Amino Acid Sequence of the Hemoglobin α Chain from a Baboon (*Papio cynocephalus*): A Product of Gene Fusion?†

Walter C. Mahoney and Peter E. Nute*

ABSTRACT: The amino acid sequence of the α chain from the adult hemoglobin of a baboon, *Papio cynocephalus*, was determined by automated sequencing of the intact chain and selected fragments generated by cleavage at the single aspartylprolyl linkage and at methionyl, glutamyl, lysyl, and arginyl residues. The chain is notable in the extent to which it differs in sequence from the corresponding chains of man and several Old World monkeys. Comparison of the human α chain with that of *P. cynocephalus* yields 11 differences in sequence, while comparison of the latter with those macaque α chains that have been fully characterized yields 9–11 differences. The similarity in the extent of divergence of baboon from macaque and human α chains is puzzling. The dispersion of structural differences (most human–baboon substitutions fall within positions 1–57) suggests that one or more events other than the accumulation of point mutations occurred during the evolutionary history of the α chains of higher primates. Perhaps crossing-over between nonallelic genes, one of which was highly divergent in nucleotide sequence, resulted in replacement of a portion of the more conservative α gene by a portion of the divergent one, thereby producing a hybrid gene that encodes α-chain structure in extant baboons.

Comparisons of the primary structures of chains from primate hemoglobins have led to several generalizations, among them that α chains have sustained fewer changes than β chains during the course of primate evolution (Hill & Buettner-Janusch, 1964; Matsuda, 1976), that rates of globin evolution have decelerated in lineages of higher primates (Goodman et al., 1975), and that the extent of structural divergence between comparable chains from members of different species tends to reflect the overall extent of phylogenetic relationships among the species in question (Dayhoff, 1972). Furthermore, it has been noted that the sequences of several homologous proteins from members of various taxa differ to an extent that is roughly proportional to the time elapsed since members of these taxa are estimated to have last shared a common ancestry (Dayhoff, 1972; Cronin, 1977).

Although the preceding statements may apply to hemoglobins and other proteins from a wide variety of primates, the results of recent analyses of the hemoglobin α chains from several Old World monkeys (Cercopithecoidea) clearly illustrate that the generalizations cited above must not be indiscriminately employed in the assessment of phylogenetic relationships among different species.

The cercopithecoid α chains that have been fully sequenced differ from the human α chain (Dayhoff, 1972) at but three (*Presbytis entellus*) (Matsuda et al., 1973a), four (*Macaca mulatta* and *Macaca fuscata*) (Matsuda et al., 1970, 1973b), or five (*Macaca nemestrina* and *Cercopithecus aethiops*) (Mahoney & Nute, 1979; Matsuda et al., 1973c) positions in their sequences. However, partial characterization of the α chains from many additional cercopithecoids (all of which are species of the genera *Cercocebus*, *Theropithecus*, and *Mandrillus*) has disclosed wider degrees of structural divergence from the human α chain, ranging from no less than 7 to at least 12 substitutions (Hewett-Emmett et al., 1976). Similarly, partial analysis of the α chain from *Papio cynocephalus* [often called *P. anubis* but, as argued by Thorington (Thorington & Groves, 1970), better designated *P. cynocephalus*] indicated that this chain differs from its human counterpart at no less than 11 sites (Sullivan et al., 1976). These data demonstrate that cercopithecoid α chains vary significantly in the extent to which they diverge from the human α chain, even though the species of Old World monkeys in question are generally presumed to share a common ancestry.

In the interest of better defining the nature and extent of structural dissimilarities between the α chains of cercopithecoids, we present the complete primary structure of the α chain from *P. cynocephalus*.

Experimental Procedures

Erythrocytes from a single, adult animal were washed 3 times with saline and lysed by mixing with 3 volumes of cold, deionized water. After centrifugation to remove cellular debris, globin was precipitated in cold acid–acetone (Rossi-Fanelli et al., 1958). The α chains were isolated on columns of carboxymethylcellulose (Whatman CM-23, Advanced Fibrous) using 8 M urea buffers (Clegg et al., 1968). The chains were desalted in 2.5 × 45 cm columns of Sephadex G-25 (fine), equilibrated and developed with 9% formic acid, and recovered by lyophilization.

Preparation and Isolation of α-Chain Fragments. Following S-pyridylethylation (Hermodson et al., 1973), 93 mg of α chains was cleaved at methionyl residues by cyanogen bromide (Nute & Mahoney, 1979a). An additional 35 mg of *S*-pyridylethyl α chains was cleaved at the single aspartylprolyl bond by 70% formic acid (Nute & Mahoney, 1979a), while 65 mg of pyridylethylated α chains was digested at glutamyl residues by *Staphylococcus aureus* strain V8 protease (Miles Laboratories) (Mahoney & Nute, 1979). Products of the cleavage by cyanogen bromide were isolated by gel filtration in a 2.5 × 195 cm column of Sephadex G-50 (superfine), using 9% formic acid as the eluant. Products generated by cleavages at the Asp–Pro bond and at glutamyl residues were isolated by gel filtration in 2.5 × 95 cm columns of Sephadex G-50 (superfine) that were equilibrated and developed with 5% acetic acid, 1% (by volume) in butanol. The desired Glu fragment (Glu-5) was further purified by a second passage through the same column.

† From the Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907 (W.C.M.), and the Departments of Anthropology (DH-05) and Medicine (Division of Medical Genetics) and the Regional Primate Research Center, University of Washington, Seattle, Washington 98195 (P.E.N.). *Received October 30, 1979.* This work was supported by Grants HL 20961 and RR 00166 from the National Institutes of Health.

0006-2960/80/0419-1529$01.00/0

The fragments generated by each of the above cleavage reactions are numbered consecutively from those containing the amino-terminal sequence to those containing the carboxyl-terminal sequence of the intact α chain. Thus, three fragments, CB-1 (residues 1–32), CB-2 (33–76), and CB-3 (77–141), were produced by cleavage with cyanogen bromide. Cleavage at Asp–Pro produced fragments AP-1 (1–94) and AP-2 (95–141), while cleavage at glutamyl residues was expected to generate six fragments, the largest of which was Glu-5 (31–116).

Subdigestion of Fragments CB-3 and Glu-5. Approximately 28 mg of the preparation of fragment CB-3 was subdigested at the sole glutamyl residue by *S. aureus* strain V8 protease (Mahoney & Nute, 1979), and one of the two products, Glu-6 (residues 117–141), was isolated by gel filtration in a 2.5 × 95 cm column of Sephadex G-50 (superfine), equilibrated and developed with 9% formic acid; this fragment was further purified by a second passage through the same column.

Subdigestion of fragment Glu-5 (22 mg) at arginyl and lysyl residues by DCC–trypsin[1] (Sigma) yielded a mixture of peptides, the largest of which was designated T-6 (residues 62–82). Peptide T-6 was isolated by gel filtration in a 2.5 × 90 cm column of Sephadex G-25 (superfine), using 9% formic acid as the eluant, and was further purified by a second passage through the same column. Tryptic peptides were detected after alkaline hydrolysis of 100-μL aliquots of effluent fractions and addition of *o*-phthalaldehyde (Durrum) as described previously (Lai, 1977; Mahoney & Nute, 1979). The intensity of fluorescence of each reaction mixture was measured in an Aminco-Bowman spectrophotofluorometer with excitation set at 340 nm and emission at 455 nm.

Amino Acid Analyses. Samples of *S*-pyridylethyl α chains and fragments derived therefrom were hydrolyzed at 110 °C in sealed, evacuated tubes with glass-distilled, 6 N HCl for 24 h (fragments) or for 24, 48, 72, 96, and 120 h (whole chains). Analyses were carried out on a Durrum D-500 amino acid analyzer according to the manufacturer's instructions. The color value for *S*-(pyridylethyl)cysteine, which is eluted as a discrete peak between ammonia and arginine, is 1.2 times that for leucine (Friedman et al., 1970).

Sequence Analyses. Intact, S-pyridylethylated α chains and selected fragments (3–6 mg) were degraded in a Beckman Model 890 C sequencer according to the procedure of Edman & Begg (1967) as modified by Hermodson et al. (1972), using the "peptide program" of Hermodson et al. (1977). Products generated by the sequencer were converted to their Pth derivatives (Hermodson et al., 1972) and identified as previously described (Nute & Mahoney, 1979a; Mahoney & Nute, 1979). From the yields of those silylated Pth amino acids that can be quantified by gas–liquid chromatography (e.g., alanine, glycine, leucine, and valine), it was calculated that 60% or more of the amount of each derivative expected (based on the amount and size of the fragment degraded) was recovered after each cycle. No Pth amino acids were identified when, in high-performance liquid chromatography, the ratio of peak to background fell below 3. Stepwise yields for the degradations ranged from 94 to 96% and, in every case, only one sequence was observed.

Identification of the Carboxyl-Terminal Residue. The first 24 residues of fragment Glu-6 were identified as described above. Following cleavage of the phenylthiocarbamoyl derivative of the penultimate residue (tyrosine), the 24th cycle was stopped before the carboxyl-terminal residue could be coupled with PITC. The material remaining in the cup of the sequencer was taken up in 50% acetic acid and identified, without prior hydrolysis, by amino acid analysis. This entire procedure was repeated, using a second portion of the preparation of fragment Glu-6.

[1] Abbreviations used: DCC, diphenylcarbamyl chloride; Pth, phenylthiohydantoin; PITC, phenyl isothiocyanate.

	5	10	15
1	Val Leu Ser Pro Asp Asp Lys Lys His Val Lys Ala Ala Trp Gly		
16	Lys Val Gly Glu His Ala Gly Glu Tyr Gly Ala Glu Ala Leu Glu		
31	Arg Met Phe Leu Ser Phe Pro Thr Thr Lys Thr Tyr Phe Pro His		
46	Phe Asp Leu Ser His Gly Ser Asp Gln Val Asn Lys His Gly Lys		
61	Lys Val Ala Asp Ala Leu Thr Leu Ala Val Gly His Val Asp Asp		
76	Met Pro Gln Ala Leu Ser Lys Leu Ser Asp Leu His Ala His Lys		
91	Leu Arg Val Asp Pro Val Asn Phe Lys Leu Leu Ser His Cys Leu		
106	Leu Val Thr Leu Ala Ala His Leu Pro Ala Glu Phe Thr Pro Ala		
121	Val His Ala Ser Leu Asp Lys Phe Leu Ala Ser Val Ser Thr Val		
136	Leu Thr Ser Lys Tyr Arg		

FIGURE 1: The complete amino acid sequence of the α chain from *P. cynocephalus*.

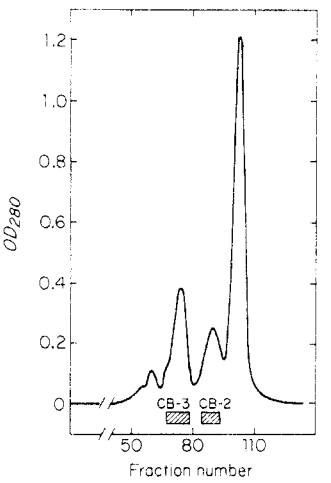

FIGURE 2: Gel filtration of fragments produced by cleavage of the *S*-pyridylethyl α chains at methionyl residues. Only fragments CB-2 and CB-3 were recovered.

Results

The amino acid sequence of the α chain from *P. cynocephalus* appears in Figure 1. Overlaps of partial sequences ranged from 4 to 14 residues and allowed unambiguous placement of all fragments. Furthermore, the close correspondence of amino acid compositions of the whole chain and fragments derived therefrom with compositions determined from the sequence indicates that the sequence, as presented, is complete (see supplementary material paragraph at end of paper).

Isolation and Identification of Fragments. Elution profiles produced by gel filtration of CB, AP, and Glu fragments from intact, *S*-pyridylethyl α chains appear in Figures 2, 3, and 4, respectively. The isolation of fragment Glu-6 from a digest of CB-3 at glutamic acid appears in Figure 5, while that of fragment T-6, produced by cleavage of Glu-5 by trypsin, is

Table I: Amino Acid Compositions of the α Chain and Selected Fragments[a]

amino acid	residues: yield (%):	whole chain[b] 1-141	CB-2 33-76 88	CB-3 77-141 76	AP-2 95-141 60	Glu-6 117-141 45	T-6 62-82 70
Asp		11.9 (12)	5.6 (6)	4.1 (4)	2.0 (2)	1.0 (1)	3.3 (3)
Thr		7.8[c] (8)	3.8 (4)	3.5 (6)	3.7 (4)	2.6 (3)	0.7 (1)
Ser		10.7[c] (11)	2.9 (3)	6.6 (7)	4.7 (5)	3.8 (4)	0.8 (1)
Glu		7.0 (7)	1.2 (1)	2.2 (2)	0.7 (1)	0.1 (0)	0.9 (1)
Pro		6.8 (7)	1.8 (2)	4.1 (4)	2.6 (3)	0.8 (1)	0.9 (1)
Gly		6.9 (7)	3.1 (3)	0.2 (0)			1.0 (1)
Ala		15.9 (16)	3.0 (3)	8.0 (8)	5.7 (6)	2.7 (3)	3,8 (4)
Val		12.6[d] (13)	4.1 (4)	6.0 (6)	4.9 (5)	2.8 (3)	2.7 (3)
Met		1.7 (2)	0.9[e] (1)	0.1[e] (1)			0.7 (1)
Leu		18.9[d] (19)	3.6 (4)	12.5 (13)	8.6 (9)	2.6 (3)	2.6 (3)
Tyr		2.9 (3)	0.8 (1)	1.1 (1)	0.8 (1)	1.1 (1)	
Phe		7.1 (7)	3.8 (4)	3.1 (3)	3.0 (3)	2.0 (2)	
His		10.6 (11)	3.8 (4)	4.9 (5)	3.1 (3)	1.0 (1)	0.8 (1)
Lys		13.2 (13)	3.6 (4)	4.9 (5)	3.1 (3)	1.8 (2)	0.8 (1)
Arg		3.0 (3)		1.8 (2)	1.0 (1)	1.0 (1)	
Cys[f]		1.2 (1)		0.8 (1)	0.8 (1)		
Trp[g]		ND (1)	ND (0)	ND (0)	ND (0)	ND (0)	ND (0)

[a] Values in parentheses refer to compositions determined from the sequence. [b] Averages over duplicate 24-, 48-, 72-, 96-, and 120-h hydrolyses, except where otherwise noted. [c] Extrapolated to zero time. [d] Average of duplicate 120-h hydrolyses. [e] Detected as homoserine and homoserine lactone. [f] Detected as *S*-(pyridylethyl)cysteine. [g] ND, not determined.

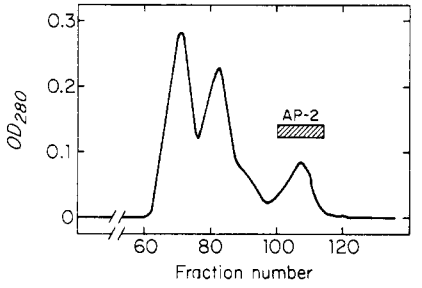

FIGURE 3: Gel filtration of fragments produced by cleavage of intact, *S*-pyridylethyl α chains at the single Asp–Pro bond by 70% formic acid. Only fragment AP-2 was recovered.

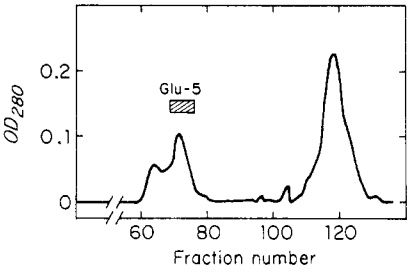

FIGURE 4: Isolation of fragment Glu-5, produced by cleavage of intact, *S*-pyridylethyl α chains at glutamyl residues, by gel filtration.

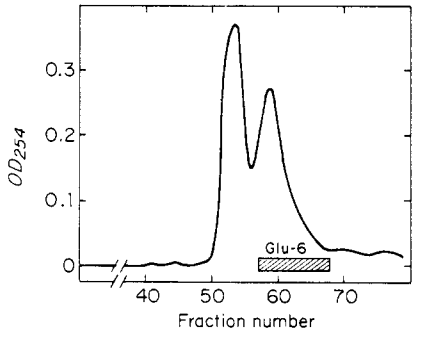

FIGURE 5: Gel filtration of fragments generated by subdigestion of fragment CB-3 at glutamic acid. The carboxyl-terminal product (Glu-6) was further purified by a second passage through the same column.

presented in Figure 6, Amino acid compositions of the whole *S*-pyridylethyl α chain and those fragments subjected to se-

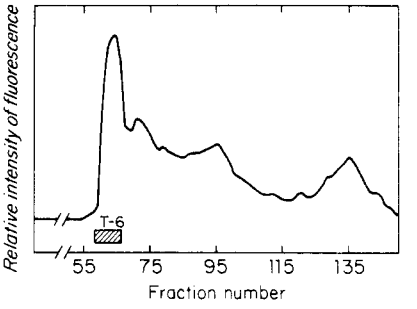

FIGURE 6: Isolation, by gel filtration, of peptide T-6, produced by tryptic subdigestion of fragment Glu-5. The ordinate is in arbitrary units. This peptide was further purified by a second passage through the same column.

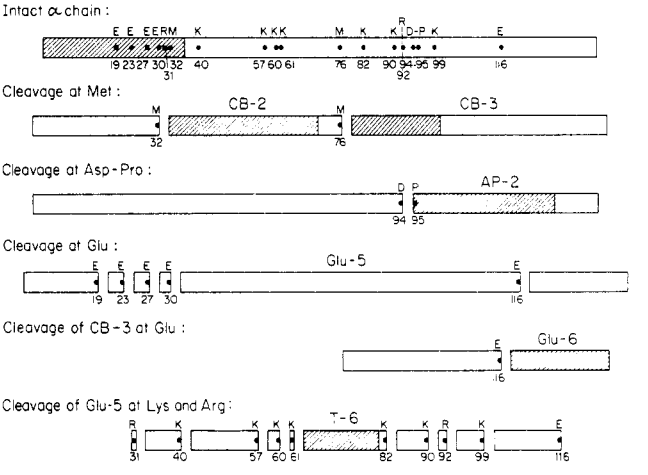

FIGURE 7: Summary of the sequencing strategy. Crosshatching denotes those portions of the intact chain and selected fragments which were degraded. Standard, one-letter abbreviations designate residues sensitive to the various methods of cleavage employed. Fragments CB-2, CB-3, AP-2, Glu-6, and T-6 were recovered and identified by amino acid analysis. Fragment Glu-5 was identified by sequence analysis of its five amino-terminal residues. Production of the remaining fragments is inferred from the primary structure of the α chain.

quence analysis (CB-2, CB-3, AP-2, Glu-6, and T-6) appear in Table I. Fragment Glu-5 was identified by determining the first five residues in its sequence; the composition of this fragment was not determined.

Table II: Structural Comparisons of Human and Papionine α Chains[a]

species	residue no. 5	8	9	12	19	23	53	56	57	68	71	78	82	133
Homo sapiens	Ala	Thr	Asn	Ala	Ala	Glu	Ala	Lys	Gly	Asn	Ala	Asn	Ala	Ser
M. mulatta	Ala	Ser	Asn	Ala	Gly	Glu	Ala	Lys	Gly	Leu	Gly	Asn	Ala	Ser
M. fuscata	Ala	Ser	Asn	Ala	Gly	Glu	Ala	Lys	Gly	Leu	Gly	Asn	Ala	Ser
M. nemestrina														
α^{I}	Ala	Thr	Asn	Ala	Gly	Glu	Ala	Lys	Gly	Leu	Asp	Gln	Ala	Gly
α^{II}	Ala	Thr	Asn	Ala	Gly	Glu	Ala	Lys	Gly	Leu	Asp	His	Ala	Gly
T. gelada[b]	*Asp*	*Lys*	*His*	Asp	*Glu*	Gln	*Asp*	Lys	*Lys*	Leu	Gly	Gln	*Lys*	Ser
P. cynocephalus	*Asp*	*Lys*	*His*	Ala	*Glu*	Glu	*Asp*	Asn	*Lys*	Leu	Gly	Gln	*Lys*	Ser

[a] See text for references. Only those positions at which structural differences occur are listed. Those residues by which the chains from both *T. gelada* and *P. cynocephalus* differ from the human and all macaque α chains are italicized. [b] The sequence of the α chain from *T. gelada* has been inferred from the amino acid compositions of small peptides (Hewett-Emmett et al., 1976).

Sequence Analysis. The strategy followed in determining the amino acid sequence of the α chain is summarized in Figure 7. The residues occupying positions 1–36 were identified by stepwise degradation of the complete chain. Analyses of fragments CB-2 and CB-3 established the sequences from positions 33–70 and 77–99, respectively. The sequence obtained from fragment AP-2 overlapped that from fragment CB-3 by five residues and continued through position 130. Determination of residues 62–80 (in fragment T-6) provided an overlap of nine residues with the sequence obtained from CB-2 and an overlap of four residues with that from CB-3. Residues 117–140 were placed by analysis of Glu-6, which overlapped the sequence from CB-3 by 14 residues. The penultimate residue of Glu-6 was identified as tyrosine by gas–liquid chromatography. The material removed from the cup of the sequencer after extraction of the anilinothiozolinone derivative of tyrosine yielded 45.7 nmol of tyrosine and 381.7 nmol of arginine upon amino acid analysis. Repetition of this procedure, using a second portion of the preparation of fragment Glu-6, yielded 93.8 nmol of tyrosine and 520.6 nmol of arginine. Thus, we conclude that arginine occupies the carboxyl-terminal site of the α chain from *P. cynocephalus*.

Discussion

Within the Cercopithecoidea, the genera *Papio*, *Mandrillus*, *Macaca*, *Cercocebus*, and *Theropithecus* constitute a natural group, all members of which have a diploid number of 42 chromosomes (Napier & Napier, 1967). Various combinations of these five genera have been set apart from other genera of Old World monkeys by their placement in the Tribe Papionini, thereby further attesting to their broad systematic affinities. There being no consensual definition of this taxon, we shall use the term in the interest of expedience and refer to members of the five genera cited above as papionines.

The results of several studies of papionine proteins stress the extent to which various papionines are equally divergent from man. Partial sequences (117 of 259 residues) of the carbonic anhydrase isozyme CA I (or CA B) from *M. mulatta* and *P. cynocephalus* differ at but three sites, while they have diverged from the comparable portion of the human isozyme at nine and seven sites, respectively (Tashian et al., 1975). The myoglobins from *P. cynocephalus* and *Macaca fascicularis* appear to differ at a single site, while three and four substitutions, respectively, distinguish these proteins from human myoglobin (Romero-Herrera & Lehmann, 1972). In addition, the sequence of the γ_{slow} chain from the major fetal hemoglobin of *M. nemestrina* is identical with those of the two γ chains found in *P. cynocephalus* at all save one or two positions (Nute & Mahoney, 1979a,b); these three papionine γ chains differ from the human $^{A}\gamma$ and $^{G}\gamma$ chains (Schroeder et al., 1963, 1968) at from three to five sites. Finally, the β chains from

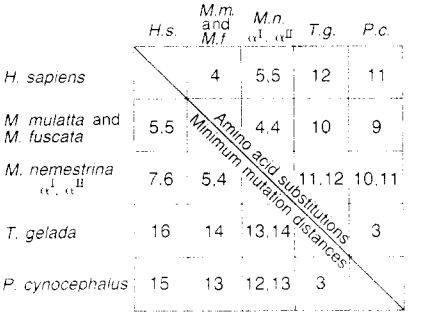

FIGURE 8: Numbers of structural differences among the α chains of *Homo sapiens* and various papionines (above diagonal) and minimum numbers of nucleotide substitutions by which the corresponding α-chain genes differ from each other (below diagonal).

the hemoglobins *M. mulatta* (Matsuda et al., 1970), *M. fuscata* (Matsuda et al., 1973b), *M. fascicularis* (Wade et al., 1970), and *M. nemestrina* (Nute & Pataryas, 1974) differ from each other and from those of *Theropithecus gelada* (Hewett-Emmett et al., 1976) and *P. cynocephalus* (P. E. Nute and W. C. Mahoney, unpublished sequence) by no more than four substitutions, while all of these papionine β chains differ from that of man at from seven to nine sites. The overall extent of genetic similarity between *Papio* and *Macaca*, as suggested by the above comparisons, is underscored when one notes that many hybrids have been produced, in captivity, by crosses between various species of these two genera (Buettner-Janusch, 1966).

To date, the only papionine α chains that have been fully sequenced are those of *M. mulatta* (Matsuda et al., 1970), *M. fuscata* (Matsuda et al., 1973b), *M. nemestrina* (Mahoney & Nute, 1979), and *P. cynocephalus*. In addition, all residues in the α chain of *T. gelada* have been placed by inference of sequence from compositions of small peptides (Hewett-Emmett et al., 1976). These sequences are compared with that of the human α chain (Dayhoff, 1972) in Table II; the numbers of differences in amino acid sequences, as well as the minimum numbers of nucleotide substitutions required to effect these differences, appear in Figure 8.

Clearly, macaque α chains differ little from that of man, the extent of divergence ranging from four to five residues (corresponding to minimum mutation distances of from five to seven). Conversely, the α chains of *P. cynocephalus* and *T. gelada* differ from the human chain at 11 and 12 sites, respectively (corresponding to minimum mutation distances of 15 and 16). Given that the partial sequences inferred for the α chains from species of *Cercocebus* and *Mandrillus* (Hewett-Emmett et al., 1976) share many peculiarities with the chains from *P. cynocephalus* and *T. gelada*, there appears to be a fundamental dichotomy in papionine α-chain structure,

with macaques on one side and the remaining papionines on the other.

In attempting to account for the close structural correspondence of macaque and human α chains and the extent to which the α chains of *P. cynocephalus* and *T. gelada* (and, most likely, those of *Cercocebus* and *Mandrillus*) have diverged from them, one must consider two possibilities. Either the type of α chain exemplified by that of *P. cynocephalus* has evolved at a very rapid rate since macaques and baboons last shared a common ancestry or one or more events other than the accumulation of nucleotide substitutions in α-chain genes have occurred during the course of primate evolution.

Natural selection has been invoked in accounting for the fixation of the large numbers of substitutions that characterize the α chains of *T. gelada* and *P. cynocephalus*. For example, most of the differences between the α chains of *P. cynocephalus* and macaques (Table II) also involve differences in charge, and residues at several of these sites in the baboon chain can form intrachain salt bridges under physiological conditions (e.g., linking Asp_5 to His_9 and Asp_{53} to Lys_{57}) (Hewett-Emmett et al., 1976). It is difficult to envision these structural peculiarities arising through the random fixation of mutations, inasmuch as their nature and distribution within the chain deviate widely from randomness. Moreover, if the differences between macaque and baboon α chains were attributable to stochastic processes, one might reasonably expect that all papionine α chains would resemble each other, if not in kind, at least in the extent of divergence from the α chains of human and nonpapionine cercopithecoid hemoglobins. Thus, if the differences between macaque and baboon α chains arose since macaques and baboons last shared a common ancestry, it seems most likely that natural selection played a dominant role in their fixation.

An alternative account of the evolutionary history of the α chains of *P. cynocephalus* and *T. gelada* is based upon the distribution of structural peculiarities within their sequences. Reference to Table II indicates that most of the sites at which the chains of both species differ from the human and all macaque α chains fall within the first 57 positions from the amino terminus. After position 57, all save one (at position 82) of the residues by which the chains of *P. cynocephalus* and *T. gelada* differ from the human α chain are shared with one or more of the macaque chains, suggesting that these substitutions arose during the evolutionary divergence of the species in question. Thus, while the α chains from *P. cynocephalus* and *T. gelada* are strikingly divergent from human and macaque chains over positions 1–57, they share similarities with macaque chains at positions 68, 71, and 78 (where they differ from the human chain) and are nearly identical with the human and macaque chains over the remaining portions of their sequences (Lys_{82} is common only to *P. cynocephalus* and *T. gelada* and Gly_{133} is unique to the α chains of *M. nemestrina*).

The distribution of substitutions described above suggests that crossing-over between nonallelic genes might have occurred in the course of α-gene evolution in the primates. Assume that two, closely linked, α-chain loci existed in the past and residing at one of them was an α gene whose evolutionary history was of sufficient duration to permit the acquisition of several structural novelties. Crossing-over between this divergent gene and an evolutionarily conservative gene at the second locus could have produced a nucleotide sequence encoding a chain whose structure was divergent over residues 1–57 but conservative in the remainder of its sequence. Such a product of gene fusion might have been ancestral to the α chains now found in *P. cynocephalus*, *T. gelada*, and, perhaps, the remaining nonmacaque papionines.

The above scheme seems, at first sight, highly speculative and leads to questions concerning the disappearance of the unusually divergent chain from macaques (if, in fact, it existed in the common papionine ancestor), as well as the means whereby the product of gene fusion achieved prominence as the only active α-chain gene in *Papio* and *Theropithecus*. However, several lines of evidence enhance the plausibility of our proposal. First, multiple α-chain loci appear to be quite common among higher primates (Boyer et al., 1973; Nute, 1974). Second, several *M. fascicularis* bear nonallelic α loci, at one of which is encoded the minor, α^X chain (Wade et al., 1967, 1970). Partial sequences, encompassing 106 residues in both the α^X and major, α^A chains of this species, were inferred from compositions of tryptic peptides; comparison of the two sequences demonstrated that the α^X chain had diverged from the α^A chain by no less than four substitutions (Wade et al., 1970). Although the α^X chain of *M. fascicularis* shares none of the structural peculiarities of the α chains from *P. cynocephalus* and *T. gelada*, its existence illustrates the presence of divergent, minor, α-chain genes in some Old World monkeys and, hence, the presence of the raw materials required for formation of products of gene fusion analogous to that posited for *P. cynocephalus* and *T. gelada*. Finally, the discovery of unusual $^3\alpha$ chains in 2 of 14 gorillas and 1 of 37 chimpanzees is noteworthy (Boyer et al., 1973). The $^3\alpha$ chains were present in trace amounts and are nearly identical in structure as defined in part by sequencing and in part by inference of sequence from peptide compositions. Each $^3\alpha$ chain differs from the normal α chain from the same species at eight or nine sites, all of which occur after position 63. As is postulated for the origin of the α chains from *P. cynocephalus* and *T. gelada*, the restriction of structural peculiarities to a portion of the $^3\alpha$ sequence (in this case, the carboxyl-terminal region) suggests that these chains might have arisen through crossing-over between a highly divergent and an evolutionarily conservative α-chain gene. How long ago such an event might have taken place is unknown, but the possibility that the $^3\alpha$ genes of African apes and the α genes of *P. cynocephalus* and *T. gelada* derive from the same event presents a tantalizing prospect. Crossing-over between closely linked, nonallelic α genes should yield two products. If one of the genes involved were highly divergent in nucleotide sequence, one product of gene fusion would have part of the divergent sequence at its 3′ end and the other would have the remainder of the divergent sequence at its 5′ end. Comparison of the $^3\alpha$ chains with the α chains from *P. cynocephalus* and *T. gelada* illustrates this kind of complementarity.

Questions concerning the operation of natural selection, stochastic processes, or both during the evolution of a highly divergent α gene prior to the crossover postulated above must remain moot. However, it is possible that a fusion gene was selected for and subjected to further modification in lineages of nonmacaque papionines while the $^3\alpha$ gene of African apes has been selected to the extent that its presence in trace amounts in but a small proportion of animals is now of little or no physiological significance. One might thus suspect that stochastic processes have been of greater consequence that natural selection in further modification of the latter chain.

Clarification of the evolutionary relationships among the α-chain genes of papionines and the $^3\alpha$ genes of gorillas and chimpanzees might be facilitated by characterization of the complete covalent structures of α chains from additional higher primates. However, further studies, directed toward defining

the numbers and nucleotide sequences of active and inactive α-chain genes in papionines and African apes, are required if the evolutionary scheme presented herein is to be confirmed or denied.

Acknowledgments

The authors thank Dr. Mark A. Hermodson for support and advice offered during the course of this investigation and Dr. Samuel H. Boyer for timely criticisms and suggestions tendered during preparation of the manuscript.

Supplementary Material Available

Table of quantitative estimates of repetitive yields, as well as a summary of the methods employed in identifying each residue in the sequence (4 pages). Ordering information is given on any current masthead page.

References

Boyer, S. H., Noyes, A. N., Boyer, M. L., & Marr, K. (1973) *J. Biol. Chem. 248*, 992.

Buettner-Janusch, J. (1966) *Folia Primatol. 4*, 288.

Clegg, J. B., Naughton, M. A., & Weatherall, D. J. (1968) *Nature (London) 219*, 69.

Cronin, J. E. (1977) *Kroeber Anthropol. Soc. Pap. 50*, 75.

Dayhoff, M. O., Ed. (1972) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Silver Spring, MD.

Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1*, 80.

Friedman, M., Krull, L. H., & Cavins, J. F. (1970) *J. Biol. Chem. 245*, 3868.

Goodman, M., Moore, G. W., & Matsuda, G. (1975) *Nature (London) 253*, 603.

Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., & Walsh, K. A. (1972) *Biochemistry 11*, 4493.

Hermodson, M. A., Ericsson, L. H., Neurath, H., & Walsh, K. A. (1973) *Biochemistry 12*, 3146.

Hermodson, M. A., Schmer, G., & Kurachi, K. (1977) *J. Biol. Chem. 252*, 6276.

Hewett-Emmett, D., Cook, C. N., & Barnicot, N. A. (1976) in *Molecular Anthropology* (Goodman, M., & Tashian, R. E., Eds.) p 257, Plenum Press, New York.

Hill, R. L., & Buettner-Janusch, J. (1964) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 23*, 1236.

Lai, C. Y. (1977) *Methods Enzymol. 47*, 236.

Mahoney, W. C., & Nute, P. E. (1979) *Arch. Biochem. Biophys. 196*, 64.

Matsuda, G. (1976) in *Molecular Anthropology* (Goodman, M., & Tashian, R. E., Eds.) p 223, Plenum Press, New York.

Matsuda, G., Maita T., Ota, H., & Takei, H. (1970) *Int. J. Protein Res. 2*, 99.

Matsuda, G., Maita, T., Nakashima, Y., Barnabas, J., Ranjekar, P. K., & Gandhi, N. S. (1973a) *Int. J. Pept. Protein Res. 5*, 423.

Matsuda, G., Maita, T., Ota, H., Araya, A., Nakashima, Y., Ishii, U., & Nakashima, M. (1973b) *Int. J. Pept. Protein Res. 5*, 405.

Matsuda, G., Maita, T., Watanabe, B., Araya, A., Morokuma, K., Goodman, M., & Prychodko, W. (1973c) *Hoppe-Seyler's Z. Physiol. Chem. 354*, 1153.

Napier, J. R., & Napier, P. H. (1967) *A Handbook of Living Primates*, Academic Press, New York.

Nute, P. E. (1974) *Ann. N.Y. Acad. Sci. 241*, 39.

Nute, P. E., & Pataryas, H. A. (1974) *Am. J. Phys. Anthropol. 40*, 75.

Nute, P. E., & Mahoney, W. C. (1979a) *Biochemistry 18*, 467.

Nute, P. E., & Mahoney, W. C. (1979b) *Hemoglobin 3*, 399.

Romero-Herrera, A. E., & Lehmann, H. (1972) *Biochim. Biophys. Acta 278*, 465.

Rossi-Fanelli, A., Antonini, E., & Caputo, A. (1958) *Biochim. Biophys. Acta 30*, 608.

Schroeder, W. A., Shelton, J. R., Shelton, J. B., Cormick, J., & Jones, R. T. (1963) *Biochemistry 2*, 992.

Schroeder, W. A., Huisman, T. H. J., Shelton, J. R., Shelton, J. B., Kleihauer, E. F., Dozy, A. M., & Robberson, B. (1968) *Proc. Natl. Acad. Sci. U.S.A. 60*, 537.

Sullivan, B., Bonaventura, J., Bonaventura, C., & Nute, P. E. (1976) in *Molecular Anthropology* (Goodman, M., & Tashian, R. E., Eds) p 277, Plenum Press, New York.

Tashian, R. E., Goodman, M., Tanis, R. J., Ferrell, R. E., & Osborne, W. R. A. (1975) in *Isozymes IV: Genetics and Evolution* (Markert, C. L., Ed.) p 207, Academic Press, New York.

Thorington, R. W., Jr., & Groves, C. P. (1970) in *Old World Monkeys: Evolution, Systematics, and Behavior* (Napier, J. R., & Napier, P. H., Eds.) p 629, Academic press, New York.

Wade, P. T., Barnicot, N. A., & Huehns, E. R. (1967) *Nature (London) 215*, 1485.

Wade, P. T., Barnicot, N. A., & Huehns, E. R. (1970) *Biochim. Biophys. Acta 221*, 450.