Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 3909–3913, May 1990
Biochemistry

Hypermutability of CpG dinucleotides in the propeptide-encoding sequence of the human albumin gene

(serum albumin/proalbumins/alloalbumins/genetic polymorphism/point mutations)

STEPHEN O. BRENNAN*, KUNIO ARAI†, JEANNE MADISON‡, CARL-BERTIL LAURELL§, MONICA GALLIANO¶, SCOTT WATKINS‡, ROBERT PEACH*, TIMOTHY MYLES*, PETER GEORGE*, AND FRANK W. PUTNAM‡‖

*Molecular Pathology Laboratory, Department of Clinical Biochemistry, Christchurch Hospital, Christchurch, New Zealand; †Department of Chemistry, Josai University, Keyakidai, Sakado, Saitama 350-02, Japan; ‡Department of Biology, Indiana University, Bloomington, IN 47405; §Department of Clinical Chemistry, Malmö General Hospital, Lund University, S-214 01 Malmö, Sweden; and ¶Department of Biochemistry, University of Pavia, 27100 Pavia, Italy

Contributed by Frank W. Putnam, February 5, 1990

ABSTRACT **An electrophoretically slow albumin variant was detected with a phenotype frequency of about 1:1000 in Sweden and was also found in a family of Scottish descent from Kaikoura, New Zealand, and in five families in Tradate, Italy. Structural study established that the major variant component was arginyl-albumin, in which arginine at the −1 position of the propeptide is still attached to the processed albumin. A minor component with the amino-terminal sequence of proalbumin was also present as 3–6% of the total albumin. After amplification of the gene segment encoding the prepro sequence of albumin, specific hybridization of DNA to an oligonucleotide probe encoding cysteine at position −2 indicated the mutation of arginine at the −2 position to cysteine (−2 Arg → Cys). This produced the propeptide sequence Arg-Gly-Val-Phe-Cys-Arg. This was confirmed by sequence analysis after pyridylethylation of the cysteine. This mutation produces an alternate signal peptidase cleavage site in the variant proalbumin precursor of arginyl-albumin giving rise to two possible products, arginyl-albumin and the variant proalbumin. Another plasma from Bremen had an alloalbumin with a previously described substitution (1 Asp → Val), which also affects propeptide cleavage. Hypermutability of two CpG dinucleotides in the codons for the diarginyl sequence may account for the frequency of mutations in the propeptide. Mutation at these two sites results in a series of recurrent proalbumin variants that have arisen independently in diverse populations.**

Many rare genetic variants of human serum albumin (alloalbumins) have recently been studied because of interest in their population distribution, site of mutation, and structural change (1–16). Three types of alloalbumins have been identified: (*i*) more than 20 different point mutants that have a single substitution in the mature processed albumin molecule, which contains 585 amino acid residues (for a list with references see ref. 1); (*ii*) several carboxyl-terminal variants (9, 10); and (*iii*) a series of proalbumins (11–16). A proalbumin is a serum albumin molecule that retains a basic amino-terminal hexapeptide because of incomplete post-translational processing. During the second stage of processing in liver a Ca^{2+}-dependent diarginyl-specific proalbumin "convertase" removes the normal propeptide Arg-Gly-Val-Phe-Arg-Arg; then the mature albumin with the amino-terminal sequence Asp-Ala-His-Lys- is secreted into the circulation. Proalbumins are characterized by their slower electrophoretic mobility and reduced binding of $^{63}Ni^{2+}$ compared to normal (common) albumin A (12, 13, 15).

A number of specific point mutations in or near the propeptide that affect the processing of proalbumins have been identified and have proved useful for study of the specificity of the convertase (12, 13). Some mutations completely block processing and result in proalbumins that make up about 50% of the total albumin—e.g., proalbumin Christchurch (arginine in position −1 replaced by glutamine, denoted −1 Arg → Gln) (11), Takefu (−1 Arg → Pro) (15), Jaffna (−1 Arg → Leu) (16), and Lille (−2 Arg → His) (14) (Fig. 1). Other mutations interfere with the processing and result in a precursor proalbumin present as 3–10% of the total albumin and a second aberrant albumin as its product (12, 13). Proalbumin Blenheim (1 Asp → Val), which retains the normal propeptide, is present in small amount, and the main product is albumin Blenheim (1 Asp → Val) (12). The gene for proalbumin Redhill contains a double mutation (−2 Arg → Cys, 320 Ala → Thr), but the predicted proalbumin product was not detected in plasma. The circulating species was arginyl-albumin with a 320 Ala → Thr substitution. This produces a glycosylation sequence (Asn-Tyr-Thr) which results in the attachment of a glucosamine glycan on Asn-318 (13).

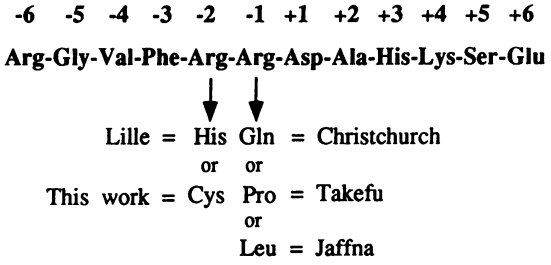

FIG. 1. Single point substitutions in the propeptide sequence of genetic variants of human serum albumin. The sequence shown is that of the propeptide (residues −6 to −1) and the amino-terminal hexapeptide (residues +1 to +6) of normal serum albumin A. Amino acid substitutions below the propeptide sequence denote previously reported substitutions that lead to secretion of proalbumins designated by geographical names (11, 14, 15, 16). The substitutions in the Lille and Christchurch types have also been reported for proalbumins given other geographical names because they were found in different populations of diverse ethnic origin. The Lille type includes proalbumins Fukuoka-2, Taipei, Somalia RS, and others (1, 6, 15); the Christchurch type includes proalbumins Honolulu-2, Shizuoka-1, Gainesville, and others (1, 3, 15). In these cases and also for the Takefu and Jaffna types the propeptide sequence is retained, and the proalbumin is secreted into the circulation. However, the −2 Arg → Cys mutation described in this work produces an aberrant signal peptidase cleavage site that results in the secretion chiefly of arginyl-albumin. The −2 Arg → Cys mutation has also been reported for albumin Redhill, which is a double mutant (−2 Arg → Cys, 320 Ala → Thr) (13).

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.

Abbreviation: PCR, polymerase chain reaction.

‖To whom reprint requests should be addressed.

In a collaborative effort involving laboratories at Malmö, Sweden; Bloomington, Indiana; Christchurch, New Zealand; Saitama, Japan; and Pavia, Italy, we studied the most common Swedish albumin variation, which is expressed in plasma as a broadened electrophoretic band indicative of a slow component at pH 8.6. This albumin variation was identified in 11 unrelated Swedes and was present with a phenotype frequency of about 1:1000 in plasmas of subjects submitted to clinical electrophoresis at Malmö General Hospital; it was also identified in a family of Scottish descent in Kaikoura, New Zealand, and in five unrelated families in Tradate, a small city in northern Italy. In all these cases a proalbumin (−2 Arg → Cys) was present in plasma as 3–6% of the total albumin, and the major variant was arginyl-albumin (about 30%). Two other albumin variants with a similar slow mobility were detected in the Malmö series, but they were not proalbumins and had a different substitution. Albumin Bremen had the 1 Asp → Val substitution described for albumin Blenheim (12).

MATERIALS AND METHODS

Sources of Sera. In a search for albumin variants a review was made of the films for 9388 plasma samples (from some 7000 patients) that had been analyzed by agarose electrophoresis at the Department of Clinical Chemistry, University of Lund, Malmö, Sweden. The most common variant observed was a slow albumin (+1 net charge relative to normal albumin, denoted slow +1). Further investigation revealed additional examples with an overall phenotype frequency of 1:1000. Altogether, plasma was obtained for study from 14 apparently unrelated Swedes heterozygous for a slow +1 albumin. These are referred to as patient 1, patient 2, etc. None of 11 subjects who answered a questionnaire recognized any of the other subjects as a relative. The carriers were patients with various diagnoses unrelated to their albumin anomaly. Family studies were not feasible; however, in one instance (patient 3) the slow albumin was present in the father of the propositus. A similar slow albumin was detected by clinical electrophoresis at Christchurch Hospital in a patient of Scottish descent from Kaikoura, New Zealand, and was present in five members of the family. Albumin Tradate was found by F. Porta (Ospedale di Circolo, Varese, Italy) as a slow-moving variant in the serum of members of five unrelated families.

Analytical Methods. Agarose gel electrophoresis was done at pH 8.6 (17) and $^{63}Ni^{2+}$ autoradiography was carried out as previously described (15, 18). Rocket electrophoresis was done by the method of Laurell (19). Cellulose acetate electrophoresis at pH 8.6, analytical isoelectric focusing, and SDS/PAGE were done as described (2–4, 13). Amino acid analysis was done on a Dionex D-300 (13) or a Beckman model 121 amino acid analyzer (2–4). Sequence analysis was performed by manual Edman degradation at Christchurch (8) and Saitama (1) and by automated analysis with the Applied Biosystems model 477A sequencer at Indiana (1). In a few instances stated in the text the Beckman model 690C sequencer was used (2–4).

Protein Purification. The total albumin and its three components (A, arginyl-albumin, and proalbumin) were purified by use of several chromatographic systems: (*i*) Some albumin fractions used for amino-terminal sequence analysis and for structural study of peptides were obtained by the two-step method of ion-exchange chromatography and gel filtration (4, 5). (*ii*) The three albumin components in plasmas from Kaikoura and Malmö were each purified on a preparative scale (up to 20 ml of plasma) by ion-exchange chromatography on DEAE-Sephadex A50 using 16 mM sodium acetate buffer and a pH gradient from pH 5.2 to 4.5 (7) (Fig. 2). This procedure was also adapted for use by HPLC on an analytical scale (0.2 ml of plasma) with a DEAE 5PW column (Bio-Rad) and use of a 20 mM sodium acetate buffer. For the Tradate plasma, in addition to DEAE-Sephadex chromatography, chromatofocusing on a Mono P HR 5/20 column was used with the Pharmacia FPLC system (16).

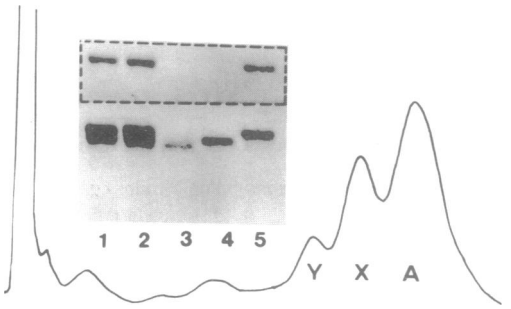

FIG. 2. DEAE-Sephadex elution profile of plasma from Malmö (father of patient 3); pH gradient 5.2–4.5. (*Inset*) Agarose gel electrophoresis (pH 8.6) of purified albumin components and $^{63}Ni^{2+}$ autoradiograph in the dashed rectangle. Lanes 3, 4, and 5 are DEAE peaks Y, X, and A, respectively. Note, neither X nor Y binds $^{63}Ni^{2+}$. Lanes 1 and 2 are diluted plasma from the Kaikoura patient and from the father of Malmö patient 3, respectively.

Protein Structural Study. In addition to amino-terminal sequence analysis an independent structural study was made of CNBr fragments and *Staphylococcus aureus* V8 protease or tryptic peptides by the standard procedure described previously (1–6). For this, the albumin fraction was reduced and carboxymethylated and then cleaved with CNBr. Analytical isoelectric focusing of the CNBr digest was used to identify the variant fragment; this was purified by HPLC and digested with a protease (4). The variant peptide was identified and purified by HPLC and submitted to amino acid analysis and automated sequence analysis.

Polymerase Chain Reaction (PCR) and DNA Analysis. The PCR and DNA analysis were carried out exactly as described by Brennan *et al.* (13) except that in the Swedish cases the DNA was extracted by using an Applied Biosystems 340A apparatus. In brief, the four oligonucleotides (TM8, TM9, TM11, and TM12) listed by Brennan *et al.* (13) were synthesized on an Applied Biosystems DNA synthesizer. A 400-base-pair fragment of the genomic DNA encoding the first exon of the albumin gene (amino acids −24 to +2) (20) was amplified by the PCR technique using TM8 and TM9 as primers. The 400-base-pair PCR product was gel-purified and sequenced as described (13). The amino acid sequence and the DNA sequence for albumin A are based on the genomic sequence of Minghetti *et al.* (20).

RESULTS AND DISCUSSION

Electrophoretic Analysis of Kaikoura, Malmö Type I, and Tradate Plasmas. Electrophoresis of the plasmas at pH 8.6 on cellulose acetate strips or on agarose showed a broad slow albumin band. An example is given in Fig. 2 for agarose electrophoresis (see *Inset*, lanes 1 and 2). Electrophoresis of the Tradate plasma on cellulose acetate at pH 5.0 (9, 10) did give resolution into three albumin components. Agarose electrophoresis of the purified proteins with concurrent autoradiography showed that the major slow albumin from the Kaikoura and Malmö plasmas ran in the same position as arginyl-albumin that had been purified from the plasma of an individual with α_1-antitrypsin Pittsburgh (13). None of these slow albumins bound $^{63}Ni^{2+}$, nor did the minor component from the Kaikoura and Malmö plasmas (Fig. 2 *Inset*, lanes 3 and 4).

Elution Profiles and Purification by Ion-Exchange Chromatography. Purified proalbumin was best obtained by ion-

exchange chromatography in 16 mM acetate buffer with a pH gradient from pH 5.2 to 4.5 (7). Fig. 2 gives an example of the elution profile of Malmö plasma (father of patient 3) on DEAE-Sephadex. The components designated Y, X, and A have been identified as proalbumin, arginyl-albumin, and albumin A, respectively, and occur in a ratio of about 5:30:65. The purified components exhibit different electrophoretic mobilities, the proalbumin (Y) migrating as slow +2 and the arginyl-albumin (X) as slow +1. Autoradiography showed that these two components did not bind $^{63}Ni^{2+}$, whereas the third component (albumin A) did (Fig. 2 *Inset*). Molecular weight estimation by SDS/PAGE gave identical values for arginyl-albumin and albumin A (65,000), but the proalbumin appeared to be slightly larger.

In most cases in the Malmö series the elution profile and component purification were accomplished on an analytical scale (0.2 ml of plasma) by HPLC on a DEAE 5 PW column with the same pH gradient as in Fig. 2 but with a 20 mM acetate buffer. The HPLC profile allowed comparison of the albumin components in a series of 14 Malmö specimens with use of purified Tradate proalbumin and arginyl-albumin as reference standards. The procedure yielded sufficient purified proalbumin, arginyl-albumin, or variant albumin for sequence analysis in the Applied Biosystems sequencer. We had already determined a different structural change in one instance (patient 5; see later), and this plasma gave a different elution profile. So also did the plasma of patient 10, for which the structural change has not yet been identified. The remaining 12 Malmö plasmas all gave similar profiles, which closely resemble the profile given in Fig. 2 by DEAE-Sephadex chromatography. In this series the distribution of components was determined by measuring the area under the peaks and was proalbumin about 3%, arginyl-albumin 20–30%, and albumin A 65–75%.

Structural Identification of Arginyl-albumin as the Major Variant Component. Arginyl-albumin was identified by amino-terminal sequence analysis of the major variant component (X) in the Kaikoura plasma, the Tradate plasma, and in 12 of the Malmö plasmas. Amino-terminal sequence analysis was done directly on the arginyl-albumin fraction either by the manual method (four to six steps), or up to 15 cycles with a protein sequencer (Applied Biosystems or Beckman), or by both the manual and the automated procedures. Individual structural study of the albumin components in these plasmas was required because four other types of proalbumins had been reported (11–16). It was also necessary to exclude the double substitution found in albumin Redhill (13).

Confirmation of Arginyl-albumin by Analysis of V8 Protease Peptides. CNBr fragment CB1 was purified by HPLC in six of the Malmö cases. A V8 protease digest was made of the CB1 fragments, and the V8 peptide profile was obtained by HPLC, which was used to isolate the variant peptide. In two cases an unequivocal sequence was obtained: *Arg*-Asp-Ala-His-Lys-Ser-Glu. This is identical to the sequence of V8 peptide S1 from albumin A except for the arginine at the amino terminus. In the other cases sequence analysis indicated the variant peptide was present together with the normal V8 peptide S1 or another normal V8 peptide from CB1. These results confirmed that arginyl-albumin was the major slow component. However, the nature of the minor proalbumin component and its relationship to the arginyl-albumin were still unclear.

Absence of the 320 Ala → Thr Substitution and of Carbohydrate in the Arginyl-albumin. A series of analytical studies showed that the arginyl-albumin purified from Malmö and Tradate plasmas was not glycosylated and did not have the substitution 320 Ala → Thr found in albumin Redhill (13). (*i*) The HPLC profile of the CNBr digest of the arginyl-albumin indicated the presence of a variant CB1 fragment, but a normal CB4 fragment (residues 299–329) rather than a variant CB4 was present. (*ii*) Amino acid analysis showed the absence of the threonine and of glucosamine in CB4 and in the four tryptic peptides isolated by HPLC from a tryptic digest of CB4. (*iii*) CB1 was the only variant CNBr fragment observed by isoelectric focusing of CNBr digests of the arginyl-albumin from the six Malmö patients examined. (*iv*) The electrophoretic patterns and the HPLC ion-exchange profiles of the albumin components in the Malmö, Kaikoura, and Tradate plasmas were all similar and differed from the patterns for the Redhill plasma. (*v*) Neuraminidase treatment of a series of Malmö plasmas to remove sialic acid produced no change in the electrophoretic mobility of the albumin components, whereas it did with Redhill plasma. (*vi*) SDS/PAGE showed that the arginyl-albumin in Malmö and Kaikoura plasmas had no detectable increase in molecular weight compared to normal albumin, but the Redhill albumin did. Despite this evidence for the identity of the arginyl-albumin components in all the cases described herein, the reason why arginyl-albumin rather than albumin A was produced from its precursor proalbumin was still uncertain.

Amino-Terminal Sequence Analysis of the Unmodified Minor Proalbumin Component. Initial amino-terminal sequence analyses of the unmodified minor component obtained by the two-step procedure indicated the characteristic propeptide sequence Arg-Gly-Val-Phe-, but the fifth residue could not be identified. Further studies were done on the minor component purified by ion-exchange chromatography either by the preparative method (7) or by the HPLC analytical method, both of which yielded proalbumin that appeared homogeneous by SDS/PAGE and by electrophoretic and chromatographic criteria. Again, manual sequence analysis of the unmodified proalbumin from all three plasma sources gave only the first four residues Arg-Gly-Val-Phe-. Automated sequence analysis for 10 cycles gave a blank at the fifth residue (the −2 position)—i.e., Arg-Gly-Val-Phe-( )-Arg-Asp-Ala-His-Lys-. This ruled out the known substitution −2 Arg → His and suggested the substitution −2 Arg → Cys, because cysteine must be alkylated to give a stable identifiable product in automated sequence analysis (21).

PCR and DNA Analysis. The discovery of the −2 Arg → Cys mutation in albumin Redhill by Brennan *et al.* (13) by use of PCR amplification and DNA analysis prompted a similar approach for study of the propeptide sequence in Kaikoura and Malmö proalbumins. DNA was extracted from whole blood from the Kaikoura propositus, from the father of Malmö patient 3, and also patients 6, 8, 13, and 15. By use of the primers TM8 and TM9 (13), PCR-amplified DNA was prepared that coded for amino acids −24 to +2, covering the entire prepropeptide sequence—e.g., the signal peptide (−24 to −7), the propeptide (−6 to −1), and the first two residues of mature albumin. The PCR-amplified DNA from several subjects was run on an agarose gel (Fig. 3 *Upper*) and was probed with ^{32}P oligonucleotide encoding cysteine at the −2 position (Fig. 3 *Lower*). The autoradiograph showed specific hybridization to the allele-specific oligonucleotide TM12. This established the substitution −2 Arg → Cys in the proalbumin from a series of unrelated individuals from Sweden and also the family from New Zealand. As in the case of albumin Redhill, sequence analysis of the amplified DNA indicated a mutation of CGT to TGT, changing the code from arginine to cysteine at position −2 (Fig. 4). This substitution accords with the slow +2 electrophoretic mobility of the proalbumin in the Kaikoura, Malmö, and Tradate cases.

Confirmation of the −2 Arg → Cys Substitution by Protein Sequence Analysis. The −2 Arg → Cys mutation was confirmed by aminoethylation of the free cysteine in the Kaikoura proalbumin without prior reduction of the albumin disulfide bonds. Electrophoretic analysis showed that the mobility of the aminoethylated proalbumin was shifted down by two positive charges owing to reaction with cysteine at the

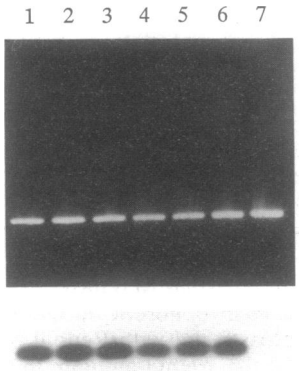

FIG. 3. (*Upper*) Agarose gel electrophoresis showing PCR amplification products of DNA encoding the prepro sequence of albumin. (*Lower*) Autoradiograph showing specific hybridization to oligonucleotide probe TM12, which encodes a cysteine at position −2. Lanes 1–4, unrelated individuals from Sweden; lane 5, patient from New Zealand; lane 6, carrier of albumin Redhill; lane 7, normal control.

−2 and +34 positions, with the result that it had a mobility of slow +4.

In further confirmation, automated sequence analysis of the pyridylethylated proalbumin established the presence of pyridylethylcysteine at the −2 position of the propeptide for the proalbumin of Malmö patient 8 and Tradate. Prior to sequence analysis the proalbumin was S-alkylated by reaction with 4-vinylpyridine *in situ* on the glass-fiber filter of the Applied Biosystems sequencer as described by Andrews and Dixon (21). Pyridylethylcysteine supplied by Applied Biosystems was used as a standard. Sequence analysis was done for 10 cycles. The amino-terminal arginine of the propeptide (−6 Arg) was unclear because of the *in situ* reaction; however, the remaining sequence was clear, including pyridylethylcysteine at the −2 position.

Other Types of Variant Malmö Albumins. In this survey of 14 cases of alloalbumins from Malmö, two other types of variants producing a broad slow albumin band were encountered. The substitution has yet to be identified in one instance (patient 10). Another variant (patient 5) involves a substitution in the second domain of the molecule and is not described here because it does not affect cleavage of the propeptide. However, albumin Bremen is described here because it was shown to have the substitution 1 Asp → Val, which is present in proalbumin Blenheim (12). No examples of the three most frequently reported European variants were observed in the Swedish population surveyed. That is, there were no cases of proalbumins of the Christchurch type (−1 Arg → Gln) or the Lille type (−2 Arg → His), or of albumin B (570 Glu → Lys). Because of their slow +2 mobility these would have readily been detected by agarose electrophoresis and easily distinguished from the slow +1 albumins. In fact, the only instance of albumin B was in a Cambodian undergoing rehabilitation at Malmö (6).

-2			-1		
Arg	CGT	Normal	Arg	CGA	Normal
Leu	C<u>T</u>T	Unknown	Gln	C<u>A</u>A	Christchurch
Pro	C<u>C</u>T	Unknown	Pro	C<u>C</u>A	Takefu
His	C<u>A</u>T	Lille	Leu	C<u>T</u>A	Jaffna
Cys	<u>T</u>GT	This work	Gly	<u>G</u>GA	Unknown
Ser	<u>A</u>GT	Unknown			
Gly	<u>G</u>GT	Unknown			

FIG. 4. Codons for the diarginyl sequence of the propeptide of normal human serum albumin A (20) and for possible effective single-base mutations. The substitutions identified to date are underlined and are discussed in the text. Potential substitutions not yet reported are designated as unknown.

Albumin Bremen (1 Asp → Val). During study of the Malmö series we investigated another serum that also exhibited a broad albumin band in electrophoresis with a mobility of slow +1, but the albumin had the exchange 1 Asp → Val, the same as in mature albumin Blenheim (12). This specimen, designated Bremen, was provided by Theodore Peters, Jr. (The Mary Imogene Bassett Hospital, Cooperstown, NY), who had received the serum from Peter Kovary of Bremen, F.R.G.

Structural study of the Bremen variant was done by the procedure described by Arai and co-workers (1, 4–6). The total albumin was purified by the two-step method, and the leading edge of the albumin peak was collected. Analytical isoelectric focusing of a CNBr digest indicated a substitution in CB1 (residues 1–87). The first cycle of automated sequence analysis gave a ratio of valine to aspartic acid of about 2:1 (allowing for the lower response of aspartic acid). This result indicated a 1 Asp → Val substitution. In confirmation, the CB1 fragment of Bremen albumin was purified by HPLC and digested with trypsin. Two forms of the amino-terminal peptide T1 were obtained by HPLC. One was the normal peptide *Asp*-Ala-His-Lys; analysis in the Beckman sequencer showed that the other was *Val*-Ala-His-Lys. This proved the 1 Asp → Val substitution, which accords with the codon change G<u>A</u>T → G<u>T</u>T.

A search for proalbumin was made after Brennan (12) discovered that proalbumin Blenheim (1 Asp → Val), which retains the normal peptide, was present in plasma (10% of the total albumin) together with mature albumin Blenheim (40%). In the Bremen case HPLC on a DEAE 5 PW column gave three components: albumin A (≈50%), the variant albumin (1 Asp → Val) (≈30%), and arginyl-albumin with the substitution 1 Asp → Val (≈20%). The latter presumably resulted from degradation of the proalbumin.

Is the Apparent Clustering of Mutations in the Propeptide Sequence an Indicator of Hypermutability or the Consequence of Ease of Ascertainment? Thus far, five different single point mutations in the diarginyl sequence of the albumin propeptide have been observed that affect processing by the convertase (Fig. 4). A single nucleotide change in the CGA codon for arginine in the −1 position can result in four different effective substitutions, three of which have been realized. Four other single-base mutations would be silent, and a fifth would produce a stop codon. Of the nine possible point mutations in the CGT codon for the −2 arginine, three would be silent and six would be effective, but only two of the latter have yet been identified. Nonetheless, the five substitutions in the diarginyl sequence equal the total number of substitutions reported to date for the first half of the 585-residue albumin molecule (1).

Is the much greater frequency of mutations identified in the diarginyl sequence a marker of hypermutability or is it the result of experimental bias such as ease or difficulty of ascertainment? Natural selection does not play an important role, because no significant adverse or beneficial effect of proalbuminemia or alloalbuminemia has yet been demonstrated, nor does the propeptide have any function once in the circulation. In the case of the −2 Cys → Arg mutation described here electrophoretic detection is difficult because of the instability and small amount of the proalbumin and the fact that the albumin band is broadened rather than split into discrete components. In fact, this variant may well have been missed in routine surveys, but it is easy to detect the other four proalbumins because of their relative stability and their slow +2 mobility, which results in discrete bands. However, electrophoretic detection is just as easy for other albumin

variants that have a slow +2 mobility resulting from a Glu → Lys exchange. In fact, there are 62 such possible sites randomly distributed in the albumin molecule, and a Glu → Lys exchange has already been reported at 9 of these sites (1). Thus, the apparent clustering of mutations in the diarginyl sequence is not attributable to experimental selection.

Proalbumin Mutations May Involve Hypermutable CpG Dinucleotides. A major factor that may contribute to the apparently higher frequency of mutation in the propeptide sequence than in the mature albumin molecule is the propensity of CpG dinucleotides to undergo recurrent mutation, most often leading to base transitions (22, 23). For example, the extent of hypermutability of CpG dinucleotides in the factor VIII gene is estimated as being 10–20 times greater than the average mutation rate for hemophilia A, and it is weighted towards transitions (22). The codons for −1 Arg and −2 Arg in the albumin propeptide are CGA and CGT, respectively (Fig. 4) (20). Five point mutations in these two codons have been observed so far (11–16) compared to some 20 point mutations reported in the exons for the 585-residue mature albumin molecule (1). Only a few of the latter involve CpG dinucleotides. It is noteworthy that two of the proalbumin mutations have apparently recurred independently in populations that are genetically distinct and geographically distant (1, 5, 15). Both of these recurrent mutations are CpG to CpA transitions on the sense strand, with the initial mutation presumably occurring on the antisense strand: CGA to CAA (−1 Arg → Gln, Christchurch type of proalbumin) and CGT to CAT (−2 Arg → His, Lille type of proalbumin) (Fig. 4). The mutation described herein is a CpG to TpG transition in the sense strand. This mutation has been found thus far only in Caucasians but is present in families of three different European nationalities—Swedish, Scottish, and Italian—and it has a phenotype frequency of 1:1000 in the Swedish group studied. The double mutant albumin Redhill (−2 Arg → Cys, 320 Ala → Thr) occurred in a family of Anglo-Saxon origin (13). Only a single family having this unusual and readily detectable glycoprotein has been reported. Because the Vikings invaded Britain, it is possible that they introduced the −2 Arg → Cys mutation, and that it was followed by the 320 Ala → Thr mutation.

Note Added in Proof. In 1966 Laurell and Nilehn (24) described a familial albumin heterogeneity characterized by a broad band in agarose electrophoresis that indicated the presence of a slow component. A family study showed that the anomalous albumin was present in 9 of 23 members representing three generations. Because of similarity of the electrophoretic pattern to that of the proalbumin described here, a search was made for the family and plasma was obtained from one of the original subjects. Purification of the albumin and protein sequence analysis by the methods described here demonstrated the presence of proalbumin (≈10%), arginyl-albumin (≈30%), and albumin A (≈60%). The −2 Arg → Cys mutation was confirmed by hybridization of the PCR-amplified DNA to an allele-specific oligonucleotide probe.

We are indebted to Dr. David R. Hathaway for use of the Applied Biosystems sequencer and to Joyce Dwulet for sequence analysis with it, and to Karen Huss for operation of the Beckman sequencer. This work was supported in part by the Medical Research Council of New Zealand (S.O.B.), by the Fundacion Federico S.A. (C.-B.L.), by a grant from the Ministero della Publica Instruzione (Rome, Italy) (M.G.), and by the National Institutes of Health (Grant DK19221 to F.W.P.).

1. Arai, K., Madison, J., Shimizu, A. & Putnam, F. W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 497–501.
2. Takahashi, N., Takahashi, Y., Blumberg, B. S. & Putnam, F. W. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4413–4417.
3. 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.
4. Huss, K., Madison, J., Ishioka, N., Takahashi, N., Arai, K. & Putnam, F. W. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6692–6696.
5. Arai, K., Madison, J., Huss, K., Ishioka, N., Satoh, C., Fujita, M., Neel, J. V., Sakurabayashi, I. & Putnam, F. W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6092–6096.
6. Arai, K., Ishioka, N., Huss, K., Madison, J. & Putnam, F. W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 434–438.
7. Brennan, S. O. (1985) *Biochim. Biophys. Acta* **830**, 320–324.
8. Brennan, S. O. & Herbert, P. (1987) *Biochim. Biophys. Acta* **912**, 191–197.
9. Galliano, M., Minchiotti, L., Iadarola, P., Zapponi, M. C., Ferri, G. & Castellani, A. A. (1986) *J. Biol. Chem.* **261**, 4283–4287.
10. Minchiotti, L., Galliano, M., Iadarola, P., Meloni, M. L., Ferri, G., Porta, F. & Castellani, A. A. (1989) *J. Biol. Chem.* **264**, 3385–3389.
11. Brennan, S. O. & Carrell, R. W. (1978) *Nature (London)* **274**, 908–909.
12. Brennan, S. O. (1989) *Mol. Biol. Med.* **6**, 87–92.
13. Brennan, S. O., Myles, T., Peach, R. J., Donaldson, D. & George, P. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 26–30.
14. Abdo, Y., Rousseaux, J. & Dautrevaux, M. (1981) *FEBS Lett.* **131**, 286–288.
15. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7403–7407.
16. Galliano, M., Minchiotti, L., Stoppini, M. & Tarnoky, A. L. (1989) *FEBS Lett.* **255**, 295–299.
17. Jeppsson, J. O., Laurell, C.-B. & Franzen, B. (1979) *Clin. Chem.* **25**, 629–638.
18. Brennan, S. O., George, P. M. & Peach, R. J. (1988) *Clin. Chim. Acta* **176**, 179–184.
19. Laurell, C.-B. (1972) *Scand. J. Clin. Lab. Invest.* **29** Suppl. 124, 21–37.
20. Minghetti, P. P., Ruffner, D. E., Kuang, W.-J., Dennison, O. E., Hawkins, J. W., Beattie, W. G. & Dugaiczyk, A. (1986) *J. Biol. Chem.* **261**, 6747–6757.
21. Andrews, P. C. & Dixon, J. E. (1987) *Anal. Biochem.* **161**, 524–528.
22. Youssoufian, H., Antonarakis, S. E., Bell, W., Griffin, A. M. & Kazazian, H. H., Jr. (1988) *Am. J. Hum. Genet.* **42**, 718–725.
23. Cooper, D. N. & Krawczak, M. (1989) *Hum. Genet.* **83**, 181–188.
24. Laurell, C.-B. & Nilehn, J.-E. (1966) *J. Clin. Invest.* **45**, 1935–1945.