

cDNA and protein sequence of polymorphic macaque albumins that differ in bilirubin binding

(point mutation/genetic polymorphism/rhesus monkey/serum albumin/evolution)

SCOTT WATKINS*, YASUSHI SAKAMOTO*[†], JEANNE MADISON*, EVELYN DAVIS*, DAVID GLENN SMITH[‡], JOYCE DWULET[§], AND FRANK W. PUTNAM*[¶]

*Department of Biology, Indiana University, Bloomington, IN 47405; [†]Department of Anthropology and California Regional Primate Research Center, University of California, Davis, CA 95616; and [§]Krannert Institute of Cardiology, Indiana University Medical Center, Indianapolis, IN 46202

Contributed by Frank W. Putnam, December 21, 1992

ABSTRACT The rhesus monkey, *Macaca mulatta*, exhibits a geographically restricted polymorphism of serum albumins Mac A and Mac B that is recognized by electrophoresis and is associated with a difference in bilirubin-binding parameters. To identify the basis of the polymorphism, the cDNA and protein sequences of serum albumin from *M. mulatta* were determined. Screening of a λ gt11 rhesus liver cDNA library yielded a 1988-bp cDNA sequence that encodes the complete amino acid sequence of mature albumin, the entire propeptide, and part of the prepropeptide. Isoelectric focusing and amino-terminal protein sequencing of CNBr fragments of albumin from A/A and B/B homozygotes were performed, and the structural difference was localized to a CNBr fragment (MCB3) spanning residues 124–264. Sequence analysis of lysyl endopeptidase peptides of MCB3 established that Mac A albumin has a glutamine residue at position 188 while the Mac B albumin has a glutamic residue at the same position. PCR amplification, subcloning, and DNA sequence analysis of clones from A/A and B/B homozygotes confirmed the protein sequence data and the codon difference of CAA versus GAA, respectively. Comparison of macaque and human serum albumin shows a 93.5% identity at the amino acid level. In human serum albumin, Glu¹⁸⁸ is located close to the IIA binding pocket for ligands, probably including bilirubin. Derivatives of coumarin compete more efficiently with bilirubin for binding sites on the Mac A albumin than on the Mac B albumin. In regions where coumarin-containing plants are important food resources, Mac B albumin may confer a selective advantage because bilirubin is less readily displaced from it.

Serum albumin is the most abundant soluble protein in the body and the dominant protein in the blood plasma of vertebrates (1, 2). It has many important functions, including maintenance of osmotic balance and highly specific binding and transport of numerous metabolites, hormones, metals, and drugs (1, 2). For example, bilirubin, which is a poorly soluble toxic end product of normal erythrocyte metabolism, is bound by albumin and transported to the liver, where it is conjugated in a water-soluble form and excreted, thereby averting jaundice and possible neurological damage. In higher vertebrates serum albumin has a highly conserved, compact structure consisting of a single polypeptide chain of about 585 amino acid residues tightly linked by 17 disulfide bonds to form three homologous domains, each of which has two subdomains (1–5). Many studies have focused on the specificity and sites of binding by albumin of ligands such as fatty acids and pharmaceutical drugs (1, 2). Recently, some of these sites, but not that of bilirubin, have been identified by He and Carter (5), who determined the crystallographic

structure of human serum albumin at a resolution of 2.8 Å. One approach to investigating binding sites is to study structural differences in genetic variants that differ in specificity of binding for a particular ligand. In this paper we report the amino acid and cDNA sequence of the serum albumin of the rhesus monkey, *Macaca mulatta*,^{||} and the difference in sequence of two polymorphic forms of macaque albumin that differ in bilirubin binding specificity (6–10).

It is a paradox that although analbuminemia is extremely rare but apparently tolerable (1, 11), genetic variants of albumin (alloalbumins) are also extremely rare in humans and other species (11–17). More than 40 different human alloalbumins have been identified by protein or DNA sequence analysis, many of them in our laboratory (11–17); however, their cumulative frequency worldwide is only of the order of 1 in 3000. Occasional reports of albumin polymorphism in domestic animals that are based on electrophoretic analysis have appeared, but none of these has apparently been substantiated by structural study. In contrast, by electrophoretic analysis Smith and co-workers (6–9) have identified a pronounced geographic distribution of albumin polymorphism in wild populations of macaque monkeys. Segregation analysis of the two phenotypes (Mac A and Mac B) in *M. mulatta* from India is consistent with the hypothesis of two codominant alleles (A_{mac}^A and A_{mac}^B) whose frequencies are approximately 0.3–0.4 and 0.6–0.7, respectively (6, 10). However, in all regional populations of *M. mulatta* outside India, A_{mac}^A is fixed, or nearly fixed. Other species of macaque vary in the degree of polymorphism in different geographic regions.

Mac A and Mac B albumins (formerly called slow and fast) differ in bilirubin binding parameters. When Mac A and Mac B are fractionated from heterozygous sera (9), Mac B exhibits a higher binding capacity (or average number of binding sites per molecule) than does Mac A. Since the binding capacity (n) of Mac A is very near 1.0, it is possible that Mac B (for which $n = 1.3$) is the only one of the two forms which binds at a secondary bilirubin-binding site. Mac A, on the other hand, exhibits a higher bilirubin-binding association constant (K) than does Mac B, indicating a greater affinity for bilirubin. The product of the two parameters, $n \times K$, which is probably a better indicator of effectiveness of bilirubin binding than either individual parameter, is about 70% higher for Mac A than for Mac B, suggesting that Mac A binds bilirubin more effectively than Mac B does.

MATERIALS AND METHODS

Electrophoretic Analysis of Phenotypes. Plasma samples representing phenotypes A/A, A/B, and B/B were obtained

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

[†]Present address: Department of Physiology, Saitama Medical School, Iruma-Gun, Saitama 350-04, Japan.

[‡]To whom reprint requests should be addressed.

^{||}The sequence reported in this paper has been deposited in the GenBank data base (accession no. M90463).

from rhesus monkeys in the California Regional Primate Research Center at Davis and were subjected to vertical polyacrylamide gel electrophoresis as described by Smith (6). Two common albumin bands that differ in mobility by about 1 mm were visualized approximately 14 cm from the origin (Fig. 1).

Protein Structural Study. Plasma samples from animals homozygous for either the A/A or B/B phenotype were pooled separately. The albumin of each phenotype was purified by HPLC on a preparative DEAE-5-PW column (21.5 × 150 mm; Bio-Rad) with use of a 20 mM sodium acetate buffer with a pH gradient from 5.2 to 4.0 at a flow rate of 5.0 ml/min for 40 min (14–16). The general strategy for protein structural study of the purified macaque albumins was similar to that described for human allobumin (12–17). However, because the amino acid sequence of the macaque albumin and its methionine content were unknown, purification and sequence analysis of the CNBr fragments of each phenotype was required to identify them by homology to human albumin. The general procedure consisted of (i) analytical isoelectric focusing of a CNBr digest of reduced and carboxymethylated albumin to identify the CNBr fragments and to ascertain which differed for the two phenotypes, (ii) purification of all CNBr fragments by size-exclusion and reversed-phase HPLC (12–15), (iii) HPLC peptide mapping of *Staphylococcus aureus* V8 protease or lysyl endopeptidase digests of the purified CNBr fragment, and (iv) automated sequence determination of the CNBr fragments and of peptides purified from the enzymatic digests (14–16). Contrary to our previous experience (12, 15), the V8 protease was not specific for glutamic residues and gave unexpected cleavage after glycine and alanine residues. Hence, lysyl endopeptidase (Wako Pure Chemical, Osaka), which is quite specific for lysine residues, was used to digest the suspected CNBr fragments.

HPLC purification of the CNBr fragment designated MCB3 from albumins Mac A and Mac B was done by gel filtration chromatography on a TSK G3000SW XL column (7.8 × 300 mm; Tosoh, Tokyo) with 0.1% trifluoroacetic acid/35% acetonitrile (vol/vol) as the elution buffer. The purified fragment was digested with lysyl endopeptidase in 0.1 M sodium phosphate, pH 8.5, at 37°C for 2 hr at an enzyme-to-substrate ratio of 1:50 (wt/wt). The peptide profile of the digest was mapped by HPLC with a TSK ODS 80 Ts column (4.6 × 150 mm; Tosoh) using buffers A (0.1% trifluoroacetic acid in water) and B (0.1% trifluoroacetic acid in acetonitrile). The buffer gradient was 0% to 50% buffer B in 100 min at a flow rate of 0.6 ml/min with a column temperature of 35°C.

cDNA Library Screening. A λ gt11 cDNA library, constructed from adult female rhesus monkey liver and consisting of 1.1×10^6 independent recombinant clones (Clontech; catalog no. OL1002b), was screened with a polyclonal anti-

serum raised in rabbits against human serum albumin (Sigma).

Heterologous antibody screening was similar to the original immunological λ gt11 screening procedure (18) with chromogenic modifications (19). A total of 250,000 λ gt11 phage were screened at a density of 25,000 phage per plate and lifted onto nitrocellulose (20). After blocking in TBS (Tris-buffered saline: 0.17 M NaCl/0.01 M Tris-HCl, pH 7.5) and 5% Carnation nonfat dry milk (10 min) (18, 21), the filters were incubated overnight in TBS/5% milk/1:100 dilution of rabbit anti-human serum albumin/100 μ g of phenylmethylsulfonyl fluoride per ml/0.05% thimerosal. Filters were washed once in TBS (5 min), once in TBS/1% Triton X-100/0.5% sodium deoxycholate/0.1% SDS (5 min) (22), and finally twice in TBS (5 min each). Goat anti-rabbit IgG conjugated to peroxidase (Sigma) or biotinylated goat anti-rabbit IgG plus an avidin/biotin-peroxidase complex (Clontech CLIK kit) was used as a secondary antibody in accordance with the manufacturer's suggested working dilutions. The secondary antibody was incubated with the filter in TBS/5% milk (1 hr) and then rinsed three times in TBS (5 min each). The positive plaques were visualized (18) and rescreened through two or three additional rounds.

Subcloning and Sequencing. Lysogens were constructed (19), phage DNA was prepared (20), and the largest fragment, about 1700 bp (Mac7-1D), was inserted (17) into the *Eco*RI site of pBluescript II (Stratagene). Subdigestions (*Alu* I or *Nla* III) of Mac7-1D were subcloned (17) into appropriate plasmid vectors. Both strands of the resulting subclones were sequenced by using double-stranded DNA, standard M13 or Bluescript primers, and Sequenase 2.0 (United States Biochemical) as previously described (17).

To obtain the complete COOH-terminal sequence, additional λ gt11 macaque cDNA clones were isolated from the original library by screening with a homologous probe from the 3' end of Mac7-1D. Phage growth, plating, and filter lifts with Schleicher & Schuell BA85 nitrocellulose were conducted using standard methodology (20). Prehybridization, probe labeling, hybridization, filter washing, and autoradiography were done as previously described (17). One clone (Mac23-1) that hybridized strongly to appropriate cDNA probes was selected for subcloning and sequencing as detailed above.

PCR Amplification, Cloning, and Sequencing. To confirm the difference between Mac A and Mac B, three primers were synthesized (17) for use in either PCR amplification or DNA sequencing (Mac-Pr5, 5'-GAAGTTGCCAGAAGACATCCTTAC-3'; Mac-Pr6, 5'-CTTTGAAAGCTCTGTCTCCGAA-3'; and Mac-Pr7, 5'-CAGCACATTCAAGCAGATCTCCAT-3'). Mac-Pr5 and -Pr7 were used to PCR amplify a 2.7-kb region of the macaque serum albumin gene encoding MCB3. PCRs were performed according to the supplier's directions for "hot start" PCR using AmpliWax PCR Gems (Perkin-Elmer/Cetus) [10 mM Tris-HCl, pH 8.3/50 mM KCl/1.5 mM MgCl₂/0.1 μ M each primer/1 μ g of genomic DNA/2.5 units of AmpliTaq DNA polymerase (Perkin-Elmer/Cetus)]. With a Perkin-Elmer/Cetus DNA thermal cycler, the products were amplified for 40 cycles using the following protocol: 1 min at 94°C, 2 min at 55°C, and 3 min at 72°C with a 5-sec increase in extension time for each successive polymerization step.

The amplified Mac A/A or Mac B/B DNA was digested with *Stu* I (New England Biolabs) and *Pst* I (Boehringer Mannheim) and then subcloned into *Sma* I-*Pst* I-cut Bluescript II KS (+) (Stratagene) as previously described (17). Approximately 2–4 μ g of double-stranded DNA was sequenced by using Sequenase 2.0 (United States Biochemical) and the Mac-Pr6 primer (17).

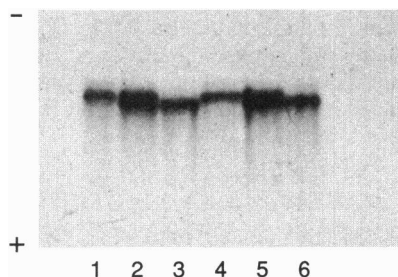


FIG. 1. Electrophoretic patterns of three common serum albumin phenotypes of rhesus macaques (*M. mulatta*) obtained by methods described in the text. Lanes 1 and 4, 2 and 5, and 3 and 6 represent phenotypes A/A, A/B, and B/B, respectively, which are products of the codominant alleles A_{mac}^A and A_{mac}^B .

RESULTS AND DISCUSSION

Protein Structural Study. Electrophoretic analysis by vertical polyacrylamide gel electrophoresis revealed two albumin bands that differed in mobility by about 1 mm approximately 14 cm from the origin, B/B being the faster-moving phenotype (Fig. 1). However, cellulose acetate electrophoresis at pH 8.6 did not distinguish the two phenotypes, and both had similar retention times in DEAE HPLC purification. Isoelectric focusing of the CNBr digests of the two phenotypes showed a unique set of bands for each (Fig. 2). Because the band pattern differed from that previously found for human albumin (11), a series of purified CNBr peptides of each phenotype was subjected to automatic sequence analysis to identify the bands. The results, together with cDNA sequence analysis described below, established that there are 9 methionines in mature macaque albumin and thus 10 possible CNBr fragments (designated MCB1 to MCB10 based on their order in the sequence); this compares with 6 methionines and 7 CNBr fragments in the human protein. Because the macaque albumin showed a very high sequence homology to the human, the MCB fragments could be numbered and aligned in sequence. The CNBr fragment in which the structural difference of the two phenotypes was located by isoelectric focusing was designated MCB3 and later found to span residues 124–264.

More than 70% of the amino acid sequence of Mac B albumin and some of Mac A albumin was determined by automated sequence analysis of the intact protein, its CNBr fragments, and many V8 protease peptides of MCB3. Although no difference in sequence of the two phenotypes was established by this extensive study, the results indicated that the structural difference had to be small and probably was limited to a single amino acid exchange in MCB3. However, because of the large size of MCB3 (141 amino acid residues) and the unexpected lack of specificity of V8 protease in this study, further protein structural study was deferred until the complete amino acid sequence of macaque albumin was established by molecular cloning and DNA sequencing.

Cloning and DNA Sequence Analysis. Through heterologous antibody screening of a λ gt11 library and subsequent rescreening with homologous DNA probes, a 1988-bp cDNA sequence for *M. mulatta* (Fig. 3) was obtained. Translation of the cDNA indicated that this sequence contains the complete amino acid sequence for mature macaque serum albumin, the entire propeptide sequence, and 10 amino acid residues of a putative 18-residue prepropeptide. Comparison of the human

and macaque serum albumin cDNAs showed a 95% identity. Furthermore, the codon usage for human serum albumin is virtually identical to that for macaque serum albumin. Of the 39 amino acid differences between human and macaque serum albumin, 36 are due to a single-base substitution and 3 arise from a two-base change.

Homology to Human and Animal Albumins. Macaque serum albumin has the characteristic structural features of human and animal albumins, consisting of three homologous domains and six subdomains linked in a compact pattern by 17 disulfide bonds (3, 4). Like the albumins of higher nonprimate mammals (e.g., rat, bovine, sheep, and pig), macaque albumin has 584 amino acid residues, rather than 585 as in the human. Comparison of amino acid sequence homology by the GAP program (23) showed 93.5% identity of macaque with human albumin (Fig. 4); like the latter, macaque albumin showed about 75% identity with the albumin of higher nonprimate mammals but only about 37% with frog albumins. As expected, most of the differences in the human and macaque albumin sequences occur at sites that have also mutated in nonprimate albumins. The 39 amino acid exchanges in the human and macaque albumins appear to be randomly distributed throughout the polypeptide chain except for an identical stretch of 75 residues from His³⁹ to Pro¹¹³. There is no significant correlation of amino acid differences in the two primate albumin sequences with substitutions in the 40 known genetic variants of human albumin A, or with the occurrence of CpG dinucleotides. The exchange of Pro¹¹⁴ in macaque by Arg¹¹⁴ in human is interesting because it occurs at the same site as the Arg¹¹⁴ → Gly substitution in albumin Yanomama, a “private” polymorphic form of human albumin present in the Yanomama Indians that is associated with decreased binding of bilirubin (11, 24). This exchange occurs in a proline-rich region that is part of a nonhelical extended intersubdomain connector—i.e., h6(I) → h7(I) in the crystal structure of the human protein (5). Four other human alloalbumins (Glu⁵⁰¹ → Lys, Asp⁵⁵⁰ → Gly, Asp⁵⁵⁰ → Ala, and Lys⁵⁷³ → Glu) have an amino acid substitution (14, 16) at the same site as a human/macaque difference, but in no case is the amino acid substitution in the alloalbumins identical to the residue found in macaque.

Amino Acid Sequence Exchange in Macaque A/A and B/B Albumins. The amino acid exchange in the macaque A/A and B/B albumins was established by HPLC purification of the MCB3 fragment of each albumin followed by lysyl endopeptidase digestion of the fragment, HPLC peptide mapping of the digest, and protein sequence analysis of the peptides in which the change was located. Two peptides differing slightly in retention time were isolated and sequenced. The peptide derived from the B/B albumin had the sequence Leu-Asp-Glu-Leu-Arg-Asp-Glu-Gly-Lys corresponding exactly to the sequence inferred from the cDNA for amino acid residues 182–190. The peptide from the A/A albumin had an identical sequence except for the change of glutamic acid to glutamine at position 188 underlined above. Thus, sequence analysis of the lysyl endopeptidase peptides of the MCB3 fragments of the respective phenotypes established that Gln¹⁸⁸ in macaque A/A albumin is exchanged for Glu¹⁸⁸ in macaque B/B albumin. The increase in net charge of +1 for the A/A albumin accounts for its slightly greater cathodal mobility in vertical polyacrylamide gel electrophoresis, which gives higher resolution than cellulose acetate electrophoresis.

To verify the location of the nucleotide change for both Mac A/A and Mac B/B homozygotes, 1 μ g of DNA was PCR-amplified with primers specific for macaque serum albumin and subcloned into pBluescript (Stratagene), and two independent subclones from each genotype were sequenced. The nucleotides from positions 610–612 (Fig. 3) were CAA for the Mac A/A subclones and GAA for the Mac

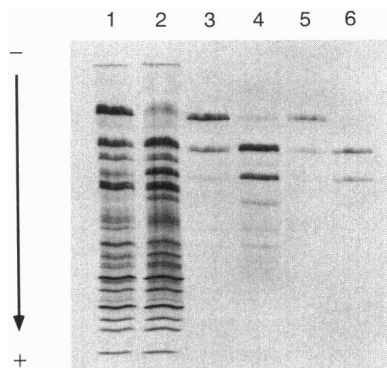


FIG. 2. Comparative isoelectric focusing of CNBr peptides of homozygous macaque albumins A/A and B/B. Lanes 1, 3, and 5 represent macaque A/A and lanes 2, 4, and 6 refer to macaque B/B. Lanes 1 and 2, CNBr peptides of the total protein digest; lanes 3 and 4, MCB3 fraction from a G3000SW size exclusion column; lanes 5 and 6, MCB3 after purification on a Vydac C₁₈ reversed-phase column. The CNBr fragments may have two charged forms because of the homoserine–homoserine lactone equilibrium.

-16	CTT CTC TTT CTC TTT AGC TCA GCT TAT TCC	AGG GGT GTG TTT CGT CGA GAC ACA CAC AAG	AGT GAG GTT GCT CAT CGG TTT AAA GAT TTG	90
	Leu Leu Phe Leu Phe Ser Ser Ala Tyr Ser	Arg Gly Val Phe Arg Arg Asp Thr His Lys	Ser Glu Val Ala His Arg Phe Lys Asp Leu	
15	GGA GAA GAA CAT TTC AAA GGC TTG GTG CTG	GTT GCT TTT TCT CAG TAT CTT CAG CAG TGT	CCA TTT GAA GAG CAT GTA AAA TTA GTG AAT	180
	Gly Glu Glu His Phe Lys Gly Leu Val Leu	Val Ala Phe Ser Gln Tyr Leu Gln Gln Cys	Pro Phe Glu Glu His Val Lys Leu Val Asn	
45	GAA GTA ACT GAA TTT GCA AAA ACA TGT GTT	GCT GAT GAG TCA GCT GAA AAT TGT GAC AAA	TCA CTT CAT ACT CTT TTT GGA GAC AAA TTA	270
	Leu Val Thr Glu Phe Ala Lys Thr Cys Val	Ala Asp Glu Ser Ala Glu Asn Cys Asp Lys	Ser Leu His Thr Leu Phe Gly Lys Lys Leu	
75	TGC ACA GTC GCA ACT CTT CGT GAA ACC TAT	GGT GAA ATG GCT GAC TGC TGT GCA AAA CAA	GAA CCT GAG AGA AAT GAA TGC TTC CTG CAA	360
	Cys Thr Val Ala Thr Leu Arg Glu Thr Tyr	Gly Glu Met Ala Asp Cys Cys Ala Lys Gln	Glu Pro Glu Arg Asn Glu Cys Phe Leu Gln	
105	CAC AAA GAT GAC AAC CCG AAC CTC CCC CCA	TTG GTG AGA CCA GAG GTT GAT GTG ATG TGC	ACT GCT TTT CAT GAT AAT GAA GCG ACA TTT	450
	His Lys Asp Asp Asn Pro Asn Leu Pro Pro	Leu Val Arg Pro Glu Val Asp Val Met Cys	Thr Ala Phe His Asp Asn Glu Ala Thr Phe	
135	TTG AAA AAA TAC TTA TAT GAA GTT GCC AGA	AGA CAT CCT TAC TTT TAT GCC CCA GAA CTC	CTT TTC TTT GCT GCA AGG TAT AAA GCT GCT	540
	Leu Lys Lys Tyr Leu Tyr Glu Val Ala Arg	Arg His Pro Tyr Phe Tyr Ala Pro Glu Leu	Leu Phe Phe Ala Ala Ala Tyr Lys Ala Ala	
165	TTT GCA GAA TGT TGT CAA GCT GCT GAT AAA	GCT GCC TGC CTC TTA CCA AAG CTT GAT GAA	CTT CGG GAT GAA GGG AAG GCT TCA TCT GCC	630
	Phe Ala Glu Cys Gln Ala Ala Asp Lys Tyr	Ala Ala Cys Leu Leu Pro Lys Leu Asp Glu	Glu Gly Lys Ala Ser Ser Ser	
195	AAA CAG AGA CTC AAG TGT GCC AGT CTC CAA	AAA TTC GGA GAC AGA GCT TTC AAA GCA TGG	GCA GTA GCT CGC TTG AGC CAG AAA TTT CCC	720
	Lys Gln Arg Leu Lys Cys Ala Ser Leu Gln	Lys Phe Gly Asp Arg Ala Phe Lys Ala Trp	Ala Val Ala Arg Leu Ser Gln Lys Phe Pro	
225	AAA GCT GAG TTT GCA GAT TTT TCC AAG TTA	GTG ACA GAT CTT ACC AAA GTC CAC ACG GAA	TGC TGC CAT GGA GAT CTG CTT GAA TGT GCT	810
	Lys Ala Glu Phe Ala Glu Val Ser Lys Leu	Val Thr Asp Leu Thr Lys Val His Thr Glu	Cys Cys His Gly Asp Leu Leu Glu Cys Ala	
255	GAT GAC AGG GCG GAC CTT GCC AAG TAT ATG	TGT GAA AAT CAA GAT TCA ATC TCC AGT AAA	CTG AAG GAA TGC TGT GAT AAA CCT CTG TTG	900
	Asp Asp Arg Ala Asp Leu Ala Lys Tyr Met	Cys Glu Asn Gln Asp Ser Ile Ser Ser Lys	Leu Lys Glu Cys Cys Asp Lys Pro Leu Leu	
285	GAA AAA TCC CAC TGC CTT GCC GAA GTG GAA	AAT GAT GAG ATG CCT GCT GAC TTG CCT TCA	TTA GCT GCT GAT TAT GTT GAG AGT AAG GAT	990
	Glu Lys Ser His Cys Leu Ala Glu Val Glu	Asn Asp Glu Met Pro Ala Asp Leu Pro Ser	Leu Ala Ala Asp Tyr Val Glu Ser Lys Asp	
315	GTT TGC AAA AAT TAT GCT GAG GCA AAG GAT	GTC TTC CTG GGC ATG TTT TTG TAT GAA TAT	GCA AGA AGG CAT CCT GAT TAC TCT GTC ATG	1080
	Val Cys Lys Asn Tyr Ala Glu Ala Lys Asp	Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr	Ala Arg Arg His Pro Asp Tyr Ser Val Met	
345	CTG CTG CTG AGA CTT GCC AAG GCA TAC GAA	GCC ACT CTA GAG AAG TGC TGT GCC GCT GCA	GAT CCT CAT GAA TGC TAT GCC AAA GTG TTT	1170
	Leu Leu Leu Arg Leu Lys Ala Tyr Glu	Ala Thr Leu Glu Lys Cys Cys Ala Ala Ala	Asp Pro His Glu Cys Tyr Lys Val Phe	
375	GAT GAA TTT CAA CCT CTT GTG GAA GAG CCT	CAG AAT TTA GTC AAA CAA AAT TGT GAA CTT	TTT GAG CAG CTT GGA GAA TAC AAA TTC CAG	1260
	Asp Glu Lys Gln Pro Leu Val Glu Glu Pro	Gln Asn Leu Val Lys Gln Asn Cys Glu Leu	Phe Glu Gln Leu Glu Tyr Lys Phe Gln	
405	AAT GCG CTA TTA GTT CGT TAC ACC AAG AAA	GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA	GAG GTC TCA AGA AAC CTA GGA AAA GTG GGT	1350
	Asn Ala Leu Leu Val Arg Tyr Thr Lys Lys	Val Pro Gln Val Ser Thr Pro Thr Leu Val	Glu Val Ser Arg Asn Leu Glu Lys Val Glu	
435	GCC AAA TGT TGT AAA CTT CCT GAA GCA AAA	AGA ATG CCC TGT GCA GAA GAC TAT CTG TCC	GTG GTC CTG AAC CGG TTA TGT GTG TGT CAT	1440
	Ala Lys Cys Cys Lys Leu Pro Glu Ala Lys	Arg Met Pro Cys Ala Glu Asp Tyr Leu Ser	Val Val Leu Asn Arg Leu Cys Val Leu His	
465	GAG AAA ACG CCA GTA AGC GAG AAA GTC ACC	AAA TGC TGC ACG GAA TCC TTG GTG AAC AGG	CGA CCA TGC TTT TCT GCT CTG GAA CTC GAT	1530
	Glu Lys Thr Pro Val Ser Glu Lys Val Thr	Lys Cys Cys Thr Glu Lys Ser Leu Val Asn Arg	Arg Pro Cys Phe Ser Ala Leu Glu Lys Asp	
495	GAA GCA TAT GTT CCC AAA GCG TTT AAT GCT	GAA ACA TTC ACC TTC CAT GCA GAT ATG TGC	ACA CTT TCT GAG AAG GAG AAA CAA GTC AAG	1620
	Glu Ala Tyr Val Ala Phe Asn Ala	Glu Thr Phe Thr Phe His Ala Asp Met Cys	Thr Leu Ser Glu Lys Glu Lys Val Lys	
525	AAA CAA ACT GCA CTT GTT GAG CTC GTG AAA	CAC AAG CCT AAG GCA ACA AAG GAA CTG	AAA GGT GTT ATG GAT AAT TTT GCA GCT TTT	1710
	Lys Gln Thr Ala Leu Val Glu Leu Val Lys	His Lys Pro Lys Ala Thr Lys Glu Gln Leu	Lys Gly Val Met Asp Asn Phe Ala Ala Phe	
555	GTA GAG AAG TGC TGC AAG GCC GAT GAT AAG	GAG GCC TGC TTT GCC GAG GAG GGT CCA AAA	TTT GTT GCT GCA AGT CAA GCT GCC TTA GCC	1800
	Val Glu Lys Cys Lys Lys Ala Asp Asp Lys	Glu Ala Cys Phe Ala Glu Glu Pro Lys	Phe Val Ala Ala Ser Gln Ala Ala Leu Ala	
	TAA AAA CAT CAC AAT TAA GAA CAT CTC AGC	CTA CCA TGA GAA CAA GAG AAA TAA AAT GAA	GAT CAA AAG CTT ATT CAT CTG TTT TTC TTT	1890
	TTT ATT GGT ATA AAG CCA ACA CCC TGT CTA	AAA AAC TAT AAA TTT CTT TAA TCA TTT TGC	CTC TTT TCT CTG TGC TTC AAT TAA TAA AAA	1980
	AAA AAA AA			1988

FIG. 3. cDNA and translated protein sequence of macaque serum albumin. Numbering on the left refers to the amino acid sequence and numbering on the right refers to the nucleotide sequence. Numbering for amino acids includes the partial sequence of the prepropeptide (−16 to −7), the propeptide (−6 to −1), and the mature protein (+1 to +584). Underlined amino acid residues have been confirmed by protein sequence analysis. The asterisks at Glu¹⁸⁸ indicate the site of amino acid exchange—i.e., Glu¹⁸⁸, macaque B/B; Gln¹⁸⁸, macaque A/A.

B/B subclones, confirming the protein sequencing data of Gln¹⁸⁸ for Mac A/A and Glu¹⁸⁸ for Mac B/B.

Bilirubin Binding. Bilirubin binds to human, macaque, and bovine serum albumins at one high-affinity site and at one or two low-affinity sites (1, 2). As reviewed by Peters (1), many investigators have attempted to identify the single strong site in human and bovine albumins by a variety of methods including chemical modification studies, covalent coupling of bilirubin, and limited proteolytic cleavage. When bilirubin is

coupled to albumin, it appears to be located in a fragment from about residue 185 (near a proteolytic cleavage site) through residue 308 and is primarily in domain II (≈ residues 195–390). Three basic residues have been implicated in the binding site: Lys¹⁹⁵ and Lys²⁴⁰ in human and Lys²²⁰ in bovine albumin (homologous to Arg²²² in human). In the crystallographic structure of human serum albumin (5), Lys¹⁹⁹ and Arg²²² interact with the carboxylate group of 2,3,5-triiodobenzoic acid, a ligand that has been shown to occupy

-24	MKWTFISLFLFFSSAYSRGVFRDRDAHKSEVAHRFKDLGEENFKALVLI	IAFAQYLOQCPFEDHVKLNVNEVTEFAKTCVADESAENCDSKHLTFGDKLCT	Human
-16	LLFLFFSSAYSRGVFRDRTHKSEVAHRFKDLGEENFKGLVLF	AFSGYLOQCPFEEDHVKLNVNEVTEFAKTCVADESAENCDSKHLTFGDKLCT	Macaque
77	VATLRRETYGEMADCCAKQPERNECFLOHKDDPNLPLRVPEDVMCTAF	HDNEETFLKKYLYEIAARRHPYFYAPELLFFAKRYKAAFTCCQAADKAA	Human
77	VATLRRETYGEMADCCAKQPERNECFLOHKDDPNLPLRVPEDVMCTAF	HDNEATFLKKYLYEVARRHPYFYAPELLFFAARYKAAFAECCQAADKAA	Macaque
177	CLLPKLDLDELDEGKASSAKQRLKASLGKGFGERAFKAMAVARLSOR	FKAEFAEVSKLVDTLTKVHTECCGDLLECCADRADLAKYICENQDSIS	Human
177	CLLPKLDLDELDEGKASSAKQRLKASLGKGFGERAFKAMAVARLSOR	FKAEFAEVSKLVDTLTKVHTECCGDLLECCADRADLAKYICENQDSIS	Macaque
277	ECCEKPLLEKSHCIAEVENDEMPADLPSLAADVSKDVCKNYAEAKD	VFLGMFLYFYARRHPDYSVMLLRLAKYETTELKCCAAADPHECYAKV	Human
277	ECCEKPLLEKSHCIAEVENDEMPADLPSLAADVSKDVCKNYAEAKD	VFLGMFLYFYARRHPDYSVMLLRLAKYETTELKCCAAADPHECYAKV	Macaque
377	FKPLVEEPONLILKQNCLEFQELGEYKFNALLVRYTKKVPQVSTPTL	VEVSRNLGKVGSKCKHPEAKRMPCAEDYLSVLLNQLCVLHEKTPVSD	Human
377	FKPLVEEPONLILKQNCLEFQELGEYKFNALLVRYTKKVPQVSTPTL	VEVSRNLGKVGSKCKHPEAKRMPCAEDYLSVLLNQLCVLHEKTPVSD	Macaque
477	CTSELVNRPRCSALEVDITYPKFNAETFTFHADICTLSEKEROIKQ	TALVELVKHKPKATKEQLKAVMDFAAFVEKCKKADKCTFAEEGKLV	Human
477	CTSELVNRPRCSALEVDITYPKFNAETFTFHADICTLSEKEROIKQ	TALVELVKHKPKATKEQLKAVMDFAAFVEKCKKADKCTFAEEGKLV	Macaque
577	AASQAALGL		Human
577	AASQAALA.		Macaque

FIG. 4. Protein sequence comparison of human serum albumin with macaque serum albumin. The alignment was generated by the GAP program (23). Vertical lines indicate identity, and double dots denote conservative changes.

a hydrophobic crevice constituting the principal binding site in subdomain IIA that is probably also one site for fatty acid binding (D. C. Carter, personal communication). From these observations it appears likely that the Gln¹⁸⁸ → Glu exchange in the macaque A and B allotypes is close to or in the primary bilirubin-binding site and thus correlates with the difference in their affinity for bilirubin.

Although bilirubin binding is also decreased in albumin Yanomama (Arg¹¹⁴ → Gly), a genetic variant restricted to the Yanomama Indians of Brazil, the lower bilirubin binding of this alloalbumin may be due to a more global protein conformational change (11).

Geographic Distribution and Potential Evolutionary Significance. Mac A is either the most common form or the only form of albumin in most species of genus *Macaca* (6–10). Since Mac A appears also to be fixed in the genus *Papio* (baboons), which is closely related to the genus *Macaca*, Mac A is likely to be the primitive form of macaque albumin. Mac B is fixed or nearly fixed in three closely related macaque species (*Macaca radiata*, *Macaca sinica*, and *Macaca assamensis*) which together make up the so-called sinica group from India and Sri Lanka. Since Mac B is found in low frequencies (i.e., <10%) in at least five other species of the genus *Macaca* which are completely unrelated to each other, the mutation probably predates the earliest speciation within the genus. Otherwise, Mac B is found in only three other species of genus *Macaca*, including *Macaca maura* and *Macaca hecki*, both from Sulawesi, with frequencies of about 50–60%.

The geographic range of species *M. mulatta* is immense, extending westward to Afghanistan and Pakistan and eastward through mainland southeast Asia into eastern China (6–10). Mac B is either rare or absent in all regional populations of *M. mulatta* except those in India, where it reaches frequencies between 60% and 70%. The high frequency of Mac B, which is otherwise rare, in unrelated species of macaques of India, but not elsewhere, suggests that forces other than genetic drift or common ancestry influence its distribution.

Derivatives of coumarin bind albumin at the major bilirubin-binding site (25) but more efficiently compete with bilirubin bound to Mac A than to Mac B (9). The substitution of glutamic acid for glutamine at amino acid residue 188, which otherwise reduces the effectiveness of bilirubin binding, enhances binding in the presence of coumarin. Coumarin and/or its derivatives, which are excreted in breast milk (26), are constituents of five of the six major floral taxa employed as food sources by *M. mulatta* from India (27). In regions where coumarin-containing plants are heavily exploited as food resources, Mac B albumin might enjoy a selective advantage because bilirubin is less readily displaced from it. In regions where such plants are rare or absent, Mac A albumin might be selectively favored due to its high bilirubin-binding association constant.

We thank Dr. David R. Hathaway for use of the Applied Biosystems sequencer. This work was supported by National Institutes of Health Grant DK 19221 to F.W.P. and Grants RR00169, HD32804, and R24RR05090 to D.G.S.

- Peters, T., Jr. (1985) *Adv. Protein Chem.* **37**, 161–245.
- Kragh-Hansen, U. (1981) *Pharmacol. Rev.* **33**, 17–53.
- Brown, J. R. & Shockley, P. (1982) in *Lipid-Protein Interactions*, eds. Jost, P. & Griffith, O. H. (Wiley, New York), Vol. 1, pp. 25–68.
- Minghetti, P. P., Ruffner, D. E., Kuang, W.-J., Dennison, O. E., Hawkins, J. W., Beattie, W. G. & Dugaiczky, A. (1986) *J. Biol. Chem.* **261**, 6747–6757.
- He, X. M. & Carter, D. C. (1992) *Nature (London)* **358**, 209–215.
- Smith, D. G. (1980) *Am. J. Phys. Anthropol.* **53**, 243–249.
- Smith, D. G. & Ahlfors, C. E. (1981) *Am. J. Phys. Anthropol.* **54**, 37–41.
- Smith, D. G. & Rolfs, B. K. (1984) *Am. J. Primatol.* **7**, 285–289.
- Lorey, F. W., Smith, D. G. & Ahlfors, C. E. (1985) *Am. J. Phys. Anthropol.* **68**, 169–171.
- Fooden, J. & Lanyon, S. M. (1989) *Am. J. Primatol.* **17**, 209–241.
- Takahashi, N., Takahashi, Y., Isobe, T., Putnam, F. W., Fujita, M., Satoh, C. & Neel, J. V. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8001–8005.
- Arai, K., Huss, K., Madison, J., Putnam, F. W., Salzano, F. M., Franco, M. H. L. P., Santos, S. E. B. & Freitas, M. J. M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1821–1825.
- Arai, K., Madison, J., Shimizu, A. & Putnam, F. W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 497–501.
- Madison, J., Arai, K., Sakamoto, Y., Feld, R. D., Kyle, R. A., Watkins, S., Davis, E., Matsuda, Y.-i., Amaki, I. & Putnam, F. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9853–9857.
- Sakamoto, Y., Davis, E., Madison, J., Watkins, S., McLaughlin, H., Leahy, D. T. & Putnam, F. W. (1991) *Clin. Chim. Acta* **204**, 179–188.
- Carlson, J., Sakamoto, Y., Laurell, C.-B., Madison, J., Watkins, S. & Putnam, F. W. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8225–8229.
- Watkins, S., Madison, J., Davis, E., Sakamoto, Y., Galliano, M., Minchiotti, L. & Putnam, F. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5959–5963.
- Huynh, T. V., Young, R. A. & Davis, R. W. (1984) in *DNA Cloning Techniques: A Practical Approach*, ed. Glover, D. M. (IRL, Oxford), pp. 49–78.
- DeWet, J. R., Fukushima, H., Dewji, N. N., Wilcox, E., O'Brien, J. S. & Helinski, D. R. (1984) *DNA* **3**, 437–448.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Plainview, NY).
- Johnson, D. A., Gautsch, J. W., Sportsman, J. R. & Elder, J. H. (1984) *Gene Anal. Tech.* **1**, 3–8.
- Helfman, D. M., Feramisco, J. R., Fiddes, J. C., Thomas, G. P. & Hughes, S. H. (1984) *Focus* **6**, 1–5.
- Devereux, J., Haerberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **4**, 387–395.
- Lorey, F. W., Ahlfors, C. E., Smith, D. G. & Neel, J. V. (1984) *Am. J. Hum. Genet.* **36**, 1112–1120.
- Brodersen, R. (1978) in *Intensive Care in the Newborn*, eds. Stern, L., Oh, W. & Fris-Hansen, B. (Masson, New York), pp. 331–345.
- Brambel, C. E. & Hunter, R. E. (1950) *Am. J. Obstet. Gynecol.* **59**, 1153–1159.
- Lorey, F. W. & Smith, D. G. (1986) in *Genetic Variation and Its Maintenance*, eds. Roberts, D. F. & DeStefano, G. F. (Cambridge Univ. Press, Cambridge), pp. 229–235.