# THE AMINO ACID SEQUENCE OF THE ALPHA CHAIN OF OPOSSUM HEMOGLOBIN

by

Peter Stenzel, M. S., M. D.

A THESIS

Presented to the Department of Biochemistry
and the Graduate Division of the University of Oregon Medical School
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
June 1974

APPROVED:

[redacted]

(Professor in Charge of Thesis)

[redacted]

(Chairman, Graduate Council)

Acknowledgments

I thank Dr. Richard Jones for his patient support which has been kind and generous throughout my graduate studies; I sincerely hope that I can be worthy of his example of professional excellence and fairness with others.

The support given me by Dr. Robert Koler was essential to this work; his interest has contributed a great deal to my personal gratification from the excitement of discovery.

The daily consultations with Dr. Bernie Brimhall on matters great and small were invaluable.

Mr. Jim Joyce helped in many ways especially with collection of opossums as did Dr. Tom Vedvick.

I thank Miss Marie Duerst for technical assistance.

I appreciate discussions with Drs. Nick Bethlenfalvay, John Black, Mike Coates, Dharam Dhindsa, Howard Mason, and James Metcalfe.

I am grateful to Mrs. Mary Buck for the editting and expert typing in addition to countless other tasks performed during the course of this work.

I was supported by National Institutes of Health grants 5-T01-HD00165 and S-P01-AM13173.

Support for supplies came from NIH 5-F03-AM32350. I was very happy to have the personal interest in my graduate program of Ms. Merilyn Hiller at NIH.

TABLE OF CONTENTS

LIST OF TABLES

# INTRODUCTION

## I. Structure-Function Relationships in Hemoglobin

Studies of the amino acid sequence of hemoglobin and of its three-dimensional structure have led to the proposal of detailed theories of the mechanism of binding and release of oxygen (1-3) and carbon dioxide (4) and of the modulation of oxygen binding by hydrogen ion and organic phosphate compounds (2,5,6). A striking feature of the proposed mechanisms is the assignment of individual roles to particular amino acid side chains, especially as they interact with the porphyrin ring, the heme iron, or amino acids on the other polypeptide chains in the tetramer (the interchain contact regions). The natures of certain amino acids are important in helix formation, folding of the subunit into its tertiary structure, and influence on solubility of the molecule. The requirement for particular amino acids at certain positions in the molecule in order for there to be structural stability and proper function is of interest, and it is here that the study of amino acid sequences of different hemoglobins is critical in the evaluation of the proposed theories of hemoglobin function.

The proposed relationships of the amino acid sequence of hemoglobin to its tertiary structure and function is principally due to the work of Perutz and his collaborators on crystals of human and horse hemoglobin in both the liganded and unliganded states and the

molecular models constructed from the results of those studies (1,2,7). Other amino acid sequences may be treated as perturbations of the human and horse hemoglobin structures. The differences in three-dimensional structure and in the interactions listed above may be predicted by changes in the molecular model necessitated by changes in amino acid residues. The structural alterations may then be correlated with functional differences and similarities. Hemoglobins for comparison are available from two sources, mutant human hemoglobins and hemoglobins of various animals.

The mutant human hemoglobins, in nearly all instances variants from normal in one amino acid, provide an opportunity to examine the results of varying two amino acids placed symmetrically in the tetramer of 574 amino acids. The amino acid substitution has been identified in more than 100 of these variants (8-11). Frequently the substitution has no detectable effect on the stability or function of the molecule; these substitutions are usually on the exterior of the molecule and not involved in critical intramolecular interactions, or the innocuous effect can be attributed to the substituted amino acid's structural similarity to the residue normally present. Other substitutions have profound effects on stability or function; when the substitution is of a similar amino acid, the structural requirement at the altered site is seen to be very specific. An explanation for the deleterious effect can usually be found by consideration of the molecular model (11,12).

In a few instances the abnormal hemoglobin has been studied crystallographically; the results have generally confirmed predictions from the model (13-16). Thus, study of the mutant hemoglobins can provide refinements of theory of hemoglobin function and, in turn, the theory frequently provides an explanation for the pathologic function of an aberrant molecule.

Complete or nearly complete amino acid sequences have been proposed for more than 40 vertebrate hemoglobin polypeptide chains (10,17-20). These sequences differ from those of horse or human by an amount which increases with decreasing relatedness of the species as estimated by conventional taxonomic considerations (10). Secondary and tertiary structures of the various vertebrate hemoglobin chains may be assumed to be very nearly identical to those in human and horse hemoglobins; this assumption is suggested by similarities in sequence and is justified by X-ray crystallographic studies of lamprey hemoglobin (21,22).[1] The variety of amino acids at corresponding sites in the several sequences may be interpreted as an index of the flexibility in the structural requirements at those sites. At the extreme of rigid requirements are the six sites which contain the same amino

[1] Comparative structural studies have been extended to invertebrates with the chemical and crystallographic investigation of hemoglobins from larval *Chironomus thummi thummi* (27,28) (a dipteran insect) and *Glycera dibranchiata* (22,29) (an annelid worm). These monomeric hemoglobins share basic secondary and tertiary structures with vertebrate hemoglobins and myoglobins.

acid in all known vertebrate hemoglobin and myoglobin sequences (23) (Dickerson [23] lists seven, but an exception has been found for the requirement of a glycine residue in the sixth position of the B helix; kangaroo alpha chain has an alanine in this position [24]). Other sites are invariant for all alpha or all beta chains from tetrameric hemoglobins. The sites at which some variability is observed could alternatively be viewed as specific in structural requirements but with different species having requirements to fit different functions. While species similarities in sequence can be explained, the differences in relation to differences in oxygen binding and modulation of binding have not been satisfactorily explained. (An exception may be diphosphoglycerate binding and the amino terminal residue of the beta chain [2,25,26]).

## II. Evolution of Hemoglobin

The central idea in the concept of molecular evolution (30) is that, within a taxonomic group of organisms, there are polypeptides the structural genes of which have descended from a common ancestral gene in the same way that organisms in that group have descended from a common ancestor in the theory of evolution of species. It may be that the members of a species apparently have only one detectable expressed copy of the descendants of that gene (e.g., cytochrome c) or two or more copies which have arisen by two or more gene duplications (e.g., the different hemoglobin chains and myoglobin in a species).

The genes are different in the several organisms as a result of the incorporation of mutations during divergent evolution. Where there are multiple copies (i.e., more than one descendant of the ancestral gene), these may be expressed as multiple versions of a polypeptide with indistinguishable functions (e.g., the two human hemoglobin gamma chains from different loci [31]) or as polypeptides which have diverged in function (e.g., myoglobin and the beta chain of hemoglobin in humans). Thus, for vertebrate hemoglobins and myoglobins, an evolutionary tree of globin chains may be pictured as branched within the branches of the evolutionary tree of vertebrate species. The set of polypeptide chains descended from a common ancestor is called a set of homologous polypeptides.

The rate of evolution of a particular polypeptide with respect to two species, for which the amino acid sequence of the polypeptide is known, is estimated as follows: (32) the number of differences in the amino acid sequences is halved and divided by the elapsed time from divergence of the evolutionary lines descending to the present species, i.e., the time since the existence of the common ancestral gene. The time of divergence is taken from conclusions of paleontological investigations. By taking the average of all the rates from pairs of homologous polypeptides in a set, a rate is derived for the entire set. Rates of evolution for different sets of homologous polypeptides may be compared if the rate for each set is divided by the length of the polypeptides in that set, so that the units for

comparison are of the type, amino acid substitution per amino acid of length per year.

Zuckerkandl and Pauling (30) first noted that, for the hemoglobin chain amino acid sequences then available, a very uniform rate of evolution is computed. Kimura (33) again noted this uniform rate, computed from more extensive sequence data, and added the observation that the number of sequence differences between any alpha chain and any beta chain are nearly the same for all comparisons, whether the compared chains are taken from the same or different species. He concludes that the alpha and beta chains have diverged at the same rate along different paths of species evolution. Recently, Air, Thompson, Richardson and Sharman (39) have compiled the number of sequence differences for hemoglobin alpha and for hemoglobin beta chains, taking care to include only sequences which have been definitely established, omitting the sequences inferred in part by homology and those not differentiating aspartic and glutamic acids from their amides in some positions. Again, these authors noted a narrow distribution of evolutionary rates for pairs of sequences compared. The table here is that of Air et al. (34), revised to include data from the recently completed kangaroo (24), dog (35), echidna (19), and catostomid fish (36) alpha chain sequences.

Kimura (37) interpreted the observed uniformity for rates of evolution of hemoglobin chains, calculated in the above manner, as

| | Human | Monkey | Horse | Bovine | Rabbit | Dog | Kangaroo | Echidna | Chicken | Carp | Catostomid fish |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Human | | 4 | 18 | 17 | 25 | 23 | 27 | 38 | 35 | 71 | 68 |
| (Rhesus) monkey | | | 16 | 16 | 25 | 24 | 26 | 36 | 35 | 71 | 68 |
| Horse | | | | 18 | 25 | 27 | 29 | 42 | 40 | 70 | 66 |
| Bovine | | | | | 25 | 28 | 26 | 44 | 38 | 68 | 66 |
| Rabbit | | | | | | 28 | 37 | 49 | 44 | 74 | 71 |
| Dog | | | | | | | 33 | 39 | 45 | 70 | 69 |
| Kangaroo | | | | | | | | 49 | 41 | 74 | 73 |
| Echidna | | | | | | | | | 47 | 79 | 73 |
| Chicken | | | | | | | | | | 75 | 71 |
| Carp | | | | | | | | | | | 16 |

Table 1. Numbers of Amino Acid Differences Between Pairs of Vertebrate Alpha Chains. The (130 Alanine) dog chain was used (35); the two relative insertions and one relative deletion of the fish chains were counted as one difference each.

evidence for point mutations becoming fixed in the genome of a species at a uniform rate (with respect to time rather than to number of generations) during the evolution of that species. He computed rates of evolution for triosephosphate dehydrogenase (38) and found the average for that rate, the rate for hemoglobin chains, and a rate computed for cytochrome c (39). He then applied the average rate to Muller's (40) estimate of the number of nucleotide pairs in the human haploid chromosome complement and concluded that, in the evolutionary history of mammals, on the average, one nucleotide pair has been substituted in the population roughly every two years. Consideration of the principle of the so-called cost of natural selection (due to Haldane [41]) brought Kimura to the conclusion that differences in amino acid sequences between homologous polypeptides of different species have come about chiefly through the fixation of selectively neutral point mutations. He developed a mathematical model for fixation of neutral mutations through a Markov process which predicts that, for neutral mutations, the rate of gene substitution in a population is equal to rate of production of new mutations per gamete.

King and Jukes (42) analyzed the amino acid substitutions at sites of the cytochrome c sequence which are different for at least two species. They found that the distribution of the number of substitutions per site fits a Poisson distribution. They offer this finding as support for Kimura's theory that the chief cause of

molecular evolution is the random fixation of selectively neutral mutants.

The derivation of the mathematical model in Kimura's theory has been both supported (43) and criticized (44) within the framework of theoretical population genetics. The basis for much of other criticism (45,46) has been chiefly that the presence of a particular amino acid at a given site in a protein cannot be judged as non-adaptive simply because no advantage can be envisioned in terms of what is known of the mechanism of action of that protein.

Kimura (33) has pointed out a prediction from his theory which may be tested: the rate of evolution of polypeptides produced by those species which, by morphologic criteria have evolved especially slowly, i.e., living fossils,[1] should be as great as the rate for polypeptides from more rapidly evolving species such as most of the mammals. He reasons that it is likely that changes in internal physiology will parallel changes in external form, at least to some degree, with respect to evolutionary rates. The species' evolutionary

[1] Kimura lists as living fossils the lobe-finned fish, the horseshoe crab, and *Lingula*, a lamp shell. Simpson (47) discussed the role of these slowly-evolved species, which he calls "immortals," in the study of evolution and mentions, in addition, the oyster, the opossum, and *Sphenodon*, a lizard-like creature of New Zealand. The conclusive criterion for inclusion of a recent species in the class of living fossils is the knowledge of a fossil from the distant past which closely resembles the present species in form. Fossil evidence reported in 1968 (48) permits the addition of a recent lamprey species to this list; this addition is important to the present discussion, since hemoglobin sequences have been determined for two species of lamprey (49,50)

rate is determined by selection intensity and is measured by the number of genera along the evolutionary line leading to a given species. Molecular evolution which proceeds mainly by fixation of neutral mutations should be a function of elapsed time only and be independent of the number of ancestral genera of the species taken for study. The foundation of Kimura's theory is the narrow distribution of computed rates of molecular evolution for a set of homologous polypeptides. Of course each newly determined amino acid sequence or a member of that homologous set permits another computation of the rate of molecular evolution and thus permits a test of the theory, but the determination of the sequence of a homologous polypeptide derived from a living fossil presents a particularly critical test. A species showing slow rates of both species and molecular evolution would provide a strong argument for an alternative explanation of molecular evolution -- that the observed mutations were fixed by natural selection. The case for the opossum as a choice of such a living fossil to be used for testing Kimura's theory will follow a synopsis of the relevant paleontological information.

A geological time scale and diagrams of vertebrate evolutionary history are attached as appendices. The paleontologic conclusions here are taken from publications of Simpson (47), Romer (51), and Stebbins (52), unless noted otherwise. The vertebrates

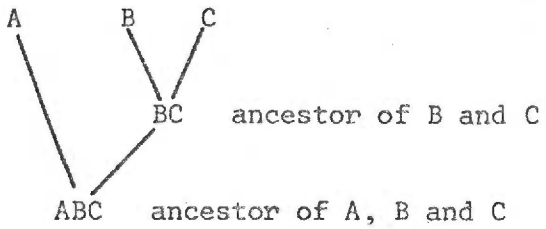
first appear in the fossil record of the Ordovician period; it is also during that period that the lamprey-like forms appear. The common ancestor of the terrestrial animals and the carp must be placed in the Devonian, since it was during this period that the bony fish appeared[1] and that the amphibia (ancestors of the reptiles and thence the mammals) arose from the bony fish. Stebbins (52) lists the dates spanning the Devonian as 405 and 355 million years ago. Since such dates must be taken as accurate only within millions of years, the figure of 350 million years used by Kimura (33) for the time elapsed from the divergence of the ancestor of the mammals and the ancestor of the carp (a recent bony fish) is reasonable. The mammals (except for the egg-laying monotremes)[2] are believed to have arisen from one reptilian group, the Pantotheria, near the division between the Triassic and Jurassic periods. The divergence of the marsupials and the placental mammals took place during the Cretaceous period, perhaps as early as the beginning of that period (34,54,55). The relationship between the marsupials in North and South America on one hand and the Australian marsupials on the other is not clear; this is due largely

[1] Romer (53) presents an argument supporting an earlier appearance of the bony fish, during the Silurian period, but the absence of such fossils from that period leaves the question problematical.

[2] Simpson (47) states, "The quaint living monotremes, platypus and echidna, of Australia probably did not have this origin, although some authorities think they did. Their origin is really quite unknown, but my own (not particularly original) suspicion is that they are rather highly modified surviving therapsid reptiles, mammals by definition rather than by ancestry."

to the lack of Australasian marsupial fossils which can be dated earlier than the late Oligocene. Martin (56) has considered recent descriptions of geotectonic zones in the Pacific Ocean and suggested that marsupials arose on a land mass situated over the Darwin rise and that land fragments subsequently drifted apart, carrying the marsupial inhabitants to Australia and to North America; this hypothesis would place the separation of the two groups of marsupials near the beginning of the Cretaceous.

Sequences taken for a living fossil test of Kimura's theory must be homologous with known sequences and must be from species whose evolutionary relationships are such that they follow the topological relationship in the evolutionary tree of the vertebrates as in this diagram:




The requirement with respect to rates of evolution of the species is that B has evolved much more slowly from BC than C has. Since the data underlying Kimura's theory have been taken largely from the globins, cytochromes c, and fibrinopeptides of vertebrates, it is appropriate to select for study one of these polypeptides from a

vertebrate species which meets the above topological requirements with respect to the species for which sequences are known. The close resemblances of the opossum and the lizard-like *Sphenodon* to their ancestral forms of 80 and 135 million years ago (47), respectively, make these two species the outstanding candidates for study. The class of vertebrates for which the greatest number of sequences is available is Mammalia. Because the opossum shares many important physiological characteristics, such as homiothermy, with the other mammals, the choice of this species rather than *Sphenodon* minimizes variables which may be unrelated to evolutionary rates, when comparisons are made with the mammalian polypeptide sequences. Furthermore, specimens of opossum are vastly more readily available for study. Of the polypeptides listed above, the hemoglobin chains contain the greatest number of variable sites and thus offer the greatest possibility for evaluating the significance of numbers of differences between sequences taken for comparison. The known hemoglobin polypeptide sequences which may be taken for point C in the above topology are chicken (57), carp (58), echidna (19), and catostomid fish (36) alpha chains; frog (59) and echidna (18) beta chains; and the beta-like chains of two species of lamprey (49,50). Since the lamprey hemoglobin differs a great deal from the mammalian hemoglobins in its quaternary structure and other properties (discussion in ref. 49), the alpha chain was chosen for sequence determination and subsequent

computations of molecular evolution rates in order to test Kimura's theory.[1]

[1] Echidna alpha and beta chains and catostomid fish alpha chain sequences were not available at the initiation of the sequence studies described in this thesis.

## MATERIALS AND METHODS

Carboxymethylcellulose (CM52) was from Reeve Angel; Aminex cation exchange resins A-5, 50W-X2 (200-325 mesh), and 50W-X4 (20-30 μ) were from Bio-Rad. Trypsin, TPCK-trypsin[1], and chymotrypsin were from Worthington; thermolysin was from Calbiochem; subtilisin BPN' and elastase were from Sigma; aminopeptidase M was from Rohm and Haas, Darmstadt, Germany. 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide for the glycinamidation reaction was from Pierce, as were the ADMA, PITC, and TFA; all were Sequanal grade. All other chemicals were reagent grade. Pyridine was redistilled over ninhydrin.

Opossums captured in the wild in the vicinity of Portland, Oregon, were anesthetized with sodium pentobarbital and blood taken by cardiac puncture. Erythrocytes were washed and hemolysate prepared according to Jones, Brimhall, and Huisman (60). The hemolysate was chromatographed on carboxymethyl-Sephadex. Globin was prepared with HCl-acetone as described by Schroeder et al. (61) from hemolysate which had been dialysed against at least 50 volumes of water at 6° for 24 h and centrifuged at high speed. Chain separation was by the method of Clegg, Naughton and Weatherall (62), using carboxymethylcellulose and 8 M urea.

[1] The abbreviations used are: TPCK, L-1-tosylamido-2-phenylethyl chloromethyl ketone; ADMA, N-allyl-N-N-dimethylamine; PITC, phenylisothiocyanate; TFA, trifluoroacetic acid.

Acid hydrolysis was done in 6N HCl containing 0.009% (w/v) phenol in ampoules sealed after evacuation and heated at 110°, for 18-24 hours unless stated otherwise. Amino acid analyses were made essentially according to Spackman, Stein and Moore (63) with a Spinco Model 120 amino acid analyzer equipped with 20mm light path flow cells (64).

Glutaminyl and asparaginyl residues were distinguished from glutamyl and aspartyl residues, respectively, by one of two methods: (1) Complete hydrolysis with aminopeptidase M essentially according to Light (65) with 16 hours of digestion at 37° followed by amino acid analysis of the digestion product. Glutamine and asparagine eluted with threonine and serine, respectively, so that the results were interpreted in view of the results of a companion HCl hydrolysate analysis. (2) Glycinamide was quantitatively coupled to free -COOH groups of a chain fragment by the method of Mross (66) as described by Brimhall et al. (67). Following subsequent enzymic cleavage of the fragment any peptide, not containing the COOH- terminus of the fragment, yielded upon acid hydrolysis one mole of glycine (in addition to any glycine present in the corresponding peptide which had not been glycinamidated) for each aspartyl or glutamyl residue present in a mole of the original peptide. The failure to find such glycine residues was taken to indicate that any aspartic or glutamic acid appearing upon acid hydrolysis arose from asparaginyl and glutamyl

residues, respectively, in the peptide. Both methods were applied to some peptides.

Determination of the $NH_2$-terminus of the chain was done by Stark's procedure (68), recovering the free N-terminal amino acid from the hydantoin after reaction of the protein with cyanate.

Digestion of peptides with dilute acetic acid in order to produce cleavage at aspartyl residues (69) was done under conditions described with the results of such applications.

Selective cleavage of the chain at argininyl and (aminoethyl) cysteinyl residues was effected by blocking of the lysine epsilon, and terminal amino groups with citraconic anhydride (70,71), aminoethylation of the cysteine groups (72), and tryptic digestion. To 500 mg of chain suspended in 8 ml water were added seven 0.1 ml portions of citraconic anhydride at 10 minute intervals; the pH was maintained at ca. 8.5 by dropwise addition of 5N NaOH. After the last addition, when the pH became stable, 12 g of urea were added to the 12 ml. Reduction with mercaptoethanol and aminoethylation was then done according to Cole (73) followed by dialysis at room temperature against two changes of 100 volumes of water adjusted to pH 8.5 with 2.5% trimethylamine over 12 hours. Four mg TPCK.trypsin were added and digestion allowed for 1 hour at 25°. The pH was adjusted to pH 3.5 with acetic acid and the resulting suspension stirred for 3 hours at 25°. The suspension was centrifuged at low speed and the precipitate

dissolved in 88% formic acid and combined with supernatant to make the resulting solution 7-15% in formic acid. This was then applied to a Sephadex G-50 column, using no more than 30 ml for a chromatogram.

Cyanogen bromide was added to an equal weight of chain (74) in 70% formic acid at 100 mg protein per ml, left in the dark at room temperature for 24 hours, diluted five-fold with water and applied to a Sephadex G-50 column. The resulting fragments were further purified by recycling on the same column.

Digestions with proteolytic enzymes were done in these buffers: trypsin, chymotrypsin, and elastase in water titrated to pH 8.4-8.6 with 2% trimethylamine; thermolysin in 28 mM Tris, 6 mM $CaCl_2$ $2H_2O$ titrated to pH 8.0 with dilute HCl; subtilisin BPN' in 1% $(NH_4)_2CO_3$ titrated to pH 8.2 with 1 M acetic acid. The volume of the digestion mixtures were generally as small as convenience of pH determination would allow (usually 1.5-5 ml). Further conditions of digestion are given with the results.

Chromatography of the peptides on sulfonated polystyrene cation exchange resins with volatile buffers and automatic detection of emerging peptides with ninhydrin either directly or following automatic alkaline hydrolysis was done as described by Jones (75). Samples known to contain large amounts of ammonia were dried by rotary evaporation and redissolved in pyridine several times before being applied to an ion exchange column.

Subtractive Edman degradations (76) followed this scheme in which coupling buffer is 0.4 M ADMA in 50% (v/v) pyridine titrated to pH 9.4 with 10% (v/v) TFA:

(1) Peptide (0.01-2 μmoles) was dissolved in 0.4 ml coupling buffer in an acid-dichromate washed, conical 15 ml centrifuge tube. Further ADMA is added if necessary to reach pH 8.5 as judged by indicator paper (pHydrion papers, Micro Essential Laboratory, Brooklyn, N.Y.); narrow pH range indicator paper was unsatisfactory so that pH 2-11 paper was used exclusively.

(2) The space above the solution was flushed well with nitrogen; 2 drops (30-60 mg) PITC were added and another brief nitrogen flush was done; the tube was covered with Parafilm and incubated for 20 minutes at 50°. The solution was vacuum dried at room temperature for at least three hours; any liquid still present was ignored. Usually the sample became a powder or a thin film on the tube wall.

(3) TFA (0.1-0.2 ml) was added; the tube was flushed with nitrogen, covered with Parafilm and incubated for seven minutes at 50°. The TFA was removed in a gentle stream of nitrogen. One-half minute after the tube no longer felt cold due to evaporation, the sample was taken to step (1) even if liquid persisted.

Sampling of the peptide was done after step (1) by adding a fraction of the solution to 1 ml of water and extracting with 2 ml of ethyl ether (determined to be free of peroxides by the absence of

color formation upon shaking with an equal volume of 4% potassium iodide). Often a precipitate formed which was discarded after centrifugation. The aqueous phase was taken for HCl or aminopeptidase digestion and amino acid analysis. If a free amino acid had been anticipated, as in the case of a tripeptide exposed to two degradation cycles, the sample was analyzed directly.

# RESULTS

## I. Preparation and Analysis of the Alpha Chain

*CM-Sephadex Chromatography of Hemolysate*--Chromatography of the hemolysate on CM-Sephadex (Figure 1) produced only one significant fraction of hemoglobin. Since variations of the shape and slope of the phosphate buffer gradient also failed to demonstrate multiple hemoglobin components, globin was prepared directly from hemolysate.

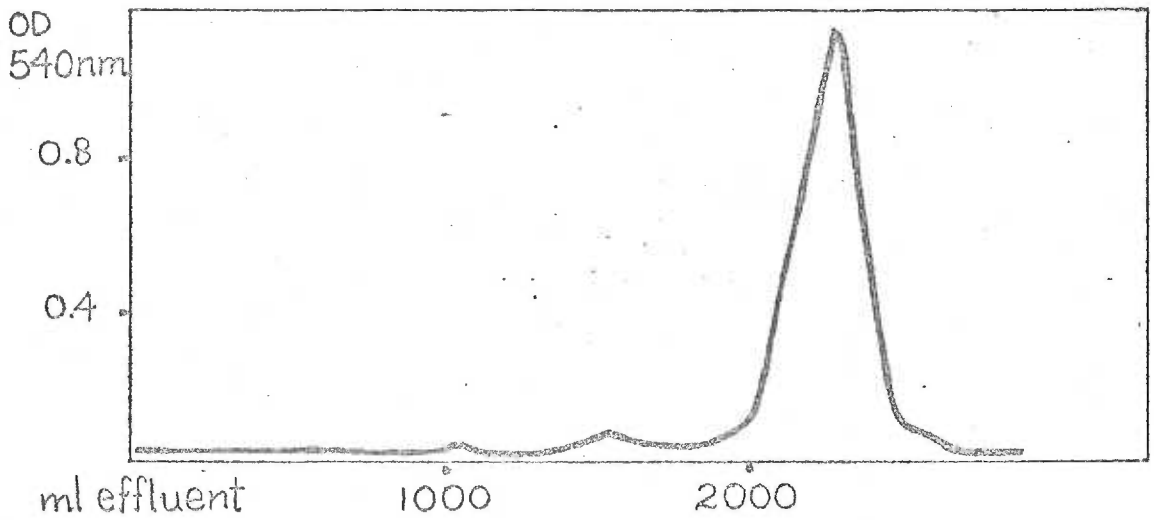
*Chain Separation*--Chromatography of globin on CM-cellulose in 8 M urea (Figure 2) yielded several components. The pattern varied somewhat, even with equal aliquots of the same globin preparation under conditions thought to be constant. The most striking variation was that of the degree of resolution of the components labeled α and α' which ranged from that of peak and barely perceptible shoulder to that shown in Figure 2. The size of the peak between components α and β also varied. The material corresponding to the α component was desalted on Sephadex G-25, lyophilized, and taken for sequence studies.

*$NH_2$-Terminus Determination*--The $NH_2$-terminal residue was shown to be valine by Stark's (68) cyanate method.

*Amino Acid Analysis of the Chain*--Amino acid analyses of aliquots of aminoethylated chain hydrolyzed in 6 N HCl for 24, 48, and 72 hours are reported in Table 2.

Figure 1. Chromatography of 2 g hemoglobin on CM-Sephadex C-50 (5 x 66 cm) at 10° developed at 80 ml per hour with a 2500 ml linear gradient from 0.05 M $NaH_2PO_4 \cdot H_2O$, 0.025 M $Na_2HPO_4$, 0.001 M KCN to 0.100 M $Na_2HPO_4$, 0.001 M KCN, followed by 500 ml of the latter buffer.

Figure 2. Separation of the chains from 600 mg of globin on CM-cellulose (5 x 10 cm) at room temperature developed at 200 ml per hour with a 3000 ml linear gradient from 0.01 to 0.05 M $Na_2H\ PO_4$. Both buffers contained 8 M urea and 0.05 M 2-mercapto-ethanol and were adjusted to pH 6.50 with $H_3PO_4$.


OD
540nm
0.8
0.4
ml effluent
1000
2000

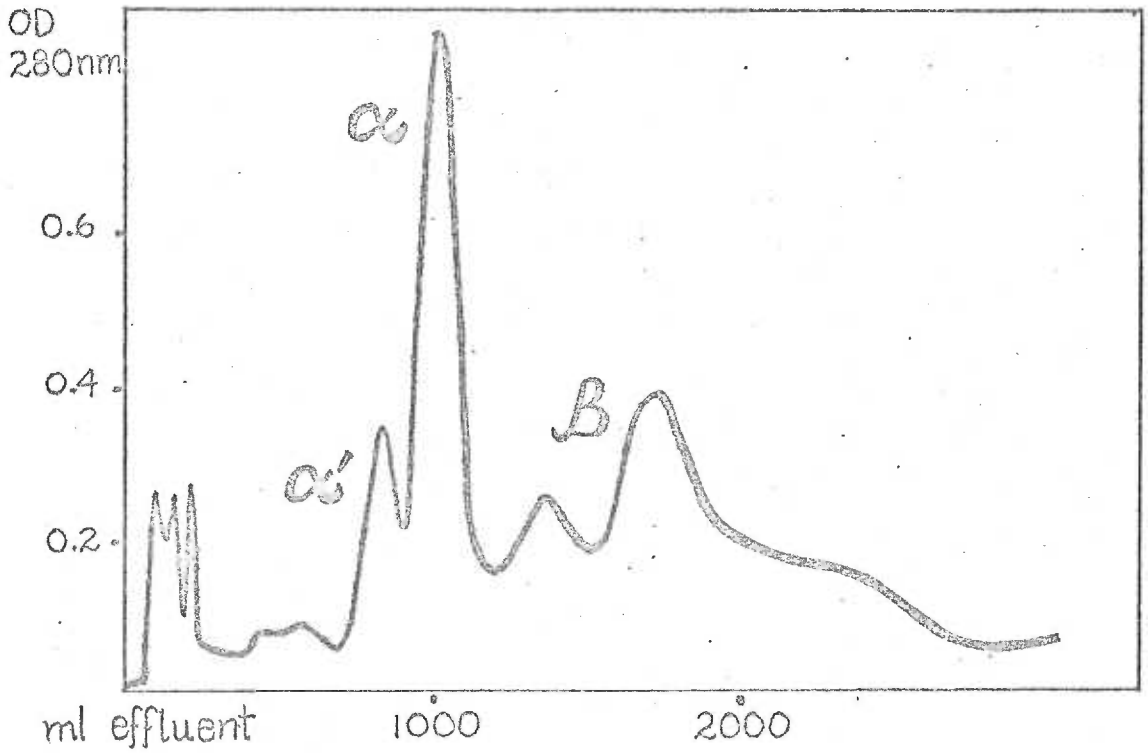

OD
280nm
0.6
0.4
0.2
α
α'
β
ml effluent
1000
2000

Table 2. Amino Acid Analyses of Aminoethylated Opossum Alpha Chain.*

| | 24 hours | 48 hours | 72 hours | Average** | Expected*** |
|---|---|---|---|---|---|
| Lys | 10.6 | 10.9 | 10.7 | 10.7 | 11 |
| His | 5.5 | 5.5 | 5.5 | 5.5 | 6 |
| AE Cys | 1.0 | 0.9 | 1.0 | 1.0 | 1 |
| Arg | 2.1 | 1.9 | 2.0 | 2.0 | 2 |
| Asp | 14.6 | 14.6 | 14.3 | 14.5 | 15 |
| Thr | 11.0 | 10.4 | 9.4 | 11.0 | 11 |
| Ser | 12.2 | 11.6 | 9.4 | 12.2 | 13 |
| Glu | 6.8 | 6.7 | 6.5 | 6.7 | 6 |
| Pro | 4.9 | 5.0 | 5.1 | 5.0 | 5 |
| Gly | 9.0 | 9.0 | 9.0 | 9.0 | 9 |
| Ala | 15.2 | 15.5 | 15.4 | 15.4 | 16 |
| Val | 10.0 | 10.9 | 10.9 | 10.9 | 11 |
| Met | 1.2 | 1.2 | 1.2 | 1.2 | 4 |
| Ile | 2.2 | 2.4 | 2.5 | 2.5 | 3 |
| Leu | 14.0 | 13.9 | 14.1 | 14.0 | 14 |
| Tyr | 5.1 | 5.1 | 4.9 | 5.0 | 5 |
| Phe | 8.0 | 8.2 | 7.8 | 8.0 | 8 |
| Try | present | present | present | | 1 |

* Expressed as molar ratios, normalized to 9 glycine residues.

** 24 hour values for serine and threonine; 72 hour values for valine and isoleucine.

*** Values derived from proposed sequence.

II. Preparation of Fragments by Selective Cleavage

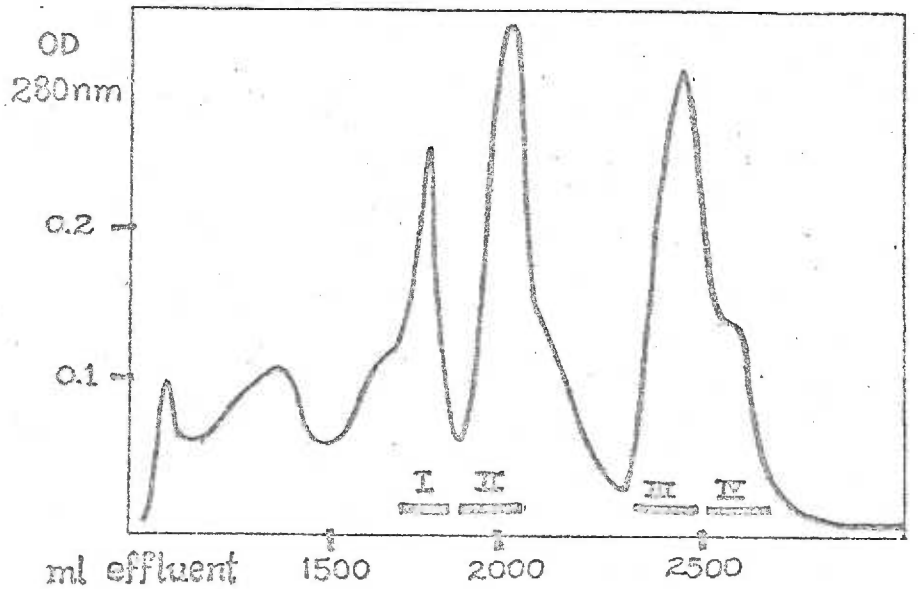
*Nomenclature*--The whole chain cyanogen bromide cleavage products, CB I, CB II, CB III, CB IVA, and CB IVB were so labelled for their positions in the chromatograms in Figure 3. Naming of the fragments produced by selective tryptic cleavage following citraconylation and aminoethylation was similarly motivated with respect to Figure 4; these became CAT I, CAT II, and CAT III. The products of cyanogen bromide cleavage of CAT I are called Ccb I and Ccb II (Figure 5); acetic acid cleavage of Ccb II yielded CcH I and CcH II (Figure 6).

*Cyanogen Bromide Cleavage*--Separation of fragments produced by cyanogen bromide treatment of chain is depicted in Figure 3; amino acid analyses of these fragments appear in Table 3.

*Selective Tryptic Cleavage*--Figure 4 shows gel filtration of the fragments produced by tryptic cleavage at (aminoethyl) cysteinyl and argininyl bonds. The corresponding amino acid analyses are in Table 4.

*Further Cleavage of CAT I*--Figure 5 shows the separation on Sephadex G50 of the products of cyanogen bromide cleavage of CAT I, one of which yielded the further cleavage products by acetic acid which were separated by the procedure illustrated in Figure 6. Corresponding amino acid analyses are in Table 4.

Figure 3. Isolation of fragments resulting from cyanogen bromide digest of chain by gel filtration on Sephadex G-50 fine (5 x 180 cm) in 7% formic acid at room temperature. The upper panel shows chromatography of the cyanogen bromide digest of 280 mg chain. The bars denote fractions taken for further purification by recycling five times through the same column after being lyophilized and redissolved in 7% formic acid. The column was put in series with a peristaltic pump and a flow cell in a spectrophotometer equipped with a recorder for monitoring absorption of light at 280 nm. The lower panel shows such recycling of zone IV from 500 mg chain resulting in the isolation of zones IVA and IVB as denoted by the bars. During collection on the fifth cycle, approximately 1% of the stream was taken for automatic alkaline hydrolysis, reaction with ninhydrin and monitoring of the product at 570 nm (beaded line). The flow rate was approximately 240 ml per hour.


OD
280nm
0.2
0.1
I
II
III
IV
ml effluent
1500
2000
2500

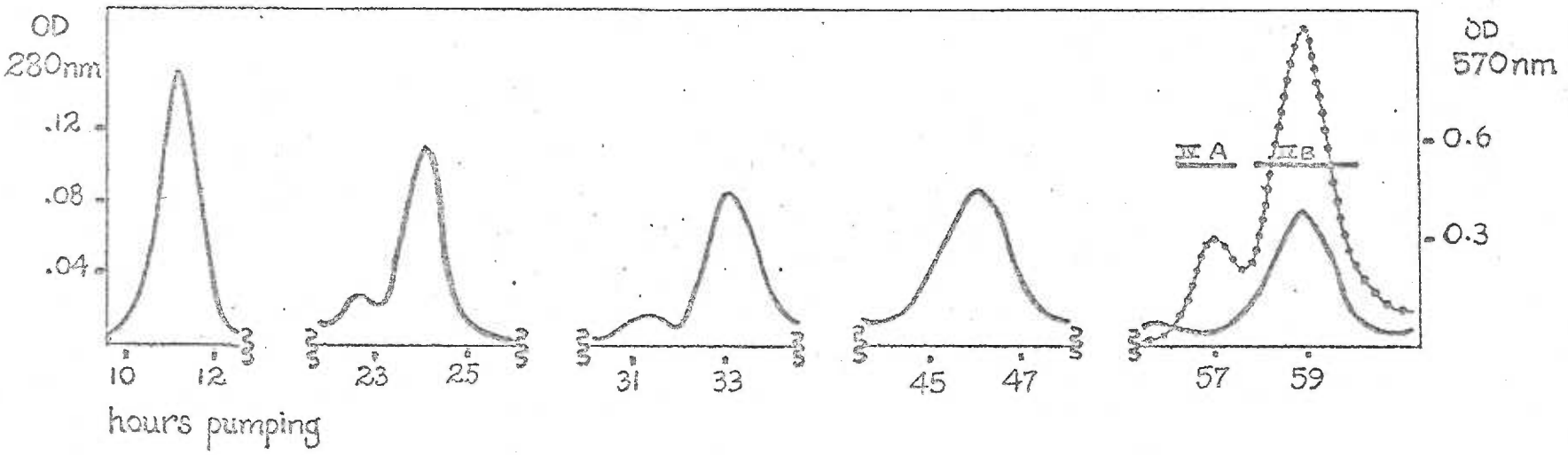

OD
280nm
.12
.08
.04
10
12
23
25
31
33
45
47
IIIA
IIIB
57
59
OD
570nm
0.6
0.3
hours pumping

Table 3. Amino Acid Analyses of Cyanogen Bromide Fragments of Opossum Alpha Chain. HCl digests of the fractions shown in Figure 3. Hsl and Hsr denote homoserine lactone and homoserine, respectively. Glutamic acid and homoserine were not resolved. Each analysis is compared with the composition of the fragment in each zone as determined by: (1) subsequent isolation and sequence studies of tryptic peptides from the material in each zone (2) similar studies of CAT fragments (3) homology with other vertebrate alpha chains. Recovery of tryptophan is not quantitative following HCl hydrolysis but frequently is sufficient for detection (77).

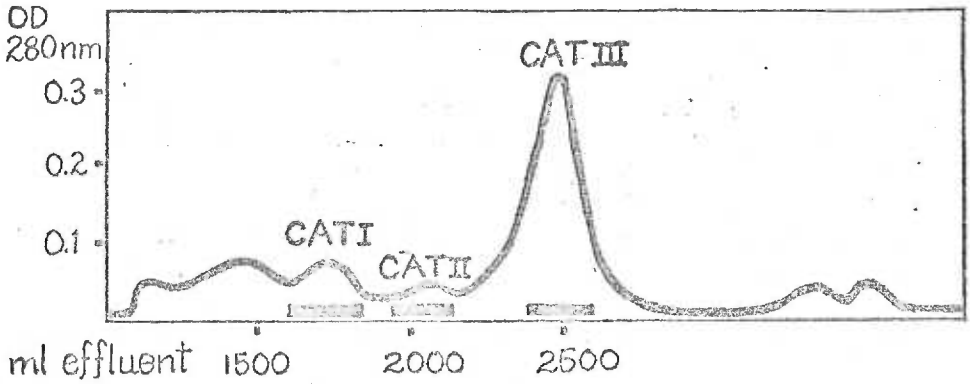
| | Zone I | [1-25,77-109] | Zone II | [26-76] | Zone III | [1-25] | Zone IVA | [110-125] | Zone IVB | [126-141] |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | 4.9 | 5 | 2.9 | 3 | 3.0 | 3 | 1.0 | 1 | 2.0 | 2 |
| His | 3.1 | 3 | 1.0 | 1 | 0.2 | 0 | 1.5 | 2 | 0 | 0 |
| Hsl | 0.6 | | 0.4 | | 0.3 | | 0.3 | | 0 | |
| Arg | 0 | 0 | 1.1 | 1 | 0 | 0 | 0 | 0 | 1.0 | 1 |
| Asp | 8.2 | 8 | 5.0 | 5 | 4.0 | 4 | 1.1 | 1 | 1.1 | 1 |
| Thr | 3.2 | 3 | 4.7 | 5 | 1.0 | 1 | 1.0 | 1 | 2.1 | 2 |
| Ser | 5.9 | 6 | 3.3 | 3 | 3.0 | 3 | 1.1 | 1 | 2.7 | 3 |
| Glu & Hsr | 2.1 | 1 | 4.6 | 4 | 0.4 | 0 | 1.3 | 1 | 0 | 0 |
| Pro | 2.0 | 2 | 2.0 | 2 | 0.1 | 0 | 1.0 | 1 | 0 | 0 |
| Gly | 4.0 | 4 | 4.1 | 4 | 4.0 | 4 | 1.0 | 1 | 0.1 | 0 |
| Ala | 5.0 | 5 | 6.7 | 7 | 3.2 | 3 | 3.0 | 3 | 1.1 | 1 |
| Cys 1/2 | 0.7 | 1 | 0 | 0 | 0 | 0 | 0.1 | 0 | 0 | 0 |
| Val | 6.7 | 7 | 2.0 | 2 | 2.7 | 3 | 0.1 | 0 | 2.0 | 2 |
| Met | 0 | 2 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| Ile | 0 | 0 | 1.9 | 2 | 0 | 0 | 0.6 | 1 | 0 | 0 |
| Leu | 7.0 | 7 | 4.3 | 4 | 1.2 | 1 | 1.0 | 1 | 2.0 | 2 |
| Tyr | 0.9 | 1 | 3.0 | 3 | 1.0 | 1 | 0 | 0 | 1.0 | 1 |
| Phe | 2.1 | 2 | 4.1 | 4 | 0 | 0 | 0.9 | 1 | 1.1 | 1 |
| Try | 0 | 1 | 0 | 0 | present | 1 | 0 | 0 | 0 | 0 |

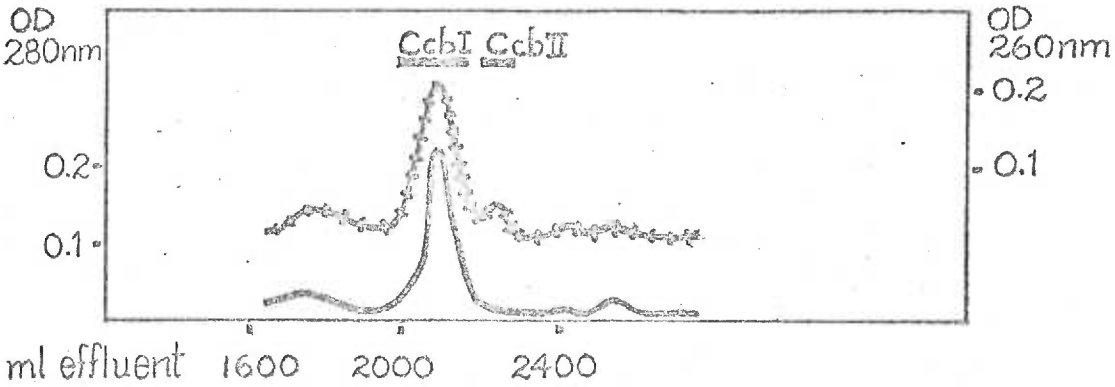
Figure 4. Gel filtration on Sephadex G-50 fine of the fragments produced by tryptic cleavage of 250 mg chain at (aminoethyl) cysteinyl and argininyl bonds. Column conditions are given with Figure 3. Bars indicate fractions taken.

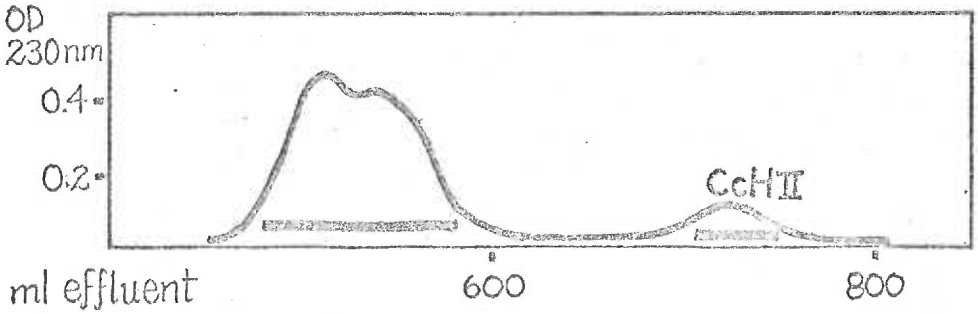
Figure 5. Gel filtration on Sephadex G-50 fine of the products of cyanogen bromide treatment of 130 mg CAT I. Column conditions are given with Figure 3. Beaded line indicates 260 nm absorption. Bars indicate fractions taken.

Figure 6A. Gel filtration of 30% acetic acid digest (48 hours, 50°) of Ccb II. The column was 2.5 x 200 cm and developed with 1.2% acetic acid at room temperature; the flow rate was 80 ml per hour. Bars indicate fractions taken. Sephadex G-25 fine was used.

Figure 6B. Gel filtration of the first fraction indicated in Figure 6A. Three Sephadex G-25 fine (2.5 x 200 cm) columns in tandem were developed with 1.2% acetic acid at room temperature; the flow rate was 20 ml per hour. The bar indicates the fraction taken.


OD
280nm
0.3
0.2
0.1
CAT III
CAT I
CAT II
ml effluent
1500
2000
2500


OD
280nm
0.2
0.1
CcbI CcbII
OD
260nm
0.2
0.1
ml effluent
1600
2000
2400


OD
230nm
0.4
0.2
CcHII
ml effluent
600
800

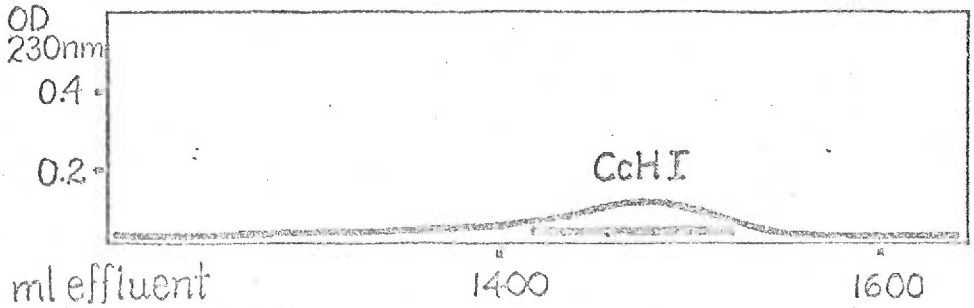

OD
230nm
0.4
0.2
CcHI
ml effluent
1400
1600

Table 4. Amino acid analyses of HCl digests of fractions indicated in Figures 4, 5, and 6. AE cys denotes aminoethylcysteine. Other abbreviations and tryptophan determination as in Table 3.

| | CAT I | [32-102] | CAT II | [103-141] | CAT III | [1-31] | Ccb I | [32-76] | Ccb II | [77-102] | CcH I | CcH II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lys | 5.0 | 5 | 3.3 | 3 | 3.0 | 3 | 2.8 | 3 | 2.1 | 2 | 1.1 | 1.0 |
| His | 3.2 | 3 | 2.7 | 3 | 0 | 0 | 1.2 | 1 | 2.0 | 2 | 1.9 | 0 |
| Hsl | 0 | | 0 | | 0 | | 0.3 | | 0 | | 0 | 0 |
| AE Cys | 0.9 | | 0 | | 0 | | 0 | | 0.9 | | 0 | 1.0 |
| Arg | 0.1 | 0 | 0.9 | 1 | 1.0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Asp | 8.0 | 8 | 3.2 | 3 | 4.0 | 4 | 5.5 | 5 | 3.2 | 3 | 2.0 | 1.1 |
| Thr | 5.6 | 6 | 3.7 | 4 | 1.2 | 1 | 5.0 | 5 | 1.1 | 1 | 1.0 | 0 |
| Ser | 5.6 | 6 | 3.9 | 4 | 2.8 | 3 | 3.0 | 3 | 2.7 | 3 | 2.6 | 0 |
| Glu & Hsr | 4.3 | 4 | 1.4 | 1 | 1.3 | 1 | 3.5 | 3 | 1.2 | 1 | 1.1 | 0 |
| Pro | 3.5 | 4 | 1.1 | 1 | 0 | 0 | 2.0 | 2 | 1.8 | 2 | 1.0 | 1.1 |
| Gly | 3.8 | 3 | 1.2 | 1 | 4.9 | 5 | 2.8 | 3 | 0.1 | 0 | 0.1 | 0 |
| Ala | 8.0 | 8 | 4.2 | 4 | 4.0 | 4 | 6.3 | 6 | 2.2 | 2 | 2.0 | 0 |
| Cys 1/2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Val | 4.1 | 4 | 3.5 | 4 | 2.5 | 3 | 2.1 | 2 | 1.9 | 2 | 1.0 | 0.9 |
| Met | 1.1 | 1 | 1.8 | 2 | 1.1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Ile | 1.6 | 2 | 0.7 | 1 | 0 | 0 | 1.7 | 2 | 0 | 0 | 0 | 0 |
| Leu | 8.4 | 8 | 4.3 | 4 | 2.1 | 2 | 3.5 | 3 | 4.9 | 5 | 4.2 | 1.0 |
| Tyr | 2.0 | 2 | 0.9 | 1 | 2.1 | 2 | 2.0 | 2 | 0 | 0 | 0 | 0 |
| Phe | 6.3 | 6 | 2.2 | 2 | 0.1 | 0 | 4.1 | 4 | 2.0 | 2 | 0.1 | 1.8 |
| Try | 0 | 0 | 0 | 0 | present | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

III. Isolation of Tryptic Peptides

The products of tryptic digestion of the CAT and CB fragments described in the previous section were separated on Sephadex G-25 fine and Aminex A-5. The resulting peptides were usually purified further by cation exchange chromatography; experimental details accompany the descriptions of each fragment. In nearly all cases, the level of contaminating amino acid residues was 0.2 residue or less; all residues determined to be present in quantities of 0.05 residue or more are reported in the tables. Some peptides were found in digests of more than one fragment since the fragments were overlapping in the sequence of the chain. The peptides were not always purified a second time to the extent of 0.2 residue background if purity was sufficient for identification.

*Nomenclature*--The tryptic peptides are named according to homology with the tryptic peptides of human alpha chain (78). Since the correspondence is not identity, additional symbolism is necessitated in order to account for trypsin - sensitive bonds present in the opossum chain and not in the human chain, and vice versa. The proposed structure for the opossum chain does, however, contain 141 residues in homology with the 141 residues of the human chain, so that identical labels for tryptic peptides from each chain imply exactly corresponding portions of sequence. The homology

is quite clear for any peptide taken by its composition alone; the association of peptides in the larger fragments and subsequent sequence determination seem to have justified this scheme of nomenclature in all cases.

Each name of a tryptic peptide begins with a capital "T" (e.g., T3). If there be neither a lysyl nor an argininyl residue corresponding to such a residue in the human chain, a hyphen is used to denote the peptide which bridges the corresponding human peptides (e.g., T7-8). Peptides containing an uncleaved lysyl or argininyl bond are named by joining its hypothetical tryptic pieces with a comma (e.g., T1,2). Peptides produced by tryptic cleavage at a lysyl, argininyl, or (aminoethyl) cysteinyl residue not corresponding to a human lysyl or argininyl residue are indicated by the corresponding human tryptic peptide number and alphabetically with lower case letters (e.g., T12a, T12b, T12c). Names of peptides arising from cleavage at sites in addition to specific trypsin sites incorporate capital letters M (cleavage at methionine, usually due to cyanogen bromide treatment), D (cleavage at aspartic acid with acetic acid), or G (gratuitous cleavage during tryptic digestion); examples are T3G1, T9-10M2, and T11D2.

By this convention the fragments CcH I and CcH II (Figures 6A and 6B) are relabelled, respectively, as T9-10M2,11D1 and T11D2,12a.

*Tryptic peptides from CAT I*--CAT I (70 mg) was treated with one mg TPCK-trypsin at 25° for one hour. The digest was adjusted to pH 3 with acetic acid, centrifuged, and reduced in volume by rotary evaporation. The peptides were separated as described in the legend to Figure 7. Rechromatography of the zones was done on Dowex 50WX4. Amino acid analyses of the resulting peptides appear in Table 5. Another 100 mg of similarly digested CAT I was subjected to gel filtration on Sephadex G-25 fine (Figure 8). The amino acid analysis of the leading zone is shown in Table 5.

*Tryptic peptides from CAT II*--CAT II (50 mg) was digested with 2 mg TPCK-trypsin at 37° for 16 hours and prepared for chromatography as described for CAT I. The separation of the peptides on Aminex A-5 is shown in Figure 9. Analyses of the peptides appear in Table 6; some peptides were rechromatographed as noted in the legend to Table 6. Fifty mg of glycinamidated CAT II were similarly digested and subjected to chromatography on Aminex A-5. The analyses of T12b and T12c after rechromatography appear in Table 6. No additional glycine residues were seen for T12b; three extra glycine residues appeared in the analyses of T12c.

*Tryptic peptides from CAT III*--CAT III (60 mg) was digested with 1 mg TPCK-trypsin at 25° for one hour and the digest chromatographed on Aminex A-5 (Figure 10). The peptides were analyzed without further purification (Table 7).

*Tryptic peptides from CB I*--CB I (50 mg) was digested with 1 mg trypsin for one hour at 25°. The peptides T2, T4M1, and T11 could be identified without further purification among the peptides isolated on Aminex A-5. The analyses are given in Table 8.

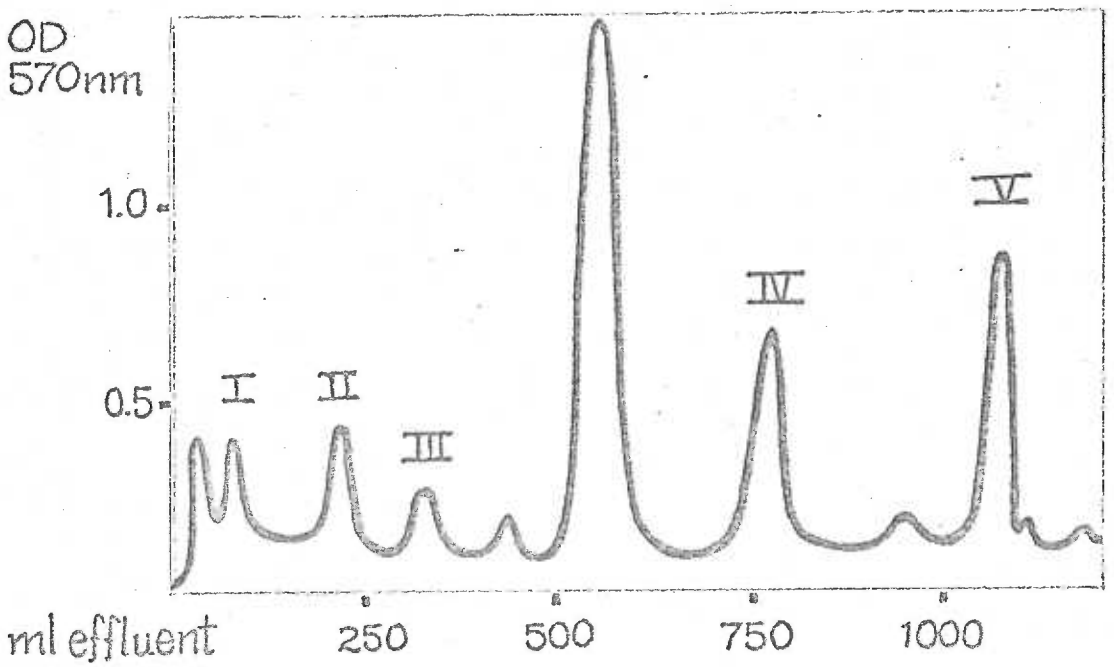
*Tryptic peptides from CB II*--CB II (140 mg) was treated with 2 mg trypsin for two hours at 25°. The digest was acidified, centrifuged, and chromatographed on Sephadex G25 fine (Figure 11). Analyses of the resulting peptides (after rechromatography in some cases as noted with the analyses) appear in Table 9.

*Tryptic peptides from CB III*--CB III (60 mg) was treated with 1.2 mg trypsin at 25° for one hour. Chromatography on Aminex A-5 under conditions given with Figure 7 was done. Analyses of the resulting peptides either directly or following rechromatography are reported in Table 10.

*Tryptic peptides from CB IVA and CB IVB*--These fragments were digested without previous separation from each other. Trypsin (2 mg) was added to 60 mg fragments and digestion carried out for one hour at 25°. Chromatography on Aminex A-5 is depicted in Figure 12; analyses of the peptides are in Table 11.

Figure 7. Separation of tryptic peptides from 70 mg CAT I on a 0.9 x 39 cm column of Aminex A-5 at 50°. Development was with 0.2 M pyridine-acetic acid pH 3.1 for 600 ml followed by a 1000 ml linear gradient to 2M pyridine-acetic acid pH 5.0. The flow rate was 30 ml per hour of which approximately 3 ml per hour were taken for continuous monitoring by reaction with ninhydrin.

Figure 8. Gel filtration of tryptic peptides from 100 mg of CAT I on Sephadex G25 fine (two tandem 2.5 x 200 cm columns) developed with 1.2% acetic acid. Fractions of 10 ml were collected and optical density read at 230 nm. The flow rate was approximately 50 ml per hour.


OD
570nm
1.0
0.5
I
II
III
IV
V
ml effluent
250
500
750
1000

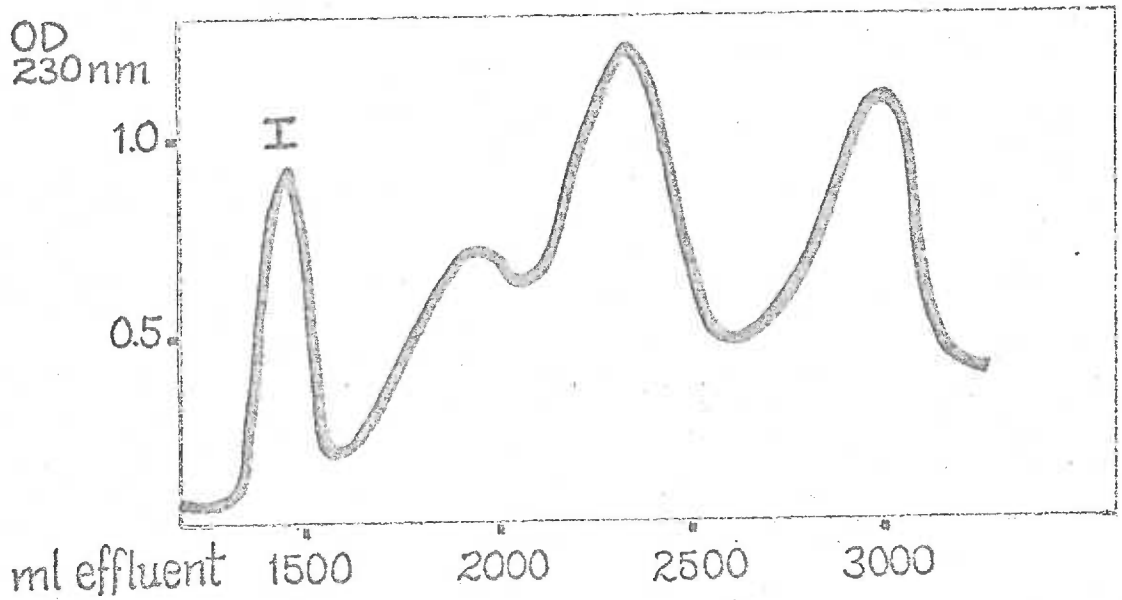
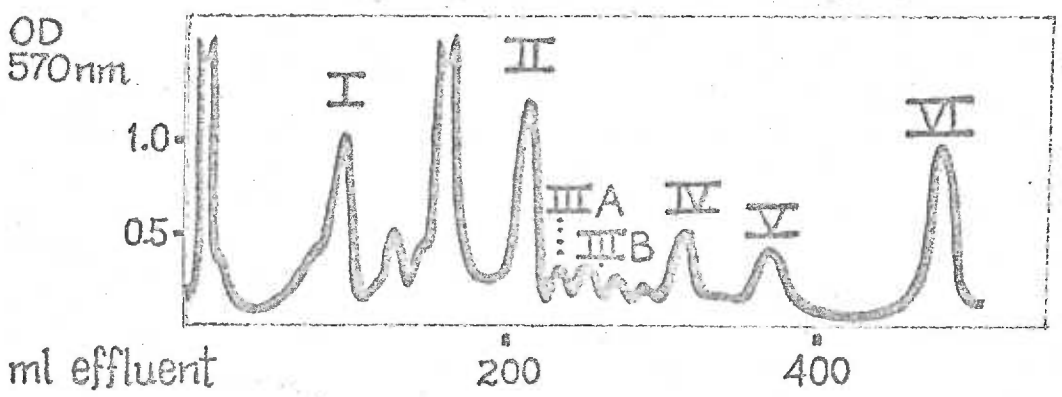

OD
230nm
1.0
0.5
I
ml effluent
1500
2000
2500
3000

Table 5. Amino acid analyses of HCl digests of tryptic peptides from CAT I. The zone numbers refer to Figure 7; each of these was rechromatographed on Dowex 50WX4 before analysis; zone II yielded two significant peptides. G-25:I refers to zone I of Figure 8. The designation of each peptide was determined after analysis as described in the text. * Analyses considered sufficient for identification but not definitive analyses.

| | Zone I | Zone II | Zone II | Zone III | Zone IV | Zone V | G-25:I |
|---|---|---|---|---|---|---|---|
| Lys | 1.0 | 1.0 | 1.3 | 1.1 | 1.0 | | 1.0 |
| His | | 0.1 | 2.7 | | | | 2.0 |
| AE Cys | | | | | | 1.0 | |
| Asp | 2.0 | 0.4 | 4.2 | 2.1 | | 0.1 | 4.3 |
| Thr | 1.1 | 2.6 | 1.1 | | 1.0 | | 1.1 |
| Ser | 1.9 | 1.2 | 2.5 | | | 0.1 | 3.0 |
| Glu | 1.1 | 0.2 | 1.7 | | 2.0 | | 1.1 |
| Pro | 1.0 | 1.0 | 1.2 | 1.0 | | | 0.9 |
| Gly | 1.0 | 0.2 | 1.7 | | 1.0 | | 1.3 |
| Ala | 2.0 | 0.4 | 5.8 | | | | 5.8 |
| Val | | 0.1 | 2.0 | 1.9 | | | 2.0 |
| Met | | | 0.4 | | | | 0.6 |
| Ile | 1.0 | | 0.9 | | | | 0.9 |
| Leu | | 1.5 | 5.8 | | | 1.0 | 5.7 |
| Tyr | 1.9 | | 0.1 | | | | |
| Phe | 2.0 | 1.7 | 0.2 | 1.0 | | 1.0 | |
| Designation | T6 | T5* | T9-10* | T11 | T7-8 | T12a | T9-10 |

Figure 9. Separation of tryptic peptides from 50 mg CAT II on Aminex A-5 (0.9 x 18 cm) at 50°. Initial 50 ml of 0.2 M pyridine-acetic acid pH 3.1 was followed by a 500 ml linear gradient between that buffer and 2 M pyridine-acetic acid pH 5.0 at a flow rate of 30 ml per hour. Detection with ninhydrin was as in the description of Figure 7.

Figure 10. Separation of tryptic peptides from 60 mg of CAT III on Aminex A-5 (0.9 x 36 cm) at 50°. Developed with 0.2 M pyridine-acetic acid pH 3.1 for 400 ml followed by 1000 ml gradient as in Figure 7 at 45 ml per hour. Detection as in Figure 7.


OD
570nm
1.0
0.5
I
II
III A
III B
IV
V
VI
ml effluent
200
400

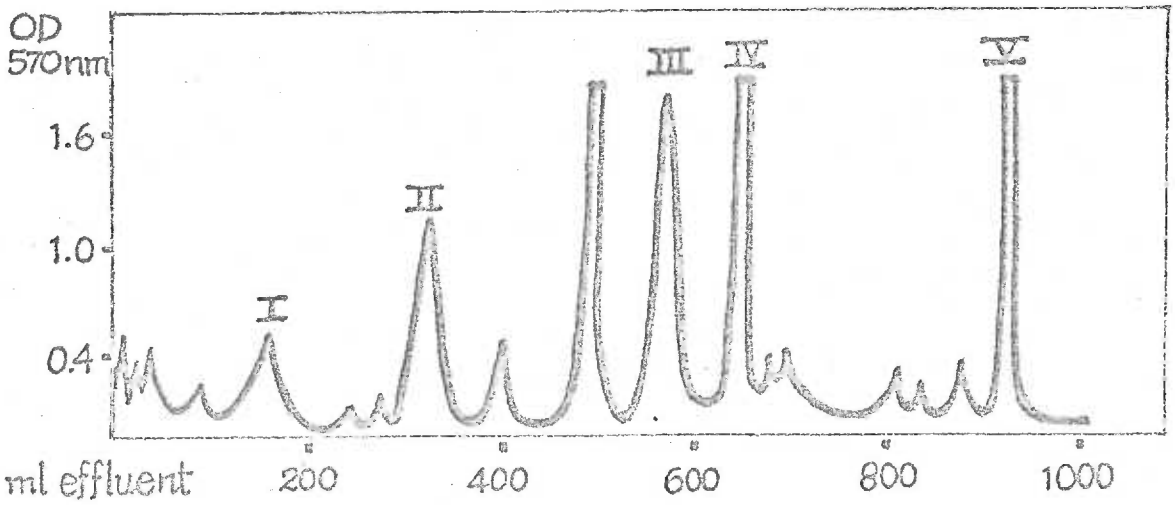

OD
570nm
1.6
1.0
0.4
I
II
III
IV
V
ml effluent
200
400
600
800
1000

Table 6. Amino acid analyses of HCl digests of tryptic peptides from CAT II. The zone numbers refer to Figure 9; zone III is the product of rechromatography on Dowex 50WX4 of combined zones III A and III B; zone I was rechromatographed successively on Dowex 50WX4 and Dowex 50WX2. GA I and GA II refer to peptides recovered from a tryptic digest of glycinamidated CAT II by chromatography on Aminex A5 followed by rechromatography on Dowex 50WX4. The designation of each peptide was determined after analysis as described in the text. Peptide identification symbol followed by /GA indicates that the peptide is derived from a glycinamidated fragment. * Analyses considered sufficient for identification but not definitive analyses. ** 36 hour HCl hydrolysis.

| | Zone I** | Zone II | Zone III | Zone IV | Zone V | Zone VI | GA I | GA II |
|---|---|---|---|---|---|---|---|---|
| Lys | 1.1 | 1.0 | 1.0 | 1.0 | 1.0 | | 1.1 | 1.1 |
| His | 1.0 | 0.1 | 2.0 | 1.0 | 1.0 | | 0.8 | 2.0 |
| Arg | | | | | | 1.0 | | |
| Asp | 2.1 | 0.3 | 1.1 | 0.2 | 0.1 | | 2.2 | 1.1 |
| Thr | 1.0 | 1.0 | 1.0 | 0.1 | 0.1 | | 1.1 | 1.1 |
| Ser | 0.9 | 1.0 | 0.1 | 0.3 | 0.2 | 0.1 | 1.1 | |
| Glu | 1.1 | 0.1 | | | | | 1.0 | |
| Pro | 1.0 | | | | | | 1.0 | |
| Gly | 0.1 | 0.1 | 1.0 | 1.0 | 1.0 | | 3.0 | 1.0 |
| Ala | 1.0 | 0.2 | 1.9 | 1.9 | 1.0 | | 1.1 | 1.9 |
| Val | 0.1 | 0.2 | 2.0 | | | | 0.2 | 1.9 |
| Met | 0.9 | 0.1 | 1.0 | | | | 0.7 | 1.0 |
| Ile | 1.0 | | | | | | 0.9 | |
| Leu | 0.1 | 0.2 | 2.1 | 1.0 | 1.0 | | 0.2 | 1.9 |
| Tyr | | | | | | 1.0 | | |
| Phe | 1.0 | | | | | | 1.0 | |
| Designation | T12c | T13G2* | T12b | T12bM2* | T12bG2 | T14 | T12c/GA | T12b/GA |

Table 7. Amino Acid Analyses of Tryptic Peptides from CAT III.*

| | Zone I | Zone II | Zone III | Zone IV | Zone V |
|---|---|---|---|---|---|
| Lys | 0.1 | 1.0 | 2.2 | 1.0 | 1.0 |
| Arg | 1.0 | | | | |
| Asp | 1.1 | 2.1 | 3.0 | 1.0 | |
| Thr | 0.1 | 0.1 | 1.0 | 1.0 | |
| Ser | 0.9 | 1.0 | 1.0 | | 1.0 |
| Glu | 1.0 | 0.1 | 0.2 | | |
| Gly | 3.9 | 0.2 | 0.3 | 0.1 | 1.0 |
| Ala | 2.6 | 1.1 | 1.1 | 0.1 | 1.0 |
| Val | 0.9 | 0.9 | 1.8 | 0.9 | |
| Met | 0.8 | | | | |
| Leu | 1.0 | 1.0 | 1.0 | | |
| Tyr | 1.8 | | 0.1 | | |
| Phe | | | | | |
| Try | | | | | 0.7 |
| Designation** | T4 | T1 | T1,2 | T2 | T3 |

* HCl hydrolysis of the zones in Figure 10.

** The designation of each peptide was determined after analysis as described in the text.

Table 8. Amino Acid Analyses of Some Tryptic Peptides from CB I.*

| | Peptide 1 | Peptide 2 | Peptide 3 |
|---|---|---|---|
| Lys | 1.1 | 1.1 | |
| Hsl** | | | 0.3 |
| Asp | 2.0 | 2.2 | 1.3 |
| Thr | | | 0.1 |
| Ser | 0.1 | 1.0 | 1.0 |
| Glu & Hsr** | | | 0.6 |
| Pro | 1.0 | | 0.2 |
| Gly | | | 2.7 |
| Ala | | 1.0 | 1.0 |
| Val | 1.9 | 0.8 | 1.1 |
| Leu | | 0.9 | 0.2 |
| Tyr | | | 0.9 |
| Phe | 1.1 | | 0.1 |
| Designation*** | T11 | T1 | T4M1 |

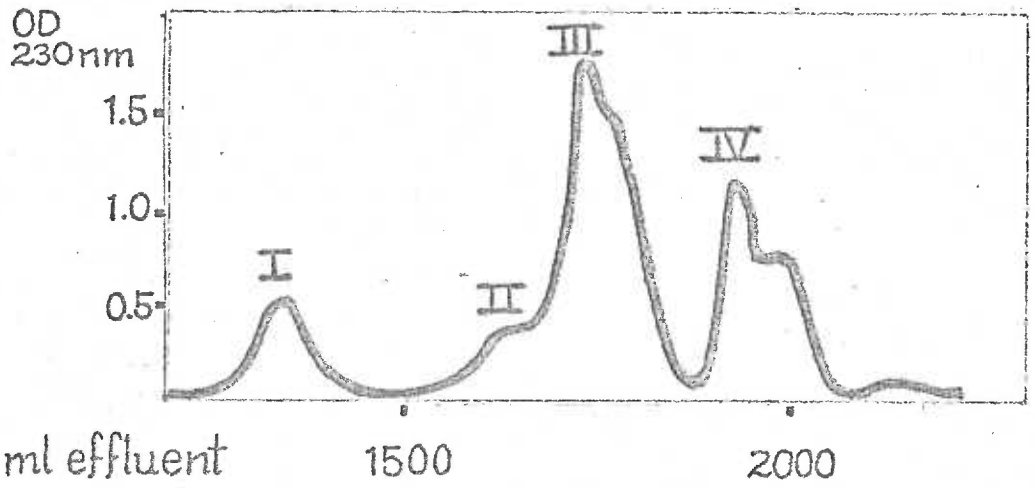
* Analyses after HCl hydrolysis.

** Abbreviations as in Table 3.

*** Designation determined after analysis as described in the text.

Figure 11. Separation of tryptic peptides from 140 mg CB II on Sephadex G25 fine. Two 2.7 x 200 cm columns in tandem were developed with 1.2% acetic acid at a flow rate of approximately 60 ml per hour. Fractions of 10 ml were collected and the optical density determined at 230 nm.

Figure 12. Separation of tryptic peptides from 60 mg CB IVA and CB IVB. Chromatography was done on a 0.9 x 35 cm column of Aminex A-5 at 50°. 50 ml of 0.2 M pyridine-acetic acid pH 3.1 was followed by a linear 1000 ml gradient from that buffer to 2 M pyridine-acetic acid pH 5.0 and finally 150 ml of the final gradient buffer. Flow rate was 40 ml per hour. Detection was as described for Figure 7.


OD
230nm
1.5
1.0
0.5
I
II
III
IV
ml effluent
1500
2000

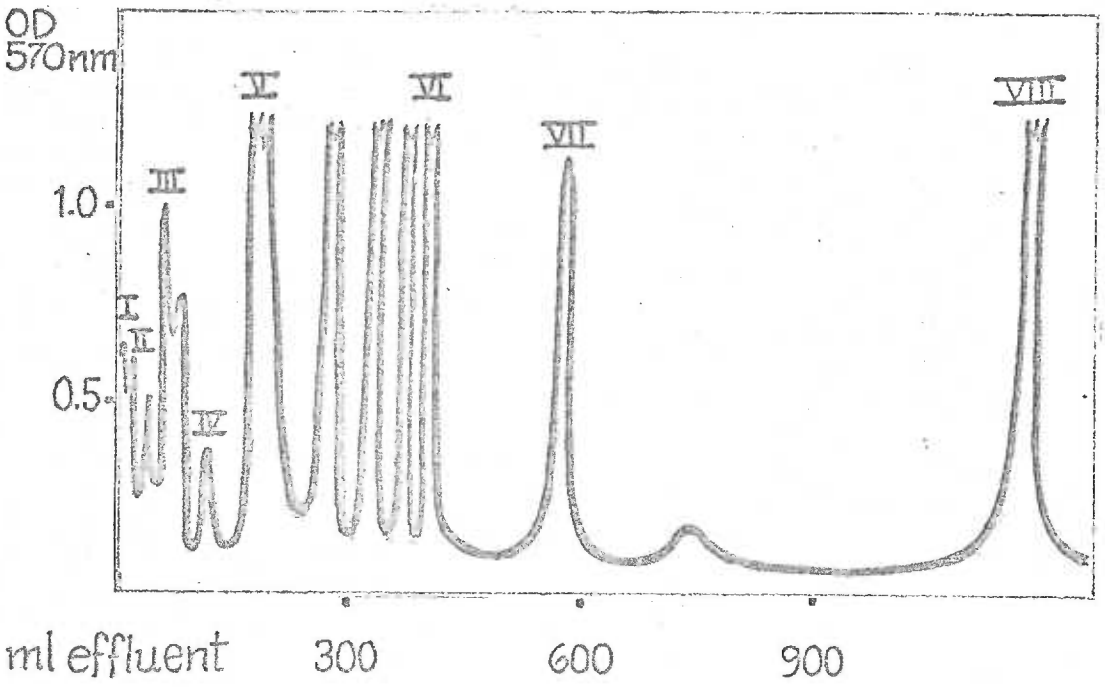

OD
570nm
1.0
0.5
I
II
III
IV
V
VI
VII
VIII
ml effluent
300
600
900

Table 9. Amino acid analyses of HCl digests of tryptic peptides from CB II. The zone numbers refer to Figure 11; zone IV was analyzed directly; the peptides from zone III and zone I were isolated by rechromatography on Aminex A-5; zone II was successively rechromatographed on Aminex A-5 and Dowex 50WX2. The designation of each peptide was determined after analysis as described in the text.

* Abbreviations as in Table 3.

| | Zone I | Zone II | Zone III | Zone III | Zone III | Zone IV |
|---|---|---|---|---|---|---|
| Lys | | 1.0 | 1.0 | 0.9 | 1.1 | |
| His | 0.9 | | | | | |
| Hsl* | 0.3 | | | | | |
| Arg | | | | | | 0.9 |
| Asp | 2.8 | | 2.0 | | | |
| Thr | 0.1 | 2.9 | 0.9 | 1.0 | 2.0 | |
| Ser | 0.1 | 1.1 | 2.0 | | 1.1 | |
| Glu & Hsl* | 0.4 | 0.1 | 1.1 | 2.0 | | 1.0 |
| Pro | 0.1 | 1.1 | 1.0 | | 0.9 | |
| Gly | 1.0 | | 1.0 | 1.0 | | 1.1 |
| Ala | 3.9 | | 2.1 | | | 1.0 |
| Val | 2.2 | | | | | |
| Ile | 1.2 | | 0.7 | | | |
| Leu | 2.0 | 1.0 | | | | 1.0 |
| Tyr | | | 2.1 | | | 1.0 |
| Phe | | 1.9 | 2.2 | | 0.9 | |
| Designation | T9-10M1 | T5 | T6 | T7-8 | T5G2 | T4M2 |

Table 10. Amino acid analyses of HCl digests of tryptic peptides from CB III. Peptides were separated on Aminex A-5 and analyzed directly except peptide 1 which was rechromatographed on Dowex 50WX2. The peptide numbers are only for labelling purposes. The designation of each peptide was determined after analysis as described in the text. * Abbreviations as in Table 3.
** Analyses considered sufficient for identification but not definitive analyses.

| | Peptide 1 | Peptide 2 | Peptide 3 | Peptide 4 | Peptide 5 | Peptide 6 |
|---|---|---|---|---|---|---|
| Lys | | | 1.0 | 1.0 | 1.0 | 1.0 |
| His | 0.1 | | | | | |
| Hsl* | 0.4 | | | | | |
| Asp | 1.0 | 0.1 | 2.0 | 0.1 | 1.0 | |
| Thr | 0.1 | | | | 1.0 | |
| Ser | 1.0 | 0.1 | 1.0 | 1.0 | | 1.0 |
| Glu & Hsr | 0.3 | | | | | |
| Pro | 0.1 | | | | | |
| Gly | 2.9 | 1.0 | | 1.0 | | |
| Ala | 1.2 | 1.0 | 1.0 | 1.0 | | |
| Val | 1.0 | | 0.6 | | 0.9 | |
| Leu | | 0.1 | 0.6 | | | |
| Tyr | 1.0 | | | | | |
| Try | | 0.3 | | 0.3 | | |
| Designation | T4M1 | T3G1 | T1** | T3 | T2 | T3G2 |

Table 11. Amino acid analyses of HCl digests of tryptic peptides from CB IVA and CB IVB. The zone numbers refer to Figure 12; zones II and IV were rechromatographed on Dowex 50WX2 before analysis. The designation of each peptide was determined after analysis as described in the text. The abbreviations are as in Table 3.

* Analyses considered sufficient for identification but not definitive analyses.

| | Zone I | Zone II | Zone III | Zone IV | Zone V | Zone VI | Zone VII | Zone VIII |
|---|---|---|---|---|---|---|---|---|
| Lys | | | 0.9 | 1.0 | 2.0 | 1.0 | 0.9 | |
| His | | 0.8 | | | | | 1.1 | |
| Hsl | | 0.2 | | | | | | |
| Arg | | | | | | | | 1.0 |
| Asp | 0.4 | 1.1 | 0.2 | 1.1 | 1.1 | | | |
| Thr | 1.1 | 1.0 | 2.1 | 1.1 | 2.0 | 1.0 | | |
| Ser | 1.9 | 0.9 | 3.0 | 2.2 | 3.0 | 1.0 | | |
| Glu + Hsr | 0.2 | 1.3 | 0.2 | | | | | |
| Pro | 0.1 | 1.2 | | | | | | |
| Gly | 0.3 | | 0.2 | | 0.1 | | 1.0 | |
| Ala | 1.0 | 0.9 | 1.0 | 0.8 | 1.0 | | 1.9 | |
| Val | 1.8 | | 1.8 | 1.8 | 1.8 | | | |
| Ile | | 0.7 | | | | | | |
| Leu | 2.2 | | 2.2 | 2.1 | 2.0 | | 1.0 | |
| Tyr | | | | | | | | 1.0 |
| Phe | 1.1 | 0.9 | 1.0 | 0.9 | 1.0 | | | |
| Designation | T13G1* | T12cM1 | T13 | T12cM2,13G1 | T12cM2,13 | T13G2 | T12bM2 | T14 |

IV. Sequence Determination of Peptides

Sequence determination of the peptides which were purified as described in the previous section was done by: (1) sequential (Edman) degradation of the peptides; (2) isolation of pieces of peptides after further enzymic or chemical cleavage and Edman degradation of the pieces; (3) assuming that lysine, arginine, or aminoethyl-cysteine is the COOH-terminus of tryptic peptides containing one such residue; (4) assuming that homoserine (when present as one residue) is the COOH-terminus of peptides derived from cyanogen bromide cleavage.

Additional abbreviations to be used in this section are: APM for aminopeptidase M; Ed. I, Ed. II, etc., for successive cycles of Edman degradations; A-5, X-4, and X-2 for the cation exchange resins Aminex A-5, Dowex 50WX4 and Dowex 50WX2; Asx for either Asp or Asn; Glx for either Glu or Gln.

Amino acid analyses of fragments and of residual peptides after Edman degradation cycles are collected into Table 12 at the end of this section. The numbers in brackets refer to analyses in that table.

Sequence of T1

APM: one Asp; one Asn.

Ed. I - Ed. V [Anal. 1-5]: Val-Leu-Ser-Ala-Asx.

Elastase (10 units enzyme, 37°, 12 hours, separation on X-4, rechromatography on X-2).

Ela. 1 [Anal. 6]: $Asx_2$,Lys.

Ela. 1 Ed. I [Anal. 7]: Asx.

Ela. 1 Ed. I APM: residual Asp.

Sequence: Val-Leu-Ser-Ala-Asn-Asp-Lys.

Sequence of T2

APM: Asn.

Ed. I - Ed. II [Anal. 8-9]: Thr-Asx.

Sequence: Thr-Asn-Val-Lys.

Sequence of T3

Ed. I - Ed. II [Anal. 10-11]: Gly-Ala.

Nonspecific tryptic peptides (Table 10) T3G1 and T3G2: Gly,Ala,Try and Ser,Lys.

Sequence: Gly-Ala-Try-Ser-Lys.

## Partial Sequence of T4M1

APM: Asn.

Ed. I - Ed. II [Anal. 12, 13] : Val-Gly.

Subtilisin BPN' (0.5 mg enzyme, 37°, 4 hours, separation on X-2).

Sub. 1 [Anal. 14] : Gly,Ala,Tyr.

Sub. 1 Ed. I - Ed. II [Anal. 15-16] : Gly-Ala

Sequence: Val-Gly-(Gly,Asn,Ser,Gly-Ala-Tyr)-Met

## Partial Sequence of T4M2

Ed. I [Anal. 17] : Gly.

Thermolysin (0.5 mg enzyme, 37°, 6 hours, separation on A-5)

Thm. 1 [Anal. 18]: Tyr,Arg.

Thm. 2 [Anal. 19]: Leu,Tyr,Arg.

Thm. 3 [Anal. 20]: Ala,Leu.

Sequence: Gly-Glx-Ala-Leu-Tyr-Arg.

## Sequence of T4

APM: Asn, Glu.

Chymotrypsin (0.5 mg enzyme, 25°, 16 hours, separation on X-4)

Chy. 1 [Anal. 21]: Val,$Gly_3$,Asx,Ser,Ala,Tyr.

Elastase of Chy. 1 (5 units enzyme, 37°, 16 hours, separation on X-4)

Chy. 1 Ela. 1 [Anal. 22]: Val,$Gly_3$,Asx,Ser,Ala.

Three Edman cycles were done on Chy. 1 Ela. 1; residual peptide was purified on X-4 and one more Edman done to give Chy. 1 Ela. 1

Ed. III-Ed. IV [Anal. 23-24]: (Val,$Gly_2$)-Asx.

The sequence is deduced from results from T4M1, T4M2, and T4:

Val-Gly-Gly-Asn-Ser-Gly-Ala-Tyr-Met-Gly-Glu-Ala-Leu-Tyr-Arg

## Sequence of T5

Subtilisin BPN' (one mg enzyme, 37°, 5 hours, separation on X2).

Sub. 1 [Anal. 25]: Thr,Phe,Leu.

Sub. 1 Ed. I - Ed. II [Anal. 26, 27]: Thr-Phe.

Nonspecific tryptic peptide (Table 9) T5G2: Ser,Phe,Pro,$Thr_2$,Lys.

T5G2 Ed. I - Ed. III [Anal. 28-30]: Ser-Phe-Pro.

Sequence: Thr-Phe-Leu-Ser-Phe-Pro-Thr-Thr-Lys.

## Partial Sequence of T6

Acetic acid treatment (0.5 M acetic acid, 100°, 16 hours, separation of A-5).

Hac 1 [Anal. 31]: Thr,$Tyr_2$,Phe,Pro,Asx.

Thermolysin (0.5 mg enzyme, 37°, 3 hours, separation on A-5).

Thm. 1 [Anal. 32]: Thr,$Tyr_2$,Phe,Pro,$Asx_2$.

Thm. 2 [Anal. 33]: Phe,$Ser_2$,$Ala_2$,Gly,Glx.

Thm. 2 APM: Gln.

Thm. 3 [Anal. 34]: Ile,Lys.

Elastase (10 units enzyme, 37°, 12 hours, separation on X-4).

Ela. 1 [Anal. 35]: Glx,Ile,Lys.

Ela. 2 [Anal. 36]: Ala,Gly.

Ela. 2 Ed. I [Anal. 37]: Ala.

Ela. 3 [Anal. 38]: Tyr,Asx.

Ela. 3 Ed. I [Anal. 39]: Tyr.

Partial Sequence: [(Thr,Tyr,Phe,Pro,Asx,Tyr),Asx]-(Phe,Ser,Ala-Gly, Ser,Ala)-Gln-Ile-Lys.

It is also shown that a Tyr-Asp partial sequence is present since Ela. 3 Ed. I left a residual Asp.

Sequence of T7-8

APM: two Gln, no Glu.

Ed. I - Ed. II [Anal. 40-41]: Thr-Glx.

Ed. II APM confirms residual Gln.

Subtilisin BPN' (0.5 mg enzyme, 37°, 6 hours, separation on X-2, rechromatography on A-5).

Sub. 1 [Anal. 42]: Thr,$Glx_2$,Gly.

Sub. 2 [Anal. 43]: Gly,Glx,Lys.

Sub. 2 Ed. I [Anal. 44]: Gly.

Sequence: Thr-Gln-Gly-Gln-Lys.

Sequence of T9-10M1

Thermolysin (0.3 mg enzyme, 37°, 5 hours, separation on X-2).

Thm. 1 [Anal. 45]: Ile,Ala,Asx.

Thm. 1 APM: Asp.

Thm. 2 [Anal. 46]: Ala,Val,Gly.

Thm. 2 Ed. I - Ed. II [Anal. 47-48]: Ala-Val.

Thm. 3 [Anal. 49]: Val,Ala,His.

Thm. 3 Ed. I - Ed. II [Anal. 50-51]: Val-Ala.

Thm. 4 [Anal. 52]: Leu,Ala.

Thm. 4 Ed. I [Anal. 53]: Leu.

Thm. 5 [Anal. 54]: Leu,$Asx_2$,Met.

Thm. 5 APM: Asp, no Asn.

Acetic acid digest (0.5 M acetic acid, 110°, 18 hours, separation on X-2).

Hac 1 [Anal. 55]: $Ala_3$,$Val_2$,Gly,$Leu_2$,His.

Hac 1 Ed. I [Anal. 56]: Ala.

Chymotrypsin (0.5 mg enzyme, 25°, 16 hours, separation on X-2).

Chy. 1 [Anal. 57]: $Ala_2$,Val,His,Leu.

Chy. 2 [Anal. 58]: Ile,$Ala_2$,Asx,Val,Gly,Leu.

Chy. 2 Ed. I - Ed. II [Anal. 59-60]: Ile-Ala.

Sequence: Ile-Ala-Asp-Ala-Val-Gly-Leu-Ala-Val-Ala-His-Leu-Asp-Asp-Met

## Partial Sequence of T9-10M2,11D1

Ed. I - Ed. III [Anal. 61-63]: Pro-Thr-Ala.

## Sequence of T9-10M2

Chymotrypsin (1.5 mg chymotrypsin, 25°, 16 hours, separation on A-5, rechromatography on X-4).

Chy. 1 [Anal. 64]: Glx,Leu,Lys.

Chy. 1 Ed. I [Anal. 65]: Glx.

Chy. 2 [Anal. 66]: Ala,His,Glx,Leu,Lys.

Chy. 2 Ed. I [Anal. 67]: Ala.

Chy. 3 [Anal. 68]: $His_2$,Ala,Glx,Leu,Lys.

Chy. 4 [Anal. 69]: Ser,Asx,$Leu_2$,$His_2$,Ala,Glx,Lys.

Chy. 4 Ed. I - Ed. II [Anal. 70-71]: Ser-Asx.

Chy. 4 Ed. II APM: residual Asp,Glu.

Chy. 5 [Anal. 72]: $Ser_3$,$Leu_3$,Asx,Ala,Glx,$His_2$,Lys.

Chy. 5 Ed. I - Ed. II [Anal. 73-74]: Ser-Ser.

Sequence (using results from T9-10M2,11D1):

Pro-Thr-Ala-Leu-Ser-Ser-Leu-Ser-Asp-Leu-His-Ala-His-Glu-Leu-Lys.

Sequence of T11D2,12a

Ed. I-II [Anal. 75-76]: Pro-Val.

Ed. II APM: residual Asn.

Thermolysin (0.3 mg enzyme, 37°, 12 hours, separation on A-5).

Thm. 1 [Anal. 77]: Leu,Cys.

Thm. 2 [Anal. 78]: $Phe_2$,Lys.

Composition of 12a (Table 5): Phe,Leu,Cys.

Sequence: Pro-Val-Asn-Phe-Lys-Phe-Leu-Cys.

## Sequence of T11

APM: One Asp, one Asn.

Ed. I [Anal. 79]: Val.

The sequence of T11D2,12a leads to the sequence of T11:

Val-Asp-Pro-Val-Asn-Phe-Lys.

## Partial Sequence of T12bM2

Ed. I-Ed. II [Anal. 80-81]: Ala-Ala.

Partial Sequence: Ala-Ala-(His,Leu,Gly)-Lys.

## Sequence of T12b

APM: confirms Asn.

Ed. I [Anal. 82]: His.

Three Edman cycles were done and the residual peptide purified on X-4 to give Ed. III [Anal. 83]: (His,Asx,Val)

Thermolysin (one mg, 37°, 12 hours, separation on X-4).

Thm. 1 [Anal. 84]: Val,Thr.

Thm. 1 Ed. I [Anal. 85]: Val.

Thm. 2 [Anal. 86]: His,Asx.

Thm. 3 [Anal. 87]: Leu,Gly,Lys.

Thm. 3 Ed. I [Anal. 88]: Leu.

Chymotrypsin (0.5 mg enzyme, 37°, 13 hours, separation on X-4).

Chy. 1 [Anal. 103]: Val,Thr,Met.

Results of partial sequence of T12bM2 and composition of T12bM2 and composition of T12bM2 (Table 6) lead to the sequence of T12b:

His-Asn-Val-Leu-Val-Thr-Met-Ala-Ala-His-Leu-Gly-Lys.

## Partial Sequence of T12c

Thermolysin (0.3 mg enzyme, 37°, 15 hours, separation on X-4).

Glycinamidated peptide.

Thm. 1 [Anal. 89]: Ile,His,Ala,Ser,Met,Asp,Lys.

Thm. 1 Ed. I [Anal. 90]: Ile.

Chymotrypsin (0.5 mg enzyme, 37°, 15 hours, separation on X-4)

Chy. 1 [Anal. 91]: Asp,Phe.

Chy. 2 [Anal. 92]: Ala,Ser,Met.

Chy. 2 Ed. I [Anal. 93]: Ala.

Chy. 3 [Anal. 94]: Thr,Pro,Glu,Ile,His,Ala.

Composition of T12cM1 leads to partial sequence:

(Asp,Phe)-(Thr,Pro,Glu)-Ile-His-Ala-Ser-Met-Asp-Lys.

## Partial Sequence of T13

Thermolysin (0.5 mg, 37°, 3 hours, separation on X-2, rechromatography on X-4).

Thm. 1 [Anal. 95]: Phe,Leu,Ala,Ser.

Thm. 2 [Anal. 96]: Leu,Ala,Ser.

Thm. 2 Ed. I-II [Anal. 97-98]: Leu-Ala.

Thm. 3 [Anal. 99]: Phe,Leu.

Thm. 4 [Anal. 100]: Ser,Lys.

Thm. 5 [Anal. 101]: Val,Leu,Thr,Ser,Lys.

Thm. 5 Ed. I [Anal. 102]: Val.

With composition of T13G2 (Table 11), the partial sequence:

[(Phe-Leu-Ala-Ser),Val,Ser,Thr]-Val-Leu-Thr-Ser-Lys.

Sequence of T14

Sequence: Tyr-Arg.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.9 | 0.9 | 0.9 | 1.0 | 1.0 |
| His | | | | | | | | | | |
| Hsl | | | | | | | | | | |
| AE Cys | | | | | | | | | | |
| Arg | | | | | | | | | | |
| Asp | 2.0 | 2.3 | 2.0 | 2.0 | 1.0 | 2.1 | 1.1 | 1.1 | 0.3 | |
| Thr | | | | | | | | 0.1 | 0.1 | |
| Ser | 1.0 | 0.8 | 0.3 | 0.1 | | | | | | 1.0 |
| Glu & Hsr | | | | | | | | | | |
| Pro | | | | | | | | | | |
| Gly | | | | | | | | | | |
| Ala | 1.0 | 0.8 | 1.1 | 0.5 | 0.3 | | | | | 1.0 |
| Val | 0 | 0 | 0 | 0 | 0 | | | 0.9 | 1.0 | |
| Met | | | | | | | | | | |
| Ile | | | | | | | | | | |
| Leu | 1.0 | 0 | 0 | 0 | 0 | | | | | |
| Tyr | | | | | | | | | | |
| Phe | | | | | | | | | | |
| Try | | | | | | | | | | + |

| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | 1.1 | | | | | | | | | |
| His | | | | | | | | | | |
| Hsl | | n.d. | n.d. | | | | | | | |
| AE Cys | | | | | | | | | | |
| Arg | | | | | | | 0.8 | 1.0 | 1.0 | |
| Asp | | 1.0 | 1.0 | | | | | | | |
| Thr | | | | | | | | | | |
| Ser | 1.0 | 1.0 | 1.1 | | | | | | | |
| Glu & Hsr | | 0.6 | 0.6 | | | | 1.0 | | | |
| Pro | | | | | | | | | | |
| Gly | 0.1 | 3.0 | 2.0 | 1.0 | 0.4 | 0 | 0.3 | | | |
| Ala | 0.2 | 1.0 | 1.0 | 1.0 | 1.1 | 0 | 1.1 | | | 1.0 |
| Val | | 0 | 0 | | | | | | | |
| Met | | | | | | | | | | |
| Ile | | | | | | | | | | |
| Leu | | | | | | | 0.9 | | 1.0 | 1.0 |
| Tyr | | 0.9 | 0.9 | 1.0 | 0.9 | 1.0 | 1.0 | 1.0 | 0.9 | |
| Phe | | | | | | | | | | |
| Try | + | | | | | | | | | |

| | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | | | | | | | | 1.0 | 1.0 | 1.1 |
| His | | | | | | | | | | |
| Hsl | | | | | | | | | | |
| AE Cys | | | | | | | | | | |
| Arg | | | | | | | | | | |
| Asp | 1.1 | 1.0 | 1.0 | 0.3 | | | | | | |
| Thr | | | | | 0.9 | 0.3 | 0 | 1.9 | 2.1 | 2.0 |
| Ser | 0.9 | 0.9 | 0.9 | 1.1 | | | | 0.2 | 0.4 | 0.2 |
| Glu & Hsr | | | | | | | | | | |
| Pro | | | | | | | | 1.0 | 0.9 | 0.5 |
| Gly | 3.1 | 3.0 | 1.0 | 1.0 | | | | | | |
| Ala | 1.1 | 1.1 | 1.0 | 0.9 | | | | | | |
| Val | 1.0 | 1.0 | 0.1 | 0.1 | | | | | | |
| Met | | | | | | | | | | |
| Ile | | | | | | | | | | |
| Leu | | | | | 1.1 | 1.1 | 1.0 | | | |
| Tyr | 0.9 | | | | | | | | | |
| Phe | | | | | 1.0 | 0.9 | 0 | 1.0 | 0.5 | 0.2 |
| Try | | | | | | | | | | |

| | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | | | | 1.0 | 1.2 | | | | | 0.7 |
| His | | | | | | | | | | |
| Hsl | | | | | | | | | | |
| AE Cys | | | | | | | | | | |
| Arg | | | | | | | | | | |
| Asp | 1.3 | 2.2 | | | | | | 1.1 | 1.0 | |
| Thr | 0.7 | 1.0 | | | | | | | | 0.1 |
| Ser | | | 1.9 | | | | | | | |
| Glu & Hsr | | | 0.9 | | 1.0 | | | | | 2.0 |
| Prc | 0.9 | 1.0 | | | | | | | | |
| Gly | | | 1.1 | | | 1.0 | 0 | | | 1.0 |
| Ala | | | 2.2 | | | 1.0 | 1.0 | | | |
| Val | | | | | | | | | | |
| Met | | | | | | | | | | |
| Ile | | | | 1.0 | 0.8 | | | | | |
| Leu | | | | | | | | | | |
| Tyr | 1.9 | 2.0 | | | | | | 1.0 | 0 | |
| Phe | 1.2 | 1.0 | 0.9 | | | | | | | |
| Try | | | | | | | | | | |

| | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | 0.8 | | 1.0 | nd. | | | | | | |
| His | | | | | | | | | 1.0 | 1.0 |
| Hsl | | | | | | | | | | |
| AE Cys | | | | | | | | | | |
| Arg | | | | | | | | | | |
| Asp | | | | | 1.0 | | | | | |
| Thr | 0 | 0.9 | | | | | | | | |
| Ser | | | | | | | | | | |
| Glu & Hsr | 1.2 | 2.0 | 1.0 | 1.0 | | | | | | |
| Pro | | | | | | | | | | |
| Gly | 0.8 | 1.1 | 1.0 | 0.3 | | 1.1 | 1.0 | 1.0 | | |
| Ala | | | | | 1.0 | 0.9 | 0.1 | 0 | 1.0 | 1.0 |
| Val | | | | | | 1.0 | 0.9 | 0 | 0.9 | 0.1 |
| Met | | | | | | | | | | |
| Ile | | | | | 0.9 | | | | | |
| Leu | | | | | | | | | | |
| Tyr | | | | | | | | | | |
| Phe | | | | | | | | | | |
| Try | | | | | | | | | | |

| | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | | | | | | | | | | |
| His | 1.0 | | | | 1.2 | 1.1 | 0.8 | | | |
| Hsl | | | | 0.3 | | | | | | |
| AE Cys | | | | | | | | | | |
| Arg | | | | | | | | | | |
| Asp | | | | 2.0 | | | | 1.0 | 1.0 | 0.9 |
| Thr | | | | | | | | | | |
| Ser | | | | | | | | | | |
| Glu & Hse | | | | 0.3 | | | | | | |
| Pro | | | | | | | | | | |
| Gly | | | | | 0.9 | 1.1 | | 1.0 | 1.0 | 1.0 |
| Ala | 0 | 0.9 | 1.0 | | 3.1 | 2.2 | 2.1 | 2.0 | 2.0 | 1.1 |
| Val | 0 | | | | 1.6 | 1.7 | 0.9 | 0.9 | 1.0 | 1.0 |
| Met | | | | | | | | | | |
| Ile | | | | | | | | 1.0 | 0 | 0 |
| Leu | | 1.1 | | 1.0 | 2.1 | 2.0 | 1.2 | 1.0 | 1.0 | 0.9 |
| Tyr | | | | | | | | | | |
| Phe | | | | | | | | | | |
| Try | | | | | | | | | | |

| | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | nd | nd | nd | 1.0 | 1.0 | 1.0 | 0.9 | 1.0 | 1.2 | nd |
| His | nd | nd | nd | | | 1.0 | 0.8 | 2.0 | 1.8 | nd |
| Hsl | | | | | | | | | | |
| AE Cys | | | | | | | | | | |
| Arg | | | | | | | | | | |
| Asp | 2.2 | 2.4 | 2.1 | | | | | | 1.0 | 0.8 |
| Thr | 1.0 | 0.7 | 0.5 | | | | | | | |
| Ser | 2.6 | 2.7 | 2.5 | | | | | | 0.8 | 0.2 |
| Glu & Hsr | 1.1 | 1.3 | 1.2 | 1.0 | 0.5 | 1.0 | 1.0 | 1.0 | 1.1 | 1.1 |
| Pro | 0.3 | 0.3 | 0.3 | | | | | | | |
| Gly | | | | | | | | | | |
| Ala | 2.2 | 2.2 | 1.8 | | | 1.0 | 0.3 | 0.9 | 0.8 | 1.0 |
| Val | 1.1 | 1.1 | 0.9 | | | | | | | |
| Met | | | | | | | | | | |
| Ile | | | | | | | | | | |
| Leu | 3.9 | 4.2 | 4.5 | 1.0 | 1.0 | 1.0 | 1.3 | 0.9 | 2.0 | 2.0 |
| Tyr | | | | | | | | | | |
| Phe | | | | | | | | | | |
| Try | | | | | | | | | | |

| | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | nd | 1.1 | nd | nd | 0.8 | 0.9 | | 0.9 | 0.8 | 1.0 |
| His | nd | 1.9 | nd | nd | | | | | | 1.0 |
| Hsl | | | | | | | | | | |
| AE Cys | | | | | 0.7 | 0.7 | 0.9 | | | |
| Arg | | | | | | | | | | |
| Asp | 0.3 | 1.1 | 1.2 | 1.0 | 1.0 | 1.1 | | | 2.1 | |
| Thr | | | | | | | | | | |
| Ser | 0.3 | 2.8 | 1.8 | 1.3 | | | | | | |
| Glu & Hsr | 1.0 | 1.1 | 1.2 | 0.9 | | | | | | |
| Pro | | | | | 0.2 | 0.2 | | | 0.8 | |
| Gly | | | | | | | | | | 1.1 |
| Ala | 1.0 | 1.0 | 1.1 | 1.0 | | | | | | 1.0 |
| Val | | | | | 0.9 | 0.1 | | | 1.1 | |
| Met | | | | | | | | | | |
| Ile | | | | | | | | | | |
| Leu | 1.7 | 3.1 | 2.7 | 3.0 | 1.2 | 1.0 | 1.1 | | | 1.1 |
| Tyr | | | | | | | | | | |
| Phe | | | | | 1.9 | 1.9 | | 2.1 | 1.1 | |
| Try | | | | | | | | | | |

| | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | 0.8 | 0.7 | 1.0 | | | | 1.1 | 1.0 | 1.2 | 1.1 |
| His | 1.0 | 1.1 | 1.3 | | | 1.0 | | | 0.8 | 0.6 |
| Hsl | | | | | | | | | | |
| AE Cys | | | | | | | | | | |
| Arg | | | | | | | | | | |
| Asp | | 0.9 | 0.1 | | | 1.0 | | | 1.2 | 1.0 |
| Thr | | 1.1 | 1.0 | 1.0 | 1.0 | | | | | |
| Ser | | | | | | | | | 0.7 | 1.0 |
| Glu & Hsr | | | | | | | | | | |
| Pro | | | | | | | | | | |
| Gly | 1.0 | 1.2 | 1.0 | | | | 1.0 | 1.0 | 1.1 | 1.5 |
| Ala | 0.2 | 2.1 | 2.2 | | | | | | 1.1 | 0.8 |
| Val | | 2.0 | 0.9 | 1.0 | 0 | | | | | |
| Met | | 1.0 | 0.8 | | | | | | 1.1 | 1.0 |
| Ile | | | | | | | | | 0.8 | 0.1 |
| Leu | 1.0 | 1.9 | 1.6 | | | | 0.9 | 0.2 | | |
| Tyr | | | | | | | | | | |
| Phe | | | | | | | | | | |
| Try | | | | | | | | | | |

| | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | | | | | | | | | | 1.0 |
| His | | | | 0.8 | | | | | | |
| Hsl | | | | | | | | | | |
| AE Cys | | | | | | | | | | |
| Arg | | | | | | | | | | |
| Asp | 1.0 | | | | | | | | | |
| Thr | | | | 1.0 | | | | | | |
| Ser | | 1.0 | 1.1 | | 1.0 | 1.0 | 1.1 | 1.0 | | 1.0 |
| Glu & Hsr | | | | 1.0 | | | | | | |
| Pro | | | | 1.0 | | | | | | |
| Gly | 1.1 | | | 1.2 | | | | | | |
| Ala | | 1.0 | 0.1 | 1.1 | 1.0 | 1.0 | 0.9 | 0.2 | | |
| Val | | | | | | | | | | |
| Met | | 1.0 | 0.9 | | | | | | | |
| Ile | | | | 0.7 | | | | | | |
| Leu | | | | | 1.1 | 1.1 | 0.3 | 0.2 | 1.0 | |
| Tyr | | | | | | | | | | |
| Phe | 0.9 | | | | 0.9 | | | | 1.0 | |
| Try | | | | | | | | | | |

| | 101 | 102 | 103 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Lys | 1.0 | 0.7 | | | | | | | | |
| His | | | | | | | | | | |
| Hsl | | | | | | | | | | |
| AE Cys | | | | | | | | | | |
| Arg | | | | | | | | | | |
| Asp | | | | | | | | | | |
| Thr | 1.1 | 1.0 | 1.0 | | | | | | | |
| Ser | 1.1 | 1.0 | | | | | | | | |
| Glu & Hsr | | | | | | | | | | |
| Pro | | | | | | | | | | |
| Gly | | | | | | | | | | |
| Ala | | | | | | | | | | |
| Val | 0.8 | 0.1 | 1.0 | | | | | | | |
| Met | | | 1.0 | | | | | | | |
| Ile | | | | | | | | | | |
| Leu | 0.9 | 1.0 | | | | | | | | |
| Tyr | | | | | | | | | | |
| Phe | | | | | | | | | | |
| Try | | | | | | | | | | |

## DISCUSSION

*Separation of Cyanogen Bromide Fragments on Sephadex G-50*--The isolation of zones I and IV (Figure 3) deserves some comment. The amino acid analysis of zone I (Table 3) after further purification by recycling almost exactly matches that to be expected from a stoichiometric mixture of residues 1-25 and residues 77-109 in the proposed sequence; this interpretation is reinforced by finding tryptic peptides from those fragments in digest of zone I (Table 8). The elution volume of zone I is compatible with a fragment of the size of the combined portions of sequence. The isolation of two fractions from zone IV upon recycling is interesting in that the usefulness of a column effectively 9 m in length (five cycles on 180 cm) is shown.

*Proposed Sequence*--The sequence shown in Figure 13 is proposed for the opossum alpha chain based on the sequence determination of several tryptic peptides, the occurrence of those peptides in larger fragments derived from chain, and by comparison with other vertebrate, especially mammalian, alpha chains. The identification of each tryptic peptide by homology was clear upon inspection except in the case of T7-8.

The identification of T7-8 rests upon its occurrence in the fragments CB II and CAT II; in each case the homology of the other tryptic peptides is clear. The composition of T7-8 balances the

```
  1                                   10
Val Leu Ser Ala Asn Asp Lys Thr Asn Val Lys Gly Ala Try Ser Lys Val

          20                                            30
Gly Gly Asn Ser Gly Ala Tyr Met Gly Glu Ala Leu Tyr Arg Thr Phe Leu

                        40                                        50
Ser Phe Pro Thr Thr Lys Thr Tyr Phe Pro Asx Tyr Asx Phe Ser Ala Gly

                                    60
Ser Ala Gln Ile Lys Thr Gln Gly Gln Lys Ile Ala Asp Ala Val Gly Leu

     70                                            80
Ala Val Ala His Leu Asp Asp Met Pro Thr Ala Leu Ser Ser Leu Ser Asp

                    90                                        100
Leu His Ala His Glu Leu Lys Val Asp Pro Val Asn Phe Lys Phe Leu Cys

                                110
His Asn Val Leu Val Thr Met Ala Ala His Leu Gly Lys Asp Phe Thr Pro

120                                         130
Glu Ile His Ala Ser Met Asp Lys Phe Leu Ala Ser Val Ser Thr Val Leu

              140 141
Thr Ser Lys Tyr Arg
```

Figure 13

A proposed sequence for opossum hemoglobin alpha chain. Underlined residues have been placed within tryptic peptides by homology. Residues 45 and 47 are either two asp or one asp and one asn.

accounting for the composition of the fragments by composition of derived tryptic peptides; the peptide is of such length to maintain homology of the entire sequence with all known mammalian alpha chains. The sequence of T7-8: thr-gln-gly-gln-lys, although not clearly homologous to vertebrate chains (which all contain a histidyl residue at position 58) is similar to the corresponding sequence from *Chironomus* (27): thr-glu-ala-asn-arg.

A sequence for T13 is proposed on the basis of the determined partial sequence. The phe-leu-ala-ser sub-sequence clearly is homologous to the $NH_2$-terminus of T13 from alpha chains from other mammals; the initial phe is invariant in all known alpha chains. The following proposed segment, val-ser-thr is identical to the homologous segment of all known mammalian chains except echidna where val-ala-thr is present.

The proposed sequence for T12c in which the segments, asp-phe and thr-pro-glu are placed by homology is derived from the findings that homologous phe-thr-pro is invariant among mammals, and that the pro is invariant among vertebrates with respect to the alpha chain.

Residues 51-53 are assigned by considering the nearly invariant gly-ser-ala sequence among mammals (rabbit shows gly-ser-glu). Ser at position 49 is constant in all alpha chains except rabbit (thr). Phe and ala are assigned to positions 48 and 50, respectively, with

only rabbit 48 phe as possible precedent, given that these two residues be assigned to positions 48 and 50 (demanded by composition and assignment of residues 49, 51, 52, 53).

All known alpha chains show the configuration:

| 41 | 42 | 43 | 45 | 47 |
|---|---|---|---|---|
| thr | tyr | phe | his | asp |

All except the fish contain pro at position 44; position 46 has been identified as phe in all alpha chains except fish (try) and echidna (met). These considerations along with the existence of a tyr-asp segment lead to the proposed sequence for residues 41-47; tyr is assigned to the nearly invariant position 46 by its chemical similarity to phe and the fact that such a substitution is frequent in evolution at sites showing restricted variability.

*Absence of Histidine at Position 58*--All vertebrate hemoglobin or myoglobin chains previously studied (except some pathologic human mutants) contain a histidyl residue at position 58 (the seventh residue of the E helix). The invertebrates *Chironomus thummi thummi* and *Glycera dibranchiata* exhibit glutamyl and leucyl residues at that position, respectively (27, 29). The opossum glutaminyl residue would seem to more closely resemble *Chironomus* in this connection, the unusual occurrence of a methionyl residue at position 25 is of interest. It is thought that position 25 need be a glycyl residue to allow close contact between the B and E helices near the position of the distal histidine (23). This notion was given support by the

finding of an abnormal, unstable human hemoglobin (Hemoglobin Riverdale-Bronx) containing an argininyl residue in the beta chain at position corresponding to alpha 25 (79). Perhaps the substitution of gln for his at residue 58 changes the requirements for the relation between the B and E helices.

*The Proposed Sequence and the Neutral Mutation Theory*-- The proposed sequence was used to expand Table 1 into Table 13. In all comparisons of the type described on page 12, it is seen that the opossum alpha chain (say point B on page 12) has evolved more rapidly than the chain against which it is compared (point C). This finding is in agreement with Kimura's (33) prediction in that a protein of a "living fossil" has not evolved more slowly than the corresponding protein from a more rapidly evolving species.

If the observation of an increased rate of molecular evolution accompanying slow morphological evolution can be shown to be general, an interesting question may be raised with respect to maximum rates of evolution. Could it be the case that during periods of slow morphologic evolution (and necessarily low rates of selective deaths with respect to traits reflected in morphology), molecular evolution may proceed more rapidly by natural selection, relatively unhampered by additive effects of simultaneous selection (44) of multiple traits?

| | Human | Monkey | Horse | Bovine | Rabbit | Dog | Kangaroo | Echidna | Chicken | Carp | Catostomid fish | Opossum |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Human | | 4 | 18 | 17 | 25 | 23 | 27 | 38 | 35 | 71 | 68 | 40 |
| (Rhesus) monkey | | | 16 | 16 | 25 | 24 | 26 | 36 | 35 | 71 | 68 | 40 |
| Horse | | | | 18 | 25 | 27 | 29 | 42 | 40 | 70 | 66 | 42 |
| Bovine | | | | | 25 | 28 | 26 | 44 | 38 | 68 | 66 | 42 |
| Rabbit | | | | | | 28 | 37 | 49 | 44 | 74 | 71 | 50 |
| Dog | | | | | | | 33 | 39 | 45 | 70 | 69 | 45 |
| Kangaroo | | | | | | | | 49 | 41 | 74 | 73 | 42 |
| Echidna | | | | | | | | | 47 | 79 | 73 | 60 |
| Chicken | | | | | | | | | | 75 | 71 | 59 |
| Carp | | | | | | | | | | | 16 | 80 |
| Catostomid fish | | | | | | | | | | | | 80 |

Table 13. Amino Acid Differences Between Pairs of Vertebrate Alpha Chains. Exactly as Table 1, but incorporating the proposed Opossum sequence.

# SUMMARY

Amino acid sequence data for proteins with the same function in many different organisms (such as data for cytochrome c and hemoglobin) along with modern concepts of genetic mechanisms have permitted questions to be asked concerning the evolution of informational macromolecules (nucleic acids and proteins). An outstanding question among these concerns the origin of interspecies amino acid sequence differences. Did each of the observed differences come about as a result of natural selection, or can the action of random processes explain much of this variation? The notion that the latter forces have been important is embodied in the neutral mutation theory of molecular evolution.

A test of the neutral mutation theory described in this thesis was done by examining the amino acid sequence of a hemoglobin polypeptide chain from the North American opossum (*Didelphis marsupialis*), a "living fossil."

Modern techniques of protein chemistry have been applied in order to propose an amino acid sequence for the alpha chain of opossum hemoglobin.

The sequence was found to have probably diverged more from an ancestral hemoglobin than have the sequences of alpha chains from other mammals. This may be interpreted as support for the neutral mutation theory under certain assumptions.

## References


1. Perutz, M. F. Stereochemistry of cooperative effects in haemoglobin. Haem-haem interaction and the problem of allostery. Nature, 1970. 228, 726-734.

2. Perutz, M. F. & Ten Eyck, L. F. Stereochemistry of cooperative effects in hemoglobin. Cold Spring Harbor Symp. Quant. Biol., 1971. 36, 295-310.

3. Anderson, L. Intermediate structure of normal human haemoglobin: Methaemoglobin in the deoxy quaternary conformation. J. Mol. Biol., 1973. 79, 495-506.

4. Kilmartin, J. V. & Rossi-Bernardi, L. Inhibition of $CO_2$ combination and reduction of the Bohr effect in haemoglobin chemically modified at its α-amino groups. Nature, 1969. 222, 1243-1246.

5. Perutz, M. F. The Bohr effect and combination with organic phosphates. Nature, 1970. 228, 734-739.

6. Arnone, A. X-ray diffraction study of binding of 2,3-diphosphoglycerate to human deoxyhaemoglobin. Nature, 1972. 237, 146-149.

7. Perutz, M. F. The Croonian lecture, 1968. The haemoglobin molecule. Proc. Roy. Soc., Ser. B, 1969. 173, 113-140.

8. Lehmann, H. & Carrell, R. W. Variations in the structure of human haemoglobin. Brit. Med. Bull., 1969. 25, 14-23.

9. Huisman, T. H. J. & Schroeder, W. A. New aspects of the structure, function, and synthesis of hemoglobins. CRC Crit. Rev. Clin. Lab. Sci., 1970. 1, 471-526.

10. Dayhoff, M. O. Atlas of Protein Sequence and Structure. Vol. 5. Silver Spring, Maryland: National Biomedical Research Foundation, 1972.

11. Morimoto, H., Lehmann, H., & Perutz, M. F. Molecular pathology of human haemoglobin: Stereochemical interpretation of abnormal oxygen affinities. Nature, 1971. 232, 408-413.

12. Perutz, M. F. & Lehmann, H. Molecular pathology of human haemoglobin. Nature, 1968. 219, 202-209.

13. Greer, J. Three-dimensional structure of abnormal human haemoglobins Kansas and Richmond. J. Mol. Biol., 1971. 59, 99-106.

14. Greer, J. Three dimensional structure of abnormal human haemoglobins M Hyde Park and M. Iwate. J. Mol. Biol., 1971. 59, 107-126.

15. Perutz, M. F., del Pulsinelli, P., Ten Eyck, L., Kilmartin, J. V., Shibata, S., Luchi, I., Miyaji, T., & Hamilton, H. B. Haemoglobin Hiroshima and the mechanism of the alkaline Bohr effect. Nature New Biol., 1971. 232, 147-149.

16. Greer, J. & Perutz, M. F. Three-dimensional structure of haemoglobin Rainier. Nature New Biol., 1971. 230, 261-264.

17. Dayhoff, M. O. Atlas of Protein Sequence and Structure. Vol. 5. Supplement I. Silver Spring, Maryland: National Biomedical Research Foundation, 1973.

18. Whittaker, R. G., Fisher, W. K., & Thompson, E. O. P. Studies on monotreme proteins I. Amino acid sequence of the β-chain in haemoglobin from the echidna, *Tachyglossus aculeatus aculeatus*. Aust. J. Biol. Sci., 1972. 25, 989-1004.

19. Whittaker, R. G., Fisher, W. K., & Thompson, E. O. P. Studies on monotreme proteins II. Amino acid sequence of the α-chain in haemoglobin from the echidna, *Tachyglossus aculeatus aculeatus*. Aust. J. Biol. Sci., 1973. 26, 877-888.

20. DeJong, W. W. W. Chimpanzee fetal hemoglobin. Structure and heterogeneity of the γ chain. Biochim. Biophys. Acta, 1971. 251, 217-226.

21. Hendrickson, W. A. & Love, W. E. Structure of lamprey haemoglobin. Nature New Biol., 1971. 232, 197-203.

22. Love, W. E., Klock, P. A., Lattman, E. E., Padlan, E. A., Ward, K. B. jr., & Hendrickson, W. A. The structure of lamprey and bloodworm hemoglobins in relation to their evolution and function. Cold Spring Harbor Symp. Quant. Biol., 1971. 36, 349-357.

23. Dickerson, R. E. & Geis, I. The structure and action of proteins. New York: Harper and Row, 1969. (page 62).

24. Beard, J. M., & Thompson, E. O. P. Studies on marsupial proteins. V. Amino acid sequence of the α-chain of haemoglobin from the grey kangaroo, *Macropus giganteus*. Aust. J. Biol. Sci., 1971. 24, 765-786.

25. Bunn, H. F. & Briehl, R. W. The interaction of 2,3-diphosphoglycerate with various human hemoglobins. J. Clin. Invest., 1970. 49, 1088-1095.

26. Riggs, A. Mechanism of the enhancement of the Bohr effect in mammalian hemoglobins by diphosphoglycerate. Proc. Nat. Acad. Sci., 1971. 68, 2062-2065.

27. Buse, G., Braig, S., & Braunitzer, G. The constitution of a hemoglobin (erythrocruorin) of an insect (*Chironomus thummi thummi;* Diptera). Z. Physiol. Chem., 1969. 349, 1686-1690.

28. Huber, R., Epp, O., Steigemann, W., & Formanek, H. The atomic structure of erythrocruorin in the light of the chemical sequence and its comparison with myoglobin. Eur. J. Biochem., 1971. 19, 42-50.

29. Imamura, T., Baldwin, T. O., & Riggs, A. The amino acid sequence of the monomeric hemoglobin component from the bloodworm, *Glycera dibranchiata*. J. Biol. Chem., 1972. 247, 2785-2797.

30. Zuckerkandl, E. & Pauling, L. Evolutionary divergence and convergence in proteins. In V. Bryson & H. J. Vogel (Eds.) Evolving genes and proteins. New York: Academic Press, 1965. pp. 97-166.

31. Schroeder, W. A., Huisman, T. H. J., Shelton, J. R., Shelton, J. B., Kleihauer, E. F., Dozy, A. M., & Robberson, B. Evidence for multiple structural genes for the γ chain of human fetal hemoglobin. Proc. Nat. Acad. Sci., 1968. 60, 537-544.

32. Zuckerkandl, E. & Pauling, L. Molecular disease, evolution, and genic heterogeneity. In M. Kasha & B. Pullman (Eds.) Horizons in biochemistry. New York: Academic Press, 1962. pp. 189-225.

33. Kimura, M. The rate of molecular evolution considered from the standpoint of population genetics. Proc. Nat. Acad. Sci., 1969. 63, 1181-1188.

34. Air, G. M., Thompson, E. O. P., Richardson, B. J., & Sharman, G. B. Amino acid sequences of kangaroo myoglobin and haemoglobin and the date of marsupial-eutherian divergence. Nature, 1971. 229, 391-394.

35. Jones, R. T., Brimhall, B., & Duerst, M. Amino acid sequence of the α and β chains of dog hemoglobin. Fed. Proc., 1971. 30, 1259. (Abstract).

36. Powers, D. A. & Edmundson, A. B. Multiple hemoglobins of catostomid fish. II. The amino acid sequence of the major α chain from *Catostomus clarkii* hemoglobins. J. Biol. Chem., 1972. 247, 6694-6707.

37. Kimura, M. Evolutionary rate at the molecular level. Nature, 1968. 217, 624-626.

38. Kaplan, N. O. Evolution of dehydrogenases. In V. Bryson & H. J. Vogel (Eds.) Evolving genes and proteins. New York: Academic Press, 1965. pp. 243-278.

39. Margoliash, E. & Smith, E. L. Structural and functional aspects of cytochrome c in relation to evolution. In V. Bryson & H. J. Vogel (Eds.) Evolving genes and proteins. New York: Academic Press, 1965. pp. 221-242.

40. Muller, H. J. cited in Kimura, M. Evolutionary rate at the molecular level. Nature, 1968. 217, 624-626.

41. Haldane, J. B. S. cited in Kimura, M. Evolutionary rate at the molecular level. Nature, 1968. 217, 624-626.

42. King, J. L. & Jukes, T. H. Non-Darwinian evolution. Science, 1969. 164, 788-798.

43. Crow, J. F. Molecular genetics and population genetics. Proc. Twelfth Intern. Cong. Genet., 1969. 3, 105-113.

44. Maynard Smith, J. "Haldanes dilemma" and the rate of evolution. Nature, 1968. 219, 1114-1116.

45. Clarke, B. Darwinian evolution of proteins. Science, 1970. 168, 1009-1011.

46. Richmond, R. C. Non-Darwinian evolution: a critique. Nature, 1970. 225, 1025-1028.

47. Simpson, G. G. The meaning of evolution. New Haven: Yale University Press, 1949.

48. Bardack, D. & Zangerl, R. First fossil lamprey: a record from the Pennsylvanian of Illinois. Science, 1968. 162, 1265-1267.

49. Li, S. L. & Riggs, A. The amino acid sequence of hemoglobin V from the lamprey, *Petromyzon marinus*. J. Biol. Chem., 1970. 245, 6149-6169.

50. Braunitzer, G. & Fujiki, H. cited in Li, S. L. & Riggs, A. The amino acid sequence of hemoglobin V from the lamprey, *Petromyzon marinus*. J. Biol. Chem., 1970. 245, 6149-6169.

51. Romer, A. S. Man and the vertebrates. Chicago: University of Chicago Press, 1941.

52. Stebbins, G. L. Processes of organic evolution. Englewood Cliffs, New Jersey: Prentice-Hall, 1966.

53. Romer, A. S. Major steps in vertebrate evolution. Science, 1967. 158, 1629-1637.

54. Sharman, G. B. Reproductive physiology of marsupials. Science, 1970. 167, 1221-1228.

55. Slaughter, B. H. Earliest known marsupials. Science, 1968. 162, 254-255.

56. Martin, P. G. Darwin rise hypothesis of dispersion of marsupials. Nature, 1970. 225, 197-198.

57. Matsuda, G., Takei, H., Wu, K. C., & Shiozawa, T. The primary structure of the α-polypeptide chain of AII component of adult chicken hemoglobin. Intern. J. Prot. Res., 1971. 3, 173-174.

58. Hilse, K. & Braunitzer, G. The amino acid sequence of the α-chain of both main components of carp hemoglobin. Z. Physiol. Chem., 1968. 349, 433-450.

59. Chauvet, J. P. & Acher, R. The β chain of frog hemoglobin (*Rana esculenta*): the complete amino acid sequence. FEBS Lett., 1970. 10, 136-138.

60. Jones, R. T., Brimhall, B., & Huisman, T. H. J. Structural characterization of two δ chain variants. Hemoglobin $A_2'$ ($B_2$) and hemoglobin Flatbush. J. Biol. Chem., 1967. 242, 5141-5145.

61. Schroeder, W. A., Shelton, J. R., Balog-Shelton, J., Cormick, J., & Jones, R. T. The amino acid sequence of the γ chain of human fetal hemoglobin. Biochemistry, 1963. 2, 992-1008.

62. Clegg, J. B., Naughton, M. A., & Weatherall, D. J. An improved method for the characterization of human haemoglobin mutants. Nature, 1965. 207, 945-947.

63. Spackman, D. H., Stein, W. H. & Moore, S. Automatic recording apparatus for use in the chromatography of amino acids. Anal. Chem., 1958. 30, 1190-1206.

64. Jones, R. T. & Weiss, G. Long-path flow cells for automatic amino acid analysis. Anal. Biochem., 1964. 9, 377-391.

65. Light, A. Leucine aminopeptidase in sequence determination of peptides. In C. H. W. Hirs (Ed.) Methods in Enzymology. Vol. 25. New York: Academic Press, 1972. (pp. 253-261).

66. Mross, G. A. jr. Solid phase peptide degradation. Ph. D. dissertation, Univ. California San Diego, 1971.

67. Brimhall, B., Duerst, M., Hollan, S. R., Stenzel, P., Szelenyi, J., & Jones, R. T. Structural characterization of hemoglobins J-Buda [α61 (E10) Lys→Asn] and G-Pest [α74 (EF3) Asp→Asn]. Biochim. Biophys. Acta, in press.

68. Stark, G. R. Use of cyanate for determining $NH_2$-terminal residues in proteins. In C. H. W. Hirs (Ed.) Methods in Enzymology. Vol. 11. New York: Academic Press, 1967. (pp. 125-138).

69. Schroeder, W. A. The primary structure of proteins. New York: Harper and Row, 1968.

70. Dixon, H. B. F. & Perham, R. N. Reversible blocking of amino groups with citraconic anhydride. Biochem. J., 1968. 109, 312-314.

71. Singhal, R. P. & Atassi, M. Z. Immunochemistry of sperm whale myoglobin. IX. Specific interaction of peptides obtained by cleavage at arginine peptide bonds. Biochemistry, 1971. 10, 1756-1762.

72. Raftery, M. A. & Cole, R. D. Tryptic cleavage at cysteinyl peptide bonds. Biochem. Biophys. Res. Commun., 1962. 10, 467-472.

73. Cole, R. D. S-Aminoethylation. In C. H. W. Hirs (Ed.) Methods in Enzymology. Vol. 11. New York: Academic Press, 1967. (pp. 315-317).

74. Gross, E. The cyanogen bromide reaction. In C. H. W. Hirs (Ed.) Methods in Enzymology. Vol. 11. New York: Academic Press, 1967. (pp. 238-254).

75. Jones, R. T. Automatic peptide chromatography. In D. Glick (Ed.) Methods of Biochemical Analysis. Vol. 18. New York: Interscience, 1970. (pp. 205-258).

76. Edman, P. Sequence determination. In S. B. Needleman (Ed.) Protein sequence determination. New York: Springer-Verlag, 1970. (pp. 211-255).

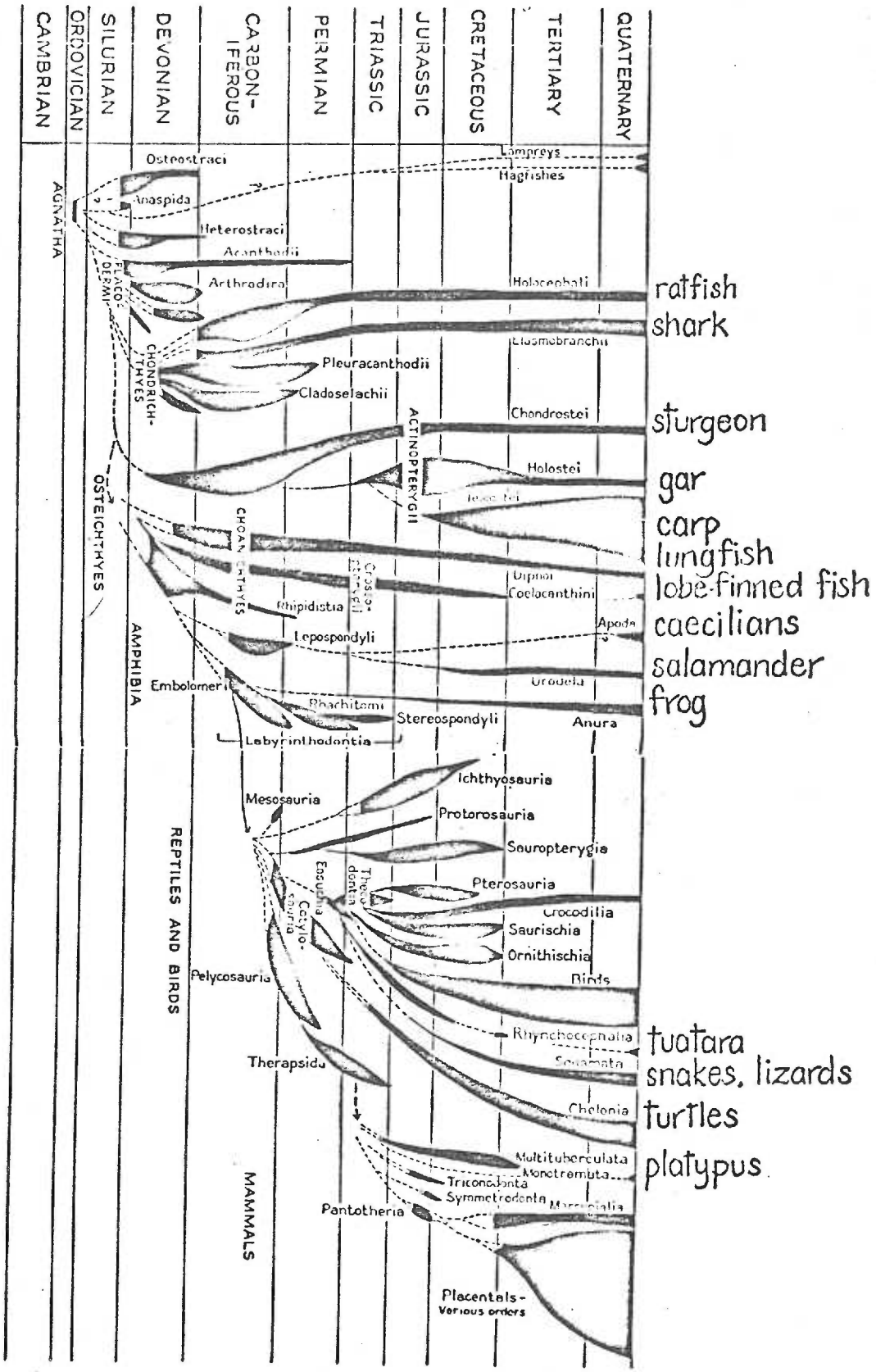
77. Salnikow, J., Liao, T., Moore, S., & Stein, W. H. Bovine pancreatic deoxyribonuclease A. Isolation, composition, and amino acid sequences of the tryptic and chymotryptic peptides. J. Biol. Chem., 1973. 248, 1480-1488.

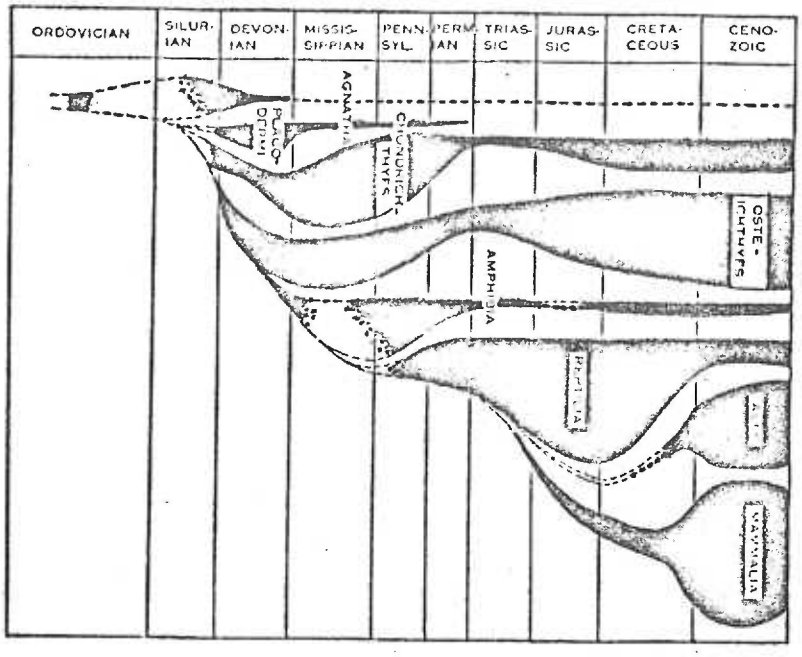
78. Jones, R. T. Structural studies of aminoethylated hemoglobins by automatic peptide chromatography. Cold Spring Harbor Symp. Quant. Biol., 1964. 29, 297-307.

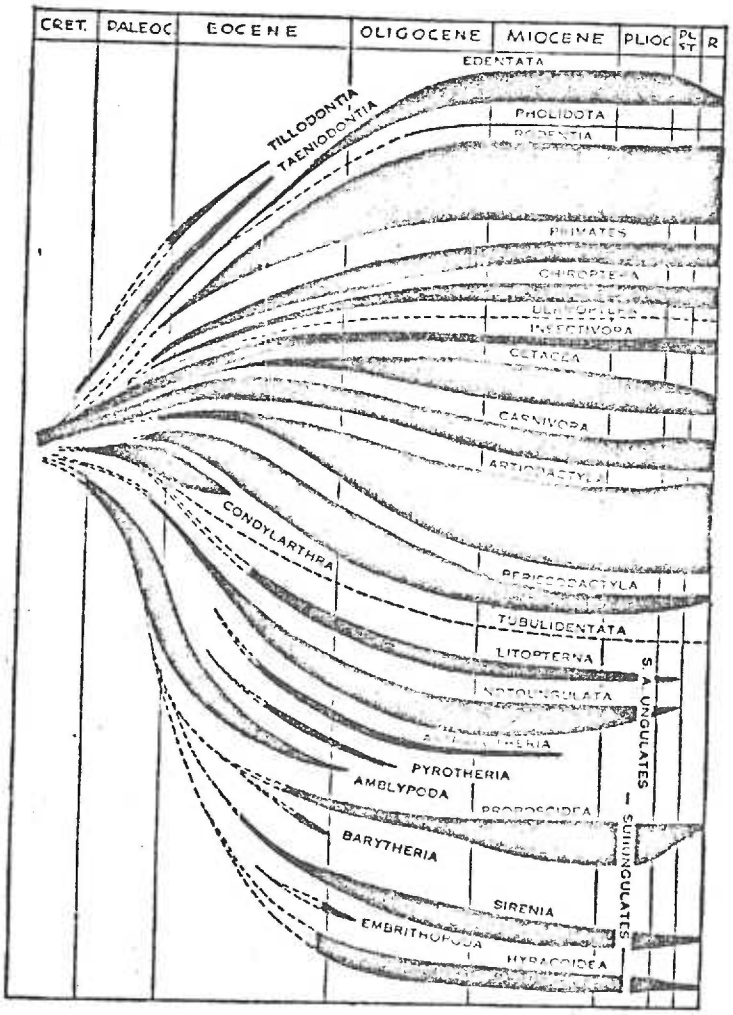
79. Ranney, H. M., Jacobs, A. S., Udem, L., & Zalusky, R. Hemoglobin Riverdale-Bronx an unstable hemoglobin resulting from the substitution of arginine for glycine at helical residue B6 of the beta polypeptide chain. Biochem. Biophys. Res. Commun., 1969. 33, 1004-1008.

# Appendix I

Diagrams of Vertebrate Evolution from Romer (51).


CAMBRIAN
ORDOVICIAN
SILURIAN
DEVONIAN
CARBON-IFEROUS
PERMIAN
TRIASSIC
JURASSIC
CRETACEOUS
TERTIARY
QUATERNARY
AGNATHA
Ostreostraci
Anaspida
Heterostraci
Lampreys
Hagfishes
PLACO-DERMI
Acanthodii
Arthrodira
Holocephali
ratfish
shark
Elasmobranchii
CHONDRICH-THYES
Pleuracanthodii
Cladoselachii
Chondrostei
sturgeon
ACTINOPTERYGII
Holostei
gar
carp
lungfish
lobe-finned fish
caecilians
salamander
frog
OSTEICHTHYES
CHOANICHTHYES
Crossopterygii
Dipnoi
Coelacanthini
Rhipidistia
Apoda
Lepospondyli
Urodela
AMPHIBIA
Embolomeri
Rhachitomi
Stereospondyli
Anura
Labyrinthodontia
Ichthyosauria
Mesosauria
Protorosauria
Sauropterygia
Pterosauria
Crocodilia
Saurischia
Ornithischia
Birds
Pelycosauria
REPTILES AND BIRDS
Rhynchocephalia
tuatara
Squamata
snakes, lizards
Therapsida
Cotylosauria
Testudinata
Chelonia
turtles
Multituberculata
platypus
Monotremata
Triconodonta
Symmetrodonta
Marsupialia
Pantotheria
MAMMALS
Placentals-
Various orders




sloth
scaly anteater

bat
flying lemur
mole

whale
dog
pig
cattle
sheep
horse
aardvark

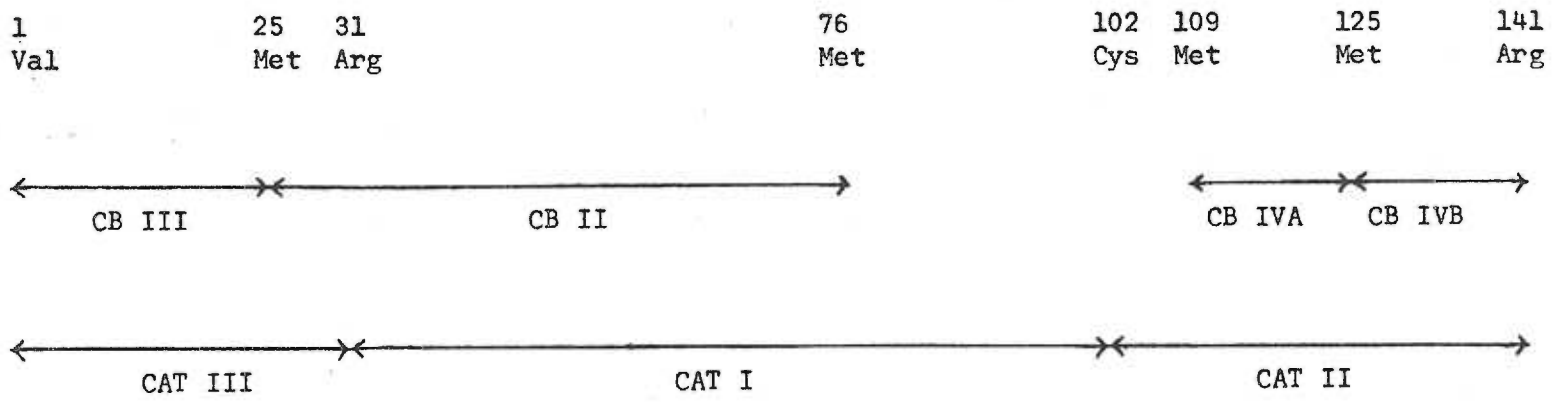
elephant

manatee
hyrax

## APPENDIX II - A Geological Timetable

| ERA | PERIOD | EPOCH | Stebbins (52) | Simpson (47) | Romer (51) |
|---|---|---|---|---|---|
| Cenozoic | Quaternary | Recent | | .025 | |
| | | Pleistocene | 1 | 1 | 1 |
| | Tertiary | Pleiocene | 11 | 12 | 7 |
| | | Miocene | 25 | 28 | 19 |
| | | Oligocene | 36 | 39 | 30 |
| | | Eocene | 54 | 58 | 45 |
| | | Paleocene | 65 | 75 | 55 |
| Mesozoic | Cretaceous | | 135 | 135 | 120 |
| | Jurassic | | 181 | 165 | 155 |
| | Triassic | | 220 | 205 | 190 |
| Paleozoic | Permian | | 280 | 230 | 215 |
| Carboniferous | (Pennsylvanian) | | 310 | 255 | |
| | (Mississippian) | | 355 | 280 | 300 |
| | Devonian | | 405 | 325 | 350 |
| | Silurian | | 425 | 360 | 390 |
| | Ordovician | | 500 | 425 | 480 |
| | Cambrian | | 600 | 505 | 550 |

A geological timetable in millions of years since the beginning of divisions of time. The dates are listed from the three given references.




Appendix III. Diagram of relationship between cyanogen bromide fragments (CB II, III,IVA, IVB), restricted tryptic fragments (CAT I, II, III) and proposed sequence for the opossum alpha chain.