Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 5959–5963, July 1991
Biochemistry

A donor splice mutation and a single-base deletion produce two carboxyl-terminal variants of human serum albumin

(alloalbumins/genetic polymorphism/frameshift mutation/exon skipping)

SCOTT WATKINS*, JEANNE MADISON*, EVELYN DAVIS*, YASUSHI SAKAMOTO*, MONICA GALLIANO†‡, LORENZO MINCHIOTTI†, AND FRANK W. PUTNAM*§

*Department of Biology, Indiana University, Bloomington, IN 47405; and †Department of Biochemistry, University of Pavia, 27100 Pavia, Italy

Contributed by Frank W. Putnam, April 15, 1991

ABSTRACT **At least 35 allelic variants of human serum albumin have been sequenced at the protein level. All except two COOH-terminal variants, Catania and Venezia, are readily explainable as single-point substitutions. The two chain-termination variants are clustered in certain locations in Italy and are found in numerous unrelated individuals. In order to correlate the protein change in these variants with the corresponding DNA mutation, the two variant albumin genes have been cloned, sequenced, and compared to normal albumin genomic DNA. In the Catania variant, a single base deletion and subsequent frameshift leads to a shortened and altered COOH terminus. Albumin Venezia is caused by a mutation that alters the first consensus nucleotide of the 5′ donor splice junction of intron 14 and the 3′ end of exon 14, which is shortened from 68 to 43 base pairs. This change leads to an exon skipping event resulting in direct splicing of exon 13 to exon 15. The predicted Venezia albumin product has a truncated amino acid sequence (580 residues instead of 585), and the COOH-terminal sequence is altered after Glu-571. The variant COOH terminus ends with the dibasic sequence Arg-Lys that is apparently removed through stepwise cleavage by serum carboxypeptidase B to yield several forms of circulating albumin.**

Approximately 35 different human serum albumin (HSA) variants, originally detected on the basis of their abnormal electrophoretic mobility, have been sequenced at the protein level (1–8). The majority reflect single-point mutations in the albumin gene. However, two chain-termination variants (designated Catania and Venezia) occur in certain population groups in Italy (1–3). In an attempt to correlate the protein change with its corresponding DNA lesion, the relevant regions of the genes of these two Italian albumin variants have been cloned, sequenced, and compared to the normal albumin genomic DNA sequence of Minghetti *et al.* (9). These variants are found only in specific geographic locations in Italy with 105 unrelated subjects carrying the Venezia trait and 62 unrelated subjects possessing the Catania allele (1) earlier designated Ge/Ct (3). Because of the relative immobility of these population groups, it has been possible to identify homozygotes for both of the variants; however, only the Venezia homozygote was available for this study.

Structural characterization showed that the HSA chain-termination mutant Catania has an altered COOH-terminal sequence in which the normal amino acid residues 580–585 (Gln-Ala-Ala-Leu-Gly-Leu) are replaced by residues 580–582 (Lys-Leu-Pro) (3). By comparison, the Venezia serum albumin also possesses a shortened polypeptide chain, having 578 residues instead of the normal 585 (2). Furthermore, residues 572 to the COOH terminus are completely variant: Pro-Thr-Met-Arg-Ile-Arg-Glu (Fig. 1). In the homozygous Venezia mutation, 80% of the protein has the mutant COOH terminus, while 20% of the albumin has the same change but has an additional COOH-terminal arginine at position 579. The major form probably results from partial proteolytic degradation of the minor form by serum carboxypeptidase B. Minchiotti *et al.* (2) proposed that this extensive modification was attributable to the deletion of exon 14 and translation to the first terminator codon of exon 15, which normally does not code for albumin.

MATERIALS AND METHODS

Southern Hybridization. Genomic DNA was isolated from Venezia and Catania whole blood by a phenol/chloroform procedure (10). Approximately 15 μg of DNA from Venezia, Catania, or human placenta was digested with *Hin*dIII and *Sca* I (GIBCO/BRL; Boehringer Mannheim) and electrophoresed in a 1% agarose gel in 1× Tris/borate buffer, pH 8.5 (89 mM Tris/89 mM boric acid/2 mM EDTA). The DNA fragments were blotted to a Zeta-Probe nylon membrane (Bio-Rad) using a NaOH transfer procedure (11). The filter was prehybridized [5× SSPE (750 mM NaCl/50 mM NaH_2PO_4/5 mM EDTA, pH 7.4), 5× Denhardt's solution, 0.1% SDS, and salmon sperm DNA at 100 μg/ml] for 2–4 hr at 65°C (12). A 1.5-kilobase (kb) *Sst* I–*Eco*RI fragment encompassing exons 13–15 of HSA was gel purified from a subclone provided by Achilles Dugaiczyk (University of California, Riverside) using GeneClean (Bio 101, La Jolla, CA). The fragment was labeled using [α-^{32}P]dCTP and random hexamers (13). After removal of unincorporated nucleotides with G-50 spin columns (5 Prime → 3 Prime, Inc.), 1 × 10^7 counts of labeled probe was added to 10 ml of hybridization solution (5× SSPE, 5× Denhardt's solution, 0.1% SDS, and salmon sperm DNA at 100 μg/ml) and incubated overnight at 65°C (12). The filter was washed twice with 2× SSPE and 0.1% SDS at room temperature for 5 min, once with 1× SSPE and 0.1% SDS at 65°C for 20 min, and finally with 0.1× SSPE and 0.1% SDS at 65°C for 20 min. Autoradiography was performed at −70°C for 3–4 days.

PCR Amplification, Cloning, and Sequencing. Two PCR primers were synthesized by the DNA synthesis facility of the Indiana University Institute for Molecular and Cellular Biology using an Applied Biosystems (Foster City, CA) model 380A DNA synthesizer (HSA primer 2, 5′ CATGCAGATGAGAATATTGAGAC 3′; HSA primer 4, 5′ GCTGTACCACTCTATTAGATTCT 3′). These primers were used to PCR amplify (14) a 1.95-kb fragment of the HSA gene encompassing the splice junctions of exons 13, 14, and 15 (9) from both Venezia and Catania. PCRs were performed according to the manufacturer's instructions [10 mM Tris·HCl,

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: HSA, human serum albumin.

‡Present address: Institute of Applied Biology, University of Sassari, 07100 Sassari, Italy.
§To whom reprint requests should be addressed.

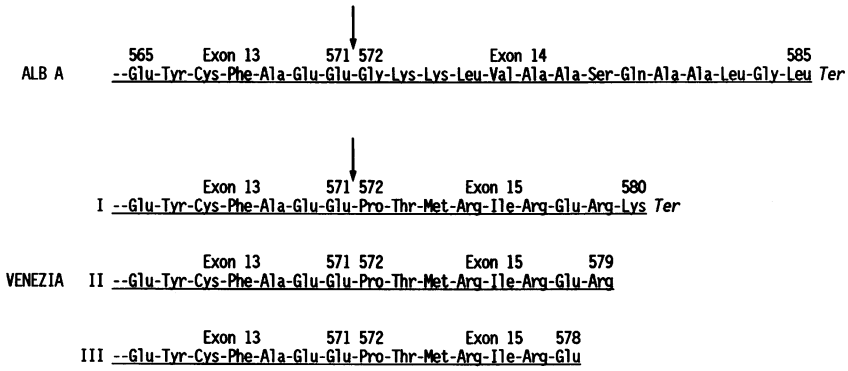

FIG. 1. Comparison of the amino acid sequences of the COOH-terminal region of normal albumin A (Alb A) and the variable mutant termini found in albumin Venezia. Venezia I represents a putative form that has not been detected in circulating serum. Venezia II comprises 20% of the circulating albumin and Venezia III makes up the remaining 80% (2). The arrows indicate exon junctions. Ter refers to a stop codon in the DNA sequence.

pH 8.3/50 mM KCl/2 mM $MgCl_2$/0.001% gelatin containing 1–2 μg of genomic DNA/1 unit of Perfect Match Enhance (Stratagene), and 2.5 units of AmpliTaq DNA Polymerase (Perkin–Elmer/Cetus)]. Using a Perkin–Elmer/Cetus DNA thermal cycler, the reactions were amplified for 30 cycles of the following protocol: denaturation for 1 min at 94°C, annealing for 2 min at 50°C, and polymerization for 3 min at 72°C with a 5-sec extension time for each successive polymerization step. The PCR products were then extracted once with phenol and twice with chloroform/isoamyl alcohol, 24:1 (vol/vol), and precipitated with 0.1 vol of 3 M NaOAc (pH 6.0) and 3 vol of 100% EtOH.

The amplified DNA from both Venezia and Catania was digested for 2 hr at 37°C with *Hin*dIII (Boehringer Mannheim) and *Sca* I (GIBCO/BRL) using buffers supplied by the manufacturers. The reaction mixture was then phenol extracted and precipitated as described above. The 1.75-kb *Sca* I–*Hin*dIII HSA gene fragment was ligated (GIBCO/BRL) into *Sma* I/*Hin*dIII-digested Bluescript KS^+ (Stratagene).

GIBCO/BRL competent cells were used for transformation as directed by the manufacturer. The transformants were plated on LB medium with ampicillin, Mg^{2+}, 5-bromo-4-chloro-3-indolyl β-D-galactoside, and isopropyl β-D-thiogalactoside and grown overnight at 37°C. White colonies were inoculated in 2XYT+Amp, and boiling minipreps (15) were used to isolate plasmid DNA. Proper insert size was confirmed by agarose electrophoresis. Approximately 4 μg of supercoiled DNA was denatured and annealed (16) to synthetic internal primers followed by double-stranded DNA sequencing using [^{35}S]dATP and Sequenase 2.0 as described in the manufacturer's (United States Biochemical) manual.

RESULTS AND DISCUSSION

Southern Analysis. Southern hybridization analysis of Catania, Venezia, and wild-type human DNA (placenta) clearly indicates a single 1.75-kb *Hin*dIII–*Sca* I band in each lane (Fig. 2) when probed with a 1.5-kb *Eco*RI–*Sst* I fragment that is completely internal to the target fragment. However, the Venezia band appears to be slightly smaller, indicating that the mutation may be due to a deletion. The comigration of the Catania 1.75-kb *Hin*dIII–*Sca* I fragment with the wild-type DNA fragment (human placenta) suggests that there is no major rearrangement in Catania genomic DNA.

DNA Sequence Analysis. To examine the DNA lesions more closely, 1-μg samples of Venezia, Catania, and human placental DNA were PCR-amplified with HSA-specific primers, subcloned into pBluescript (Stratagene), and sequenced.

Catania. Sequence analysis of six independent clones from an individual heterozygous for the Catania mutation indicates deletion of the cytosine residue at position 15985 (Fig. 3). Three of the clones matched the published sequence of normal HSA (9), and the other three clones exhibited the single-base deletion. This deletion leads to a frameshift mutation in exon 14, generating a mutant COOH terminus as predicted by Galliano *et al.* (3), who reported the variant protein sequence to be a replacement of residues 580–585 (Gln-Ala-Ala-Leu-Gly-Leu) (Fig. 1) by residues 580–582 (Lys-Leu-Pro).

Venezia. For the homozygous Venezia variant of HSA, three independent clones from the same PCR-amplification event were sequenced in their entirety on both strands of the 1.75-kb *Hin*dIII–*Sca* I fragment for a total of 10.5 kb. The major change in Venezia is a mutation that involves the last 29 nucleotides of exon 14 and the first consensus nucleotide of intron 14 (Fig. 4). This mutation may be due either to deletion of bases 16,000–16,029 (9) and a five-base insertion consisting of 5′ AAAAT 3′ or to multiple small deletions in this region. In sequencing these three mutant clones, we found five "isolated" nucleotides that were identical to the human placental DNA clone (wild-type) but different from the published sequence of the normal HSA gene (9), suggesting that several polymorphisms may exist in the general population. The following nucleotide changes were seen: position 15229, C → T; 15542, G → A; 15557, A → C; 15748, T → G; and 15813, A → G. All position numbers are based on Minghetti *et al.* (9), who also noted a polymorphism at nucleotide 15229. The polymorphism we observed at position 15229 in both the human placenta and the Venezia DNA leads to the loss of an *Sst* I site, causing the HSA gene to be present on a single 20.7-kb *Sst* I fragment instead of on the two expected 16.3-kb and 4.4-kb fragments. An additional nucleotide change (not listed) that we found in all three mutant clones proved to be the result of an early PCR error as

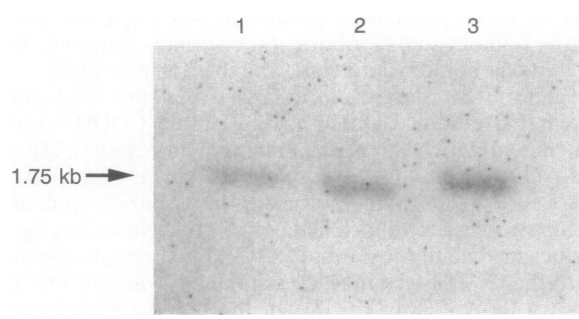

FIG. 2. Hybridization of a 1.5-kb *Sst* I–*Eco*RI probe encompassing exons 13 and 14 of the gene encoding HSA to a Southern blot of genomic DNA cut with *Hin*dIII–*Sca* I. Lanes: 1, wild-type (human placenta); 2, Venezia; 3, Catania.

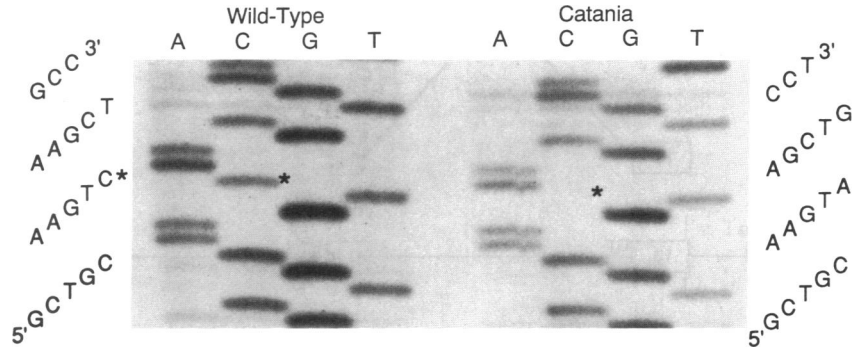

FIG. 3. Sequence ladder of the Catania HSA variant showing deletion (*) of the single cytosine at position 15985. [Numbering is according to Minghetti *et al.* (9).]

evidenced by two independent repeat experiments that gave results identical to the published DNA sequence for HSA.

The mutant HSA found in Venezia has been shown by protein sequence analysis to result from replacement of the last 14 amino acid residues by 7 or 8 different amino acids, leading to a shortened COOH terminus (Fig. 1) (2). This abnormality could have been accounted for either by genomic deletion of exon 14 or by an aberrant splicing event during mRNA maturation (2). The splicing of higher eukaryotic pre-mRNAs is believed to require three elements: the 5′ donor splice site, the 3′ acceptor splice site, and a pyrimidine-rich region just upstream of the 3′ splice site (17). The 5′ and 3′ splice sites conform to well-established consensus sequences in which the G-T and A-G dinucleotides, respectively, are universally conserved. Mutations at either of these sites may abolish or greatly reduce normal splicing, activate cryptic splice sites, or lead to exon skipping. From our DNA sequencing, it is clear that both the donor and the acceptor splice junctions of intron 13, as well as the acceptor splice junction of intron 14, are unmutated and intact, but the donor splice junction of intron 14 has been altered from the consensus G-T to T-T. Apparently, this splicing mutation causes exon 14 to be skipped, resulting in direct joining of exon 13 to exon 15 (Fig. 5).

It has been proposed that effective splicing requires a 5′ splice junction sequence of 5′ (C/A)AG/GU(G/A)AGU 3′ (18). Within this consensus splice sequence, the G at intron position 1 is obligatory and its mutation can abolish normal splicing. A lariat intermediate is formed between this G and an invariant A in the pyrimidine-rich region at the branch point. Complementary base pairing of the 5′ donor splice site with nucleotides 4–11 of the U1 small nuclear RNA (5′ CUUACCUG 3′) is one of the key steps in the splicing mechanism (19, 20). The most frequently formed pairings comprise only 5–7 bases. As a result of the mutation affecting the 5′ donor splice junction between exon 14 and intron 14, the mutated Venezia pre-mRNA has fewer possible base pairings with the U1 RNA.

Exon Skipping in Donor Splice Mutations. Other examples of mutated donor splice junctions causing exon skipping have been reported—e.g., Ehlers–Danlos syndrome, in which a shortened pro-α2(I) collagen chain is produced (21), one type of β-thalassemia (22), and classical phenylketonuria (23). In that case, a G-T → A-T transition at the 5′ donor splice site of intron 12 results in deletion of exon 12, causing direct splicing of exon 11 to exon 13. Furthermore, this deletion in phenylketonuria produces a frameshift mutation that shortens phenylalanine hydroxylase by 52 amino acids, leading to protein instability and complete loss of enzymatic activity (23).

In the liver of Nagase analbuminemic rats (which do not produce albumin), a seven base-pair deletion in the albumin gene, extending from base 5 to base 11 in the 5′ donor splice site of intron H-I, leads to the precise deletion of exon H (24).

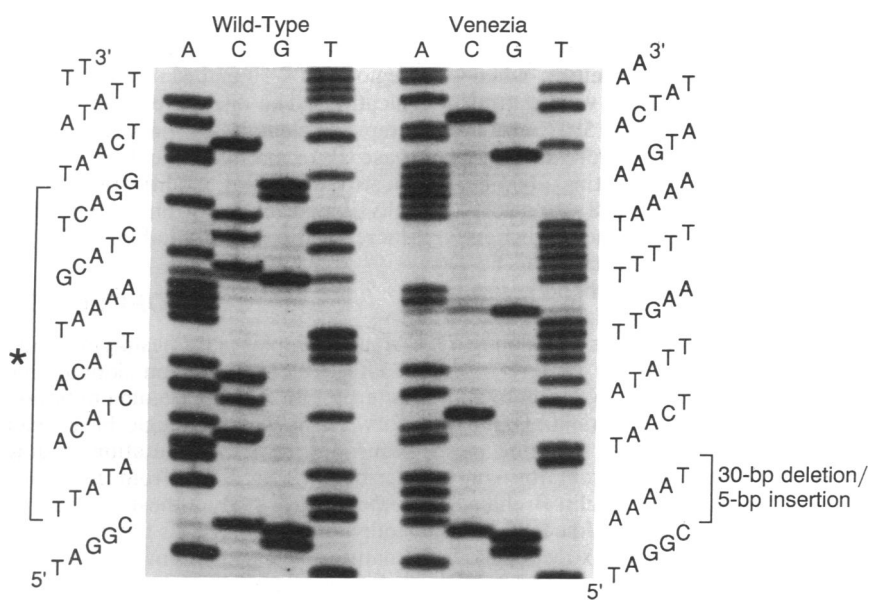

FIG. 4. Sequence ladder of the Venezia HSA variant showing the 30-base deletion and the 5-base insertion (right-hand bracket). The region in brackets on the left (*) shows the 30-base sequence that is present in wild-type DNA (human placenta) but missing in the Venezia clones.

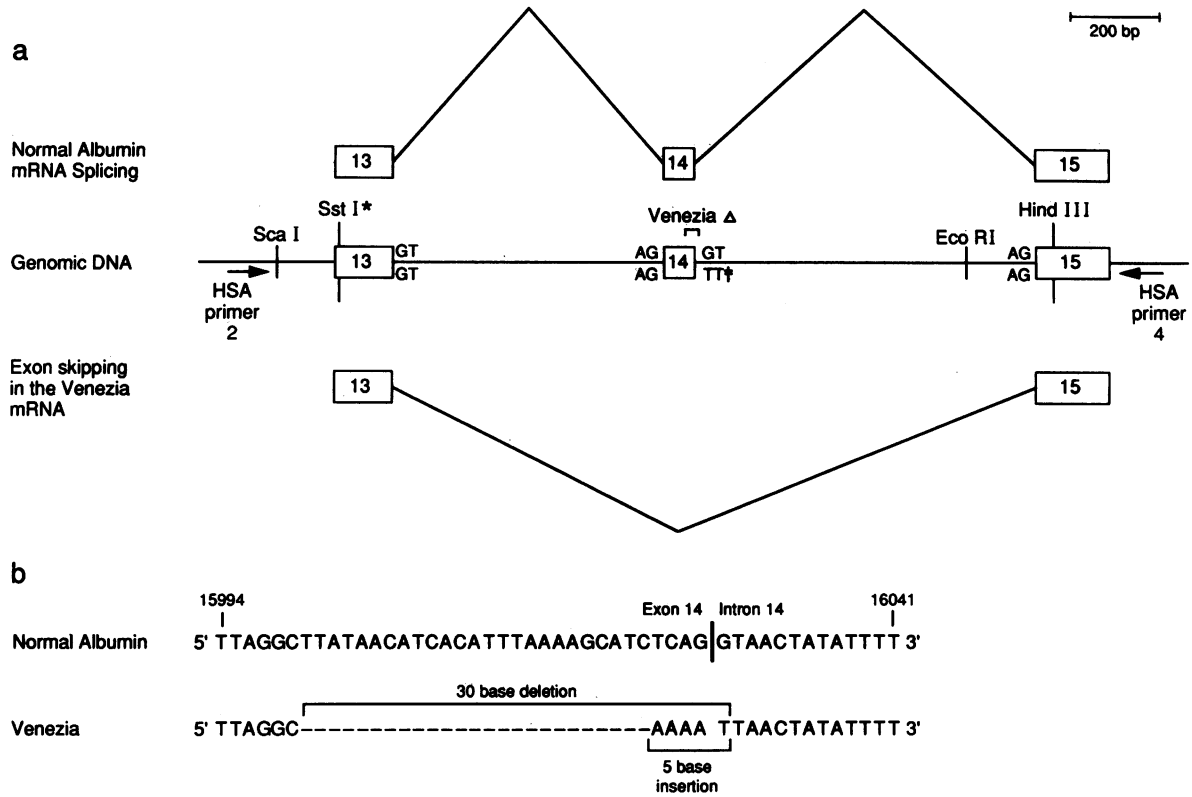

FIG. 5. (*a*) Schematic representation of an alternative splicing event: the skipping of exon 14 caused by deletion of the 5′ donor splice junction of exon 14. The approximate position and extent of the deletion is indicated by the square bracket. The obligatory donor and acceptor nucleotides for normal albumin are shown above the line on the genomic DNA map. The altered Venezia splice junction is shown below the line as indicated by the double dagger (‡). Relevant restriction sites and locations of the PCR primers are shown on the genomic map. The polymorphic *Sst* I site is indicated by a star (*). The figure is drawn to scale except for the PCR primers. (*b*) Comparison of the nucleotide sequence of normal albumin with that of Venezia at a region near the exon 14/intron 14 junction. [Numbering is according to Minghetti *et al.* (9).]

This produces a frameshift in the mRNA resulting in a premature termination codon.

So far as we know, the only previously reported mutation in the human albumin gene found to cause a splicing error is a single base substitution changing the universal A-G dinucleotide to G-G at the 3′ end of intron 6 (25). In that case, no serum albumin was produced, and analbuminemia resulted from the splicing defect.

Putative Effects of Serum Carboxypeptidase B. In Venezia, the DNA sequence analysis of exon 15 matches the published sequence for normal serum albumin (9). Therefore, when exon 13 is spliced to exon 15 in the Venezia variant, the albumin should end in a lysine residue at position 580 (Fig. 1). However, analysis of a homozygous Venezia individual has previously demonstrated that 80% of the circulating albumin ends with a glutamic acid residue at position 578 while 20% of the circulating albumin terminates in an arginine at position 579 (2). Furthermore, no circulating albumin has been detected with a lysine at the COOH terminus. We hypothesize that these basic COOH-terminal amino acid residues (lysine and arginine) are proteolytically degraded by a circulating carboxypeptidase B (2).

There are several known examples of proteolytic removal of basic COOH-terminal residues by carboxypeptidase B-like enzymes: anaphylatoxins C3a, C4a, and C5a (arginine is released) (26, 27); vasoactive peptides derived from fibrin or fibrinogen (lysine is released) (28); and the insulin B chain (arginine is released) (29). Part of the activation mechanism of proinsulin to its mature form involves an exopeptidase with carboxypeptidase B-like specificity. This activity leads to the sequential removal of COOH-terminal basic amino acid residues from the insulin B chain generating functional insulin (29).

Stability and Conformation of the COOH-Terminal Variants of Serum Albumin. Although subjects homozygous for the Catania and Venezia variants seem to be asymptomatic, the variants appear to differ in stability from normal albumin A. In heterozygous subjects the point mutants of albumin are generally present in the same concentration as normal albumin. However, both COOH-terminal variants are present in lower amount than the normal albumin (30:70%) (2). The currently available crystal structure map (resolution, 4.0 Å) shows residues 579 and 580 to fall in or near a disordered region (refs. 30 and 31; Daniel C. Carter, personal communication). The disordered conformation probably facilitates stepwise cleavage of the Venezia albumin by carboxypeptidase B to produce the forms illustrated in Fig. 1. This post-translational modification may affect the binding of physiologically important ligands such as fatty acids and also increase the rate of catabolism (2).

CONCLUSIONS

Many examples of frameshift mutations are known, but fewer confirmed cases of exon skipping have been reported. The interpretation of the Catania mutation is essentially straightforward in that the single base deletion at the DNA level causes a frameshift mutation leading to an altered COOH terminus as predicted from the protein sequence (3). On the other hand, as hypothesized from the protein data for an individual homozygous for the Venezia trait (2), most of the mature RNA is apparently generated by the skipping of exon 14. The current work has confirmed that this exon skipping is due to an extensive modification that involves the 3′ end of exon 14 and the first consensus nucleotide of the 5′ donor splice junction of intron 14. Mutation of the exon 14/intron

14 junction inactivates the selection of this splice site and, as a consequence, exon 13 is spliced directly to exon 15. The protein product of the mutated Venezia gene exists in several forms probably because of stepwise processing of the disordered polypeptide tail by carboxypeptidase B.

We are indebted to Dr. Achilles Dugaiczyk for providing albumin genomic and cDNA clones and restriction maps and for helpful advice and criticism. We thank Lawrence Washington for synthesis of the DNA primers. This work was supported in part by grants (40% and 60%) from the Ministero dell' Universita e della Ricerca Scientifica (Rome) (M.G. and L.M.) and a North Atlantic Treaty Organization grant for international collaboration (CRG 910029 to F.W.P.) and by National Institutes of Health Grant DK19221 (to F.W.P.).

1. Galliano, M., Minchiotti, L., Porta, F., Rossi, A., Ferri, G., Madison, J., Watkins, S. & Putnam, F. W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8721–8725.
2. Minchiotti, L., Galliano, M., Iadarola, P., Meloni, M. L., Ferri, G., Porta, F. & Castellani, A. A. (1989) *J. Biol. Chem.* **264**, 3385–3389.
3. Galliano, M., Minchiotti, L., Iadarola, P., Zapponi, M. C., Ferri, G. & Castellani, A. A. (1986) *J. Biol. Chem.* **261**, 4283–4287.
4. Arai, K., Madison, J., Shimizu, A. & Putnam, F. W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 497–501.
5. Takahashi, N., Takahashi, Y., Blumberg, B. S. & Putnam, F. W. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4413–4417.
6. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7403–7407.
7. 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.
8. Brennan, S. O., Arai, K., Madison, J., Laurell, C.-B., Galliano, M., Watkins, S., Peach, R., Myles, T., George, P. & Putnam, F. W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3909–3913.
9. 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.
10. Gross-Bellard, M., Oudet, P. & Chambon, P. (1973) *Eur. J. Biochem.* **36**, 32–38.
11. Reed, K. C. & Mann, D. A. (1985) *Nucleic Acids Res.* **13**, 7207–7221.
12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
13. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **132**, 6–13.
14. Mullis, K. B., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986) *Cold Spring Harbor Symp. Quant. Biol.* **51**, 263–273.
15. Holmes, D. S. & Quigley, M. (1981) *Anal. Biochem.* **114**, 193–197.
16. Zagursky, R. J., Baumeister, K., Lomax, N. & Berman, M. L. (1985) *Gene Anal. Tech.* **2**, 89–94.
17. Lewin, B. (1990) *Genes IV* (Cell, Cambridge, MA), pp. 595–604.
18. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
19. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. (1980) *Nature (London)* **283**, 220–224.
20. Rogers, J. & Wall, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1877–1879.
21. Weil, D., Bernard, M., Combates, N., Wirtz, M. K., Hollister, D. W., Steinmann, B. & Ramirez, F. (1988) *J. Biol. Chem.* **263**, 8561–8564.
22. Treisman, R., Proudfoot, N. J., Shander, M. & Maniatis, T. (1982) *Cell* **29**, 903–911.
23. Marvit, J., DiLella, A. G., Brayton, K., Ledley, F. D., Robson, K. J. H. & Woo, S. L. C. (1987) *Nucleic Acids Res.* **15**, 5613–5628.
24. Shalaby, F. & Shafritz, D. A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2652–2656.
25. Ruffner, D. E. & Dugaiczyk, A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2125–2129.
26. Bokisch, V. A. & Muller-Eberhard, H. J. (1970) *J. Clin. Invest.* **49**, 2427–2436.
27. Gorski, J. P., Hugli, T. E. & Muller-Eberhard, H. J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5299–5302.
28. Belew, M., Gerdin, B., Lindeberg, G., Porath, J., Saldeen, T. & Wallin, R. (1980) *Biochim. Biophys. Acta* **621**, 169–178.
29. Kemmler, W., Peterson, J. D. & Steiner, D. F. (1971) *J. Biol. Chem.* **246**, 6786–6791.
30. Carter, D. C., He, X.-m., Munson, S. H., Twigg, P. D., Gernert, K. M., Broom, M. B. & Miller, T. Y. (1989) *Science* **244**, 1195–1198.
31. Carter, D. C. & He, X.-m. (1990) *Science* **249**, 302–303.