J. Mol. Evol. 10, 167–182 (1977)

Journal of Molecular Evolution

Molecular Evolution of Snake Venom Toxins

T.H. Hseu[1] *, E.D. Jou[2] **, C. Wang[2] *, and C.C. Yang[1]**

[1] Institue of Molecular Biology and
[2] Department of Physics, National Tsing Hua University, Hsinchu, Taiwan, Republic of China

Summary. Phylogenetic trees were constructed for 62 venom toxins of snakes of *Proteroglyphae* suborder using matrix method. The resulting tree from *Minimum Spanning Tree-Cluster Analysis* technique had the lowest "percent deviation" (8.55). The taxonomic relationship of these toxins agrees very well with zoological opinions. However, the appearance of the tree did not directly provide a plausible evolutionary model for the toxins. A model was derived from nodal ancestral sequence calculations, comparisons between intra- and inter-generical rates of amino acid change, and generally held ideas about protein evolution. According to the model, short neurotoxin is the ancient form of snake venom toxins. The courses of evolution leading to the present intraspecific homologous toxins are explained by gene duplication and allelomorphism.

Key words: Molecular evolution – Phylogenetics – Numerical taxonomy – Snake venom toxins – Elapidae – Hydrophiidae

In the past decade intensive studies have been made of the snake venom toxins. To date there have been over 62 amino acid sequences determined from venoms of *Proteroglyphae* (Fig. 1). They are all small basic polypeptides with different pharmacological activities, namely neurotoxins, cardiotoxins, and so-called angusticeps-type toxins of unknown pharmacological activity. All the neurotoxins whose amino acid sequences have been determined are classified into two groups. Although the two types of neurotoxins are pharmacologically and structurally related (the short neurotoxins consist of 60–62 amino acid residues with four disulfide bonds; the long neurotoxins have 71–74 amino acid residues with five disulfide bridges), their immunochemical properties are completely different (Botes, 1972). Homologous cardiotoxins (60–61 amino acid residues with four disulfide bridges) are serologically distinct from both types of neurotoxins (Viljoen and Botes, 1973).

In the absence of paleontological records, knowledge of the classification, the origin, and subsequent evolution of the venomous snakes developed years of effort by morphologists, and though broadly understood, has not yet attained unequivocal

* To whom correspondence should be addressed
** Present address: Department of Computer Sciences, University of Texas, Austin, Texas, USA
*** Present address: Biological Division, California Institute of Technology, Pasadena, California, USA

agreement. The difficulties arise from the limitations of the comparative method and the high degree of specialization of these snakes (Johnson, 1956). However, where morphological changes accompanying speciation have been few, molecular evidence is useful in helping unravel relationships of closely related species. One of the more informative molecular approaches is the construction of phylogenetic trees using amino acid sequences.

Studies on phylogenetic relationships of snake venom toxins have previously been attempted. Based on trees from 11 (Strydom, 1972a), 16 (Strydom, 1973) and 43 (Strydom, 1974) sequences, Strydom has subsequently altered his initial conclusion that a cardiotoxin-like structure was the ancient form of the snake toxins to his recent

Genus	Species	Common Name	Origin	ID.NO.	Toxins	
Family: Elapidae						
Naja	N. naja atra	Taiwan cobra	Taiwan	1	Cobrotoxin	(62-4)
				47	Cardiotoxin	(60-4)
				53	Cardiotoxin I	(60-4)
	N. nigricollis	Black neck spitting cobra	S. Africa	3	Toxin α	(61-4)
	(N. mossambica pallida)*			43	Toxin γ (F14)	(60-4)
	N. mossambica mossambica			44	V^{II}_1	(60-4)
				45	V^{II}_2	(60-4)
				46	V^{II}_3	(60-4)
				52	V^{II}_4	(60-4)
	N. haje haje	Egyptian cobra	Nile valley, Egypt	2	Toxin α	(61-4)
	(N. haje annulifera)*			56	V^{II}_1	(60-4)
	N. melanoleuca	Forest cobra	S. Africa	7	Toxin d	(61-4)
				28	Toxin b	(71-5)
				38	Toxin 3,9,4	(71-5)
				55	V^{II}_1	(60-4)
				58	V^{II}_2	(61-4)
				59	V^{II}_3	(61-4)
	N. nivea	Cape cobra	S. Africa	2a	Toxin δ	(61-4)
				8	Toxin β	(61-4)
				29	Toxin α	(71-5)
	N. naja siamensis	Monocellate Thai cobra	Thailand	21	Toxin 3	(71-5)
	N. naja	Indian cobra	India	25	Toxin A	(71-5)
				26	Toxin B	(71-5)
				27	Toxin C	(71-5)
				48	Cytotoxin II	(60-4)
				49	Cytotoxin IIa	(60-4)
				50	CM-XI	(60-4)
				51	Cytotoxin I	(60-4)
	N. naja naja	Indian spectacle cobra	India	23	Toxin 3	(71-5)
				24	Toxin 4	(71-5)
	N. naja naja	Black cobra	W. pakistan	22	Toxin 3	(71-5)
	N. naja	Cambodian cobra	Cambodia	54	Cardiotoxin (F8)	(60-4)
	N. naja oxiana		Iran	6	Oxiana α	(61-4)
				34	Neurotoxin I	(72-5)
Ophiophagus	O. hannah	King cobra	Thailand	35	Toxin a	(73-5)
				36	Toxin b	(73-5)
Hemachatus	H. hemachatus	Ringals	S. Africa	4	Toxin II	(61-4)
				5	Toxin IV	(61-4)
				57	DLF (F12b)	(61-4)
Dendroaspis	D. polylepis polylepis	Black mamba	E. Africa	9	Toxin α	(60-4)
				31	Toxin γ	(72-5)
				32	Toxin δ	(72-5)
	D. jamesonii kaimosae	Tropical mamba	Africa	10	V^{I}_{n1}	(60-4)
				30	V^{III}_{n1}	(72-5)
	D. viridis			11	Toxin 4.11.3	(60-4)
				33	4.7.3 & 4.9.3	(72-5)
				42	Toxin 4.9.6	(60-4)
	D. angusticeps			40	Toxin FvII	(61-4)
				41	Toxin Ta2	(60-4)
Bungarus	B. multicinctus	Banded Krait	Taiwan	37	α-bungarotoxin	(74-5)
Family: Hydrophiidae						
Laticauda	L. semifasciata	Erabu-umihebi	Amami Island, Japan	16	Erabutoxin a	(62-4)
				17	Erabutoxin b	(62-4)
				18	Erabutoxin c	(62-4)
				39	Ls III	(66-5)
	L. laticaudata	Hiroo-umihebi	Amami Island, Japan	19a	Laticotoxin a	(62-4)
				20	Laticotoxin b	(62-4)
	L. colubrina	Aomadara-umihebi	Amami Island, Japan	19	Laticotoxin a	(62-4)
Enhydrina	E. schistosa	Common-sea snake	Penang Island	12	Toxin 4	(60-4)
				13	Toxin 5	(60-4)
Hydrophis	H. cyanocinctus	Annulated sea snake	Taiwan	13a	Hydrophitoxin b	(60-4)
				15	Hydrophitoxin a	(60-4)
Pelamis	P. platurus		Taiwan	14	Pelamitoxin a	(60-4)

Fig. 1. Toxins isolated from snake venoms of *Proteroglyphae* suborder. Only those with amino acid sequences determined are listed. References to sequence data are given in Figure 2. A (*) indicates Broadly's proposed reclassification of *Naja* species

recognition of two mutually exclusive alternative evolutionary pathways in the course of snake toxin evolution. Chang (1972) suggested that the complicated appearance of the phylogenetic tree constructed from 19 sequences is imperfection rather than anomaly.

The difficulties encountered in phylogenetic studies on snake venom toxins are different from many of those encountered by investigators of cytochrome c and hemoglobins. It is known that multiple toxin varieties are present in the venom of a single individual snake, and only some of them have been sequenced. Moreover, the number of toxin varieties may change from species to species. Furthermore, sequence alignment for all these toxins (Fig. 2) requires the assumption of a large number of deletions or insertions. The mutation distances, calculated in terms of minimum number of mutation required (Fitch, 1966), may not provide a sufficiently close picture of the evolutionary history of the gene loci encoding for the toxins. In addition, the study of toxin protein evolution is an attempt made to examine the phylogenetic relationships of a set of homologous proteins all below a suborder level.

All of the above complexities indicate snake venom toxins are rather unique proteins for the study of protein evolution. The present study attempted to understand the implication of the above distinct characteristics for toxin phylogeny and to provide some insights into the origin and evolution of toxin molecules and species containing them.

Phylogenetic Tree Construction

The matrix method (Fitch and Margoliash, 1967; Hartigan, 1973; Moore et al., 1973) was chosen primarily on account of its general applicability and simplicity of idea. It has been shown (Peacock and Boulter, 1975), using a computer simulation, that when dissimilarity among sequences became greater the results of the method were slightly more accurate than those obtained by the ancestral sequence method (Dayhoff and Eck, 1966).

Three approaches were used. All were based on a distance matrix calculated from Minimum Number of Multations Required (MNMR), in terms of nucleotide changes, to convert the codon for one amino acid into that for another (Fitch, 1966). The MNMR for amino acid pairs were taken from Table 1 of Fitch and Margoliash (1967), except that the values for amino acid pairs involving isoleucine with glutamine, glutamic acid and lysine are one less than in that table. Deletions which produce the gaps, represented by —'s, were treated as if they were a 21st amino acid. Their MNMR in pairing with the rest of the 20 amino acids was assigned a value of 1 by the assumption of simultaneous appearance or disappearance of all three nucleotides encoding an amino acid.

The first approach used was the method of Fitch and Margoliash (1967) (FM). A slightly different algorithm was employed (Jou, 1975). However, extensive search for a best tree was not attempted. Another approach is what we called Prelimset Ancestral Sequence (PAS) procedure. The tree building algorithm was the same as that used in FM procedure. The principal difference is in the calculation of averaged values for those elements of the distance matrix affected when two subsets are joined into a nodal

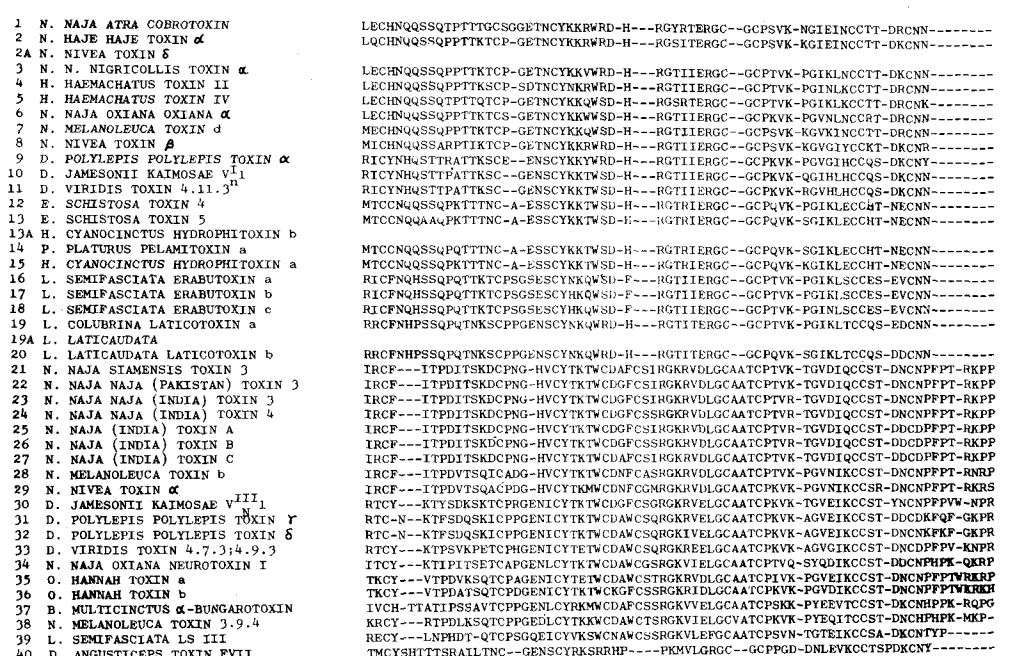


Fig. 2. Amino acid sequence alignment chart for snake venom toxins. The one-letter IUPAC-IUB codes for amino acid residues are used. The number in front of a toxin name is the identification (ID) number of the toxin as given in Figure 1. The sequence data shown are from: *1* (Yang et al., 1969); *2* (Botes and Strydom, 1969); *2a* (Botes et al, 1971); *3* (Eaker and Porath, 1967); *2a* (Botes et al., 1971); *3* (Eaker and Porath, 1967); *4, 5* (Strydom and Botes, 1971); *6* (Arnberg et al., 1974; Grishin et al., 1973); *7* (Botes, 1972); *8* (Botes, 1971); *9* (Strydom, 1972); *10* (Strydom, 1973); *11* (Banks et al., 1974); *12* (Fryklund et al., 1972); *13* (Sato, 1974); *13a* (Liu et al., 1973); *14* (Wang et al., 1976); *15* (Liu and Blackwell, 1974); *16, 17* (Sato and Tamiya, 1971); *18* (Tamiya and Abe, 1972); *19* (Sato and Tamiya, 1971); *19a, 20* (Sato, 1974); *21* (Karlsson et al., 1972); *22, 23, 24* (Karlsson, 1974); *25* (Nakai et al., 1971); *26* (Ohta and Hayashi, 1973); *27* (Hayashi, 1974); *28* (Botes, 1972); *29* (Botes, 1971); *30* (Strydom, 1973); *31, 32* (Strydom, 1972); *33* (Banks et al., 1974); *34* (Grishin et al., 1974); *35, 36* (Joubert, 1973); *37* (Mebs et al., 1972); *38* (Shipolini et al., 1974); *39* Maeda and Tamiya, 1974); *40* (Viljoen and Botes, 1973); *41* (Viljoen and Botes, 1974); *42* (Shipolini and Banks, 1974); *43* (Botes, 1974); *44, 45, 46* (Louw, 1974a); *47* (Nakai and Lee, 1970); *48* (Takechi and Hayashi, 1972); *49, 50* (Takechi et al., 1973); *51* (Hayashi et al., 1971); *52* (Louw, 1974b); *53* (Hayashi et al., 1975); *54* (Botes, 1974); *55* (Carlsson and Joubert, 1974); *56* (Weise et al., 1973); *57* (Fryklund and Eaker, 1973); *58, 59* (Carlsson, 1974)

subset. Each time a new nodal subset was formed, a prelimset nodal ancestral sequence was reconstructed from its two immediate descendant subsets according to the rules set by Fitch (1971). The third approach is the Minimum Spanning Tree-Cluster Analysis (MSTCA) procedure. The method has been described and discussed in detail by Gower and Ross (1969) and Zahn (1971). We used the algorithm of Dijkstra (1959) in finding a minimum spanning tree from a known distance matrix. Clustering was then performed by successively breaking the linkages starting from the one with longest linking path length.

Among the three procedures employed in the present study both FM and MSTCA algorithms were easily utilizable for computer calculations. The PAS procedure required considerable core storage and computer time.

Statistical Evaluation of a Tree

The statistical evaluation of a tree and calculations of internodal 'path' lengths were made by a least-squares fitting. With N source subsets (the present day amino acid sequences), the number of paths is $2N$-3 (in an unrooted tree). The total number of relations for connecting two source subsets with the sum of path lengths set equal to input distance is $N(N\text{-}1)/2$. The least-squares calculations then correspond to finding $2N$-3 parameters from $N(N\text{-}1)/2$ observations. The function minimized was

$\sum_{i<j} w_{ij} \; (D^o_{ij} - D^c_{ij})$, where D^o_{ij} is the element (observed) of

the distance matrix calculated from MNMR between the ith and the jth source subsets, D^c_{ij} is the corresponding calculated distance, and w_{ij} is the weight associated with D^o_{ij}. Three quantities were calculated to indicate the 'fit' of a tree to the 'observed' distances:

$$(1) \quad \text{GOF} = \left\{ \sum_{i<j}^{N} w_{ij} \; (D^o_{ij} - D^c_{ij})^2 / \; [N(N-1)/2 - (2N-3)] \right\}^{\frac{1}{2}},$$

$$(2) \quad R_w = \left[\sum_{i<j}^{N} w_{ij} \, (D^o_{ij} - D^o_{ij})^2 \; / \sum_{i<j}^{N} w_{ij} \, (D^o_{ij})^2 \right]^{\frac{1}{2}} \times 100 \; \%,$$

$$(3) \quad \text{SD} = \left\{ \sum_{i<j}^{N} [(D^o_{ij} - D^c_{ij}) \; / \; D^o_{ij}]^2 \; / \; [N(N-1)/2-1] \right\}^{\frac{1}{2}} \times 100 \; \%,$$

where GOF stands 'goodness of fit', R_w,% 'weighted residual', and SD, % 'standard deviation'.

There are several criteria exterior to tree building processes that may be used in defining a statistical best tree. The percent 'standard deviation (SD)' based on reconstructed internodal path lengths is commonly used when distance matrix methods are employed in phylogenetic tree construction. SD, to be statistically meaningful, requires the assumption that errors associated with the 'observed' distances are Poisson

distributed. Both GOF and R_w also need the supposition that the distribution is random. GOF is the external estimate of the standard error of an observation of unit weight. A correct model will give GOF a value of 1 only when weights are correctly estimated for all 'observed' distances. In assuming a Poisson distribution, R_w and GOF are further related to SD as follows.

$$(4) \quad \text{GOF} = \left\{[N(N-1)/2-1] \;/\; [N(N-1)/2(2N-3)]\right\} \tfrac{1}{2} \times (\text{SD}/100),$$

$$(5) \quad R_w = \left\{[N(N-1)/2-1] \;/\; [N(N-1)/2]\right\} \tfrac{1}{2} \times (\text{SD}).$$

The indices of fit calculated from unit-weight least-squares calculations in the absence of a proper weighting scheme are at best only suggestive.

After the path lengths were calculated, the center of the tree (point of earliest time), which is the point on a path of the tree such that the distance from any source subset to the point is about equal, was located. The nodal ancestral amino acid sequences were then determined by minimum mutation fit (Hartigan, 1973); the computations are in terms of amino acid residues.

The resulting tree topologies of snake venom toxins from all three procedures are given in Figure 3. Their corresponding indices of fit from unit-weight least-squares calculations are: GOF = 2.92, R_w = 5.15, SD = 8.82 for FM tree; GOF = 2.87, R_w = 5.06, SD = 8.85 for PAS tree; and GOF = 2.86, R_w = 5.04, SD = 8.55 for MSTCA tree. Figure 2 gives the source amino acid sequences.

Apparently, the tree resulting from the MSTCA calculations is the best one, with all three indices of fit the lowest. However, it does not exclude the possibility of finding a more statistically optimal alternative tree, as not all of the alternative trees possible from FM and PAS calculations were tested. Also, a statistically best tree (or phenogram) based on knowledge of the present amino acid sequences coded by few genes may not be the most genealogically significant one (Fitch and Margolish, 1968).

Least-squares calculations also provide estimated errors for all path lengths. Estimated errors of path lengths are related partly to cluster separability. They could result from assumptions of a constant rate of evolution, unknown multiple mutations, improper sequence alignment, inexact calculated mutation distances, or erroneous tree topology. Unit-weight least-squares calculations gave estimated errors in calculated path lengths for all three toxin ranging from a maximum of 1.8 to a minimum of 0.5. When estimated errors are considered, all three trees differs insignificantly. The statistical uncertainty involved in the separation of a cluster from other thus leaves room for different interpretations of the clusters concerned.

Two other trees using mutation distances estimated from a formulation (Fitch and Margoliash, 1967) which does not treat gaps as 21st amino acid, and a scheme using the number of amino acid differences were also constructed employing the MSTCA procedure. They are essentially identical to those given in Figure 3.

Toxin Phylogenetic Relationship

Most of the toxins are well clustered at the genus level (MSTCA tree, Fig. 3) except *H. cyanocinctus* hydrophitoxin b (13a, the identification number in Fig. 2), which is identical in sequence to *E. schistosa* toxin 5 (13), and *N. naja oxiana* neurotoxin I

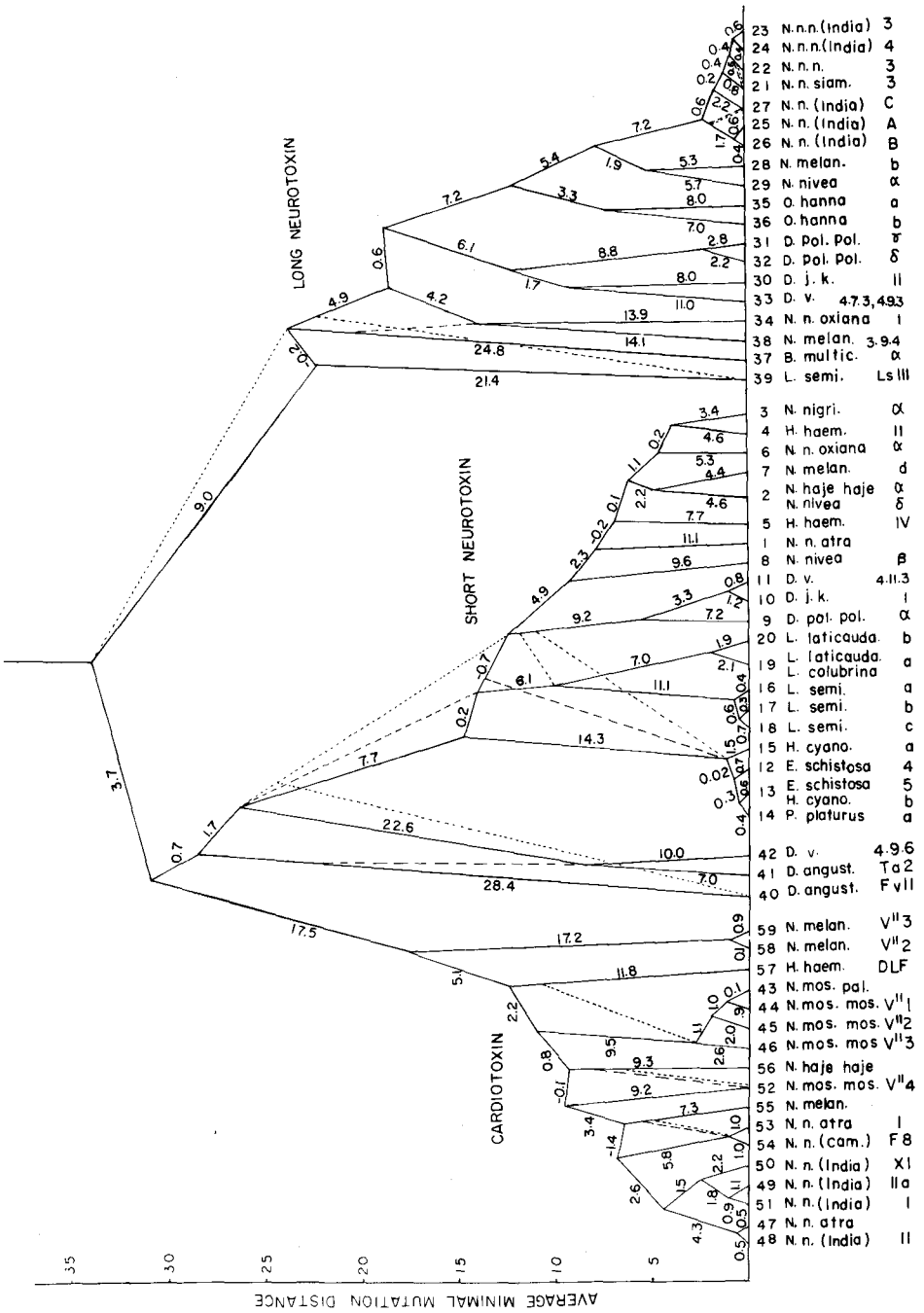


Fig. 3. Toxin phylogenetic trees. Solid lines indicate the MSTCA tree. The dashed lines apply only to FM tree and the dotted lines to PAS tree

(34) and *N. melanoleuca* toxin 3.9.4 (38), both of which are not clustered together with other *Naja* toxins. At the species level, there are two more coincidences: *N. haje haje* toxin α (2) is identical to *nivea* toxin δ (2a); and the only toxin from *L. colubrina* coincides with *L. laticaudata* laticotoxin a (19). All these concurrences are of the short neurotoxin type. A number of obvious oddities can also be found at the species level in the tree. The principal anomalies are the two cardiotoxins (47, 53) from *N. naja atra,* two short neurotoxins (2a, 8) from *N. nivea* and two long neurotoxins (28, 38) from *N. melanoleuca.* They were not grouped together as would be expected.

Toxins from genus *Naja,* therefore, constitute the major source of problems in the interpretation of the tree at the species level. They comprise over half of the present group of 62 snake toxins of known amino acid sequence. In view of the large number of toxin varieties that may be present in the venom of a single snake, it is likely that the anomalous toxins in the phylogenetic tree could be the products of gene duplications at different times or a sudden burst in mutation rates in the course of evolution of genus *Naja*, which is the most recently diverged, and a geologically widespread one.

Irrespective of the above complications, the present phylogenetic tree, however, manifestly classifies these toxins into four corresponding to the four commonly used type names. The classification is the same whether the distance matrix was based on MNMR or in terms of amino acid difference between sequences.

The so-called angusticeps-type toxin is a new addition to the recent outpouring of snake toxin sequences. So far, only three of them are known. Their pharmacological activities have not yet been clarified. Cardiotoxins have an irreversible depolarizing effect on cell membrane of muscle, especially cardia and nervous tissues (Lee, 1972). They have not been isolated from sea-snake venoms. On the other hand, the neurotoxins block neuromuscular transmission by the competitive inhibition of acetylcholine at the postsynaptic membrane of the neuromuscular junction (Chang and Lee, 1966). The short neurotoxins are the most toxic components of snake venoms; they are present in almost all of the snake venoms of both *Elapidae* and *Hydrophiidae* studied to date. The long neurotoxins, containing one additional disulfide bond, are thought to be the result of an unequal crossing over of the toxin genes (Strydom, 1973). The results of a statistical calculation (Fitch, 1970) based on the present tree topology and its reconstructed nodal amino acid sequences indicated that all these four types of toxins are undoubtly homologous.

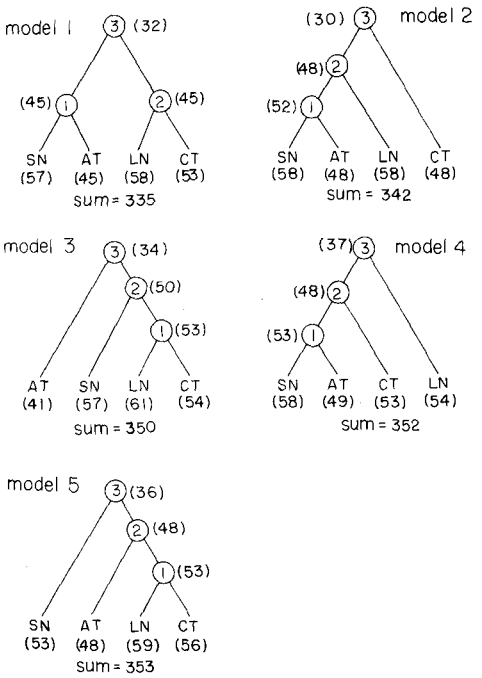
Homologous toxins, however, can be either orthologous or paralogous (Fitch, 1973). From inspecting the phylogenetic tree of Figure 3, we fail to tell immediately orthologous gene products from paralogous ones. As mentioned earlier, multiple toxin varieties are present in the venom of a single individual snake. Those toxins from the same snake and with large number of amino acid residue differences among sequences are very likely the products of gene duplication, for example, toxin a and b of *O. hanna* (35, 36); toxin δ and β of *N. nivea* (2, 8); toxin II and IV of *H. haemachatus* (4, 5); cardiotoxin and cardiotoxin I of *N. naja atra* (47, 53). However, a number of toxins isolated from pooled venom material of snakes of the same species and demonstrated to have only a small number of amino acid residue differences among sequences are probably controlled by allelic genes (Fitch, W.M., personal communication), such as erabutoxins of *L. semifasciata* (16, 17, 18). It is very unlikely that they were the

products of a series of recent gene duplication events. A more detailed analysis will have to wait till more sequences are available.

The simultaneous presence of orthologous and paralogous toxins in the phylogenetic tree as well as the possibility of a high frequency of gene duplication leading to the present great varieties of homologous toxins in the venom of snakes complicated the explanation of the course of evolution of these four types of toxins from the tree topologies. The center of the tree given in Figure 3 may not be the point of earliest time in the divergence of these toxins. Strydom (1974) gave two mutually exclusive postulates for the evolution of these toxins from digestive tract enzyme based on the number of amino acids assigned to reconstructed nodal ancestal sequences. One had a short neurotoxin and the other had a cardiotoxin as the ancestral venom protein.

Evolutionary Models of Toxins

Mathematically, there are five different evolutionary pathways possible from an unrooted tree containing four different types of toxins. Those models based on the present unrooted tree (Unrooted MSTCA tree) of 62 snake toxins are shown in Figure 4. Only the most remote ancestral nodes of each type of toxins are given. In distinguishing the probable model from others, we assumed that the most elegant evolutionary pathway would have the largest sum of amino acids assigned to all nodes by minimum mutation fit. We called it 'minimax' criterion, meaning maximum number of amino acids assigned by minimum mutation fit. The number of amino acids assigned to each node of the five models together with their sums are also given in Figure 4. The results




```
     1          10         20         30
ND3  M+CYN++++QPA+T+TC++GENSCY+K+WR+++--
SN   M+CYNQQSSQPATT+TC++GENSCYKKTWRD-H--
ND2  M+CYN+++TQPA+T+TC+EGENSCYRK+WRK+---
AT   M+CYSH+TTQPA+T+TC+EGENSCYRKS+RK+---
ND1  I+CYN--KT+P++S+TCPEGQNLCYRKMW+KA---
LN   I+CY---KTPPDTSQTCP+GQNLCYRKMWC+AWCS
CT   IKCHN--KL+PF+SKTCPEGQNLCY-KMT+KA---
     40         50         60         70
-+G+++ERGC--GCP+VK-P+++++CC++-DKCN+-----
-RGT+IERGC--GCP+VK-PGIKL+CC++-DKCNN-----
-PGVV+ERGC--GCP+VK-PLVEIKCC++-DKCN+-----
-PGVV+GRGC--GCP++++-+LVEIKCC+S-DKCN+-----
+PGVVVERGCAATCP+VK-PLVEIKCC+T-DKCN+-----
SRGKVVE+GCAATCPSVK-P++EIKCCST-DKCN++P---
+PKV+VKRGCAATCPKSS-LLVKYVCCNT-DKCN------
```

Fig. 4. Alternative topologies of the MSTCA tree describing the evolutionary relationship of the four types of toxins, short neurotoxins (SN), long neurotoxins (LN), cardiotoxins (CT), and angusticeps-type toxins (AT). The number in parentheses are the number of ancestral nodal amino acids assigned from minimum mutation fit. Only ancestral amino acid sequences for Model 5 are shown. The numbers in the circles are the node numbers (ND). A (+) symbol indicates that amino acid residue at the position could not be unambiguously assigned

thus indicate Model 5 is the most probable one. However, the sum of Model 5 is only 1 greater than Model 4 and 3 greater than Model 3. Model 5 and Model 3 are essentially the same in view of the small path length separating AT and SN nodes. Therefore, Model 5 and Model 4 are the two most plausible models based on the 'minimax' criterion. Model 5 was considered the most likely one by Strydom (1974). He rejected Model 4 simply on the basis of the proposition that a plausible mechanism exists for the insertion of the extra disulfide loop of the long neurotoxin into a short neurotoxin- or cardiotoxin-type of chain.

The relevant evidence supporting one model instead of others is presently unascertainable. Different numbers of toxin varieties in snake venoms, the possibility of different gene duplication frequencies and other genetic events happening in the evolution of snakes of different groups, and the presence of both orthologous and paralogous toxins in the phylogenetic tree constructed indicate that a more detailed explanatory model simulating the course of evolution of snake toxins cannot be inferred simply from the appearance of a phylogenetic tree.

Rates of Toxin Evolution

We measured rate of evolution in terms of amino acid differences between sequences normalized to a common base of 100 amino acids and corrected for the possibility that two or more amino acid changes have occured at the same place (Dickerson, 1971). All comparisons were made for species of which at least two types of toxins have been sequenced. When more than one toxin of the same type were present in the species compared, an average rate was calculated. These values together with mutation distances calculated from MNMR are given in Table 1.

The analysis indicates, intragenerically, that the rates of amino acid change of short neurotoxins, on the average, are relatively slower than those of long neurotoxins and cardiotoxins. Similarly, the rates of cardiotoxins are faster than those of long neurotoxins. However, intergenerically, the rates of amino change of all those three types of toxins are comparable. Although only a small fraction of all toxins have been compared, the observations indicate that evolution of snake toxins is rather peculiar. While uniform rates of amino acid change can be assumed for parallel evolution of the three types of toxins from genetically more distantly related species, intragenerical different rates of amino acid change seems to require the supposition that long neurotoxins and cardiotoxins have been undergoing a more rapid change in the selection of the functions of the two proteins. It would be contradictory to the observation of comparable rates of amino acid change above the genus level if we assumed that the more rapidly evolving (below the genus level) cardiotoxin-type protein is the ancient form of the snake toxins. A more meaningful discussion in terms of effects of amino acid changes on the physical and physiological properties will be possible only when information from X-ray structure determination or chemical modification is available relative to all those three types of toxins. However, a uniform rate of amino acid change for homologous proteins would be expected if amino acid changes are selectively neutral. This seems in accord with the comparable number of common amino acid residues from sequence alignment: 26 in short neurotoxins; 21 in long neurotoxins;

Table 1. Intra- and inter-generical rates of amino acid change

Genus[a]	Type of toxin[b]	Toxins compared[c]	Amino acid changes per 100 residues corrected for multiple hits[d]	Minimum mutation number per 100 residues[d]
		(A) Intragenerical		
N.	SN	(2,6,7)	17.6(5.2)	19.4(5.8)
	LN	(28,29,34)	37.7(22.1)	40.4(22.5)
N.	SN	(2,3,5)	18.3(4.9)	19.9(4.1)
	CT	(43,53,56)	35.7(2.7)	36.0(3.4)
D.	SN	(9,10,11)	12.1(7.7)	13.4(8.8)
	LN	(30,32,33)	27.1(2.2)	28.8(2.5)
		(B) Intergenerical		
D./N.	SN	(9,10,11)/(2,6,7)	42.2(4.9)	43.0(5.6)
D./N.	LN	(30,32,33)/(28,29,34)	49.0(2.5)	51.0(4.2)

a Abbreviations for genera Naja (N.) and Dendroaspis (D.)
b Abbreviations for short neurotoxins (SN), long neurotoxins (LN), and cardiotoxins (CT)
c Toxins compared are represented by identification numbers as given in Figure 1
d Estimated errors are in parentheses

and 22 in cardiotoxins. It also agrees with the fact that these toxins assume similar biological roles in snakes.

The above discussion, therefore, implies that short neurotoxins had been relatively stabilized since acquiring their toxicity. The apparent absence of cardiotoxins in venoms of sea snakes and the isolation of only a long neurotoxin in venom of *L. semifasciata* further suggest that short neurotoxin-type molecules are the most primitive form of toxins and that other types of toxins have evolved in one way or another from these molecules. This agrees with Model 5, chosen on the basis of the 'minimax' criterion, in one aspect – the ancient form of toxins.

Cladistic Relationships of Snakes

Toxins used in the present study were isolated from venoms of twenty three subspecies of eighteen species representing nine genera from two families of the *Proteroglyphae* suborder. The tree topologies indeed show acceptable genetic relatedness above the species level. Figure 5 gives a conservative depiction of the genus phylogeny inferred both from the tree and an analysis using composite data where they were available.

The venom contents of sea snakes are simpler. No cardiotoxin-type proteins have yet been isolated from them. Only one type toxin (Ls III), similar to long neurotoxin has been isolated and sequenced from *L. semifasciata.* The branching topologies of the short neurotoxin portion of the tree depict a very close relationship among sea snakes. The taxonomic relationship is in remarkable agreement with zoological opinions. The

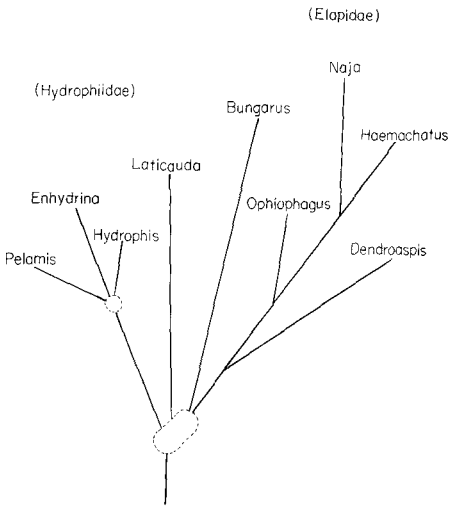


Fig. 5. Cladistic relationship of snakes at genus level as inferred from toxin phylogeny

phylogenetic scheme, however, does not conclusively agree with the suggestion that *Laticaudinae* is the most primitive form of the *Hydrophiidae* (Smith, 1926). Of course, the implication from toxin phylogeny is uncertain as the errors associated with the path lengths separating *Laticaudinae* from *Hydrophinae* is larger (± 1.6). It only indicates that *Laticaudinae* and *Hydrophinae* diverged very early in the evolutionary history of *Hydrophiidae.*

If one accepts the supposition that *Elapidae* arose from aquatic forms and that no cardiotoxin-like components in sea snake venoms will be observed, the evolutionary model of toxins is such that cardiotoxins were evolved after *Elapidae* were established as a terrestrial group. This also agrees with the result that the *Hydrophinae*, including the ocean-going *Pelamis platurus*, diverged long before the short neurotoxin portion of the toxin phylogenetic tree. Alternatively, if one considers that *Hydrophiidae* had originated from terrestrial *Elapidae* and that *Hydrophiidae* were the aquatically adapted group of their ancient terrestrial counterpart, the above suggested model for toxin evolution would indicate that cardiotoxins branched off late after the divergence of sea snakes. The long neurotoxins' intermediate rate of amino acid change (intragenerical) would imply that they evolved earlier than cardiotoxin did.

Members of *Laticaudinae* have usually been described as partly marine and partly terrestrial and the name 'seakrait' was suggested because of their close relation to the *Elapidae* (Burger and Natsuno, 1974). Those of *Hydrophinae* were identified as true sea snakes (Burger and Natsuno, 1974). McDowell (1967) hypothesized that *Laticaudinae* and *Hydrophinae* had separate origins and independent histories. The short neurotoxin part of the present toxin tree indicates only that *Laticaudinae* diverged between *Hydrophinae* and *Elapidae.*

Venom contents of *Elapidae* are much more complicated; nevertheless the taxonomic relationship of the toxins is as expected from zoological classification except for a few kraits, which appear to be closer to the ancient form of *Elapidae.* The highly developed 'spitting' cobra (*H. haemachatus*) and genus *Naja* diverged only recently. The time after their divergence is not long enough to allow their genetic differentiation to be clearly distinguished in the phylogenetic tree. In the absence of further information,

we consider the present cladistic relationship at the genus level and as given in Figure 5 to be the best possible consistent with the present evidence.

Strydom (1974) has suggested, from sequence comparisons between cobrotoxin and bovine pancreatic ribonuclease, that toxins evolved from an ancestor of ribonuclease or a homologous enzyme in the snake digestive tract through gene duplication. The relatively slower rates of evolution of short neurotoxins and their presence in both *Elapidae* and *Hydrophiidae*, therefore, allow an estimate to be made for the branching date of the two families. Their averaged number of amino acid differences per 100 amino acids with masking effect being corrected is 45.4 (± 6.6). Assuming short neurotoxins evolved at the same rate as ribonuclease, 2.1 UEP (unit evolutionary period per million years) (Dickerson, 1971), the estimated diverging time between *Elapidae* and *Hydrophiidae* is about 100 (± 14) million years. This is not an unreasonable figure considering the branch dates estimated from paleontological records for mammals and reptiles (Dickerson, 1971; Yang, 1962), 300 million years, and birds and reptiles, 240 million years.

Course of Toxin Evolution

So far we have not considered explicitly the so-called angusticeps-type toxins. There are currently only three of them isolated, all from a single genus, *Dendroaspis.* These three toxins differ greatly from each other in terms of number of amino acid differences. The number of amino acid differences indicate they would be more properly classified into two types: angusticeps-F-type (*D. angusticeps* toxin FvII) and angusticeps-T-type (*D. angusticeps* toxin Ta2 and D. viridis toxin 4.9.6). The intragenerical rate of amino acid change between the two angusticeps-T-type toxins is intermediate (20.8) as compared to their short neurotoxins (12.1, averaged value) and long neurotoxins (26.4) which may indicate that they diverged before the long neurotoxins. The assignment of a branch point for angusticeps-type toxins would be as indicated in Model 5 (Fig. 4).

Therefore, Model 5 is the most plausible one for the evolution of toxin. It has been implicitly expressed in the unrooted phylogenetic tree and was chosen based on the 'minimax' criterion. The choice of the model, however, is principally based on the two generally held ideas that uniform rate of evolution may be assumed for homologous proteins from genetically more distantly related species and that a protein underwent a more rapid change in amino acid composition when it was selecting for a new function.

Although the suggested model for the evolution of toxins is very plausible, the total complexity of the relationships between intraspecific homologous toxins required further clarification. Gene duplication and allelomorphism have been considered the two principal mechanisms leading to intraspecific homologous proteins (Neurath et al., 1967). Extending the discussion given by Chang (1972), we explain the course of evolution of toxins leading to the present great number of toxin varieties by the simplified diagrammatic illustration given in Figure 6. The ancestral S gene duplicated before separation of the two most ancient branching points. In one lineage, the descending duplicated genes underwent unequal crossover. The event of unequal crossover introduced the L. gene. In the other lineage, the first ancestral separation resulted in one of the descending duplicated genes coding for a new protein (A gene). Further indepen-

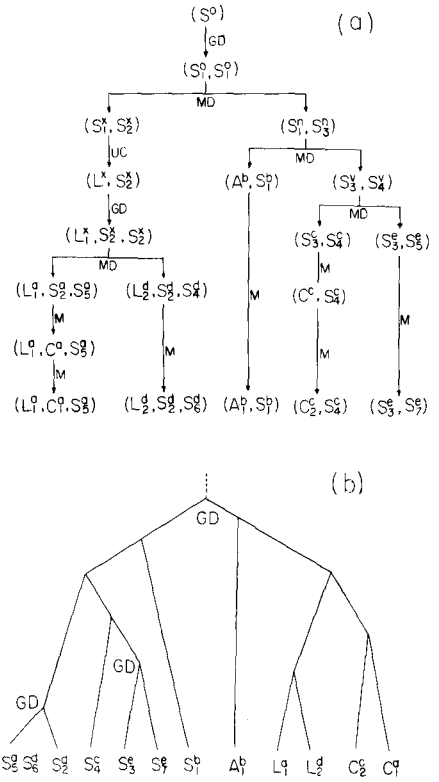


Fig. 6. A simplified diagrammatic illustration of genetic events which might have happened in the evolutionary history of toxins. (*S*) represents short neurotoxin gene, (*L*) long neurotoxin, (*C*) cardiotoxin, and (*A*) angusticeps-type toxin. The alphabetic superscripts indicate distinct species, the numerical subscripts are for identification purpose. *GD* denotes gene duplication. *M* means mutation. *MD* indicates mutation divergence accompanying speciation. (a), the course of evolution; (b), one of the possible phylogenetic trees

dent evolution in separated lines may ultimately have produced C genes by accumulated mutational changes. One of the possible existing phylogenetic trees of all these genes then may have the appearance shown in Figure 6b. However, it will be possible to reproduce the entire series of genetic events leading to a toxin phylogenetic tree only when sequence data from most of the toxins coded by homologous genes in all species studied are available.

Acknowledgment. We thank Mr. C. C. Uh for help in data compilation and Chung Shan Institute of Science and Technology for providing generous computer time. This work was supported in part by National Science Council, Republic of China.

References

Arnberg, H., Eaker, D., Fryklund, L., Karlsson, E. (1974). Biochem. Biophys. Acta **359**, 222

Banks, B.E.C., Miledi, R., Shipolini, R.A. (1974). Eur. J. Biochem. **45**, 457

Botes, D.P. (1971). J. Biol. Chem. **246**, 7383

Botes, D.P. (1972). J. Biol. Chem. **247**, 2866

Botes, D.P. (1974). Quoted by Maeda, N., Tamiya, N. (1974). Biochem. J. **141**, 389

Botes, D.P., Strydom, D.J. (1969). J. Biol. Chem. **244**, 4147

Botes, D.P., Strydom, D.J., Anderson, C.G., Christensen, P.A. (1971). J. Biol. Chem. **246**, 3132
Broadley, D.G. (1968). Arnoldia, 3, 1
Burger, W.L., Natsuno, T. (1974). The Snake **6**, 61
Carlsson, F.H.H. (1974). Biochem. Biophys. Res. Commun. **59**, 269
Carlsson, F.H.H., Joubert, F.J. (1974). Biochem. Biophys. Acta **336**, 453
Chang, T.W. (1972). M. S. Thesis, National Tsing Hua Univ., Hsinchu, Taiwan.
Chang, C.C., Lee, C.Y. (1966). Brit. J. Pharmacol. Chemother. **28**, 172
Dayhoff, M.O. (1969). Atlas of protein sequence and structure, Vol. 3. National Biochemical Research Foundation, Silver Spring, Md.
Dayhoff, M.O., Eck, R.V. (1966). Atlas of protein sequence and structure, Vol. 2. National Biochemical Research Foundation, Silver Spring, Md.
Dickerson, R.E. (1971). J. Mol. Evol. **1**, 26
Dijkstra, E.W. (1969). Numer. Math. **1**, 269
Eaker, D.L., Porath, J. (1967). Jap. J. Microbiol. **11**, 353
Fitch, W.M. (1966). J. Mol. Biol. **16**, 9
Fitch, W.M. (1970). Syst. Zool. **19**, 99
Fitch, W.M. (1971). Syst. Zool. **20**, 406
Fitch, W.M. (1973). Ann. Rev. Genet. **7**, 343
Fitch, W.M., Margoliash, E. (1967). Science **155**, 279
Fitch, W.M., Margoliash, E. (1968). Brookhaven Symp. on Biology **21**, 217
Fryklund, L., Eaker, D. (1973). Biochemistry **12**, 661
Fryklund, L., Eaker, D., Karlsson, E. (1972). Biochemistry **11**, 4633
Gower, J.C., Ross, G.J.S. (1969). Appl. Statistics **18**, 54
Grishin, E.V., Lukyanchu, A.P., Slobodyan, L.N., Lipkin, V.M., Orchinnikov, Y.A., Sorokin, V.M. (1973). FEBS Lett. **36**, 77
Grishin, E.V., Sukhikh, A.P., Orchinnikov, Y.A. (1974). 4th Intern. Symp. on Animal, Plant and Microbial Toxins, Tokyo. Abst. p. 60
Hartigan, J.A. (1973). Biometrics **29**, 53
Hayashi, K. (1974). Quoted by Maeda, N., Tamiya, N. (1974). Biochem. J. **141**, 389
Hayashi, K., Takechi, M., Sasaki, T. (1971). Biochem. Biophys. Res. Commun. 45, 1357
Hayashi, K., Takechi, M., Sasaki, T. Lee, C.Y. (1975). Biochem. Biophys. Res. Commun. **64**, 360
Johnson, R.G. (1956). Evolution **10**, 56
Jou, E.D. (1975). Proc. Intern. Computer Symp. Vol. 1. p. 90
Joubert, F.J. (1973). Biochem. Biophys. Acta **317**, 58
Karlsson, E. (1974). Quoted by Maeda, N., Tamiya, N. (1974). Biochem. J. **141**, 389
Karlsson, E., Eaker, D., Ponterius, G.H. (1972). Biochem. Biophys. Acta **257**, 235
Lee, C.Y. (1972). Ann. Rev. Pharmacol. **12**, 265
Liu, C.S., Blackwell, R.Q. (1974). Toxicon **12**, 543
Liu, C.S., Huber, G.S., Lin, C.S., Blackwell, R.Q. (1973). Toxicon **11**, 73
Louw, A.I. (1974a). Biochem. Biophys. Acta **336**, 481
Louw, A.I. (1974b). Biochem. Biophys. Res. Commun. **58**, 1022
Maeda, N., Tamiya, N. (1974). Biochem. J. **141**, 389
McDowell, S.B. (1967). J. Zool. London **151**, 497
Mebs, D., Narita, K., Iwanaga, S., Samejima, Y., Lee, C.Y. (1972). Hoppe-Seyler's Z. Physiol. Chem. **353**, 243

Moore, G.W., Goodman, M., Barnabas, J. (1973). J. Theoret. Biol. **38**, 459
Nakai, K., Lee, C.Y. (1970). Biochem. Biophys. Res. Commun. **41**, 339
Nakai, K., Sasaki, T., Hayashi, K. (1971). Biochem. Biophys. Res. Commun. **44**, 893
Neurath, H., Walsh, K.A., Winter, W.P. (1967). Science **158**, 1968
Ohta, M., Hayashi, K. (1973). Biochem. Biophys. Res. Commun. **55**, 431
Peacock, D., Boulter, D. (1975). J. Mol. Biol. **95**, 513
Sato, A. (1974). Quoted by Maeda, N., Tamiya, N. (1974). Biochem. J. **141**, 389
Sato, S., Tamiya, N. (1971). Biochem. J. **122**, 453
Shipolini, R.A., Bailey, G.S., Banks, B.E.C. (1974). Eur. J. Biochem. **43**, 203
Shipolini, R.A., Banks, B.E.C. (1974). Eur. J. Biochem. **49**, 399
Smith, M.A. (1926). A Monograph of the sea snake (Hydrophiidae). Brit. Mus. (Nat. Hist.) London
Strydom, A.J.C. (1973). Biochem. Biophys. Acta **328**, 491
Strydom, A.J.C., Botes, D.P. (1971). J. Biol. Chem. **246**, 1314
Strydom, D.J. (1972a). Toxicon **10**, 39
Strydom, D.J. (1972b). J. Biol. Chem. **247**, 4029
Strydom, D.J. (1973). Comp. Biochem. Physiol. **448**, 269
Strydom, D.J. (1974). Syst. Zool. **22**, 596
Takechi, M., Hayashi, K. (1972). Mol. Pharmacol. **8**, 446
Takechi, M., Sasaki, T., Hayashi, K. (1973). Proc. Symp. 11th Peptide Chem. p. 108
Tamiya, N., Abe, H. (1972). Biochem. J. **130**, 547
Viljoen, C.C., Botes, D.P. (1973). J. Biol. Chem. **248**, 4915
Viljoen, C.C., Botes, D.P. (1974). J. Biol. Chem. **249**, 366
Wang, C.L., Liu, C.S., Hung, Y.O., Blackwell, R.Q. (1976). Toxicon **14**, 459
Weise, K.H.K., Carlsson, F.H.H., Joubert, F.J., Strydom, D.J. (1973). Hoppe-Seyler's Z. Physiol. Chem. **354**, 1317
Yang, C.C., Yang, H.J., Huang, J.S. (1969). Biochem. Biophys. Acta **188**, 65
Yang, J.Z. (1962). The life of vertebrates. Cambridge: Oxford University Press
Zahn, C.T. (1971). IEEE Trans. Computers **C-20**, 68

Received December 16, 1976