COMPARISONS OF PRIMATE CATALASE TRYPTIC PEPTIDES AND IMPLICATIONS FOR THE STUDY OF MOLECULAR EVOLUTION[1,2]

SANDRA L. HORNE

Department of Human Genetics, University of Michigan, Ann Arbor, Michigan

With the development of new techniques in protein chemistry, the analysis of genetic differences between species can be carried out utilizing characters only a few steps removed from the actual genetic material. Our present understanding of the mechanisms involved in organic evolution is only to a very minimal degree based upon information about the many mutations acquired by a single homologous gene in the course of its divergent evolution in different species. If the observable differences in the amino acid composition of a given protein are due to differences in the base pairs within the structural gene, then these amino acid differences reflect some of the intragenic changes which have become incorporated in species during their evolutionary development. On the basis of specific amino acid substitutions in the various polypeptide chains of hemoglobin (Buettner-Janusch and Hill, 1965a; Zuckerkandl and Pauling, 1965; Ingram, 1961; Braunitzer et al., 1965) and cytochrome *c* (Margoliash and Smith, 1965), a number of speculations on evolutionary relationships and time scales have been presented.

It is the purpose of this study to provide preliminary data on another protein, erythrocyte catalase, from several primate species. These data were obtained by comparing the two-dimensional patterns, frequently referred to as "fingerprints," formed by tryptic peptides when subjected to chromatography and high voltage electrophoresis. Insofar as differences in peptides reflect differences in the structural gene, these data, as well as those for the amino acid sequences in hemoglobin and cytochrome *c* are used to illustrate the extension of analyses such as those by Zuckerkandl (1965) and Margoliash and Smith (1965). A new application of the statistical methods of Cavalli-Sforza and Edwards (1964) is also presented which permits the development of evolutionary schemes for single genes with estimated rates of change in each phylogenetic line.

MATERIALS AND METHODS

Catalase was isolated from the erythrocytes of spider monkey (*Ateles belzebuth*), cynomolgus macaque (*Macaca irus*), rhesus macaque (*Macaca mulatta*), green monkey (*Cercopithecus aethiops*), baboon (*Papio cynocephalus*), and man.

Isolation and Purification of Catalase.—Blood was drawn in approximately 0.20 volume of 4% sodium citrate. Samples of 80 ml or more were treated separately, and smaller ones were pooled. Efforts to determine the intraspecific homogeneity of these samples by vertical starch gel electrophoresis were unsuccessful because of the time lag between drawing and receipt of the samples and the consequent variability in the electrophoretic mobility of catalase (Baur, 1963). While this variability may have obscured some intraspecies differences, such differences are probably few in comparison with those seen between species.

The catalase was isolated by a modified version of the method of Herbert and Pinset (1948) in which the hemoglobin was removed by precipitation with chloroform and ethanol. The crude catalase solution
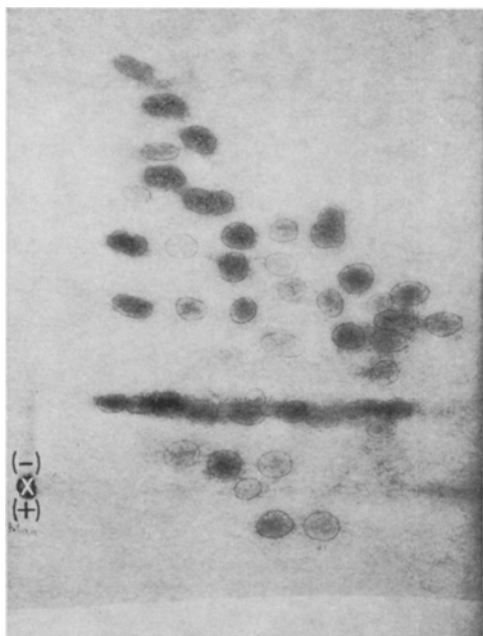
was concentrated in collodion bag membrane filter concentrators (Schleicher and Schuell Co.) to 0.01 of its original volume. Then most of the proteins with molecular weight less than 100,000 were removed by gel filtration with Sephadex G-100 using 0.1 M phosphate buffer. This was followed by precipitation at pH 4 and pH 5.6 of some remaining contaminating protein. After vertical starch gel electrophoresis of the preparation, the gels were sliced and stained for protein with nigrosin and for catalase activity (Thorup et al., 1961). Absence of all protein bands but that corresponding to the catalase activity suggested that no single contaminating protein was likely to be present as more than 2 to 3% of the total protein. Following purification, the solution was lyophilized and stored at −20 C. Yields were on the order of 60 to 80 mg per 500 ml of whole blood.

Peptide Analysis.—Fingerprinting of tryptic digests of catalase was carried out by the method of Chernoff and Liu (1961) using 3 mg of catalase per fingerprint and

performing the descending chromatography with butanol : acetic acid : water : pyridine, 15 : 3 : 12 : 10 (Swenson et al., 1962). Methods used for detection of peptide spots and some amino acids contained therein included scanning with ultraviolet light and staining with ninhydrin, ninhydrin-collidine, the Ehrlich stain for tryptophan, Sakaguchi stain for arginine, Pauly stain for histidine and tyrosine (Chernoff and Liu, 1961), and a stain for bivalent sulfur of cystine, cysteine and methionine (Toennies and Kolb, 1951). For some stains it was possible to perform two or three staining procedures successively on one fingerprint with results comparable to those on which only one stain was done (Chernoff and Liu, 1961). Following each procedure, the positive spots were circled and scored for intensity (+, ++, and +++) and any color variation noted. Thus, though some of the spots fade with time, a permanent record was retained and the fingerprint sheets with the marked peptide spots could be used for comparison long after they had faded.

RESULTS

From six to 40 fingerprints were prepared from the catalase of each species. A photograph of a representative human fingerprint stained with ninhydrin is shown in Figure 1. The peptide patterns of all the species considered are quite similar. Therefore, special stains were used to permit the identification of differences not discernible from the peptide positions alone. Because of the relatively large number of spots on each fingerprint, and the small number of these which occupied different positions from species to species, the following system of identification was developed. A standard pattern of peptide spots was established which included all the spots observed for all the species considered. Each spot was then assigned a number based solely on its position regardless of its staining characteristics (Fig. 2). For example, peptide number 17 which occurs in all the primate catalases examined, stains
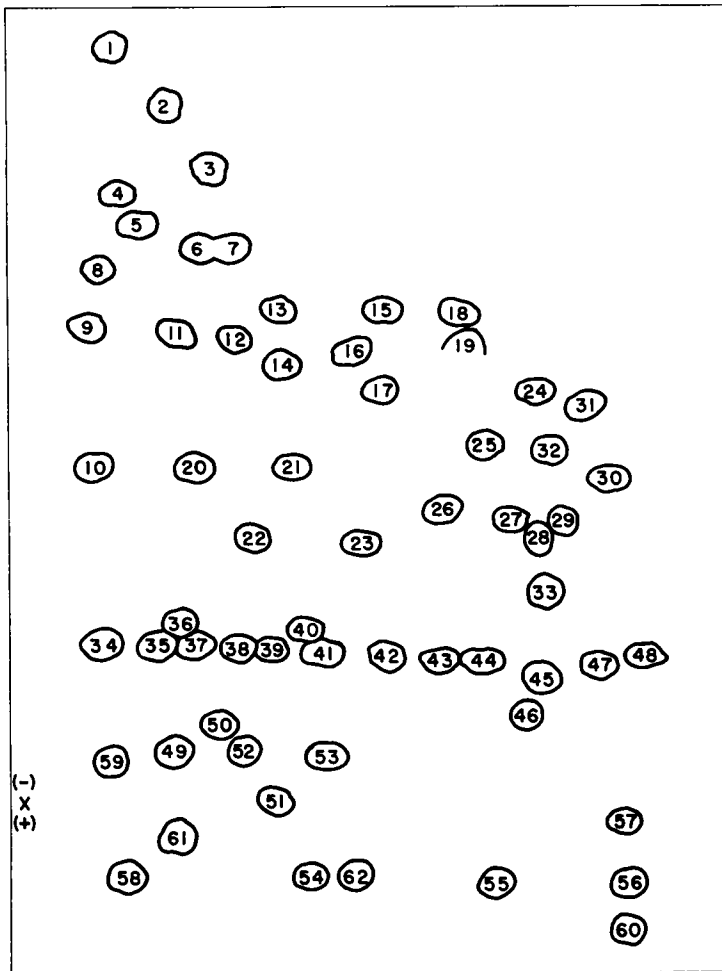
FIG. 2. Diagram of the composite tryptic peptide pattern.

only with ninhydrin in human catalase, whereas in all the monkey catalases it stains strongly with both Ehrlich and Pauly stains.

Table 1 lists all the peptides by number and gives their staining characteristics for each species. Peptides which were positive at any intensity for the stain in question in 67–100% of the fingerprints are ++, those positive in 34–66% are +, and those positive in 0–33% are blank. If only one fingerprint was stained with a particular stain, a + was used. This approximation in the method of scoring was made because

of the variability in the staining reaction of some peptides and the infrequent appearance of others.

The special stains have probably not identified all the peptides containing the amino acid for which they give positive stains, particularly tryptophan and bivalent sulfur. This is suggested by the variability in staining intensity, as well as by comparison with equine and bovine liver catalases which contain 44 moles of tryptophan and 70 moles of half cystine and methionine (Schroeder et al., 1964). If there are four identical subunits as suggested by Tanford

TABLE 1. *Staining characteristics of the primate catalase peptides. Stains used were the Sakaguchi stain for arginine, Pauly for histidine and tyrosine, Ehrlich for tryptophan, bivalent sulfur, ninhydrin and ninhydrin-collidine. Spots staining with a frequency of 67–100% are designated ++; those staining 34–66% are +; and those staining 0–33% are left blank. The number at the head of each column is the number of fingerprints evaluated with each type of stain.*

Peptide number	Human						Rhesus						Cynomolgus						Cercopithecus						Baboon						Spider					
	His		Cys				His		Cys				His		Cys				His		Cys				His		Cys				His		Cys		
	Arg	Tyr	Try	Met	Nin	Coll	Arg	Tyr	Try	Met	Nin	Coll	Arg	Tyr	Try	Met	Nin	Coll	Arg	Tyr	Try	Met	Nin	Coll	Arg	Tyr	Try	Met	Nin	Coll	Arg	Tyr	Try	Met	Nin	Coll
	7	8	6	2	8	9	7	7	6	4	6	7	2	2	3	3	2	2	4	6	2	1	4	5	2	2	1	1	2	1	1	1	1	1	1	1

TABLE 1. Continued.

Peptide number	Human						Rhesus						Cynomolgus						Cercopithecus						Baboon						Spider				
	His		Cys				His		Cys				His		Cys				His		Cys				His		Cys				His		Cys		
	Arg	Tyr	Try	Met	Nin	Coll	Arg	Tyr	Try	Met	Nin	Coll	Arg	Tyr	Try	Met	Nin	Coll	Arg	Tyr	Try	Met	Nin	Coll	Arg	Tyr	Try	Met	Nin	Coll	Arg	Tyr	Try	Met	Nin	Coll
	7	8	6	2	8	9	7	7	6	4	6	7	2	2	3	2	2	2	4	6	2	1	4	5	2	2	1	1	2	1	1	1	1	1	1

TABLE 2. *Classification of peptides according to similarity or difference.*

Peptides which are probably identical		Peptides with definitive differences	Peptides with undefined differences	Peptides which seldom stained
#1	15	30	8	4	22
#2	16	31	17	14	26
#3	18	32	37	19	43
#5	20	33	38	29	48
#6	21	34	39	36	55
#7	23	35	41	42	56
#9	24	40	44	45–46	57
10	25	47	49	53	60
11	27	51	50	58	62
12	28	54	52		
13		61	59		

and Lovrien (1959), then there should be a maximum of 11 tryptophan spots and 17 cysteine or methionine spots. But human catalase has only two peptides which stain consistently for tryptophan and one which stains for bivalent sulfur. On the other hand, there may be instances of false positive staining, especially with the Pauly and Sakaguchi stains. This variability in the ability of a stain to detect a given amino acid makes it impossible to make definitive statements about peptide composition on the basis of these stains. However, it is not unreasonable to expect them to produce consistent results for identical homologous peptides from species to species.

In view of these limitations, the data presented in Table 1 are most suitable for making tentative comparisons between species rather than for establishing actual composition of the peptides.

Another problem in efforts at comparisons is introduced by the fact that the large number of peptides made it impossible to identify homologous peptides with altered positions due to amino acid substitutions. In order to estimate the number of peptide differences between species, a scoring method was employed based on comparison of spots at the same position. The staining characteristics of the spot at one position were compared only to those of other spots with the same number. In

evaluating the data in Table 1, a score of ++ was considered to be different from a corresponding blank; however, a score of + was not considered to differ significantly from either ++ or blank. Also, differences in staining with ninhydrin or ninhydrin-collidine were not considered significant by themselves. The numbers of peptides differing between species obtained in this way provide an estimate of the relative differences between species. A single amino acid substitution which produces on observable change will usually affect two peptide positions. However in the case of lysine and arginine substitutions, three may be affected, whereas, in other cases only one altered peptide may be detected. Also, the larger the peptide, the more likely it is to contain more than one amino acid substitution. Therefore, the assumption must be made that these two factors are of similar magnitude in the species being compared.

When compared as outlined above, 33 of the 62 peptide positions appear to be identical in all the species examined, 11 show definitive differences, 10 others differ equivocally, and 8 did not stain frequently enough to be characterized. Table 2 lists the peptides in each group.

Considering only those peptides whose differences are reasonably clear cut, two (No. 8 and 17) are basic, six (No. 37, 38, 39, 41, 44, and 49) are neutral and three (No. 50, 52, and 59) are acidic. Of these altered peptides, numbers 8, 38, and 44, though present in all species, differ only in the baboon. These differences are defined by a frequent reaction with one of the special stains for the baboon catalase as opposed to the absence of such a reaction in all the other species. Peptides number 37 and 59 are uniquely different for man. Number 59 is absent in man but present in all monkeys examined; number 37 stains definitely with two special stains in man, but not in any of the monkeys. Numbers 17, 41, and 50 are apparently alike in man and the spider monkey but different from all the old world monkeys
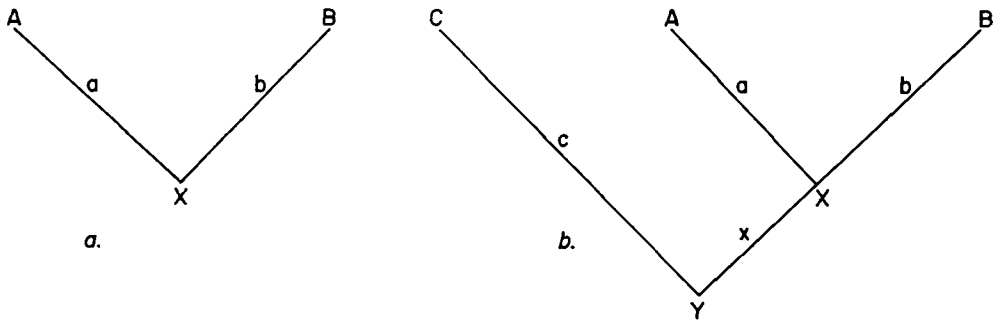
FIG. 3. Hypothetical lines of descent. See explanation in text.

which are, in turn, all alike. Peptide number 39 differs only in cynomolgus. Peptide number 49 is alike in rhesus, cynomolgus, baboon, and spider monkey but different from man and green monkey which in turn appear to be the same. Finally, number 52 is alike in man, rhesus, cynomolgus, and spider, but different in the green monkey and baboon, which are again alike.

METHOD OF STATISTICAL ANALYSIS

In hemoglobin, a protein whose amino acid sequences are known for a series of animals, partial reconstructions of its phylogenetic history have been made by comparing the amino acids at each position in the chain (Zuckerkandl, 1965). A more general method of devising phylogenetic schemes from contemporary organisms, and applicable to an entire protein, instead of single amino acid positions, could be advantageous. The following method, developed independently but then found to be based on essentially the same model as that employed by Cavalli-Sforza and Edwards (1964) for use with gene frequency data, is an attempt to do this.

If species are related to each other through descent from a single common ancestor, then a single protein in these resultant species should differ from species to species by the total of the independently acquired amino acid substitutions, less those which occur at the same position in excess of one substitution. The latter may be different substitutions occurring at the same

amino acid position in divergent lines, or convergent changes in divergent lines. In the following, these two classes are considered negligible.

In the simplest case where X is the common ancestor, one population will have undergone "a" changes to produce species A protein, another "b" changes to produce species B protein. Then, if these changes occur at different positions, the total difference (i.e. number of discordants at corresponding amino acid positions) between species A and B will be $a + b$ (Fig. 3a). The total $a + b$ is subject to experimental observation. Species X lies somewhere along the scale $a + b$, but its location can not be determined with only one observation. Addition to the scheme of species C, arising from an earlier common ancestor Y, with c the number of substitutions between Y and C, and x the number between Y and X, (Fig. 3b) allows the determination of the values a and b and hence the location, in terms of numbers of substitutions, of X. The following set of simultaneous equations expresses the relation between the observed numbers of differences and the distance between any two species:

$$
\begin{aligned}
a + b &= n_1 \\
a + x + c &= n_2 \\
b + x + c &= n_3 ,
\end{aligned}
\tag{1}
$$

where

$n_1 =$ differences between A and B
$n_2 =$ differences between A and C
$n_3 =$ differences between B and C,

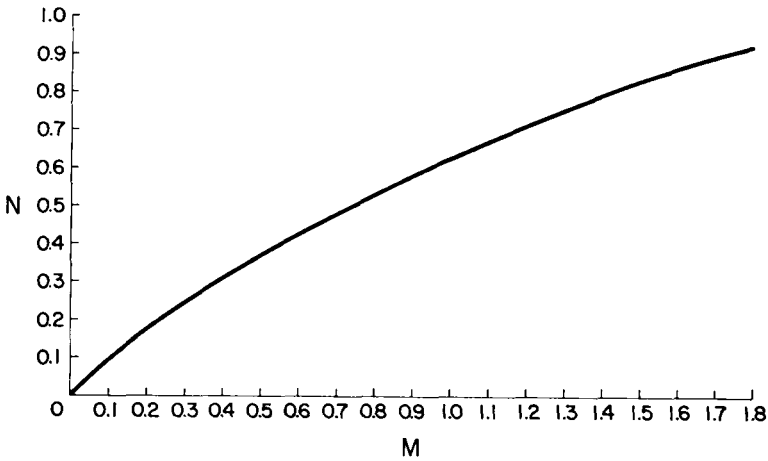and a, b, c, and x are as defined above.

FIG. 4. Graphical conversion table for N to M. See text for explanation.

Solution of these equations then will give values for a, b, and $x + c$ and permit the proper relative positioning of common ancestor X. However, there is not sufficient information to position Y properly, and only a total value for $x + c$ can be obtained. In a similar fashion, more complex schemes may be worked out, permitting the construction of a many-branched phylogenetic tree for the one protein under consideration. The construction of more complex trees requires the use of least squares methods (Kempthorne, 1952) for evaluating all the unknowns because the presence of four or more terminal points on the tree produces more equations than unknowns. In some cases there may be a perfect fit and the sum of the squares, S, then equals 0; usually however, S is greater than 0. This method will provide a means of obtaining reproducible values for each unknown, (a, b, etc.). Also if there is a question about which of many tree forms best fits the data, the S value provides a measure of goodness of fit; the smaller the value, the better the fit of the tree to those data. Trees so constructed should correspond to the organisms' phylogenetic tree, though details such as branch segments between points of divergence may be lost, and the relative lengths of the branches may differ considerably.

Within the framework of this method it is impossible to determine the proper position of the apex, or earliest point of divergence within the system. An approximation for the earliest branch point may be borrowed from paleontological evidence, or an additional assumption of equal mean numbers of changes in both lines from the point of divergence to the present may be made, and the earliest point of divergence placed so that this assumption is fulfilled. It is probably better to apply the first method, since in attempting to make the system of analysis completely independent by using the second method, evidence for real differences in rates of change may be obscured. Although with the former method, the system loses its total independence from other methods of analysis, this seems a minor loss compared with that incurred with the independent evaluation within the system.

As the percentage of differences increases (number of differences compared with the total number of units available for change) the chance that an observed difference is due to two or more independent changes at one site increases. This may be partially corrected in the following manner:

$$N = x + x^2 + x^3 + \ldots x^n \qquad (2)$$

and
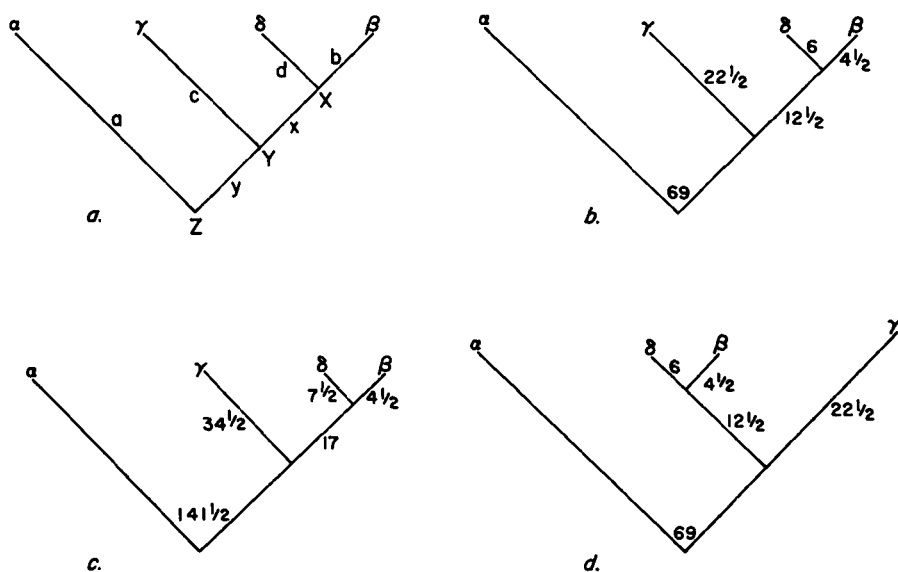$$M = kx + 2kx^2 + 3kx^3 + \ldots nkx^n, \qquad (3)$$

FIG. 5. Lines of descent for the human hemoglobin genes based on amino acid differences. (a) hypothetical scheme (b) solutions for branch segment values (c) values corrected for multiple changes at single positions (d) values for a different branching pattern.

where n_1, n_2...are the observed number of differences between two species, n being the general case; k is the total number of amino acid residues being considered

and $N = n/k$, the observed frequency of differing residues between any two species, or, the frequency of amino acid positions at which there has occurred one or more amino acid substitutions;

and x is the frequency of amino acid positions at which there has occurred one amino acid substitution.

Then x^2 is the frequency of amino acid positions at which there has oc-curred two amino acid substitutions, etc.

and M is the most probable total number of changes for all the amino acid positions under consideration.

Representative values of N and M have been calculated and plotted as shown in Figure 4, from which the M value can be read directly for any N. As values for n become larger, the difference between them and M becomes increasingly large. This correction will not alter the general shape of the phylogenetic tree obtained, it will however, alter the relative lengths of the branches.

The validity of the results obtained by this method is limited by the following assumptions: (1) no back mutation, (2) equal likelihood of substitution at any position, (3) equal likelihood of substitution of any amino acid for any other, and (4) no polymorphic systems at the time of divergence. The accuracy of the approximation will also depend on the units whose differences are being compared and the magnitude of the differences between them.

TABLE 3. *(a) Numbers of amino acid differences between human hemoglobin chains. (b) Numbers corrected for multiple changes at single positions.*

	3a			3b		
	α	β	γ	α	β	γ
β	87			165		
γ	91	39		175	55	
δ	87	10	42	165	11	61

TABLE 4. *Numbers of amino acid differences between cytochromes c of various species (Chan et al., 1963; Kreil, 1963; Narita et al., 1963; Yasunobu et al., 1963; Margoliash, 1963; Margoliash et al., 1965).*

	Man	Horse	Sheep Cow Pig	Chicken	Tuna
Horse	12				
Cow	10	3			
Chicken	11	9	7		
Tuna	21	19	17	17	
Yeast	44	44	39	39	42

APPLICATIONS OF THE METHOD OF STATISTICAL ANALYSIS

Application to Amino Acid Data

The foregoing method of analysis is best applied to amino acid sequence data. Appropriate data are available for the human hemoglobin chains, a few mammalian hemoglobin β chains and cytochrome c from a diverse series of organisms. The application of this method to the four major human hemoglobin chains is given in detail as an example.

On the basis of identical amino acid residues at many of the positions in the four major human hemoglobin chains, Ingram (1961) has postulated that their genes have all arisen from a single common ancestoral gene. These four major human hemoglobin chains most probably jointly possess 148 positions (determined by aligning the amino acid sequences side by side leaving gaps in order to maximize the number of concordant positions) though seven of these are missing in the α chain and two are missing in the β, γ, and δ chains. When the four chains are compared in all possible pairs, numerous discordant pairs are found as shown in Table 3a; values corrected for multiple changes are given in Table 3b. As Ingram (1961) has pointed out, the β and δ chains have the fewest differences, suggesting that they were the last to diverge. The β and γ, and δ and γ have an intermediate number of differences suggesting an earlier divergence of the β-δ

TABLE 5. *Numbers of differences between hypothetical peptides from the human hemoglobin chains.*

	α	β	γ
β	11		
γ	11	8	
δ	11	5	8

and γ lines. From these inferences a tentative scheme may be drawn as in Figure 5a, and the following set of simultaneous equations, based on the uncorrected values, set down (for simplicity a and y are together denoted as a):

$$a + x + b = 87$$
$$a + c = 91$$
$$a + x + d = 87$$
$$c + x + b = 39$$
$$d + b = 10$$
$$c + x + d = 42.$$

In this case there are six equations and five unknowns, so the method of least squares must be employed to obtain the best evaluation of the unknowns. In this case:

$$S = (a + x + b - 87)^2 + (a + c - 91)^2 + (a + x + d - 87)^2 + (c + x + b - 39)^2 + (d + b - 10)^2 + (c + e + d - 42)^2.$$

Differentiating with respect to each unknown and equating to zero (thus minimizing the S value) produces the following set of simultaneous equations:

$$3a + b + c + d + 2x - 265 = 0$$
$$a + 3b + c + d + 2x - 136 = 0$$
$$a + b + 3c + d + 2x - 172 = 0$$
$$a + b + c + 3d + 2x - 139 = 0$$
$$2a + 2b + c + d + 3x - 213 = 0.$$

The solution of these equations gives:

$$a = 69$$
$$b = 4\frac{1}{2}$$
$$c = 22\frac{1}{2}$$
$$d = 6$$
$$x = 12\frac{1}{2}$$
and $\qquad S = 3.$

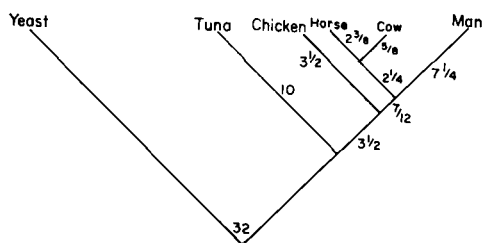Replacing these values in the scheme gives the results shown in Figure 5b, or utilizing

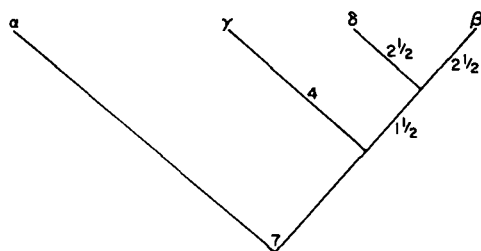FIG. 6. Lines of descent for cytochrome *c* based on amino acid differences.

FIG. 7. Lines of descent for human hemoglobin genes based on peptide differences.

the corrected values, those in Figure 5c. Note that the various lines may be pivoted around their branch points (as in Fig. 5d) so that the linear order of the terminal points may be different. It is, however, not this order that represents degrees of relationship, but the underlying branch work.

Any tree in which the branches are joined together differently will result in a higher value for S, suggesting that this scheme does represent the one which best fits the data. For example, if the δ chain is represented diverging from the δ line instead of the β line, $S = 467$, instead of $S = 3$.

If myoglobin, presumably also a descendant of the primitive hemoglobin gene (Ingram, 1961), is considered, the relative lengths of the α versus the β-δ and γ lines can be estimated. The results (25 changes in the α line and 27 in the β-δ and γ lines), based only on those amino acid residues which have been identified with a fair degree of certainty (Schroeder, 1963), indicate approximately equal rates of change in both lines, contrary to the suggestion of Buettner-Janusch and Hill (1965*b*) that the α chain is more conservative and should have acquired fewer changes than the others.

A similar analysis for cytochrome *c* is shown in Table 4 and Figure 6.

Application to Peptide Data

Amino acid sequence data is available only with obvious difficulty. Can similar results be obtained through an analysis of the component tryptic peptides of a protein, rather than of its amino acid residues? And if the results with well-identified peptides are comparable to those obtained with amino acids, will the same method work with peptides identified only to the extent to which they were in this study? The following analysis of the peptides of human hemoglobin demonstrates that in this case at least, the results are comparable. The results obtained with the catalase peptides described above are consistent with evolutionary schemes established for primates by other methods, suggesting the feasibility of analysing the evolution of single genes in this manner.

Though the human hemoglobin chains are well studied, it has not been possible to compare the peptide patterns of the four major chains on the basis of observed position and staining alone (as is done here with the catalase peptides) because of variations between reports on different chains by different workers. Therefore, the analysis presented in Figure 7 and Table 5 has been performed on hypothetical peptides, the differences having been estimated on the basis of the amino acid sequence. Though the relative lengths of the various branches is different from those obtained from the amino acid sequence (Fig. 5) the branches have the same basic relationship.

Analysis of the data of Hill et al. (1963) on peptides (as they appeared in fingerprints with aid of special stains) from the α and β chains of various primate hemo-

Man
Hylobates
Saimiri
Cacajao
Macaca
Papio   Cercopithecus Saimiri Cacajao Papio
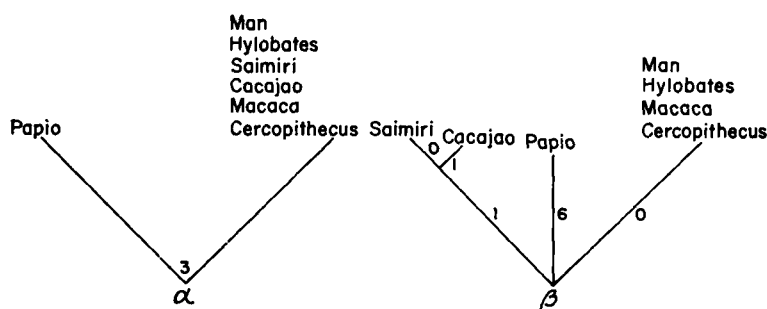
Man
Hylobates
Macaca
Cercopithecus

FIG. 8. Lines of descent for primate α and β hemoglobin chains based on tryptic peptides.

globins is presented in Table 6 and Fig. 8. Most striking is the pronounced difference of the baboon from the other old world monkeys, new world monkeys, apes, and man.

The α chain seems at this late point in its evolution to have changed considerably less than the β chain. Such a difference might be explained (in light of the evidence presented earlier for equal overall rates of the α and other line) by an early rapid change for the α chain, followed by little change after its divergence from the β line.

Table 7 gives the results of the peptide comparisons from catalase. Tabulation of the total numbers of differences is given in Table 8; the final scheme in Figure 9 gives two of the many alternative ways of pivoting the branches about their branch points. Because the data from the spider monkey are not complete, the value attached to the spider monkey branch is not representative. However, enough spots could be compared with those of man and the old world monkeys to use it as a tentative reference point for establishing the point of divergence of human catalase from that of the old world monkeys (as in using C to establish the location of X in Fig. 3). The spread at the base of the old world monkey branches is very small and small changes in the data could easily produce a scheme with somewhat different configuration at this point.

Interestingly, in both catalase and hemoglobin, the baboon protein appears to have acquired many more differences than any of the other species considered. In the other species considered, there are no marked differences for either protein.

DISCUSSION

Recent genetic and biochemical evidence suggests that there may be a need to reconsider the taxonomic arrangement of some primate species. Among the old world
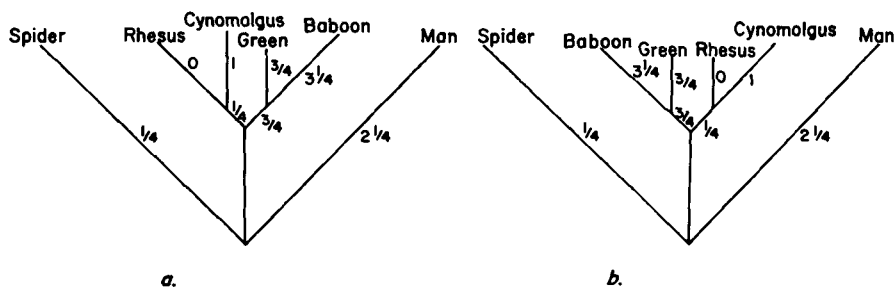
Spider   Rhesus   Cynomolgus Baboon   Man   Spider   Baboon Green Rhesus Cynomolgus Man
                        Green

a.   b.

FIG. 9. Alternative representations of the lines of descent for catalase based on tryptic peptides. The lowest vertical branch has a value of 2¾.

TABLE 6. *Numbers of differences between pep-tides from the α and β chains of various primate hemoglobins (Hill et al., 1963).*

		Man	Hylob-ates	Sai-miri	Caca-jao	Ma-caca	Cer-copi-thecus
	Hylobates	0					
	Saimiri	0	0				
α	Cacajao	0	0	0			
	Macaca	0	0	0	0		
	Cercopithecus	0	0	0	0	0	
	Papio	3	3	3	3	3	3
	Hylobates	0					
	Saimiri	1	1				
β	Cacajao	2	2	1			
	Macaca	0	0	1	2		
	Cercopithecus	0	0	1	2	0	
	Papio	6	6	7	8	6	6

monkeys preliminary data suggest that *Papio* is a more distant cousin to *Macaca* and *Cercopithecus* than is apparent from the accepted taxonomic scheme in which they are each a separate genus in the sub-family Cercopithecinae (Williams, 1964). On the other hand, Pilbeam and Simons (1965) cite results of breeding experiments

TABLE 8. *Numbers of differences between cata-lase tryptic peptides.*

	Man	Rhesus	Cyno-molgus	Green	Baboon
Rhesus	6				
Cynomolgus	7	1			
Green	6	2	3		
Baboon	10	4	5	4	
Spider	3	3	4	5	7

which suggest that *Macaca, Cercocebus,* and *Papio* should all be classified in a single genus *Papio* because of their ability to produce hybrids. With conflicting re-sults such as these, insight is gained, not so much into the actual relationships, as into the criteria for defining a genus and species. As a more accurate reflection of genetic constitution, biochemical evidence has great potential, but at this point it would be premature to draw any sweeping taxonomic conclusions from it.

However, biochemical data can yield valuable information about rates of change in systems very near the actual genetic material. If both the formation of hybirds

TABLE 7. *Catalase peptide differences. L symbolizes likeness of the corresponding peptides of the two species in question; U represents a difference.*

Peptide No. 8, 38, 44						Peptide No. 39					
	Man	Rh	Cyn	Green	Bab		Man	Rh	Cyn	Green	Bab
Rh	L					Rh	L				
Cyn	L	L				Cyn	U	U			
Green	L	L	L			Green	L	L	U		
Bab	U	U	U	U		Bab	L	L	U	L	
Spi	L	L	L	L	U	Spi	L	L	U	L	L

Peptide No. 37, 59						Peptide No. 49					
	Man	Rh	Cyn	Green	Bab		Man	Rh	Cyn	Green	Bab
Rh	U					Rh	U				
Cyn	U	L				Cyn	U	L			
Green	U	L	L			Green	L	U	U		
Bab	U	L	L	L		Bab	U	L	L	U	
Spi	U	L	L	L	L	Spi	U	L	L	U	L

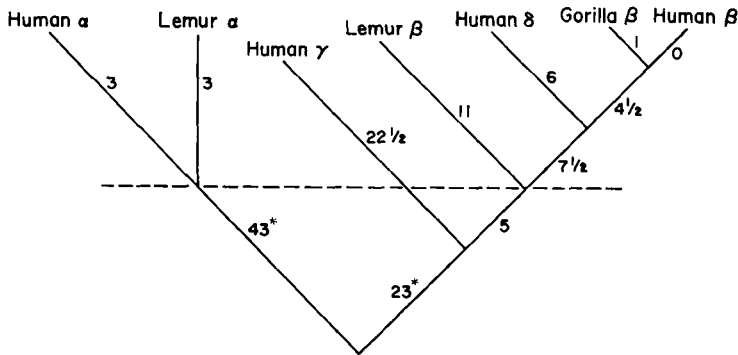| Peptide No. 17, 41, 50						Peptide No. 52					
	Man	Rh	Cyn	Green	Bab		Man	Rh	Cyn	Green	Bab
Rh	U					Rh	L				
Cyn	U	L				Cyn	L	L			
Green	U	L	L			Green	U	U	U		
Bab	U	L	L	L		Bab	U	U	U	L	
Spi	L	U	U	U	U	Spi	L	L	L	U	U

Fig. 10. Composite diagram of lines of descent of hemoglobin chains from man and lemur. The dotted line is drawn through the two points at which the human and lemur α and β chains diverged. Presumably these two events occurred at approximately the same time. These values have been obtained by estimation from myoglobin data (Schroeder, 1963).

and significant differences in the serum proteins in *Papio* and *Macaca* can be verified, this may be evidence for greatly different rates of evolution in different systems [i.e., mosaic evolution (Pilbeam and Simons, 1965)]. It would suggest that in the baboon, as opposed to the other cercopithecoids, the genetic isolating mechanisms have been slow to evolve, whereas, there has apparently been relatively rapid evolution of some proteins (notably hemoglobin and catalase) in *Papio* when compared with *Macaca* and *Cercopithecus*. If this is further verified by amino acid analysis this does not necessarily mean that the taxonomic status of the baboon need be changed, or that it is more distantly related, in terms of time elapsed from the common ancestor of this subfamily. It does, however, indicate that certain proteins have evolved at a more rapid rate in the baboon than in its cousin species.

As shown in the analysis of hemoglobin amino acid substitutions, the human hemoglobin chains appear to have acquired amino acid substitutions at about the same rate, with the possible exception of the γ chains which may have evolved at a somewhat faster rate. This is contrary to Buettner-Janusch and Hill's (1965a) calculation based on human and *Lemur fulvus* data that the α and β chains differ in rate

of change and that the α chain has been a conservative element in hemoglobin evolution. This discrepancy is explicable if α chains evolved very rapidly following their divergence from the myoglobin gene, and that once established, there was a slower rate of change. If a composite diagram (Fig. 10) is made for human and lemur hemoglobin chains, an imaginary time line can be drawn through the points of divergence of the two species lines in the α and β chains. Pictorially this suggests that indeed, since the appearance of the lemur, the α chain has been conservative in its evolutionary rate, but that its early development was not. This diagram is suggestive of more rapid evolution of the α chain than of the β-γ-δ line during most of their history, while in the primates the α appears to have been stabilized while the γ, β, and δ chains have changed relatively rapidly. On the basis of amino acid data, there appears to be no difference in the rate of change of human, gorilla, or lemur β hemoglobin chains.

There does then seem to be some evidence, based on amino acid sequence analysis, for differences in rates of substitution in different evolutionary lines. Zuckerkandl (1965) has postulated that " 'living fossils' such as the cockroach, the horseshoe crab, the shark, and . . . the lemur probably

manufacture a great many polypeptide molecules that differ only slightly from those manufactured by their ancestors millions of years ago. . . . it is unlikely that selective forces would favor the stability of morphological characteristics without at the same time favoring the stability of biochemical characteristics, which are more fundamental." This, despite the argument that both "primitive" and "advanced" forms have been "at risk" for evolutionary changes equally long. With the data cited above, if the β hemoglobin chain of *L. fulvus* is compared with that of man, there is no difference between their rates of change. Additional data should prove most interesting with respect to this point. Though these data suggest that there are differences in substitution rates of various lines, the low rates, so far, do not correlate with close morphological resemblance to ancestral forms as exemplified by the lemur. If this observation is borne out, it will contribute more evidence for mosaic evolution in terms of morphological versus biochemical traits.

For the purpose of comparing proteins, the results obtained with peptide data would be expected to be considerably more tentative than those obtained with amino acid data. In order for a peptide change to be comparable from one protein to another, peptides must be of the same average size and composition. Furthermore, all the differing peptides should be identified and corrections must be made for different sizes of the proteins. Because of these problems, meaningful comparisons of rates of change based on peptide data can not be made at this time.

In the foregoing analysis, there are examples of variation in the rate of fixation of mutant genes in different proteins and in different species. At present, many of these substitutions appear to be neutral. Hence, an explanation of these changes must await further elucidation of their roles in their respective organisms as well as a re-examination of our concept of the role of natural selection. For, if these changes are neutral, their fixation can not be accounted for by simple selection and if they are not neutral, selective forces must have discriminating powers of which we are not yet aware.

SUMMARY

Comparisons of the tryptic peptide patterns of erythrocyte catalase from man, baboon, and cynomolgus, rhesus, green, and spider monkeys have been made by utilizing both position and reaction with stains specific for particular amino acids. The results of these comparisons indicate very few differences between rhesus and cynomolgus (which are both species of the same genus), intermediate numbers of differences among the rest of the old world monkeys, large numbers of differences between man and the old world monkeys, and an even larger number of differences between man and baboon.

A method has been adopted whereby an evolutionary family tree, with approximations of the number of changes in each branch can be constructed, based on this comparison of differences between contemporary species. Results of this analysis, done not only with the catalase peptide data from this study, but also with amino acid sequence data for hemoglobin and cytochrome *c* and primate hemoglobin peptide data suggest that: (1) evolutionary change, when measured in units of amino acid substitutions, has not always occurred at a constant rate in a given evolutionary line, (2) some phylogenetic lines may acquire evolutionarily effective mutations more rapidly than others during the same period of time, and (3) close morphological resemblance to ancestral forms need not be accompanied by relatively few changes in protein structure.

LITERATURE CITED

BAUR, E. W. 1963. Catalase abnormality in a

Caucasian family in the United States. Science **140:** 816–817.

BRAUNITZER, G., F. BRAUN, K. HILSE, G. HOBOM, V. RUDLOFF, AND G. VON WETTSTEIN. 1965. Constancy and variability of protein structure in respiratory and viral proteins. *In:* Evolving genes and proteins, V. Bryson and H. Vogel, Eds. New York: Academic Press.

BUETTNER-JANUSCH, J., AND R. L. HILL. 1965*a*. Molecules and monkeys. Science **147:** 836–842.

—— AND ——. 1965*b*. Evolution of hemoglobin in primates. *In:* Evolving genes and proteins, V. Bryson and H. Vogel, Eds. New York: Academic Press.

CAVALLI-SFORZA, L. L., AND A. W. F. EDWARDS. 1964. Analysis of human evolution. *In:* Genetics today. p. 923–933. Oxford: Pergamon Press.

CHAN, S. K., S. B. NEEDLEMAN, J. W. STEWART, O. F. WALASEK, AND E. MARGOLIASH. 1963. Amino acid sequences of various cytochromes *c*. Federation Proc. **22:** 658.

CHERNOFF, A. I., AND J. C. LIU. 1961. The amino acid composition of hemoglobin. II. Analytical technics. Blood **17:** 54–70.

HERBERT, T., AND T. PINSET. 1948. Crystalline human erythrocyte catalase. Biochem. J. **43:** 203.

HILL, R. L., J. BUETTNER-JANUSCH, AND V. BUETTNER-JANUSCH. 1963. Evolution of hemoglobin in primates. Proc. Nat. Acad. Sci. **50:** 885–893.

INGRAM, V. M. 1961. Gene evolution and the hemoglobins. Nature (Lond.) **189:** 704–708.

KEMPTHORNE, O. 1952. Design and analysis of Experiments. p. 30–32. New York: John Wiley and Sons.

KREIL, G. Z. 1963. Uber die Artspezifitat von Cytochrome; Vergleich der Aminosaure sequenz des Thunfisch-Cytochroms *c* mit der des Pferde-Cytochroms *c*. Hoppe Seyler Z. Physiol. Chem. **334:** 154–166.

MARGOLIASH, E. 1963. Primary structure and evolution of cytochrome *c*. Proc. Nat. Acad. Sci. **50:** 672–679.

MARGOLIASH, E., AND E. L. SMITH. 1965. Structural and functional aspects of cytochrome *c*

in relation to evolution. *In:* Evolving genes and proteins, V. Bryson and H. Vogel, Eds. New York: Academic Press.

NARITA, K., K. TITANI, Y. YAO, H. MURAKAMI, M. KIMURA, AND J. VANECEK. 1963. Peptides from a tryptic digest of baker's yeast cytochrome *c*. Biochim. Biophys. Acta **73:** 670–673.

PILBEAM, D. R., AND E. L. SIMONS. 1965. Some problems of hominid classification. Amer. Sci. **53:** 237–259.

SCHROEDER, W. A. 1963. The hemoglobins. Ann. Rev. Biochem. **32:** 301–320.

SCHROEDER, W. A., J. R. SHELTON, J. B. SHELTON, AND B. M. OLSON. 1964. Some amino acid sequences in bovine-liver catalase. Biochim. Biophys. Acta **89:** 47–65.

SWENSON, R. T., R. L. HILL, H. LEHMANN, AND R. T. S. JIM. 1962. A chemical abnormality in hemoglobin G from Chinese individuals. J. Biol. Chem. **237:** 1517–1520.

TANFORD, C., AND R. E. LOVRIEN. 1959. Abstracts, American Chemical Society, 15th Meeting. Boston, April 5–10, p. 14C.

THORUP, O. A., W. B. STROLE, AND B. S. LEAVELL. 1961. A method for the localization of catalase on starch gels. J. Lab. Clin. Med. **58:** 122–128.

TOENNIES, G., AND J. J. KOLB. 1951. Techniques and reagents for paper chromatography. Anal. Chem. **23:** 823–826.

WILLIAMS, C. A. 1964. Immunochemical analysis of serum proteins of the primates: A study in molecular evolution. *In:* Evolutionary and genetic biology of primates, J. Buettner-Janusch, Ed. New York: Academic Press.

YASUNOBU, K. T., T. NAKASHIMA, AND H. HIGA. 1963. The amino acid sequence of bovine heart cytochrome c. Biochim. Biophys. Acta **78:** 791–794.

ZUCKERKANDL, E. 1965. The evolution of hemoglobin. Sci. Amer. **212:** 110–118.

ZUCKERKANDL, E., AND L. PAULING. 1965. Evolutionary divergence and convergence in proteins. *In:* Evolving genes and proteins, V. Bryson and H. Vogel, Eds. New York: Academic Press.