Eur. J. Biochem. **268**, 344–352 (2001) © FEBS 2001

A nucleotide insertion and frameshift cause albumin Kénitra, an extended and O-glycosylated mutant of human serum albumin with two additional disulfide bridges

Lorenzo Minchiotti[1], Monica Campagnoli[1], Antonio Rossi[1], Maria E. Cosulich[1], Maria Monti[2], Piero Pucci[2], Ulrich Kragh-Hansen[3], Brigitte Granel[4], Patrick Disdier[4], Pierre-Jean Weiller[4] and Monica Galliano[1]

[1]*Dipartimento di Biochimica 'A. Castellani', Università di Pavia, Italy;* [2]*Centro Internazionale di Servizi di Spettrometria di Massa e Dipartimento di Chimica Organica e Biologica, 2° Facoltà di Medicina e Chirurgia, Università di Napoli Federico II, Napoli, Italy;* [3]*Department of Medical Biochemistry, University of Aarhus, Aarhus, Denmark;* [4]*Hopital d'adultes de la Timone, Marseille, France*

Albumin Kénitra is a new type of genetic variant of human serum albumin that has been found in two members of a family of Sephardic Jews from Kénitra (Morocco). The slow-migrating variant and the normal protein were isolated by anion-exchange chromatography and, after treatment with CNBr, the digests were analyzed by two-dimensional electrophoresis in a polyacrylamide gel. The CNBr peptides of the variant were purified by reverse-phase high performance liquid chromatography and submitted to sequence analysis. Albumin Kénitra is peculiar because it has an elongated polypeptide chain, 601 residues instead of 585, and its sequence is modified beginning from residue 575. DNA structural studies showed that the variant is caused by a single-base insertion, an adenine at nucleotide position 15 970 in the genomic sequence, which leads to a frameshift with the subsequent translation to the first termination codon of exon 15. Mass spectrometric analyses revealed that the four additional cysteine residues of the variant form two new S–S bridges and showed that albumin Kénitra is partially O-glycosylated by a monosialylated HexHexNAc structure. This oligosaccharide chain has been located to Thr596 by amino-acid sequence analysis of the tryptic fragment 592–597.

Keywords: human serum albumin; extended genetic variant; O-glycosylation; disulfide bridges; frameshift.

Bisalbuminemia is a rare inherited trait characterized by the presence of a circulating variant of human serum albumin (HSA) that is usually detected during screening of plasma proteins either in routine clinical electrophoresis or in systematic genetic surveys [1]. The occurrence of two albumin bands in electrophoresis, representing normal and variant protein, has been observed with a frequency of 0.0003–0.0010 in the average population, but is not clearly associated with disease [1]. The only reported cases of clinical relevance are the Arg218→His and Arg218→Pro mutations that have been found to be responsible for familial dysalbuminemic hyperthyroxinemia [2–4], and Leu66→Pro, which is responsible for familial dysalbuminemic hypertriiodothyroninemia [5].

In the last two decades, several genetic variants of HSA have been studied in different laboratories in an attempt to correlate the changes in the protein sequence with the functional properties and stability of the molecule. These studies have allowed the characterization by protein and/or DNA sequence analysis of more than 60 different genetic mutations [6,7] (see the continuously updated website at http://www.albumin.org) and have provided a unique insight into albumin's ligand binding sites [6,8,9]. Nearly all the known albumin mutations occur on the molecular surface and most of them are clustered in three regions: the propeptide and N-terminal region, and the C-terminal parts of subdomains IIB and IIIB [10]. The vast majority of them reflects the existence of single-base changes in the structural gene, and the recurring alloalbumins are associated with mutations in hypermutable CpG dinucleotides [11,12].

Albumin synthesis is governed by a single copy gene codominantly expressed why heterozygous subjects usually show the presence of the normal and the variant proteins in a 1 : 1 ratio. However, four different genetic events have been reported to give rise to low-concentration chain termination mutants. These include: (a) single nucleotide deletions and subsequent frameshifts in albumins Catania [13,14] and Bazzano [15]; (b) a 25-bp deletion that results in the skipping of exon 14 in albumin Venezia [14,16]; (c) a GT to CT 5′ splice mutation in intron 13 of albumin Rugby Park [17]; and (d) the introduction of an alternative 3′ splice site in intron 13 of albumin Banks Peninsula [7]. Normal HSA is not glycosylated but there are three point mutations that give rise to the canonical Asn-Xaa-Thr/Ser tripeptide acceptor sequence for N-glycosylation. In every case, the mutated sequences turned out to be N-glycosylated: Asp63→Asn [18], Ala320→Thr [19] and Asp494→Asn [10,20].

In the present work we have identified, using both protein and DNA sequence analysis, the molecular changes of a new type of slow-migrating alloalbumin found in a

Correspondence to M. Galliano, Dipartimento di Biochimica 'A. Castellani' Università di Pavia, via Taramelli 3B, I-27100 Pavia, Italy. Fax: + 39 0382423108, Tel.: + 39 0382507724, E-mail: galliano@unipv.it

Abbreviations: HSA, human serum albumin; SSCP, single strand conformation polymorphism; ESMS, electrospray mass spectrometry; pI, isoelectric point; Hex, hexose; HexNAc, *N*-acetyl hexosamine.

Enzymes: trypsin (EC 3.4.21.4); proteinase K (EC 3.4.21.64).

(Received 5 October 2000, accepted 9 November 2000)

66-year-old patient and her sister, Sephardic Jews from Kénitra (Morocco) [21]. This variant, which represents about 15% of the total plasma albumin, is caused by a single-base insertion, an adenine at nucleotide position 15 970 in the genomic sequence [22], which leads to a frameshift and subsequent translation to the first termination codon of exon 15. This results in an elongated polypeptide chain, 601 residues instead of 585, with a modified sequence from residue 575. Mass spectrometric analyses reveal that the abnormal C-terminal sequence of this mutant contains two additional disulfide bridges and an O-linked oligosaccharide chain. Thus, albumin Kénitra represents the first example of an extended and O-glycosylated circulating mutant of HSA.

MATERIALS AND METHODS

Albumin purification

The serum sample (2 mL) was desalted through a PD10 column, freeze dried, dissolved in 500 μL of 20 mM sodium acetate, pH 5.2 (buffer A), and injected on a preparative DEAE-5PW column (21.5 × 150 mm; Bio-Rad). Elution was achieved using a 40-min linear pH gradient from 100% buffer A to 100% buffer B (20 mM sodium acetate, pH 4.5) at a flow rate of 5 mL·min^{-1} [18]. Peaks were analyzed by polyacrylamide gel electrophoresis at pH 8.3 and the fractions containing the variant and the normal albumins were pooled and lyophilized.

Protein structural studies

After reduction and carboxymethylation, normal and variant albumins were cleaved with CNBr and the digests were compared by two-dimensional electrophoresis to identify the fragment containing the substitution site. The first dimension, isoelectric-focusing, was performed on laboratory-made polyacrylamide gels cast on GelBond, dried and rehydrated to the original thickness of 0.5 mm with a solution containing 8 M urea and carrier ampholytes in the pH range 2.5–8.0 [13]. After isoelectric focusing, the gel strips were equilibrated with SDS and placed on top of vertical 17% polyacrylamide SDS gels, and the second dimension was carried out using a Mini PROTEAN II cell (Bio-Rad).

The major part of the CNBr digests were then analysed by RP-HPLC on a Vydac C18 column (4.6 × 250 mm; The Separation Group). The lyophilized digests were dissolved in 50 μL of 70% formic acid + 150 μL of 0.1% aqueous trifluoroacetic acid (solvent A), and 100 μL (corresponding to about 10 nmol) were injected onto the Vydac C18 column equilibrated with solvent A. Peptides were eluted at a flow rate of 1 mL·min^{-1} with the equilibration buffer for 8 min followed by a two-step linear gradient: from 0 to 30% of solvent B (0.1% trifluoroacetic acid in acetonitrile/2-propanol, 2 : 1, v/v) in 3 min and from 30 to 50% of solvent B in 60 min.

Amino-acid sequence analysis of the two purified variant peptides were performed by an Applied Biosystems model 477 A (Aarhus) or by a Hewlett-Packard model G 1000 A sequencer (Centro Grandi Strumenti, Università di Pavia).

The lyophilized CB7M was dissolved in 0.1 M NH_4HCO_3 and digested with trypsin (Roche) for 3 h at 37 °C using an enzyme/substrate ratio of 1 : 50. The reaction was terminated by addition of trifluoroacetic acid and the digest was directly injected onto the Vydac C18 column equilibrated in 0.1% aqueous trifluoroacetic acid (solvent A). Tryptic peptides were eluted with a 60-min linear gradient from 0 to 30% of solvent B (0.1% trifluoroacetic acid in acetonitrile) at a flow rate of 1 mL·min^{-1} and submitted to amino-acid sequence analysis.

DNA structural studies

Genomic DNA was extracted from whole blood by using the lysis buffer and proteinase K digestion protocol [23]. A 339-bp fragment containing exon 14 of the albumin gene was amplified by PCR using primers A27A and A28A [23]. PCR was performed in a 25-μL reaction volume using Ready to Go Beads (Amersham Pharmacia Biotech) with a final $MgCl_2$ concentration of 1.5 mM. Conditions for amplification included an initial DNA denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 30 s, primer annealing at 64 °C for 30 s and elongation at 72 °C for 30 s. Finally an extension at 72 °C was performed for 3 min. Single strand conformation polymorphism (SSCP) and heteroduplex analysis was performed either on the whole PCR product or on restriction fragments generated by digestion with *Alu*I (Amersham Pharmacia Biotech). To test for SSCP and heteroduplexes, PCR products and their restriction digests were denatured at 95 °C for 3 min and placed on ice before the electrophoretic separation. Samples were then loaded on horizontal ultra-fine gels (gel thickness 0.3 mm) composed of 15% acrylamide (acrylamide/piperazine diacrylamide 85 : 1) with 8% glycerol in 120 mM Tris/formate buffer, pH 9.0. The electrodes consisted of paper wicks soaked in 1.04 M Tris/borate, pH 9.0, and the gels were run in a Pharmacia Multiphore II apparatus at 6 °C for 90 min at 0.8 Watt·cm^{-1} [24]. Bands were visualized by silver staining.

In order to confirm the mutation by PCR, a restriction site for the mutated allele was introduced using a modified primer 5′-GTCTTTGTGTTCAGGGTAAATTA-3′ (A31M; the modified nucleotides are underlined). A31M was used with primer 5′-AGCATCTCAGGTAACTATAT-3′ (A32M) to amplify a 92-bp PCR fragment. The thermocycle profile was as follows: initial DNA denaturation at 94 °C for 3 min, followed by 35 cycles of 30 s at 94 °C, 30 s at 52 °C and 20 s at 72 °C. The PCR products were digested with *Tru*9I (Roche) according to the manufacturer instructions and the restriction digests were analyzed by electrophoresis in nondenaturing 15% polyacrylamide gels in 0.09 M Tris/borate, 2 mM EDTA, pH 8.0 in a Mini PROTEAN II cell (Bio-Rad) followed by silver staining.

Reduction and alkylation of cysteine residues

An aliquot of albumin Kénitra was reduced in 0.25 M Tris/HCl, pH 8.5, containing 1.25 mM EDTA and 6 M guanidinium chloride with a 10 : 1 molar excess of dithiothreitol over the SH groups, for 2 h at 37 °C under nitrogen atmosphere. Alkylation of cysteine residues on the native or reduced protein was carried out in the denaturing buffer with 10 : 1 molar excess of iodoacetamide over the total SH groups for 30 min at room temperature, in the dark under nitrogen atmosphere. Alkylated protein samples were desalted by RP-HPLC on a Vydac C4 column

(25 × 0.46 cm, 5 μm) using 0.1% trifluoroacetic acid (solvent A) and 0.07% trifluoroacetic acid in 95% acetonitrile (solvent B) by means of a linear gradient from 20 to 95% solvent B in 10 min. The fractions were manually collected and directly analysed by electrospray mass spectrometry (ESMS).

Isolation and chemical derivatization of oligosaccharide chains

O-Linked oligosaccharides were released from the glycosylated CB7M fraction by reductive elimination carried out in 50 mM NaOH containing 1 M sodium borohydride at 45 °C for 18 h [25]. The oligosaccharides were permethylated with methyl iodide in dimethylsulfoxide using the sodium hydroxide procedure to increase the sensitivity of MALDI MS analysis [26]. Derivatized samples were diluted with water, extracted in chloroform and dried in a Speed-Vac centrifuge.

Mass spectrometric analyses

Intact proteins or individual peptide fractions from CNBr digestion of albumin Kénitra were submitted to ESMS analysis using either a VG Bio-Q triple quadrupole (Micromass) or an API 100 single quadrupole (PerkinElmer Applied Biosystems Inc) mass spectrometer. Samples were dissolved in 1% acetic acid in 50% acetonitrile and injected into the mass spectrometer at a flow rate of 5 $\mu L \cdot min^{-1}$. The quadrupole was scanned from *m/z* 600–1800 at 10 s per scan and the spectra were acquired using MASSLYNX or the BIOMULTIVIEWER software. Calibration was performed by a separate injection of myoglobin (*m* 16951.5 Da). All mass values are reported as average masses.

MALDI MS analysis of oligosaccharides was carried out using a Voyager DE mass spectrometer (PerSeptive Biosystems). The mass range was calibrated with the molecular ions of two peptides, WMDF-NH2 at *m/z* 598.7 and angiotensin I at 1297.5 as internal standards. Samples were dissolved in methanol and a 1-μL aliquot was mixed with 1 μL of a solution of 2,5-dihydroxybenzoic acid (10 $mg \cdot mL^{-1}$) in acetonitrile/0.2% trifluoroacetic acid 2 : 1 (v/v), applied to a sample slide and allowed to air-dry. Raw data were analysed by using a computer software provided by the manufacturer and are reported as average masses.

RESULTS AND DISCUSSION

Protein structural studies

The variant protein and its normal counterpart were isolated from the patient's serum by anion-exchange chromatography on a DEAE-5PW column. As expected on the basis of the electrophoretic mobility on cellulose acetate where it comigrates with α_1 antitrypsin [21], the variant peak eluted at a pH value of about 0.2 units higher than that of the corresponding normal albumin (Fig. 1). The ratio between the quantities of the two albumins obtained, 0.175, is in

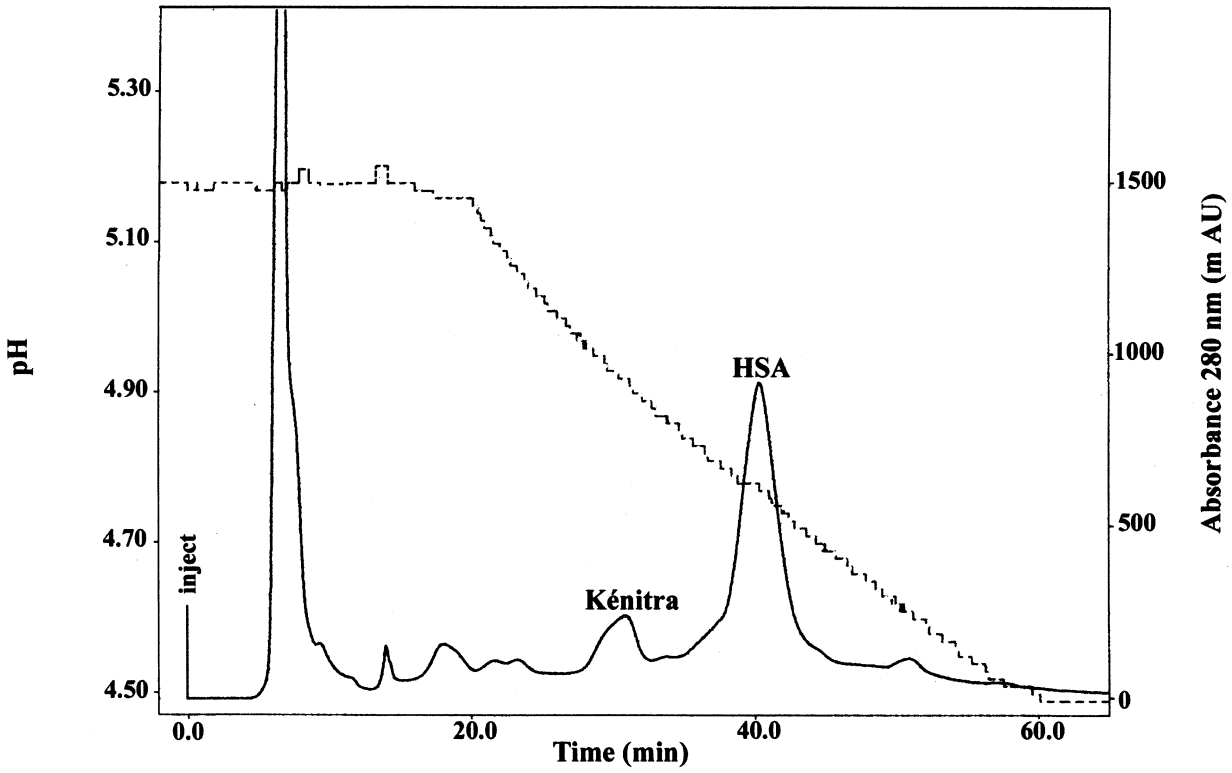

Fig. 1. HPLC elution profile on a DEAE-5PW column of the Kénitra serum. The serum sample (2 mL) was desalted by a PD10 column, freeze dried, dissolved in 500 μL of 20 mM sodium acetate, pH 5.2 (buffer A) and injected on the DEAE-5PW column equilibrated with the same buffer. Elution was achieved using a 40-min pH-linear gradient (dashed line) from 100% buffer A to 100% buffer B (20 mM sodium acetate, pH 4.5) at a flow rate of 5 $mL \cdot min^{-1}$.

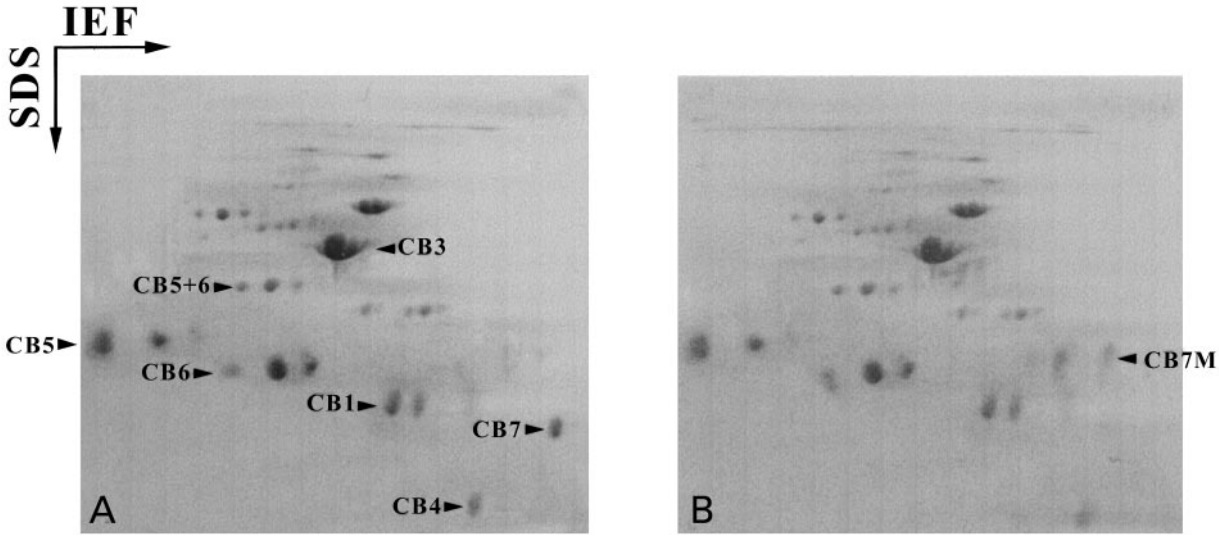

Fig. 2. Two-dimensional electrophoresis of the CNBr fragments from carboxymethylated normal (A) and Kénitra (B) albumins. The first dimension (horizontal) is isoelectric focusing in the presence of 8 M urea in the pH range 2.5–8.0. The second dimension (vertical) was carried out as SDS/PAGE on 17% polyacrylamide gels which were stained with Coomassie Brilliant Blue. Fragments CB2 and CB8M were not identified, probably owing to their high solubility and low affinity for the dye. Microheterogeneity, mostly in the case of larger fragments, is probably due to partial cleavage, oxidation of unreacted cysteines or to deamidation. CB7M denotes the variant peptide of albumin Kénitra.

keeping with the value determined by scanning the two albumin bands on cellulose acetate electrophoresis [21]. In order to localize the mutation, Kénitra and normal albumins were reduced, carboxymethylated, cleaved with cyanogen bromide and examined by two-dimensional electrophoresis in a polyacrylamide gel (Fig. 2). This analysis showed that all the CNBr fragments from the two albumins were identical with the exception of fragment CB7 in the variant, which gave rise to three spots with values of isoelectric point (pI) in the pH range 4.55–4.65, while the corresponding normal peptide migrated as a single spot with a pI value of 4.3. The finding of multiple spots suggested the presence of different forms of the abnormal fragment. In addition, their slower mobilities in the second dimension indicated a significant increase in the molecular mass (Fig. 2). The CNBr mixtures from the two proteins were resolved by RP-HPLC on a Vydac C18 column (Fig. 3) and all the fractions were identified by N-terminal sequence analysis. In the profile of albumin Kénitra, fragment CB7M is clearly eluted earlier than normal and appears as a broad and heterogeneous peak. In addition, a new and very hydrophilic peak, that we have labelled CB8M, shows up. Both CB7M and CB8M were directly submitted to amino-acid sequence determination.

Fragment CB7M gave a single 49 amino-acid sequence: DDFAAFVEKCCKADDKETCFAEEGKKTCCCKSSCLRLITSHLKASQPTM, which is completely different from that of the normal protein following Thr575 (underlined).

Sequence analysis showed that the hydrophilic peak, CB8M, consists of a new basic tetrapeptide, RIRE. This additional fragment must be positioned at the C-terminus of the bisalbumin molecule because of the presence of a methionine at the C-terminal position of CB7M.

The results obtained so far indicate that albumin Kénitra has an elongated polypeptide chain, 601 residues instead of 585, completely variant from residue 575 to the C-terminal end: TCCCKSSCLRLITSHLKASQPTMRIRE. The C-terminal sequence of the variant is reported in Fig. 4, together with the known nucleotide sequence of the corresponding region of the structural gene of HSA [22], starting from residue 572, which corresponds to the first codon of exon 14. This exon, which is 68 bp long and normally only partially translated, specifies the last 14 residues of HSA and contains the terminator codon (TAA) (Fig. 4, upper part). The amino-acid sequence we have found for albumin Kénitra can be accounted for by a single-base insertion: an adenine at nucleotide position 15 970, which leads to a frameshift with the subsequent translation of all of exon 14 and until the first termination codon of exon 15 (163 bp), which normally does not code for the protein but is conserved in the mature mRNA [22]. The seven C-terminal residues of the variant correspond to the first seven codons of this exon, while the eighth and the ninth codons specify for arginine and lysine, respectively, and the tenth is a stop codon (TGA) (Fig. 4, lower part). We did not find evidence for the presence in the circulating variant of the C-terminal dipeptide Arg-Lys. These two amino acids have probably been removed by one of the basic carboxypeptidases, carboxypeptidase B or carboxypeptidase N, which are active in the plasma compartment [27]. This phenomenon has previously been observed in another C-terminally modified but shortened mutant of HSA, albumin Venezia, which has the same seven C-terminal residues as albumin Kénitra. In that case, there was a 25-bp deletion resulting in the skipping of exon 14 and in a continued translation of exon 15 until the terminator codon was encountered [14,16]. However, in albumin Venezia the digestion was only partial, as about 20% of the circulating variant retained the arginine residue at its C-terminal end.

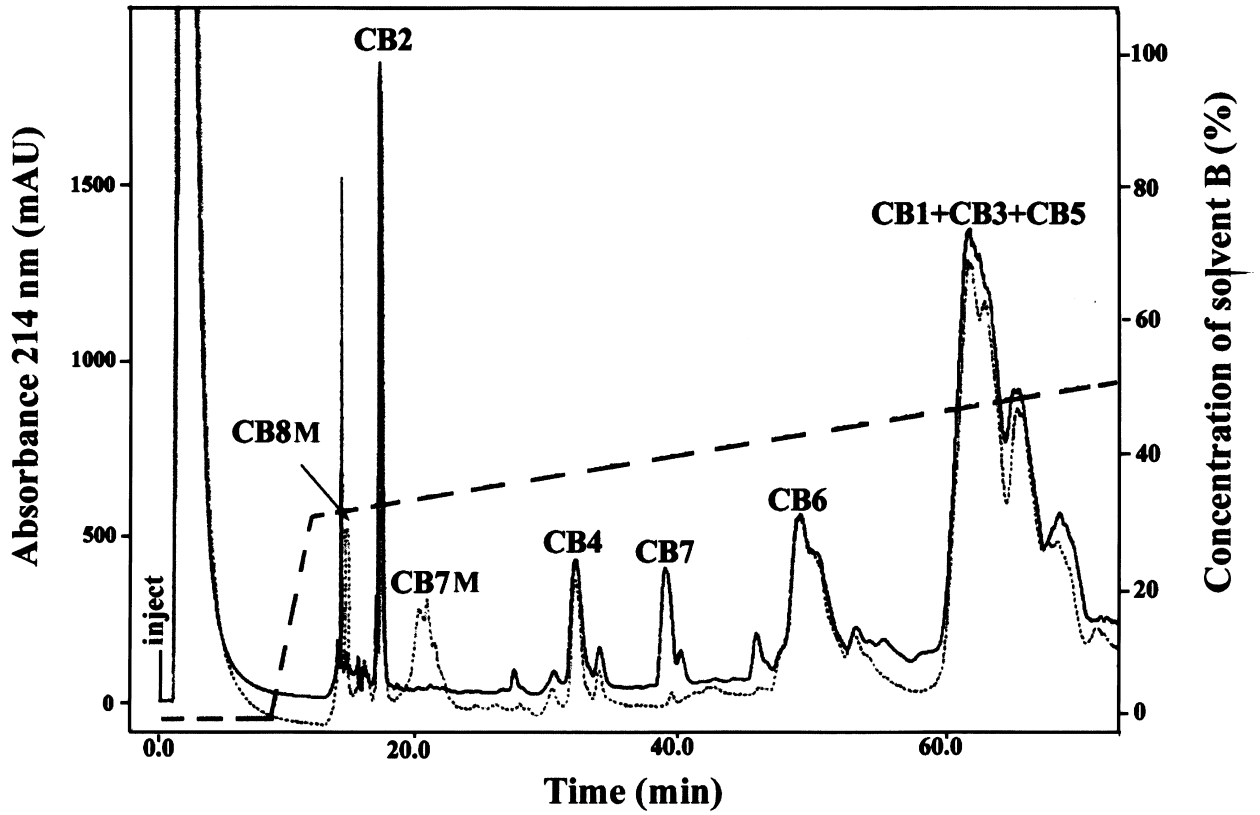

Fig. 3. RP-HPLC elution profile on a Vydac C18 column of the CNBr fragments from carboxymethylated Kénitra (dotted line) and normal (full line) albumins. The lyophilized CNBr digests were dissolved in 50 μL of 70% formic acid + 150 μL of 0.1% aqueous trifluoroacetic acid (solvent A), and 100 μL (corresponding to about 10 nmol) were injected on the Vydac C18 column equilibrated with solvent A. Peptides were eluted at a flow rate of 1 mL·min^{-1} using the gradient indicated by the dashed line. Solvent B was 0.1% trifluoroacetic acid in acetonitrile/2-propanol 2 : 1, v/v. CB7M and CB8M denote the two variant peptides resulting from the mutation of albumin Kénitra.

DNA structural studies

In order to check for the adenine insertion at nucleotide position 15 970, genomic DNA was extracted from whole blood and a 339-bp fragment, containing exon 14 of the albumin gene, was amplified by PCR using primers A27A and A28A [23] and submitted to SSCP and heteroduplex analysis. The PCR product was also digested with *Alu*I, which gave rise to three fragments of 94, 104, 141 bp, respectively. Figure 5 shows the electrophoretic patterns obtained for the native and denatured samples from the patient and two normal controls. A clear heteroduplex band

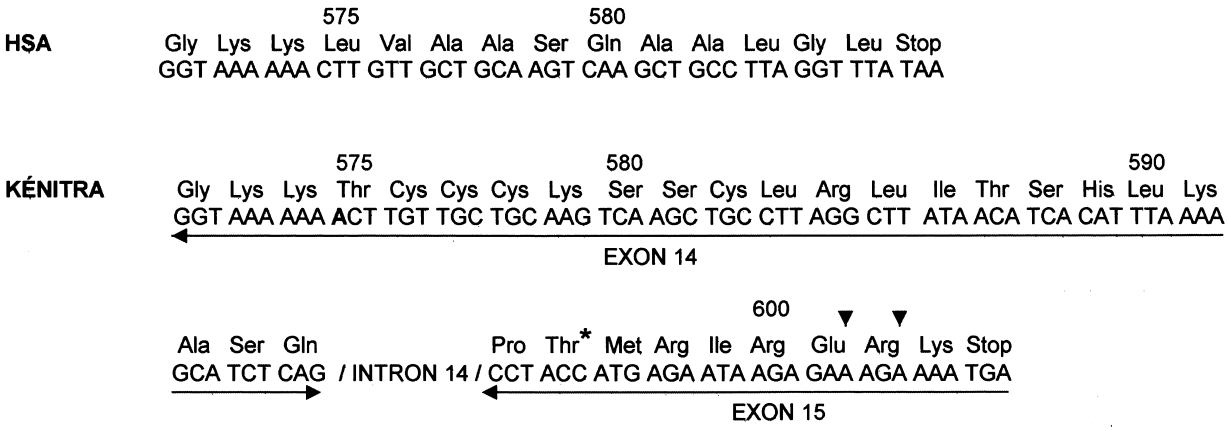

Fig. 4. Nucleotide and amino-acid sequences of the C–terminal regions of normal albumin and albumin Kénitra. The additional adenine at nucleotide position 15 970 in the genomic sequence of albumin Kénitra is indicated in bold. The O-glycosylated Thr596 is marked by an asterisk. The two arrowheads mark the cleavage sites for the two basic C-terminal residues, which are absent in the circulating protein.

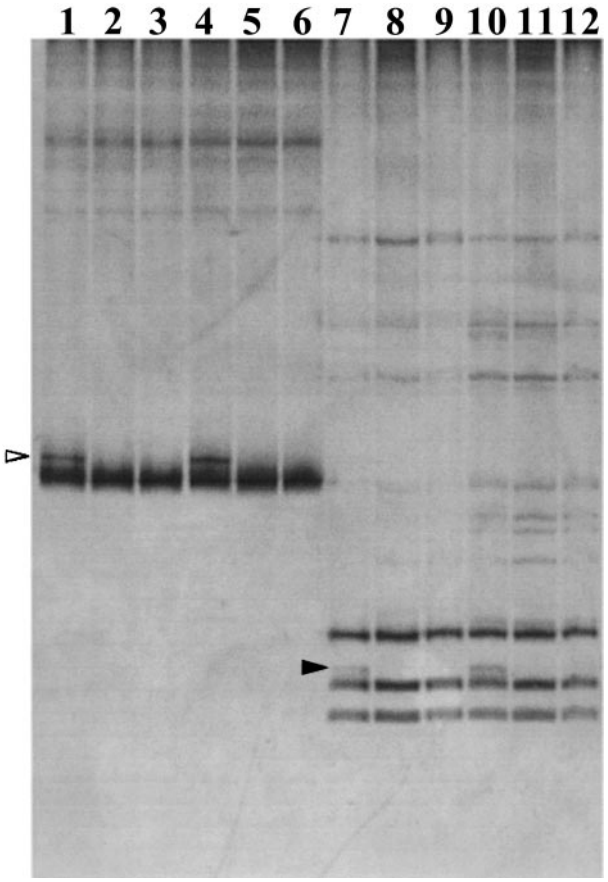

Fig. 5. Heteroduplex and SSCP analysis of exon 14. DNA from Kénitra and two controls were amplified with primers A27A and A28A and the PCR fragments were electrophoresed onto a nondenaturing polyacrylamide gel (lanes 1, 2, and 3, respectively). The same samples were denatured and cooled before loading (lanes 4, 5, and 6). The same PCR fragments were analyzed after digestion with *Alu*I (lanes 7, 8, and 9) and after digestion with *Alu*I with prior denaturation (lanes 10, 11, and 12). The open arrowhead denotes the abnormal heteroduplex in lanes 1 and 4. The closed arrowhead indicates the single abnormal heteroduplex after digestion with *Alu*I in lanes 7 and 10.

is present in the amplified abnormal exon 14 (lanes 1 and 4), as well as in the 104-bp *Alu*I fragment (lanes 7 and 10). This result confirms that the mutation is located within the region 15 886–15 989 of exon 14. Then, the variant and control DNA were amplified using primer A32M and the modified primer A31M, which introduces at the point of the presumed insertion a restriction site for *Tru*9I in the mutated allele. The PCR products, encompassing nucleotides 15 947–16 038 of the HSA gene, were digested with the enzyme and the restriction fragments analysed by nondenaturing 15% PAGE. The wild-type gave rise to two fragments of 70 and 22 bp because of the presence of a restriction site for *Tru*9I at nucleotide position 16 014, while in the mutant the 71-bp fragment was further cleaved into two fragments of 48 and 23 bp, thus confirming the adenine insertion in nucleotide position 15 970 (Fig. 6).

Mass spectrometric analysis of the intact variant

The status of the four additional cysteine residues present in the C-terminal region of albumin Kénitra was investigated by mass spectrometric analysis. The variant protein was either alkylated with iodoacetamide without any reduction step or submitted to dithiothreitol treatment followed by carboxyamidomethylation under denaturing conditions. Both the alkylated and the reduced and carboxyamidomethylated samples were desalted by RP-HPLC on a Vydac C4 column and directly analysed by ESMS. The results are summarized in Table 1. When the reduction step was omitted, the ES spectra showed the presence of two main components whose molecular masses were measured as 69199.2 ± 4.7 Da and 68538.9 ± 5.6 Da, respectively. The lower mass value is in good agreement with that expected on the basis of the new 601 amino-acid sequence of the variant albumin with a single alkylated cysteine residue (68544.1 Da), confirming the removal of the C-terminal Arg-Lys dipeptide. This result demonstrates that the four extra cysteine residues present in the C-terminal region of the variant are all involved in S-S bridges. The obligatory pairings of the two new disulfide bonds are Cys576–Cys578 and Cys577–582, because three cysteine residues are contiguous in the primary structure. The only cysteine alkylated by iodoacetamide should then be Cys34, the single unpaired cysteine occurring in normal HSA. The second component exhibited a mass increase of about 660–670 Da as compared to the variant albumin, an increment which we could not account for at this stage of the investigation.

Two molecular species were also detected in the ES spectra of the reduced and alkylated abnormal albumin, showing molecular masses of 71416.5 ± 4.3 Da and 70745.9 ± 4.8 Da, respectively. The lower mass value corresponds to the modified protein in which all the 39 cysteine residues have been alkylated by iodoacetamide, confirming the oxidation state of the new four cysteines. The second species showed again a mass increase of 660–670 Da, as previously determined for the alkylated variant. Because the mass difference between the two albumins is conserved even after reduction and alkylation under denaturing conditions, these results suggest that the variant sample consists of two forms which differ by a still unknown covalent modification. The ratio between these two forms can be estimated as ≈ 1 : 1.

Table 1. Electrospray mass analysis of the intact variant. CAM, carboxyamidomethylated; REDCAM, reduced and carboxyamidomethylated.

	Molecular mass (Da)	
	Measured	Theoretical
CAM albumin Kènitra	68 538.9 ± 5.6	68 544.1
REDCAM albumin Kènitra	69 199.2 ± 4.7 70 745.9 ± 4.8 71 416.5 ± 4.3	70 752.6

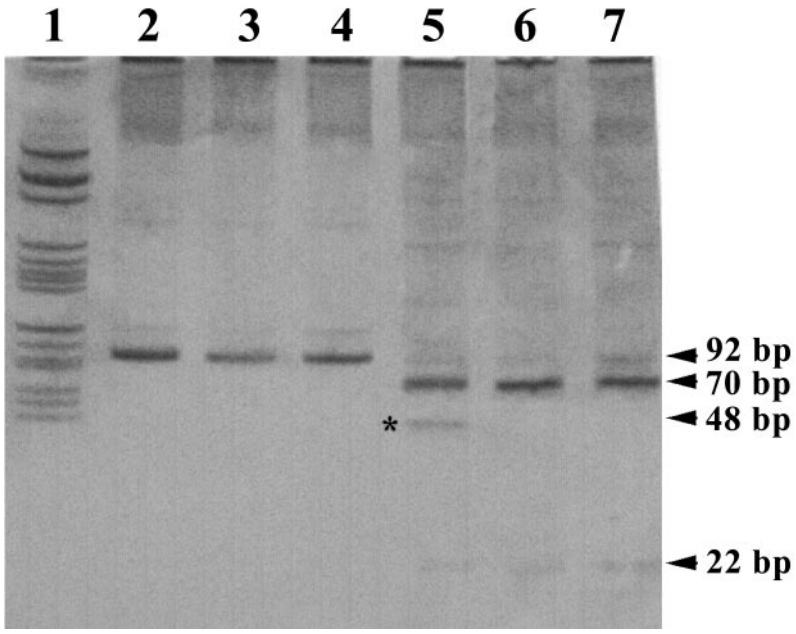

Fig. 6. Artificial insertion restriction site analysis of exon 14 in polyacrylamide gel. Exon 14 of the albumin gene was amplified by PCR using a mismatch primer 5′-GTCTTTGTGTTCAGGGTAAATTA-3′ (A31M), which introduces an additional restriction site for *Tru*9I in the mutated allele. A31M was used with primer 5′-AGCATCTCAGGTAACTATAT-3′ (A32M) to amplify a 93-bp PCR fragment, which was then digested with *Tru*9I. PCR products from Kénitra and two controls were electrophoresed onto a nondenaturing polyacrylamide gel before (lanes 2, 3, and 4, respectively) and after digestion with *Tru*9I (lanes 5, 6, and 7, respectively). Lane 1: molecular mass markers. The asterisk marks the 48-bp fragment produced by the cleavage at the artificial restriction site.

Mass spectrometric characterization of the CNBr peptides from albumin Kénitra

The exact molecular masses of the individual CNBr peptides, fractionated by RP-HPLC (Fig. 3), were determined by direct ESMS analysis. All but one of the peptides analysed by ESMS displayed the expected molecular mass, including the newly detected CB8M. However, mass spectral investigation of the broad chromatographic peak corresponding to the CB7M fragment showed the presence of two components, whose molecular masses were measured as 5779.6 ± 0.4 Da and 6438.8 ± 0.8 Da. The lower mass value perfectly matched that expected on the basis of the new C-terminal sequence of albumin Kénitra. The second species showed a molecular mass about 659.2 Da higher than that of the variant CB7M. This mass difference corresponds to the entire mass difference observed in the analysis of the intact variant, thus confirming that the covalent modification is confined within this fragment.

The variant CB7M was then deglycosylated by reductive elimination following the procedure reported by Carlson [25]. The putative oligosaccharides released from the polypeptide chain were permethylated and submitted to MALDI MS analysis. The resulting MALDI spectra showed the occurrence of a mass signal at *m*/*z* 895, corresponding to a monosialylated HexHexNAc structure. This finding demonstrates that albumin Kénitra exists as two molecular species differing in the presence of O-glycosylation at level of CB7M.

Localization of the O-glycosylation site

As the mutated C-terminal region of the variant contains four serine and three threonine residues, in order to localize the O-glycosylation site the variant CB7M was cleaved with trypsin and the fragments were resolved by RP-HPLC (data not shown). All the expected peptides were identified and their amino-acid sequences were in total agreement with the results obtained on the whole CB7M. The only exception was the C-terminal fragment (residues 592–597), that gave rise to four different peaks, due to partial glycosylation and to the homoserine/homoserine lactone equilibrium. Amino-acid sequence analysis of these peaks showed that in two of the fractions the threonine residue at position 596 was clearly detectable, while a blank was found at the same position in the other two species, thus indicating that the oligosaccharide chain is O-linked to Thr596.

The specificities of the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferases which link the carbohydrate GalNAc to the side chain of certain serine and threonine residues in mucin type glycoproteins are not yet defined, but there is a widely used program that predicts the O-glycosylation sites in mammalian proteins on the basis of sequence context, secondary structure and surface accessibility [28]. Use of this program on the C-terminal region of albumin Kénitra shows that only Thr596, among the possible sites, has a glycosylation potential close to, although lower than, the threshold value. This finding is in full accordance with the results of the amino-acid sequencing of the variant CB7M tryptic fragments. Thr596 is adjacent to a Pro residue, which has been reported to favor oligosaccharide addition to threonine residues [29]. The partial glycosylation of this residue indicates that it is easily accessible to the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase and therefore reflects the high exposure to the solvent of the C-terminal region of the variant.

CONCLUSIONS

Albumin Kénitra is a new type of genetic variant of HSA that arises from the insertion of an adenine residue at nucleotide position 15 970 in the structural gene. This mutation causes a shift of the reading frame and the translation of all of exon 14 and part of exon 15, which normally does not code for protein [22], until the first

termination codon. This is in full accord with the primary structure of the protein, which has an elongated polypeptide chain, 601 residues instead of 585, and is completely different from the normal molecule starting from residue 575. In addition the new sequence contains two extra disulfide bridges and is partially O-glycosylated at Thr596, the glycosylated form representing about 50% of the variant. These features make albumin Kénitra unique among the genetic variants of HSA that have so far been characterized. The low plasma concentration of the variant, about 15% of the total albumin content, indicates that the extensive modifications decrease its *in vivo* stability. In the known three-dimensional structure of HSA [30], residues 575–583 correspond to the C-terminal part of helix h10(III). Because of the strain imposed by the two new disulfide bonds, a premature breakdown of this helix is likely to occur in proximity of position 570 of the variant. This is in agreement with secondary structure predictions that suggest the presence of an additional helix encompassing residues 580–590 [31]. Because this new elongated variant as well as all the known truncated alloalbumins are present in serum at levels ranging from only 2–30%, any major change from the native structure of the C-terminal region of HSA appears to be crucial for the stability of the protein.

ACKNOWLEDGEMENTS

The authors wish to thank Dr Patrizia Arcidiaco (Centro Grandi Strumenti, Università di Pavia) for sequence analysis, Marco Bellaviti for computer elaboration of the figures and Antonio Mortara for excellent technical assistance. This work was supported by PRIN (Progetto di ricerca di interesse nazionale) grant 'Structural studies on hydrophobic molecule-binding proteins' from the Ministero della Università e della Ricerca Scientifica e Tecnologica (Rome, Italy) and by CNR grant to P. P.

REFERENCES

1. Peters, T. Jr (1996) All about albumin. *Biochemistry, Genetics and Medical Applications.* Academic Press, San Diego, CA.
2. Petersen, C.E., Scottolini, A.G., Cody, L.R., Mandel, M., Reimer, R. & Bhagavan, N.V. (1994) A point mutation in the human serum albumin gene results in familial dysalbuminaemic hyperthyroxinaemia. *J. Med. Genet.* **31**, 355–359.
3. Sunthornthepvarakul, T., Angkeow, P., Weiss, R.E., Hayashi, Y. & Refetoff, S. (1994) An identical missense mutation in the albumin gene results in familial dysalbuminemic hyperthyroxinemia in 8 unrelated families. *Biochem. Biophys. Res. Commun.* **202**, 781–787.
4. Wada, N., Chiba, H., Shimizu, C., Kijima, H., Kubo, M. & Koike, T. (1997) A novel missense mutation in codon 218 of the albumin gene in a distinct phenotype of familial dysalbuminemic hyperthyroxinemia in a Japanese kindred. *J. Clin. Endocrinol. Metab.* **82**, 3246–3250.
5. Sunthornthepvarakul, T., Likitmaskul, S., Ngowngarmratana, S., Angsusingha, K., Kitvitayasak, S., Scherberg, N.H. & Refetoff, S. (1998) Familial dysalbuminemic hypertriiodothyroninemia: a new, dominantly inherited albumin defect. *J. Clin. Endocrinol. Metab.* **83**, 1448–1454.
6. Minchiotti, L., Kragh-Hansen, U., Nielsen, H., Hardy, E., Mercier, A.Y. & Galliano, M. (1999) Structural characterization, stability and fatty acid binding properties of two French genetic variants of human serum albumin. *Biochim. Biophys. Acta* **1431**, 223–231.
7. Brennan, S.O., Fellowes, A.P. & George, P.M. (1999) Albumin Banks Peninsula: a new termination variant characterised by electrospray mass spectrometry. *Biochim. Biophys. Acta* **1433**, 321–326.
8. Kragh-Hansen, U., Pedersen, A.O., Galliano, M., Minchiotti, L., Brennan, S.O., Tàrnoky, A.L., Franco, M.H.L.P. & Salzano, F.M. (1996) High-affinity binding of laurate to naturally occurring mutants of human serum albumin and proalbumin. *Biochem. J.* **320**, 911–916.
9. Nielsen, H., Kragh-Hansen, U., Minchiotti, L., Galliano, M., Brennan, S.O., Tàrnoky, A.L., Franco, M.H., Salzano, F.M. & Sugita, O. (1997) Effect of genetic variation on the fatty acid-binding properties of human serum albumin and proalbumin. *Biochim. Biophys. Acta* **1342**, 191–204.
10. Minchiotti, L., Watkins, S., Madison, J., Putnam, F.W., Kragh-Hansen, U., Amoresano, A., Pucci, P., Cesati, R. & Galliano, M. (1997) Structural characterization of four genetic variants of human serum albumin associated with alloalbuminemia in Italy. *Eur. J. Biochem.* **247**, 476–482.
11. Brennan, S.O., Arai, K., Madison, J., Laurell, C.B., Galliano, M., Watkins, S., Peach, R., Myles, T., George, P. & Putnam, F.W. (1990) Hypermutability of CpG dinucleotides in the propeptide-encoding sequence of the human albumin gene. *Proc. Natl Acad. Sci. USA* **87**, 3909–3913.
12. Galliano, M., Watkins, S., Madison, J., Putnam, F.W., Kragh-Hansen, U., Cesati, R. & Minchiotti, L. (1998) Structural characterization of three genetic variants of human serum albumin modified in subdomains IIB and IIIA. *Eur. J. Biochem.* **251**, 329–334.
13. Galliano, M., Minchiotti, L., Iadarola, P., Zapponi, M.C., Ferri, G. & Castellani, A.A. (1986) Structural characterization of a chain termination mutant of human serum albumin. *J. Biol. Chem.* **261**, 4283–4287.
14. Watkins, S., Madison, J., Davis, E., Sakamoto, Y., Galliano, M., Minchiotti, L. & Putnam, F.W. (1991) A donor splice mutation and a single-base deletion produce two carboxyl-terminal variants of human serum albumin. *Proc. Natl Acad. Sci. USA* **88**, 5959–5963.
15. Madison, J., Galliano, M., Watkins, S., Minchiotti, L., Porta, F., Rossi, A. & Putnam, F.W. (1994) Genetic variants of human serum albumin in Italy: point mutants and a carboxyl-terminal variant. *Proc. Natl Acad. Sci. USA* **91**, 6476–6480.
16. Minchiotti, L., Galliano, M., Iadarola, P., Meloni, M.L., Ferri, G., Porta, F. & Castellani, A.A. (1989) The molecular defect in a COOH-terminal-modified and shortened mutant of human serum albumin. *J. Biol. Chem.* **264**, 3385–3389.
17. Peach, R.J., Fellowes, A.P., Brennan, S.O. & George, P.M. (1992) Albumin Rugby Park: a truncated albumin variant caused by a G → C splice-site mutation in intron 13. *Biochim. Biophys. Acta* **1180**, 107–110.
18. Sakamoto, Y., Kitamura, K., Madison, J., Watkins, S., Laurell, C.B., Nomura, M., Higashiyama, T. & Putnam, F.W. (1995) Structural study of the glycosylated and unglycosylated forms of a genetic variant of human serum albumin (63Asp→Asn). *Biochim. Biophys. Acta* **1252**, 209–216.
19. Brennan, S.O., Myles, T., Peach, R.J., Donaldson, D. & George, P.M. (1990) Albumin Redhill (-1Arg, 320Ala→Thr): a glycoprotein variant of human serum albumin whose precursor has an aberrant signal peptidase cleavage site. *Proc. Natl Acad. Sci. USA* **87**, 26–30.
20. Peach, R.J. & Brennan, S.O. (1991) Structural characterization of a glycoprotein variant of human serum albumin: albumin Casebrook (494 Asp→Asn). *Biochim. Biophys. Acta* **1097**, 49–54.
21. Granel, B., Serratrice, J., Galliano, M., Minchiotti, L., Disdier, P. & Weiller, P.J. (2000) Une nouvelle bisalbuminémie: la bis-albumine Kénitra. *Presse Med.* **29**, 1224.
22. Minghetti, P.P., Ruffner, D.E., Kuang, W.J., Dennison, O.E., Hawkins, J.W., Beattie, W.G. & Dugaiczyk, A. (1986) Molecular

structure of the human albumin gene is revealed by nucleotide sequence within q11–22 of chromosome 4. *J. Biol. Chem.* **261**, 6747–6757.
23. Watkins, S., Madison, J., Galliano, M., Minchiotti, L. & Putnam, F.W. (1994) A nucleotide insertion and frameshift cause analbuminemia in an Italian family. *Proc. Natl Acad. Sci. USA* **91**, 2275–2279.
24. Superti-Furga, A., Hastbacka, J., Wilcox, W.R., Cohn, D.H., van der Harten, H.J., Rossi, A., Blau, N., Rimoin, D.L., Steinmann, B., Lander, E.S. & Gitzelmann, R. (1996) Achondrogenesis type IB is caused by mutations in the diastrophic dysplasia sulphate transporter gene. *Nat. Genet.* **12**, 100–102.
25. Carlson, D.M. (1968) Structures and immunochemical properties of oligosaccharides isolated from pig submaxillary mucin. *J. Biol. Chem.* **243**, 616–626.
26. Dell, A. (1990) Preparation and desorption mass spectrometry of permethyl and peracetyl derivatives of oligosaccharides. *Methods Enzymol.* **193**, 647–660.
27. Sakharov, D.V., Plow, E.F. & Rijken, D.C. (1997) On the mechanism of the antifibrinolytic activity of plasma carboxypeptidase B. *J. Biol. Chem.* **272**, 14477–14482.
28. Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L. & Brunak, S. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.* **15**, 115–130.
29. Gooley, A.A., Classon, B.J., Marschalek, R. & Williams, K.L. (1991) Glycosylation sites identified by detection of glycosylated amino acids released from Edman degradation: the identification of Xaa-Pro-Xaa-Xaa as a motif for Thr-*O*-glycosylation. *Biochem. Biophys. Res. Commun.* **178**, 1194–1201.
30. Carter, D.C. & Ho, J.X. (1994) Structure of serum albumin. *Adv. Prot. Chem.* **45**, 153–203.
31. Geourjon, C. & Deleage, G. (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* **11**, 681–684.