Journal of Chromatography, 359 (1986) 181–191
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 777

APPLICATION OF AN AUTOMATED TANDEM HIGH-PERFORMANCE LIQUID CHROMATOGRAPHIC SYSTEM TO PEPTIDE MAPPING OF GENETIC VARIANTS OF HUMAN SERUM ALBUMIN

NOBUHIRO TAKAHASHI*, YOKO TAKAHASHI and NORIAKI ISHIOKA
Department of Biology, Indiana University, Bloomington, IN 47405 (U.S.A.)
BARUCH S. BLUMBERG
Fox Chase Cancer Center, Philadelphia, PA 19111 (U.S.A.)
and
FRANK W. PUTNAM
Department of Biology, Indiana University, Bloomington, IN 47405 (U.S.A.)

SUMMARY

An automated tandem high-performance liquid chromatographic system was developed for peptide mapping of very large proteins. The method was applied systematically to the peptide mapping analysis of four genetic variants of human serum albumin, and amino acid substitutions in three of them could be identified. In two variants the amino acid substitutions were identified in both homozygote and heterozygote specimens. By this peptide mapping method at least 80% of the total amino acid residues can be screened rapidly. The method is especially useful for establishing the molecular basis of genetic relationships among protein variants found in different populations.

INTRODUCTION

Serum albumin is the most abundant protein in human serum and consists of a single polypeptide chain with 585 amino acid residues. The predominant albumin in all populations is known as albumin A, but over twenty electrophoretic variants have been reported[1]. However, only three genetic variants have been characterized with respect to amino acid substitutions, *e.g.*, albumin Oliphant (Glu-570 → Lys-570)[2], albumin Mexico-2 (Asp-550 → Gly-550)[3] and albumin Mi/Fg (Lys-573 → Glu-573)[4]. Although a large amount of albumin is present in serum, usually only a small amount of serum can be collected during genetic studies of some populations. In addition, the large size of the protein has prevented identification of the site of substitution in many genetic variants. Hence, a systematic procedure is needed for the identification of the amino acid exchanges in the genetic variants of albumin and other large proteins.

Franklin *et al.*[3] used preparative polyacrylamide gel electrophoresis after cyanogen bromide (CNBr) digestion of the protein to identify the amino acid substitution

0021-9673/86/$03.50

in albumin Mexico-2. Recently, reversed-phase high-performance liquid chromatography (HPLC) has been used to separate all the CNBr fragments of human serum albumin by Iadarola *et al.*[5], who have identified the amino acid substitution in albumin Mi/Fg[4]. However, their methods seem most suitable for detecting substitutions within a small CNBr fragment.

In this paper we describe a procedure for systematic peptide mapping analysis by automated tandem HPLC and its application to genetic variants of human serum albumin. For this purpose, the alkylated protein is first digested with trypsin to produce peptides of appropriate size, and the peptide profile is determined by HPLC. This allows identification in one step of peptides that have an amino acid substitution comparable to that of albumin A. Four genetic variants have been analyzed systematically by the method, and the identification of the variant peptides in three was achieved by analysis of the tandem HPLC profile. Whereas previous reports of the site of substitution in albumin variants have been based only on amino acid composition[3] or partial sequence analysis[4], our HPLC procedure provided sufficient peptide to identify the site of substitution by amino acid analysis and to verify it by sequence analysis.

EXPERIMENTAL

Materials

Human sera from individuals with an albumin genetic variant were screened electrophoretically[1,6]. The variants studied were Mexico-2 (Me-2), Naskapi (Na), Mersin and Adana. Six sera from individuals with albumin genetic phenotypes, A/Me-2, Me-2/Me-2, A/Na, Na/Na, A/Mersin and A/Adana, and also a commercial albumin (Lot 102578, Calbiochem-Behring, La Jolla, CA, U.S.A.), were used for peptide mapping analysis. Although the commercial albumin was of unknown phenotype, it had been prepared from pooled plasma of an American population and is designated the standard albumin (albumin A). To purify the albumin, each human serum sample (0.5–1.0 ml) was first subjected to ion-exchange chromatography (TSK DEAE-5PW column, 75 × 7.5 mm I.D.). The albumin fraction was purified further by gel chromatography (TSK 3000 SW column, 600 × 7.5 mm I.D.). All albumins were at least 95% pure, as judged by SDS–polyacrylamide gel electrophoresis.

The sources of enzymes, reagents and columns were as follows: trypsin, treated with L-(tosylamido-2-phenyl)ethyl chloromethyl ketone (TPCK) (Worthington, Freehold, NJ, U.S.A.); methanesulfonic acid (Aldrich, Milwaukee, WI, U.S.A.); ammonium hydroxide, ultrapure, (J. T. Baker, Phillipsburg, NJ, U.S.A.); trifluoroacetic acid (TFA) (Pierce, Rockford, IL, U.S.A.); and acetonitrile (chromatography grade) (Burdick & Jackson, Muskegon, MI, U.S.A.). The TSK DEAE-2SW column (25 × 0.4 cm I.D.) and an Ultrasphere ODS column (15 × 0.4 cm I.D.) were obtained from Altex (Berkeley, CA, U.S.A.).

Tryptic digestion

All albumins were reduced and carboxymethylated and then digested with TPCK-treated trypsin (enzyme-to-substrate ratio = 1:50) in 0.1 *M* ammonium bicarbonate solution at 37°C for 15 h. The digest was lyophilized to remove the ammonium bicarbonate, and the lyophilizate was dissolved in 0.5 ml of 8 *M* urea and

diluted with 0.5 ml of 0.02 *M* Tris–acetic acid buffer (pH 8.0). The insoluble material was removed by centrifugation and the supernatant was used as a sample for tandem HPLC.

Tandem high-performance liquid chromatography

Tandem HPLC was performed according to the method of Takahashi *et al.*[7], except that 1.0 *M* ammonium hydroxide–methanesulfonic acid–0.02 *M* Tris–acetic acid (pH 8.0) was used as B2 buffer instead of 0.4 *M* ammonium hydroxide–methanesulfonic acid–0.02 *M* Tris–acetic acid (pH 8.0). The pumping ratio of B2 for the first chromatographic step was increased in a stepwise manner as follows: 0, 5, 10, 15, 20, 25, 30, 35, 40, 50 and 100% B2.

Identification and nomenclature of peptides

The peptides were identified by their amino acid composition or by amino acid sequence. Amino acid analysis was performed with a Beckman Model 121M amino acid analyzer and sequence determination with a Beckman Model 890C sequencer, as described[8]. Tryptic peptides of human albumin are given the prefix T and are numbered consecutively in their order in the sequence[9].

RESULTS AND DISCUSSION

Peptide mapping of a standard albumin

Although the commercial albumin is assumed to be phenotype A/A, which is predominant in all populations, this albumin was prepared from pooled serum and the phenotype is not certain. The pattern of peptide mapping by the tandem HPLC is shown three-dimensionally in Fig. 1. Human albumin has a potential yield of 80 tryptic peptides. About 100 peaks are obtained. The peptides identified are indicated by the numbers in Fig. 1. The number of peaks exceeds the expected 80 theoretical peptides, because some peptides are duplicated in the adjacent stepwise elution, and also because of irregular or incomplete cleavage. The sum of the amino acid residues identified was about 80% of the total amino acid residues of the protein.

Fig. 2 shows the distribution of charge and hydrophobicity of all of the tryptic peptides expected from a completely specific cleavage of albumin. The amino acid sequence reported by Lawn *et al.*[9] is used for this purpose. The charges of the theoretical peptides are plotted against $\ln(1 + H)$ in Fig. 2, where H indicates the hydrophobicity of the peptides, calculated by the method of Sasagawa *et al.*[10] using their list of non-weighted retention constants.

The charges of the peptides were calculated from the constituent amino acids of the peptide at pH 8.0 by the microcomputer method of Manabe[11]. The unidentified peptides include the free amino acids arginine and lysine, most of the di-, tri- and tetrapeptides and three very hydrophobic peptides T5, T44 and T67. The unidentified short peptides have a value of $\ln(1 + H)$ less than 3, so these peptides are not separated by reversed-phase chromatography, as was the case for the tryptic peptides of ceruloplasmin[7]. The three very hydrophobic peptides that were not identified may be present in the unanalyzed precipitate that was removed by centrifugation or may be bound strongly to the columns in the system. Many incomplete cleavages with trypsin were observed, because an acidic amino acid is adjacent to

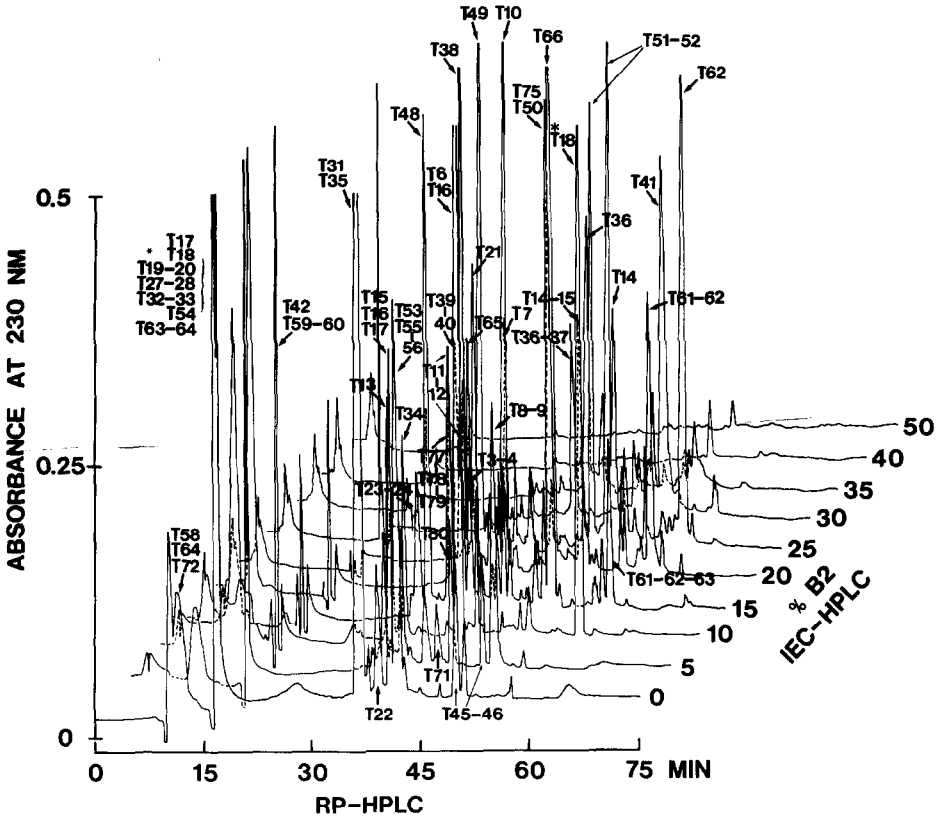

Fig. 1. Three-dimensional representation of the automated tandem HPLC of standard human serum albumin. Each horizontal profile represents one cycle in the chromatography; the first 15 min indicate stepwise elution from the ion-exchange column (IEC-HPLC), and the remaining 60 min show linear gradient elution from the reversed-phase column (RP-HPLC). The concentration of the buffer for the stepwise elutions of the ion-exchange column is shown at the right-hand side of each horizontal profile. The nomenclature for peptides, described under Experimental, is used to illustrate the elution positions of the tryptic peptides identified. Peptides identified as two peaks are shown by two fine arrows and peptides identified as a single peak by a thick arrow. The numbers connected by a dash (*e.g.*, T51–52) indicate that the peptides were produced by incomplete cleavage with trypsin, probably because of adjacent acidic amino acids. An asterisk shows that the peptide was obtained by chymotrypsin-like cleavage.

lysine or arginine at many positions in the sequence of albumin. Also, there are six instances in which cysteine is adjacent to a basic amino acid but is converted into carboxymethylcysteine during reduction and alkylation of the albumin.

Peptide mapping applied to albumin Mexico-2 with known amino acid substitution

Albumin Mexico-2 is known to have an amino acid substitution at position 550 in peptide T75 (residues 546–557) where aspartic acid is replaced with glycine[3]. Tandem HPLC was applied to isolate a peptide containing the amino acid substitution in albumin Mexico-2. With standard albumin, peptide T75 was eluted at 15% B2 from the ion-exchange column (Fig. 3A). Unexpectedly, in reversed-phase chromatography T75 was eluted at the same position as T50 (residues 373–389), even

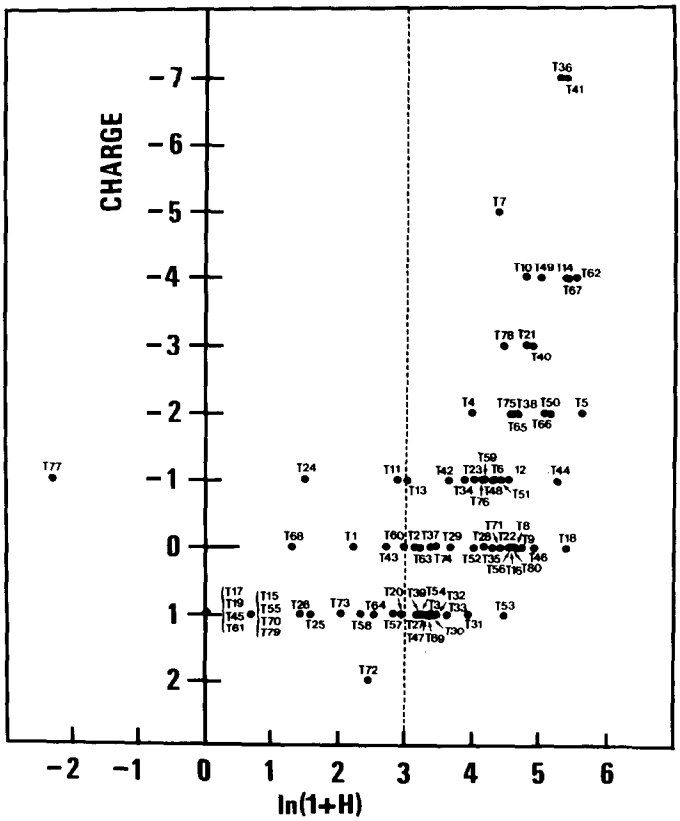

Fig. 2. Distribution between charge and ln(1 + H) of the tryptic peptides of human serum albumin. All theoretical tryptic peptides are designated by the nomenclature described under Experimental. A dotted line is drawn at ln(1 + H) = 3.

though the two peptides differed in hydrophobicity, as predicted by the method of Sasagawa *et al.*[10] (see Fig. 2). The chromatogram of the homozygote specimen of Mexico-2 is shown in Fig. 3C. Because the elution position of peptide T75 was identified by the analysis of the standard albumin, the corresponding peak of albumin Mexico-2 was picked out to analyze the substitution in peptide T75. Although a difference in the elution position was not observed for peptides T75 and T50, the amino acid composition indicated that Asp was substituted by Gly in either peptide T50 or T75. The two peptides, therefore, were separated by reversed-phase HPLC with 0.1% heptafluorobutyric acid. Sequence analysis of T75 confirmed that Asp-550 was replaced with Gly-550 in albumin Mexico-2. The yield of the peptide was 70%. However, because there was disagreement in the amino acid sequences of human serum albumin that have been reported[9], it was necessary to confirm the presence of aspartic acid at position 550 in albumin A and the other variants. Therefore, the experiment was repeated for a heterozygote specimen (A/Me) of Mexico-2 (Fig. 3B), and the presence of both peptides in the same peak, one with and the other without the substitution, was confirmed by amino acid analysis. The yields of the one with and the other without the substitution were 31% and 44%, respectively.

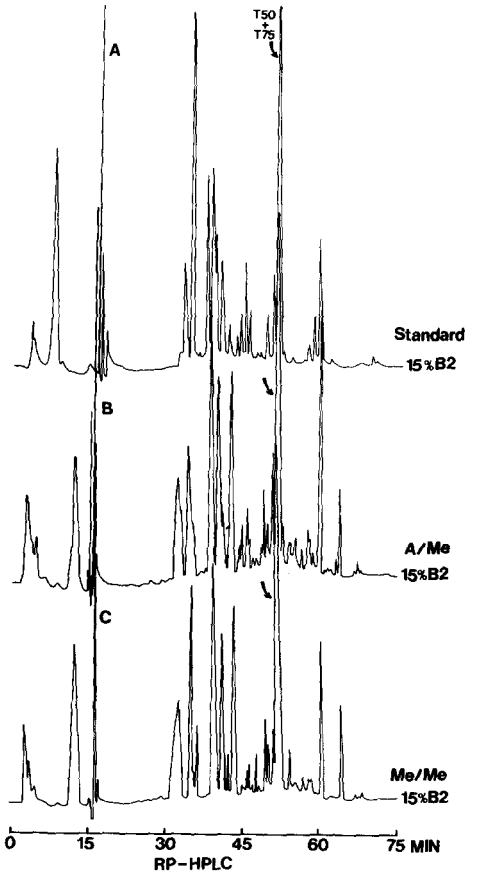

Fig. 3. Partial comparison of the peptide maps among (A) standard, (B) Mexico-2 heterozygote (A/Me) and (C) Mexico-2 homozygote (Me/Me) albumins. The peaks containing T50 and T75 are indicated by arrows. The concentration (% B2) for the stepwise elution of the ion-exchange column is shown at the right-hand side of each horizontal profile.

The analyses of the amino acid composition or the sequence of T75 from standard albumin and all the other variants also indicated that position 550 was occupied only by aspartic acid in those albumins. Thus, once the elution position of the peptide with the substitution is identified in the chromatogram, it is relatively easy to screen for a known amino acid substitution in many albumin samples by use of tandem HPLC.

Peptide mapping analysis applied to albumin Naskapi with unknown amino acid substitution

Of about two dozen electrophoretic variants of human serum albumin, albumin Naskapi, which is found mainly in North American Indians, is one of the few that occur in polymorphic frequency[12]. Although the amino acid substitution site in the variant was assumed to be located between residues 330 and 446 in the protein sequence by the analysis of the CNBr fragments[6], the site had not been determined.

Therefore, tandem HPLC was applied to peptide mapping analysis of albumin Naskapi in order to test whether a peptide with an amino acid substitution can be detected by the comparison of peptide maps of standard albumin and Naskapi.

When the profiles of the tandem HPLC of the standard albumin and the homozygote of Naskapi were compared (Figs. 1A and 4A), five major peaks, corresponding T10, T11–12, T48, T49 and T50 + T75, disappeared from the chromatogram of Naskapi homozygote, and three new peaks appeared in the elution pattern of the ion-exchange column at 30%, 35% and 40% B2. These peaks were analyzed by both amino acid composition and sequence analysis.

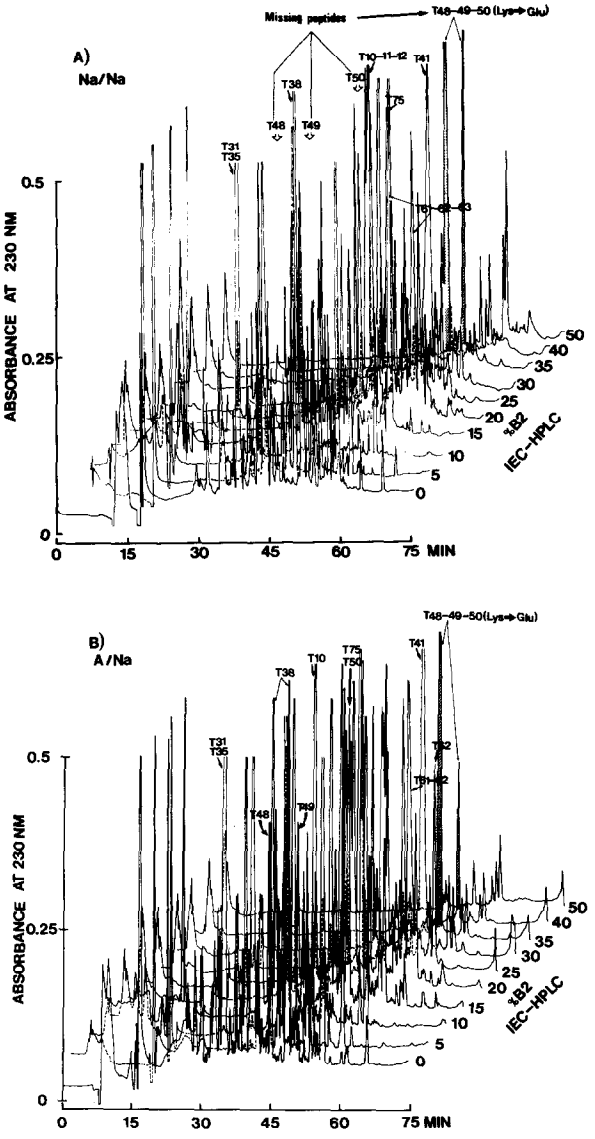

Fig. 4.

(Continued on p. 188)

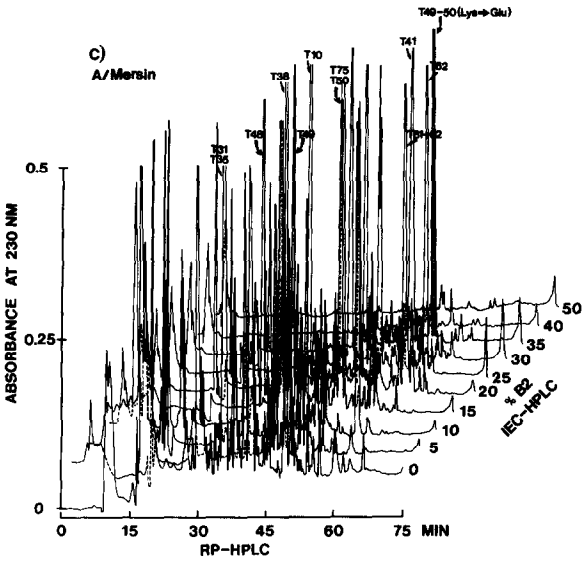

Fig. 4. Three-dimensional representation of the peptide maps of (A) Naskapi homozygote (Na/Na), (B) Naskapi heterozygote (A/Na) and (c) Mersin heterozygote (A/Mersin) albumins. Peptides from the region from residues 352 to 389 in the sequence of human serum albumin are indicated, and the peak containing the peptide with the substitution (Lys → Glu) is shaded. Some of the peptides are indicated as internal markers to facilitate comparison of the chromatograms. In A the missing peptides are indicated by open arrows.

The new peak that was eluted at 30% B2 contained peptide T10–11–12, which was produced by incomplete cleavage between T10 and T11–12 and which did not have any amino acid substitution in the peptide sequence. However, a peptide that was eluted at 35% B2 (T48–49–50, the shaded peak in Fig. 4A) was found to cover the region from residues 352 to 389 in the protein sequence and to contain one amino acid substitution at residue 372 where lysine was replaced with glutamic acid (Fig. 5). This substitution explains why peptides T49 and T50 were not obtained with trypsin, and therefore they disappeared from the chromatogram of the Naskapi homozygote. The lack of cleavage between T48 and T49 is probably caused by the acidic environment around Lys-359 due to the presence of the adjacent glutamic acid

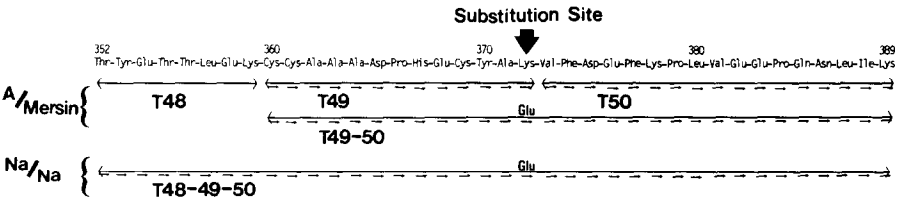

Fig. 5. Amino acid sequence analysis of the tryptic peptides with the substitution obtained from Naskapi homozygote (lower; Na/Na) and Mersin heterozygote (upper; A/Mersin) albumins. A single peptide (T48–49–50) with Lys → Glu interchange was obtained from the Naskapi homozygote, whereas both peptides, one (T49–50) with and the other (T49) without substitution, were obtained from the Mersin heterozygote. Amino acid residues identified by automated Edman degradation are indicated by small arrows.

at 358 and two carboxymethylcysteines at 360 and 361. The peak that was eluted at 40% B2 (second shaded peak in Fig. 4A) had the same retention time and amino acid composition as peptide T48–49–50. The combined yield of the peptide T48–49–50 was 75%.

The corresponding tryptic peptides of the Naskapi heterozygote were also analyzed by tandem HPLC in order to obtain both peptides, one with and the other without the substitution (Fig. 4B). Peptides T48, T49, T50 and T48–49–50 were easily picked out by comparison of the chromatograms of the standard albumin and the Naskapi homozygote and heterozygote. The amino acid and sequence analyses showed that T49 (yield 10%) contained a lysine residue at residue 372, that the amino acid sequence of T50 (yield 20%) started from Val-373 and that in peptide T48–49–50 (yield 25%) lysine was replaced with glutamic acid at position 372. Hence it was proved that both lysine and glutamic acid were present at position 372 in the heterozygote specimen of albumin Naskapi.

In the Naskapi homozygote, T75 was found at an elution position (Fig. 4A, 25% B2) that was different from the position where both T75 and T50 were eluted in the case of the standard, Mexico-2 homozygote and heterozygote albumins (Fig. 3). Because the Naskapi homozygote albumin does not yield a T50 peptide due to the Lys-372 → Glu-372 substitution, the coelution of T50 and T75 in the case of the other albumins is attributable to interaction of the two peptides.

Peptide mapping applied to albumin Mersin

Because of their similar electrophoretic behavior, albumin Mersin, found in the Eti Turks of southern Turkey, is assumed to be identical with albumin Naskapi[6]. Therefore, albumin Mersin was also analyzed by tandem HPLC (Fig. 4C). Only the heterozygote specimen of the variant was available for our study, because the homozygote has not yet been reported for this genetic variant. As shown in Fig. 4C, peptide T49–50 was obtained from the ion-exchange column at 35% B2, and in RP-HPLC T49–50 was eluted 40 s earlier than T48–49–50 of Naskapi. However, the sequence analysis showed that Lys-372 also was replaced with Glu-372 in peptide T49–50, as for Naskapi (Fig. 5). The yield of the peptide T49–50 was 25%. The cleavage between T48 and T49 with trypsin did occur with albumin Mersin. Peptides T48, T49 and T50 were also recovered from the Mersin heterozygote (Fig. 4C). Sequence analysis verified that peptide T49 had Lys at the carboxy-terminal position and that the sequence of peptide T50 started from Val-373 (Fig. 5). The yields of T49 and T50 were 22% and 30%, respectively.

Non-random location of substitutions

The three amino acid substitutions reported previously for Oliphant, Mexico-2 and Mi/Fg albumin variants are all located in the carboxy-terminal region of albumin[2–4]. This may be fortuitous, as the carboxy-terminal region is much easier to analyze than the middle of the molecule, because a small fragment can be obtained easily by CNBr cleavage of the protein, as with the Mexico-2 and Mi/Fg variants. However, the automated HPLC described here can be used to identify amino acid substitutions not only in the carboxyl-terminal region (Mexico-2) but also in the middle of the molecule (Naskapi and Mersin). Our continuing study of other variant albumins may show a more random distribution of substitutions.

Limitations of the method

Although the peptide mapping method was applied to another variant, Adana, we have not yet succeeded in the identification of the amino acid substitution in this variant. The substitution has not been found in peptides representing about 80% of the total amino acid sequence of albumin Adana. Therefore, it is probably present in the small or very hydrophobic peptides that are not identified easily by the method. In this instance, those peptides will have to be separated by the other procedures. Because about 80% of the total amino acid residues of the albumin sequence can be screened easily by our method, there is a high probability of finding the substitution site without resorting to additional procedures. However, because most substitutions found in albumin variants seem to involve Glu → Lys or Lys → Glu interchanges,the chance of substitution may not be equal throughout the whole protein sequence. We are now applying the method to the other albumin variants to clarify this point.

The main problem encountered during this study was a small degree of irregular or incomplete tryptic cleavage of albumin samples. This occurs most often when an acidic amino acid is adjacent to lysine or arginine, and this is the case at many positions in the sequence of albumin, which has a high content of glutamic and aspartic acids. Therefore, in studies of a variant it was sometimes necessary to analyze several peptide peaks that did not appear in the HPLC profile of standard albumin, even though the substitution turned out to be present in only one peptide. Nonetheless, the method was successfully applied to the identification of the amino acid substitutions in three out of four albumin genetic variants.

Impact on the amino acid sequence analysis of human serum albumin by nucleic acid sequencing

There are discrepancies of the protein sequences[13,14] compared with the nucleic acid sequences and also between the two nucleic acid sequences[9,15]. Even the number of differences in the recent nucleic acid sequences[16] is greater than the single substitutions found so far in the genetic variants. Obviously, there is a serious problem in identifying substitutions in genetic variants if there is no standard albumin sequence. This problem should be resolved by determining the sequence by methods of both protein sequencing and gene sequencing for the albumin of the same individual whose genetic phenotype is known.

Application to population migration studies by using tandem HPLC

Once the elution position of the peptide with the substitution has been identified in the chromatogram, it is easy to screen for a known amino acid substitution by use of tandem HPLC. Indeed, in two variants, Mexico-2 and Naskapi, the amino acid substitutions were easily identified in both homozygote and heterozygote specimens. This approach can be extended to many albumin variants with the potential substitution. Therefore, the method is very useful for establishing the molecular basis of genetic relationships among protein variants found in different populations. Because a possible molecular relationship between albumin Naskapi found mainly in North American Indians and Mersin found in the Eti Turks of southern Turkey was suggested by the analysis described, it would be especially interesting to screen the amino acid substitution in many populations along the possible migration pathway from Turkey to North America from the point of view of population migration studies.

ACKNOWLEDGEMENTS

We thank K. Huss, J. Madison and S. A. Dorwin for technical assistance. This work was supported by a grant from the NIH (AM19221).

REFERENCES

1 L. M. Schell and B. S. Blumberg, in V. M. Rosenoer, M. Oratz and M. A. Rothschild (Editors), *Albumin Structure, Function and Uses*, Pergamon Press, Oxford, 1977, p. 113.
2 W. P. Winter, L. R. Weitkamp and D. L. Rucknagel, *Biochemistry*, 11 (1972) 889.
3 S. G. Franklin, S. I. Wolf, A. Zweidler and B. S. Blumberg, *Proc. Natl. Acad. Sci. U.S.A.*, 77 (1980) 2505.
4 P. Iadarola, L. Minchiotti and M. Galliano, *FEBS Lett.*, 180 (1985) 85.
5 P. Iadarola, G. Ferri, M. Galliano, L. Minchiotti and M. C. Zapponi, *J. Chromatogr.*, 298 (1984) 336.
6 S. G. Franklin, S. I. Wolf, Y. Ozdemir, G. T. Yuregir, T. Isbir and B. S. Blumberg, *Proc. Natl. Acad. Sci. U.S.A.*, 77 (1980) 5480.
7 N. Takahashi, N. Ishioka, Y. Takahashi and F. W. Putnam, *J. Chromatogr.*, 326 (1985) 407.
8 N. Takahashi, Y. Takahashi and F. W. Putnam, *Proc. Natl. Acad. Sci. U.S.A.*, 82 (1985) 1906.
9 R. M. Lawn, J. Adelman, S. C. Bock, A. E. Franke, C. M. Houck, R. C. Najarian, P. H. Seeburg and K. L. Wion, *Nucleic Acid Res.*, 9 (1981) 6103.
10 T. Sasagawa, T. Okuyama and D. C. Teller, *J. Chromatogr.*, 240 (1982) 329.
11 T. Manabe, *Kagaku no Ryoiki*, 36 (1982) 470.
12 F. E. Johnston, B. S. Blumberg, S. S. Agarwal, L. Melartin and T. Burch, *Hum. Biol.*, 41 (1969) 263.
13 J. R. Brown, in V. M. Rosenoer, M. Oratz and M. A. Rothschild (Editors), *Albumin Structure, Function and Uses*, Pergamon Press, Oxford, 1977, p. 27.
14 B. Meloun, L. Moravek and V. Kostka, *FEBS Lett.*, 58 (1975) 134.
15 A. Dugaiczyk, S. W. Law and O. E. Dennison, *Proc. Natl. Acad. Sci. U.S.A.*, 79 (1982) 71.
16 M. Mariotti, C. Salmon and G. Lucotte, *Protides Biol. Fluids, Proc. Colloq.*, 33 (1985) 177.