1961

Structural studies on the [beta]-lactoglobulins

Albert Peter Pfuderer
Iowa State University

Mic 61-2267

PFUDERER, Albert Peter. STRUCTURAL STUDIES ON THE β-LACTOGLOBULINS.

Iowa State University of Science and Technology
Ph.D., 1961
Chemistry, biological

University Microfilms, Inc., Ann Arbor, Michigan

STRUCTURAL STUDIES ON THE β-LACTOGLOBULINS

by

Albert Peter Pfuderer

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

Major Subject: Physical Chemistry

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 University
Of Science and Technology
Ames, Iowa

1961

TABLE OF CONTENTS

I. INTRODUCTION AND REVIEW OF LITERATURE

A. Introduction

β-Lactoglobulin is the second most abundant protein of milk. Bovine β-lactoglobulin has been the subject of many investigations because of its availability and the ease with which it crystallizes. Despite this, not very much is known about its actual structure.

In 1955 β-lactoglobulin was separated by electrophoresis into two proteins, called β-lactoglobulins A and B. The presence of each protein was later shown to be under the control of the same genetic locus, and it became feasible to obtain each protein in a pure form.

Because β-lactoglobulins A and B are one of the few pairs of pure proteins controlled by one locus, the difference in structure between A and B is of great interest.

It was thought that the "fingerprinting" technique which was recently developed by V. M. Ingram (42) could be used to discover differences in amino acid sequence which might account for the different properties of these two proteins.

In this technique peptides obtained from an enzymatic digestion of the protein are separated by two dimensional paper chromatography and electrophoresis. The peptide fragments which differ are eluted, and the differences in

structure between the original proteins are inferred from them.

In order to explain the type of anticipated results and the reason for their value, a brief discussion of the genetic determinations of proteins, protein synthesis, and protein structure will be presented.

B. Genetic Determination of Protein Structure

The following terms to be used throughout this discussion are defined here for the benefit of those unfamiliar with genetics:

gene: A discrete heritable unit concerned with the transmission and development of some hereditary characteristic such as eye color.

locus: A small region on a chromosome which is occupied by a given gene.

allele: Any one of several alternative types of gene which may be present at a given locus on a chromosome. For example, alleles for eye color might be B and b, standing for brown and blue respectively.

diploid: Having twice the basic number of chromosomes, and therefore having two genes to determine most characteristics.

genotype: The type of organism as specified by its gene constitution, such as B/b for eye color in a diploid organism.

phenotype: The type of organism as determined by characteristics, such as blue-eyed.

mutation: A permanent heritable alteration of a gene, detectable if it causes a recognizable phenotypic change in the organism.

recombination: The relocating of genes once linked on the same strand of a chromosome to different homologous chromosomes as if the strands broke and rejoined differently.

Hereditary traits are controlled by genes. These genes are arranged in linear fashion on chromosomes, which are strands in the cell nucleus sometimes visible microscopically. The chromosomes are composed of deoxyribonucleic acid (DNA) and a basic protein. DNA has been shown to be "the genetic material," in that all the hereditary information is coded in its structure.

Much evidence has accumulated in recent years to show that the genetic control of protein synthesis is not a side effect but is probably the primary way in which a gene exerts its effect.

Perhaps the most direct evidence comes from the bacterial viruses, specifically the T even bacteriophages of Escherichia coli. Here we have a clear cut case of almost pure "genes" being added to a bacterial cell system, causing this system to subsequently produce virus proteins which it was previously incapable of making.

These viruses are fairly simple biochemically, being composed of an inner core which is 97% DNA and 3% protein and an outer coat which is all protein. In an infection the virus injects its inner core into the bacterial cell and the synthesis of several virus proteins starts immediately. By four minutes time the enzymes necessary to stop production of bacterial DNA and start production of more virus DNA appear in completed form (54). At ten minutes virus coat proteins appear (98), and by fifteen minutes intact virus has been synthesized.

This same system has been used by Brown (14) for in vitro protein synthesis. Digested bacterial protoplasts were incubated with purified virus DNA, and virus tail proteins were subsequently produced by the system. However, the DNA preparations still contained 0.3% protein which was not virus tail protein.

More evidence that genes direct protein synthesis is also furnished by an array of 54 enzymes isolated from microorganisms. The activities of all these enzymes have been changed by mutations. In each case a mutation at one locus has been demonstrated to lead to an altered enzyme. They have been listed by Fincham in his review article (26). None of these enzymes has been isolated in pure form, however. It is also not experimentally clear whether the altered enzyme is a direct or indirect result of the mutation.

Enzymes are the catalysts the cell uses to synthesize other cellular constituents. In order to control the reactions an organism can carry out and the substances present in the organism, genes must control the enzymes in it. One usually finds a one to one correspondance between gene and enzyme. Logically one would expect that every phenotypic difference between two organisms would be traceable ultimately to just such a difference in enzymes or proteins. A difference in enzymes would then lead to other differences between the two organisms. This theory is summed up in the "one gene-one enzyme hypothesis" of G.W. Beadle (6) which states that the structure of one protein is controlled by one gene.

The specific effects of allelic replacements on protein structure to data have only been worked out for one protein, human hemoglobin. In the famous series of investigations by Ingram and Hunt (43, 39) on the abnormal hemoglobins, hemoglobin A (normal adult), S (sickle cell), and C (C disease) have been shown to differ by the replacement of amino acids at one site of the β chain of the globin molecule. The sequence val-his-leu-thr-pro-glu-glu-lys in hemoglobin A is replaced by.....pro-val-glu-lys in hemoglobin S, and by.....pro-lys-glu-lys in hemoglobin C. The valine residue is thought to be N terminal on the chain, making the substitution at position 6. All these

hemoglobins, as well as more than 9 other abnormal hemoglobins, are allelic (44). That is, they are caused by different alleles at the same hemoglobin locus. There are four chains in the normal hemoglobin molecule, which may be represented as $\alpha\alpha\beta\beta$. This molecule can dissociate into two almost identical half molecules, $\alpha\alpha$ and $\beta\beta$. A heterozygote for sickle cell hemoglobin, which would have the sickle trait rather than the disease, would be of genotype Hb^A/Hb^S; and it would produce both hemoglobin A, $\alpha\alpha\beta^A\beta^A$, and hemoglobin S, $\alpha\alpha\beta^S\beta^S$ (97). Hemoglobin C has a substitution of a glycine for the glutamic acid at position 7 in the β chain (37) while hemoglobins Do and T differ in their α chains (36). Nevertheless, the hemoglobin locus controls all four chains, and all these hemoglobins are allelic. This is a good experimental verification of the "one locus-one protein" hypothesis, as we have rephrased it. It should also be mentioned that there is another hemoglobin disease called thalassemia which involves a different locus. However, it does not produce an abnormal hemoglobin molecule, but seems to inhibit the synthesis of normal hemoglobin (44).

Species differences have also been found in many proteins and peptides. It is not known how many loci are involved here since the species will usually no longer interbreed. One would postulate one locus per protein, however.

Beef ribonuclease has been shown to differ from sheep ribonuclease by the substitution in the sheep enzyme of a serine for a threonine at residue 3 and a glutamic acid for a lysine at residue 37 (2). This work was also done by the fingerprinting technique.

The following are the residues at positions 8, 9, and 10 in the A chain of insulin, as found in the species listed (15, 35). The rest of the molecule is identical in each case.

beef : ala-ser-val
swine: thr-ser-leu
sheep: ala-gly-val
horse: thr-gly-ileu
whale: thr-ser-ileu

The peptide hormones have also been shown to vary from species to species. Swine ACTH differs from sheep and beef ACTH by a substitution of 8 amino acids, residues 25 to 33 (58). Bovine β MSH differs from the swine hormone by a substitution of serine for glutamic acid at residue 2 (30). Similarly, beef vasopressin contains an arginine residue where the swine vasopressin contains lysine (24).

The ribonuclease and insulin molecules are the only two proteins to date whose amino acid sequence has been completely worked out. This makes Ingram's work on hemoglobin all the more remarkable since his technique was able to determine an amino acid difference without determining the entire sequence of both molecules, a project likely to

take a research team a good many years. The peptide hormones, being peptides, are of much smaller molecular weight and the elucidation of their entire sequence is usually feasible.

When genetic mapping was recently extended to the microorganisms it was seen that genetic loci themselves have fine structure, as do the proteins whose synthesis they control. This fine structure of a locus was capable of modification at many points through rare mutations and recombinations.

By using these microorganisms the existence of one rare recombinant organism in approximately a million progeny could be detected. One could then show that mutations occurring in the same locus were affecting different parts of this locus since the affected sites would recombine (9). These recombination frequencies within a locus were usually very much smaller than those between loci.

Since the gene was supposedly indivisible, current usage has been to call the functional unit which specifies a protein by the name "locus."

The locus is observed to have fine structure just as an enzyme has amino acid sub-units. It is highly conceivable that these sub-units of a locus determine the arrangement of amino acids on the peptide chain of the enzyme.

C. Protein Synthesis

The mechanism by which DNA transfers its information into protein structure is not known. Although DNA is the primary genetic material, RNA (ribonucleic acid) is involved at the site of protein synthesis and it seems likely that amino acids are transported to the site of synthesis by soluble RNA (SRNA). DNA is required to start protein synthesis, but once synthesis is started it can continue for a long time without requiring the presence of DNA in the cell. Unicellular organisms have been known to synthesize protein for two weeks after being enucleated. Once their RNA is destroyed however, protein synthesis stops (78). RNA is the immediate participant in protein synthesis.

As the cell multiplies its hereditary materials must be exactly duplicated and passed on to the daughter cells. The mechanism by which this occurs is just beginning to be understood for both DNA and RNA. For each there is an enzyme system which will synthesize the respective nucleic acid given the proper substrate of nucleotides and a nucleic acid "primer" (55, 67). The nucleic acid synthesized has been shown in the case of DNA to resemble the primer exactly, even to the frequency of neighbor nucleotides (53).

These two enzyme systems apparently insure the proper reproduction of the cell's genetic information. However,

the way in which this information, coded in the sequence of nucleotides on DNA, is transferred from DNA to RNA and from RNA to protein is not known.

D. Protein Structure

The preceding discussion has dealt with the way in which a locus, composed of DNA, presumably transfers the information coded in its nucleotide sequence into the sequence of amino acids in a protein.

This amino acid sequence is but one aspect of protein structure and is called its primary structure. There is also a secondary and a tertiary structure. The secondary structure, as proposed by Linderstrom-Lang, is the way in which this chain of residues is folded. The secondary structure characteristic of most proteins is the Pauling-Corey hydrogen bonded α-helix which changes the flexible chain into a coiled rigid rod. The tertiary structure is the configuration in which this rigid rod is bent and folded to form the elipsoids of revolution characteristic of most globular proteins. These ellopsoids are many times thicker than a single α-helix. This tertiary structure is stabilized by hydrogen bonding, hydrophobic bonds, and covalent disulfide bonds between cysteine residues (49). When the tertiary structure is unraveled, the special biological activity of the protein or enzyme

usually vanishes.

We have seen that the hereditary information is probably transferred into a linear peptide sequence. There is very little evidence to date whether the secondary and tertiary structure of a protein is left to chance rearrangements, or whether it also has been coded into the hereditary material. No general hydrogen-bonded structure has been found for RNA as was the case with DNA. Each RNA molecule could have a specific folded structure which acts as a template for the tertiary structure of the protein it synthesizes.

An alternative to this template mechanism which is currently in favor is that the RNA specifies the amino acid sequence of a protein, and this sequence then assumes its most thermodynamically stable form. A difficulty in this argument is that the unfolding, or denaturation of most proteins is irreversible. The folded form is barely stabilized in aqueous solutions by the heat content of its secondary and tertiary bonds, which is offset by the large positive entropy of unfolding. As a result most proteins will denature slowly and irreversibly on standing in solution for a period of days. If the protein were to be formed in a completely unfolded state it is doubtful if it would ever attain its "native" state, especially in a cellular environment containing proteolytic enzymes, which

act extremely rapidly on denatured proteins.

If the protein were synthesized directly as an alpha helix, it would have a better chance of attaining its native form. The entropy change going to the ordered native molecule would be less, and the reaction therefore more spontaneous. A proline residue inserted in the sequence would cause a bend, since proline cannot hydrogen bond in the α-helix. The direction of the bend could be directed by electrostatic forces or even hydrophobic bonds.

If the protein were allowed to seek, its most thermodynamically stable state, it is possible that two forms might be equivalent. If the protein's tertiary structure were formed on a template it is also possible that it might recoil into another configuration. This would lead to a new type of isomerization. The term "configuration isomer" has been proposed by Anfinsen (1) for this type of protein isomer that differs in its tertiary structure. "Sequential isomers" would be those that differed in their amino acid sequence. Genetic changes could produce both types of isomerization as far as we know at the present.

There are also some regions of a protein not necessary for activity. In the section on genetic determination of protein structure, many species variations in hormones, enzymes, and proteins were listed. These altered proteins were still able to perform their function in the other

species. In papain, a classical example, up to two thirds of the molecule, 120 of the 180 residues, may be removed by aminopeptidase without loss of enzymatic activity (38). A mutation at one of these non-functional portions of the molecule might ordinarily go undetected. Any protein obtained from a diploid organism could easily be a mixture of two sequential isomers at one of these non-functional positions. This fact could be the basis of various articles which have been written stating that no protein is homogeneous (19).

In summary it is not known whether a pair of alleles could produce configurational isomers in a protein. They could certainly produce sequential isomers, and any protein isolated from a diploid organism could be a mixture of two sequential isomers.

E. Structural Features of β-Lactoglobulin

In 1955 two English workers, Aschaffenburg and Drewry, examined the β-lactoglobulin of cow's milk by obtaining milk from individual cows (3). They discovered that when β-lactoglobulin was precipitated and subjected to paper electrophoresis the protein separated into two distinct bands which they called A and B, A being the band migrating more rapidly at pH 8.6.

The initial evidence suggested that there was one locus for β-lactoglobulin, and two alleles at this locus, lg A and lg B. Further studies on pedigrees and breeding experiments showed this to be true (4). The cows could be of genotype lg A/lg A, lg B/lg B, or lg A/lg B, producing β-lactoglobulin A, B, or a mixture of A and B respectively.

Although each breed had a characteristic gene frequency, the lg B allele always predominated ($p = 0.72$), even though it led to the production of the smallest amount of β-lactoglobulin. Genotype lg A/lg A characteristically produced almost twice as much β-lactoglobulin as lg B/lg B, with the hybrid intermediate in level.

Studies on the X-ray unit cell of crystals of these two proteins by Green et al. (33) showed that both could exist in orthorhombic crystal forms with differences in lattice constants which were not much greater than the errors involved in measuring these constants. The authors felt that these differences were real. The differences tended to cancel each other in computing the molecular weights. Using their data, one arrives at a molecular weight for A of 35,800±400 and for B of 35,700±400. B was also found to exist in a monoclinic crystal, which gave a molecular weight of 35,800±400. This reverted to the less soluble orthorhombic on standing. Their data lead us to believe that both β-lactoglobulins have a molecular weight

near 35,800, although A may be heavier than B by about 100. The authors also found that both A and B could crystallize together when solutions of A were seeded by crystals of B, and vice versa.

The above data led Green et al. to conclude that the two β-lactoglobulins were almost identical proteins, probably differing only by the substitution of different amino acids at a few sites within the molecule.

Ogston and Tombs have reported that both A and B have identical sedimentation coefficients of 2.77S at 25° C (69). This is in contradiction to the findings of Klostergaard and Pasternak (52) who report that B has a value of 2.88S, while A has a value of 2.68S at 29° C. The ultracentrifuge data in Klostergaard's paper, however, has been questioned by Townend et al. (94). Both the above determinations of the sedimentation constant were done under conditions where no aggregation would be likely to occur according to Townend and Timasheff (93), and so the values are indicative of molecular differences, rather than tendencies to aggregate.

Tombs, in testing the homogeneity of A and B by the variable solvent and constant solvent solubility tests reports that A contains 10% of another protein (91), and that B contains two minor constituents, probably not exceeding 10% of the total protein (92). Townend and

Timasheff (93) also found that 10% of A could not aggregate to form a tetramer, while B could not aggregate at all by itself. However, 30% of B could aggregate in the presence of A to form the tetramer.

We may conclude then, that A is 90% homogeneous, and B either 90% or 70%, homogeneous. These other constituents could be either configurational or sequential isomers.

Tanford and Nozaki (89) have titrated both proteins, and found that form A contains two more carboxyl groups than does B. They also found no significant difference in the U.V. absorptions, the salt binding capacities, or the rotary dispersions of the two proteins.

It may be noted that the substitution of two aspartic or glutamic residues for glycine or alanine would add about one hundred to the molecular weight of A. If the two extra carboxyl groups on A were produced by B having two extra amide groups, but the same total number of carboxyls, there would be virtually no change in molecular weight.

Tanford *et al.* report that there are 51 titratable carboxyl groups/35,500 in native pooled milk β-lactoglobulin (87). Two are α-COOH's, and 49 are side chain COOH's. Two more carboxyl groups are "locked in," and are released irreversibly when pH 7.5 is reached. They do not titrate until then and seem to be in an inaccessible region in the

interior of the molecule. After denaturation at pH 11 these two carboxyl groups titrate in the normal acid range. Both form A and form B were found to have these two anomalous groups, A having 52 normal COOH's and B having 48, two of which were α-COOH's in each case (66).

The presence of these two anomalous carboxyl groups and their subsequent release has been corroborated by the differential infrared spectral study of Suzi et al. (85) on β-lactoglobulin from pooled milk. They were able to show the presence of just two of these carboxyl groups in their sample, which was presumably composed of almost equal amounts of forms A and B.

It is only possible for A and B to be configurational isomers if B contains 2 additional "locked in" carboxyl groups making 4 in all. Both the above experiments show that this is not the case, since they would be expected to be released by pH 11, and they were not. A and B are probably sequential isomers then, and differ either in the number of amide groups, or in the total number of acidic residues.

Data concerning the number of aspartic and glutamic residues in the protein is usually done on the acid hydrolyzed protein in which all amide linkages as well as peptide bonds have been hydrolyzed. We will call this data the "total" number of aspartic or glutamic groups. It

includes those acidic residues present as amides and those "locked in" in the native protein.

Chibnall et al. (18) have compared the number of carboxyl groups present as aspartic and glutamic acid, and the number present in amide bonds as asparagine and glutamine in pooled milk β-lactoglobulin with that in "β1" lactoglobulin. Their "β1" fraction bears no immediate resemblance to either A or B however (93). Of the thirty "total" aspartic and forty eight "total" glutamic residues, ten and seventeen were bound in amide linkages/35,500 M. W. respectively.

Stein and Moore (83) report 31 total aspartic and 46 total glutamic residues/35,500, and 27 moles of amide nitrogen.

The best accepted value for the molecular weight of β-lactoglobulin is 35,500±400 (77). This 35,500 species is composed of two sub-units, which dissociate below pH 3.5, and above pH 7.5 (95). These two half-molecules are either identical or close to identical.

The end group analysis of β-lactoglobulin shows two peptide chains/35,500 both having N terminal leucine and C terminal isoleucine with penultimate histidine (28).

The two halves into which form A splits at low pH will not recombine with the two halves into which form B splits. This was determined by Kiddy et al. (51) by labelling A

with C^{14} and mixing it with unlabelled B. The solution was acidified until the molecules dissociated, and then neutralized. When A and B were separated by paper electrophoresis, no hybrid molecules with one charge difference were found, and none of the radioactivity had passed to form B. This shows that both chains of A differ from both chains of B and suggests that the two extra carboxyls on A are distributed one to each chain, making each molecule a union of two identical halves.

X-ray analysis has revealed that these sub-molecules are nearly spherical and are associated in a double molecule of axial ratio 2:1 (32). Townend et al. (95) have confirmed this data with their values of the frictional coefficient which lead to axial ratios of 1.1:1 and 2:1 for the monomer and dimer respectively.

F. Summary

β-lactoglobulins A and B are almost identical proteins. The only detected difference between A and B is that A contains two more carboxyl groups/35,500. These two carboxyl groups are probably distributed one to each peptide chain since both chains of A differ from both chains of B. The two extra groups could originate by A having two more aspartic or glutamic residues than B or by B having the same total number of acidic residues but having two more present

as amides.

It is doubtful that A and B are configurational isomers, since B does not appear to have more than two "locked in" carboxyl groups.

Neither A nor B are pure proteins, and are at best 90% homogeneous. These impurities could be configurational isomers of A and B or they may be sequential isomers with no detectable charge difference.

II. ISOLATION OF PROTEIN

A. Experimental

1. Typing of Herd

Unpasteurized whole milk was collected from cows in the Iowa State dairy herd and the β-lactoglobulin isolated from it for typing by a modification of Aschaffenburg's method (4).

A 50 ml. sample of whole milk was warmed to 37° C and 10 g. of Na_2SO_4 added with stirring until all the salt has dissolved ($\frac{1}{2}$ hr.). The casein was filtered off on S. and S. 598 filter paper and the sample cooled to 25° C. The whey proteins were precipitated by adding 5 g. of $(NH_4)_2SO_4$ to 25 ml. of the filtrate. This precipitate was filtered on Whatman #50 filter paper, skimmed off the filter paper while still viscous and transferred into $\frac{1}{4}$" Visking dialysis tubing, with no entrapped air. This was dialyzed overnight against a large volume of distilled water.

A strip of Whatman #3 filter paper of the appropriate size was dipped into the dialyzed sample and inserted into a starch gel for electrophoresis by the method of Smithies (79).

Baker reagent grade soluble starch was used without further modification at a concentration of 14% ($\frac{W}{V}$) in the

electrophoresis. The gel was made up in .0015 M sodium barbital solution, and had a final pH of 6.7. The bridge buffer used was composed of 10.3 g. sodium barbital and 111.8 g. boric acid per 5 liters of solution, and had an ionic strength of .01 and a pH of 6.7.

As Smithies had shown previously (79) β-lactoglobulin can be separated into two bands in starch gel electrophoresis if extremely low ionic strengths are employed. Although barbiturate gives much better separation than the other buffers tried, the acid form was not soluble enough to give a pH of 6.7 at the ionic strength necessary to stop the gel from swelling, and so a mixed bridge buffer of borate and barbituate was ultimately used.

The starch gel experiments were conducted in horizontal plexiglas trays 12" x 1" x $\frac{1}{4}$". An EC Apparatus Co. #453 power supply and # 401 electrophoresis unit were used with filter paper strips to conduct current to the trays. The sample was run at 3.5 volts/cm. for five to six hours and excellent separation of the two bands resulted. Pure samples of A and B were kindly furnished by Dr. Ashaffenburg, and our two bands coincided with his forms A and B.

2. Isolation

β-lactoglobulin A was isolated from a Brown Swiss cow by ammonium sulfate fractionations using the procedure of

Ogston and Tilley (68). The oil which appeared after adjusting the pH to the isoelectric point and dialyzing extensively was seeded with crystals from Aschaffenburg's sample of A. Crystals formed within 15 minutes. These crystals were separated from the oil by centrifuging, dissolved in dilute NaCl solution and redialyzed. They came out directly as crystals on this dialysis.

β-lactoglobulin B was first isolated from Holstein Friesian cow H-4073 using the same procedure used on A. After one week of dialysis, no oil had appeared. The solution was then electrodialyzed, and 50 g. of oil were collected from 47 liters of milk. This oil only yielded 2 g. of wet crystals, or about 0.7 g. of dry β-lactoglobulin B. During the time spent in working up the oil, cow H-4073 was sold, and an alternate B cow, H-4016 had to be located.

Since β-lactoglobulin B could not readily be isolated by normal procedures because of its high solubility, the new procedure of Ashaffenburg and Drewry (4) was used. This involved an acid fractionation at pH 2 to eliminate α-lactalbumin, and kept the volume of all solutions at an absolute minimum. The protein crystallized readily when seeded with orthorhombic crystals of Aschaffenburg's form B.

B. Results and Discussion

1. Typing of Herd

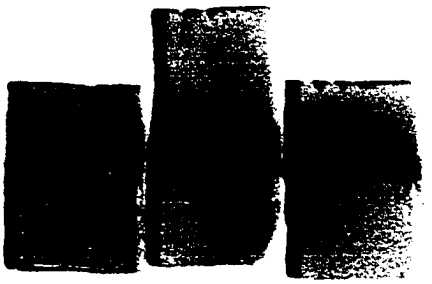
A few representative starch gel patterns of β-lactoglobulin from individual cows of the Iowa State herd are shown in Plate 1. Two samples were commonly run side by side in the same gel. A is the faster migrating band of the two. The origins are at the botton of the gel.

The 6 samples in the two figures on the lower right had a commercial sample from Swift and Co. run as a control between the two herd samples. This sample was dated 1950, and had presumably been stored in the cold room at 4° C since then. It shows a new band which migrated ahead of A. This sample also exhibits unique denaturation properties not normally observed with β-lactoglobulin (7).

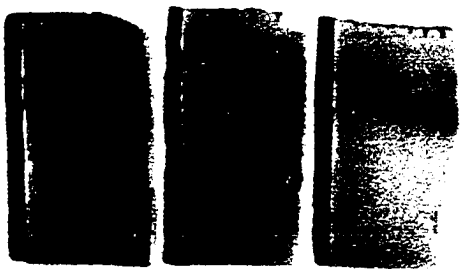
Sample #H-4275 is also of interest since it shows a component migrating about one half as far as β-lactoglobulin B. This is characteristic of α-lactalbumin, another whey protein. All of the electrophoresis samples except the crystallized protein contained other whey proteins which were usually not present in large enough quantities to be visible.

Samples S-4228 and H-4016 were finally selected as the purest A and B available. Another B sample, H-4073, was also selected. The impure herd samples of these proteins

Plate 1. Starch gel electrophoresis of herd samples of β-lactoglobulin. The gels were run in .0015 M sodium barbital, pH 6.7 at 3.5 volts/cm for 5 or 6 hours



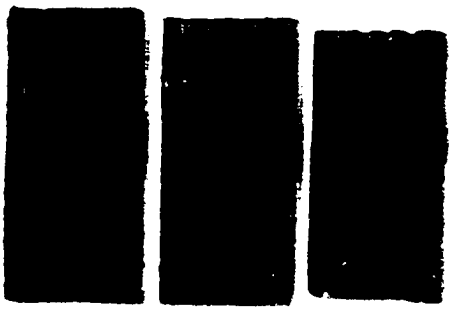
H-4201	H-3746	H-4093
H-4139	H-4275	H-4268



H-3964	H-3111	H-4055
H-4104	H-3953	H-3461

H-3993	H-4052	H-4201
H-3957	H-4170	H-4281

G-3348	S-4033	G-4167
S-4225	S-4208	G-4262



J-3470	S-3945	J-4117
G-3707	S-4112	J-4243

continually showed complex edge effects during electrophoresis and no clear patterns were ever obtained. However, a cyanogum gel electrophoresis of the twice crystallized material is shown in Plate 5, together with Ashaffenbrug's A and B.

Traces of B were found in the A sample and vice versa, but these were of the order of 10% or less and would not be expected to interfere with the fingerprinting analysis. It is hard to reconcile the fact that no absolutely pure A or B samples were ever found in the 42 cows examined with the simple genetic mechanism of two alleles at one locus controlling the synthesis of A and B. Perhaps the lg A allele suppresses the synthesis of B, or converts B to A, either operation of which could be incomplete.

2. Isolation

Approximately 35 g. of dry crystalline β-lactoglobulin A were obtained from 37 liters of milk. This protein contained only trace amounts of other proteins as can be seen from the patterns in Figure 4 of Plate 5.

A much lower yield of β-lactoglobulin B was obtained. This was probably due to the greater solubility of B, and the fact that cows producing B have a lower concentration of this protein in their milk. B also contained only trace amounts of other protein, as is shown in Figure 5 of Plate 4.

Plate 1. Starch gel electrophoresis of herd samples of β-lactoglobulin. The gels were run in .0015 M sodium barbital, pH 6.7 at 3.5 volts/cm for 5 or 6 hours

A twice crystallized sample of B was tested for purity on a Sober-Peterson DEAE cellulose column (60). The elution diagram is shown in Graph 1. The column dimensions were 8.7 x 2.0 cm. The column was equilibrated with pH 5.9, .05 M. phosphate buffer, .05 M. in NaCl. The 100 mg. of B applied was eluted with pH 3.4 phosphate buffer .05 M. in NaCl by a pH gradient elution using a 200 ml. mixing flask.

III. PEPTIDE MAPPING

A. Introduction

The two crystalline samples of A and B were subjected to the "fingerprinting" technique of Ingram (42) as it has been modified by Anfinsen and coworkers (2, 48).

In this technique the protein is digested into peptides with a proteolytic enzyme of high specificity, such as trypsin or chymotrypsin. The digestion mixture is chromatographed on filter paper in one dimension, and then the paper is electrophoresed in the second dimension at ninety degrees to this. The pattern of peptide spots, revealed by spraying with appropriate reagents, is characteristic for a given protein; if a second protein differs from this protein by only a few amino acid residues, its pattern will differ in only a few of the peptide spots. Efforts can then be concentrated on the small portion of the structure shown to be different. Amino acid analyses performed on the peptides separated by the technique will indicate any differences in composition; amino acid sequence studies of the peptides which are different may locate the structural differences in the proteins.

B. Procedure

1. Digestions

Both native and heat denatured β-lactoglobulin proved fairly resistant to the action of proteases, the former because of its folded structure, and the latter because of its insolubility.

Accordingly, the enzymatic digestions were carried out in 2M urea after initially treating the protein with 6M urea in a test tube held in an 85° C water bath for five minutes. Optical rotation measurements taken on the protein solutions after dilution to 2M urea at 2,4 and 10 minutes incubation times showed that the protein had not completely unfolded at 2 minutes but that its rotation had reached a maximum value at 4 and 10 minutes. Both trypsin and chymotrypsin are active in 2M urea, but not in higher molarities.

The procedure of Katz, Preyer and Anfinsen (48) was followed for removing the urea and for fingerprinting the peptides. Six digestions were done in this way: one trypsin, one chymotrypsin, and one trypsin followed by chymotrypsin for each protein.

Two mg. of protein, 50 γ of trypsin and/or 25 γ of chymotrypsin in 0.6 ml. of total solution were used for each fingerprint. Both enzymes were crystalline and salt free

and both were obtained from Armour. Digestions were done at 25° C and the pH was kept near 8.2 by making the solution 0.1 M in ammonium carbonate. Each enzyme was allowed to act for 90 minutes; the reaction was then stopped by placing the reaction mixture on a small Dowex 50 column in the H^+ form.

Urea was removed by washing the column with deionized water, the $CO_3^=$ removed as CO_2 gas, and the peptides eluted from the column with 4M NH_3. The eluate was lyophilized to remove water and ammonia, a drop of water added to the dry product, and this peptide solution applied to one corner of a sheet of Whatman #3 filter paper. Fingerprints of these six digestions are shown in Plate 2, Figures 1-6.

Two control digestions at pH 8.0 were also run in the Carlsberg pH-stat (45): trypsin and chymotrypsin digestions of β-lactoglobulin B. The sample had been previously heat treated in 6M urea, and was diluted to 2 M urea. The pH was adjusted by automatically adding 0.1 M NaOH. A urea blank was run, and the urea hydrolysis subtracted from each run. The protein solutions were titrated to pH 10 initially, and again at the end of 90 minutes of enzymic digestion. Concentrated HCl was added to give a pH of 8.0, and the digestion run at this pH. The increase in the final titration over the initial one, added to the uptake during the reaction gives an estimate of the number of bonds broken by the

enzyme since the average pK of the amino groups liberated in the peptides is about 7.7.

Subsequent digestions and fingerprints were performed on protein which had its disulfide bonds reduced and alkylated, rather than on urea denatured protein. This disulfide treatment, as shown in the subsequent section on physical studies, unfolded much of the native structure of the protein.

The first method used to reduce and alkylate was that of Sela, White and Anfinsen (76), worked out for ribonuclease.

The protein was dissolved in 8 M urea and reduced with thioglycolic acid at pH 8.5 for 4.5 hours at room temperature. The reduced protein was precipitated with an acetone - 1 M HCl mixture (39:1) at -5° C, and the sample washed with this reagent and with cold ether. The precipitate was then reacted with iodoacetic acid for 2 hours at room temperature and pH 8.5. The carboxymethylated protein, so called because its cysteine residues have been changed to carboxymethyl (CM) cysteines, was then precipitated and wahsed as before with acetone-HCl and ether.

β-Lactoglobulin proved to be so soluble in 8 M urea that large volumes of acetone-HCl were needed to precipitate it. Several hours were required to centrifuge out the precipitate, and during this time there was ample opportunity

for air oxidation of the reduced protein.

The CM derivatives of both A and B were nevertheless made using the above procedure, and enzymatic digestions of these derivatives were performed in the pH-stat in order to determine if the same number of bonds were broken in each protein. A sample of 1.5 mg. of CM protein proved best for each chromatogram, and a sample of from 6 to 12 mg. of protein was usually digested at one time. The same ratio of enzyme to protein previously used was employed, and the digestions were done at pH 8.0 at 25° C. The pH was maintained by automatically adding either 0.05 M diethyl amine solution or 0.1 M KOH. The small amount of base added did not seem to interfere in the subsequent chromatograms.

Control digestions of this material were also run to determine the total number of bonds broken during the digestion.

Fingerprints of these trypsin and chymotrypsin digestions are shown in Plate 3, Figures 6-10.

Further attempts at complete reduction and alkylation were based on the newly reported method of Moore et al. (63) involving reduction with sodium borohydride and subsequent destruction of the reagent by a pH change at the time the iodoacetate is added. The excess reagents were removed by overnight dialysis against distilled water with added mixed bed ion exchange resin. This method has been reported to

completely alkylate β-lactoglobulin and several other proteins.

The only change in the procedure was the addition of a few drops of n. decanol to stop foaming during the reduction. The entire treatment was carried out in a three neck flask, into which N_2 was bubbled continuously.

Trypsin digestions were then performed on this material in the same fashion as for the other CM β-lactoglobulins, and 4 representative chromatograms are shown in Plate 6, Figures 11-14.

2. Paper chromatography

The solvent employed in all the paper chromatography was butanol-acetic acid-water (4:1:5). The solvent was usually prepared an hour before each run. Paper chromatography was performed first because this solvent gave a cleaner initial separation of the peptides than did paper electrophoresis.

Large 18 1/4 x 22 1/2" sheets of Whatman #3 filter paper were used for the experiments. Although Whatman #3 mm occasionally would give better separations, the thick #3 paper held more material, which was essential for subsequent amino acid analyses of the peptides. All the chromatography was done in large chromatocabs, of the type made by Research Specialties Co., Berkeley, Calif. Since the solvent front

in the outer papers regularly moved faster, blank sheets were run on the outside troughs. The paper was removed just before the solvent front reached the bottom, and air dried in a hood for about three hours before electrophoresis was run in the next dimension.

3. Paper electrophoresis

In order to do paper electrophoresis on large sheets of filter paper, a high voltage power supply capable of handling about a hundred milliamperes of current is needed to reduce the time required for migration of the peptides. The passage of this current at high voltages generates considerable heat on the paper, so that an efficient cooling system is also needed.

For the initial experiments conducted at the National Institute of Health during the summer of 1958, the old procedure in use at that time in Dr. C. B. Anfinsen's laboratory was utilized. A 2,000 volt D. C. power supply was connected to electrodes in two rectangular glass museum jars, set up so that the small jar was inside the larger jar. The electrodes were in contact with a low ionic strength pH 3.55 pyridine-acetic acid-water (1:10:289) buffer at the bottom of each jar. The rest of each jar was filled with toluene into which was dipped a glass cooling coil. The paper, first saturated with buffer, was draped so that one

end was in each buffer solution with the majority of the paper in contact with the cooled toluene. The buffer level on the cathode was approximately 11" higher than the anode to oppose electroosmosis. The papers (Plates 1-10 of Figs. 1 and 2) were run at 2,000 volts for 60 minutes.

Later experiments were conducted in a smaller modification of the water cooled plexiglass tank described in reference (48), which was built at Iowa State University. The two electrode chambers were filled with the pH 3.55 buffer, and a purified kerosene fraction called "Oleum" manufactured by Standard Oil layered over them. The Oleum was cooled by tap water circulating through two glass coils on the sides of the tank, with water pressure regulated by a diaphragm valve manufactured by the Halsey-Taylor Co. of Warren, Ohio. The oleum was also stirred with an air powered mixer to get better thermal equilibrium. A 2,500 volt D. C. power supply was also built utilizing an old gas tube transformer, two Amperex 872 AX mercury rectifier tubes, and two oil condensers.

Several runs were made on the finished apparatus with uncharged dextran to check on solvent flow by electroosmosis. It was found that the dextran moved toward the fold on the top (middle) of the paper from both ends, and moved farther the closer it was placed to the end of the paper. This flow may have been caused by evaporation of the solvent from the

warmer top (middle) of the paper. The flow of solvent to the cathode by electroosmosis, evident near the middle fold of the paper, was slight in comparison to this effect.

The buffer levels on both sides were made equal and the paper always placed so that the fold occurred exactly in the center. The distance migrated was then increased for a spot near the origin, and decreased for a spot near the end in a reproducible manner which tended to prevent material from running off the bottom of the paper.

The fingerprints run in this new apparatus (Figures 11-14 of Plate 6) were run at 2,400 volts for 120 minutes. More voltage and longer times were required for results comparable to those obtained in the old apparatus. The ionic strength of the buffer evidently decreased with time over a period of months such that the same applied voltage eventually produced higher mobilities.

4. Development of spots

All the spots on the chromatograms in Figures 1, 2 and 3 were developed by dipping the chromatogram in a 0.5% solution of ninhydrin in ethanol and heating in a 75° C oven for 8 minutes, or a 70° C oven for 10 minutes.

Care must be taken that the papers are ethanol free before being put in the oven; in this investigation, one oven exploded when moist chromatograms were placed in it.

The papers to be used for amino acid analysis were sprayed lightly with .01% ninhydrin in ethanol, heated at 75° C for five minutes, and then hung away from direct light at room temperature until maximum color developed, which took from 3-12 hours.

Several other color tests were used as a check on the ninhydrin reagent.

A starch iodine spot test sensitive to peptide bonds was made by combining the procedure of Reindel (72) for chlorinating the peptide imido group with the starch-KI reagent of Rydon (74). This combination proved to be more sensitive than either the benzidine test of Reindel or the Cl_2-KI test of Rydon, and was more sensitive than ninhydrin for several of the larger peptides.

A modification of the Pauli test developed by Sanger and coworkers (15) was used to locate peptides containing histidine or tyrosine. Peptides containing arginine were located by the sensitive Sakaguchi reagent of Acher listed on page 92 of reference (11).

5. Amino acid analysis

The first qualitative amino acid analyses were performed with the two dimensional paper chromatography systems reported by Anfinsen et al. (2) on sheets of Whatman #1 filter paper 18 1/4 x 11 1/4". The first dimension was

B.A.W. (4:1:5) and the second was 80% pyridine. This method failed to resolve leucine from isoleucine or valine from methionine, and also required considerable sample.

The next method tried used the E/O solvent systems of Hardy et al. (34). This method consistently separated valine from methionine and separated leucine from isoleucine about 50% of the time. The size of the papers could be reduced to 8 x 8" so that less sample was required, and these small papers could easily be run in one gallon wide mouth bottles so that as many as 12 or more papers could be run simultaneously.

The peptide spot to be analyzed was located by staining with 0.01% ninhydrin as described previously, cut from the chromatogram, and the peptide material eluted directly into a 10 x 70 mm. thick walled ignition tube with constant boiling HCl. Only about 5 drops of material were collected in the tube.

The mechanics of elution were as follows. The paper was cut so that one end came to a point. The other end was slit lengthwise and a clear strip of Whatman #3 filter paper inserted into the slit. This paper strip was placed between two halves of a broken glass microscope slide held together by a clip made of polyethylene tubing. The glass slide was leaned against one side of the bottom half of a petri dish which had been glued to an inverted 150 ml. beaker. The

ignition tube was placed under the slide in a 30 ml beaker such that the pointed end of the peptide spot rested on the lip of the tube. When the petri dish was filled with constant boiling HCl, the acid rose by capillary action through the paper strip, then washed the contents of the peptide spot into the ignition tube.

The tube was then sealed and placed in an oven at 110° C for 18 hours. At the end of this time the tube was broken open and the contents dried under vacuum in a desiccator filled with KOH flakes. A few drops of water were added to the dried residue, and the resulting solution applied to a paper chromatogram for analysis.

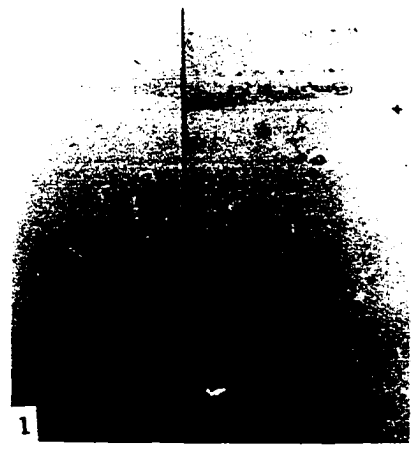
C. Results and Discussion

1. Heat and urea denaturation

Fingerprints of the heat and urea denatured proteins are shown in Figures 1-6 of Plate 2. β-Lactoglobulin A is in Figures 1, 3 and 6, while B is in the others. Figures 1 and 2 are trypsin digestions, Figures 3 and 4 are chymotrypsin digestions, and Figures 5 and 6 are trypsin digestions followed by chymotrypsin. The electrophoresis tanks used were not large enough to contain the whole chromatogram so each fingerprint has been cut in half and the halves electrophoresed separately. The reproducibility of the electrophoresis in these runs was poor because of temperature

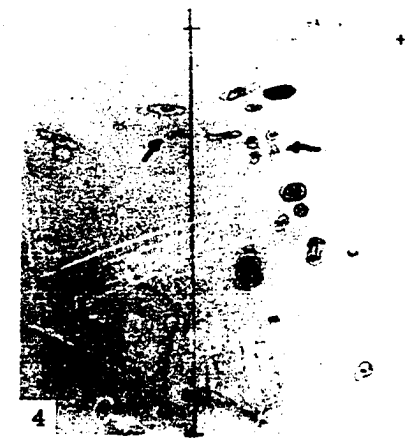
Plate 2. Fingerprints of urea denatured β-lactoglobulins A and B

Figure 1. A, trypsin digestion
Figure 2. B, trypsin digestion
Figure 3. A, chymotrypsin digestion
Figure 4. B, chymotrypsin digestion
Figure 5. B, trypsin and chymotrypsin digestion
Figure 6. A, trypsin and chymotrypsin digestion

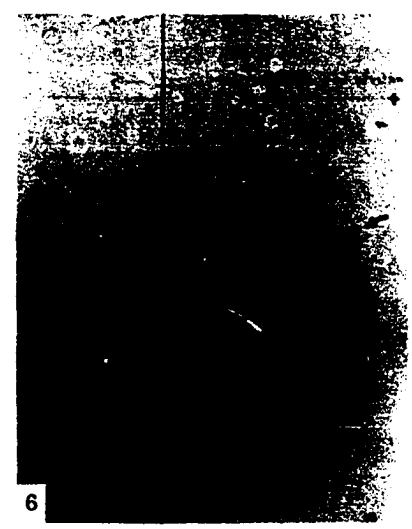

1


2


3


4


5


6

variations in the tanks, so the patterns are occasionally shifted in the vertical direction. The fingerprints have been deliberately moved in this direction so that the patterns correspond as much as possible, with the result that the origins (crosses in the upper right hand corner) are not aligned. Since we are searching for a few peptides that are different among many that coincide, it is not essential that the origins be aligned.

A careful study of all six chromatograms shows that the spots which differ between A and B are all present in very low concentrations, as judged by the extent with which they stain with ninhydrin. These spots have been indicated by arrows. The two doubled peptides in A, Figure 6, are the only exceptions to this, and on close examination the corresponding spots on the B chromatogram, which are blurred, show evidence of being double also.

The trypsin digestions resulted in about 24 peptide spots, and the trypsin-chymotrypsin digestions in about 53 peptide spots. The number of splits with trypsin determined from the number of arginine and lysine residues per 17,750 M. W. should have been 17, leaving 18 peptide fragments. The number of splits with chymotrypsin judged by the number of tyrosine, tryptophan, phenylalanine and methionine residues per 17,750 M.W. should have been 13, leaving 14 peptide fragments. Enzymatic digestions of both insulin and

ribonuclease show a close correspondence between this calculated number of breaks and the actual observed number, although the bonds broken by chymotrypsin do not always correspond to its specified action (13, 82a).

The occurrence of more peptides than usual is typical of an incomplete enzymatic digestion. When a control trypsin digestion of β-lactoglobulin B, previously heat-treated in 6M urea and diluted to 2M urea, was titrated in the pH stat it was found that about 51% of total digestion had been obtained. The chymotrypsin digestion of heat and urea denatured B was 58% of the total.

The optical rotatory constants of solutions of A and B in 2M urea, heated for the specified times in an 85° C bath in 6M urea, are shown in Table 1. Readings were taken at 5 wavelengths, and the concentration was approximately 1%. Since the concentrations were not determined exactly but the same stock solutions of A and B diluted for each sample, the specific rotations are only approximate, but are useful for comparison purposes. The uncertainty in the λ_c values from only five readings is about 5 mμ. When these values are compared with other rotatory constants of native and denatured β-lactoglobulins in Graph 10, we see that the protein heated for 0 and 2 minutes is about half denatured, while the protein which has been heated for 4 and 10 minutes is about two thirds denatured. In this graph, the origin

Table 1. Optical rotary constants of β-lactoglobulin

Time	λ_c (mμ) A	λ_c (mμ) B	$[\alpha]_D$ A	$[\alpha]_D$ B
0 min.		233		
2 min.		231	-58°	-53°
4 min.	235	225	-83°	-75°
10 min.	229		-85°	
60 min.		214		

represents the completely denatured protein while the farthest point in the upper right hand corner is the native protein.*

2. Thioglycolate reduction and alkylation

It was decided to obtain more complete enzymatic digestion of the protein by further denaturation rather than investigate the minor components shown on these fingerprints, which could easily represent peculiarities of each incomplete digestion.

Accordingly the disulfide bridges of A and B were reduced and alkylated using thioglycolic acid and iodoacetic acid, and the procedure of Sela et al. (76). It was suspected that much reoxidation had occurred during the

*The physical significance of these optical constants is discussed in the section on Optical Rotation.

procedure, as explained in the preceding section, but the CM material was chromatographed nevertheless. Since it was not water soluble, the material was first dissolved in 8 M urea, then diluted to 2 M urea, and the urea eliminated after the digestion by the same procedure used previously.

Control digestions were run in the pH stat in 2 M urea solution of the trypsin digestion of CMA. The kinetics were typical of two reactions first order in H^+ uptake with widely different rates. The first reaction was essentially over in 30 minutes while the second reaction had a half time of 60 minutes. The time for the first reaction to end was estimated by the time at which the plot of ln (moles H^+) vs. time coincided with the line of the linear second reaction. The number of bonds broken during the digestion was calculated assuming a pK of 7.7 for the liberated amino groups. At the end of 15 minutes the digestion had proceeded to 51% of completion, while the final uptake at 6 hours of digesting was 109% of the total. This extra 9% may have been caused by traces of other proteolytic enzymes in the trypsin or by nonspecific action of the trypsin itself at this long time.

The chymotrypsin digestions of CMA also were typical of two first order reactions. The fast reaction was over at 30 minutes and the slow reaction had a half time of 61 minutes. The digestion at the end of 30 minutes had proceeded to 50% of completion, while the final value, obtained

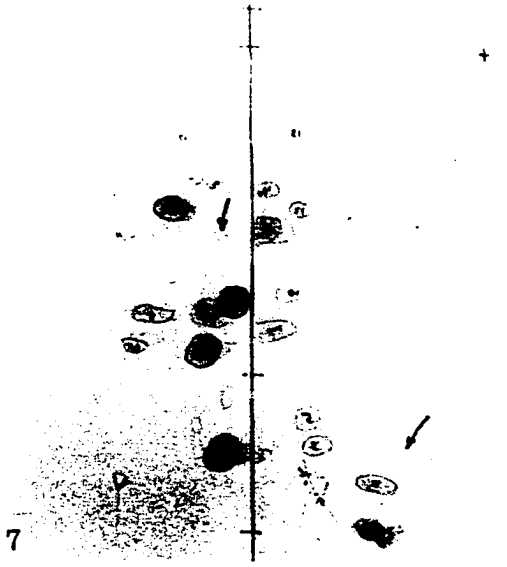
by graphical extrapolation from a plot of moles H^+ vs. 1/t was 130% of the total. Chymotrypsin is known to break a spectrum of bonds at varying rates and the four listed (tyr, try, phe, met) were those which had the fastest rates.

The trypsin and chymotrypsin digestions of CMA and CMB were accordingly run for 3.5 hours at which time both digestions should have proceeded to better than 90% of completion. It was felt that a longer time might give appreciable non-specific enzymatic attack. The pH was kept at 8.5 by the addition of a 5% aqueous solution of trimethylamine using an autotitrator, and the resulting fingerprints are shown in Plate 3. Figures 7 and 8 are trypsin digestions while 9 and 10 are chymotrypsin digestions. Peptide spots present in one chromatogram but not the other corresponding one have been indicated by arrows, and a dashed circle has sometimes been drawn in the corresponding place in the other chromatogram.

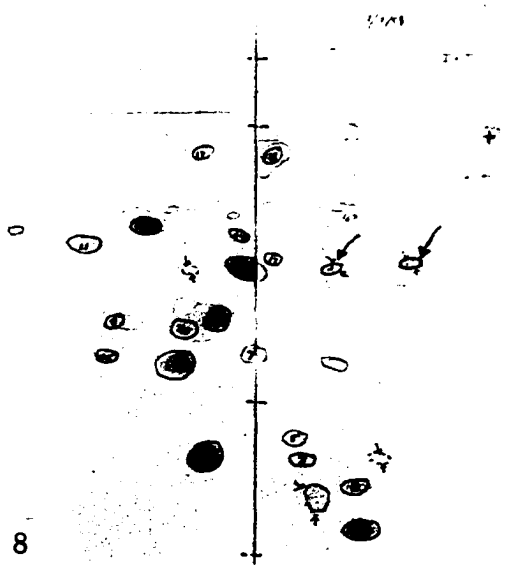
Spot #2, indicated by an arrow in Figure 9 of the chymotrypsin digestion of CMA was cut out, eluted and hydrolyzed, and showed a trace of lysine and a faint trace of another amino acid which may have been alanine, valine, or methionine but which was not present in high enough concentration to identify. Lysine alone on a fingerprint moves approximately as far as spot #1 in the vertical (electrophoresis) direction but lies close to the origin in the horizontal (BAW) direction

Plate 3. Fingerprints of thioglycolate reduced and alkylated β-lactoglobulins A and B

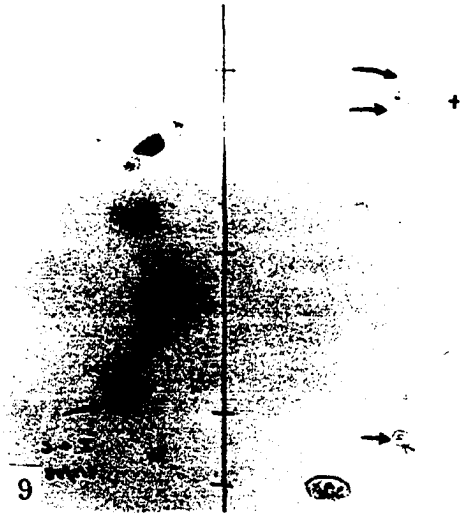
Figure 7. CMB, trypsin digestion
Figure 8. CMA, trypsin digestion
Figure 9. CMA, chymotrypsin digestion
Figure 10. CMB, chymotrypsin digestion



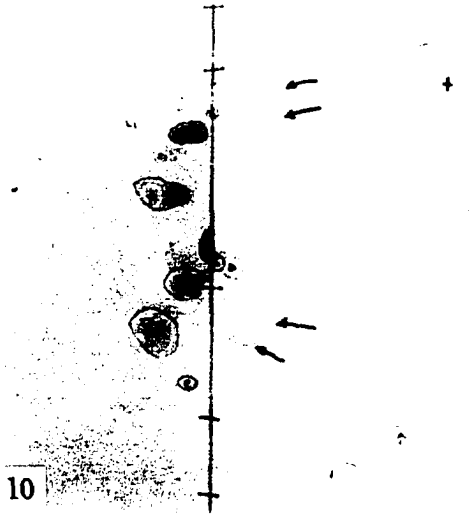
7



8



9



10

than does spot #2, so spot #2 was probably not just lysine. This same peptide spot showed up in subsequent chromatograms of CMB made from this same chymotryptic digestion. The two spots near the origin in CMA, Figure 9, also showed up in CMB on subsequent chromatograms, while the double spot labelled 4 and 5 in Figures 9 proved to be two spots present in both A and B. The two very faint spots between spot #2 and spot #4 in Figure 10 proved to have a third spot below them all three of which were present in both CMA and CMB in subsequent chromatograms. The higher of the two very faint spots near the origin in CMB, Figure 10 also showed up in subsequent chromatograms of CMA, while the lower one did not. Further work on these fingerprints was halted because the supply of CMA and CMB had been used up.

The trypsin digestions of CMA and CMB are shown in Figures 7 and 8. In CMA, Figure 8, the two faint spots on the right hand side indicated by arrows did not show up in CMB, Figure 7, where dotted lines have been drawn. In other chromatograms of this same digestion the spot nearest the origin showed up in B while neither of the two spots showed up in A.

The faint spot above peptide 2 in B, Figure 7, also appeared in both A and B in subsequent chromatograms, while the spot below peptide 3 in Figure 8 never showed up again in either A or B. The spot between peptides 9, 12 and 15

in B, Figure 7, showed up on A but not on B in other chromatograms.

To summarize, the only reproducible peptide spot found to be different in this trypsin digestion was the faint spot indicated by the arrow nearest the origin in CMA, Figure 8. Two faint spots were found in the chymotrypsin digestion, spot #2 in CMA, Figure 9, and the lower of the two arrows near the origin in CMB, Figure 10. All these peptide spots, with the possible exception of spot #2, were so faint that they would not always appear in the fingerprint, and they may easily have been either products of incomplete enzymatic digestion, peptides produced at a slower rate in the digestion, or peptides produced because of a difference in the reduction and alkylation, rather than because of a difference in primary structure between A and B. Accordingly, it was decided to carry out a more satisfactory reduction and alkylation rather than continue on with the old procedure.

3. Borohydride reduction

Both β-lactoglobulin A and B were reduced and alkylated using $NaBH_4$ and iodoacetate according to the procedure of Moore et al. (63). Borohydride would not be expected to react with anything in a protein except cystine (29). The final products were water soluble, in contrast to the thioglycolate reduced proteins. An amino acid analysis, a

rotary dispersion, an ultracentrifuge run, and a gel electrophoresis experiment were run on the carboxymethylated proteins to check on their purity. The results of these tests will be reviewed before the digestion and fingerprinting evidence is discussed.

The results of the amino acid analysis are shown in Table 2. The CM derivatives were first dissolved in glass distilled water which was then diluted 200 times with constant boiling HCl, and the hydrolysis was done in evacuated pyrex tubes at 110° C for 22 hours. The method of Moore and Stein (64, 65, 81) was used in the analysis.

Table 2. Amino acid analysis

Amino acid residue	CMA moles/35,500	CMB moles/35,500	Literature values[a] moles/35,500
cys-SO_3H	.04	.02	
CM cys	7.4	7.6	
asp	28.1	26.3	29.4
thr	9.1	9.1	17.3
ser	4.8	5.1	13.9
glu	39.8	40.6	51.9
pro	14.2	13.4	16.0
gly	5.2	6.9	7.1
ala	24.6	25.7	26.3
½ cys	0.4	0.2	10.1
val	19.2	16.0	17.4

[a]Stein and Moore (83).

The carboxy-methyl cysteine values have been corrected for the 90% loss in hydrolysis under these conditions reported by Moore et al. (63), and as is seen in Table 2, only 7.5 of the 10 half cysteines appear as CM cysteine. The values for serine, threonine, proline and glutamic acid are also extremely low. Moore et al. (82b) have very recently reported that there is some cyanate present in all urea solutions, and that this reacts with cysteine and lysine to form carbamyl derivatives which can be recovered after acid hydrolysis. Since 2 of the 10 half cystines are present in the native protein as cysteine residues, this is presumably what has occurred here. A similar reaction may have occurred in the threonine, serine, and glutamic residues also, since cyanate is known to react with acids and alcohols also.

The differences in aspartic acid, glycine, alanine, and valine content between CMA and CMB are evidently real, and have been confirmed by the recently published work of Gordon et al. (31b) who report that A contains 2.0 more moles of aspartic acid and 1.6 moles of valine, while B contains 1.4 more moles of glycine and 1.3 moles of alanine per 35,000. We found 1.8 more moles of aspartic and 3.2 moles of valine in CMA, 1.7 more moles of glycine and 1.1 moles of alanine in CMB per 35,500. Since the error in these determinations

Plate 4. Starch gel and cyanogum gel electrophoresis of β-lactoglobulins A and B and their derivatives

Figures 1, 2 and 3 are cyanogum gels of β-lactoglobulins A and B, previously reduced with 4% thioglycolic acid, run in 6 M urea-1% thioglycolic acid which had been adjusted to pH 6.5 with tris, and were run at 20 volts/cm.

Figure 1. A and B after 1 hour
Figure 2. A and B after 4 hours
Figure 3. A and B after 6 hours

Figures 4, 5, 6 and 7 are gels run in .001 M sodium barbital at 3.5 volts/cm. for 6 hours

Figure 4. A and Ashaffenburg's A; cyanogum gel, pH 7.7
Figure 5. B and Ashaffenburg's B; cyanogum gel, pH 7.7
Figure 6. CMA and CMB; cyanogum gel, pH 7.7
Figure 7. CMA and CMB; starch gel, pH 6.7

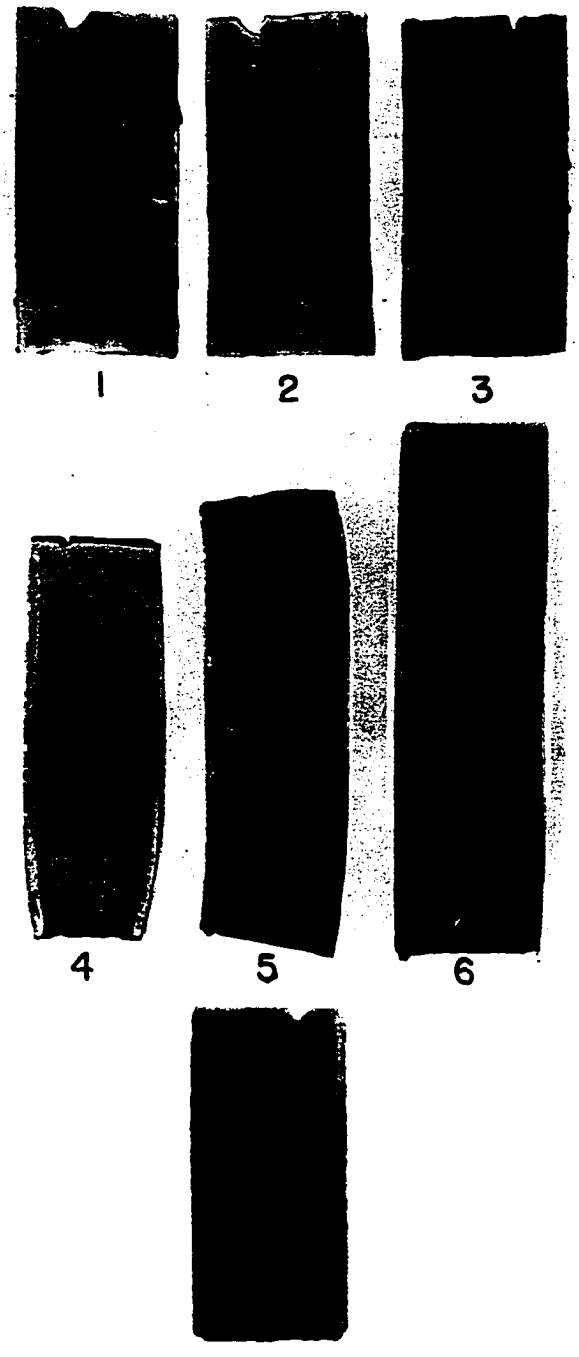
is of the order of one mole of amino acid per 35,500, the agreement is remarkable.

The optical rotatory constants of CMB were determined, and, as can be seen in Graph 10 where CMB is represented by the hexagon on the abscissa, this derivative is almost completely unfolded.

An ultracentrifuge run on a 1% solution of CMA in 0.1 M KCl at 25° C showed a small amount of an aggregate having a sedimentation coefficient near 10 x 10^{13} and the main material which had an S^{o}_{20} of 2.33 x 10^{13}.

The 35,500 species of β-lactoglobulin has a S^{o}_{20} of 2.87 x 10^{13}, while the 18,000 monomer has a S^{o}_{20} of 2.25 x 10^{13}. The CMA material was evidently composed of the 18,000 species with a slightly changed shape.

A series of gel electrophoresis runs on the CMA and CMB material disclosed that there were three species present in both CMA and CMB, and that each species in CMA had the same mobility as the corresponding species in CMB. This is shown in Figures 6 and 7 of Plate 4. Figure 6 is a starch gel electrophoresis in which CMA and CMB were run side by side under the same conditions used on the previous starch gel electrophoresis runs in Plate 1. A third band in between the two shown in Figure 7 appeared early in the dyeing procedure but would then disappear, leaving only the faint shadow shown in Figure 7.



CMB CMA

A cyanogum* gel electrophoresis of CMA and CMB in the same gel is shown in Figure 6. All three bands are clearly visible here, and do coincide. The heavy band which was first in the starch gel now appears to be last. It is suspected from this and other results that mobilities in the starch gel may be influenced by other factors as well as charge. β -Lactoglobulin A in cyanogum electrophoresis is shown in Figure 4, and β-lactoglobulin B in Figure 5. The samples under the notches on the left in Figures 4 and 5 are our samples of A and B respectively, and the samples run beside them in the same gel are samples of A and B furnished by Dr. Aschaffenburg. The gels in Figures 4, 5, and 6 were all run at the same time under identical conditions. We see from these three gel experiments that our samples of β-lactoglobulins A and B have identical mobilities with those isolated by Dr. Ashaffenburg, that there does not seem to be any β-lactoglobulin A in the B sample, or vice versa, and that the three bands in CMA and CMB do not seem to coincide with either A or B.

Since it was rather disconcerting to start with two proteins which were distinguished by their charge differences, and end up with two derivatives which had identically charged species, several gel electrophoresis experiments were

*The cyanogum #41 gelling agent as well as instructions for preparing it were all obtained from the E-C Apparatus Co., Swarthmore, Pa.

conducted on β-lactoglobulin A and B to determine the agent causing the three identically charged components. A cyanogum electrophoresis in 1% NH_3 at pH 11 showed that native A was still the faster after standing for 3 hours at this pH. The pH of the borohydride reduction was 10.5.

A series of cyanogum electrophoresis runs in 8 M urea using thioglycollic acid as the reducing agent are shown in Figures 1, 2 and 3 of Plate 4. Native β-lactoglobulin A or B was exposed to 4% thioglycollate for 16 hours at pH 8.5 in 8 M urea, conditions under which one would expect complete reduction (76), and the reduced protein was then electrophoresed in 1% thioglycollic acid adjusted to pH 8.5 with tris(trihydroxymethylaminomethane). The gel after 1 hour of electrophoresis is shown in Figure 1. Although only 1 band is visible in the photograph, the negative of the photograph showed there were apparently two bands present in both A and B. The electrophoresis at 4 and 6 hours is shown in Figures 2 and 3, where apparently 4 bands are present in both A and B. β-lactoglobulin A has been run on the side of the gel which is notched. The bands all coincide in the middle of the gel where A and B meet.

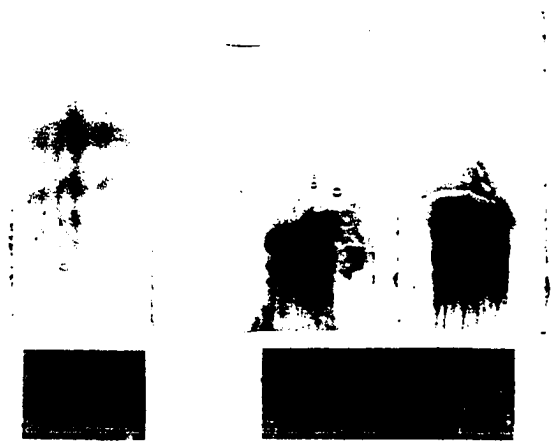
The cyanogum electrophoresis of A and B in 8 M urea alone is shown in the top two figures of Plate 5. In all cases β-lactoglobulin A is indicated by a notch in the gel. The dry native protein was dissolved in 8 M urea and the

Plate 5. Cyanogum gel electrophoresis of β-lactoglobulins A and B under varying conditions

Upper figure: A and B dissolved in 8 M urea at zero time, and electrophoresed in 8 M urea--.01 M pH 8.6 barbital buffer for 3 and 6 hours at 20 volts/cm.

Lower figure: A and B were dissolved in "cyanate free" 7 M urea, left for 30 hours in this reagent, and were then electrophoresed in 7 M urea--0. M pH 8.3 barbital buffer at 20 volts/cm. for band 10 hours. The same protein-urea solution was then made 0.1 M in potassium cyanate and electrophoresed under the same conditions for 8 hours

8 MOLAR UREA



electrophoresis started immediately afterwards. After 3 hours of exposure to 8 M urea during electrophoresis, A is migrating ahead of B, and no fine structure is visible. After 6 hours, however, what appear to be 3 bands are visible, and at least two of the bands in B coincide with those in A. A trace of a band in B is also visible which coincides with the fastest band in A.

It was pointed out in the section on amino acid analysis that urea has very recently been shown to contain trace amounts of cyanate in equilibrium with it in water solution. If we destroy the cyanate present in the urea according to Moore and co-workers (82b), we get the patterns in the two figures below. The two proteins were left standing in this "cyanate free" urea for 30 hours before the electrophoresis was started. The urea-cyanate equilibrium reaction started immediately of course, so that a small amount of cyanate was present, but the patterns show that the bands in A never coincide with those in B. If the same protein-urea solutions used in the two previous gels are now made 0.1 M in potassium cyanate, left at room temperature for 24 hours, and then run in the cyanogum the 4 or more bands in A now appear to coincide with those in B. It is not known whether the presence of more than one band is an artifact caused by traces of cyanate, or whether several species of different mobility are present in both A and B once they are unfolded

in urea solutions.

The cyanate could change the charge on a protein either by reacting with an positively charge species such as a lysine ϵ-amino group to form an uncharged group, or by reacting with a negative carboxyl group such as aspartic acid to form an uncharged group. In either case, the reaction is different in the two β-lactoglobulins, and is probably related to their difference in amino acid sequence.

The fact that cyanate reacts with the amino group of lysine complicates the analysis of a trypsin digestion, since trypsin cleaves bonds next to positively charged lysine residues. The fact that A and B have reacted differently with cyanate could also lead to further differences in the fingerprints of these two proteins.

Trypsin digestions of CMA and CMB produced by the borohydride method were run at various pH's between 8 and 9, at 25 and at 37° C. The best digestions were usually obtained at pH 8.0 at 25° C. Data obtained on the pH stat from some of the 25° C digestions is shown in Table 3. Both A and B were reduced and alkylated twice using the borohydride method and many trypsin digestions were run on each product in the pH stat. In every case CMB was split more by the enzyme in a given time interval than was CMA. The pH 8.6 trypsin digestions in Table 3 show this clearly, that of the order of three more bonds per 17,750 monomer were broken in

Table 3. Kinetics of trypsin digestion

pH	Time for first reaction to end	Half time of second reaction	Percent of total digestion	Dif. in digestion uptake moles/35,500
8.0	10 min.	86 min. (A)	58% (22 min., A)	
8.5	15 min.	75 min. (B)	77% (25 min., B)	5.1
	15 min.	97 min. (A)	62% (25 min., A)	
8.8	10 min.	113 min. (A)	39% (30 min., A)	
			Time A B	
8.6	18 min. (A)	168 min. (A)	20 min. 52% 60%	2.6
	14 min. (B)	89 min. (B)	30 min. 56% 65%	2.7
			60 min. 65% 76%	3.7
			120 min. 71% 88%	5.5

CMB than in CMA. This result was not just characteristic of one digestion or of one CM derivative, but of all those examined.

The trypsin digestions were commonly run at either pH 8.0 or 8.5 for from 2.5 to 3 hours at 25° C. This would result in about 85 to 90% of total digestion, and the product would give a good fingerprint. Raising the temperature to 37° C resulted in a slower second reaction. The reaction was found to go faster if the trypsin was added in two equal aliquots, one at zero time and one at 15 minutes. Representative chromatograms of the trypsin digestions of these CM derivatives are shown in Figures 11-14 of Plate 6.

The spots labelled 20A and 20B in Figures 11 and 12 do not coincide. Peptide spot 20A is almost opposite spot 12/19, while 20B is slightly below 12/19. The spots have been mislabelled in Figures 11 and 12 and the 20A and 20B should be reversed. Spot 20B shows up in almost every fingerprint of CMB while spot 20A has shown up in only 3 of more than 30 fingerprints, and was mistakenly labelled 20B in Figure 11. Spot 20B can also be seen in the fingerprint of CMB in Figure 14, while both spots are absent in CMA, Figure 13.

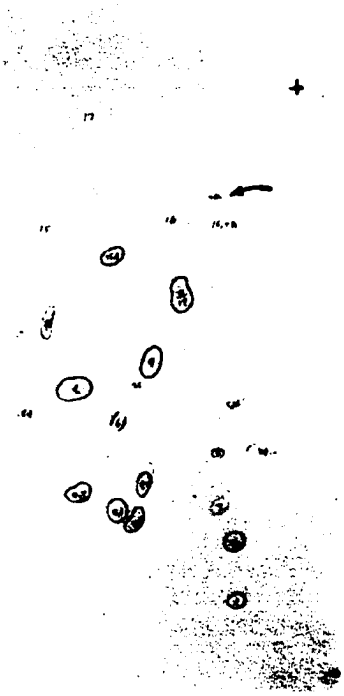
There is usually a difference between CMA and CMB in the region above spot 12/19 also. Figures 13 and 14 are

Plate 6. Fingerprints of borohydride reduced and alkylated β-lactoglobulins A and B

Figure 11. CMA, trypsin digestion
Figure 12. CMB, trypsin digestion
Figure 13. CMA, trypsin digestion
Figure 14. CMB, trypsin digestion

11

12



13

14

more representative of the usual result. Both CMA and CMB will show a faint streaky spot which is probably 3 peptides, labelled 15, 16, and 16.5A in Figures 13 and 14. A peptide spot below this will be present in B, but not A, and is labelled 16.4 in Figure 14. It is also present in CMB in Figure 12. A third spot above the three streaky peptides, labelled 22 in Figure 13, has shown up in chromatograms of both CMA and CMB. To summarize, the only peptide spots which have shown up reproducibly different in CMA and CMB fingerprints are the faint spots 20B and 16.4 present in CMB, and 20A present in CMA.

Longer electrophoresis has revealed that spot 12/19 is indeed two peptides, 12 and 19, present in both CMA and CMB. Spot 2.5, in dotted lines in CMB, Figure 14, has only shown up once before on a CMA fingerprint, and was not present in Figure 14.

Other stains than ninhydrin have also been used. The starch-KI reagent showed that the faint spot below peptide 9 in Figures 11 and 12 is a large peptide present in both A and B. Spots 4.2, and 4.3 also stained with this reagent, as did an extremely fast spot in the corner diagonally opposite from the origin which was present in both CMA and CMB, and may be an artifact. It is moving extremely far in the BAW direction for a peptide, and moves farther in the electrophoresis direction than does free lysine.

A spot in CMA in the same place that 20B is in CMB gave a Sakaguchi reaction for arginine. A few minutes after this, a second spot directly below this spot also turned pink. No corresponding color was present in CMB. Spots 1, 2 and 12/19 in CMB also turned color in CMB but not in CMA. Amino acid analyses reveal argine in both CMA and CMB in these three peptides, however. Spots 4, 6 and 9 also showed color with the reagent in both CMA and CMB.

Spots 4.1, 15.1 and the faint spot below spot No. 9 all gave a color reaction with the Sanger reagent for histidine and tyrosine.

Spots 12/19, 9, 6, 4.2, 20B, 15, 15.1, 16, and 16.4 all showed up under U.V. flourescence, but no new spots were noticed using this technique.

Three complete amino acid analyses of every spot on both fingerprints have been made. The results left much to be desired, since in most of the amino acid chromatograms from each peptide the amino acid spots were just observable, even when material from two fingerprints was pooled. One would expect about 0.1 μ moles of each amino acid from a fingerprint using 1.5 mg of protein, and the final hydrolysis gave about 1/4 th this amount. The amino acids could have been destroyed in the formation of humin in the hydrolysis tube, or by reacting with the breakdown products of ninhydrin decomposition. Ninhydrin eluted from the paper always left

a tarry mass in the bottom of the hydrolysis tubes. The amino acid chromatogram method itself proved to be extremely sensitive, however.

Bearing in mind the limitations of the procedure, the three analyses showed that all the peptide spots from CMA had the same composition as the corresponding spots in CMB with the following exceptions.

Spot 20B in CMB has been assigned the tentative composition lys, gly, cys. Spots 20A, 16.4, and the spot corresponding to 20B in CMA have not been analyzed.

Spot 4.1 on one analysis showed a faint alanine and an amino acid which could have been either valine or methionine in CMA, while CMB showed no valine or methionine and a darker alanine spot. Spot 4.0 showed a faint alanine in CMA and a darker alanine in CMB. In the two subsequent analyses, spots 4 and 4.1 were eluted together since they had not separated sufficiently. The combination of 4/4.1 contained neither valine nor alanine on one analysis, while CMB contained a distinct alanine and CMA a distinct valine and a distinct aspartic residue in the next analysis. The nearby spot 4.2 contains both valine and alanine, while spot 4.3 has shown an aspartic residue in CMA on one analysis.

In summary, fingerprints of the tryptic digestions of CMA and CMB have shown that they differ in the very faint peptide spots 20A, 20B, and 16.4 as well as one or two

arginine containing spots in the part of the CMA fingerprint corresponding to 20B. Spot 4.1 may contain a valine residue in CMA and an alanine residue in CMB, while spot 20B has been assigned the tenative composition lys, gly, cys.

IV. OPTICAL ROTATION STUDIES

A. Introduction

The optical rotation of a protein has been used as a sensitive but empirical measure of its folding for a long time. Native globular proteins show residue rotations near -30 to -80°. Their denatured form usually has a residue rotation of -100 to -120°, as can be seen from Table 4. The residue rotation is related to the specific rotation by

$$[R] = [\alpha] \cdot \frac{\text{Residue M.W.}}{100} .$$

The rotary dispersions of both native and denatured globular proteins are simple. This means that the wave length dependence of the rotation can be represented by a one term Drude equation of the form

$$[R] = \frac{a_o^* \lambda_c^2}{\lambda^2 - \lambda_c^2} , \qquad (1)$$

where λ_c and a_o are constants. The parameters $[\alpha]_D$ or $[R]_D$ and λ_c, which completely characterize a simple dispersion, are the constants usually reported for such dispersions. The

*Since we will want both [R] and a_o independent of solvent, we have incorporated the refractive index term, $(n^2 + 2)/3$, into a_o. Throughout the remainder of the thesis all residue rotations have been corrected to make them, and therefore a_o, relative to water.

Table 4. Optical constants of proteins

Protein	$[R]_D^F$	$[R]_D^U$	$\Delta[R]_{D\,U \rightarrow F}$	λ_c^F	λ_c^U	a_o^F	b_o^F	$b_{o\,calc'd}^F$ [b]	Ref.
				Globular proteins					
Ribonuclease	-90.7° [a]	-126.5° [a]	35.8°	233	220	-610°	----	-128°	(75)
β-Lactoglobulin	-30.9	-124.5	93.6	251	225	-208	----	-83	(75)
Insulin	-36.5	-98.3	61.8	267	226	245	----	-144	(75)
Ovalbumin	-33.9	-106.8	72.9	270	226	228	----	-141	(75)
Bovine serum albumin	-69.6	-122.0	54.4	265	221	-468	----	-262	(75)
γ-Globulin	-49.4	-92.7	43.3	200-220	---	-332	----	----	(75,47)
Lysozyme	-53.3	-96	42.7	254	---	-358	----	-154	(75)
Pepsin	-64.5	-87	22.5	213	---	-434	----	----	(75,47)
Chymotrypsin	-71.9	-115.9	44.0	241	220	-484	----	-140	(75)
Chymotrypsinogen	-84.6	-120.9	36.3	239	224	-569	----	-154	(75)
				Fibrous proteins					
Fibrinogen	-63° [c]	-120° [c]	57	256	213	-390°	-264°	-178°	(99,75)
L-Meromyosin I	-14° [c]	-131° [c]	117	---	212	18.7	-830	----	(86)
Tropomyosin	-17° [c]	-131° [c]	114	---	213	-6.4	-780	----	(86)
Myosin	-31° [c]	-119° [c]	88	---	218	-129	-470	----	(86)

[a] Relative to H_2O at 25° C.

[b] By equation 3.

[c] Calculated from dispersion data.

fact that a dispersion appears simple usually implies only that the rotation being measured is fairly far from the active absorption band giving rise to the rotation.

The increase in positive rotation encountered in going from the unfolded to the folded protein has been postulated to arise from the formation of an α-helix. An α-helix would be expected to have a configurational rotation because of its asymmetric structure, or alternatively because it restricts rotation about the peptide bonds, producing a different asymmetry in the residues.

There have been two recent theoretical treatments of the optical rotation due to an α-helix, one by Moffitt (60) and one by Fitts and Kirkwood (27). The results of both treatments have been incorporated into a more general theory (61), which predicts that if the normal dispersing terms are separated from the anomalous dispersing terms, the residue rotation of the helix will be given by

$$[R] = \frac{a_o \lambda_o^2}{(\lambda^2 - \lambda_o^2)} + \frac{b_o \lambda_o^4}{(\lambda^2 - \lambda_o^2)^2} \qquad (2)$$

where $[R]$ is the residue rotation, and λ the wavelength. For a protein the residue molecular weight is taken as $\frac{\text{(M.W. protein)}}{\text{(no. of residues)}}$. The symbol λ_c will be reserved for the experimentally obtained quantity in the one term Drude

equation, and the symbol λ_o for the constant in the above equation.

This equation was previously suggested by Moffitt and Yang on a more empirical basis (62). The coefficients a_o, b_o, and λ_o cannot be predicted by the theory as yet.

Doty and coworkers in a series of papers have attempted to evaluate a_o, b_o, λ_o, and therefore $[R]_D$ experimentally for an α-helix (22, 23, 99).

Doty found that silk fibroin which is a fibrous protein and two polyamino acids, poly-L-glutamic acid and poly-γ-benzyl-L-glutamate are completely helical in organic solvents such as dioxane, chloroform, or m-cresol (22). The helix formed consisted of a single long rod. The specific rotation at any one wave length could not be used as a measure of the helical content since it varied with the solvent even after correcting for changes in index of refraction (23). The dispersion of the helical form was always anomalous, while the random chain form followed a one term Drude expansion.

The polyamino acid, poly-L-glutamic acid, could be changed from a random coil to a helix in dioxane by changing the pH from 6.6 to 4.7, i.e., by partially protonating the γ-carboxyl group. The change in residue rotation accompanying

this change was 106°.* By combining parts of both the random and the helical dispersion curves it was found that poly-L-glutamic acid would exhibit simple dispersion up to 40% helix, with an apparent change in λ_c from 212 mμ for the helical form to 254 mμ at 30% helix.

When globular proteins were examined none of them exhibited anomalous despersion in any solvents. It was concluded that the presence of proline residues, side chain interactions and cystine bridges prevented the formation of more than 40% α-helix (99).

Both Doty (23) and Elliot (25) found that the residue rotation of an α-helix alone could be obtained by a method of graphical extrapolation. Table 5 summarizes their data for the residue rotation of an α-helix in organic solvents.

It is possible that the organic solvents stabilize the helix by interacting with the non-polar portions of the side chain, while the polar groups are "hidden" by their hydrogen bonds in the helical portion of the molecule. It is significant that these polyamino acids do not form helices in aqueous solutions.

Both Doty (23) and Moffitt (62) found the anomalous dispersion of poly-γ-benzyl-L-glutamate to fit equation 2

*All rotations reported have been made relative to water at 25° C by applying the Lorentz correction, $[R]_{H_2O} = [R]_{\text{solvent}} \times 3/(n^2 + 2)_{\text{solvent}} \times (1.776+2)/3_{H_2O}$.

Table 5. Residue rotation of an α-helix

Polyamino acid	$[R]_D^{Helix}$	Solvent	Reference
γ-benzyl-L-glutamate	+113°	dioxane	(23)
γ-benzyl-L-glutamate	+93°	chloroform	(23)
L-leucine	+125°	benzene (60° C)	(25)

and found λ_o = 212 mμ to be the best fit. This is the same value that was previously found for λ_c in the normal dispersion of the random chain form of this polyamino acid.

The values of a_o' = +680°, and b_o' = -630° were also determined from the dispersion of poly-γ-benzyl-L-glutamate.* Although a_o' is dependent on solvent and the rotation of the amino acid residues, b_o' ranges from -600 to -650° for the right handed helical forms of poly-γ-benzyl-L-glutamate, poly-L-glutamate, poly-L-alanine, poly-L-leucine and poly-L-lysine (71b). The value of b_o' for the randomly coiled forms is 0°.

Unfortunately, none of the globular proteins in aqueous solution show anomalous dispersion, so we cannot estimate their helical content from b_o, which must be close to zero.

*$a_o = (n^2+2)/3\ a_o'$, $b_o = (n^2+2)/3\ b_o'$. Since all our values are relative to H_2O, $(n^2+2)/3$ = 1.259.

At least four ways have been proposed for estimating the helical content of a protein from optical rotation data. The estimates for β-lactoglobulin range from 10% to 100% helix. The following methods will be reviewed and their limitations discussed:

Method one: the first involves using the change in the specific rotation at the sodium D line going from the unfolded to the folded protein as an index of the helical content, using $+100^{o}$ as the change in going from a random chain to a helix, Yang and Doty (99).

$$\% \text{ Helix} = \frac{[\alpha]_D^F - [\alpha]_D^U}{+100^{o}} \times 100. \qquad (3)$$

Method two: we can use residue rotations instead of specific rotations in equation 3, and let the change for β-lactoglobulin, which is $+93^{o}$, be a 100% helix to random coil transition, Schellman and Schellman (75).

$$\% \text{ Helix} = \frac{[R]_D^F - [R]_D^U}{+93^{o}} \times 100. \qquad (4)$$

Method three: the third method uses the change in λ_c as measure of the helical content. Since Doty has determined the change in λ_c as poly-L-glutamate goes from random chain to a mixture with 30% helix, we can use this to set up a linear scale, Yang and Doty (99).

$$\begin{array}{l} 254\ m\mu\ \text{--}30\%\ \text{helix} \\ \underline{212\ m\mu\ \text{--}\ 0\%\ \text{helix}} \\ \Delta\ \lambda_c = 42\ m\mu; \quad \dfrac{42\ m\mu}{30\%} = \dfrac{140\ m\mu}{100\%} \end{array}$$

and

$$\%\ \text{Helix} = \frac{(\lambda_c)_F - (\lambda_c)_U}{140\ m\mu} \qquad (5)$$

where F = folded state, and U = unfolded.

Method four: if our protein shows anomalous dispersion, we determine b_o' for it, letting the b_o' for a unique protein, light meromyosin I be 100% helix. The constant b_o' is zero for a random chain, is -630° for poly-γ-benzyl-L-glutamate, and is -660° for light meromyosin I, Szent-Gyorgyi (86).

$$\%\ \text{Helix} = \frac{b_o'}{-660^\circ} \quad x \quad 100. \qquad (6)$$

Methods one and two are similar, since they both involve using the change in rotation at one wavelength as a measure of the amount of helix. Method two, which uses residue rotations, is independent of residue weight, and would seem to be the better of the two.

Methods three and four both involve using data from the dispersion, rather than the rotation itself. Values from both methods are compared in Table 6. If we consider the contribution of the b_o term to λ_c, the similarity of the two methods is easily explained. Let us represent the two

Table 6. Helical estimates from b_o' calculated from equation 4

Protein	$\frac{\Delta \lambda c}{140 \text{ m}\mu}$	$\frac{b_o'}{-660^o}$
Ribonuclease	14%	15%
β-Lactoglobulin	28%	10%
Insulin	39%	18%
Ovalbumin	39%	17%
Bovine serum albumin	38%	32%
Lysozyme	30%	18%
Chymotrypsin	21%	17%
Chymotrypsinogen	19%	18%
Silk fibroin	25%	11%
Fibrinogen	31%	22%

term Drude expansion by equation 2

$$[R] = \frac{a_o \lambda_o^2}{\lambda^2 - \lambda_o^2} + \frac{b_o \lambda_o^4}{(\lambda^2 - \lambda_o^2)^2}.$$

Expanding the second term

$$\frac{b_o \lambda_o^4}{(\lambda^2 - \lambda_o^2)^2} = b_o\left(\frac{\lambda_o}{\lambda}\right)^4 - 2b_o\left(\frac{\lambda_o}{\lambda}\right)^6 + \cdots$$

we can disregard everything but the first term since $\frac{\lambda}{\lambda_o} < 1$.

Equation 2 then becomes

$$[R] = \frac{a_o \lambda_o^2}{\lambda^2 - \lambda_o^2} + b_o\left(\frac{\lambda_o}{\lambda}\right)^4 = \frac{a_o \lambda_o^2\left(1 + \frac{b_o}{a_o}\frac{\lambda_o^2}{\lambda^2} - \frac{b_o}{a_o}\frac{\lambda_o^4}{\lambda^4}\right)}{\lambda^2 - \lambda_o^2}$$

$$= \frac{a_o \lambda_o^2}{\lambda^2 - \left(1 + \frac{b_o}{a_o}\right)\lambda_o^2 + \left(1 + 2\frac{b_o}{a_o}\right)\frac{\lambda_o^4}{\lambda^2} - \ldots}$$

$$[R] = \frac{a_o \lambda_o^2}{\lambda^2 - \left(1 + \frac{b_o}{a_o}\right)\lambda_o^2} \quad \text{if } o < \frac{b_o}{a_o} < 1; \quad \lambda_o = 212 \text{ m}\mu \quad \lambda \Rightarrow 350 - 750 \text{ m}\mu \tag{7}$$

From equation 7 we find that

$$\lambda_c^2 = \left(1 + \frac{b_o}{a_o}\right)\lambda_o^2 \tag{8}$$

so that as the amount of anomalous dispersion measured by b_o increases, there is an apparent increase in λ_c^2. Equation 4 has been used to calculate values of b_o' assuming λ_o = 212 mμ, and these have been tabulated in the last column of Table 7. The amount of helix, computed from $\frac{b_o}{-600}$ and compared with the values computed from $\frac{\Delta\lambda_c}{140 \text{ mu}}$ is shown in Table 6. The values agree quite well for about half the proteins in Table 6. The half for which the two estimates do not agree all have low values of a_o (Table 4). The values of b_o' computed in this fashion are evidently influenced by the a_o value of the native protein which is involved in equation 4.

We can thus limit our problem to contrasting methods one and two which depend on the first dispersion constant*, a_o, with methods three and four that depend on the second dispersion constant, b_o.

The amount of helix when calculated from b_o data is invariably smaller than that calculated from a_o data, as can be seen from the first two columns of Table 7. The b_o method has been shown by Doty to apply to long helices in organic solvents, where b_o' is zero for the unfolded form, and near -630^o for the helical form.

The value of 212 mμ was taken as the λ_c^u of the unfolded form in computing $\Delta\lambda_c$ in Table 7.

The estimates of helix given by the two optical methods can be compared with other estimates of the amount of α-helix in Table 7. The amount of hydrogen bonding is related to the amount of α-helix by the following equation:

(% H-bonded residues) x F x (L/L-4) = % helical residues. (9)

L is the average length of the helices in residues and F is the fraction of the total hydrogen bonds which are in the α-helix. The L/L-4 term arises because there are four hydrogens at the N terminal end of every helix which are not bonded. The smallest α-helix that can form is then five residues long.

Table 7. Comparison of hydrogen bonding and helical data

Protein	% α-Helix: b_o / −660°	% α-Helix: Δλc / 140 mμ	% α-Helix: Δ[R] / 117°	% α-Helix: X-ray	% Hydrogen bonding: H_2O adsorption	% Hydrogen bonding: D_2O exchange	% α-Helix forming residues: Avg. comp.	% α-Helix forming residues: Sequence
Globular proteins								
Ribonuclease	15	14	31			38[a]	55	55
β-Lactoglobulin	10	28	80		74		70	
Insulin	18	39	53	(50)		59	67	75
Ovalbumin	17	41	62		67		75	
Bov. serum alb.	32	38	45		66	45[b] 51[c]	76	
γ-Globulin		0-6	37				55	
Lysozyme	18	30	37				69	
Pepsin		0	19				56	
Chymotrypsin	17	21	38					
Chymotrypsinogen	18	19	31				51	
Myoglobin				70		51	76	
Fibrous proteins								
Fibrinogen	32		49				69	
L. Meromyosin I	100		100					
Tropomyosin	94		97				87	
Myosin	56		75				75	

[a] 3 hydrogen bonded tyrosines have been subtracted from this data. The figure would have been 40% for all hydrogen bonding.

[b] Computed from I.R. data; should only include peptide imido nitrogens.

[c] D_2O exchange by density detm. of Bresler (13).

Titration data has shown that most of the polar side chain groups in the globular proteins are not hydrogen bonded, so that almost all the hydrogen bonding is in the peptide carbonyl and imide group, which is the group which hydrogen bonds in the α-helix. Most of the hydrogen bonding may then be intrahelical in the globular proteins. We will assume that both F and L/L-4 are close to unity and compare hydrogen bonding directly with the amount of helix. If they do differ from unity, each effect does tend to cancel the other as equation 9 shows.

The hydrogen bonding data in Table 7 was determined in three ways. The values of hydrogen bonding from water adsorption were computed using the analytical data in Tristam's review (96) coupled with the water adsorption data of Bull (16) using a method somewhat similar to that used by Pauling (71a).

Bull found that the water adsorption of proteins followed a "type three isotherm" which could be divided into three parts by points A_1 and A_2. Point A_1 is the water adsorbed in the B.E.T. monolayer, while point A_2, which can be estimated closely from the minimum of a plot of A/x vs. x, is claimed by Vull to be the point at which the polar groups of the molecule are saturated with water. (The amount of water adsorbed is A, and x is the relative vapor pressure).

Table 8 shows that the number of moles of water adsorbed in the B.E.T. monolayer is slightly greater than the number of oxygen-containing polar groups on the side chains, but is less than the total number of side chain polar groups except in bovine serum albumin, where it is slightly greater. At the B.E.T. point the peptide carbonyls which are not internally hydrogen bonded have probably not bound appreciable water.

However at point A_2 some peptide groups do not seem to be binding water. We can easily calculate the number of residues which are not hydrogen bonded if we know how many water molecules are bound by each polar group, and by the peptide bond. Models were made of all the polar groups and of a peptide and it was found that all of the groups listed in footnotes b and c of Table 8 accommodated one water molecule with two hydrogen bonds, except the arginine side chain which could fit three water molecules. The peptide bond also only bound one water molecule.

Using this data, we subtract the calculated number of water molecules adsorbed at saturation by the side chains (row 5) from the experimental quantity in row 2 to get the water molecules adsorbed by the peptide carbonyls. This figure is then divided by the total number of residues (row 6) to give the percent of the peptide carbonyls which are not internally hydrogen bonded.

Table 8. H_2O adsorption data

Row		Bov. serum alb. Moles/10^5 g protein	Ovalbumin Moles/10^5 g protein	β-Lactoglobulin Moles/10^5 g protein
1	Water adsorbed in first monolayer B.E.T. plot[a] A_1	474	341	370
2	Water adsorbed when "polar groups are saturated"[a] A_1	810	694	705
3	Number of O containing polar groups on side chains[b]	395	271	341
4	Number of polar groups on side chains[c]	454	367	449
5	Calc'd. number of water molecules at saturation of side chain polar groups[d]	522	433	482

[a]From data of Bull (16).

[b]Includes ser, thr, tyr, asp, glut, asp-NH_2, glut-NH_2, COOH; analytical data from Tristam (96).

[c]Includes all in footnote b plus lys, his, arg, try.

[d]Includes all in footnote c but arg counted 3X.

Table 8. Continued

Row		Bov. serum alb. Moles/10^5 g protein	Ovalbumin Moles/10^5 g protein	β-Lactoglobulin Moles/10^5 g protein
6	Total number of residues	522	433	482
7	Percent residues not H-bonded from row 5	$\frac{810-522}{852}$ =34%	$\frac{694-433}{796}$ =33%	$\frac{705-482}{852}$ =26%
8	Percent H-bonding from row 5	66%	67%	74%

This method would not be valid if a large fraction of the polar side chains are hydrogen bonded.

Titration data shows that two carboxyl groups per mole in β-lactoglobulin (87) and 7-8 tyrosine groups in ovalbumin (20, 17) are not titratable in the normal range for these groups, nor is their titration reversible. These polar groups would be expected to be in the interior of the molecule, and not available to bind water (46). All the other polar side chain groups in these two molecules titrate normally and reversibly. All the polar groups in bovine serum albumin titrate reversibly, but the pK's of the carboxyl, amino and phenolic groups are all abnormal (90), suggesting they might be involved in some sort of internal bonding and not available to bind water. However, further evidence from deuterium exchange to be explained shortly would seem to show that most of the hydrogen bonding in serum albumin is indeed in the imido groups, and not in the side chains.

Comparison of the amount of hydrogen bonding calculated from water adsorption with the other values in Table 8 shows that the figures seem to be of the right order of magnitude. If any meaning can be attached to them, it is that these three proteins have almost the same amount of hydrogen bonding.

Deuterium exchange experiments can also be utilized to

give estimates of hydrogen bonding which are far more precise than the values just calculated from water adsorption. It was found by Linderstrom-Lang and his collaborators that hydrogens which were not hydrogen bonded would exchange with deuterium in D_2O solution within 0.5 minutes, while poly-DL-alanine, a helical polymer of chain length 30, had 26 hydrogens which exchanged at a much slower rate (41). One would expect just 26 hydrogen bonds from a helix of 30 residues.

All the values from deuterium exchange in Table 7 were determined by Linderstrom-Lang using density measurements (41, 8, 10), except for the two values for serum albumin which will be discussed separately. The number of hydrogens which exchanged slowly were divided by the total number of residues in the molecule.

Since the infrared absorption of the peptide imido nitrogen shifts as it is deuterated, the I.R. spectrum of a protein can also be utilized to give the number of hydrogens which exchange with deuterium. The difference in this case is that we are selecting the imido group from all the other hydrogen bonds in the molecule. We have used the two I.R. spectra of bovine serum albumin published by Lenormant and Blout (57) in the following manner.

The N-H bending frequency at 1550 cm characteristic of a cis peptide bond in H_2O (56) shifts to 1450 cm^{-1} as the

protein is deuterated. If the protein is denatured by base in D_2O the optical density at this band increases when the hydrogen bonded imido hydrogens exchange. Then; (O.D. denatured) - (O.D. native) is proportional to the imido hydrogens which do not exchange, while (O.D. denatured) is proportional to the total imido hydrogens, and

$$\text{Percent H-bonding} = \frac{\text{(O.D. denatured) - (O.D. native)}}{\text{(O.D. denatured)}} \times 100.$$

All O.D.'s were measured using the minima at 1375 cm^{-1} and 1500 cm^{-1} for a baseline. The value found was 46% H-bonding, and probably has an uncertainty of 10% because of the uncertainty in the base line. It is the percent of the peptide imido hydrogens which do not exchange, and it can be compared with the value of 51% found by Bresler (13) for the percent of all hydrogen bonds which do not exchange. Bresler used a procedure similar to that of Linderstrom-Lang. The closeness of these two estimates shows that most of the hydrogen bonding in serum albumin is in the peptide group and not in the side chains. The figure from I.R. data would be more closely related to helical content. The fact that the number of groups exchanging in serum albumin is virtually independent of time at room temperature (13) makes this one I.R. measurement valid.

The estimate of helix labelled "X-ray" in Table 7 is the actual amount of α-helix in the myoglobin molecule in the

solid state (50). The length of the helical regions in the structure was divided by the length to be expected if the entire chain was helical. This number, 70%, is much larger than the amount of intramolecular hydrogen bonding in solution inferred from deuterium exchange data, which is 51%. The molecule could be composed of helices with an average length of twenty residues, or it is possible that the molecule contains more helix when it is in the solid state. A preliminary estimate of the helical content of insulin is also listed (71b).

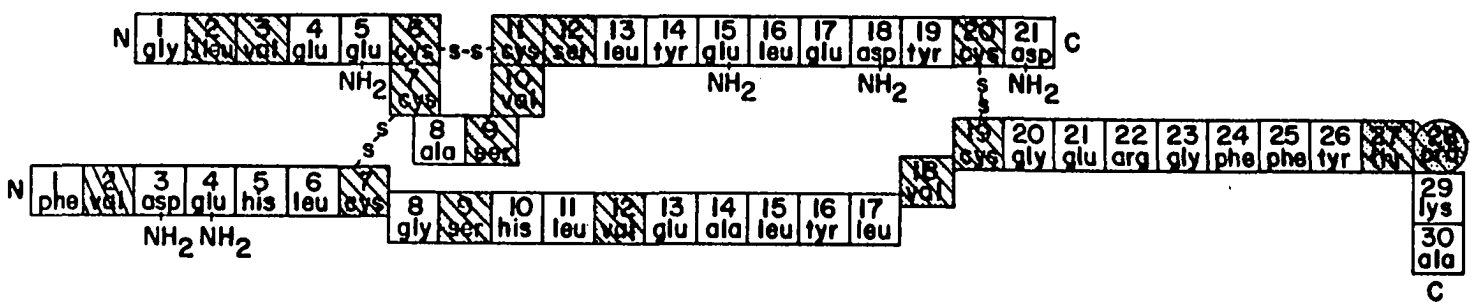
The fifth estimate of helical content in Table 7 was made by computing the percent of helical forming residues in the proteins using the analytical data of Tristam (96). Blout et al. (12) have shown by I.R. studies of dried films of various polyamino acids that residues which were disubstituted on the β carbon atom (valine, isoleucine) or those which had a hetero-atom on the β carbon (cystine, cysteine, serine, threonine) would not form an α-helix in the film, probably because of steric hindrance. The proline residue also cannot normally fit in an α-helix because of its ring structure and lack of an amido hydrogen. The column on the left in Table 7 is 100% minus the percentage of these seven amino acids in the protein, computed from the analytical data.

Models were made of these amino acids in the α-helix

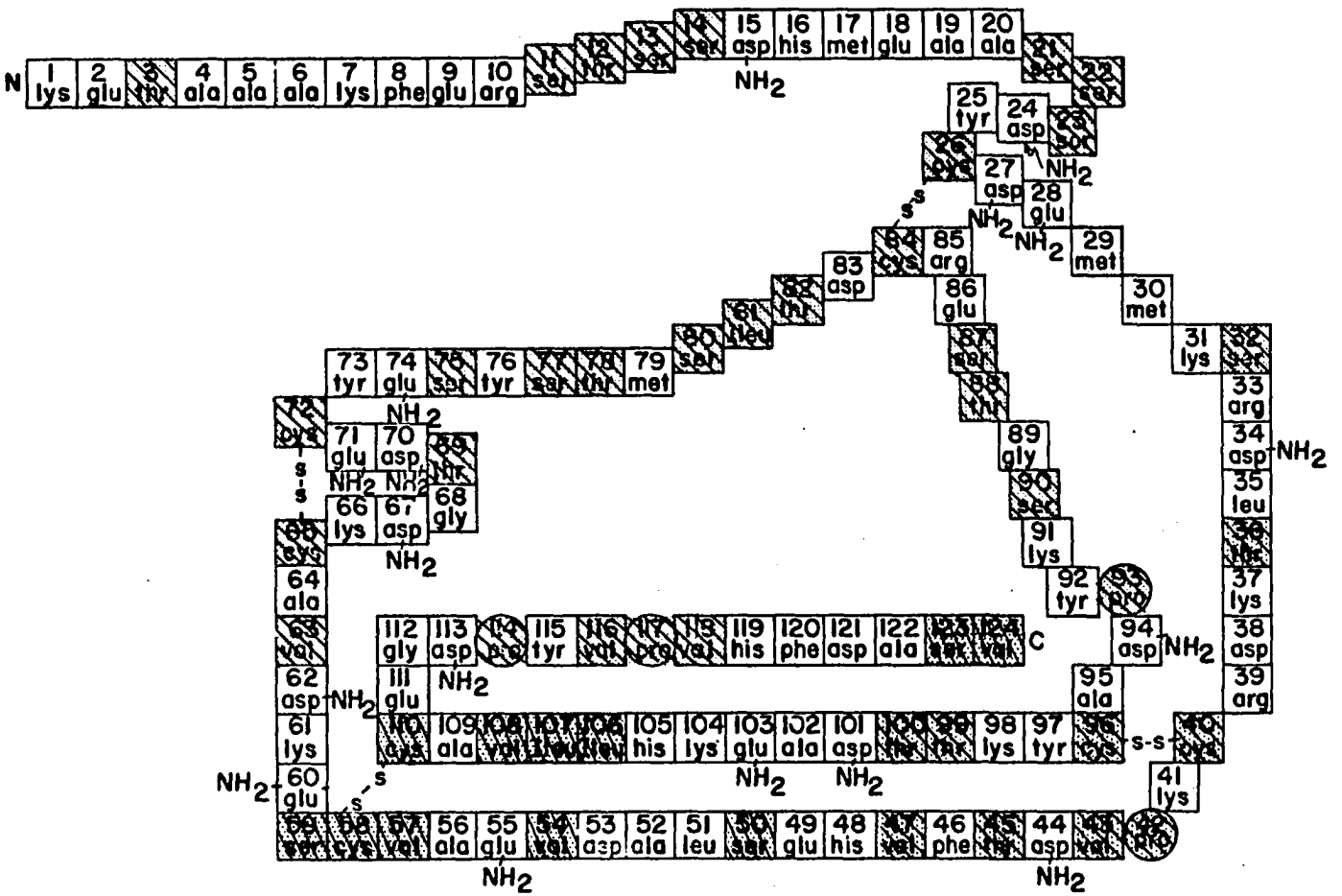
and it was found that all but proline could be formed into a helix. Two nearby peptide carbonyls in the helix completely interferred with rotation about the β-carbon atom in the other six amino acids and the fit between these two groups and the β-carbon was very close. Valine and isoleucine were the most hindered in this respect, followed by threonine, then cysteine, with serine the least hindered. The only way a cystine bridge could be incorporated intrahelically was by having three other residues between the two half cystines. If the cystine was near the end of the helix considerably more freedom was allowed, and zero to three residues could be between the two half-cystines.

The amino acid sequences of insulin and ribonuclease are shown in Graphs 2 and 3. The non-α-forming residues have been shaded while proline residues are indicated by a shaded circle. There seem to be obvious places in the two structures where α-helices would be likely to form and other places where one would not expect a helix. Examples of the former would be residues 1-26 in the B chain of insulin or residues 1-10 in ribonuclease, while residues 6-12 in insulin, 106-110 and 114-118 in ribonuclease would be examples of the latter. Residues 6-12 in insulin can fitted in a left handed helix, however (59).

We may hypothetically fold both structures into α-helices, trying to form the longest and therefore most stable



Graph 2. Amino acid sequence of insulin

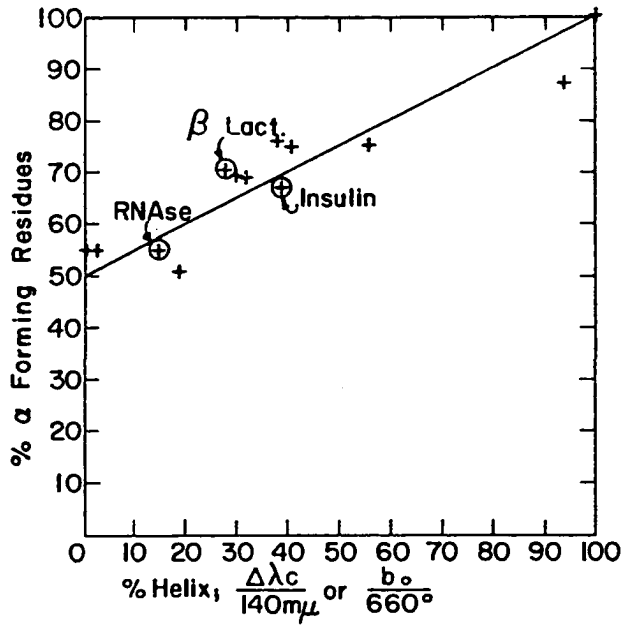


Graph 3. Amino acid sequence of ribonuclease

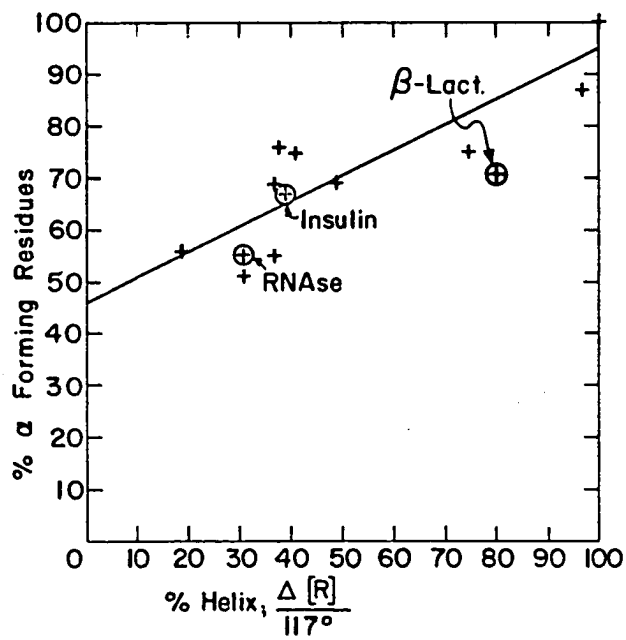
helices. In order to get the required number of hydrogen bonds, 30 in insulin and 50 less 3 phenolic hydrogens in ribonuclease (68b), insulin must be made 75% helical and ribonuclease 55% helical. These figures have been entered in the last column of Table 7. The figures are only approximate since there was not one unique structure in either case. The average length of the helices formed was about 20 residues. We have assumed all hydrogen bonds but three in ribonuclease to be helical. These results are somewhat higher than those given by the two optical methods in Table 7.

A correlation between the amount of helix calculated by the two optical methods and the percentage of α-forming residues is suggested by the curves in Graphs 4 and 5 where the amount of helix calculated from a_o data in Graph 4 or b_o data in Graph 5 is plotted against the percent of α-forming residues. In both cases we obtain approximately linear relationships. The curves imply that no α-helices are formed until the composition is near 50% α-forming residues but that the degree of helix approaches 100% as the composition approaches 100% α-formers. The wide scatter of the individual points is presumably due to the non-random distributions of the α-forming residues.

In summary, several methods of treating optical rotation data to get helical content have been reviewed. No clear cut



Graph 4. Helical estimates vs. composition: methods 3 and 4



Graph 5. Helical estimates vs. composition: method 2

decision could be made as to the suitability of any one method because of the lack of model compounds with specified amounts of α-helix in aqueous solutions. The polyamino acids which might have been ideal models only formed helices in non-aqueous solvents such as benzene. In this respect the polyamino acid probably acted as a typical colloid and formed a structure which had all its polar peptide carbonyl and imido groups buried only when it was in a nonpolar solvent. Three estimates of the helical content of β-lactoglobulin using various methods are:

10% helix ($b_o/660^o$)

30% helix ($\Delta\lambda_c/140$ mμ)

80% helix (Δ [R]/117^o)

The method using λ_c has been shown to depend on the amount of anomalous dispersion present (which is better estimated by b_o) and this method has no theoretical basis.

The method using b_o seems to be independent of solvent, and works very well for long helices in organic solvents and for completely unfolded molecules in any solvent.

The methods using Δ[R], Δ[α], or a_o are only applicable to one reference solvent, and give higher estimates of helical content for the globular proteins than do either the b_o or λ_c methods.

The amount of helix in β-lactoglobulin judged by the amount of α-forming residues present is 40-50%.

B. Experimental

A series of rotation measurements were made on β-lactoglobulin B, which was the first of the two proteins to be isolated. The results for the most part seemed characteristic of β-lactoglobulin in general and probably do not reflect a difference between A and B. All measurements unless otherwise specified were taken using crystals of B that had not been freeze dried, and were taken at the isoelectric point of the solution which was approximately pH 5.3, at 25° C. The uncertainty in the determination of $[\alpha]_D$ was probably about ± 1°, and the uncertainty in λ_c about ± 1 mμ.

The specific extinction coefficient of β-lactoglobulin B at 280 mμ in 0.1 M KCl determined from an average of 8 readings is $\epsilon^{1\ cm}_{1\%} = 10.05 \pm 0.12$. The concentration was determined by drying to constant weight under vacuum at 105° C. Extreme care had to be used to prevent surface denaturation of the sample during transfers of solution before the O.D. measurements were taken.

1. Dispersions

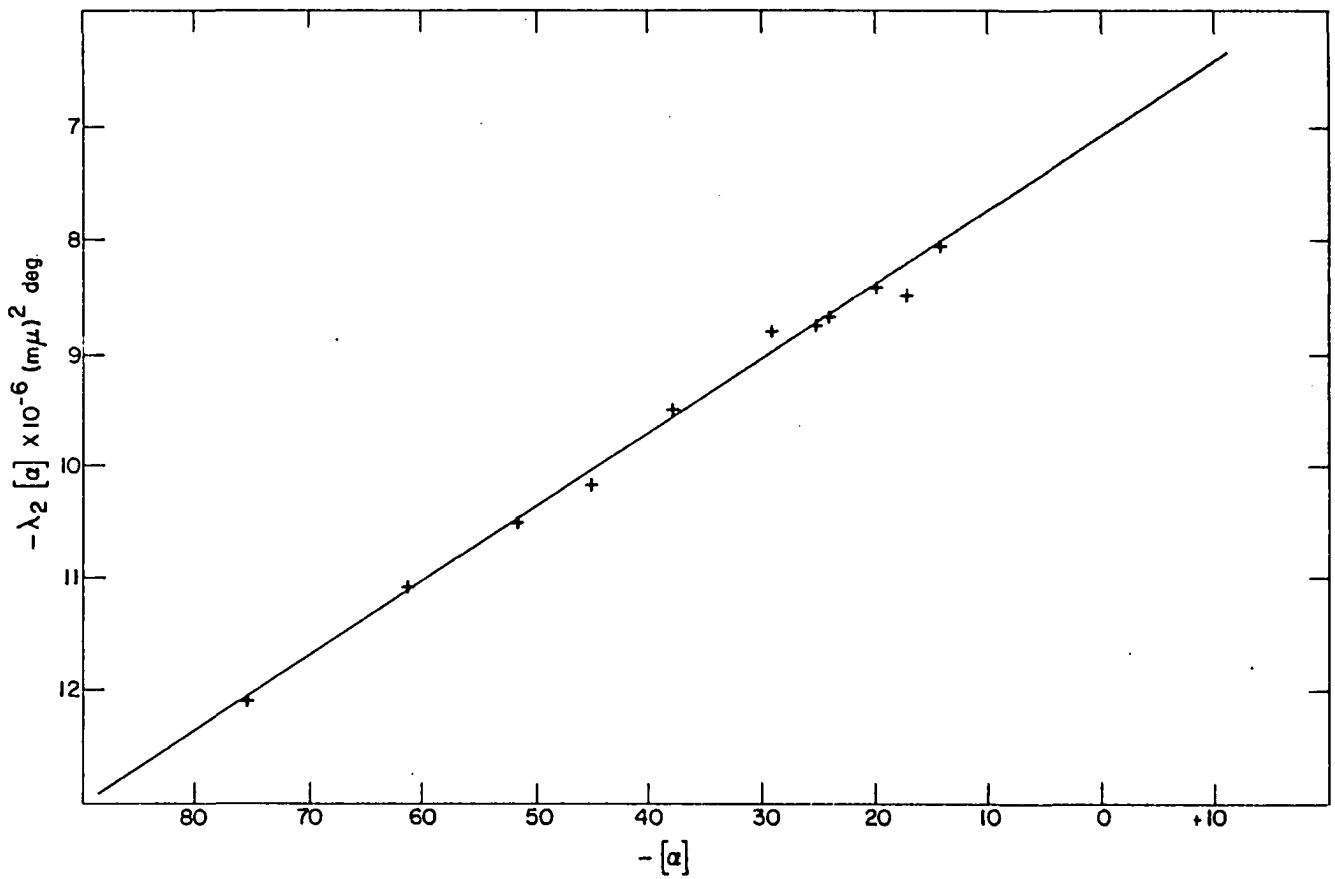
A 1% solution of β-lactoglobulin B in 0.1 M KCL was prepared, centrifuged and the rotary dispersion taken immediately. The concentration was obtained from the optical density at 280 mμ. The dispersion follows a simple

one term Drude expansion, $[R] = \frac{a_o \lambda_c^2}{\lambda^2 - \lambda_c^2}$ as can be seen from Graph 6, with the values of 256.5 mu for λ_c and -25.2° for $[\alpha]_D$ or -28.6° for $[R]_D$. If the protein solution was allowed to stand after centrifugation or to surface denature $[\alpha]_D$ and λ_c could become as low as -33° and 244 mμ respectively.

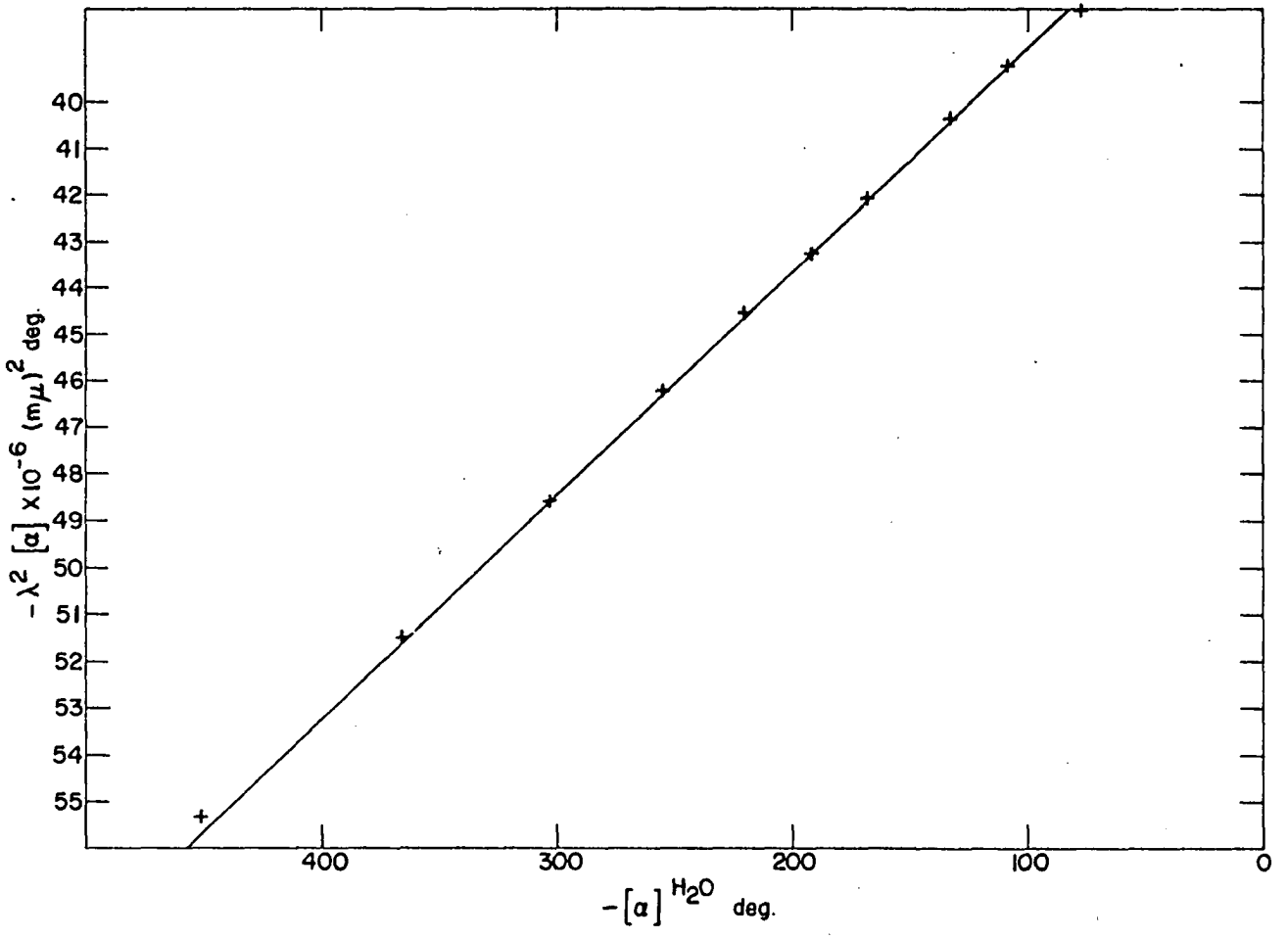
The dispersion of a 1% solution of B in 4.8 guanidine · HCl was also taken. The solution was made by pipetting 1 ml of an approximately 5% solution of B in 0.1 M KCl into 4 ml of 6 M urea. The optical density of an aliquot of the 5% solution was used to determine the concentration. The dispersion is shown in Graph 7, and the values of 215.6 mμ for λ_c and -113.7° for $[\alpha]_D^{H_2O}$ or -129.0° for $[R]_D^{H_2O}$ are obtained from it. The latter two constants have been corrected for index of refraction. Although these constants are typical of an unfolded protein, the following experiment shows that λ_c can assume even lower values.

The disulfide bridges were reduced and subsequently alkylated in a sample of β-lactoglobulin B by reacting with $NaBH_4$ followed by iodoacetate in a procedure outlined in the section on Peptide Mapping.

The dispersion of CMB in 0.1 M KCl was also taken, at a concentration of 2.14% by weight of undried protein. The



Graph 6. Rotary dispersion of B in 0.1 M KCl



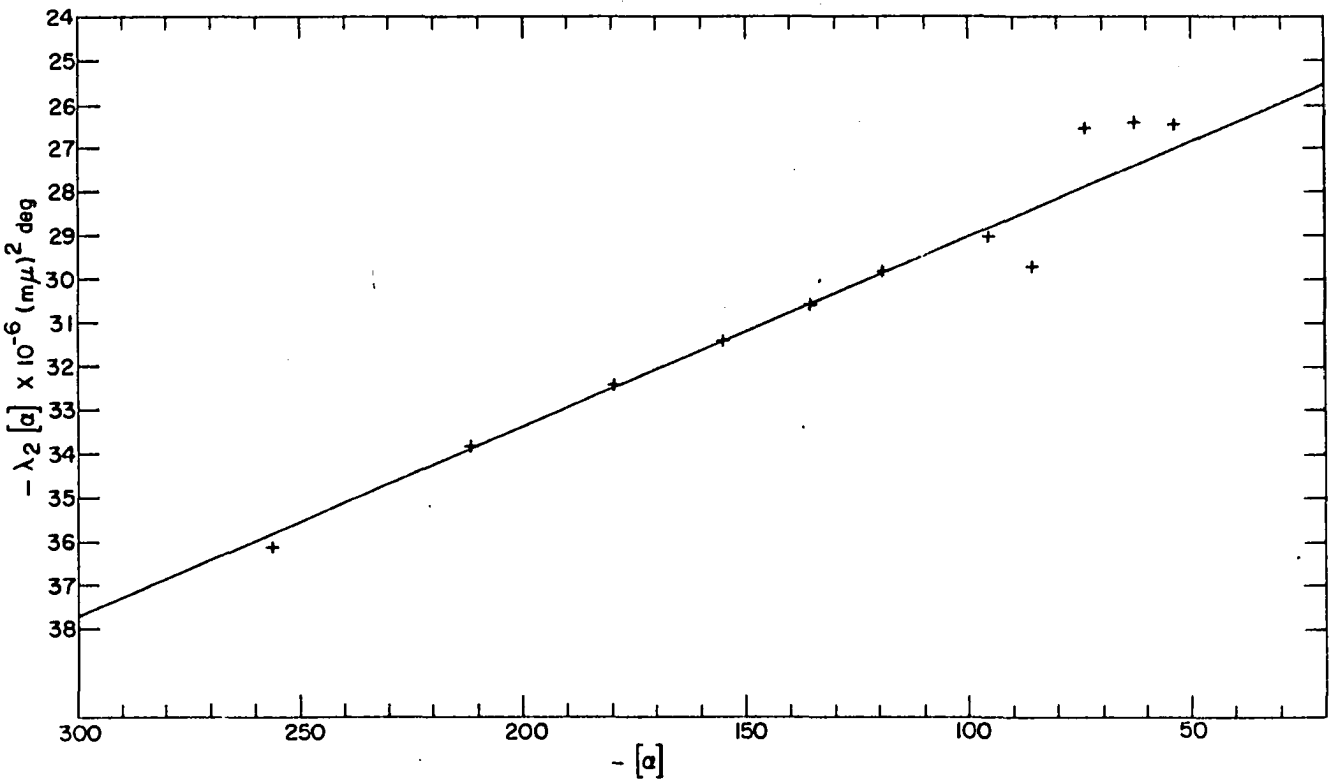
Graph 7. Rotary dispersion of B in 4.8 M guanidine • HCl

actual protein concentration obtained by dry weight was .64%, showing that this derivative had absorbed considerable water. The concentration determined from the optical density at 280 mμ was 0.49%, assuming that this derivative had the same molar tyrosine absorption at this wavelength as the native protein*, but that the molecular weight had increased from 35,500 to 36,000 with the assumed addition of eight carboxymethyl groups.

As can be seen from the curve of this dispersion in Graph 8, the values at the four highest wavelengths, 700, 650, 600 and 589 mμ, deviated widely from the line determined by the next seven points. A dispersion of native β-lactoglobulin A taken as a check on this same machine also exhibited this behavior for the points at 700, 650 and 600 mμ. Accordingly, these four points were rejected from the following computations. The λ_c of this CM derivative determined from the slope of a plot of $\lambda^2 [\alpha]$ vs. $[\alpha]$ by the method of least squares was 210.4 ± 1.9 mμ. When λ_c is determined from this plot, the value is independent of protein concentration.

This value is some what lower than the value of 215 mμ found by us for guanidine denatured β-lactoglobulin, or the

*The tyrosine side chains of β-lactoglobulin are known to titrate normally (90), and the spectrum of β-lactoglobulin shows no evidence of distorted tyrosine groups.



Graph 8. Rotary dispersion of CMB in 0.1 M KCl

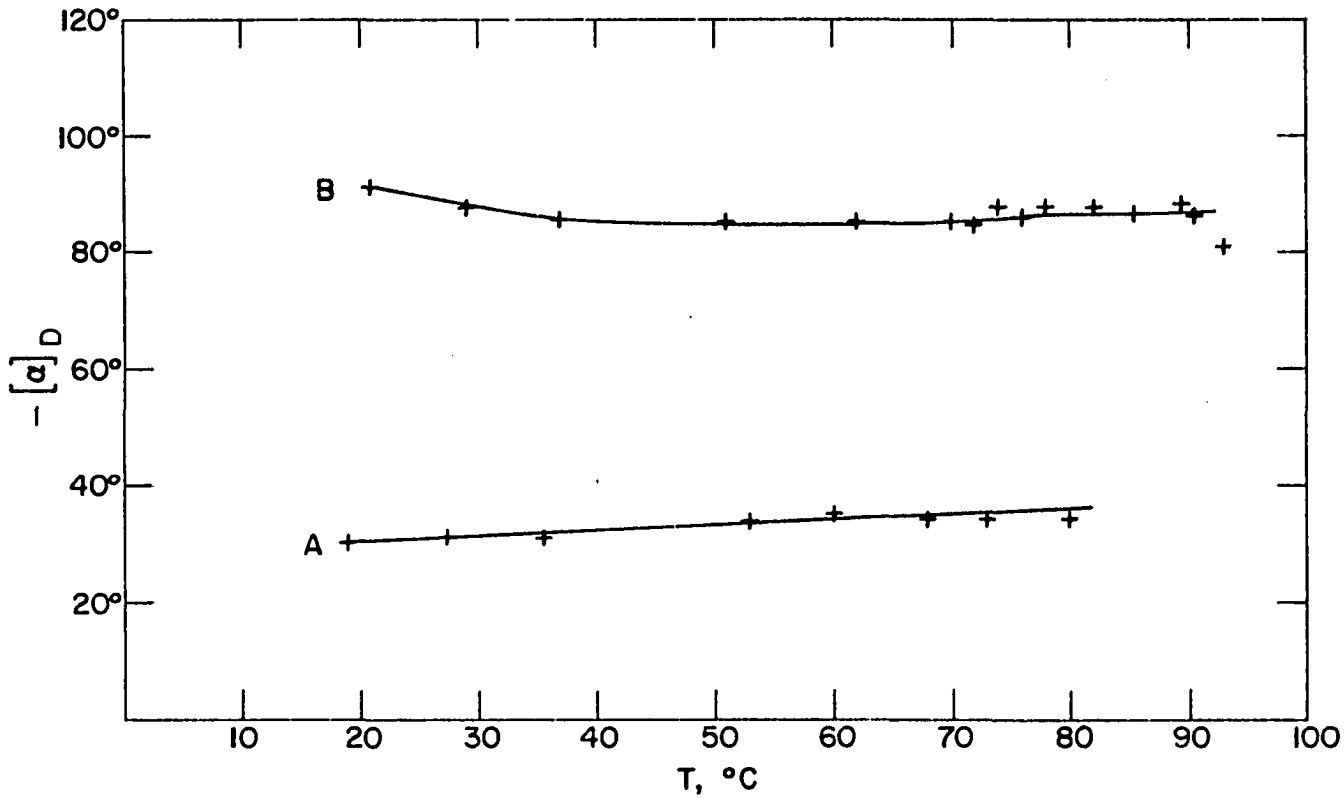
value of 220 mμ found by Schellman for the urea denatured protein (75).

The value of $[\alpha]_D = -90^\circ$ was found for CMB by extrapolution from Graph 8, assuming a protein concentration of 0.49%. The value would be -69° for a concentration of 0.64%.

We have reduced and alkylated the cystine residues which have a specific rotation of -214°. Although the specific rotation of CM cysteine is not known, judging from other cysteine derivatives it should be between $+30^\circ$ and -30°. The change from cystine to cysteine in the protein should then only raise the specific rotation by about $+6^\circ$. The difference between the value of -113.7° found for the $[\alpha]_D$ of the guanidine denatured protein and the value of -90° for the CM derivative is not completely explained by the breaking of the cystine bridges.

2. Temperature dependence

The temperature dependence of the rotation of the CM derivative and of native β-lactoglobulin B are shown in Graph 9. The rotation of the native protein, curve A, increases very slightly with temperature presumably as the secondary structure unfolds somewhat, but the protein can be heated to approximately 82° C without any great change taking place.



Graph 9. Temperature dependence of specific rotation

At this temperature the protein aggregates and precipitates out of solution in 0.1 M KCl. If the KCl concentration is increased to 0.25 M, the irreversible aggregation occurs at 84° C, and in 1 M KCl it occurs at 87° C. The protein is evidently unfolding very much in this narrow temperature range, exposing groups which lead to strong electrostatic attractions between molecules.

The CM derivative, curve B, shows a loss of levorotation initially as the temperature is increased. This is a characteristic behavior of optically active compounds, and corresponds to increased freedom of rotation about the asymmetric atom (75). The native protein does not show this behavior probably because its rotation is restricted by its secondary structure. The rotation of the derivative then levels off showing a slight discontinuity at 75° which may not be real.

The sample was removed from the polarimeter tube after the measurement at 93° and was subsequently boiled without any aggregation taking place. The addition of extra negative charge on the molecule by the carboxy-methyl groups may have prevented the irreversible aggregation of this derivative, which is evidently completely unfolded.

3. Organic solvents

A 1% solution of β-lactoglobulin B in 0.1 M KCl is precipitated out of solution by adding ethanol to a

concentration of 33% V/V. Normal propanol achieves the same result at a concentration of 20% V/V. The protein is denatured rather than salted out of solution by the treatment since the precipitate will not redissolve in 0.1 M KCl solution after being centrifuged and washed free of alcohol. An optical rotation measurement taken on a solution in 16% n. propanol gave an $[\alpha]_D = -42^\circ$, showing that the molecule had unfolded somewhat at this point. This concentration is just below the point of incipient cloudiness. As more alcohol is added more and more protein is precipitated just as if it were being titrated with the reagent. However, if the "titration" is stopped at any one point the amount of precipitate slowly increases with time.

We could be titrating the type of hydrophobic bonding postulated by Kauzman (49). This bonding consists of parafin-like regions which are unable to participate in hydrogen bonding to water molecules and so avoid the aqueous phase, adhereing to one another in an intramolecular "micelle" stabilized by the strong hydrogen bonds of water. As organic reagents such as alcohol are added to the protein-water solution, a point is reached where the hydrophobic groups are attracted to the solvent, and the interior of the protein unfolds. The α-helix could form a large part of these hydrophobic regions besides the hydrocarbon side chains of some of the amino acids, since we have seen that the helix

is formed preferentially in organic solvents, and has all its polar peptide groups well hidden in the interior of the helix.

4. Hydrogen bonding reagents

β-Lactoglobulin B seems to be completely unfolded by 4.8 M guanidine hydrochloride as was shown in Graph 7. Concentrations of guanidine hydrochloride 2 molar and under have virtually no effect on it, while 3 and 4 molar give a kinetic unfolding of the molecule as measured by the increase in optical rotation. This behavior parallels the urea denaturation of β-lactoglobulin as described by Kauzmann (49).

It is interesting to note that the sample of β-lactoglobulin used as a control in the herd sampling from Swift showed anomalous behavior in that it was partially unfolded by 2 M guanidine • HCl solutions.

The kinetics of unfolding of this sample increased with decreasing pH between pH 4 and 6, and were first order with respect to hydrogen ion. It is not known whether this pH dependence is also characteristic of the denaturation of other β-lactoglobulins at higher guanidine concentrations.

C. Discussion

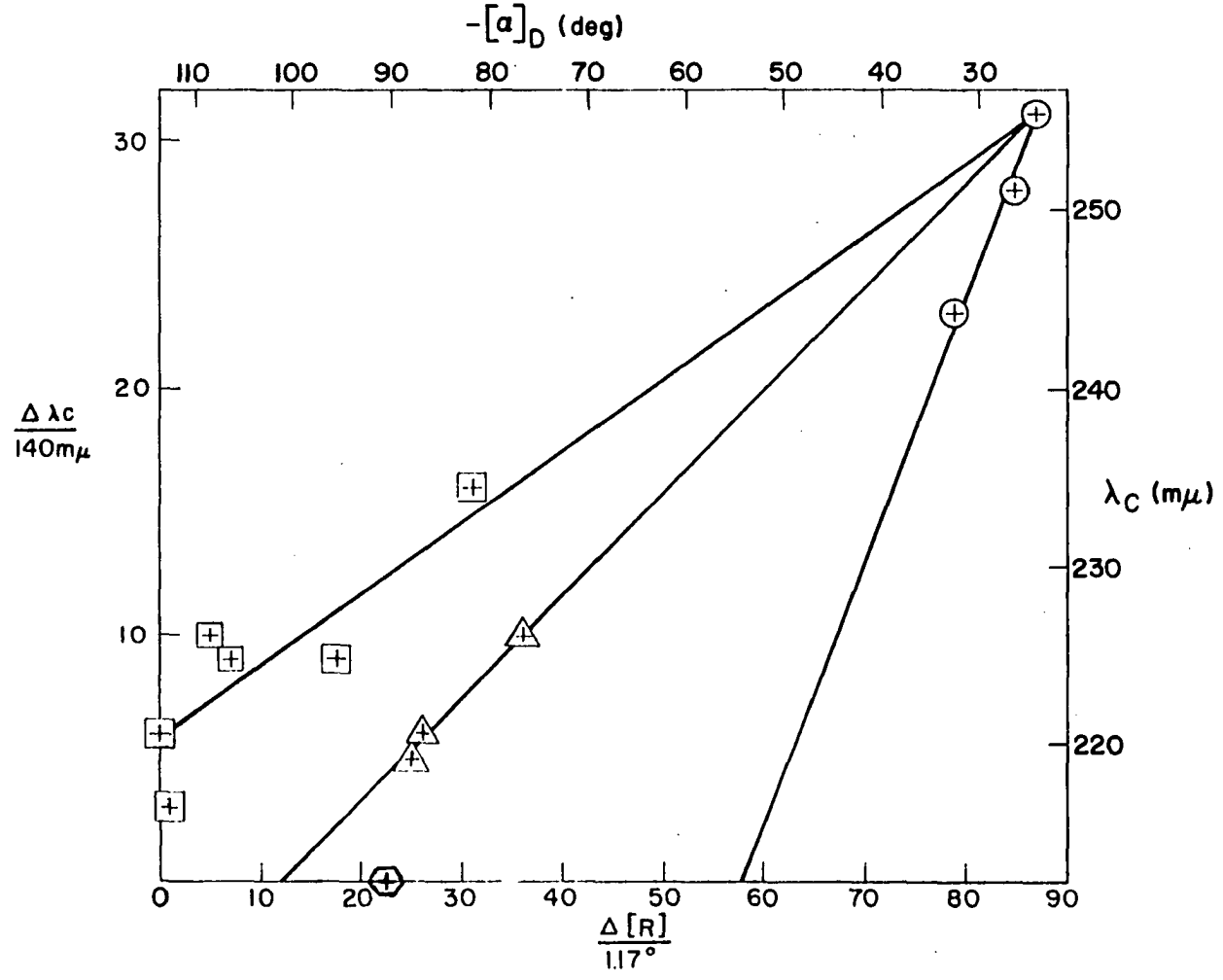
The λ_c and $[\alpha]_D$ values of various native and denatured forms of β-lactoglobulin determined by us and by Schellman

(75) have been plotted in Graph 10. The helical estimates corresponding to these measurements are also shown in percentages on the bottom and left scales.

The circle in the upper right hand corner is our value for native β-lactoglobulin B, the circle below it is Schellman's value for native β-lactoglobulin and the circle below this is our value for a β-lactoglobulin B solution which had been left standing before the rotation measurements were taken. The three triangular points are β-lactoglobulin solutions denatured by base which were left at pH 9.7, 10.80 and 10.88 presumably for several days at 20° C. These values are also Schellman's (75), and the more basic values lie closer to the origin. The hexagon on the abscissa represents our sample of CMB. This point should lie about $\frac{1}{4}$" below the abscissa which represents 212 mμ, while the point was 210 $\pm$ 2 mμ.

The squares are urea and guanidine denatured β-lactoglobulins. The value nearest the origin is our value for 4.8 M guanidine, while the rest are Schellman's values for varying urea molarities.

Th hydrogen bond breaking reagents, urea and guanidine · HCl, produce the largest change in $[\alpha]_D$ and the smallest change in λ_c, while standing at room temperature produces the converse effect. This can be taken to mean that the change in $[\alpha]_D$ (or $[R]_D$) is measuring the disappearance of



Graph 10. Optical constants of native and denatured β-lactoglobulins

an asymmetric hydrogen bonded structure.

The two constants, λ_c and $[\alpha]_D$, are certainly shown by this graph to be measuring different parts of the unfolding of the protein, and each method of denaturation would seem to follow a different path to a different product.

The asymmetric hydrogen bonded structure we have postulated to be unfolding does not have the rotatory properties of the α-helices encountered in the synthetic polyamino acids, since it shows almost no anomalous dispersion as measured by either b_o or λ_c.

An alternative viewpoint to the above held by Tanford (88a) is that the change in $[\alpha]_D$ as β-lactoglobulin unfolds is caused by a change from a hydrophobic environment in the protein to an aqueous environment around the asymmetric centers. This viewpoint assumes that b_o is the only true measure of helical content. This change in environment has to be more than just a simple change in index of refraction since the magnitude of the rotation change is too great.

Graph 10 must then be interpreted as urea and guanidine causing the greatest unfolding of the hydrophobic areas into an aqueous environment, which is not as satisfactory an explanation of this graph.

Tanford has shown that if the denaturation by organic solvents is done at pH 3 so that the β-lactoglobulin stays in solution after unfolding, the protein will first unfold

somewhat, as we have reported, and will then refold in the organic solvent into a presumably α-helical form. This form has an anomalous dispersion with a b_o value near -380° (53% helix). During the whole process the $[\alpha]_D$ remains around -30 to -40° (88a).

It seems extremely unusual that a random coil to helix transformation such as this could take place in somewhat similar nonpolar environments without a large change in $[\alpha]_D$ occurring also. We would propose that a form somewhat similar to the α-helices of the polyamino acids is present in the native protein, and that this rearranges into the typical long α-helix with anomalous dispersion when the protein rearranges in the organic solvent.

In conclusion, new values of $[R]_D$ for native and denatured β-lactoglobulin B lead to a helical estimate for the native protein of 86%. The b_o value of the native protein leads to a 10% estimate of helix. Although no clear cut decision can be made between these widely different estimates, the former seems more defensible. An alternative theory is that some helical structure resembling the α-helix is present, such as the 2_{8b} or 3_{10} folding discussed by Robinson (73).

V. SUMMARY

Milk from cows in the Iowa State dairy herd was sampled until one was found whose β-lactoglobulin consisted of only β-lactoglobulin A and a corresponding one of B. Crystalline A and B were then isolated from the milk of these animals.

These two proteins were then denatured either by a heat and urea treatment or by one of two methods of reducing and alkylating their cystine bridges in urea. Trypsin and/or chymotrypsin digestions were performed on the denatured proteins and the resulting peptides were separated by two dimensional paper chromotography and paper electrophoresis. Although all subsequent work was complicated by trace amounts of cyanate present in the urea, several peptides which stained only faintly with ninhydrin were found to be different in the proteolytic digestions of A and B.

A Moore and Stein column analysis of the CM proteins prepared by the borohydride method revealed that CMA contained 1.8 more moles of aspartic acid and 3.2 more moles of valine than CMB, while CMB contained 1.7 more moles of glycine and 1.1 more moles of alanine per 35,500 MW. The error in these determinations was of the order of 1 mole/35,500.

Peptide 20-B, a faint peptide which showed up reproducibly in CMB has been assigned the tentative composition lys, gly, cys. This peptide was not present in CMA.

In addition, preliminary experiments have shown that peptide 4.1 may contain a valine residue in CMA and an alanine in CMB.

Optical rotatory dispersion measurements on β-lactoglobulin B indicate that its optical constants are very similar to mixed milk β-lactoglobulin. Its rotatory dispersion is puzzling in that the protein shows almost no anomalous dispersion, yet a very low specific rotation in the native state. An α-helical secondary structure would be expected to give the native protein a low specific rotation, but would also produce considerable anomalous dispersion. Experiments have indicated that the low levorotation seems to be caused by some sort of hydrogen bonded structure. The small amount of anomalous dispersion present may not be caused by a hydrogen bonded structure. This finding is discussed in the light of recent work by Tanford, who has shown that organic solvents will rearrange the native structure of β-lactoglobulin into one showing a typical helical dispersion with little change in specific rotation but with a large increase in the amount of anomalous dispersion.

The amount of hydrogen bonding in β-lactoglobulin and two other proteins has been calculated utilizing water adsorption data. A method of computing the amount of hydrogen bonding of the peptide imido nitrogen from the I.R.

spectrum of a protein-deuterium oxide solution has also been suggested. The results from serum albumin have been compared with other hydrogen bonding data

Equation 8, $\lambda_c^2 = \left(1 + \frac{b_o}{a_o} \right) \lambda_o^2$ has been derived for low values of λ_c which relates the two measures of anomalous rotatory dispersion, λ_c and b_o, to one another.

VI. BIBLIOGRAPHY

1. Anfinsen, C. B., The Molecular Basis of Evolution, John Wiley and Sons, New York, N. Y. (1959).

2. Anfinsen, C. B., Aqvist, S. E. G., Cooke, J. P., and Jonsson, B., J. Biol. Chem., 234, 1118 (1959).

3. Aschaffenburg, R. and Drewry, J., Nature, 176, 218 (1955).

4. Aschaffenburg, R. and Drewry, J., Biochem. J., 65, 273 (1957).

5. Aschaffenburg, R. and Drewry, J., Nature, 180, 276 (1957).

6. Beadle, G. W., The Role of the Nucleus in Heredity, in McElroy, W. D. and Glass, B., eds., The Chemical Basis of Heredity, pp. 3-22, The Johns Hopkins Press, Baltimore, Md. (1957).

7. Beckwith, A. C., Titration studies of the unfolding of beta lactoglobulins in the presence of guanidine hydrochloride, Unpublished M.S. Thesis. Library, Iowa State University of Science and Technology, Ames, Iowa (1959).

8. Benson, E. E. and Linderstrom-Lang, K., Biochim. et Biophys. Acta, 32, 579 (1959).

9. Benzer, S., The Elementary Units of Heredity, in McElroy, W. D. and Glass, B. eds., The Chemical Basis of Heredity, pp. 70-93, The Johns Hopkins Press, Baltimore, Md. (1957).

10. Berger, A. and Linderstrom-Lang, K., Arch. Biochem. Biophys., 69, 106 (1957).

11. Block, R. J., Durrum, E. L., and Zweig, G., A Manual of Paper Chromatography and Paper Electrophoresis 2nd Ed. Academic Press, Inc., New York, N. Y. (1958).

12. Blout, E. R., de Loze, C., Bloom, S. M., and Fasman, G. D., J. Am. Chem. Soc., 82, 3787 (1960).

13. Bresler, S. E., Discussions Faraday Soc., 25, 158 (1958).

14. Brown, G. L. and Brown, A. V., Symposia Soc. Exptl. Biol., 12, 6 (1958).

15. Brown, H., Sanger, F., and Kitai, R., Biochem. J., 60, 556 (1955).

16. Bull, H. B., J. Am. Chem. Soc., 66, 1499 (1944).

17. Cannan, R. K., Kibrick, A. C., and Palmer, A. H., Ann. N. Y. Acad. Sci., 41, 243 (1941).

18. Chibnall, A. C., Haselbach, C., Manzan, J. L., and Rees, M. W., Biochem. J., 68, 122 (1958).

19. Colvin, J. R., Smith, D. B., and Cook, W. H., Chem. Revs., 54, 687 (1957).

20. Crammer, J. L. and Neuberger, A., Biochem. J., 37, 302 (1945).

21. Crick, F. H. C. and Watson, F. D., Proc. Roy. Soc. (London), A 223, 80 (1954).

22. Doty, P. Bradbury, J. H., and Holtzer, A. M., J. Am. Chem. Soc., 78, 947 (1956).

23. Doty, P. and Lundberg, R. D., Proc. Natl. Acad. Sci. U. S., 43, 213 (1957).

24. du Vigneaud, V., Ressler, C., and Trippett, G., J. Biol. Chem., 205, 949 (1953).

25. Elliot, A., Hanby, W. E., and Malcolm, B. R., Nature 178, 1170 (1956).

26. Fincham, J. R. S., Ann. Rev. Biochem., 28, 343 (1959).

27. Fitts, D. D. and Kirkwood, J. G., Proc. Natl. Acad. Sci. U. S., 42, 33 (1956).

28. Fraenkel-Conrat, H., Symposium on Structure of Enzymes and Proteins, J. Cell. and Comp. Physiol., 47, Supplement 1, p. 133-149, (1956).

29. Gaylord, N. G., Reduction with Complex Metal Hydrides, Interscience Publishers, Inc., New York, N. Y. (1956).

30. Geschwind, I. J., Li, C. H., and Barnafi, L., J. Am. Chem. Soc., 78, 4494 (1956).

31a. Giles, N. H., Partridge, C. W. H., and Nelson, N. J., Proc. Natl. Acad. Sci. U. S., 43, 305 (1957).

31b. Gordon, W. G., Basch, J. J., and Kalan, E. B., Biochem. Biophys. Research Comm., 3, 672 (1960).

32. Green, W. D. and Aschaffenburg, R. J., Molecular Biology, 1, 54 (1959).

33. Green, W. D., North, A. C. T., and Aschaffenburg. R., Biochim. et Biophys. Acta., 21, 583 (1956).

34. Hardy, T. L., Holland, D. O., and Nayler, J. H. C., Anal. Chem., 27, 971 (1955).

35. Harris, J. I., Sanger, F., and Naughton, M. D., Arch. Biochem. Biophys., 65, 427 (1956).

36. Hill, R. L., Kimmel, J. R., and Smith, E. L., Ann. Rev. Biochem., 28, 87 (1959).

37. Hill, R. L. and Schwartz, H. C., Nature, 184, 641 (1959).

38. Hill, R. L. and Smith, E. L., Biochim. et Biophys. Acta, 19, 376 (1956).

39. Hunt, J. A. and Ingram, V. M., Nature, 181, 1062 (1958).

40. Hvidt, A. and Linderstrom-Lang, K., Biochim. et Biophys. Acta., 16, 168 (1958).

41. Hvidt, A. and Linderstrom-Lang, K., Biochim. et Biophys. Acta, 18, 307 (1958).

42. Ingram, V. M., Nature, 178, 792 (1956).

43. Ingram, V. M., Biochim. et Biophys. Acta, 36, 402 (1959).

44. Itano, H., Advances in Protein Chem., 12, 215 (1957).

45. Jacobsen, C. F., Leonis, J., Linderstrom-Lang, K., and OHesen, M., in Glick, D. ed., Methods of Biochemical Analysis, Vol IV, pp. 171-210, Interscience Publishers, Inc. New York, N. Y. (1957).

46. Jacobsen, C. F. and Linderstrom-Lang, K., Nature, 164, 411 (1949).

47. Jirgensons, B., Arch. Biochem. Biophys., 91, 123 (1960).

48. Katz, A. M., Dreyer, W. J., and Anfinson, C. B., J. Biol. Chem., 234, 2897 (1959).

49. Kauzmann, W., Denaturation of Proteins and Enzymes, in McElroy, W. D. and Glass, B., eds., The Mechanisms of Enzyme Action, pp. 70-109, The Johns Hopkins Press, Baltimore, Md. (1954).

50. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Syckoff, H., and Phillips, D. C., Nature, 181, 662 (1958).

51. Kiddy, C. A., Plowman, R. D., Townend, R., and Timasheff, S. N., Agricultural Research, 8 No. 6, 10 (1959).

52. Klostergaard, H. and Pasternak, R. A., J. Am. Chem. Soc., 79, 567 (1957).

53. Kornberg, A. and Josse, J., Federation Proc., 19, 305 (1960).

54. Kornberg, A., Zimmerman, S. B., Kornberg, S. R., and Josse, J., Proc. Natl. Acad. Sci., U. S., 45, 772 (1959).

55. Lchman, I. R., Bessman, M. J., Simms, E. G., and Kornberg, A., J. Biol. Chem., 233, 171 (1958).

56. Lcnormant, H., Ann. de Chim. (Paris), Ser. 12, 5, 30 (1950).

57. Lenormant, H. and Blout, E. R., Nature, 172, 770 (1953).

58. Li, C. H., Dixon, J. G., and Chang, D., J. Am. Chem. Soc., 80, 2587 (1958).

59. Lindley, H. and Rollett, J. S., Biochim. et Biophys. Acta, 18, 183 (1955).

60. Moffitt, W., Proc. Natl. Acad. Sci., U. S., 12, 736 736 (1956).

61. Moffitt, W., Fitts, D. D., and Kirkwood, J. G., Proc. Natl. Acad. Sci., U. S., 43, 723 (1957).

62. Moffitt, W. and Yang, J. T., Proc. Natl. Acad. Sci., U. S., 42, 596 (1956).

63. Moore, S., Cole, R. D., Gundlach, H. G., and Stein, W. H., Intern. Congr. Biochem., Symposium on Proteins, 4th Meeting, (Vienna, Austria, Sept. 1958).

64. Moore, S., Spackman, D. H., and Stein, W. H., Anal. Chem., 30, 1185 (1958).

65. Moore, S. and Stein, W. H., J. Biol. Chem., 211, 893 (1954).

66. Nozaki, Y., Bunville, L. G., and Tanford, G., J. Am. Chem. Soc., 81, 5523 (1959).

67. Ochoa, S. and Mii, S., Biochim. et Biophys. Acta, 26, 445 (1957).

68. Ogston, A. G. and Tilley, J. M. A., Biochem. J., 59, 644 (1955).

69. Ogston, A. G. and Tombs, M. D., Biochem. J., 66, 399 (1957).

70. Pateman, J. A. and Fincham, J. R. S., Heredity, 12, 317 (1958).

71a. Pauling, L., J. Am. Chem. Soc., 67, 555 (1945).

71b. Perlman, G. E. and Diringer, R., Ann. Rev. Biochem., 29, 151 (1960).

72. Reindell, Fand Hoppe, W., Naturwiss, 40, 221 (1953).

73. Robinson, C. and Ambrose, E. J., Trans. Faraday Soc., 48, 854 (1952).

74. Rydon, H. N. and Smith, D. W. G., Nature, 169, 922 (1952).

75. Schellman, C. and Schellman, J. A., Compt. rend. trav. lab. Carlsberg Ser. Chim., 30 No. 27, 463 (1958).

76. Sela, M., White, F. H. Jr., and Anfinsen, C. B., Science, 125, 691 (1957).

77. Senti, F. R. and Warner, R. C., J. Am. Chem. Soc., 70, 3318 (1948).

78. Simpkin, J. L., Ann. Rev. Biochem., 28, 145 (1959).

79. Smithies, O., Biochem. J., 61, 629 (1955).

80. Sober, H. A. and Peterson, E. A., J. Am. Chem. Soc., 76, 1711 (1954).

81. Spackman, D. H., Stein, W. H., and Moore, S., Anal. Chem., 30, 1190 (1958).

82a. Spackman, D. H., Stein, W. H., and Moore, S., J. Biol. Chem., 235, 648 (1960).

82b. Stark, G. R., Stein, W. H., and Moore, S., J. Biol. Chem., 235, 3177 (1960).

83. Stein, W. H. and Moore, S., J. Biol. Chem., 178, 79 (1949).

84. Steinburg, D. and Mihalyi, E., Ann. Rev. Biochem., 26, 873 (1957).

85. Suzi, H., Zcll, T., and Timasheff, S. N., Arch. Biochem. Biophys., 85, 437 (1959).

86. Szent-Gyorgyi, A. G. and Cohen, C., Science, 126, 697 (1957).

87. Tanford, C., Bunville, L. G., and Nozaki, Y., J. Am. Chem. Soc., 81, 4032 (1959).

88a. Tanford, C., De, P. K., and Taggert, V. G., J. Am. Chem. Soc., 82, 6028 (1960).

88b. Tanford, C., Hauenstein, J. D., and Rands, D. G., J. Am. Chem. Soc., 77, 6409 (1955).

89. Tanford, C. and Nozaki, Y., J. Biol. Chem., 234, 2874 (1959).

90. Tanford, C., Swanson, S. A., and Shore, W. S., J. Am. Chem. Soc., 77, 6414 (1955).

91. Tombs, M. P., Biochem. J., 67, 517 (1957).

92. Tombs, M. P., Biochem. J., 69, 491 (1957).

93. Townend, R. and Timasheff, S. N., J. Am. Chem. Soc., 82, 3168 (1960).

94. Townend, R., Timasheff, S. N., and Winterbottom, R. J., J. Am. Chem. Soc., 82, 3161 (1957).

95. Townend, R., Weinberger, L., and Timasheff, S. N., J. Am. Chem. Soc., 82, 3175 (1960).

96. Tristram, G. R., Advanc. Protein Chem., 5, 83 (1949).

97. Vinograd, J. and Hutchinson, W. D., Federation Proc., 18, 344 (1959).

98. Watanabe, I., Biochim. et Biophys. Acta, 25, 665 (1957).

99. Yang, J. T. and Doty, P., J. Am. Chem. Soc., 79, 761 (1957).

VII. ACKNOWLEDGMENTS

The author wishes to express thanks to:

Dr. L. S. Bartell for his enlightening criticism and his patience during the preparation of this manuscript, and for help during the course of the research.

Dr. W. R. Hearn for guidance in the research project, and also for his strong belief that proteins obey organic chemistry.

Dr. W. F. Harrington for the lesson that nature is, above all else, rational.

Dr. C. B. Anfinsen in whose laboratory the finger-printing project was started.

Dr. Arnold Katz and Juanita Cook under whom the author learned many of the techniques used here.

Francisco Puchal for the use of his Moore and Stein column which behaved so nicely.

(Dr.) Eugene Lazzari for many helpful midnight discussions.

This work has been supported in part by research grants RG-4830 and A-2352 of the National Institute for Arthritis and Metabolic Diseases, National Institutes of Health, and by the Iowa Agricultural and Home Economics Experiment Station, Project No. 1384.

VIII. APPENDIX

The following abbreviations not commonly encountered have been used in this work.

A ..... β-lactoglobulin A

B ..... β-lactoglobulin B

CM ..... carboxymethyl

DNA ..... deoxyribonucleic acid

RNA ..... ribonucleic acid

SRNA ..... soluble ribonucleic acid

I.R. ..... infra red

O.D. ..... optical density

cys ..... cystine

½ cys ..... half cystine residue

gly ..... glycine

ala ..... alanine

val ..... valine

leu ..... leucine

Ileu ..... isoleucine

ser ..... serine

thr ..... threonine

met ..... methionine

asp ..... aspartic acid

asp-NH_2 ..... asparagine

glu ..... glutamic acid

glu-NH_2 ..... glutamine

lys ..... lysine

arg ..... arginine

his ..... histidine

phe ..... phenylalanine

tyr ..... tyrosine

try ..... tryptophane

pro ..... proline