THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 256, No. 9, Issue of May 10, pp. 4301-4309, 1981
Printed in U.S.A.

Amino Acid Sequence of a Myoglobin from Lace Monitor Lizard, *Varanus varius*, and Its Evolutionary Implications*

(Received for publication, August 1, 1980, and in revised form, January 15, 1981)

Nobuyo Maeda and Walter M. Fitch

From the Department of Physiological Chemistry, University of Wisconsin, Madison, Wisconsin 53706

Myoglobin was purified from a muscle extract of lace monitor lizard, *Varanus varius*, by Sephadex G-75, followed by DEAE-cellulose column chromatography. The apomyoglobin was cleaved with cyanogen bromide. The largest fragment was further digested with pepsin, trypsin, and α-chymotrypsin. From the amino acid sequence of the cyanogen bromide fragments, together with those of tryptic peptides of apomyoglobin, the complete amino acid sequence of lizard myoglobin was deduced.

To investigate the tetrapod and amniote origins, many possible phylogenetic trees were constructed using the myoglobin sequences, including those of map turtle and lace monitor lizard. The tree that requires the minimum number of nucleotide substitutions in their genes for the myoglobin sequences to have evolved from a common ancestor was different from the similarly most parsimonious trees for cytochrome *c* or for α-hemoglobin. The trees were different from each other and from the tree that best reflects current biological opinions.

Problems in evolution or taxonomy are classically studied by the analysis of fossil records and the morphology of current representatives. Fossils directly illuminating the origin of tetrapods are scarce, particularly for the amphibians. Opinions regarding the early phylogeny of extant Amphibia, subclass Lissamphibia, fall into two groups. Most workers (Romer, 1966; Porter, 1972) regard the Lissamphibia as monophyletic, having diverged into three suborders (frogs, salamanders, and apodons) from a common amphibian ancestor. The other opinion (Jarvic, 1963) suggests that the frogs and amniotes (reptiles, birds and mammals) had a common ancestor not shared with the salamanders.

The early evolution of the amniotes is considered less uncertain than that of the tetrapods and, following Romer (1966), most zoologists agree that the Anapsida, leading to turtle and tortoise, and the Synapsida, leading to the Mammalia, separated from the main diapsid Lepidosaurians nearly 300 million years ago. Though the evidence is weak, turtles are considered to have branched off first. In the mid-Permian, the Archosaurians, leading to the Crocodilians, Dinosaurs and birds, diverged from the Lepidosaurians, which soon branched into Rhyncocephalia and Squamata. By the end of the Jurassic, the Squamata had produced most modern lizard families and the snakes.

Since Fitch and Margoliash (1967) first used the amino acid sequences of orthologous proteins for the systematic study of evolutionary relationships, a variety of proteins have been examined with the method. To aid in the study of the tetrapod and amniote origins systematically using the amino acid sequence of proteins, the myoglobin was isolated from the muscle of the lace monitor lizard, *Varanus varius*. This paper describes the amino acid sequence determination of lizard myoglobin and its evolutionary implication.

EXPERIMENTAL PROCEDURES

Materials—Proteinases, reagents, and chromatographic media were the same as described previously (Maeda and Fitch, 1981).

Isolation of Myoglobin—Muscle extract preparation from the lizard, *V. varius*, was a kind gift of Dr. E. Margoliash. That preparation was the fraction of the muscle extract which did not adsorb on an IRC-50 column. Ammonium sulfate was added to the solution to 55% saturation. The mixture was centrifuged at 9000 × *g* for 30 min and the supernatant was dialyzed against water containing 1 mM KCN. The purification of myoglobin and the removal of the heme group were carried out by the same method used for turtle myoglobin (Maeda and Fitch, 1981).

Pyridylethylation of Apomyoglobin—Apomyoglobin (108 mg) was dissolved in 5 ml of 0.15 M Tris-HCl buffer, pH 7.4, containing 6 M guanidine hydrochloride and 0.1% EDTA and reacted with 0.01 M dithiothreitol for 3 h at 37 °C. To the reaction mixture 4-vinylpyridine (30 μl) was added and the mixture stirred continuously for 1.5 h at room temperature. The salt and excess reagents were removed by gel filtration on a Sephadex G-50, fine grade, column (1.8 × 93 cm) using 30% (v/v) formic acid, 10% (v/v) acetic acid as eluant.

Automatic Sequence Analysis—Automatic sequence analyses of the apomyoglobin and CNBr-cleavage fragments were carried out by Beckman Sequencer 890C with the Quadrol or the dimethylallylamine program. Polybrene (Pierce, 0.2 mg) was added to the reaction cup together with protein or peptides (0.2-0.3 μmol) to help retention in the cup.

Digestion with Staphylococcus aureus Protease V8—Peptides (approximately 0.5 μmol) were dissolved in 0.2 ml of 1% (w/v) ammonium bicarbonate and digested with the protease (Miles, 85 μg) for 16 h at 37 °C.

Other Sequence Methods—The other methods for derivatizing, digesting, separating, and sequencing peptides were the same as described previously (Maeda and Fitch, 1981).

Phylogenetic Reconstructions—The myoglobin sequences of the lizard (this paper) and of the turtle (Maeda and Fitch, 1981) were compared to 48 other myoglobins and their most parsimonious tree was compared to the tree currently preferred by biologists and to those of two other orthologous sets of sequences, *viz.*, 25 cytochromes *c* and 40 α-hemoglobins. Trees were examined by the method of Fitch (1971) and Fitch and Farris (1974). A program[1] was used that will perform the preceding task on all 10,395 different unrooted trees for 8 taxa.

RESULTS

The purification of myoglobin by Sephadex G-75 and DE52 column chromatography is shown in Figs. 1 and 2, respectively. The main fraction of heme protein eluted from DE52 was concentrated and used for sequence analysis without

* This work was supported by Grants DEB78-04291 and DEB78-14197 from National Science Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] W. M. Fitch, unpublished data.

further purification, although the material gave a very faint band of contaminant on disc gel electrophoresis at pH 8.9. The amino acid composition of myoglobin is given in Table I. The values were obtained from the analysis of 10 separate 22-h hydrolysates. Values of valine and isoleucine are from 72-h hydrolysates (3 analyses). Tryptophan content was determined by 3 N mercaptoethanesulfonic acid hydrolysis of the protein and cysteine was analyzed as pyridylethyl cysteine.

Cyanogen bromide cleavage of *S*-pyridylethylated apomyoglobin produced 6 fragments, CNBr-I, -II, -III-2, -III-3, -IV-1, and IV-2, of length ranging from 11 to 76 residues. The amino acid composition of peptides CNBr-I, -III-2, -III-3, -IV-1, and -IV-2 explained that of apomyoglobin. Peptide CNB-II, which was a product of incomplete cleavage, was composed of peptides CNBr-III-2 and -III-3. The NH_2-terminal amino acid sequence of peptides CNB-III-2 and CNBr-I was determined by automatic Edman degradation. NH_2-terminal residues of CNBrIII-3 degraded manually (6 cycles) were identical with that of whole apomyoglobin determined automatically. The CNBr-cleavage peptides were ordered as shown in Fig. 3 with the aid of analyses of peptides isolated from a tryptic digest of *S*-pyridylethylated apomyoglobin. The tryptic digest was separated into 5 fractions, TI-TV, by Sephadex G-25 column chromatography in 0.05 M ammonium bicarbonate. Fractions TII-TV were further separated into their components by DE52 or phosphocellulose column chromatography or by thin layer chromatography and thin layer electrophoresis.

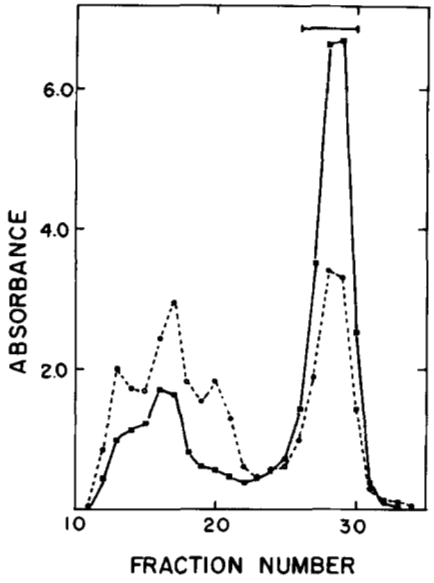


FIG. 1. **Gel filtration of lizard muscle extract on Sephadex G-75 column.** The extract (10 ml) was applied on a Sephadex G-75 column (1.8 × 60 cm) equilibrated with 0.05 M Tris/HCl buffer, pH 8.5, containing 1 mM KCN, and eluted with the same buffer. The elution was monitored by measuring the absorbance at 280 nm (○- - -○) and 410 nm (■——■). Each 3.6-ml fraction was collected at a flow rate of 9 ml/h.

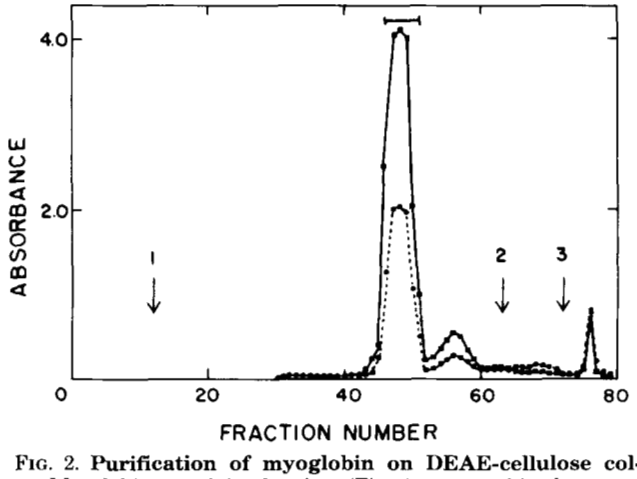


FIG. 2. **Purification of myoglobin on DEAE-cellulose column.** Myoglobin containing fractions (Fig. 1) were combined, concentrated, and changed to 0.01 M Tris/HCl buffer, pH 8.5, containing 1 mM KCN by ultrafiltration. Protein solution (30 ml) was applied to a DEAE-cellulose (DE52) column (1.8 × 11 cm) equilibrated with the same buffer. At *1*, linear gradient elution was begun with 200 ml of buffer in the mixing chamber and 200 ml of buffer containing 0.1 M NaCl in the reservoir. At *2* and *3*, the NaCl concentration in the buffer was abruptly stepped up to 0.1 and 1 M, respectively. Each 4.8-ml fraction was collected at a flow rate of 72 ml/h. *Symbols* are same as in Fig. 1.

TABLE I

Amino acid composition of lizard myoglobin

The results are expressed in molar proportions of the amino acids, assuming the sum of the numbers of aspartic acid, alanine, leucine, and histidine to be 47. Values in parentheses give the numbers obtained from sequence study.

Amino acid	Molar proportion	
Aspartic acid	13.73	(14)
Threonine	6.69	(7)
Serine	6.23	(6)
Glutamic acid	21.19	(22)
Proline	4.54	(5)
Glycine	10.08	(10)
Alanine	12.31	(12)
Valine[a]	8.03	(8)
Methionine	4.13	(4)
Isoleucine[a]	9.27	(10)
Leucine	12.72	(13)
Tyrosine	1.72	(2)
Phenylalanine	7.12	(7)
Lysine	15.97	(16)
Histidine	8.11	(8)
Arginine	6.16	(6)
Cysteine[b]	1.08	(1)
Tryptophan[c]	2.35	(2)
Total		153

[a] The value obtained from 72-h hydrolysates are given.
[b] Determined as pyridylethylcysteine.
[c] Tryptophan was determined by the analysis of the hydrolysates with 3 N mercaptoethanesulfonic acid.

Eight cycles of manual Edman degradation of fraction TI indicated that this fraction is a mixture of at least three peptides whose NH_2 termini were positions 80, 99 and 103 of apomyoglobin. None of the three peptides could be recovered from a phosphocellulose column, presumably because of tight adsorption or precipitation. These three peptides comprise the hydrophobic region of apomyoglobin (residue 99–118). Residues 103 to 118 were purified as tryptic peptide tCNBr Ic from a tryptic digest of peptide CNBrI in approximately 10% yield. The amino acid sequence of this region was deduced from the study of peptides and the α-chymotryptic peptides of peptide CNBrI (Fig. 3).

The whole sequence was determined as shown in Fig. 3. The experimental details of fragmentation of peptide chains, purification, and amino acid composition of peptides are given in the miniprint supplement.[2]

[2] Portions of this paper (including part of "Results," Figs. 1 to 10, and Tables I to IX) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20014. Request Document No. 80M-1601, cite author(s), and include a check or money order for $10.40 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

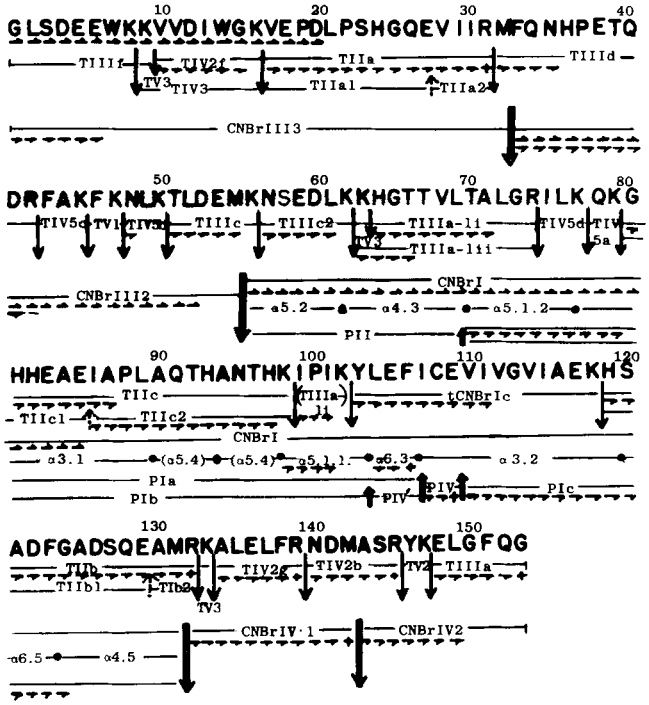


FIG. 3. **Amino acid sequence of lizard myoglobin.** Peptides are shown with names. —, amino acids detected by automatic Edman degradation; —, amino acids detected by manual Edman degradation; +, a residue determined by amino acid analysis of the remainder after several cycles of Edmans; ⇓, cyanogen bromide cleavage; ↓, ⇑, and ●, tryptic, *S. aureus* V8, peptic, and α-chymotryptic digestion, respectively. The *peptide in parentheses* was obtained as a mixture.

The amino acid sequence from lace monitor lizard was compared with those of other animals available. Lizard myoglobin shares some unique residue replacements with map turtle myoglobin, namely Ile-29, Gln-40, Arg-42, Ala-44, Arg-74, and Cys-108. In addition to these, the interesting features of residue replacement in the lizard myoglobin are as follows.

Glu-5—This is a unique replacement, and the negative charge of the side chain is common only to turtle (aspartic acid).

Lys-8—This is common to alligator with histidine in turtle and in harbour seal. All other mammal and bird myoglobins have glutamine.

Lys-9—This is a unique replacement. Histidine occurs in alligator, turtle, and shark myoglobins. The rest have non-charged residues at this position.

Val-11—This is unique in lizard. Asparagine occurs here in shark and leucine occurs in all others.

Asp-12—Alligator is the only other species which has a negative charge in its side chain at this position.

Met-32—This is a unique replacement. All others have leucine.

Leu-52—Hydrophobic residues occur at this position also in turtle (isoleucine) and shark (valine). Most others have glutamic acid with variations of proline, serine, and glutamine at this position.

Gln-78—This is common to alligator and shark. All mammal and bird myoglobins have lysine. Leucine occurs in turtle and alanine occurs in carp and tuna.

Ala-87—This is common to hedgehog. All others have lysine.

Thr-96—This is common to carp. Glutamine occurs in alligator and lysine occurs in all others.

Glu-129—The negative side chain at this position is common only to penguin and shark.

Arg-145—This is a unique replacement, although many others have lysine at this position.

As is the case of turtle myoglobin, the effect of these replacements on the function of molecule is uncertain in the absence of functional data.

DISCUSSION

The determination of the sequences of lizard and turtle myoglobin was undertaken in order to throw light on the origins of the amniotes and tetrapods. There is a problem in the paleontological record that leads to uncertainties about the order of divergence of several of the major groups and sequence data appear to be powerful in their ability to recognize these features (Fitch and Margoliash, 1967).

In analyzing amino acid sequences in an evolutionary context, the differences in those sequences requires that mutations appeared and were fixed in the diverging populations. Thus nucleotide substitutions are required to account for the observed amino acid replacements. But no postulated nucleotide substitution has to be true since a postulated adenine to guanine substitution could, for example, have been in reality adenine to cytosine to guanine. This places each postulated substitution in the category of an assumption made to explain the data. The computer programs used in this study are based on the proposition that we should minimize the number of assumptions necessary to explain the sequence differences. This minimum then becomes the criterion for the acceptability of a tree depicting the ancestral relations of the taxa whose orthologous sequences are being examined. That tree is best that requires the fewest nucleotide substitutions and is called the most parsimonious tree.

While the trees obtained with one set of sequences are largely congruent with those obtained from a different set, the truth is that inconsistencies are not rare and occur most often precisely where the questions are most interesting and the issue most in doubt. We therefore felt obligated to compare the myoglobin results with those from other sequences. We chose cytochrome *c* and α-hemoglobin.

The number of possible unrooted trees goes up as the product of successive odd numbers and at 8 taxa there are

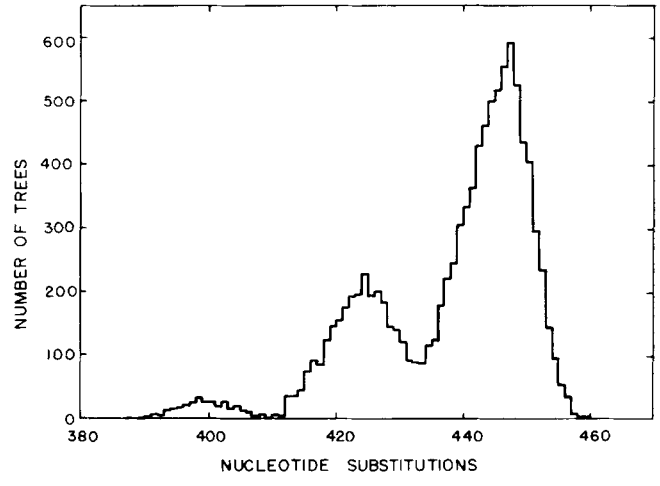


FIG. 4. **Distribution of the 10,395 possible unrooted trees for myoglobins.** The sequence data employed are lizard (*V. varius*; present paper), turtle (*G. geographica*; Maeda and Fitch, 1981), and alligator (Dene *et al.*, 1980), shark (Fisher and Thompson, 1979), and ancestral sequences computed for 40 mammals, chicken and penguin, tuna, and carp and hemoglobins of lamprey and sea lamprey. Nucleotide substitutions ranged from 387–460. Positions that require the same number of substitutions regardless of tree topology were omitted.

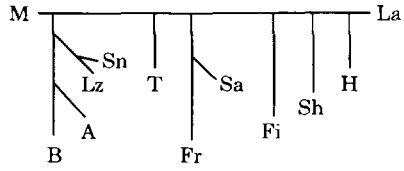
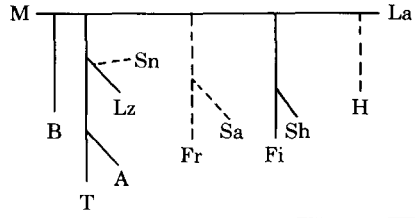
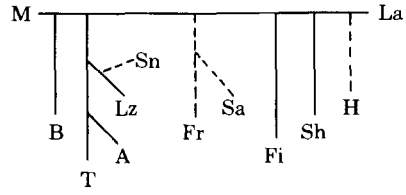
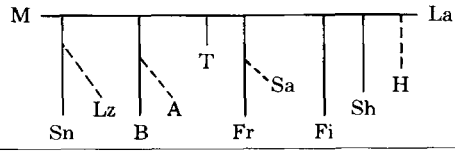
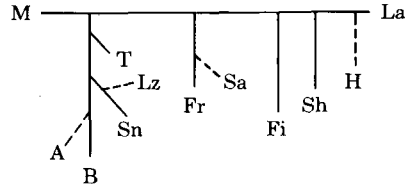
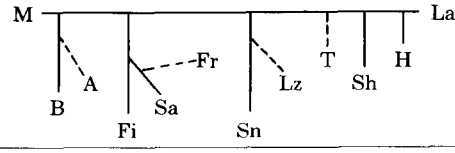
already 10,395 different trees. It is thus difficult to substantiate for large numbers of taxa that any given tree is indeed the most parsimonious from among the universe of possible trees. But we are only interested in the origins of the major taxa of lower vertebrates, which never number more than 8 for any 1 of the 3 proteins being examined. To obtain the full utility of the large number of sequences at hand, all sequences for a single major taxon are examined separately to provide a putative ancestral sequence for that group. Thus, the 40 mammalian myoglobins were used to construct an ancestral mammalian representative. This requires an assumption regarding the phylogeny of the mammalian taxa represented in the data set. We used the same phylogeny for all three proteins in such cases and chose, as being least biased with respect to the questions being asked, the tree that seems best to reflect current biological opinion. The major taxa for which there were 2 or more sequences for these proteins were mammals, birds, and fish. All other taxa were represented by a single sequence.

With 8 sequences in hand, all 10,395 possible trees were

TABLE II

Number of nucleotide substitutions in three protein sets

Myoglobin (Mb), cytochrome *c* (Cyt *c*), and α-hemoglobin (α-Hb) for various possible phylogenetic trees. Topology of each tree (*a*, biological; *b*, myoglobin, best; *c*, myoglobin, modified, *d*, cytochrome *c*, best; *e*, cytochrome *c*, 2nd best; *f*, α-hemoglobin, best) is shown schematically. *M*, ancestral mammal; *B*, ancestral bird; *A*, alligator (*Alligator mississippiensis*), Sn, snake; Lz, lace monitor lizard (*V. varius*); T, turtle; Fr, bull frog (*Rana catesbeiana*); Sa, salamander (*Taricha granulosa*); F, ancestral fish; Sh, shark; H, hagfish (*Myxine glutinosa*); La, lamprey (ancestral hemoglobin of lamprey and sea lamprey used for both myoglobin and α-hemoglobin). Branches shown in broken lines are species whose sequences are not available in the protein set. Numbers of taxa compared are 50, 25, and 40 for myoglobin, cytochrome *c*, and α-hemoglobin, respectively. The numbers in parentheses are those obtained when the number up to each ancestor are subtracted from the totals above in order to get values related only to differences among the relevant taxa. The species used and their sequences can be found in Dayhoff (*1978*). Others are: alligator (Dene *et al.*, 1980); snake (cytochrome *c*, *Crotalus adamanteus* (Bahl *et al.*, 1965); hemoglobin, *Vipera aspis* (Duguet *et al.*, 1974)); turtle (cytochrome *c*, *Chelydra serpentina* (Chan *et al.*, 1966); myoglobin (*Graptemys geographica* (Maeda and Fitch, 1981)): lizard (this work); frog (Chan *et al.*, 1967); salamander (Coates *et al.*, 1977); shark (cytochrome *c*, *Squalus sucklii*, Goldstone and Smith (1967); myoglobin and hemoglobin, *Heterodontus portusjacksoni*, Fisher and Thompson (1979) and Nash *et al.*, (1976), respectively); Hagfish (Liljeqvist *et al.* 1979); lamprey (hemoglobin *Lampetra fluviatilis* (Braunitzer and Fujiki, 1969) and *Petromyzon marinus* (Li and Riggs, 1970)); cytochrome *c* (*Entosphenus tridentatus*, Nolan *et al.*, 1973).

	Tree topology of higher taxa	Mb	Cyt *c*	α-Hb	Total
Biological tree		813 (431)	145 (99)	950 (536)	1908 (1066)
Mb-best tree		801 (419)	152 (106)	948 (534)	1901 (1059)
Mb-best-modified tree		803 (421)	147 (101)	941 (527)	1891 (1049)
Cyt *c* best tree		813 (431)	143 (97)	955 (541)	1911 (1069)
Cyt *c*-2nd best tree		814 (432)	144 (98)	950 (536)	1908 (1066)
α-Hb-best tree		838 (456)	155 (109)	940 (526)	1933 (1091)

examined and the number of nucleotide substitutions required of each determined. The distribution of these trees by the number of substitutions required is shown for the myoglobins in Fig. 4. The range is from 387–460. The extreme trimodal shape of this curve is unlike anything we have observed before (see, for example, Fitch, 1979). This probably reflects the fact that certain branching patterns have a particularly drastic affect on the number of substitutions but we have so far been unable to specify those patterns in a general way.

Because the all-possible program can miss an occasional nucleotide substitution, the lengths of the better trees from that program were verified with a more powerful program and any necessary changes in the substitution counts were made in all reported data except those in Fig. 4. That figure shows fewer counts because all positions that add precisely one nucleotide substitution regardless of the tree topology have been removed.

It is instructive to note that 10,395 alternative phylogenetic hypotheses are compressed into a range of only 73 nucleotide substitutions. Thus, the practice of only revealing the best tree hides that statistically unpleasant reality that there may be 1 or more trees only a few substitutions less parsimonious. Since there is no agreed upon way of estimating the likelihood of the most parsimonious tree being true relative to that of a less parsimonious tree, it behooves one to be cautious in interpreting the results. It should be noted that a similar caveat is in order for all the other methods and all other types of data, including those that led to the trees we currently believe to be the best estimates of evolutionary history.

The most parsimonious tree for each protein is shown in Table II along with the tree currently considered best in biological opinion. Each tree is, for simplicity shown unrooted (*i.e.* in this case, without indication of where the invertebrates should be attached) since the location of the root does not affect the number of nucleotide substitutions required. For each parsimonious tree, the *dashed lines* indicate species whose sequence was not available for the particular protein but, since not all the taxa were the same for all 3 proteins, whose location is important since the total tree represented is the one used when assessing how well the other 2 proteins fit the best phylogeny of the one. The lengths of the lines are of no significance, only their branching order. The *numbers not in parentheses* are the minimum number of nucleotide substitutions required for all sequences, not just the 8 used to examine the 10,395 possible trees for the major groups. The *numbers in parentheses* are the preceding number of nucleotide substitutions less those required to form the ancestral sequences that reduce the set to 8 taxa.

The most parsimonious myoglobin tree required 12 fewer nucleotide substitutions than the biological tree using these sequences. It requires 801 nucleotide substitutions and is shown in greater detail in Fig. 5 (*top*). If one interchanges the alligator and lizard, the resulting tree also requires only 801 nucleotide substitutions. The presumably somewhat more biologically realistic joining of the alligator to the lizard to make them a sister group of the turtle costs four extra nucleotide substitutions. It is interesting that the biological tree (Fig. 5, *bottom*) has a much greater disparity in evolutionary rates than does the "best" tree. Depending upon your ideological preferences in the neutral evolution controversy, this might lend support to the belief that there are some historical realities more accurately reflected in the upper tree than in the lower tree.

There is a second myoglobin tree that differs only in the location of lamprey hemoglobin. Romero-Herrera *et al.* (1979) have isolated a myoglobin from the lamprey, proving that its hemoglobin is paralogous and should only be used as an

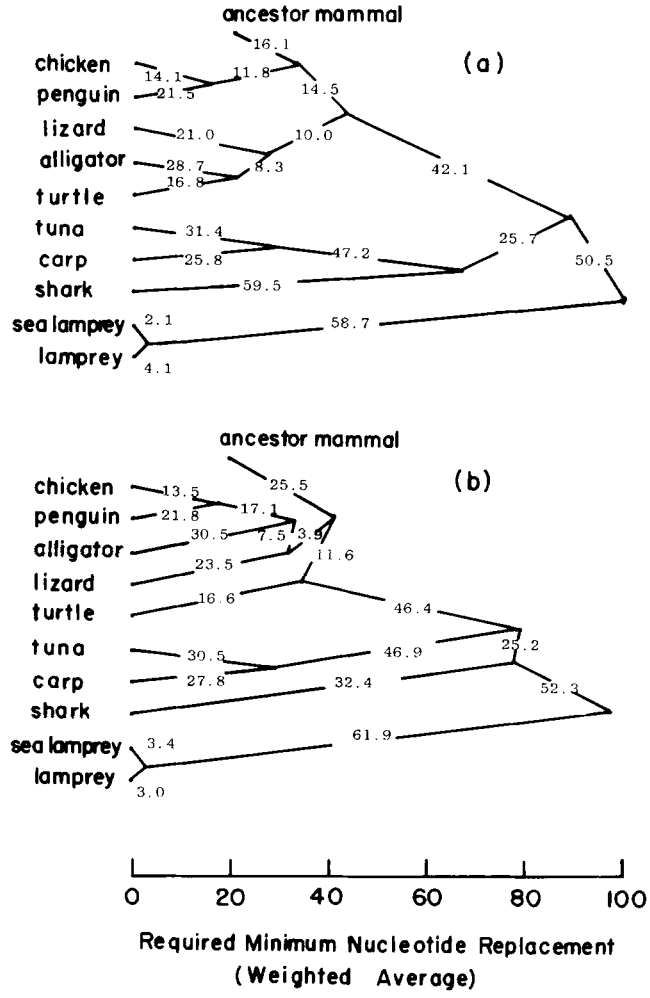


FIG. 5. **Examples of phylogeny for lower vertebrate myoglobins.** The most parsimonious tree (*a*) and the tree that reflects biological opinion (*b*) are shown. Branching of mammalian part was fixed according to biological opinions, and only the ancestor position is given in each part. *Numbers* on legs are numbers of nucleotide substitutions required to account for the descent from the ancestor and its immediate descendant. The total number of substitutions for *a* and *b* are 801 and 813, respectively.

outside marker and should not be permitted to affect the phylogeny of the orthologous myoglobins. In such a case, the lamprey hemoglobin sequence must be placed closer to the shark than to the teleost fish. This only costs two additional nucleotide substitutions but its effect is to create a tree that, compared to the best myoglobin tree, saves five nucleotide substitutions in the cytochromes *c* and seven in the α-hemoglobins, giving the lowest total for the three sequences combined among those examined. Its only differences from the biological tree are that it makes birds a sister group of mammals and turtles a sister group of alligators.

Two cytochrome trees are shown because they differ by only a single nucleotide substitution but are rather different among the reptiles. The second best tree is very like the biological tree (except for the displacement of the turtle) so that it is not surprising to find the total for this topology over all three proteins to be the same as for the biological tree. It requires, however, the mammals to be the sister group of all other reptiles while the best tree requires the mammals to be the sister group of the snakes.

The most parsimonious α-hemoglobin tree is truly bizarre, having the fish and salamanders diverging only after the snakes had separated from the sharks. This could be the result

of these sequences not all being orthologous. But, in the absence of outside evidence for an ancient α gene duplication, it is perhaps wiser to believe that α-hemoglobin genes are less reliable indicators of phylogenetic relationships (see also Fitch, 1979) than to postulate the presence of paralogous sequences. It is also noteworthy that such a vast phylogenetic change as required to convert the best hemoglobin tree into the topology of the second best myoglobin tree costs only one additional nucleotide substitution in the alpha hemoglobins.

Because the ancestral relationships suggested by the examination of protein sequences have not been in precise agreement with biological opinion, it is important to see if the different proteins keep giving the same answer. If so, we may have some confidence that the current opinion may be wrong. It is unfortunately true that these sequences do not give a self consistent set of answers and we must face the fact that, as close as protein sequences are to the genes themselves, there are important evolutionary problems that they (and perhaps nucleotide sequences as well?) may not permit us to solve.

It is of course reasonable to suppose that combining several different protein sequences into one long artificial sequence will be helpful by permitting us to average out the noise and converge upon the biological truth. It might be better, however, to recognize the nature of the fundamental problem. Current opinion is that the synapsids (which gave rise to the mammals) separated from the diapsids (which gave rise to birds and most reptiles) within perhaps 10 million years following the divergence of the anapsids (which gave rise to the turtles) in the late carboniferous nearly 300 million years ago. The recognition that two groups diverged subsequent to a third group's divergence is recognized by those character states (amino acids) that are common to the first two but not shared with the third and more distantly related outside taxa. These are the amino acid replacements that occurred in the line descending to the first 2 groups in the 10-million-year interval after the third group diverged but prior to the divergence of the first two. Now 10 million years is not such a long time that we may expect a sizeable number of such replacements. Thus, discrimination between 2 species-diverging events occurring closely in time may necessarily be based upon a limited number of differentiating events.

Confusing the recognition of these few events is the possibility that that same event may also occur in the other line as well (parallel evolution). Moreover, subsequent amino acid replacements in the same position as the differentiating events may obliterate the evidence of that event's occurrence. Finally, any event in the first line to separate that is in fact parallel to an event in 1 of the other 2 major lines will tend to mislead us into believing that these are the differentiating events and that these 2 lines have the more recent common ancestor. There have been 290 million years for such confusion to be generated and wipe out the evidence of the preceding 10 million years. It should hardly surprise us, then, that some problems are not readily resolved by sequences, protein or nucleic acid. Furthermore, while "averaging" across several proteins may reduce the noise, this effectively reduces the data to 1 large sequence and we no longer see the phylogenetic variability in the results from the different sequences. The so-called improved estimate is bought at the price of being much less certain of the degree of relative confidence one can place in alternative phylogenetic hypotheses.

Acknowledgments—We are grateful to Dr. H. F. Deutsch and to K. M. Hansen for automatic sequence assistance and to Dr. R. L. Niece and to D. Smith for technical help. We thank Dr. J. W. Warren for collecting the lizards and Drs. E. Margoliash and G. E. Tarr for providing us with the muscle extract.

REFERENCES

Bahl, O. P., and Smith, E. L. (1965) *J. Biol. Chem.* **240,** 3585–3593
Braunitzer, G., and Fujiki, H. (1969) *Nature (Lond.)* **56,** 322–323
Chan, S. K., Tulloss, I., and Margoliash, E. (1966) *Biochemistry* **5,** 2586–2597
Chan, S. K., Walasek, O. F., Barlow, G. H., and Margoliash, E. (1967) *Fed. Proc.* **26,** 723
Coates, M., Brimhall, B., Stenzel, P., Hermodson, M., Gibson, D., Jones, R. T., and Vedvick, T. (1977) *Aust. J. Biol. Sci.* **30,** 1–19
Dayhoff, M. O. (1978) *Atlas of Protein Sequence and Structure,* Vol. 5., Suppl. 3, National Biomedical Research Foundation, Washington, D. C.
Dene, H., Sazy, J., Goodman, M., and Romero-Herrera, A. E. (1980) *Biochim. Biophys. Acta* **624,** 397–408
Duguet, M., Chauvet, J.-P., and Acher, R. (1974) *FEBS Lett.* **47,** 333–337
Fisher, W. K., and Thompson, E. O. P. (1979) *Aust. J. Biol. Sci.* **32,** 277–294
Fitch, W. M. (1971) *Syst. Zool.* **20,** 406–416
Fitch, W. M. (1979) *Syst. Zool.* **28,** 375–379
Fitch, W. M., and Farris, J. S. (1974) *J. Mol. Evol.* **3,** 263–278
Fitch, W. M., and Margoliash, E. (1967) *Science* **155,** 279–284
Goldstone, A., and Smith, E. L. (1967) *J. Biol. Chem.* **242,** 4702–4710
Jarvic, E. (1963) *Q. Rev. Biol.* **38,** 26–53
Li, S. L., and Riggs, A. (1970) *J. Biol. Chem.* **245,** 6149–6169
Liljeqvist, G., Braunitzer, G., and Paléus, S. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360,** 125–135
Maeda, N., and Fitch, W. M. (1981) *J. Biol. Chem.* **256,** 4293–4300
Nash, A. R., Fisher, W. K., and Thompson, E. O. P. (1976) *Aust. J. Biol. Sci.* **29,** 73–97
Nolan, C., Fitch, W. M., Uzzell, T., Weiss, L. J., and Margoliash, E. (1973) *Biochemistry* **12,** 4052–4060
Porter, K. D. (1972) *Herpetology,* W. B. Saunders, Philadelphia
Romer, A. S. (1966) *Vertebrate Paleontology,* 3rd Ed, University of Chicago Press, Chicago
Romero-Herrera, A. E., Lieska, N., and Nasser, S. (1979) *J. Mol. Evol.* **14,** 259–266

Amino Acid Sequence of a Myoglobin from Lace Monitor Lizard, *Varanus varius*, and Its Evolutionary Implications

Nobuyo Maeda and Walter M. Fitch
Department of Physiological Chemistry
University of Wisconsin
Madison, Wisconsin 53706

Tryptic Digestion of S-Pyridylethylated Apomyoglobin

Pyridylethylated protein (35 mg) was suspended in 3 ml of 0.05M-tris/HCl buffer, pH 8.0. Upon incubation with trypsin (0.6 mg) at 37°C for 16h, the initial precipitates disappeared. The digest was applied on a Sephadex G-25 column (fine grade, 1.5 cm x 110 cm) which was connected to a second column (1.5 cm x 97 cm) and eluted with 0.05M ammonium bicarbonate (Fig. 1). The effluent was separated into five fractions, TI-TV, and freeze dried separately.

Amino acid composition of fraction I (Table 1) indicated that the fraction was not a pure peptide. Eight cycles of Edman degredation showed that this is a mixture of at least three peptides whose N-termini were positions 80, 99 and 103 of apomyoglobin. Fraction I was dissolved in 0.025M phosphoric acid and applied on the phosphocellulose column (1.5 cm x 6.8 cm). The peptides adsorbed so tightly or precipitated, that even elution with 0.025M phosphoric acid containing 1M KCl didn't recover the peptide from the column.

Fraction TII was dissolved in 0.025M phosphuric acid, applied on a phosphocellulose column (1.5 cm x 6.8 cm) and eluted with 0.025M phosphoric acid raising KCl concentration in the eluent gradually (Fig. 2). Three peptides IIa, IIb and IIc were separated. The digestion of each peptide (approx. 0.5 μmole) with Staphylococcus aureus protease V8 (85 μg) was carried out in 1% (w/v) ammonium bicarbonate (0.2 ml) for 16h at 37°C. The components of each digest were separated by thin layer chromatography (IIa) or thin layer electrophoresis (IIb and IIc). The amino acid compositions of the resulting peptides together with those of TIIa, TIIb, TIIc are given in Table 1. In peptide IIa, peptide bond of Glu-Val was cleaved effectively but Glu-Pro was not digested at all. Also in peptide IIc, Glu-Ala was very poorly digested compared to Glu-Ile.

Fraction TIII was dissolved in 0.01M ammonium bicarbonate and chromatographed on a DEAE cellulose, DE-52, column (1.5 cm x 8 cm) raising the concentration of ammonium bicarbonate linearly (Fig. 3). Fraction IIIf was a pure peptide. Fractions IIIa-IIIe were separated into their components by thin layer chromatography. IIIa-1 was a mixture of three peptides, one of which (IIIa-1-I1) was separated by thin layer electrophoresis but the other two remained together in IIIa-1-I. Each of the first three cycles of Edman degradation of peptide mixture TIIIa-1-I gave two PTH-amino acids. As the N-terminal sequence of peptide TIIIa-1-I1 indicated that the peptide had one extra lysine residue at its N-terminus compared to one of the peptides in TIIIa-1-I, the sequence of the other peptide was deduced to be Ile-Pro-Ile-(Lys). The amino acid composition of the mixture was explained by these two peptides (Table 2).

The components of fraction TIV were separated by thin layer electrophoresis followed by thin layer chromatography. Fraction IV gave three peptides by thin layer chromatography. The amino acid compositions of these peptides are given in Table 3. Peptides TIV2c and TIV5a are equally composed of one each glutamic acid (or glutamine) and lysine. The ratio of recovery of these peptide was 3 to 1. Electrophoretic mobility of TIV5a (Rarg=0.75) suggested the sequence of Gln-Lys and that of TIV2c as pyroGlu-Lys.

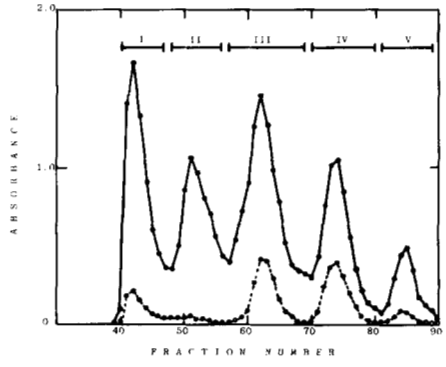


Fig. 1. Gel filtration of tryptic digest of pyridylethylated apomyoglobin.

The digest was applied to a column (1.5 cm x 110 cm) of Sephadex G-25, fine grade which was connected in series to a second column (1.5 cm x 99 cm), and eluted with 0.05M ammonium bicarbonate. Each 4 ml of fraction was collected at a flow rate of 16.5 ml/h. Elution was monitored by measuring the absorption at 230 nm (●—●) and 280 nm (○···○). Five fractions I-V were collected and freeze-dried.

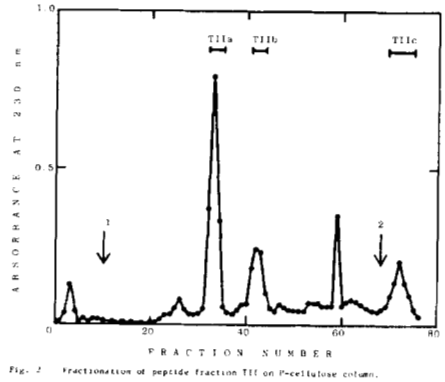


Fig. 2. Fractionation of peptide fraction TII on P-cellulose column.

Fraction TII (Fig. 1) was dissolved in 0.025M phosphoric acid and applied to a phospho-cellulose column (1.5 x 6.8 cm) equilibrated with the same solution. At arrow 1, a linear gradient elution with increasing KCl concentration was begun with 125 ml of 0.025M phosphoric acid in the mixing chamber and 125 ml of the same solution containing 0.3M KCl in the reservoir.

At arrow 2, the concentration of KCl in the eluent was changed abruptly to 0.6M. Each 4 ml fraction was collected at a flow rate of 36 ml/h.

Table 1. Amino Acid Composition of Tryptic Peptides – TIIa, TIIb, TIIc and Their Fragments by Staphylococcus aureus Protease V8 Digestion.

Values in parentheses are obtained from sequence study.

	TI	TIIa	TIIa-1	TIIa-2	TIIb	TIIb-1	TIIb-2	TIIc	TIIc-1	TIIc-2
D	0.77	1.06 (1)		1.00*(1)	2.03 (2)		2.00*(2)	1.12 (1)	1.00*(1)	
T	0.77							2.50 (2)	1.83 (2)	
S	0.41	0.88 (1)		0.93 (1)	1.61 (2)		1.82 (2)			
E	3.19	2.85 (3)		2.83 (3)	2.00 (2)		2.04 (2)	1.52 (3)	0.99 (1)	2.00*(2)
P	1.44	1.80 (2)		2.40 (2)				0.98 (1)	0.87 (1)	
G	1.41	1.00 (1)		1.01 (1)	0.98 (1)		0.91 (1)	1.00 (1)		0.86 (1)
A	2.37				2.93 (3)	1.00* (1)	2.24 (2)	4.08 (4)	2.80 (3)	1.65 (1)
V	1.50	1.58 (2)	0.58 (1)	1.03 (1)						
M					0.47 (1)	0.98 (1)				
I	2.79	1.38 (2)	1.30 (2)					1.01 (2)	0.93 (2)	
L	1.36	1.05 (1)		0.98 (1)				1.07 (1)	0.92 (2)	
Y	0.80									
F	1.12				0.90 (1)		0.96 (1)			
K	1.66							1.08 (1)	1.11 (1)	
H	1.50	0.97 (1)		0.98 (1)	0.95 (1)		0.96 (1)	1.71 (2)	2.17 (2)	2.09 (2)
R		1.03 (1)	1.00*(1)		0.98 (1)	1.00 (1)				
P&C W		trace								
Total		15	4	11	14	3	11	19	13	6
Yield** (%)		90%	66%	78%	53%	47%	39%	40%	40%	24%
Residue Number		17-31	28-31	17-27	119-132	130-132	114-124	90-98	86-98	80-85
TEP (Rarg)***		0.27	0.61	0	0.27	0.69	0	0.79	0.77	0.21
TLC (Rf)		0.67	0.82	0.52	0.30	0.59	0.27			

*The amount of these amino acids was taken as the reference value.

**Recovery of each tryptic peptide is given as the ratio to pyridylethylated apomyoglobin and of S. a. protease digested peptides to each tryptic peptide.

***Mobilities of peptides on thin layer electrophoresis are given relative to the mobility of arginine (1.00). Under this condition mobilities of glycine and glutamic acid are 0.27 and -0.54 respectively.

Table 2. Amino Acid Composition of Tryptic Peptides Isolated from Fraction TIV. Results are expressed by molar ratio of amino acids.

	IIIa-1-I	IIIa-1-I1	IIIa-2	IIIb-1	IIIb2	IIIb3	IIIc-1	IIIc-2	IIId	IIIe-1	IIIf
D			0.89 (1)		2.00*(2)	2.00*(2)	1.00*(1)	1.93 (2)	1.96 (2)		1.10 (1)
T	2.72 (3)	2.73 (3)					1.00 (1)		1.19 (1)		
S					0.90 (1)	0.95 (1)		0.89 (1)			0.90 (1)
E				1.00*(1)		1.91 (1)	1.65 (1)	1.22 (1)	3.09*(3)	2.00*(2)	2.01 (2)
P	[0.96.....(1)]#								1.14 (1)		
G	2.00 (2)	2.00*(2)								1.41 (2)	1.00*(1)
A	1.21 (1)	1.09 (1)		1.00 (1)	0.98 (1)						
V	1.06 (1)	0.97 (1)									
M					0.98 (1)		1.08 (1)		0.98 (1)		
I	[2.40.....(2)]#										
L	2.85 (2)	1.13 (1)	1.00*(1)	2.20 (2)		1.21 (1)	0.99 (1)	1.00*(1)		1.07 (1)	1.07 (1)
Y											
F				0.90 (1)					1.03 (1)	1.04 (1)	
K	[2.03 (1).....(3)]#		1.10 (1)			2.29 (2)	1.02 (1)	1.10 (1)			1.02 (1)
H	0.95 (1)	0.89 (1)							1.06 (1)		
R	1.00 (1)	1.07 (1)		1.00 (1)	0.93 (1)				1.19 (1)		
P&C W											(1)
Total	12 4	11	3	6	6	7	6	6	11	6	4
Yield(%)***	(10) (15)	13	11	19	23	10	22	26	43	23	51
Residue	63-74 99-102	94-74	48-50	134-139	140-145	57-63	51-56	57-62	32-42	148-153	1-4
TEP (Rarg)	0.45 0.45	0.38	0.65	0.33	0.32					0.06	
TLC (Rf)	0.85 0.85	0.85	0.71	0.93	0.57	0.37	0.65	0.48	0.67	0.82	0.67
Same peptide (yield)			IV5b (26)	IV2g (26)	(10)				IIIe-2 (3)	IV-1 (3)	

*The amount of these amino acid residues was taken as the reference value.

**The tryptophan was detected by spot test with Ehrlich reagent.

***Recovery of each peptide is given by percentage to the pyridylethylated apomyoglobin.

#IIIa-1-I was a mixture of two peptides (see text).

Table 3. Amino Acid Composition of Tryptic Peptides (Fraction TIV, TV). Results are expressed by molar ratio of amino acids.

	IV2c	IV2e	IV2d	IV3	IV3c	IV3d	IV3	V1	V2
D			1.00*(1)	1.00*(1)					
T									
S									
E	1.00*(1)	2.00*(2)							
P									
G		1.48 (2)	1.08 (1)	1.26 (1)					
A					0.47 (1)				
V			1.41 (2)	1.83 (2)					
M									
I			0.96 (1)	0.92 (1)		1.05 (1)			
L		1.06 (1)				0.82 (1)			
Y		0.91 (1)							0.52 (1)
F		1.04 (1)			1.00*(1)				
K	0.90 (1)	1.19 (1)	1.09 (1)	2.38 (2)	1.04 (1)	1.00*(1)	1.00*(1)	0.95 (1)	1.00*(1)
H								1.00*(1)	1.00*(1)
R									
P&C W									
			(1)	(1)					
Total	4	8	7	8	1	3	1	2	2
Yield (%)	23	13	16	*	26	24	29	32	16
Residue	78-79	146-153	10-16	9-16	43-45	75-77	*	46-47	146-147
TLE (Rarg)	0.30	0.30	0.33	0.52	0.73	0.75	1.05	0.26	0.17
TLC (Rf)	0.31	0.52	0.71	0.54	0.54	0.69	0.10		
Same peptide (yield)	IV5a (2.5)						V3 (3)		

*The amount of these amino acids was taken as the reference value.

**Tryptophan was detected by spot test with Ehrlich reagents.

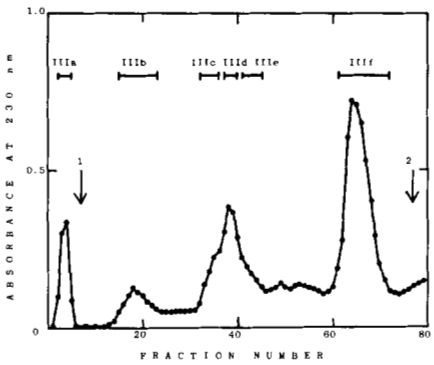


Fig. 3. Fractionation of cryptic peptide fraction TIII on DE-52 column.

Peptide fraction TIII (Fig. 1) was freeze-dried, dissolved in 0.01M ammonium bicarbonate and subjected to DE-52 column (1.5 × 8 cm) chromatography. At arrow 1, a linear gradient elution with increasing ammonium bicarbonate concentration was begun with 150 ml of 0.01M ammonium bicarbonate in the mixing chamber and 150 ml of 0.3M ammonium bicarbonate in the reservoir. At arrow 2, the concentration of ammonium bicarbonate was changed abruptly to 0.6M. Each 4 ml fraction was collected at a flow rate of 35 ml/h.

Cyanogen bromide cleavage of pyridylethylated myoglobin.

Pyridylethylated Mb (approx. 100 mg) was dissolved in 70% w/w) formic acid (10 ml) and incubated with CNBr (150 mg) at room temperature for 16h. The reaction mixture was diluted with 90 ml of water and freeze-dried. The freeze-drying was repeated after the addition of 0.2M acetic acid (20 ml). The material was chromatographed on a Sephadex G-50 column (1.8 cm x 93 cm) in 0.5M-acetic acid. The eluate was separated into four functions, I-IV (Fig. 4). Fraction I was a single peptide of CNBrI, although the elution pattern seldom showed a fine peak shape. Fraction II was evaporated to dryness, dissolved in 0.5M acetic acid (4 ml) and rechromatographed on the same column. Fraction III was freeze-dried and a portion (5 mg) of dried material was dissolved in 0.02M ammonium bicarbonate and applied on a DLAE cellulose column (1.2 cm x 12 cm) and separated into three peptides III(-1, III-2, and III-3, of which III-1 and III-2 gave the same amino acid composition. The separation was caused by the difference between C-terminal homoserine and homoserine lactone (Fig. 5). Fraction IV was evaporated to dryness and chromatographed on Sephadex G-25, fine grade, column (1.5 cm × 110 cm + 1.6 cm x 97 cm) in 0.05M ammonium bicarbonate. Four peptides, III-3, III-2 (and III-1), IV-1 and IV-2 were separated from each other (Fig. 6). The amino acid compositions of resulting peptides are given in Table 4. Peptide II is composed of peptides III-3 and III-2 indicating incomplete cleavage at Met-32.

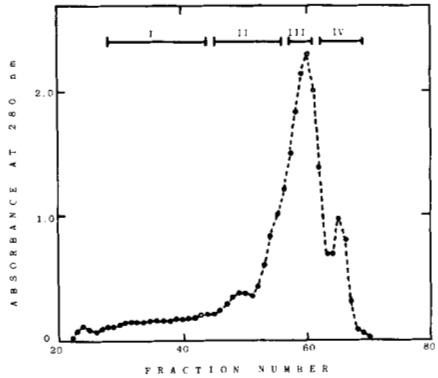


Fig. 4. Gel filtration of CNBr-cleaved peptides of pyridylethylated apomyoglobin.

The peptide mixture from CNBr cleavage was applied on a column (1.8 cm x 91 cm) of Sephadex G-50, fine grade, and eluted with 0.5M acetic acid. Each 3.5 ml fraction was collected at a flow rate of 25 ml/h.
●---● A_{280}.

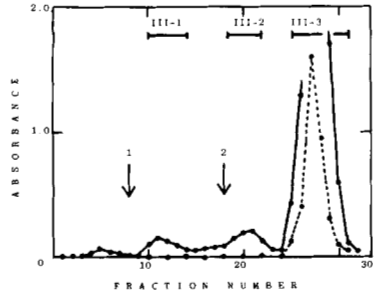


Fig. 5. Further purification of CNBr cleavage fraction CNBrIII on DE-52.

Fraction CNBrIII (Fig. 4) was freeze-dried and 5 mg of dried material was dissolved in 0.025M ammonium bicarbonate and applied to a column (1.5 cm 8 cm) of DE-52. Concentration of the ammonium bicarbonate was raised stepwise at fraction number 9 and 18, to 0.1M and 0.3M, respectively. Elution was monitored by measuring A_{280} (●---●) and A_{230} (●——●). Each 3.6 ml fraction was collected at a flow rate of 30 ml/h.

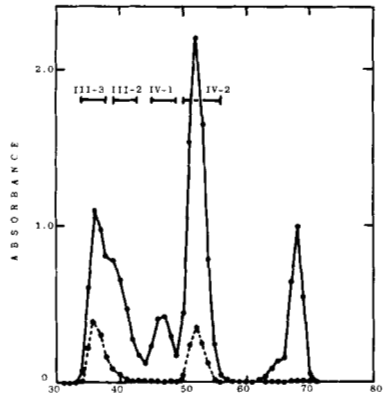


Fig. 6. Gel filtration of CNBr-cleaved peptides fraction CNBrIV on Sephadex G-25 column.

Fraction CNBrIV (Fig. 4) was freeze-dried dissolved in 0.05M ammonium bicarbonate (2.0 ml) and applied on a column (1.5 cm x 110 cm) of Sephadex G-25, fine grade, which was connected to the second column (1.5 cm x 97 cm). Elution was monitored by measuring A_{230} (●——●) and A_{280} (●---●). Amino acid composition of the first two peaks are the same as those of III-3 and III-2 (III-1), respectively. Each 5 ml of fraction was collected a a flow rate of 25 ml/h.

Table 4. Amino Acid Composition of Cyanogen Bromide Cleaved Peptide of Lizard PE Mb.
Results are given as molar ratio of amino acids.

	I	II	III(1) III(2)	III-3	IV-1	IV-2
Asp	5.00* (5)	7.00* (7)	4.00* (4)	3.00* (3)	2.00* (2)	
Thr	3.86 (3)	1.80 (2)	2.05 (2)			
Ser***	2.99 (3)	1.57 (2)		1.74 (2)		0.98 (1)
Glu	9.68 (10)	10.60 (9)	4.69 (4)	5.13 (5)	1.45 (1)	2.10 (2)
Pro	1.27 (2)	2.40 (3)	0.80 (1)	1.84 (2)		
Gly	4.19 (5)	2.63 (3)		3.12 (3)		2.00* (2)
Ala	8.50**,# (9)	1.34 (1)	1.10 (1)		1.06 (1)	1.11 (1)
Val***	4.09 (4)	2.05 (4)		3.05 (4)		
Met	**(1)	0.30 (1)	(1)	(1)	(1)	
Ile	5.51 (7)	1.85 (3)		2.07 (3)		
Leu	5.57 (6)	4.00 (4)	2.16 (2)	1.82 (2)	1.95 (2)	1.06 (1)
Tyr	0.51 (1)					0.77 (1)
Phe#	1.73 (2)	2.87 (3)	3.00 (3)		1.04 (1)	1.05 (1)
Lys	8.09 (8)	5.68 (6)	3.72 (3)	3.37 (3)	1.51 (1)	1.05 (1)
His	5.68# (6)	2.14 (2)	1.03 (1)	1.04 (1)		
PRO##	2.06#					
Trp		1.74 (2)		1.83 (2)		
Arg	1.03 (1)	1.94 (2)	1.02 (1)	0.81 (1)	1.86 (2)	0.99 (1)
total	76	55	24	32	11	11
Yield (%)	6b	8	42	33	23	84
Residue	56-131	1-55	33-55	1-32	132-142	143-153

*The amount of these amino acids was taken as the reference value.

**The values obtained from the analysis of 72h hydrolysate.

***Homoserine and homoserine lactone coeluted with glutamic acid and lysine respectively.

#Buffer change overlapped and gave larger pyridylethyl cysteine amount.

##Tryptophan contents were determined by 3N-mercaptoethene sulphonate hydrolysis.

Peptic digestion of CNBrI

Peptide CNBrI (22.2 mg) was dissolved in 0.01N HCl (1.0 ml) and digested with pepsin at an enzyme/substrate ratio of 1:100 (w/w) at 24°C for 30 min.

The digests were applied on Sephadex G-25, fine grade, column (1.5 x 110 + 1.5 x 95) and eluted with 0.05M ammonium bicarbonate (Fig. 7).

The fraction PI was freeze-dried and dissolved in 0.025M ammonium bicarbonate and separated into three components, PIa, PIb and PIc, by DEAE cellulose, DE52, column (1.5 cm x 8 cm) chromatography (Fig. 8). Fraction PII was a pure peptide. PV was two tri-peptides one of which was purified further by TLC (Rf 0.5). PIV was considered to be a mixture of two peptides Thr-Ala-Leu and Val-Ile-Val-Gly-Val from its amino acid composition. No further purification was performed. The amino acid compositions of purified peptic peptides are given in Table 5.

Peptide PIc was not pure. Fifteen cycles of manual Edman degradation showed that PIc is contaminated with the peptide whose N-terminus is shorter by five residues than PIc. Table 6 gives the yield of amino acid recovered when PIA derivatives of each cycle were hydrized with 0.2M-NaOH at 150°C for 4h and subjected to amino acid analyses.

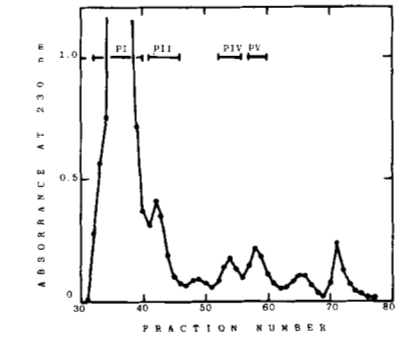


Fig. 7. Gel filtration on Sephadex G-25 of peptic digest of CNBrI.

The digest was applied on a Sephadex G-25, fine grade, column (1.5 cm x 110 cm), which was connected in series to a second column (1.5 cm x 95 cm) and eluted with 0.05M ammonium bicarbonate.

Each 5 ml fraction was collected at a flow rate of 25 ml/h.

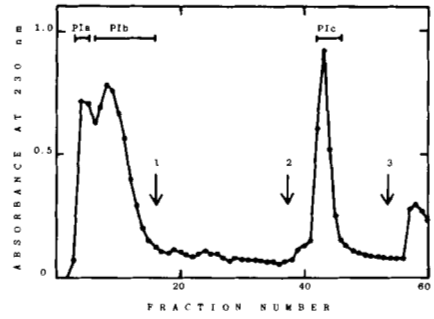


Fig. 8. Separation of peptic fraction PI on DE-52 column.

Fraction PI (Fig. 7) was freeze-dried, dissolved in 0.025M ammonium bicarbonate and applied on a column (1.5 cm x 8 cm) of DE-52 column. At the arrow, a linear gradient with increasing ammonium bicarbonate concentration was begun with 0.025M and 0.3M ammonium bicarbonate in mixing chamber and reservoir respectively (150 ml each) at fraction tube number 37 and 53, the concentration of eluent was changed stepwise to 0.3M and 1.0M respectively. Each 4 ml fraction was collected at a flow rate of 40 ml/h.

Table 5. Amino Acid Compositions of Peptic Peptides of CNBrI.

	PIa	PIb	PIc	PII	PVa
D	1.07 (1)	1.00* (1)	2.00* (2)	2.00* (2)	
T	2.67 (3)	2.80 (3)		1.75 (2)	
S			1.91 (2)**	1.05 (1)	
E	4.00* (4)	5.00 (5)	3.39 (3)	1.46 (1)	1.09 (1)
P	1.84 (2)	1.96 (2)			
G	1.92 (2)	2.01 (2)	1.76 (2)	1.19 (1)	
A	4.93 (5)	4.82 (5)	3.74 (4)	0.55	
V			1.87 (3)	1.18 (1)	
I	3.50 (4)	3.55 (4)	1.53 (2)	0.20	1.00* (1)
L	2.84 (3)	3.92 (4)		1.85 (2)	
Y	0.65 (1)	0.53 (1)			
F		0.61 (1)	1.15 (1)**	2.98 (3)	
K	4.04 (4)	4.33 (4)	1.67 (1)	1.01 (1)	
H	3.39 (4)	3.87 (4)	1.23 (1)		
PLCys					0.90 (1)
R**	0.93 (1)	0.77 (1)			
M			(1)**		
Total	34	37	22	14	3
Yield (%)	16	38	49	42	27
Residue	70-103	70-108	110-131	56-69	107-109

*The amounts of these amino acids were taken as standards.

**Homoserine and homoserine lactone coeluted with glutamic acid and lysine respectively.

Table 6. Manual Edman degradation of peptide PI_c (0.24 μmole).

Step number	Residue TLC	Yield 0.2M NaOH Hydrolysis nmole	Contaminant Residue	Yield
1	Val	104.0	Ile	33.5
2	Ile	104.0	Ala	27.2
3	Val	74.8	Glu	23.3
4	Gly	66.5	Lys	22.2
5	Val	75.6	His	9.6
6	Ile	38.2		
7	Ala	39.2		
8	Glu	39.1		
9	Lys	32.6		
10	His	25.9		
11	(Ser)	-		
12	Ala	27.1		
13	Asp	23.1		
14	Phe	18.4		
15	Gly	19.3		

Tryptic digestion of CNBrI

Amino acid composition and N-terminal sequences of CNBr peptide and tryptic peptides of apomyoglobin provided the information to align the peptides. The tryptic peptide which failed to purify from the digest of apomyoglobin (residue 103-118) and was described before was purified from the tryptic digest of peptide CNBr-I as follows.

Peptide CNBrI (13 mg) was dissolved in 1% (w/v) ammonium bicarbonate (1.0 ml) and digested with trypsin at an enzyme/substrate ratio of 1:50 (w/w) for 16h at 37°C. The digest was chromatographed on a DEAE cellulose, DE-52, column (1.5 cm 8 cm, Fig. 9). Amino acid composition (Table 7) and N-terminal sequence of 9 residues of peptide t-CNBrI-c explained residues 103-118. Peptides t-CNBrI-b and tCNBrI-d are considered to be T11c and T11b respectively from their amino acid compositions.

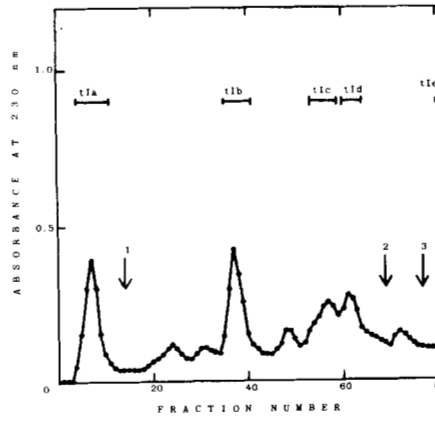


Fig. 9. Separation of tryptic peptides of CNBrI on DE-52 column.

The digest was applied to a column (1.5 cm x 8 cm) of DE-52 equilibrated with 0.02M ammonium bicarbonate. At arrow 1 a linear gradient elution with ammonium bicarbonate was begun with 0.02M and 0.3M ammonium bicarbonate in the mixing chamber and the reservoir respectively (100 ml each). At arrows 2 and 3, the ammonium bicarbonate concentration was changed abruptly to 0.3M and 1.0M respectively. Each 3 ml fraction was collected at a flow rate of 45 ml/h.

Table 7. Amino acid compositions of tryptic peptides of CNBrI.

	tCNBrI-b	tCNBrI-c	tCNBrI-d
D	1.19 (1)		2.00*(2)
T	1.59 (2)		
S			1.70 (2)
E	3.00*(3)	2.81 (3)	2.59 (3)
P	0.80 (1)		
G	0.99 (1)	1.17 (1)	1.14 (1)
A	3.73 (4)	3.91 (3)	3.15 (3)
V		1.76 (2)	
I	0.97 (1)	1.09 (2)	
L	1.44 (2)	1.00 (1)	
Y		0.38 (1)	
F		0.94 (1)	1.04 (1)
K	1.16 (1)	0.91 (1)	0.23
H	1.33 (2)		0.83 (1)
R			
PtCys		0.72 (1)	
M			(1)
Total	19	16	13
Yield (μmole)	0.35	0.16	0.49
Residue	80-98	103-118	119-131

*The amount of these amino acids was taken as the reference value.

**Homoserine and homoserine lactone eluted with glutamic acid and lysine respectively.

Table 8. Amino acid composition of α-chymotryptic peptides of CNBrI.

	α3-1-2	α3-2-3	α4-3	α4-5	α5-1-1	α5-1-2	α5-2	α3-4	α6-1	α6-2
D				1.00*(1)			2.20 (2)	0.96 (1)		1.80 (1)
T			1.83 (2)			1.21 (1)		1.96 (2)		
S				0.65 (1)			0.77 (1)			0.62 (1)
E	2.65 (3)	2.24 (2)		2.15 (2)			2.14 (2)	1.13 (1)	0.95 (1)	
P	- (1)				0.99 (1)					
G	1.00*(1)	1.00*(1)	1.34 (1)	0.76 (1)		1.00*(1)				
A	1.72 (2)	1.17 (1)		1.72 (2)		1.45 (1)		2.00*(2)		1.22 (1)
V		2.05 (2)	0.92 (1)							
I	0.70 (1)	1.85 (2)	1.07 (1)		1.45 (2)	0.81 (1)				
L	1.26 (1)					1.53 (2)	1.00*(1)		1.45 (1)	
Y					0.98 (1)					
F									1.00*(1)	1.12 (1)
K	2.55 (2)	1.96 (1)	2.26*(2)	0.41	2.66*(2)		1.22 (1)			
H	1.87 (2)	1.12 (1)	1.00 (1)					1.87 (2)		
R						0.97 (1)				
PtCys		0.84 (1)								
M				(1)**						
Total	13	13	8	8	6	7	8	6	3	4
Yield (μmole)		0.13	0.16	0.23	0.20	0.31	0.13	0.40	0.22	0.21
Residue	77-89	107-119	92-69	124-131	98-103	70-76	96-81	90-93*** (96-97)	104-106	120-121
TLE (Karg)	0.18	0.09								
TLE (Rf)	0.27	0	0.64	0.34	0.69	0.69	0.58	0.27	0.76	0.45

*The amount of these amino acids was taken as the reference value.

**Homoserine and homoserine lactone eluted with glutamic acid and lysine respectively.

***See text.

α-chymotryptic Digestion of CNBrI

Peptide CNBrI (Ca 10 mg) was dissolved in 0.05M-tris/HCl buffer, pH 7.8, (10 ml) and incubated with 0.1 mg of α-chymotrypsin for 16h at 37°C. The digest was applied on a Sephadex G-25 column (fine grade, 1.3 cm x 116 cm) and eluted with 0.02M ammonium bicarbonate (Fig. 10). Peptides were further purified by thin layer chromatography and by thin layer electrophoresis. The amino acid compositions of the peptides that explain that of CNBrI are given in Table 6. Fraction α-5-4 was subjected to 5 cycles of Edman degradation. At the second cycle both PTH-Gln and PTH-Asn were identified on t.l.c. and nothing was found at the fifth cycle. These facts indicate that α-5-4 was a mixture of two extra peptides, Ala-Gln-Thr-His and Ala-Asn-Thr-His, produced by a digestion at C-terminal site of histidine residues.

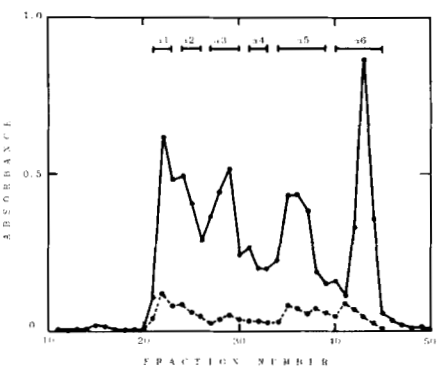


Fig. 10. Gel filtration of α-chymotryptic digest of CNBrI on Sephadex G-25 column.

The digest was applied to a column (1.3 cm x 116 cm) of Sephadex G-25, fine grade, and eluted in 0.02M ammonium bicarbonate. Elution was monitored by measuring A_{210} (●——●) and A_{280} (●---●). Each 1.6 ml fraction was collected at a flow rate of 12 ml/h.

Sequence Analysis

Apomyoglobin (0.2 μmole), peptide CNBrI (0.3 μmole) and peptide CNBrIII-2 were subjected to automatic sequence analysis on Beckman sequencer 890B. Polybrene (0.2 mg) was added in the reaction cup to retain peptide. Quadrol program was used for apomyoglobin (20 cycles) and CNBrI (30 cycles) and a dimethylallylamine program was for CNBrIII-2 (20 cycles). Table 9 summarizes the recovery of the amino acid residue of each cycle after hydrolysis with 6M HCl/0.1% (w/v) $SnCl_2$ at 150°C for 4h.

Table 9. Automatic Sequence Analysis.

Nanomoles of each amino acid recovered was calculated as the difference between the nanomoles analyzed on a given cycle and the number of nanomoles for that particular residue on the previous cycle.

Cycle Number	Apomyoglobin (0.2 μmole) Residue	Recovery (nmole)	TLC	CNBrI (0.3 μmole) Residue	Recovery (nmole)	TLC	CNBrIII-2 (0.2 μmole) Residue	Recovery (nmole)
1	Gly	115.4	Gly	Lys	90.3		Phe	37.4
2	Leu	102.2	Ile/Leu	Asp	111.7	Asn	Glu	102.6
3	"	"	Ser	"	"		Asp	93.4
4	Asp	79.7	Asp	Glu	107.5	Glu	His	51.9
5	Glu	84.3	Glu	Asp	98.7	Asp	Pro	102.6
6	Glu	60.4	Glu	Leu	82.7		Glu	91.7
7	"	"	Trp	Lys	72.0		Thr	34.4
8	Lys	23.7	Lys	Lys	67.3		Glu	20.4
9	Lys	49.0	Lys	His	54.6		Asp	68.4
10	Val	53.0	Val	Gly	46.2		Arg	14.8
11	Val	74.1	Val	Thr	Trace		Phe	51.1
12	Asp	42.7	Asp	Thr	Trace		Ala	35.1
13	Ile	28.0	Ile/Leu	Val	17.1		Lys	34.1
14	"	"	Trp	Leu	9.1		Phe	17.7
15	Gly	43.2	Gly	Thr	Trace		Lys	22.0
16	Lys	12.0	Lys	Ala	7.6		Asp	16.9
17	Val	16.0	Val	Leu	4.7		Leu	24.1
18	Glu	28.2	Glu	Gly	22.7		Lys	15.9
19	Pro	26.1	Pro	Arg	6.6		Thr	9.9
20	Asp	16.8	Asp	Ile	4.0		Leu	14.5
21				Leu	9.2			
22				Lys	2.5			
23				Glu	11.0			
24				Lys	4.3			
25				Glu	11.1			
26				His	2.9			
27				His	2.5			
28				Glu	5.8			
29				Ala	5.6			
30				Glu	3.1			
Program		Quadrol			Quadrol			Dimethylallylamine
Initial Yield		57%						
Repetitive Yield	R_{Gly}: 92%				R_{Lys}: 91.5%			R_{Asp}: 94.1%