pepsin, four or fewer against thirty-six respectively near pH 6, the electrostatic repulsive forces might be expected to play a contributory role in the denaturation of pepsin. However, increased ionic strength of the medium should reduce the effect of such electrostatic forces, yet an acceleration of inactivation was observed when the ionic strength was increased (54, 56). In the first of these studies (54), *p*-nitrophenol was the buffer salt used, and this might bind with

TABLE II  
PHYSICAL PROPERTIES OF SWINE PEPSIN

| Property                              | Method or symbol       | Value                    | Reference                         |
|---------------------------------------|------------------------|--------------------------|-----------------------------------|
| Molecular weight                      | Sedimentation equation | 35,500                   | Philpot and Erickson-Quensel (98) |
|                                       |                        | 34,500                   | Herriott (27)                     |
|                                       | Phosphorus content     | 34,800                   | Kroman and Stern (99)             |
|                                       |                        | 37,600                   | Yasnoff and Bull (100)            |
|                                       |                        | 35,000                   | Dieu (65)                         |
| Specific optical rotation             | Osmotic pressure       | 35,000                   | Northrop (19)                     |
|                                       |                        | -65                      | Perlmann (42)                     |
| Specific rotatory dispersion constant | (pH 4.6)               | 216 m $\mu$              | Jirgensons (41, 52)               |
|                                       |                        |                          | Jirgensons (41)                   |
| Intrinsic viscosity                   |                        | 0.032 dl-g <sup>-1</sup> | Perlmann (42)                     |
| Sedimentation constant                | $s_{20,1w}$            | $3.2 \times 10^{13}$     | Edelhoch (43)                     |
| Diffusion constant                    | $d_{20,1w}$            | $9.0 \times 10^7$        | Philpot (98)                      |
| Frictional coefficient                |                        | 1.1                      | Philpot <i>et al.</i> (98)        |
| Axial ratio                           | $a/b$                  | 3.0                      | Edelhoch (43)                     |
| Isoelectric point                     | Electrophoresis        | pH 1.0                   | Edelhoch (43)                     |
|                                       |                        |                          | Tiselius <i>et al.</i> (50)       |

the protein and increase the repulsive forces, but such an argument may not be so readily applied to the NaCl used by Edelhoch (56). If no negative ions are adsorbed by the protein, a totally different explanation for pepsin's alkaline denaturation must be sought.

It was suggested some time ago (58) that during the alkali denaturation of pepsin, acidic groups are liberated. This has been confirmed (56, 59, 60), and the *pK* of these groups is almost certainly in the region assigned to carboxyl groups (pH 3-5). No detectable amino or basic groups were liberated simultaneously. There are several possible explanations for this, and very little experimental evidence. A prolyl peptide, an imide, or an ester are possible covalently bonded structures which on cleavage would release an acid equivalent without a corresponding basic one. None of these bonds as usually encountered in small molecules would break at neutrality and room



temperature, but too little is known to exclude them from consideration. Other possibilities currently receiving attention are hydrogen-bonded tyrosine phenol-carboxyl groups (45) and hydrogen-bonded carboxyl-carboxyl groups (45, 56, 60). These are certainly weaker than the covalent bonds, and it is possible that carboxyl-carboxyl hydrogen bonding in pepsin is important in acid solution (61), but it is not clear at what pH this bonding will tend to break. Steinhardt's (62) studies on hemoglobin are pertinent in this field.

The presence of tyrosine phenol-carboxyl hydrogen bonds in pepsin from spectrophotometric evidence (63) cannot be said to be established in the light of other studies (64), but neither are they excluded. More than usual consideration must be given to this type bonding, for Li (49) showed many years ago from studies of the rate of iodination of pepsin at pH 6 that some (about six) of the tyrosines of pepsin were not iodinated unless the protein was denatured. Alkali denaturation of pepsin made available for iodination some, but not all, these tyrosines, but the combination of heat and alkali liberated all of them. With both phenolic groups and carboxyl groups being liberated on denaturation, the plausibility of a phenolic-carboxyl bonding is increased.

The denaturation of pepsin is accompanied by a several fold increase (32, 43) in intrinsic viscosity which means that the axial ratio of an unhydrated prolate ellipsoid (66) changed from 3 to 16 in an ionic strength of 0.01 (43). Disulfide bonds are not ruptured in this process (32). No change in optical rotation is observed on denaturation (67, 68), but there is a small drop in the rotatory dispersion constant,  $\lambda c$  (41, 42).

It has been reported (43) that alkali denaturation of pepsin is accompanied by a cleavage resulting in a decrease in molecular weight as measured by sedimentation and light scattering. This should produce a doubling in osmotic pressure if the products are non-diffusible, or a dialyzable component should be detectable if there is no change in osmotic pressure. Neither of these conditions was observed by the present writer when the denaturation was carried out rapidly at pH 8 and 65°C to insure that no residual pepsin remained which could digest the denatured protein.

#### *Acid Stability*

Pepsin is remarkably stable to acid. The writer (69) has kept it in normal H<sub>2</sub>SO<sub>4</sub> at 5°C for weeks with only minor losses of peptic activity. In view of its low basic amino acid content, there are perhaps too few electrostatic forces in strong acid to induce disruption of this protein. In addition, it seems quite reasonable to expect a fair amount of carboxyl-carboxyl hydrogen bonding in strong acid (60-62), which will tend to stabilize the protein. Lindley (24) has pointed out that it is probably very important for an enzyme to be quite stable at its pH optimum, for the distances between components of the active

site are probably critical to its functioning as a catalyst. If the molecule is pliable, or extensible by the application of stresses normally encountered in the environment, the critical distances might change, producing sharp discontinuities in the catalytic efficiency.

### *The Active Site of Pepsin*

In this discussion no distinction is made between structures which bind the substrates and those which have catalytic activity as was discovered in chymotrypsin (18).

It has been known from chemical inactivation studies of pepsin that one or more of the carboxyl groups (presumably of aspartic or glutamic acids) and phenolic groups of tyrosine must be in some way associated with the active structure. Thus, acetylation (70, 71), iodination (72-74), and nitration (75), which modify the tyrosine moieties, and sulfur mustard (47) which esterifies carboxylate ions, destroyed the enzymatic activity. In one of these studies recovery of the activity was demonstrated by hydrolytic removal of the acetyl radicals from the phenol groups (69). In all these studies many groups of a given type react before the activity is reduced to a low value.

Such a result might be expected if the protein groups which react with the chemical reagent are related only secondarily to the active center. Closer analysis, however, showed that the activity fell off logarithmically from the beginning as the number of amino acid residues reacting increased (70, 72, 76) (see Fig. 2). This is the expected course if one member of a particular type of group is directly involved in the activity and the remaining members bear no direct relationship to it (77). This is only a correlation, however, and is not compelling or decisive, since the results are average values of many enzyme molecules and do not yield information about individual molecules. The correlation does not distinguish between the case of multiple independent enzymic sites on a molecule, and multiple chemically reactive groups, only one of which is a component of the active site. Invertase and urease are reported to have four active sites per molecule (24, 78), 3-phosphoglyceraldehyde dehydrogenase, two sites (79), and yeast alcohol dehydrogenase, four sites (80). Recent studies (by Mr. Richard Pharo) in the writer's laboratory with a dipeptide inhibitor, virtually establish that pepsin has but one active site. A thorough discussion of similar problems in enzymes in general is found elsewhere (23) (see also reference 81).

Support for the thesis that a carboxyl group is essential for the activity of pepsin is found in its low pH activity maximum. Only carboxyl and phosphate groups have  $pK$ 's in the region of pH 2-3 where the activity of this

enzyme changes with pH, and Perlmann's (48) removal of the one phosphorus<sup>1</sup> from pepsinogen without loss of capacity to form pepsin suggests the phospho-structure plays no role in the activity. The maximum activity at pH 1.8 suggests that the carboxyl group must be discharged or uncharged. The amino groups which would be charged at this pH cannot be implicated in the active site, for their acetylation (70) or deamination (75) failed to alter the peptic activity.

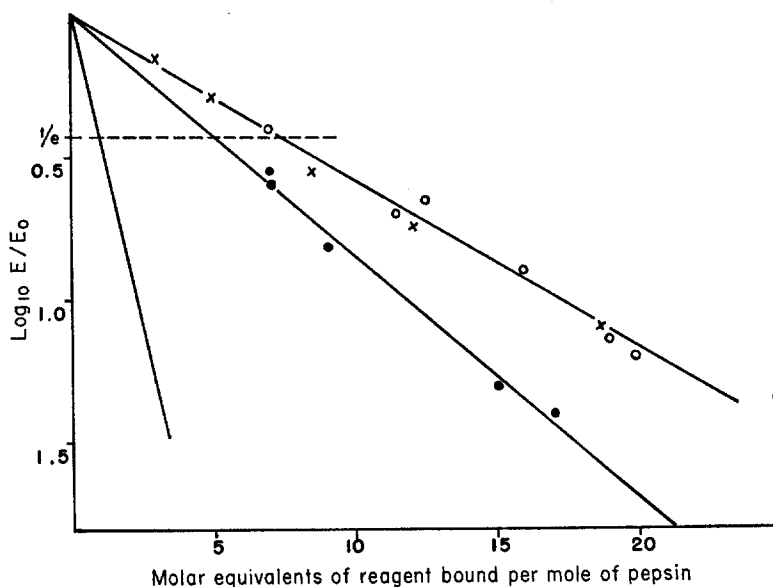


FIGURE 2. Log decrease in activity with chemical reaction. The open circles are values obtained when iodine was applied; the closed circles, when sulfur mustard was used; and the crosses, when acetylation by ketene was studied. The curve to the left is the expected curve if one reagent molecule inactivates one pepsin molecule. The average number of reagent equivalents needed for inactivation of one pepsin molecule is obtained by interpolating at the  $1/e$  level.

Evidence that the terminal carboxyl group of the protein chain is *not* the all important one for peptic activity comes from studies using carboxypeptidase (30). The sequence of amino acids in the C-terminal position of pepsin and pepsinogen is the same (alanine, leucine, valine), yet the activity of pepsin formed from carboxypeptidase-digested pepsinogen was the same as from normal pepsinogen. The essential carboxyl must be found among the

<sup>1</sup> Flavin(82) showed that one of the phospho ester bonds in pepsin is linked with the hydroxyl of serine in the amino acid sequence, -thr-ser-glu-.

$\gamma$ -carboxyl groups of glutamic acid and the second one of aspartic acid in pepsin.

In Fig. 3, one phenol and one carboxyl group have been tentatively assigned to the active group.

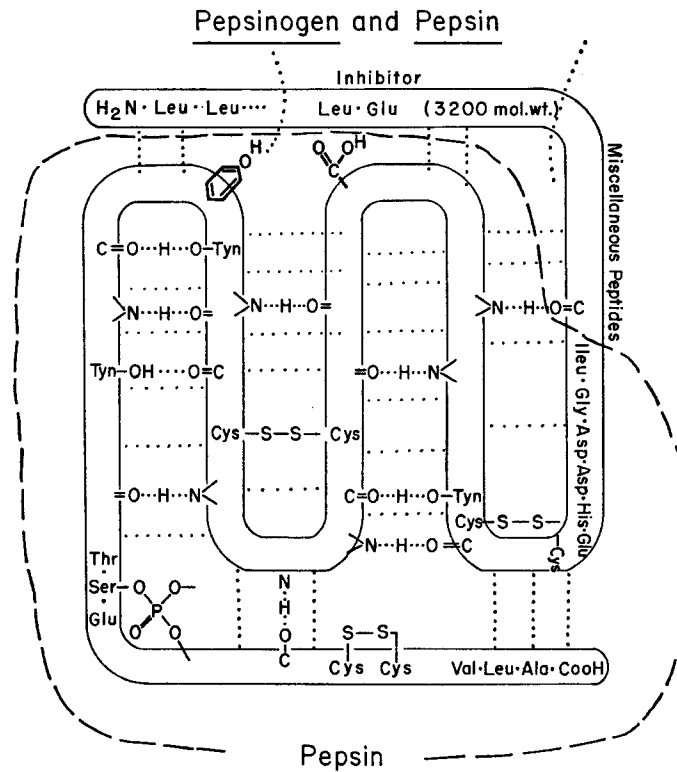


FIGURE 3. Diagrammatic sketch of pepsinogen and pepsin.

#### *Special Requirements of the Active Site*

Pepsin, like the other proteases, is known to split certain dipeptides as well as large proteins. Unless one assumes different active groups for these substrates, and there is not much evidence on this point (23), the components of the one active site will of necessity be close together in order to react with a dipeptide.

As noted earlier, if the active site consists of components of two different amino acids located at distant points in the peptide chain, but brought into close juxtaposition by the folding of the protein chain, then any spacial change, such as accompanies denaturation, would be expected to destroy the catalytic power of the protein. This is the case and it serves to explain the structural representation of the active site in Figs 3 and 4. Since nearly all proteins have carboxyl and phenol groups, it would appear reasonable that

the unique feature of pepsin may lie in the juxtaposition of these two structures, determined by the folding and bridging of the polypeptide chain. The distance between the two amino acid groups supplemented by certain restricting "contours" close to the active site may limit the structures which can interact with the enzyme. The union of enzyme and substrate must in some manner weaken the bond to be split, perhaps by making a new bond or by applying a strain (24, 83) (lowered activation energy) which then allows the bombardment of water molecules or hydronium ion to be much more effective. The great advantage of two fixed groups on a single molecule at just the

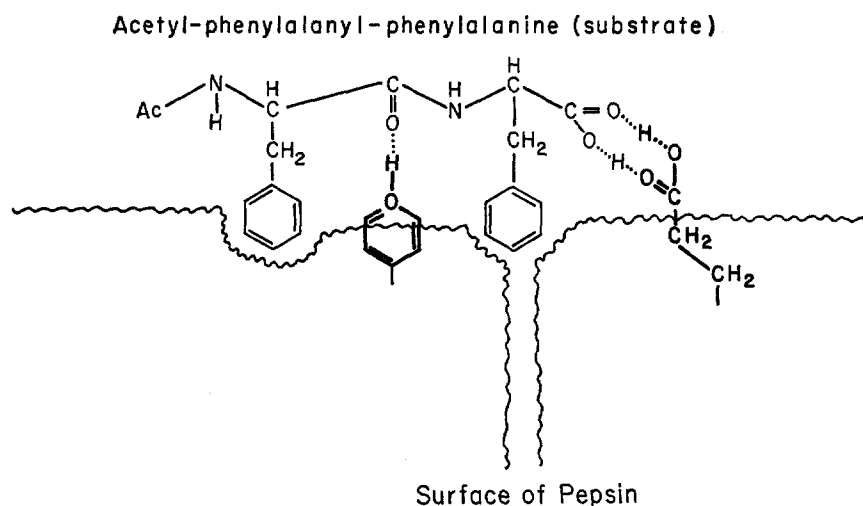


FIGURE 4. A sketch of peptic interaction with a substrate.

appropriate distance for substrate union has been discussed by Swain and Brown (84). The placement of the active site in Fig. 3 is only representative of how it may be. Just how the active site is kept from functioning in pepsinogen is not understood, but perhaps a shape change during conversion brings the two amino acid components of the active site into the required position. An obvious mechanism for inhibiting the active site in the precursor involves an interaction with that portion of the pepsinogen chain which is lost on activation. This could take place through the inhibitor component of pepsinogen, but there is no direct evidence for this proposal.

#### *Perlmann's Active Dialyzable Autolytic Fragment*

A few years ago Perlmann reported (85) that following incubation of crystalline pepsin at pH 5.5 and 37°C a fraction which dialyzed through a cellophane bag and had very little activity on proteins (hemoglobin) carried nearly full

specific activity for the synthetic substrate, acetyl-phenylalanyldiiodo-*l* tyrosine. A dialyzable product would be expected to be smaller than molecular weight 5,000. Studies of this general type had been made both by Northrop and the writer, but this was before synthetic substrates were known for this enzyme and the pH of autolysis was usually more acid. The great discrepancy between the action of the dialyzable fragment on protein and the synthetic acyldipeptide is not readily understood.

Certain bits of evidence from other sources point somewhat in the direction of Perlmann's results. Thus, it was observed some time ago that a small fraction of commercial pepsin or that prepared from pepsinogen had nearly double the specific activity (per milligram protein nitrogen) of the usual crystalline product. In addition, it has been observed (30) that at least three amino acids can be liberated by carboxypeptidase from the C-terminal end of pepsin without loss of catalytic activity. In a similar, though more dramatic manner, Hill and Smith (11) report the removal of nearly two-thirds of the amino acids from the *N*-terminal end of crystalline papain with amino polypeptidase without loss of protease activity. More recently, a small active fragment of pepsin has been reported to pass through a dialysis membrane at pH 4 (86).

#### *Inhibitors and the Active Structure*

The inhibitor isolated from activated pepsinogen combines reversibly with pepsin on a mole for mole basis<sup>2</sup> (87). Since inhibition in this case is favored at pH 5–6, it is measured as routine by the milk-clotting method at pH 5–5.5 and confirmed by protein digestions in this same region of acidity. At pH 2, dissociation of the inhibitor complex is nearly complete.

A glance at the amino acid composition of the inhibitor (88) reveals a lysine content that is higher than usual—nearly 15 per cent. This by itself has no special significance, but with the observations of Katchalski *et al.* (89) that polylysine inhibits peptic activity, the lysine in the inhibitor was of greater interest. Van Vunakis (unpublished results) has found that partially hydrolyzed polylysines also inhibit peptic activity, but the pure amino acid does not. The polylysines inhibit pepsin at pH 2 as well as at pH 5. At pH 2, charged amino ions would not be expected to combine with essentially unionized phenolic and acidic groups, so that inhibition must be through hydrogen or hydrophobic bond formation. It would be interesting to examine acetylated or deaminated polylysine as inhibitor.

The mole for mole relationship of the inhibitor from pepsinogen to pepsin

<sup>2</sup> The equivalent weight of inhibitor noted earlier (87) should have been 3500 instead of 7000. Failure to include the dilution of inhibitor in the test mixture led to recording the inhibitor concentration incorrectly too high.

(87) might have been thought to answer the question of the number of active sites were it not for the large size of the inhibitor (3100 molecular weight). Conceivably, this inhibitor could cover several active sites simultaneously. Recently, however, Mr. Pharo in the writer's laboratory has found that *N-ac-d* phenylalanyl-*l*-diiodotyrosine,<sup>3</sup> an isomer of the sensitive substrate, inhibits pepsin. Using the procedure described by Northrop (90) in which a constant quantity of inhibitor is mixed with increasing quantities of enzyme, it was found that  $5 \times 10^{-5}$  M inhibitor blocked very close to  $5 \times 10^{-5}$  M pepsin. If this action is confirmed for other substrates, including proteins, it will establish the unitary nature of the enzymic site on pepsin and lend strong support to the kinetic data that the site is made up of one carboxyl and one tyrosine phenol group.

#### SUBSTRATE STRUCTURAL REQUIREMENTS

The early studies on synthetic substrates of pepsin (91, 92) indicated maximal digestion at pH 4 and the rates were generally low, so that the relationship to the more rapid protein digestion with a maximum near pH 2 was difficult to interpret. The finding of Baker (93) that the peptide bond between adjacent aromatic amino acids was particularly susceptible to peptic cleavage and that the rate was maximal near pH 2 was, therefore, an important discovery. The rate of peptic splitting of acetyl-*l*-phenylalanyl-*l*-tyrosine at pH 1.8 was at least tenfold faster than any previously studied synthetic substrate of pepsin and iodination of this peptide to make acetyl-*l*-phenylalanyl-*l*-diiodotyrosine increased the sensitivity to pepsin almost another tenfold. Some preliminary observations in the writer's laboratory suggest that the increased lability on iodination is due to a "closer" or "better" fit to the enzyme and not to a greater strain (83) on the peptide bond as a result of iodination or to an increased "electron flow" which then labilizes the bond. This tentative conclusion was reached on finding that the rates of acid hydrolysis of these two peptides did not differ appreciably.

A rough approximation of the rate of peptic action on a number of peptides is shown in Table III. Due to differences in substrate concentrations, an occasional use of *dl* mixtures, and the assumption that the hydrolysis follows first-order kinetics throughout, only the order of magnitude of the rate constant deserves comparison.

To judge from these data, it appears that splitting is more rapid if the phenolic group is free and not acetylated. Although there were no data or comment in Baker's paper, it is presumed that acetylation of the amino group of the substrate has some importance though Harington and Pitt-Rivers (92) failed to observe any effect in their studies. An amide in place of the terminal

<sup>3</sup> Kindly given to the writer by Dr. Lilian E. Baker.

carboxyl may have rendered the structure insusceptible, but this case is complicated by its low solubility.

The prospects for a clear understanding of the substrate specificity of pepsin seemed bright from the work of Baker (93), but when Sanger *et al.*, (94, 95) examined the peptides liberated by pepsin from the A and B chains of oxi-

TABLE III  
SYNTHETIC SUBSTRATE SPECIFICITY

| No.  | Substrate                                      | Reference* | Substrate concentration | Pepsin         | Time        | Per cent hydrolyzed | $k'$ |
|------|--|------------|-------------------------|----------------|-------------|---------------------|------|
|      |  |            |                         | <i>mg N/ml</i> | <i>hrs.</i> |                     |      |
| 1    | Cbz- <i>l</i> -glu- <i>l</i> -tyr              | B          | 0.002                   | 0.06           | 72          | 12                  | 0.0  |
| 2    | Cbz- <i>l</i> -glu- <i>l</i> -diiodotyrosine   | B(pc)      | 0.00025                 | 0.04           | 24          | 19                  | 0.2  |
| 3    | Cbz- <i>l</i> -phe- <i>l</i> -phe-amide        | B          | 0.002                   | 0.1            | 144         | 0                   | 0.0  |
| 4    | Cbz- <i>l</i> -tyr- <i>l</i> -phe              | B          | 0.0005                  | 0.05           | 5           | 29                  | 1.4  |
| 5    | Cbz-O-ac- <i>l</i> -tyr- <i>l</i> -phe         | B          | 0.002                   | 0.12           | 24          | 0                   | 0.0  |
| 6    | N-ac- <i>l</i> -tyr- <i>l</i> -glu             | B(pc)      | 0.003                   | 0.06           | 72          | 5                   | 0.0  |
| 7    | N-ac- <i>l</i> -tyr- <i>l</i> -tyr             | B          | 0.001                   | 0.12           | 4           | 34                  | 0.9  |
| 8    | N-ac- <i>l</i> -phe- <i>l</i> -phe             | B          | 0.0005                  | 0.05           | 3.25        | 68                  | 7    |
| 9    | N-ac- <i>l</i> -phe- <i>l</i> -tyr             | B          | 0.003                   | 0.06           | 1           | 56                  | 14   |
| 10   | N-ac- <i>l</i> -phe- <i>l</i> -tyr             | P and H    | 0.0006                  | 0.012          | 0.33        | 6.7                 | 18   |
| 11   | N-ac- <i>l</i> -phe- <i>l</i> -diiodotyrosine  | B          | 0.0005                  | 0.012          | 0.25        | 100                 | 200  |
| 12   | N-ac- <i>l</i> -phe- <i>l</i> -diiodotyrosine  | P and H    | 0.0006                  | 0.012          | 0.33        | 44                  | 145  |
| 13   | N-ac- <i>d</i> -phe- <i>l</i> -diiodotyrosine  | B          | 0.0005                  | 0.12           | 48          | 0                   | 0.0  |
| 13 a | N-ac- <i>d</i> -phe- <i>l</i> -diiodotyrosine  | P and H    | 0.0006                  | 0.012          | 2           | 0                   | 0.0  |
| 14   | N-ac- <i>l</i> -diiodotyrosine- <i>l</i> -glu  | B(pc)      | 0.003                   | 0.06           | 5           | 33                  | 1.3  |
| 15   | N-ac- <i>dl</i> -diiodotyrosine- <i>l</i> -leu | B(pc)      | 0.001                   | 0.04           | 24          | 78                  | 1.6  |

\* B, L. E. Baker (93, 101).

B(pc), Baker, personal communication.

P and H, Pharo and Herriott (unpublished results).

†  $k'$  is a first order hydrolysis constant per milligram pepsin N at 37°C and corresponds to Baker's  $k_u$  (101).

In the calculation of  $k'$  no correction for differences in substrate concentration has been made.

dized insulin, it showed that our understanding was far from clear. It is true that the bonds between two phenylalanyl residues and phenylalanine and tyrosine (B, 24–25–26) were split rapidly, but so also was the bond joining a leucine to valine (B, 11–12). This high lability of the leucyl-valyl bond may have been unusual, since another leucyl-valyl bond (B, 17–18) was not split. From this evidence it would appear that either the surrounding amino acids have an influence or there is some special strain on the B, 11–12 leucyl-valyl bond which makes it labile.



When pepsinogen is converted to pepsin, the enzyme must cleave bonds involving leucine and isoleucine, for these amino acids occupy the *N*-terminal positions of the formed pepsin and the inhibitor.

In other proteins or polypeptides of known amino acid sequence, similar unforeseen results have been reported. In ribonuclease (96), the two -ala-ala-bonds in -ala-ala-ala-sequence were split by pepsin along with -phe-glutamine and -phe-glutamic acid. In  $\alpha$ -corticotrophin (97), -asp-asp-, -ala-ser-, -ala-phe-, and -phe-ala- bonds were split by pepsin, and in ACTH (97), it was the -ala-ser-bond. This shows that the specificity of pepsin is very broad and suggests that local conditions of stress on these bonds may be critical.

#### *The Bonds Linking Pepsin and Substrate*

Let us assume the active site of pepsin contains a carboxyl and tyrosine phenolic group. How then do they bind the substrate? At pH 1.8, both these groups will be largely unionized. The structures in pepsin likely to be charged at this pH are the amino and guanido groups, and as noted earlier, the amino groups have been ruled out of consideration (70).

The substrates of known structure also show an interesting picture at this pH. Those most rapidly split at pH 2 (see Table III) are acetyl-phenylalanyl-tyrosine (93), acetyl-phenylalanyl-phenylalanine, and acetyl-phenylalanyl-diiodotyrosine (93). None of these is ionized at pH 1.8, so even if there are charged structures (guanido) on pepsin, there are no complementary charges on the substrates. This excludes electrostatic interaction and leaves hydrogen and hydrophobic bonding (33) as the means of forming the enzyme-substrate complex necessary for subsequent hydrolysis. The acetyl-phenylalanyl-phenylalanine is particularly interesting in this regard, for it is limited in its hydrogen bonding to carboxyl and peptide structures. A possible arrangement may be visualized in Fig. 4 where the proposed active site of pepsin is hydrogen-bonded to the simplest of the sensitive substrates. Studies on the action of materials interfering with hydrogen bonding are contemplated to determine their effect on the rate of splitting of these peptides. Hydrophobic bonding between pepsin and its substrates may be critical to catalysis, but there is too little information to comment further.

At pH 5.5, where pepsin exhibits milk-clotting action, it must be supposed either that the enzyme and milk proteins hydrogen-bond without regard to electrostatic charges or that a peptic carboxylate may find a cationic structure in the casein to serve as one bond in the enzyme-substrate complex.

This work was supported in part by Training Grant 2G-73 from the United States Public Health Service.

## REFERENCES

1. LINDERSTROM-LANG, K., *Proteins and Enzymes*, Lane Medical Lectures, Stanford University Press, 1952, 58.
2. HIRS, C. H. W., MOORE, S., and STEIN, W. H., *J. Biol. Chem.* 1960, **235**, 633.
3. ANFINSEN, C. B., *Molecular Basis of Evolution*, New York, John Wiley & Sons, 1959.
4. RICHARDS, F. M., and VITHAYATHIL, P. J., *J. Biol. Chem.*, 1959, **234**, 1459.
5. NEURATH, H., and DIXON, G. H., *Fed. Proc.*, 1957, **16**, 791.
6. BALLS, A. K., and JANSEN, E. F., *Advances Enzymol.*, 1952, **13**, 321.
7. SCHAEFFER, N. K., LANG, R. P., SIMET, L., and DRISKO, R. W., *J. Biol. Chem.*, 1958, **230**, 185.
8. DESNEUELLE, P., in *The Enzymes*, (P. D. Boyer, H. Lardy, and K. Myrbäck, editors), New York, Academic Press, Inc., 2nd edition, 1960, **4**, 93.
9. DESNEUELLE, P., in *The Enzymes*, (P. D. Boyer, H. Lardy, and K. Myrbäck, editors), New York, Academic Press, Inc., 2nd edition, 1960, **4**, 119.
10. BALLS, A. K., and WOODS, H. N., *J. Biol. Chem.*, 1956, **219**, 245.
11. HILL, R. L., and SMITH, E. L., *J. Biol. Chem.*, 1958, **231**, 117.
12. SMITH, E., *J. Biol. Chem.*, 1958, **233**, 1392.
13. WILSON, I. B., in *The Enzymes*, (P. D. Boyer, H. Lardy, and K. Myrbäck, editors), New York, Academic Press, Inc., 1960, **4**, 501.
14. ALBERTY, R., *Advances Enzymol.*, 1956, **17**, 1.
15. SANGER, F., *Symp. Soc. Exp. Biol.*, 1955, **9**, 10.
16. JANSEN, E. F., NUTTING, M. D. F., JANG, R., and BALLS, A. K., *J. Biol. Chem.*, 1949, **179**, 189, 201.
17. VISWANATHA, T., WONG, R. C., and LIENER, I. E., *Biochim. et Biophysica Acta*, 1958, **29**, 174.
18. DOHERTY, D. G., and VASLOW, F., *J. Am. Chem. Soc.*, 1953, **75**, 928.
19. NORTHROP, J. H., *J. Gen. Physiol.*, 1930, **13**, 739.
20. KAUZMANN, W., *Advances Protein Chem.*, 1959, **14**, 1.
21. Protein Structure and Function, Brookhaven Symposium in Biology No. 13, Upton, New York, 1960.
22. ANFINSEN, C. B., and REDFIELD, R. R., *Advances Protein Chem.*, 1956, **11**, 1.
23. KOSHLAND, D. E., Jr., *Advances Enzymol.*, 1960, **22**, 45.
24. LINDLEY, H., *Advances Enzymol.*, 1954, **15**, 271.
25. BOVEY, F. A., and YANARI, S. S., in *The Enzymes*, (P. D. Boyer, H. Lardy, and K. Myrbäck, editors), New York, Academic Press, Inc., 2nd edition, 1960, **4**, 63.
26. HERRIOTT, R. M., and NORTHROP, J. H., *Science*, 1936, **83**, 469.
27. HERRIOTT, R. M., *J. Gen. Physiol.*, 1938, **21**, 501.
28. VAN VUNAKIS, H., and HERRIOTT, R. M., *Biochim. et Biophysica Acta*, 1957, **23**, 600.
29. OREKHOVICH, V. N., LOKSHINA, L. A., MANT'EV, V. A., and TROITSKAYA, O. V., *Doklady Akad. Nauk S.S.S.R.*, 1956, **110**, 1041.
30. VAN VUNAKIS, H., and HERRIOTT, R. M., unpublished results.

31. LOKSHINA, L., and OREKHOVICH, V., *Biokhimiya*, 1957, **22**, 699.
32. KERN, H. L., Doctoral Dissertation, Biochemistry Department, Johns Hopkins University, 1953.
33. KAUZMANN, W., in *Mechanism of Enzyme Action*, (W. D. McElroy and H. B. Glass, editors), Baltimore, The Johns Hopkins Press, 1954, 70.
34. BOVEY, F. A., and YANARI, S. S., *Abstr. Am. Chem. Soc.*, 134th meeting, Chicago, 1958, 33C.
35. HERRIOTT, R. M., *J. Gen. Physiol.*, 1938, **22**, 65.
36. WILLIAMSON, M. B., and PASSMANN, J. M., *J. Biol. Chem.*, 1952, **199**, 121.
37. WILLIAMSON, M. B., and PASSMANN, J. M., *J. Biol. Chem.*, 1956, **222**, 151.
38. HEIRWEGH, K., and EDMAN, P., *Biochim. et Biophysica Acta*, 1957, **24**, 219.
39. WILLIAMSON, M. B., and PASSMANN, J. M., *Biochim. et Biophysica Acta*, 1954, **15**, 246.
40. BLUMENFELD, O., and PERLMANN, G., *J. Gen. Physiol.*, 1959, **42**, 553.
41. JURGENSONS, B., *Arch. Biochem. and Biophysics*, 1958, **74**, 70.
42. PERLMANN, G., *Proc. Nat. Acad. Sc.*, 1959, **45**, 915.
43. EDELHOCH, H., *J. Am. Chem. Soc.*, 1957, **79**, 6100.
44. STEINHARDT, J., *J. Biol. Chem.*, 1938, **123**, 543.
45. PERLMANN, G., *Arch. Biochem. and Biophysics*, 1956, **65**, 210.
46. MEARS, W. H., and SOBOTKA, H., *J. Am. Chem. Soc.*, 1939, **61**, 880.
47. HERRIOTT, R. M., ANSON, M. L., and NORTHROP, J. H., *J. Gen. Physiol.*, 1946, **30**, 185.
48. PERLMANN, G., *Advances Protein Chem.*, 1955, **10**, 22.
49. LI, C. H., *J. Am. Chem. Soc.*, 1942, **67**, 1065.
50. TISELIUS, A., HENSCHEN, G. E., and SVENSSON, H., *Biochem. J.*, 1938, **32**, 1814.
51. HERRIOTT, R. M., DESREUX, V., and NORTHROP, J. H., *J. Gen. Physiol.*, 1940, **23**, 439.
52. JURGENSONS, B., *Arch. Biochem. and Biophysics*, 1952, **39**, 261.
53. NORTHROP, J. H., *J. Gen. Physiol.*, 1931, **14**, 713.
54. STEINHARDT, J., *K. Danske Vidensk. Selsk. Mat. fys. Medd.*, 1937, **14**, 11.
55. BUZZELL, A., and STURTEVANT, J. M., *J. Am. Chem. Soc.*, 1952, **74**, 1983.
56. EDELHOCH, H., *J. Am. Chem. Soc.*, 1958, **80**, 6640.
57. CASEY, E. J., and LAIDLER, K. J., *J. Am. Chem. Soc.*, 1951, **73**, 1455.
58. CONN, J. B., GREGG, D. C., KISTIAKOWSKY, G. B., and ROBERTS, R. M., *J. Am. Chem. Soc.*, 1941, **63**, 2080.
59. HERRIOTT, R. M., unpublished results.
60. EDELHOCH, H., *J. Am. Chem. Soc.*, 1956, **78**, 2644.
61. JOHNSON, E. W., and NASH, L. K., *J. Am. Chem. Soc.*, 1950, **72**, 547.
62. STEINHARDT, J., and ZAISER, E. M., *Advances Protein Chem.*, 1955, **10**, 151.
63. BLUMENFELD, O., and PERLMANN, G., *J. Gen. Physiol.*, 1959, **42**, 563.
64. WETLAUFFER, D. B., EDSALL, J. T., and HOLLINGSWORTH, B. R., *J. Biol. Chem.*, 1958, **233**, 1421.
65. DIEU, H., *Bull. Soc. chim. belg.*, 1956, **65**, 603.
66. MEHL, J. W., ONCLEY, J. L., and SIMHA, R., *Science*, 1940, **92**, 132.
67. JURGENSONS, B., *Arch. Biochem. and Biophysics*, 1952, **41**, 333.
68. KAUZMANN, W., and SIMPSON, R. B., *J. Am. Chem. Soc.*, 1953, **75**, 5154.

69. HERRIOTT, R. M., *J. Gen. Physiol.*, 1935, **19**, 283.
70. HERRIOTT, R. M., and NORTHROP, J. H., *J. Gen. Physiol.*, 1934, **18**, 35.
71. HOLLANDER, V., *Proc. Soc. Exp. Biol. and Med.*, 1943, **53**, 179.
72. HERRIOTT, R. M., *J. Gen. Physiol.*, 1937, **20**, 335.
73. HERRIOTT, R. M., *J. Gen. Physiol.*, 1941, **25**, 185.
74. HERRIOTT, R. M., *J. Gen. Physiol.*, 1947, **31**, 19.
75. PHILPOT, J. ST. L., and SMALL, P. A., *Biochem. J.*, 1938, **32**, 534.
76. HERRIOTT, R. M., in *Chemical Basis of Heredity*, (W. D. McElroy and H. B. GLASS, editors), Baltimore, The Johns Hopkins Press, 1954, 24.
77. HERRIOTT, R. M., *J. Cell Comp. Physiol.*, 1956, **47**, suppl. 1, 239.
78. AMBROSE, J. F., KISTIAKOWSKY, G. B., and KRIDL, A. G., *J. Am. Chem. Soc.*, 1951, **73**, 1232.
79. VELICK, S. F., HAYES, J. E., and HARTING, J., *J. Biol. Chem.*, 1953, **203**, 527.
80. HAYES, J. E., and VELICK, S. F., *J. Biol. Chem.*, 1954, **207**, 225.
81. BOTTS, J., and MORALES, M., *Tr. Faraday Soc.*, 1953, **49**, 696.
82. FLAVIN, M., *J. Biol. Chem.*, 1954, **210**, 771.
83. PAULING, L., *Chem. Eng. News*, 1946, **24**, 1375.
84. SWAIN, C. G., and BROWN, J. F., *J. Am. Chem. Soc.*, 1952, **74**, 2538.
85. PERLMANN, E. G., *Nature*, 1954, **173**, 406.
86. FUNATSU, M., and TOKUJASEU, K., *Proc. Japan Acad.*, 1959, **35**, 139.
87. HERRIOTT, R. M., *J. Gen. Physiol.*, 1941, **24**, 325.
88. VAN VUNAKIS, H., and HERRIOTT, R. M., *Biochim. et Biophysica Acta*, 1956, **22**, 537.
89. KATCHALSKI, E., BERGER, A., and NEUMANN, H., *Nature*, 1954, **173**, 998.
90. NORTHROP, J. H., *Harvey Lectures*, 1925-26, **21**, 36.
91. FRUTON, J. S., and BERGMANN, M., *J. Biol. Chem.*, 1939, **127**, 627.
92. HARRINGTON, C. R., and PITT-RIVERS, R. V., *Biochem. J.*, 1944, **38**, 417.
93. BAKER, L. E., *J. Biol. Chem.*, 1951, **193**, 809.
94. SANGER, F., and TUFFY, H., *Biochem. J.*, 1951, **49**, 481.
95. SANGER, F., and THOMPSON, E. O. P., *Biochem. J.*, 1953, **53**, 336.
96. BAILEY, J. L., MOORE, S., and STEIN, W. H., *J. Biol. Chem.*, 1956, **221**, 143.
97. LI, C. H., *Advances Protein Chem.*, 1956, **11**, 101.
98. PHILPOT, J. ST. L., and ERICKSSON-QUENSEL, I. B., *Nature*, 1933, **132**, 932.
99. KROMAN, M. J., and STERN, M. D., *J. Physic. Chem.*, 1955, **59**, 969.
100. YASNOFF, D. S., and BULL, H. B., *J. Biol. Chem.*, 1953, **200**, 619
101. BAKER, L. E., *J. Biol. Chem.*, 1954, **211**, 701.