chapter 7

SPECIES VARIATION

IN

PROTEIN STRUCTURE

One of the major questions to be answered in arriving at a clear understanding of the phylogenetic relationships between different forms of life is whether there exist identical, or closely *homologous,* genes in widely separated species, or whether similarities in phenotype are due to *analogous* genes which determine equivalent appearance or function by different pathways. The techniques of experimental genetics permit us to compare the genetic makeup of only those organisms that can be successfully crossed. We know, for example, that the eye pigments of a wide variety of species contain the same light-sensitive compound. However, we have no genetic way of testing whether the synthesis of this compound in different organisms is under the control of the same set of genes, structurally modified perhaps in some slight manner but still essentially identical, or whether completely different genes are involved which act in concert to achieve the same end result.

All genetic analyses depend on the availability of some recognizable phenotypic character, be it morphological, functional, or meta-

bolic. When the character being used as the criterion for the presence or absence of a functional gene or set of genes is a gross morphological one, we cannot attempt to distinguish homology of genes from analogy of genes. This is true even in those instances in which we can demonstrate the presence, in widely separated species, of identical chemical structures such as the creatine phosphate in the tissues of all vertebrates. The production of such a substance could be carried out by analogous, rather than homologous, enzyme systems in the different species, and the genes that exert the basic control on the synthetic process might conceivably be quite different in terms of chemical organization. There does appear to be one approach, however, which might give definitive information about the persistence of particular genes throughout the phyla. The techniques of isolation and structural analysis now available enable the protein chemist to make exact comparisons of proteins isolated from a wide variety of biological sources. If we accept the hypothesis that the proteins represent a primary, if perhaps a fuzzy, "print" of genetic information, we may then conclude that two organisms have the same gene, or gene set, when they both contain the same protein molecule. (Many readers will not be willing to swallow, whole, the thesis that proteins represent the *direct* translation of genetic information. We shall consider some of the arguments, pro and con, in Chapter 8.)

It is clear that we must expect to find differences between the "same" protein in various species. This is particularly true since, as we have discussed in the past chapter, certain parts of biologically active protein molecules are relatively more "dispensable" than others from the standpoint of function. Mutations that lead to changes in the sequence of amino acids in the last three C-terminal amino acids of ribonuclease, for example, might cause little change in the life expectancy or fertility of the affected animal. On the other hand, a mutation that led to a critical modification in the sequence of the "active center" of the enzyme might well be lethal, and the gene, so mutated, would not be perpetuated.

A comparison of the structures of homologous proteins (i.e., proteins with the same kinds of biological activity or function) from different species is important, therefore, for two reasons. First, the *similarities* found give a measure of the minimum structure which is essential for biological function. Second, the *differences* found may give us important clues to the rate at which successful mutations have occurred throughout evolutionary time and may also serve as an additional basis for establishing phylogenetic relationships.

The proteins and polypeptides for which *complete* comparisons of covalent structure can be made at the present time are relatively few in number. We can list, in this category, insulin, adrenocorticotropin, melanotropin, vasopressin, oxytocin, and hypertensin. The complete structure of glucagon (hypoglycemic factor) has been elucidated, but no comparisons of this hormone from different species have as yet appeared. Among the enzymes, only ribonuclease has been sufficiently studied to permit essentially complete comparison of species differences. The structure of a tetradecapeptide portion of cytochrome c, in the vicinity of the heme prosthetic group, has been examined for a fairly large number of organisms. Finally, there exists a large literature on the composition, end group, and end sequence analysis of various sets of homologous proteins, and on their physical, enzymatic, and immunologic properties, which allows us to make at least some educated guesses about similarities and differences.

Ribonuclease and the "Fingerprinting" Technique

The detection and study of chemical differences between homologous proteins is generally carried out by one or another variety of "fingerprinting" technique. In essence, this involves the use of reproducible physical methods for the separation of peptide fragments produced by digestion of the proteins with proteolytic enzymes. After establishing the distribution pattern of these fragments for the original "reference standard" protein, differences obtained on digests of the protein from other biological sources may then be easily detected, and the nature of the chemical modification may be determined by classical methods of amino acid and sequence analysis. An example of a "fingerprint" comparison is presented in Figures 71 and 72. This figure shows the patterns for beef and sheep ribonucleases. The differences are quite clear and are completely reproducible from digest to digest. In this instance, digestion was carried out first with trypsin and subsequently with chymotrypsin. The protein was oxidized with performic acid prior to digestion to avoid steric complications which might be introduced by the disulfide bridges.

The sort of fingerprinting used here is rapid and technically simple and serves, adequately, for the preliminary detection of differences. Indeed, if we are happy with micro techniques of the sort that served Sanger and his colleagues so well in their studies of insulin structure, a complete comparison of the corresponding peptides

Figure 71. A "fingerprint" of a proteolytic enzyme digest of oxidized bovine pancreatic ribonuclease. An aliquot of the digest was applied in a small spot at the left of the figure. This material was then subjected to descending paper chromatography, as indicated by the arrow, and, after allowing the solvent to evaporate, the paper was moistened with buffer solution and subjected to high-voltage electrophoresis. The sheet of paper was then sprayed with ninhydrin solution to stain those areas containing peptides. These areas were cut out (from a lightly stained paper), and the peptides were eluted. The amino acid composition of each was determined, after hydrolysis with acid, by paper chromatography. The composition of the various peptide components is given in Table 8. For further details consult an article by C. B. Anfinsen, S. E. G. Åqvist, Juanita P. Cooke, and Börje Jönsson, *J. Biol. Chem.*, **234**, No. 5 (1959).
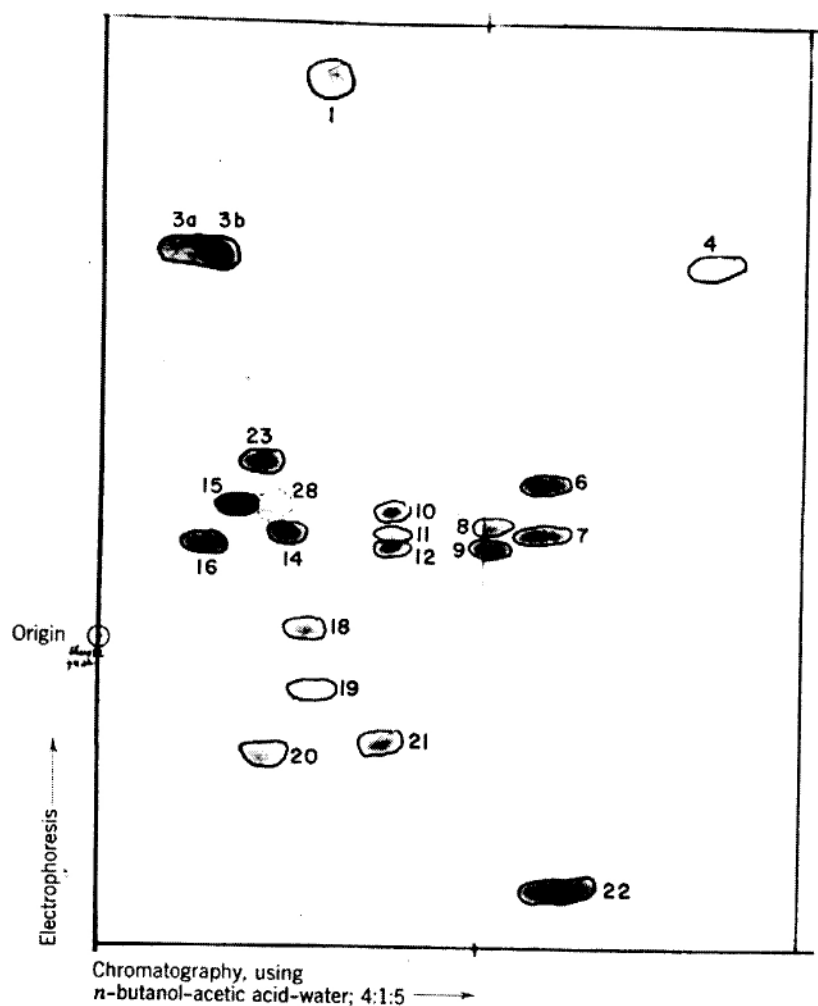
Figure 72. A "fingerprint" of performic acid-oxidized ovine pancreatic ribonuclease. The techniques employed were the same as those described in Figure 71, and the composition of the various peptide components is given in Table 8.

shown in such a fingerprint can be made with relative assurance. The use of such methods always involves an uncertainty in regard to the minor ninhydrin-positive components that routinely plague the paper chromatographer, and a certain amount of personal judgment is frequently involved in deciding whether a trace peptide compo-

nent is due to a fleck of dirt on the paper or to a bona fide structural fragment. For this reason the purist will often prefer the use of ion exchange columns over paper chromatography and electrophoresis, since he can then make more quantitative estimates of the recovery of fragments in relation to theoretical expectation. The latter course is obviously to be recommended in principle. However, when a large series of proteins are to be compared, and when experience has indicated the limits of error involved, the more rapid and flexible paper method will probably be employed to establish the major generalities of structure.

The study of species differences in the structure of enzymes is of special interest because with these proteins we can, in many instances, consider such variations in terms of the ability of the enzyme to catalyze a specific, chemical reaction. When, for example, a suc-

TABLE 7

Amino Acid Analyses of Beef and Sheep Pancreatic Ribonuclease[a]

Amino Acid	"Theory"	Beef Enzyme Average Observed Number of Residues per Mole (mol. wt. 13,683)	Sheep Enzyme Average Observed Number of Residues per Mole (mol. wt. 13,683)	Estimated Changes in Sheep Enzyme
Asp	15	15.65	15.20	−1 (?)
Thr	10	9.77	9.06	−1
Ser	15	13.95	16.15	+2
Glu	12	11.75	13.3	+1 to +2
Pro	4	3.95	3.96	
Gly	3	3.03	3.14	
Ala	12	11.87	11.1	−1
Cys	4	3.36	3.66	
Val	9	8.48	8.87	
Met	4	3.62	3.73	
Ileu	3	1.83	1.76	
Leu	2	2.05	1.99	
Tyr	6	5.7	5.71	
Phe	3	3.27	3.41	
Lys	10	11.25	9.88	−1
His	4	3.51	3.83	
Arg	4	3.76	3.78	

[a] From unpublished experiments of S. Åqvist, C. B. Anfinsen, J. Cooke, and B. Jönsson.

TABLE 8

Analyses of Peptides from Fingerprints of Trypsin and Chymotrypsin
Digestions of Oxidized Pancreatic Beef and Sheep Ribonuclease[a]

Beef No.	Composition	Sheep No.
3b	Lys$_2$,Glu,Thr,Ala$_3$ (beef)	--
--	Lys,Glu,Ser,Ala (sheep)	3a and 3b
6[b]	Phe+Glu,Arg.	6[b]
_[c]	Asp,Glu,Thr,Ser$_3$,Met,His	_[c]
_[c]	Asp,Ala$_2$,Ser$_3$,Tyr	_[c]
16	Cys,Asp,Glu,Met$_2$,Lys	16
1	Ser,Arg	1
2	Asp,Leu,Thr,Lys	--
13	Asp,Arg	--
--	Asp,Leu,Thr,Glu,Arg (sheep)	10
9	Cys,Asp,Val,Pro,Thr,Lys,Phe	9
18	Glu$_3$,Val$_3$,Leu,Ser$_2$,His,Asp,Ala$_2$,Cys	18
14	Asp,Cys,Ala,Val,Lys	14
20	Cys,Asp$_2$,Glu,Gly,Thr,Tyr	20
7	Glu,Ser,Tyr	7
12	Thr,Ser,Met	12
21	Cys,Asp,Ileu,Thr,Ser,Arg	21
15	Glu,Gly,Ser$_2$,Thr,Lys	15
8[b]	Cys,Asp,Ala,Tyr$_2$,Pro,Lys	8[b]
3c	Asp,Glu,Ala,Thr$_2$,Lys	--
--	Lys,Glu,Ala,Thr	23
28	His,Val,Ileu$_2$,Cys,Asp,Glu,Gly,Ala,Pro,Tyr	28
4	Val$_2$,Pro,His,Phe	4
22	Asp,Ala,Ser,Val	22
From carbobenzoxy-oxidized sheep ribonuclease	Lys,Ser,Glu,Ala,Phe,Arg	S3
	Cys,Lys,His,Asp,Glu,Ser,Met,Thr,Ala,Tyr,Arg	S1

[a] Amide nitrogens cannot be assigned to specific glutamic or aspartic acid residues since they are split off by the acid hydrolysis prior to chromatography.

[b] According to earlier observations (see Figure 60, Chapter 5), cleavages should have occurred between phenylalanine and glutamic acid in peptide 6 and in such a way as to remove the lysine residue from peptide 8. However, only traces of free phenylalanine and lysine were detected on the fingerprint patterns.

[c] These two peptides were not detected by the ninhydrin-staining reaction but have been accounted for in peptide S1, which was prepared by trypsin digestion of the carbobenzoxylated polypeptide chain (see Chapter 5), rather than by combined trypsin-chymotrypsin digestion of oxidized ribonuclease.

The spots labeled 5 and 17 in the fingerprint of digests of beef ribonuclease were present in such small quantities that their amino acid compositions could not be determined with any assurance. The same was true for the component labeled "X" in the beef ribonuclease fingerprint.
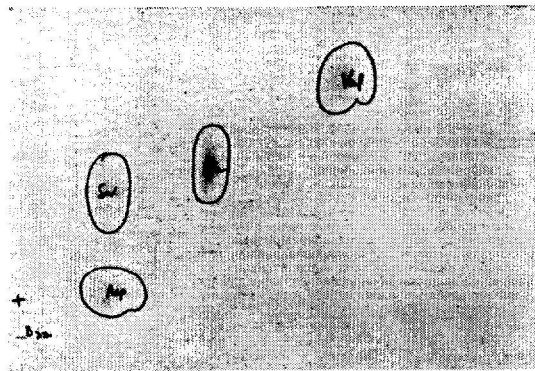
The amino acids on the paper chromatograms (Figure 73) which gave an unusually strong ninhydrin reaction are italicized. The subscripts, when present, indicate the number of moles of each amino acid in the peptide under consideration, as

cessful mutation has occurred which leads to the substitution of a charged amino acid residue for an uncharged one in a sequence, this information permits us to make some helpful conclusions regarding the nature and location of the binding site for substrate molecules on the enzyme surface.
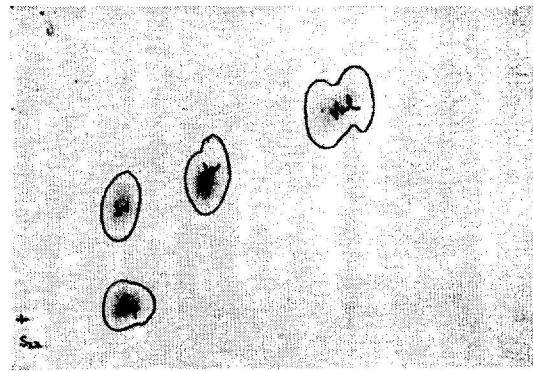
The amino acid analyses for bovine and sheep pancreatic ribonuclease are shown in Table 7. These data show that sheep ribonuclease contains less lysine, threonine (and perhaps less aspartic acid), and more serine and glutamic acid than does the beef enzyme. End group analysis indicates that both proteins contain N-terminal lysine, and sedimentation constants determined in the ultracentrifuge are essentially identical. When the peptides separated by the fingerprinting procedure illustrated in Figures 71 and 72 were eluted and analyzed, qualitatively, for amino acid composition, the results summarized in Table 8 were obtained. Examples of the chromatographic comparison of hydrolysates of a few sheep and beef peptides are shown in Figure 73. The peptides obtained from hydrolysates of the beef protein are those to be expected from the combined action of trypsin and chymotrypsin on oxidized ribonuclease, as may be deduced from the partial formula for this polypeptide shown in Figure 60. Corresponding peptides from the sheep material are also easily assignable to particular areas of the formula. Those sequences in sheep ribonuclease which differ from the beef structure may be allocated to specific areas of the chain on the basis of their compositions. At one point in the chain of the sheep enzyme, the absence of a trypsin-sensitive sequence involving lysine had precluded cleavage where the beef enzyme *was* split, and a single, longer peptide (sheep peptide 10 Figure 73e), embodying two of the beef fragments, resulted.

The studies relating various aspects of structure to function which were reviewed in the last chapter have suggested that the disulfide bridge joining half-cystines 1 and 6 may be reduced without complete destruction of catalytic activity. The species variation occurring at residue 37, where a positively charged amino acid, lysine, has been replaced by a negatively charged amino acid, glutamic acid, also suggests the unessentiality for substrate adsorption and hydrolysis of this part of the polypeptide chain.
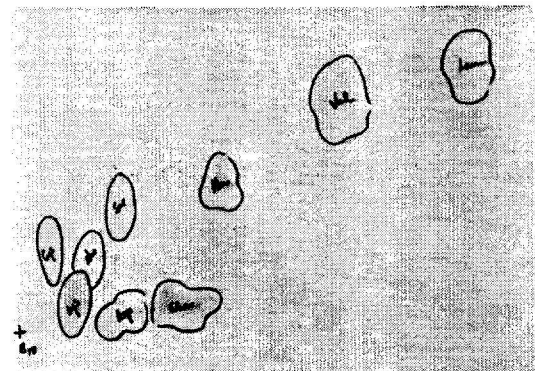
determined for beef ribonuclease (Figure 60, Chapter 5). Thus for beef peptide 3b, the earlier quantitative analyses, which demonstrated the presence of two residues of lysine and three of alanine for each residue of glutamic acid and threonine, were confirmed by the staining reaction which was correspondingly stronger for the two former amino acids.
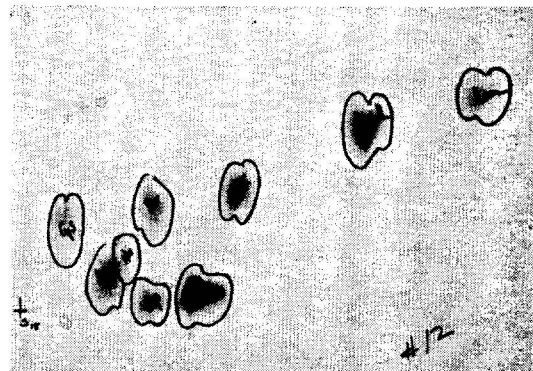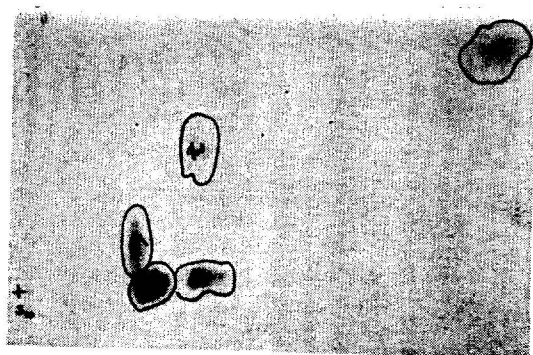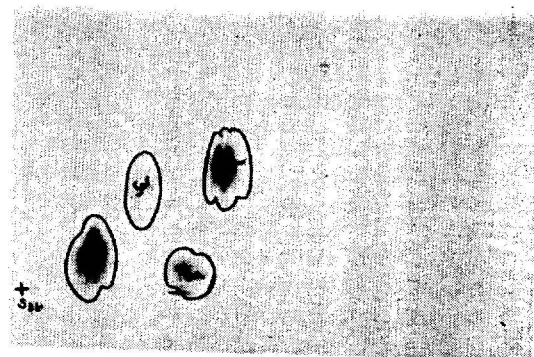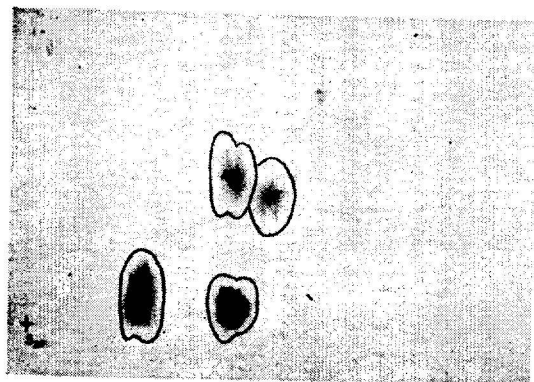
(a)

(b)

(c)

(d)

(e)

(f)

(g)

Figure 73. Two dimensional chromatography of the amino acids produced by acid hydrolysis of some of the peptide components shown in Figures 71 and 72. (*a*) Peptide 22 from beef ribonuclease. This is the *C*-terminal tetrapeptide sequence of the enzyme and has the structure Asp.Ala.Ser.Val. (*b*) The *C*-terminal tetrapeptide of sheep ribonuclease, having the same composition as that obtained from the beef enzyme. (*c*) Peptide component 18 from bovine pancreatic ribonuclease. This peptide is derived from residues 47 through 61 in the polypeptide chain. (*d*) The amino acid composition of peptide 18 from the ovine enzyme. Peptide 18 appears to be identical from both species. (*e*) Peptide component 10 from sheep ribonuclease. This peptide is derived from the amino acids in positions 34 through 39. The lysine residue, present in the structure of bovine ribonuclease within this portion of the sequence, has been replaced by glutamic acid in the sheep enzyme. The cleavage with trypsin which occurs at residue 37 of the beef enzyme can thus not occur for the sheep protein, and a single hexapeptide sequence is obtained instead of a tetrapeptide and dipeptide. (*f*) Peptide component 3b from sheep ribonuclease. The peptide represents the *N*-terminal heptapeptide sequence of the enzyme. (*g*) Component 3b from the beef fingerprint, the *N*-terminal heptapeptide *sequence*. The beef enzyme differs, in this region, from the sheep enzyme by the replacement of serine by threonine. See Figure 62 for details of structure.

Adrenocorticotropin (ACTH)

The complete amino acid sequences are known for corticotropins isolated from the anterior pituitary glands of three different species, pig, beef, and sheep. The structure of sheep ACTH was discussed in the last chapter, and the sequences shown in Table 9 include only those areas of the three molecules where differences are to be found. Although some difference between the content of amide nitrogen groups has been reported for the three species, these are not included in the figure since it has not been possible to rule out, with certainty, the possibility that these variations are due, in part, to the rigors of the isolation and purification techniques employed.

TABLE 9
Variations in Amino Acid Sequences Among Different Preparations of ACTH

Preparation	Species	Residue No.								
		25	26	27	28	29	30	31	32	33
β-Corticotropin	sheep beef[a]	Ala.	Gly.	Glu.	Asp.	Asp.	Glu	Ala.	Ser.	Glu.NH$_2$
Corticotropin A	pig	Asp.	Gly.	Ala.	Glu.	Asp.	Glu	Leu.	Ala.	Glu

[a] Identity with sheep hormone not absolutely certain but very probable as judged from the nearly complete sequence analysis by J. S. Dixon and C. H. Li (personal communication to the author).

Two points are of particular interest in regard to the sequences shown. First, the corticotropins of sheep and beef are identical and differ from that of the pig. This finding is consonant with the closer phylogenetic relationship of sheep and cows to each other than of either to pigs. Second, chemical differences are found only in that portion of the ACTH molecule which has been shown to be unessential for hormonal activity. Genetic mutations leading to such differences might, therefore, not be expected to impose significant disadvantages in terms of survival, and these genes could become established in the gene pools of the species.

Melanotropin (MSH)

Melanotropin, like the other hormones considered in this chapter, is a typically chordate polypeptide. Indeed, the demonstration of melanocyte-stimulating activity in extracts of tunicates constitutes an
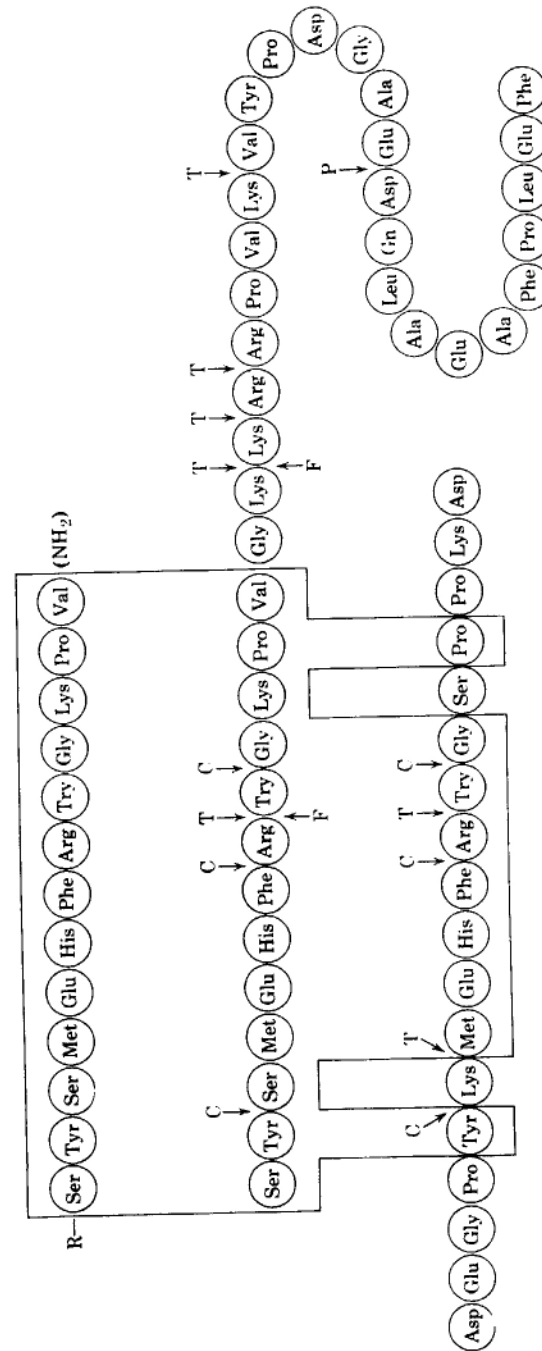

Figure 74. The structures of porcine α-ACTH (center), α-MSH (lower), and β-MSH (upper). Points of cleavage by chymotrypsin (C), trypsin (T), pepsin (P), and fibrinolysin (F) are indicated by the arrows. The common sequences are indicated within the boxed area.

important bit of evidence supporting the assignment of these organisms to the main thoroughfare of evolution between invertebrates and chordates.

Melanotropins have been isolated in pure form from both pig and beef pituitaries (posterior-intermediate lobes). Only a single polypeptide having MSH activity has been isolated from beef tissues, whereas two different chemical entities termed α and β-MSH have been isolated and characterized from hog pituitaries. The structures of these substances is given in Figure 74 together with that for porcine ACTH. We shall consider further the provocative similarity between the structures of MSH and ACTH in Chapter 10 in relation to protein biosynthesis. This similarity is undoubtedly responsible for the fact that adrenocorticotropic hormone exhibits marked melanocyte-stimulating activity.

Beef β-MSH differs from porcine β-MSH only in the replacement of the glutamic acid residue in position 2 by serine. Porcine α-MSH, however, is considerably different from both the other hormones and is actually identical with the sequence of the first thirteen amino acids in pig ACTH, except for the presence of a masking, acyl, group on the N-terminal amino group and an amide nitrogen group at the C-terminus. Lee and Lerner,[1] who isolated α-MSH, have suggested that this form of the hormone is the major one in pituitary extracts since it accounts, in their experiments, for the largest share of activity, although other investigators have not so far confirmed the presence of α-MSH.[*]

Bovine β-MSH possesses considerably less biological activity than does porcine β-MSH, and, until comparisons can be made of synthetic samples of these two materials, this difference in potency must be ascribed to the amino acid substitution at position 2.

Insulin

Sanger and his colleagues have determined the amino acid sequences for insulins derived from five different species. Differences

[*] C. H. Li and his colleagues have also recently isolated α-MSH from both porcine and bovine pituitary glands. The structures were found to be the same in both species. (Personal communication.)

The sequence of α-MSH has been confirmed by total synthesis in the laboratories of Klaus Hofmann and of R. Boissonnas. The activity of the synthetic material is critically dependent on the presence of the acetyl group on the N-terminal serine residue. Thus, in the experiments of Boissonnas and his coworkers, the activity of the acetylated polypeptide was approximately 70 times greater than before acetylation. (Personal communications from Hofmann and Boissonnas.)
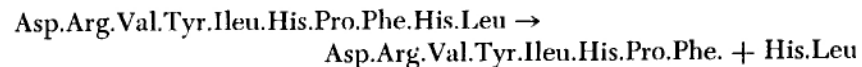
. . . CySO$_3$H.Ala.Ser.Val . . . (beef)
. . . CySO$_3$H.Thr.Ser.Ileu . . . (pig)
. . . CySO$_3$H.Ala.Gly.Val . . . (sheep)
. . . CySO$_3$H.Thr.Gly.Ileu . . . (horse)
. . . CySO$_3$H.Thr.Ser.Ileu . . . (sperm whale)
. . . CySO$_3$H.Ala.Ser.Thr . . . (sei whale)

Figure 75. Species differences in the amino acid sequences of insulins from various biological sources. These differences all occur within the disulfide "loop" of the A chain.
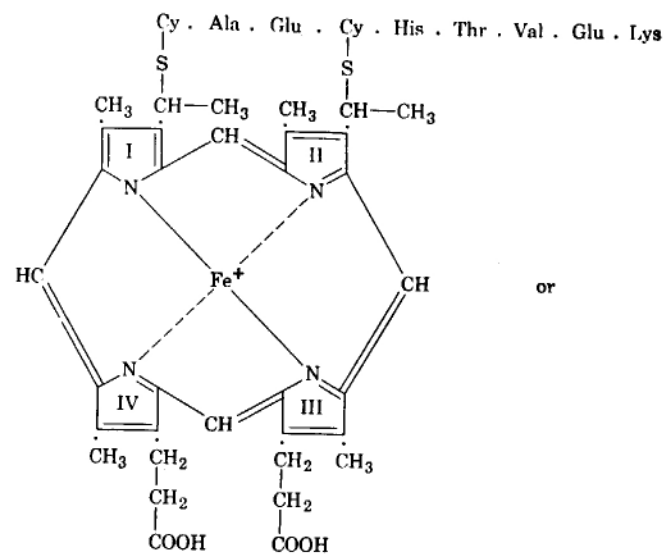
were limited to the amino acids within the disulfide "loop" of the A chain (Figure 75) and the B chain was identical in all instances (see Figure 65). Of the five insulins examined, only those from the pig and the sperm whale exhibited the same structure. The fact that all the observed differences were restricted to the sequence within the "loop" suggests that the amino acids in this region of the insulin molecule are not particularly critical ones from the standpoint of hormonal activity. On the other hand, several investigators have obtained evidence indicating that insulin loses its biological activity when disulfide bridges are reductively cleaved. The species difference results with insulin suggest that only the steric configuration of the loop is essential and that the "spacers" between the half-cystine residues may be varied through mutation of the corresponding gene or genes. It is of interest that sequence variations have *not* been observed in the C-terminal region of the B chain, an area which does appear to be essential for activity as shown by the inactivation of insulin following the removal of the last seven residues in the chain.

Hypertensins

The hypertensins are peptides present in serum which possess pressor activity. Two forms have been isolated from horse serum,[2] the first being convertible to the second by the action of an enzyme in plasma according to the equation:

Asp.Arg.Val.Tyr.Ileu.His.Pro.Phe.His.Leu →
          Asp.Arg.Val.Tyr.Ileu.His.Pro.Phe. + His.Leu

The precursor compound is not active in an *in vitro* test system, but after cleavage of the critical peptide bond it becomes active both *in vivo* and *in vitro*. The precursor form has also been isolated from
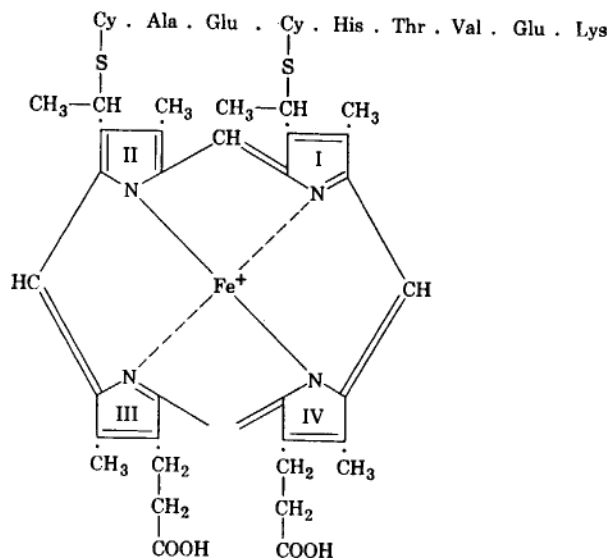
Cy . Ala . Glu . Cy . His . Thr . Val . Glu . Lys

CH₃ CH—CH₃ CH₃ CH—CH₃

(porphyrin structure with rings I, II, III, IV, Fe⁺)

or

CH₃ CH₂ CH₂ CH₃
CH₂ CH₂
COOH COOH

Cy . Ala . Glu . Cy . His . Thr . Val . Glu . Lys

CH₃—CH CH₃ CH₃—CH CH₃

(porphyrin structure with rings II, I, III, IV, Fe⁺)

CH₃ CH₂ CH₂ CH₃
CH₂ CH₂
COOH COOH

Figure 76. Structure of the peptide-porphyrin compound isolated from trypsin digests of cytochrome c. After H. Tuppy and G. Bodo, *Monatsch. Chem.,* **85,** 1024 (1954).

bovine serum[3] and is identical with that from horse serum except for the substitution of *isoleucine by valine.* These two amino acids are extremely similar in structure and the substitution in this case represents one of the more minimal changes possible given the available selection of naturally occurring amino acids.

Cytochrome c

The electron-transporting enzyme, cytochrome c, furnishes one of the most interesting examples of species variations in protein structure, since it has been isolated in pure form from a particularly wide assortment of species. Unfortunately, studies on variations in sequence have been carried out for only a relatively small portion of the total chain, and preliminary amino acid analyses indicate that there may be modifications elsewhere in the molecule as well. Never-
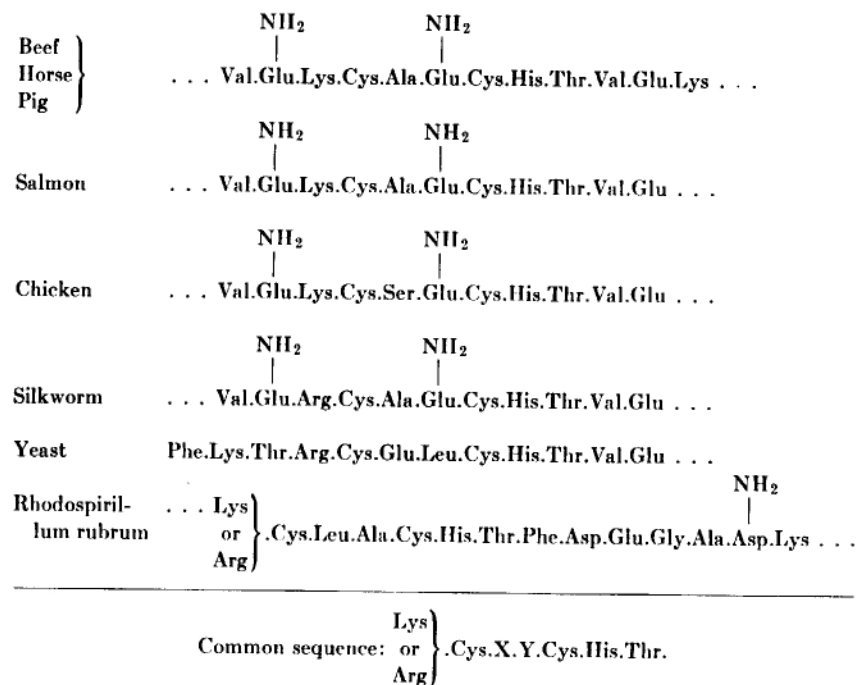
Beef }
Horse }
Pig } ... Val.Glu.Lys.Cys.Ala.Glu.Cys.His.Thr.Val.Glu.Lys ...
(NH₂ on Lys and Glu)

Salmon ... Val.Glu.Lys.Cys.Ala.Glu.Cys.His.Thr.Val.Glu ...
(NH₂)

Chicken ... Val.Glu.Lys.Cys.Ser.Glu.Cys.His.Thr.Val.Glu ...
(NH₂)

Silkworm ... Val.Glu.Arg.Cys.Ala.Glu.Cys.His.Thr.Val.Glu ...
(NH₂)

Yeast Phe.Lys.Thr.Arg.Cys.Glu.Leu.Cys.His.Thr.Val.Glu ...

Rhodospiril- ... Lys }
lum rubrum or } .Cys.Leu.Ala.Cys.His.Thr.Phe.Asp.Glu.Gly.Ala.Asp.Lys ...
  Arg } (NH₂ on Lys)

Common sequence:
Lys }
or } .Cys.X.Y.Cys.His.Thr.
Arg }

Figure 77. Variations in the sequence of the polypeptide chain of cytochrome c from species to species. From H. Tuppy, *Symposium on Protein Structure* (A. Neuberger, editor), John Wiley & Sons, 1958.

theless, the investigations of H. Tuppy, S. Paléus, and G. Bodo,[4] on this ubiquitously distributed enzyme, lend the most convincing support to the argument that certain units of the universal gene pool may be extremely ancient.

In the degradative studies of the enzyme, advantage was taken of the finding of H. Theorell[5] that the heme prosthetic group of cytochrome c is attached through stable thioether linkages to the protein moiety. After proteolytic degradation with trypsin (and in later studies with pepsin), that portion of the polypeptide chain which is attached to the heme nucleus was isolated. The structure of the heme-peptide compound as determined for cytochrome c from horse heart tissue is shown (in two alternatively possible forms) in Figure 76.

In subsequent investigations corresponding sequences from cytochrome obtained from a variety of other species have been elucidated as shown in Figure 77.

Somatotropins (Growth Hormones) and Prolactin

Pure growth hormone has been isolated from the pituitaries of the species listed in Table 10 by C. H. Li and his colleagues. These proteins have been subjected to both physical and chemical study, and, although even partially complete sequences are not yet available, a great deal can already be said about species variability. Molecular weights vary over nearly a twofold range, and the differences in the number of chains and the cystine content are striking. The beef and sheep hormones are, as in the case of several other proteins we have discussed earlier, quite similar, reflecting once again the close phylogenetic relationship between these two species.

The prolactins of sheep and beef are also extremely similar (Table 11), the only difference between them so far observed being a slightly greater tyrosine content in the beef hormone. The absence of a chemically detectable C-terminal amino acid residue is another example of "masked" end groups. The nature of the masking is unknown.

Hemoglobin

The hemoglobins have been studied, from the phylogenetic point of view, perhaps more than any other class of proteins. Most of

TABLE 10

N- and C-Terminal Sequences of Somatotropins from Various Species[*]

Somatotropins	Terminal Sequences	
	Amino End	Carboxyl End
Bovine	Phe.Ala . . .	. . . Ala.Phe.Phe
	Ala.Phe.Ala . . .	
Ovine	Phe . . .	. . . Try.Ala.Phe
	Ala . . .	
Whale	Phe.AspNH$_2$.Lys . . .	. . . Leu.Ala.Phe
Monkey	Phe.Ala.Thr . . .	. . . Ala.Gly.Phe
Human	Phe.Ser.Thr . . .	. . . Tyr.Leu.Phe

Some Physicochemical Properties of Various Somatotropins

Physicochemical Characteristics[b]	Bovine	Ovine	Whale	Monkey	Human
$S_{20,w}$	3.19	2.76	2.84	1.88	2.47
$D_{20,w} \times 10^7$	7.23	5.25	6.56	7.20	8.88
V_{20}	0.76	0.733	0.737	0.726	0.732
Molecular weight	45,000	47,800	39,900	25,400	27,100
f/f_0	1.31	1.68	1.45	1.57	1.23
P_I	6.85	6.8	6.2	5.5	4.9
Cystine	4	5	3	4	2
N-Terminal Residue(s)	Phe,Ala	Phe,Ala	Phe	Phe	Phe
C-Terminal Residue	Phe	Phe	Phe	Phe	Phe

[*] From C. H. Li, *Symposium on Protein Structure* (A. Neuberger, editor), John Wiley & Sons, 1958.

[b] $S_{20,w}$ in Svedbergs; $D_{20,w}$ in cm.^{2}/sec.; $\overline{V}$ in cc./gram; f/f_0, dissymmetry constant; P_I, isoelectric point; cystine in residues per mole. For details on the determination and significance of the physical constants consult the volumes entitled *The Proteins* (H. Neurath and K. Bailey, editors), Academic Press, 1953, 1954.

the available information on the hemoglobins has to do with the chemical and spectrophotometric characteristics of the various prosthetic groups and with oxygen and carbon dioxide-combining properties. Consequently, this information is not directly pertinent to our present discussion of species differences in protein structure. Re-

TABLE 11

Some Physical and Chemical Properties of Prolactin from Ovine and Bovine Pituitary Glands[a]

Physical and Chemical Properties	Ovine	Bovine
Molecular weight		
Sedimentation-diffusion	24,200	
Osmotic pressure	26,500	26,000
Analytical data	24,100	
Diffusion coefficient (D_{20})	8.44×10^{-7}	
Sedimentation constant ($S_{20,w}$)	2.19	
Partial specific volume (V_{20})	0.739	
Isoelectric point, pH	5.73	5.73
Specific rotation	$-40.5°$	$-40.5°$
Partition coefficient (2-butanol/0.35% aqueous trichloroacetic acid)	1.58	2.07
Tyrosine, %	5.26	6.62
Tryptophan, %	1.69	1.75
Cystine, residue/mole	3	3
N-terminal amino acid	Threonine	Threonine
C-terminal amino acid	none	none

[a] From C. H. Li, *Symposium on Protein Structure* (A. Neuberger, editor), John Wiley & Sons, 1958.

cently, however, following the important initial studies of R. Porter and F. Sanger,[6] a number of investigators have begun to examine amino acid sequences in the hemoglobins. Such studies are becoming increasingly more meaningful as the result of physicochemical investigations on the number of peptide chains per molecule and on the size of the monomer subunit. It now appears that, with the possible exception of foetal hemoglobin and the hemoglobin of the chicken (perhaps birds in general), the vertebrate hemoglobins contain two types of chains. These are present, under physiological conditions, in the form of a molecule with a molecular weight of about 65,000, composed of four chains, two of each type, held together through noncovalent linkages. (The earlier results, which indicated the presence of *six* chains in the hemoglobin of the horse, are probably incorrect on the basis of recent electrophoretic and ultracentrifugal investigations.) Valine is the N-terminal amino acid residue on both polypeptide chains of the vertebrate hemoglobins so far examined, except for the goat, sheep, and cow, in which one of the chains begins with methionine. A summary of the available end group data is given in Table 12. It is far too early to attempt to make any

TABLE 12

End Group Data on Vertebrate Hemoglobins

Species	N-Terminal Amino Acids or Sequences[a]		
Human adult[1,3]	Val.Leu	Val.–	
Human foetal[1,3]	Val.–	Val.–	
Dog[2]	Val.Leu	Val.Gly	(Val.Asp)[2]
Horse,[1,2,4] pig[2]	Val.Leu	Val.Glu.Leu	(Val.Gly)[2]
Cow,[1,2] goat,[1,2] sheep[1,2]	Val.Leu	Met.Gly	
Guinea pig[2]	Val.Leu	Val.Ser	(Val.Asp)[2]
Rabbit,[2] snake[2]	Val.Leu	Val.Gly	
Chicken[2]	Val.Leu		

[a] The presence of a third N-terminal sequence has been reported only by Ozawa and Satake.[2]

1. K. Porter and F. Sanger, *Biochem. J.*, **42**, 287 (1948).
2. H. Ozawa and K. Satake, *J. Biochem. (Japan)*, **42**, 641 (1955).
3. M. S. Masri and K. Singer, *Arch. Biochem. Biophys.*, **58**, 414 (1955).
4. D. B. Smith, A. Haug, and S. Wilson, *Federation Proceedings*, **16**, 766 (1957).

evolutionary sense out of this information. However, it might be pointed out that the N-terminal sequence, Val.Leu, seems to be present throughout the species examined, including the representatives of the reptiles and the birds. We may speculate on the possibility that the "valyl-leucyl" chain represents a relatively early invention of evolution, and that the addition (and modification) of a second type of chain accompanied later differentiations in the vertebrate phylum.

The elegant studies of Pauling, Itano, and their colleagues, and of Ingram on the normal and abnormal human hemoglobins, are discussed in the next chapter. The information gained from these studies should be of great value as a baseline for the investigation of hemoglobin structure in other species.

Species Comparisons of Serum Proteins

The chemist, interested in comparative protein chemistry, generally studies proteins that have unique and interesting biological activities. This is the reason why most of our knowledge of protein structure concerns enzymes, hormones, and pigment-associated proteins. The odds in favor of being chosen for study are, in a sense, fixed in favor of exactly those proteins that Nature might find it necessary to preserve in reasonably unmodified form during the evolutionary process.

TABLE 13

Precipitin Tests with Antihuman Serum[*]

(Antihuman serum was prepared in rabbits by periodic injection
with human serum)

Origin of Serum	Amount of Precipitate Relative to Human
Primates	
Man	100
Chimpanzee	130 (loose precipitum)
Gorilla	64
Orang	42
Mandrill	42
Guinea baboon	29
Spider monkey	29
Carnivores	
Dog	3
Jackal	10 (loose precipitum)
Himalayan bear	8
Genet	3
Cat	3
Persian lynx	3
Tiger	2
Ungulates	
Ox	10
Sheep	10
Water buck	7
Hog deer	7
Reindeer	7
Goat	2
Horse	2
Swine	0
Rodents	
Guinea pig	0
Rabbit	0
Insectivores	
Tenrec	0
Marsupials	
Six species: rock and nail-tailed wallabies, kangaroo, Tasmanian wolf	0

[*] After G. H. F. Nuttall, from *Biochemical Evolution*; G. Wald in *Trends in Physiology and Biochemistry* (E. S. G. Barron, editor), Academic Press, 1952.

On the other hand, the "permissible" degree of change in proteins that require less rigid engineering, such as certain of the serum proteins, the less dynamic elements of tissues such as the collagens and elastins, and various proteins of the hair and skin, is likely to be quite large.

One of the earliest studies of the comparative biochemistry of proteins was carried out by Nuttall and his collaborators. These investigators used immunological techniques to study the phylogenetic relationships between the serum proteins of a wide variety of species. They employed the extent of the precipitin reaction between antihuman serum and the serums of other species as a measure of similarity (Table 13). We know that the precipitin reaction is not an absolutely specific one and, therefore, that cross reactions which do occur do not require the presence of molecules identical to human serum protein molecules. The results suggest that the serum proteins of the species examined form a graded series of macromolecules in which only serums from phylogenetic "neighbors" can cross react significantly. Nuttall strengthened this conclusion by cross-reacting serums from more closely related animals. A dramatic example is his observation that antifrog serum reacts strongly with serums of other tail-less amphibia, but not at all with those of tailed amphibia.

In spite of the complete lack of immunochemical similarity between the serum albumins of distant species, the more obvious functional aspects of the protein may nevertheless be retained. For example, the serum albumins of rat and man carry out such physiological functions as fatty acid binding and transport and osmotic pressure regulation, in essentially the same manner and with equal facility. We may guess, until experimentation has a chance to prove us wrong, that the modifications which led to immunological differences spared, or at least only slightly remodeled, the functionally critical parts of serum albumin structure.

REFERENCES

1. T. G. Lee and A. B. Lerner, *J. Biol. Chem.*, **221**, 943 (1956).
2. L. T. Skeggs, Jr., K. E. Lentz, J. R. Kahn, N. P. Shumway, and K. R. Woods, *J. Exptl. Med.*, **104**, 193 (1956).
3. D. F. Elliott and W. S. Peart, *Nature*, **177**, 527 (1956).
4. These studies are summarized by H. Tuppy in *Symposium on Protein Structure* (A. Neuberger, editor), Methuen, London, 1958.
5. H. Theorell, *Biochem. Z.*, **298**, 242 (1938); *Enzymologia*, **6**, 88 (1939).
6. R. R. Porter and F. Sanger, *Biochem. J.*, **42**, 287 (1948).