Iowa State University
Digital Repository

Retrospective Theses and Dissertations | Iowa State University Capstones, Theses and Dissertations

1953

Sequences of amino acid residues in insulin

Donald J. Blaney
Iowa State College

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Biochemistry Commons

Recommended Citation
Blaney, Donald J., "Sequences of amino acid residues in insulin " (1953). *Retrospective Theses and Dissertations*. 12403.
https://lib.dr.iastate.edu/rtd/12403

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

المنارة للاستشارات

www.manaraa.com

NOTE TO USERS

This reproduction is the best copy available.

UMI®

SEQUENCES OF AMINO ACID
RESIDUES IN INSULIN

by

Donald J. Blaney

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Biochemistry

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College

1953

UMI Number: DP11802

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

UMI Microform DP11802

Copyright 2005 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

QP801.A5
B61s
c.1

TABLE OF CONTENTS

Page

I. INTRODUCTION 1

II. HISTORICAL 3

A. Purity of Insulin. 3
B. Molecular Weight 4
C. Amino Acid Composition 5
D. Sequence of Amino Acid Residues. 8

III. METHODS. 17

A. The Amino Acid Residue Sequence Procedure. 17
B. Conditions for Reaction of Amino Acids and Proteins with Phenylisothiocyanate . 17
C. Formation of Substituted Hydantoins from Thiocarbamyl Derivatives 19
D. Microbiological Assay Procedure. 20
E. Ninhydrin Procedure for the Determination of Amino Acids 23

IV. EXPERIMENTAL AND RESULTS 25

A. Sequence of Amino Acid Residues in Insulin. 25

1. Electrophoresis of the insulin sample 25
2. Moisture determination 26
3. Microbiological determination of certain amino acids before and after phenylisothiocyanate treatment of insulin. 27
4. Repetitions of the residue sequence procedure. 30
5. Attempted extraction of the phenylthiohydantoins formed from the terminal amino acids of insulin. 32

T10694

Page

6. Ether extraction of phenylthiohydantoins derived from amino acids of insulin. 33

7. Attempted extraction of the hydantoins of the terminal amino acids of insulin 34

B. Formation and Properties of Substituted Hydantoins. 35

1. Solubility of phenylthiohydantoins in various solvents 35

2. Other substituted hydantoins. 37

3. Hydrolysis of phenylthiohydantoins. . . 37

4. Estimation of phenylthiohydantoins without hydrolysis. 39

5. Effect of pH on the extraction of phenylthiohydantoins from buffered protein solutions. 40

6. Extent of the reaction of phenylisothiocyanate with leucine and DL-leucyl-DL-valine 41

V. DISCUSSION. 44

A. Sources of Error in the Sequence Procedure 44

B. Sequence Data and Its Comparison to Literature Reports. 48

C. Identification of the Hydantoin as a Sequence Procedure. 54

VI. CONCLUSIONS 56

VII. SUMMARY 57

VIII. ACKNOWLEDGEMENTS. 58

IX. BIBLIOGRAPHY. 59

I. INTRODUCTION

It seems probable that a better understanding of the mode of physiological action of proteins and peptides will be possible when their chemical structure is known. That this is well recognized is evidenced by the repeated attempts to develop methods for establishing the relation of one amino acid to another in these complex substances. One such method, originally employed by Abderhalden and Stix (1) and recently improved and extended by Sanger (2), has made possible a proposed sequence for all the amino acid residues in insulin. A structure for only one other protein, clupein, has been suggested (3), but the residue sequence was based on an hypothesis shown to be untenable (4).

There have appeared a number of reports of the chemical properties of insulin that cannot be readily correlated with the structure advanced by Sanger. Because of this, and because of the desirability of investigating the structure by an independent method, the present work was undertaken.

It was believed that some information could be obtained by means of an amino acid residue sequence method previously developed in these laboratories (5). In addition, it was hoped that a procedure more direct and rapid than either this or that of Sanger could be evolved. In this connection, the

reaction between an isocyanate and the free amino groups of amino acids and proteins was investigated, since previous work in these laboratories indicated that phenylisothiocyanate can be made to react with proteins in a stepwise fashion, one amino acid residue being removed in each successive step.

II. HISTORICAL

A. Purity of Insulin

With the crystallization of bovine insulin by Abel (6) in 1926, the hormone became available for the first time in a high state of purity. Investigations of crystalline samples by modern techniques have generally indicated complete homogeneity. Electrophoresis, ultracentrifuge, and solubility curve measurements gave evidence for no more than a single substance (7), nor could insulin be resolved into more than one component by chromatographic procedures (8). However, countercurrent distribution studies, in which 900 or more transfers were made, produced two biologically active crystalline proteins from several samples of crystalline pork, beef, and sheep insulin (9). The major component in each species represented from 50% to 90% of the sample (10) and was usually present in twice the amount of the other biologically active protein. There was evidence also for small amounts of material other than the two principal constituents.

If this investigation can be taken as proof that there exist more than one molecular species in crystalline insulin, then structural work to date must be viewed with considerable caution. The questions arise whether established amino acid

sequences have all originated from the same molecule, and if so, whether the major or minor component of insulin has been studied.

B. Molecular Weight

The molecular weight of insulin has been extensively investigated. Through the use of the ultracentrifuge, Sjögren and Svedberg (11) obtained a value of 35,000. Crystallographic work indicates that three identical units of molecular weight 12,000 aggregate to form a trimer of molecular weight 36,000 in the solid state (12). Gutfreund (13) has also presented evidence that insulin has a minimum molecular weight of 12,000 and that, in aqueous solution, this monomer may aggregate into tri- or tetrameric moieties depending on the pH and protein concentration. On the basis of ultracentrifuge and light scattering experiments, Tietze and Neurath (14) are in accord with a minimum value of 12,000, but Harfenist and Craig (10) have suggested a minimum weight of 6,500 for bovine insulin purified by countercurrent distribution. Their method for molecular weight determination consisted in treating insulin with 2,4-dinitrofluorobenzene for fifteen minutes. A product believed to represent insulin substituted with only one dinitrophenyl group per molecule of insulin was isolated by countercurrent

distribution. Colorimetric estimation of the amount of dinitrophenyl groups in a given weight of the derivative provided the necessary data for calculation of the molecular weight.

C. Amino Acid Composition

No prosthetic group has been found in insulin. With the exception of a small and variable amount of zinc, crystalline samples appear to be composed entirely of amino acids, all except methionine, tryptophan, and hydroxyproline being present. Reports on the amino acid composition have come largely from three laboratories. Fromageot (15) analyzed a sample of crystalline bovine insulin by means of a combination of quantitative paper chromatography and individual chemical determinations for some of the amino acids. Tristram (16) has summarized the results from Chibnall's laboratory, where most of the values were obtained by titration of the acetyl derivatives of the amino acids separated on silica gel columns. Microbiological assay of the amino acids of insulin was carried out by Brand (17). The results of these investigations are listed in Table 1. For comparison, the numbers of residues of each amino acid in the structure proposed by Sanger (18,19) are included.

Table 1

The Amino Acid Composition of Insulin

(Results expressed as number of residues per molecule of insulin, M.W. = 12,000)

Amino acid	Tristram (16)	Brand (17)	Fromageot (15)	Sanger (18,19)
glycine	7	7	7	8
alanine	6	-	6	6
valine	8	9	8	10
leucine	12	12	13	12
isoleucine	3	3	2	2
proline	3	3	2	2
phenylalanine	6	6	6	6
half-cystine	12	11	12	12
arginine	2	2	2	2
histidine	4	4	4	4
lysine	2	2	2	2
aspartic acid	6	6	5	6
glutamic acid	15	16	16	14
serine	6	6	6	6
threonine	2	3	2	2
tyrosine	9	8	8	8

The data of Tristram, Fromageot, and Sanger were obtained with samples of bovine insulin. Brand does not mention the species from which his sample was obtained. Harfenist and Craig (20) examined the amino acid content of the principal component obtained by countercurrent distribution of beef, pork, and sheep insulins and found that while most of the amino acids were present in identical amounts in the three species, there were significant variations in the quantities of six amino acids. Their results are presented in Table 2.

Table 2

Variation in Amino Acid Composition of Insulin Obtained from Different Species

(Results expressed as number of residues per molecule of insulin, M.W. = 12,000)[a]

Amino Acid	Beef	Pork	Sheep
isoleucine	2	4	2
valine	10	8	10
glycine	8	8	10
alanine	6	4	6
threonine	2	4	2
serine	6	6	4

[a]The results of Harfenist and Craig (20) are rounded to the nearest whole number and converted to M.W. = 12,000.

Sanger (21) had previously published results of a "semi-quantitative" analysis of four amino acids in pig, ox, and sheep insulin. His experiments, while less precise than those summarized in Table 2, indicated similar differences. It seems obvious that at least small variations in the arrangement of amino acid residues are present in insulins obtained from different species, but as yet no great differences in physical or physiological properties have been detected.

D. Sequence of Amino Acid Residues

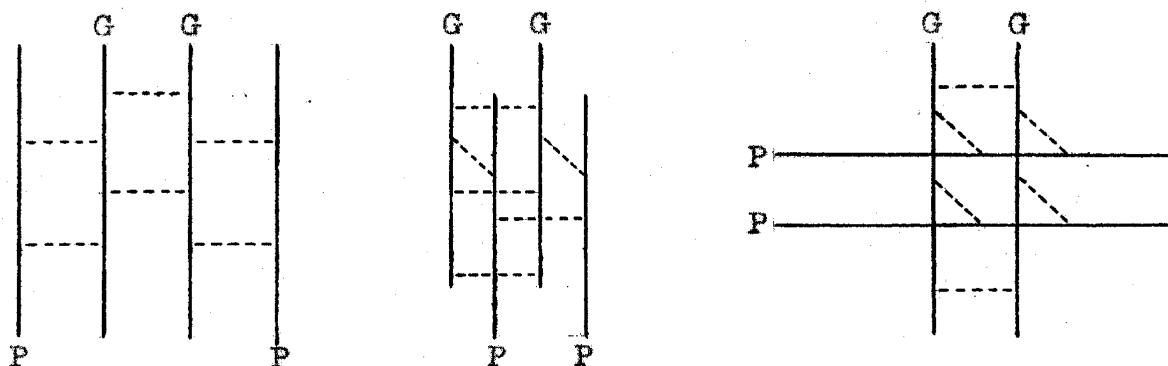
The problem of the sequence of the amino acid residues was attacked by Sanger (2) in 1945 with the aid of an improved reagent which reacts with free amino groups. Eliminating hydrogen fluoride, 2,4-dinitrofluorobenzene reacts rapidly in mildly alkaline solutions with the free amino, hydroxyl, phenolic, thiol, and imidazole groups of proteins to yield arylation products that are resistant to hydrolysis under conditions which result in hydrolysis of peptide bonds. In contrast to the free amino acids of the hydrolysate, those acids that have been substituted on the α-amino group are ether-soluble and highly colored. They can be extracted, separated chromatographically, and estimated colorimetrically. By this procedure Sanger found that in a molecule of insulin

(M.W. = 12,000), there were six free amino groups. Two of these were the amino groups of glycine, two the groups of phenylalanine, and two the ε-amino groups of lysine. It was concluded that there were four peptide chains in insulin, two of these having glycine as aminoid terminus (N-terminal) and two having phenylalanine.

Reasoning that the peptide chains might be held together by the dithio linkage of cystine, Sanger attempted to separate the chains by oxidizing the -S-S- bond with performic acid (22). He succeeded in isolating several fractions, two of which A and B, were believed to represent the entire insulin molecule. The yield of fraction A was greater than 25% of the starting material, of fraction B, somewhat less. Each of the two fractions was individually submitted to the 2,4-dinitrofluorobenzene reagent, and each product was partially hydrolyzed with concentrated hydrochloric acid for eight days at 37°. This treatment afforded a number of low molecular weight dinitrophenyl-peptides which could be separated chromatographically. The structures of the peptides were studied individually. Sanger was able to show that fraction A contained the pentapeptide glycyl-isoleucyl-valyl-glutamyl-glutamic acid, while fraction B contained the following sequences: phenylalanyl-valyl-aspartyl-glutamic acid and threonyl-prolyl-lysyl-alanine. Since no peptides were obtained which did not fit into these sequences, he concluded

that "the insulin molecule is built up of two pairs of very similar, if not identical, polypeptide chains" (22).

In addition to dinitrophenyl-peptides, partial hydrolysis of the insulin fractions A and B yielded a number of other peptides which were separated by means of paper chromatography, ionophoresis, and ion-exchange columns (23,24). The structures of many of these, as well as some additional peptides obtained by enzymic digestion of the insulin fractions (18, 19), were examined. The tedious investigations enabled Sanger to conclude that the insulin molecule was made up of four open peptide chains which were held together by the dithio bonds of the cystine residues. Two of the chains were reported to have the structure phe, val, asp, glu, his, leu, cys, gly, ser, his, leu, val, glu, ala, leu, tyr, leu, val, cys, gly, glu, arg, gly, phe, phe, tyr, thr, pro, lys, ala. For the other two chains, also identical, the reported sequence was gly, iso, val, glu, glu, cys, cys, ala, ser, val, cys, ser, leu, tyr, glu, leu, glu, asp, tyr, cys, asp. An examination of the order in either fraction reveals no simple periodic arrangement of the amino acids. On the basis of the cystine distribution in the four chains of insulin, Sanger has proposed possible configurations for the molecule (25). These are illustrated in Figure 1.



G represents a glycyl chain.

P represents a phenylalanine chain.

••••• represents -S-S- bond of cystine.

Figure 1

Proposed Configurations for Insulin

There has been a reluctance on the part of some investigators to accept Sanger's conclusions, and in fact there have been a number of reports which actually contradict his findings. Bull (26) has suggested that the enzymes Sanger used to obtain certain peptides may have caused some rearrangement in amino acid sequence. He also feels that a more precise determination of the molecular weight of the fractions as well as their isolation in a purer form is desirable. There is also the difficulty that some dinitrophenyl-amino acids are rather easily hydrolyzed (27), and

the effect of performic acid oxidation on a native protein is not completely understood. Some peptides obtained in the partial hydrolysis of fraction B were not examined, and three small peptides did not fit into the proposed sequence (23). Probably the greatest reason for questioning the validity of Sanger's conclusions is the fact that there is no assurance that all of the insulin molecule has been investigated. In the performic acid oxidation two fractions, labeled X and M and representing 20 to 30% of the oxidized insulin, were not studied for amino acid sequences. Sanger admits that these fractions may "contain another chain of the insulin molecule" (22).

Contrary to Sanger's findings, Udenfriend and Velick (28) detected only one terminal glycine and one phenylalanine molecule per molecule of insulin. They obtained their results by treating insulin with I^{131}-labeled p-iodophenylsulfonyl chloride which reacts with free amino groups to form a derivative which upon hydrolysis of the protein can be isolated and estimated. These authors suggest that their lower results may be due to incomplete reaction of p-iodophenylsulfonyl chloride with the protein.

The Fraenkel-Conrats (29,30) treated insulin with phenylisothiocyanate in aqueous pyridine. The thiocarbamyl derivative that was formed was cleaved in 1 N hydrochloric

acid at 36° to yield a phenylthiohydantoin derived from the terminal amino acid. The authors claim that the residual protein remains intact, no additional peptide bonds being broken by the acid treatment. The hydantoins are extracted with ether and estimated by means of their optical density at 268 mu. Three to 3.5 moles of amino acids per mole of insulin were found in the terminal position. The extracted hydantoins were shown to have been derived from glycine and phenylalanine and, unexpectedly, a small amount from alanine.

A similar experiment wherein the cleavage was carried out in 0.1 N hydrochloric acid at 75° yielded only glycine and phenylalanine, four moles per mole of insulin (31). The phenylisothiocyanate treatment and acid cleavage were applied five times successively to an insulin sample. After each acid treatment the hydantoins were extracted, estimated, and identified. In all cases the amino acids predicted by Sanger's structure were encountered, but in addition after the first acid cleavage other amino acids and more than six moles of hydantoin per mole of insulin were obtained. These results could be explained on the basis of Sanger's structure if it is assumed that interior peptide bonds are hydrolyzed.

Investigating the carboxyl ends of the insulin chains, Fromageot and co-workers (32) found that in one mole of

insulin there were two moles of alanine and two moles of glycine having free carboxyl groups. Their procedure consisted in reducing the free acid groups with lithium aluminum hydride, hydrolyzing the product, and separating and estimating the alcohols that are liberated. While the results are in accord with Sanger's suggestion of four chains, two being terminated at the carboxyl end with alanine, the appearance of two moles of glycine cannot be correlated with the structure proposed by Sanger.

Glycine was also found as carboxoid terminus by Chibnall and Rees (33) who esterified and then reduced with lithium borohydride the free carboxyl groups of insulin. The reduced protein was hydrolyzed and the alcohols derived from alanine (two moles), glycine (one mole), and one unidentified amino acid were isolated.

Using the enzyme carboxypeptidase, Lens (34) found that after 2.6 amino groups had been liberated as determined by the Van Slyke procedure, only alanine could be detected. After six hours 3.16 amino groups had been liberated and there could be detected glycine, valine, leucine or iso-leucine, and traces of tyrosine and either aspartic or glutamic acid. The author concluded that at least one and possibly three of the chains of insulin were terminated with alanine residues.

Harris (35) reinvestigated the action of carboxypeptidase on insulin and on the individual fractions, A and B, obtained by performic acid cleavage. After two minutes digestion, alanine, aspartic acid, and asparagine could be detected, and after eight hours lysine, glutamic acid, tyrosine, and leucine could also be found. Alanine was liberated eight times more rapidly than asparagine. Enzymic action on fractions A and B confirmed Sanger's observation that two moles of alanine and two moles of asparagine were terminal.

Two groups of workers have attempted to determine the nature of the carboxyl ends of the insulin chains by reaction of the acetylated protein with ammonium thiocyanate. In the presence of acetic anhydride, the amino acids bearing free α-carboxyl groups are involved in a cyclization to form a thiohydantoin. When severed from the remainder of the peptide chain by acid or alkali, the hydantoin can be recovered and identified. Waley and Watson (36) found alanine and a trace of aspartic acid, while Baptist and Bull (37) found alanine, tyrosine, phenylalanine, and traces of glycine in both pork and beef insulin. Since only alanine and asparagine should be detected according to Sanger's structure, Baptist and Bull suggest that it is conceivable that tyrosine and phenylalanine may occupy terminal positions in

impurities common to both pork and beef insulin. The authors state that aspartic acid in terminal position cannot be detected by their method.

III. METHODS

A. The Amino Acid Residue Sequence Procedure

The method developed by Fox, et al. (5), consists in the following steps.

1) The amino acids in an hydrolysate of the peptide in question are determined microbiologically.

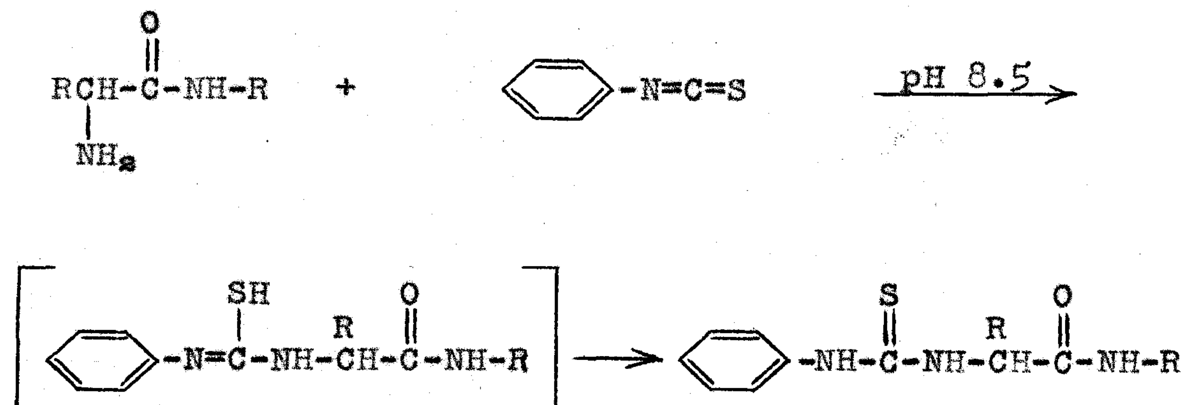
2) The amino acid bearing a free α-amino group (aminoid terminal amino acid) is made nutritionally unavailable to microorganisms by virtue of its reaction with phenylisothiocyanate. Acidic conditions convert the derivatized amino acid to a phenylthiohydantoin.

3) Microbiological determination of the amino acids in an hydrolysate of the phenylisothiocyanate-treated peptide reveals a reduction in the amount of amino acid that has reacted with the blocking agent.

B. Conditions for Reaction of Amino Acids and Proteins with Phenylisothiocyanate

The procedure was essentially that of Edman (38). The amino compound was dissolved in 50% aqueous pyridine; usually the concentration was about one mg. per ml. A small amount of bromthymol blue and enough 0.1 N sodium hydroxide solution

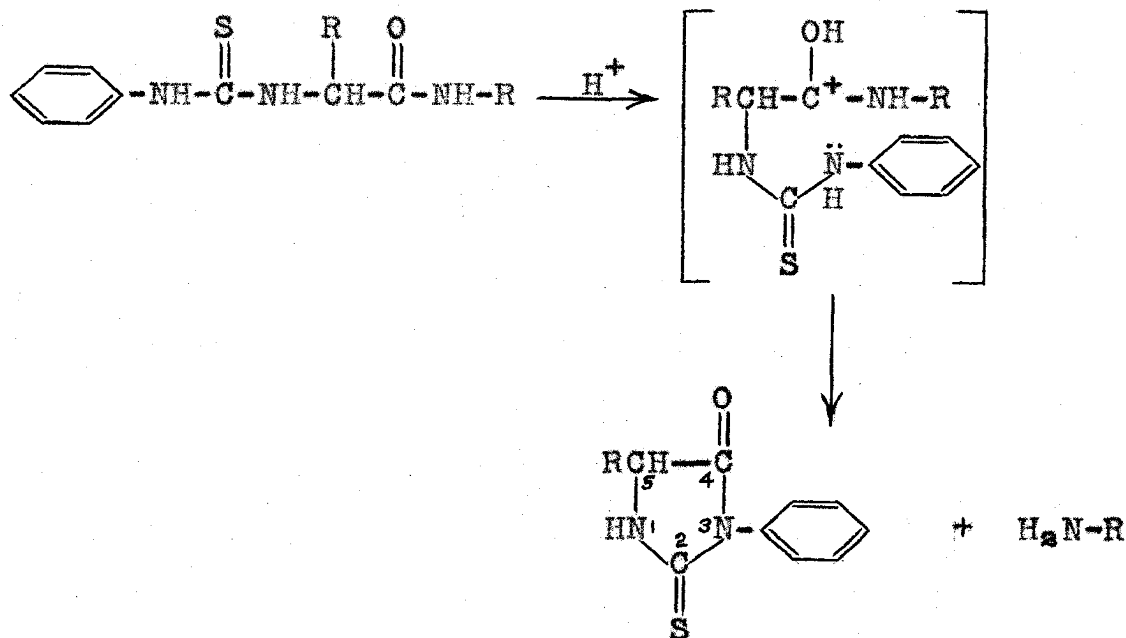
to produce a dark emerald green were added. The pH was approximately 8.5. Phenylisothiocyanate in the amount of twenty moles per mole of amino group was shaken with the solution until all dissolved. It was sometimes necessary to add additional pyridine*. The reaction was allowed to proceed four to six hours at 37° with additional 0.1 N sodium hydroxide solution being added periodically to maintain a constant pH. The product is a substituted thiourea or thiocarbamyl derivative of the amine.



*An alternate procedure consisted in adding the required amount of phenylisothiocyanate dissolved in pyridine. An homogeneous solution was effected immediately and reaction was more rapid.

C. Formation of Substituted Hydantoins from Thiocarbamyl Derivatives

Acids catalyze the cyclization of thiocarbamyl amino acids and peptides to yield substituted hydantoins (38,39).



In the present experiments, several methods were used to effect this reaction. When, in addition to forming the hydantoin, it was desired to hydrolyze the entire protein molecule, the thiocarbamyl derivative was autoclaved at fifteen pounds pressure for sixteen hours with 6 N hydrochloric acid. When it was desired to form the hydantoin while not causing hydrolysis of any interior peptide bonds,

the thiocarbamyl derivative was treated with dry dioxane saturated with gaseous hydrochloric acid. The reaction was carried out at room temperature in a desiccator containing calcium chloride and gaseous hydrochloric acid (5). When thiocarbamyl-amino acids were to be converted to hydantoins, some additional acidic conditions were employed.

D. Microbiological Assay Procedure

For the determination of the amino acids of interest in this work, four microorganisms were used. Lactobacillus arabinosus 17-5 (ATCC No. 8014), Lactobacillus brevis (ATCC No. 8287), and Streptococcus faecalis R (ATCC No. 8043) were obtained from the American Type Culture Collection. Leuconostoc mesenteroides P-60 (ATCC No. 8042) was obtained from the bacteriology department of Iowa State College. Stab cultures of these microorganisms were maintained in nutritionally complete agar medium, with transfers to fresh medium being made every two weeks. When an organism was to be used for inoculation, a sample was transferred to a complete medium containing no agar and was incubated approximately sixteen hours at 37°. The cells were then centrifuged, washed free of the medium with 0.9% sodium chloride, and finally suspended in 0.9% sodium chloride.

For the determination of all the amino acids the procedure of Kuiken, et al. (40), was used with the following exceptions. Two and one half ml. of the assay medium were used in each of the tubes containing graded amounts of the particular amino acid to be assayed and in the tubes containing the unknown. All solutions were made to a total volume of five ml., and after inoculation with the appropriate microorganism the tubes were incubated at 37° for seventy-two hours or, in the case of glycine, forty-eight hours. The growth response was then measured by titrating the lactic acid produced in the entire solution with 0.05 N sodium hydroxide to a pH of 6.8 as indicated by a Beckman pH meter or by means of bromthymol blue indicator. The standard curve was constructed from the titration values of duplicate tubes which usually contained 0.0, 2.0, 4.0, 6.0, 8.0, 10.0, 14.0 and 20.0 ug of amino acid. The unknown samples were analyzed at two levels and in duplicate.

The assay medium varied from that of Kuiken in that 400 mg. of glycine and 0.2 mg. of folic acid were added per liter, and the tomato eluate was omitted. Forty gm. of glucose per liter of medium were normally present, but when Lactobacillus brevis was used, this was replaced by 10 gm. of arabinose and 30 gm. of glucose. When glutamic acid was to be determined, the aspartic acid content of the medium was reduced to 200 mg. per liter, and 200 mg. of L-asparagine and

50 mg. of glutamine were added. Four times the usual amount of phosphate buffer was used when aspartic acid was assayed, and the glutamic acid was reduced to 200 mg. per liter of medium.

The enumerated modifications of Kuiken's procedure are changes which have been used in these laboratories and have been found to be most consistent in giving satisfactory results.

The organisms used for assay of a particular amino acid are recorded in Table 3.

Table 3

Organisms Used for Assay of Various Amino Acids

Amino Acid	Microorganism
isoleucine	*Lactobacillus arabinosus* 17-5
glutamic acid	*Lactobacillus arabinosus* 17-5
phenylalanine	*Lactobacillus brevis*
glycine	*Lactobacillus brevis*
aspartic acid	*Leuconostoc mesenteroides* P-60
valine	*Streptococcus faecalis* R

E. Ninhydrin Procedure for the Determination of Amino Acids

The method used was similar to that of Fowden (41), but a standard curve was run in conjunction with each determination. This practice makes it necessary that the amino acid to be determined be separated completely from all others on a one-dimensional paper chromatogram. The method has the advantage of being less susceptible to errors due to variations in the quality of the ninhydrin reagent, the temperature and time used to develop the color, the filter paper used, and the distance traveled by the amino acid. When few amino acids were present, a solvent such as the butanol, acetic acid, water solution of Partridge (42) was satisfactory. Solvent systems suggested by McFarren (43) were more suitable for complex mixtures of amino acids.

The standard curve was prepared by spotting in duplicate 5, 10, and 20 µg. of the amino acid on a 7.5 x 22 inch Whatman No. 1 filter paper sheet. The amino acids were applied 2.5 inches from the narrow edge of the paper by means of a 5 µl. pipet. Solutions of the amino acids were prepared such that the desired amount of the acid could be spotted with one 5 µl. application. The solution to be assayed was also spotted in duplicate on the same sheet.

Chromatograms were developed by the descending technique in a large glass jar.

Fowden (41) located the amino acids on the paper sheet by means of their fluorescence after the papergrams were heated in an oven at 100°. Because this method fails completely for low concentrations of an amino acid and did not permit accurate outlining of spots containing larger amounts, it was found preferable to locate the amino acid by spraying the chromatogram with dilute ninhydrin solution (0.05% ninhydrin in ethanol).

The paper squares were cut out, ammonia removed, and color developed according to Fowden's procedure. The solutions were made to a final volume of 10 ml., and the optical densities at 570 mµ were measured in a Beckman DU spectrophotometer.

IV. EXPERIMENTAL AND RESULTS

A. Sequence of Amino Acid Residues in Insulin

This problem was approached in two ways: 1) by the procedure of Fox, *et al.* (5), outlined in the Methods Section, p. 17, and 2) by attempts to quantitatively remove and estimate the terminal amino acids derivatized with phenylisothiocyanate.

1. Electrophoresis of the insulin sample

The insulin sample used in all experiments described in this thesis was obtained from Armour and Co., Chicago, Illinois. It was crystalline beef insulin, sample 473-53AB, and had an activity of 26 units per mg. In order to acquire additional information regarding the purity of the protein, two electrophoretic patterns were obtained. A 30 mg. sample was dissolved in 15 ml. of glycine hydrochloride buffer of pH 2.7. Enough sodium chloride was added to make the total ionic strength of the solution 0.1 μ. Under these conditions the insulin sample moved as a single sharp peak in both the ascending and descending limbs of the electrophoresis cell. The same 30 mg. sample was also studied in

acetate buffer of pH 4.0 and a low but undetermined ionic strength. The protein concentration was again 0.2%. The peak in the electrophoretic pattern was not as sharp under these conditions, but again there was no indication of more than one component in the sample.

Before being used for the electrophoretic studies, the 30 mg. insulin sample was used for a moisture determination (Experiment 2). Phosphorous pentoxide drying appeared to have no disruptive effect on the homogeneity of the sample.

2. Moisture determination

Because a limited amount of insulin was available and because the dry protein is quite hygroscopic (44.), the moisture determination was performed in the following manner. A small weighing bottle was dried in a desiccator over phosphorous pentoxide. At two and four minutes after being removed from the desiccator, the bottle was weighed on a microbalance. A graph was constructed with time as abscissa and weight as ordinate. By extrapolating to zero a line connecting the weights at two and four minutes, it was possible to estimate the weight of the bottle in the desiccator. A 31.828 mg. sample of insulin was then weighed in the flask, which had been allowed to come to equilibrium with the balance and surroundings. After being dried six days

in a desiccator over phosphorous pentoxide, the insulin sample was removed, weighed at two and four minutes, and the weight at zero time was calculated as before. The loss in weight, 2.403 mg., represents a moisture content of 7.5%.

In order to maintain a constant moisture content, the insulin sample was kept in a tightly closed vial which in turn was kept within another tightly closed container.

3. Microbiological determination of certain amino acids before and after phenylisothiocyanate treatment of insulin

Five samples of insulin were weighed on a microbalance. Sample 0, 6.29 mg., was hydrolyzed with three ml. of 6 N hydrochloric acid in the autoclave at 15 pounds pressure for 16 hours. The acid was removed from the hydrolysate by concentration to dryness on a steam bath; the dry residue was taken up in 50 ml. of water, and the pH of the solution was adjusted to 6-7.

The other four samples, all weighing approximately one mg., were treated with phenylisothiocyanate as discussed in the section on Methods, p. 17. Samples 1, 2, 3, and 4 were given one, two, three, and four treatments respectively. After each phenylisothiocyanate reaction, except the last,

the derivatized protein was treated with dry hydrochloric acid-dioxane solution. Only the aminoid terminal peptide bond is susceptible to hydrolysis under these conditions. After each sample had received the last phenylisothiocyanate treatment, the entire protein was hydrolyzed with 6 N hydrochloric acid. The hydrolysate of each of the four samples was taken to dryness and then dissolved in 10 ml. of water. Phenylalanine, isoleucine, valine, aspartic acid, and glutamic acid were determined in the hydrolysate of Sample 0. Phenylalanine was determined in the hydrolysate of Sample 1, isoleucine and valine in Sample 2, valine and aspartic acid in Sample 3, and glutamic acid in Sample 4. These particular assays were performed because the structure advanced by Sanger (Historical, p. 10) indicates that these amino acids occupy positions near the aminoid terminus of the molecule. For example, phenylalanine and glycine are reported terminal; isoleucine and valine are in second positions.

As was frequently found to be the case during the course of these investigations, the microbiological assay of glycine was unsatisfactory. Values for glycine are therefore not included in Table 4, where are recorded the results of the assays. When duplicate determinations were not in close agreemant, an average was taken, but these

Table 4

Amino Acid Values Before and After Treatments
with Phenylisothiocyanate, First Experiment

	Sample 0		Sample 1		Sample 2		Sample 3		Sample 4	
	%[a]	res.[b]	%[a]	res.[b]	%[a]	res.[b]	%[a]	res.[b]	%[a]	res.[b]
phenylalanine	8.1	5.9	5.1	3.7	---	---	---	---	----	----
isoleucine	3.8	3.4	---	---	0.0	0.0	---	---	----	----
valine	10.1	10.4	---	---	6.8	7.0	4.8	4.9	----	----
aspartic acid	(6.9	6.2)	---	---	---	---	(6.2	5.5)	----	----
glutamic acid	(16.3	13.3)	---	---	---	---	---	---	(13.2	10.8)

[a]In gm./100 gm. protein.

[b]Number of amino acid residues per molecule of M.W. 12,000.

figures should be interpreted with caution. They are enclosed in parentheses. The calculations are made on a moisture free basis.

4. Repetitions of the residue sequence procedure

Attempts were made to substantiate the results obtained in the above experiment. In the first repetition similar quantities of insulin were taken, and the phenylisothiocyanate treatments and hydrolyses were carried out as before. In this experiment, considerable difficulties were encountered with the microbiological assays, and as a result there was collected only a limited amount of reliable data. The figures are recorded in Table 5.

Table 5

Amino Acid Values Before and After Treatments with Phenylisothiocyanate, Second Experiment

	Sample 0		Sample 1		Sample 2		Sample 3	
	%[a]	res.[b]	%[a]	res.[b]	%[a]	res.[b]	%[a]	res.[b]
phenylalanine	---	---	5.1	3.7	---	---	---	---
isoleucine	2.5	2.2	---	---	---	---	---	---
valine	9.2	9.5	---	---	5.9	6.1	5.3	5.4

[a]In gm./100 gm. protein.

[b]Number of amino acid residues per molecule of M.W. 12,000.

A second repetition of the sequence procedure was carried out as follows: A 21.82 mg. insulin sample was dissolved in dilute hydrochloric acid of pH 3, since the protein is not soluble between pH 4.3 and pH 7.8. Four 2.5 ml. aliquots (4.20 mg. of insulin each) were given zero, one, two, and three phenylisothiocyanate treatments as described previously. The dioxane-hydrochloric acid and 6 N hydrochloric acid treatments were as before. Since larger samples were used in this experiment, it was possible to assay a greater number of amino acids in each sample. The results are found in Table 6.

Table 6

Amino Acid Values Before and After Treatments with Phenylisothiocyanate, Third Experiment

	Sample 0		Sample 1		Sample 2		Sample 3	
	%[a]	res.[b]	%[a]	res.[b]	%[a]	res.[b]	%[a]	res.[b]
phenylalanine	7.2	5.3	5.4	3.9	5.5	4.0	5.2	3.8
isoleucine	3.2	3.0	3.2	3.0	0.6	0.6	0.6	0.6
valine	9.8	10.0	8.8	9.0	7.5	7.7	5.6	5.7
aspartic acid	5.1	4.6	4.0	3.6	4.4	4.0	2.7	2.5

[a]In gm./100 gm. protein.

[b]Number of amino acid residues per molecule of M.W. 12,000.

5. Attempted extraction of the phenylthiohydantoins formed from the terminal amino acids of insulin

An 11.9 mg. sample of insulin was dissolved in 4 ml. of 50 per cent aqueous pyridine and treated with phenylisothiocyanate in the usual manner. After eight hours, the reaction mixture was taken to dryness, and five ml. of dioxane-hydrochloric acid solution were added to form the hydantoin. After ten hours the suspension was again taken to dryness, and the dry residue was stirred four times with five ml. portions of methylene chloride. Each of the methylene chloride extracts was transferred by means of a filter stick directly to a small test tube, and the combined extracts were taken to dryness by evaporation. In order to regenerate the free amino acid from extracted hydantoin, 0.4 ml. of 0.25 N barium hydroxide solution was added to the tube. It was sealed and heated at 140° for 48 hours. At the end of this time the tube was opened, excess barium hydroxide was precipitated by a gentle stream of carbon dioxide, and paper chromatograms were run using 10 and 20 µl. of the hydrolysate. Spraying the chromatograms with ninhydrin revealed no trace of amino acids.

6. Ether extraction of phenylthiohydantoins derived from amino acids of insulin

The insulin sample used in the previous experiment was treated again with phenylisothiocyanate and dioxane-hydrochloric acid solution. After the suspension had been taken to dryness to remove dioxane, the residue was dissolved as well as possible in water and the solution was adjusted to pH 1. The solution was extracted repeatedly with ether at both this pH and after being made alkaline with 0.1 N sodium hydroxide solution. The concentrated ether extracts were taken to dryness by evaporation in a small test tube to which was added 0.4 ml. of 0.25 N barium hydroxide solution. The hydrolysis procedure and chromatography were carried out as in the previous experiment. With butanol/acetic acid/water (4/1/5) as developing solvent, the papergram showed five ninhydrin spots. The darkest three of these were found to have the same R_f values as glycine, valine and either or both phenylalanine and isoleucine. This solvent system does not separate phenylalanine and isoleucine. Two faint ninhydrin spots were located between the valine and glycine spots. When the barium hydroxide hydrolysate was chromatographed with a different solvent system, butanol/methylethylketone/water (4/4/1.5), the same

results were obtained. Neither of the faint spots had the R_f values of tyrosine or alanine, amino acids which could be expected to be located between valine and glycine on these chromatograms. As no more hydrolysate was available, it was impossible to identify the substances further, nor could it be established whether both phenylalanine and isoleucine were present in the fast moving spot. An approximate visual estimation of the amount of amino acids on the papergram indicated there was only 10-20 per cent the amount that should have been extracted if all reactions were quantitative.

7. <u>Attempted extraction of the hydantoins of the terminal amino acids of insulin</u>

Because the amino acids identified in the preceding experiment were derived from both the first and second positions in the insulin molecule, the experiment was repeated on a sample treated only once with phenylisothiocyanate. After dioxane-hydrochloric acid cleavage, fifteen ether extractions were carried out. The extracts were combined and evaporated as before, but no amino acids could be detected in the barium hydroxide hydrolysate of the residue remaining after evaporation.

B. Formation and Properties of Substituted Hydantoins

Both in this work and in reports in the literature (29,36) difficulties have been encountered in the attempt to quantitatively derivatize as the hydantoin, extract, and estimate the terminal amino acids of proteins. Because of this, it was decided to investigate the formation and properties of substituted hydantoins in simpler systems. Edman's (45) procedure for the preparation of hydantoins in 80-90 per cent yield was used to obtain products from a number of amino acids*.

1. Solubility of phenylthiohydantoins in various solvents

Estimations were made of the ether solubility of phenylthiohydantoins derived from eleven common amino acids. While most of the compounds appeared to dissolve eventually, it was generally true that they were not readily soluble. The hydantoin from glycine (3-phenyl-2-thiohydantoin) was one of the more insoluble compounds, while the hydantoin from valine (3-phenyl-5-isopropyl-2-thiohydantoin) seemed to be

---

*The preparation of these hydantoins was carried out by Robert S. Baker and William F. Serat.

typical of the more soluble compounds. Accordingly, these two substances were selected for solubility tests in various other solvents with a view toward finding a more effective extracting agent. The results are tabulated in Table 7.

Table 7

Solubility of Phenylthiohydantoins in Various Solvents

Solvent	Phenylthiohydantoin of Glycine	Phenylthiohydantoin of Valine
	(structure: H_2C–C(=O), HN, N–phenyl, C=S)	(structure: $(CH_3)_2CH$–C–C(=O), HN, N–phenyl, C=S)
ethyl ether	-	+
ethyl acetate	+	++
methylene chloride	-	++
skelly A	-	-
carbon disulfide	-	+
pyridine	++	++
benzene	-	+
carbon tetrachloride	-	+
amyl acetate	+	++
tributyl phosphate	+	++
triethanolamine	-	-
N-methylaniline	+	++
tributylamine	-	+
dl-β-phenylethylamine	++	++
tributyl citrate	-	+
dioxane	+	++
N,N-dimethylformamide	++	++

++....readily soluble
+.....slow or limited solubility
-.....insoluble or very slowly soluble

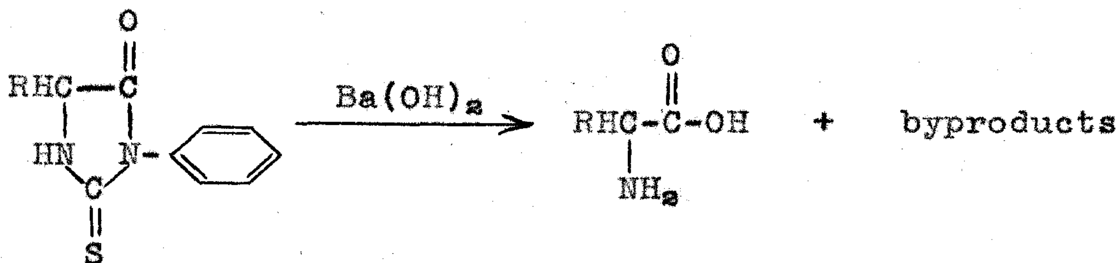
2. Other substituted hydantoins

In order to find a more soluble and easily extracted hydantoin, the reaction between amino acids and two isocyanates other than phenylisothiocyanate was investigated. Valine and leucine were treated with α-naphthyl isocyanate under the conditions and in the amounts used by Edman (45) in the preparation of phenylthiohydantoins. The reaction with α-naphthyl isocyanate was more rapid, as evidenced by immediate precipitation and liberation of hydrogen ions, but the products were more insoluble in ether than the corresponding phenylthiohydantoins. However, solubility in pyridine and particularly N,N-dimethylformamide was equally as good as that of the phenylthiohydantoins.

Octadecyl isocyanate reacted with leucine much more slowly than either phenylisothiocyanate or α-naphthyl isocyanate, so the reagent was not further investigated.

3. Hydrolysis of phenylthiohydantoins

One method for identifying and estimating the hydantoin removed from a peptide chain consists in recovering the free amino acid from the hydantoin by hydrolysis with barium hydroxide (36).



In order to determine the extent of this reaction, 9.7 mg. (0.05 millimole) of 2-phenyl-3-thiohydantoin (hydantoin from glycine) was placed in a small test tube with one ml. of 0.25 N barium hydroxide solution. The tube was sealed and heated 48 hours at 140°. At the end of this time, Ba^{++} was precipitated by means of a gentle stream of carbon dioxide. Glycine in 5 µl. aliquots of the neutralized hydrolysate was determined by means of the ninhydrin procedure described in the section on Methods, p. 23. The chromatographic solvent was butanol/acetic acid/water. According to this determination, the hydrolysis of the hydantoin proceeded to the extent of 76 per cent. Microbiological assay of a 5 µl aliquot of the hydrolysate indicated a somewhat greater hydrolysis, but not quantitative.

4. Estimation of phenylthiohydantoins without hydrolysis

Phenylthiohydantoins exhibit an absorption peak at 268 mμ and have an extinction coefficient of about 16,000 (30). If all other materials that absorb in this region of the ultraviolet are absent, it is possible to assay a solution of the hydantoin by determining its optical density in a spectrophotometer. For the preparation of a standard curve, the phenylthiohydantoin of leucine was dissolved in ether and five concentrations varying from 2 to 10 μg. per ml. were read in a Beckman DU spectrophotometer at 268 mμ. Over this range there was no variation from Beer's law. The hydantoin can be dissolved in ethanol, 50 per cent aqueous ethanol, and dry or water-saturated ether, and in all these solvents the extinction coefficient is approximately the same.

In order to determine whether pH has any effect on the position or height of the absorption peak, 50 per cent aqueous ethanol solutions of the phenylthiohydantoin derived from valine were adjusted to pH values of 1, 3, 5, 7, and 9. The absorption peak for all solutions was a wide band with a maximum near 268 mμ, and at all pH values the peaks were of nearly the same height. The peak in the solution of pH 9 was shifted very slightly toward the longer wavelengths.

5. Effect of pH on the extraction of phenylthiohydantoins from buffered protein solutions

It was of interest to determine whether a known amount of phenylthiohydantoin could be extracted quantitatively from a protein solution, since one report in the literature (36) suggests there is some interaction between hydantoins and proteins. At the same time the effect of pH on the extraction was studied.

McElvain's citrate-phosphate buffers of pH 2.2, 4.0, and 6.0, and Clark and Lubs phosphate buffer, pH 8.0, and borate buffer, pH 10.0, were prepared. None of these exhibited absorption at 268 mμ. Two mg. of bovine serum albumin and 0.2 mg. of 3-phenyl-5-isopropyl-2-thiohydantoin were thoroughly mixed in one ml. of water. To each of five such mixtures was added three ml. of one of the buffers. The solutions were then extracted ten times with equal volumes of ether, and the extracts were combined and diluted to 50 ml. A standard consisting of 0.2 mg. of the hydantoin in 50 ml. ether was prepared, and all solutions were read in the spectrophotometer at 268 mμ. The optical densities and percents recovery are recorded in Table 8.

Table 8

Recovery of Hydantoin from Buffered Protein Solutions

Solution	Optical Density	% Recovery
Standard	0.290	-
pH 2.2	.290	100
4.0	.295	101
6.0	.295	101
8.0	.290	100
10.0	.205	69

6. <u>Extent of the reaction of phenylisothiocyanate with leucine and DL-leucyl-DL-valine</u>

To each of two 0.476 mg. samples of <u>DL</u>-leucyl-<u>DL</u>-valine dissolved in 2 ml. aqueous pyridine was added 5 µl. of phenylisothiocyanate. The reaction was carried out in the usual manner for five hours. Pyridine and excess phenylisocyanate were extracted with cyclohexane because their presence interferes with the subsequent estimation of the hydantoin. After ten extractions, the odor of pyridine was still present so the last traces were removed by

evaporation of the thiocarbamyl solution to dryness. The dipeptide was cleaved with 6 N hydrochloric acid and the resulting solution was extracted ten times with ether. The amount of hydantoin in the extracts was calculated from the optical density of the ether solutions. Duplicate experiments appeared to indicate that the reaction proceeded to the extent of only 23 and 28 per cent.

In Table 9 are recorded results of experiments with a simple amino acid in place of the dipeptide.

Table 9

Extent of Formation of the Phenylthiohydantoin of Leucine

Reaction	Acid Reagent	No. of extractions[c]	Extent of rx.(%)
1	a	10	58
2	a	10	69
3	b	10	66
4	b	10	79
5	b	20	92

[a] 1 N HCl, autoclaved 6 hours at 15 lbs. pressure.

[b] 1 N HCl, 12 hours at 37°.

[c] Ether extraction of hydantoin formed after acid treatment.

One-tenth mg. of leucine was derivatized with 10 µl. of phenylisothiocyanate. In this case, however, the reaction was allowed to proceed 10-12 hours and the pyridine and excess reagent were removed by ether extractions. Two types of acid treatment were used to convert phenylthiocarbamyl-leucine into the phenylthiohydantoin. These are indicated in Table 9 along with the apparent extent of the reaction as determined by the optical density of ether extractions performed after acid treatment. The total volume of the reaction mixture was slightly greater than one ml., and the ether extractions were carried out with equal volumes.

The reaction was also carried out in a small all-glass continuous ether extractor. After the aqueous phase had been extracted seventy-two hours, only that amount of hydantoin was extracted which indicated a 55% reaction.

V. DISCUSSION

A. Sources of Error in the Sequence Procedure

It soon became apparent during the course of this investigation that the validity of conclusions concerning the sequence of amino acid residues in insulin would be limited in greatest part by the microbiological assays. The reproducibility (and probably accuracy) of the assays varied from one amino acid to another; in some cases the reproducibility was entirely satisfactory, in others it was quite poor. By inspection of the standard curves, the agreement between duplicate determinations, and the results of assays carried out at different levels of unknown, it was obvious that the values for phenylalanine, valine, and isoleucine were more trustworthy than those for glycine, and aspartic and glutamic acids. The accuracy of the values can be inferred from reports in the literature which indicate that microbiological assays can be expected to be reliable within five to ten per cent (46,47). This degree of accuracy has the most serious effect on amino acids present in large amounts. When there are more than ten to twelve residues of a particular amino acid in the protein under investigation, conclusions about its sequence position become less valid. For example, a ten per cent error in

the glutamic acid value (Table 4, p. 29) would alter by more than one the calculated number of residues, while the same error in the phenylalanine assay would produce no significant change in the number of residues. Although most of the assay values in the tables are probably more accurate than within ten per cent, interpretation of the data should be made with the possible deviations from true values clearly in mind.

On first consideration, the degree of accuracy attainable in microbiological assay may appear unduly poor unless one recalls the difficulties entailed in the determination of amino acids in protein hydrolysates. There are only two other commonly employed assay methods, quantitative paper chromatography (48) and the ion-exchange column method (49), that could be expected to give results of comparable accuracy. These methods, however, have not as yet been adapted to the determination of a large number of samples as must be performed in the sequence procedure employed here.

In this sequence method the designation of an amino acid as terminal depends upon detection of a substantial reduction in the original quantity of that acid after the protein has been treated first with the blocking agent, phenylisothiocyanate, and then with hydrochloric acid to effect hydrolysis. It is assumed that the terminal amino

acid residue is converted quantitatively, or nearly so, to a hydantoin or some derivative nutritionally unavailable to the microorganism used in assay. It may not be true that the reaction is quantitative, but in addition to published results indicating essentially quantitative reaction (5), there have been collected in these laboratories considerable data which are most logically interpreted on a basis of a quantitative reaction (50,51).

Perhaps a more serious question than the extent of the reaction is one concerning the nutritional value of the phenylthiohydantoins to the assay organism. Can growth be stimulated by the hydantoin as well as by the free amino acid? If this were found to be the case, then the conversion of a terminal amino acid residue to a hydantoin might be overlooked. There is indication that the phenylthiohydantoins derived from leucine and isoleucine are utilized in part by the organism when the assay tubes contain five per cent ethanol (52). That there is utilization under the usual conditions of the experiment is contraindicated by experiments in which a known amount of leucine was added to a peptide mixture. After phenylisothiocyanate and acid treatments, there was no growth stimulation due to the added leucine (53).

In order to locate the position of an amino acid residue that is other than terminal, the thiocarbamyl insulin was

treated with dry hydrochloric acid-dioxane solution. It is necessary that this reagent causes no appreciable hydrolysis of internal peptide bonds; otherwise newly liberated amino groups will be derivatized in the subsequent phenylisothiocyanate reaction and an amino acid may be assigned a position incorrectly near the aminoid terminus of the protein molecule. It is well known that certain peptide bonds are more labile than others, and it is conceivable that under the conditions of the experiment bonds involving certain amino acids (such as the amino group of serine or threonine) may be severed to some extent. However, since the sequence procedure commonly leads to whole number decrements in the number of amino acid residues before and after phenylisothiocyanate treatment, it appears that cleavage of internal peptide bonds is negligible. That some hydrolysis does occur might be construed from Experiment 6, p. 33. Ether extraction of insulin treated with phenylisothiocyanate after one hydrochloric acid-dioxane cleavage produced two faint ninhydrin spots in addition to spots expected on the basis of Sanger's structure. Since the faint spots are probably due to amino acids, their presence may be interpreted as being indicative either of partial hydrolysis of internal bonds or, alternatively, the presence of additional peptide chains in insulin. Probably no peptide bonds are cleaved during the phenylisothiocyanate treatment at pH 8.5 since Harrington and Neuberg (54)

showed that no irreversible change takes place in insulin subjected to pH variations from 2.5 to 11.4.

The possibility that all assay values recorded in the tables may be a few per cent low is suggested by the work of Miller and du Vigneaud (44). Only fifty per cent of the moisture in their sample of insulin was removed by a drying procedure similar to the one used in this investigation. With the exception of glutamic acid (since it is present in large amounts), a few per cent error in the moisture value would have negligible effect on most of the calculated residue numbers. Neither is there significant error introduced by failing to make a correction for ash content, since values for insulin are usually less than one per cent.

B. Sequence Data and Its Comparison to Literature Reports

In Table 10 are summarized the values found by various authors and in the present work for the number of residues of five amino acids in the insulin molecule of M.W. 12,000. The assay results are in quite close agreement with one another and with the number of residues demanded by Sanger's structure. The data of this thesis is slightly more in accord with Sanger's structure than is any of the previously reported assays. The number of isoleucine residues is not in agreement.

Table 10

The Number of Residues of Five Amino Acids in Insulin

Amino acid	Tristram (16)	Brand (17)	Fromageot (15)	Sanger[b]	Present work I	Present work II	Present work III	Average of I,II,III
phenylalanine	5.9	5.8	6.0	6	5.9	---	5.3	5.6
isoleucine	2.5	2.7	1.5	2	3.4	2.2	3.0	2.9
valine	7.9	9.0	7.6	10	10.4	9.5	10.0	10.0
aspartic acid	6.1	6.1	5.0	6	6.2	---	4.6	5.4
glutamic acid	15.2	16.4	16.0	14	13.3	---	---	13.3

[a]Calculated from data presented.

[b]Indicated by Sanger's structure (18,19), not an analysis.

After the terminal residues have been removed from the proposed structure, there are two less phenylalanine and two less glycine residues. All others remain the same. The results of the present work are in excellent agreement in the case of phenylalanine, but in addition one valine and one aspartic acid residue are lost (Table 6, p. 31). While it is entirely possible that the aspartic acid value is in error, the valine determination is generally reliable. The valine figure of nine residues was checked by a repeated assay. In this connection it is interesting to note that Chibnall (4) found by a Van Slyke determination six free amino groups in addition to the ε-amino group of lysine in an insulin molecule of M.W. 12,000. Sanger (2) suggests, however, that this value may be attributable to the fact that glycyl peptides give high values in the Van Slyke determination (55).

In subsequent determinations of the number of residues remaining after one or more treatments with phenylisothiocyanate, there was usually obtained rather close agreement with the residue numbers predicted by Sanger's structure. To illustrate this, the results of the assays, averaged and rounded to the nearest whole number, are tabulated in Table 11 along with the residue numbers of Sanger. The latter values are in parentheses.

Table 11

Number of Residues Before and After

Phenylisothiocyanate Treatments

Amino acid	Number of treatments				
	0	1	2	3	4
phenylalanine	6 (6)	4 (4)	4 (4)	4 (4)	- (4)
isoleucine	3 (2)	3 (2)	0-1 (0)	0-1 (0)	- (0)
valine	10 (10)	9 (10)	8 (8)	6 (6)	- (6)
aspartic acid	5 (6)	4 (6)	11 (6)	2 (4)	- (4)
glutamic acid	13 (14)	- (14)	- (14)	- (14)	11 (10)

A limitation in the sequence method employed in this investigation is clearly illustrated by the experiments with insulin, a protein of more than one peptide chain. While it can be established that both isoleucine and valine occupy second positions, it is obviously impossible to state which of these acids was joined to the terminal phenylalanine residue.

The more direct approach to sequence analysis which was attempted in Experiments 5, 6, and 7, p. 32-34, was not highly successful. It was certainly not quantitative. The phenylthiohydantoins of the terminal residues are either incompletely formed, or more probably, not easily extracted. Methylene chloride extraction (Experiment 5) and on one occasion fifteen ether extractions (Experiment 7) were unsuccessful in recovering the phenylthiohydantoins. However, in one case (Experiment 6) ether extraction removed the hydantoins derived from glycine, valine, and either or both phenylalanine and isoleucine. Since two phenylisothiocyanate reactions had been performed in this experiment, the extracted acids represent the terminal and second positions in insulin. While the sequence method furnished some evidence that aspartic acid was terminal in insulin (p. 50), the fact that it was not found in this experiment makes it appear doubtful that aspartic acid is terminal. That there

is an error in the aspartic acid assay value is the most probable explanation, although it is true that limited ether solubility of the phenylthiohydantoin of aspartic acid could account for its absence in the ether extracts.

Finally it should be stated that these attempts to interpret and correlate the data should be made with two points in mind. One of these is that there is a possibility that insulin samples obtained from different animals and prepared by different laboratories may have slightly different composition. Certainly there is species variation (20); there may actually be individual variation.

The other fact to be considered is that there is no assurance that only one molecular type is present in crystalline insulin. The investigations of Craig (9) make it appear that at least two molecules have been examined simultaneously in this and all other sequence work. If this is the case, it is impossible to state whether the indicated sequences occur in the same molecule. A tentative interpretation is that the majority of such sequences are identical parts of very similar molecules.

The presence of companion molecules in insulin samples may account for the frequent occurrence of odd residue numbers in reported analyses of insulin (Historical, p. 6). No uneven value can be correlated with the symmetrical structure proposed by Sanger.

C. Identification of the Hydantoin as a Sequence Procedure

In a satisfactory sequence procedure in which the hydantoin derivative of the terminal amino acid is identified directly, the following steps are necessary:

1) the formation of the hydantoin in essentially quantitative yield,

2) the quantitative isolation of the derivative from the remainder of the protein,

3) the estimation of the hydantoin extracted.

That Step 1 takes place in the sequence procedure used in these experiments is suggested by the fact that integral decrements in residue numbers are encountered before and after phenylisothiocyanate treatment. Supporting though less convincing evidence is the fact that eighty to ninety per cent yields of the hydantoins can be isolated in their preparation from amino acids and phenylisothiocyanate (45). It is the extraction of the hydantoin (Step 2) that presents difficulties. In Experiment 6, p. 41, parallel preparations gave quite different apparent extents of reaction, the highest value being obtained when ether extractions were carried out twenty times. It is probable that in all cases the reaction proceeded to near completion but that all hydantoin

was not extracted. A result that is somewhat difficult to reconcile with this view was obtained in Experiment 5, p. 40. When preformed hydantoin was added to protein solutions, it was quantitatively extracted by ten ether extractions except at a pH of 10. (The weakly acidic property of hydantoins derived from neutral amino acids makes their affinity for alkaline solutions readily understandable.) It would seem that there is a difference in the extractability of hydantoins depending upon whether the hydantoin is formed in a solution or added to it.

A more effective extracting solvent, such as N,N-dimethylformamide, would be of much help, but because of interference with the ultra-violet determination of the hydantoin, this solvent and many others cannot be used. Estimation of the hydantoins by means of their conversion to the amino acid is time-consuming and not quantitative (Experiment 3, p. 37). If a highly colored hydantoin were prepared, its extraction could be easily followed and its colorimetric determination could be performed without the many interferences encountered in the ultraviolet region. It may be that nitro or azo derivatives of phenylisothiocyanate would be found satisfactory.

VI. CONCLUSIONS

1. A quantitative sequence procedure which is based on detection of decrements in the number of amino acid residues before and after treatment of a peptide with phenylisothiocyanate is applicable to proteins of the size of insulin.

2. In an electrophoretically pure sample of bovine insulin, there are ten valine residues, six phenylalanine residues, and three isoleucine residues per M. W. 12,000.

3. There are two aminoid terminal phenylalanine residues. This is in agreement with some literature reports and contrary to others.

4. Valine and two or three isoleucine residues occupy second positions; aspartic acid and two valine residues occupy third positions. An undetermined number of glutamic acid residues are in the fourth position.

5. Difficulties encountered in the extraction of hydantoins from protein solutions are not due to an interaction between the hydantoin and the protein, but rather to the fact that hydantoins formed in aqueous solution are dissolved slowly by ether.

VII. SUMMARY

A quantitative procedure for establishing the sequence of amino acid residues in peptides and low molecular weight proteins has been applied to an electrophoretically pure sample of crystalline bovine insulin. Results were obtained that are in rather close agreement with the structure advanced by Sanger on the basis of his investigations of fragments of the molecule.

Attempts were made to use an alternate more direct procedure in which the terminal amino acids were isolated as their phenylthiohydantoin derivatives. The method was not quantitative nor consistently dependable.

With a view toward developing a new and more rapid sequence procedure, some information regarding the formation and properties of phenylthiohydantoins derived from amino acids was obtained.

VIII. ACKNOWLEDGEMENTS

The author is pleased to express his gratitude to Dr. Sidney W. Fox for his suggestion of this problem, for his guidance, and for many helpful discussions.

Thanks are due Mr. Joseph L. Kucera for performing the electrophoretic studies.

The author is also grateful to the American Cyanamid Company for a fellowship held part of the time during which this work was carried out.

The Armour Company generously supplied the crystalline insulin used in the investigation.

IX. BIBLIOGRAPHY

1. Abderhalden, E. and Stix, W., Hoppe-Seyler's Z. physiol. Chem., 129, 143 (1923).

2. Sanger, F., Biochem. J., 39, 507 (1945).

3. Felix, K. and Mager, A., Hoppe-Seyler's Z. physiol. Chem., 249, 111 (1937).

4. Chibnall, A. C., Proc. Roy. Soc. (London), B. 131, 136 (1942).

5. Fox, S. W., Hurst, T. L. and Itschner, K. F., J. Am. Chem. Soc., 73, 3573 (1951).

6. Abel, J. J., Proc. Natl. Acad. Sci. U. S., 12, 132 (1926).

7. Fredericq, E. and Neurath, H., J. Am. Chem. Soc., 72, 2684 (1950).

8. Porter, R. R., Biochem. J., 53, 320 (1953).

9. Harfenist, E. J. and Craig, L. C., J. Am. Chem. Soc., 74, 3083 (1952).

10. Harfenist, E. J. and Craig, L. C., ibid., 74, 3087 (1952).

11. Sjögren, B. and Svedberg. T., ibid., 53, 2657 (1931).

12. Crowfoot, D., Proc. Roy. Soc. (London), A. 164, 580 (1938).

13. Gutfreund, H., Biochem. J., 42, 544 (1948).

14. Tietze, F. and Neurath, H., J. Am. Chem. Soc., 75, 1758 (1953).

15. Fromageot, C., Cold Spring Harbor Symposia Quant. Biol., 14, 49 (1950).

16. Tristram, G. R., Adv. in Prot. Chem., 5, 130 (1949).

17. Brand, E., Ann. N. Y. Acad. Sci., 47, 187 (1946).

18. Sanger, F. and Tuppy, H., Biochem. J., 49, 481 (1951).

19. Sanger, F. and Thompson, E. O. P., ibid., 53, 366 (1953).

20. Harfenist, E. J. and Craig, L. C., J. Am. Chem. Soc., 74, 4216 (1952).

21. Sanger, F., Nature, 164, 529 (1949).

22. Sanger, F., Biochem. J., 44, 126 (1949).

23. Sanger, F. and Tuppy H., ibid., 49, 463 (1951).

24. Sanger, F. and Thompson, E. O. P., ibid., 53, 353 (1953).

25. Sanger, F., Cold Spring Harbor Symposia Quant. Biol., 14, 153 (1950).

26. Bull, H. B., Ann. Rev. Biochem., 21, 197 (1952).

27. Mellon, E. F., Korn, A. H. and Hoover, S. R., J. Am. Chem. Soc., 75, 1675 (1953).

28. Udenfriend, S. and Velick, S. F., J. Biol. Chem., 190, 733 (1951).

29. Fraenkel-Conrat, H. and Fraenkel-Conrat, J., Acta Chem. Scand., 5, 1409 (1951).

30. Fraenkel-Conrat, H., Federation Proc., 11, 214 (1952).

31. Christensen, H., Acta Chem. Scand., 6, 1555 (1952).

32. Fromageot, C., Jutisz, M., Meyer, D. and Penasse, L., Biochim. et Biophys. Acta, 6, 283 (1950).

33. Chibnall, A. C. and Rees, M. W., Biochem. J., 48, xlvii (1951).

34. Lens, J., Biochim. et Biophys. Acta, 3, 367 (1949).

35. Harris, J. I., J. Am. Chem. Soc., 74, 2944 (1952).

36. Waley, S. G. and Watson, J., J. Chem. Soc., 1951, 2394.

37. Baptist, V. H. and Bull, H. B., J. Am. Chem. Soc., 75, 1727 (1953).

38. Edman, P., Acta Chem. Scand., 4, 283 (1950).

39. Marckwald, W., Neumark, M. and Stelzner, R., Ber., 24, 3278 (1891).

40. Kuiken, K. A., Norman, W. H., Lyman, C. M., Hale, F. and Blotter, L., J. Biol. Chem., 151, 615 (1943).

41. Fowden, E. G., Biochem. J., 48, 327 (1951).

42. Partridge, S. M., ibid., 42, 238 (1948).

43. McFarren, E. F., Anal. Chem., 23, 168 (1951).

44. Miller, G. L. and du Vigneaud, V., J. Biol. Chem., 118, 101 (1937).

45. Edman, P., Acta Chem. Scand., 4, 277 (1950).

46. Schweigert, B. S., McIntire, J. M., Elvehjem, C. A. and Strong, F. M., J. Biol. Chem., 155, 183 (1944).

47. Henderson, L. M. and Snell, E. E., ibid., 172, 15 (1948).

48. Block, R. J., Le Strange, R. and Zweig, G., Paper Chromatography. New York, Academic Press Inc., 1952.

49. Moore, S. and Stein, W. H., J. Biol. Chem., 192, 663 (1951).

50. De Fontaine, D. Quantitative Chemical Distribution of Amino Nitrogen from Leaves of Two Strains of Corn. Unpublished M. S. Thesis. Ames, Iowa, Iowa State College Library, 1952.

51. Hurst, T. L. Terminal Amino Acid Residues Appearing During the Proteolysis of Lysozyme. Unpublished Ph.D. Thesis. Ames, Iowa, Iowa State College Library, 1953.

52. Serat, W. F. Properties of Phenylthiohydantoins Derived from Amino Acids. Unpublished M. S. Thesis. Ames, Iowa, Iowa State College Library, 1953.

53. Fox, S. W. and Warner, C. W. Unpublished experiments. Iowa State College, Ames, Iowa.

54. Harington, C. R. and Neuberg, A., Biochem. J., 30, 809 (1936).

55. Schmidt, C. L. A., J. Biol. Chem., 82, 587 (1929).