COMPLETE AMINO ACID SEQUENCE OF HUMAN SERUM ALBUMIN

B. MELOUN, L. MORÁVEK and V. KOSTKA
Institute of Organic Chemistry and Biochemistry, Czechoslovak Academy of Sciences, 166 10 Prague 6, Czechoslovakia

Received 21 July 1975

1. Introduction

The complete understanding of the binding and transport functions of human serum albumin, the main protein constituent of plasma, requires the knowledge of its chemical structure. The results of sequence studies on human serum albumin reported till the end of 1970 have been summarized in the Atlas of Protein Sequence and Structure [1]; the largest continuous part of the sequence known by then was that of the 24-residue N-terminal region [2,3]. These studies had shown that the heterogeneity of albumin preparations [4,5] does not originate in its primary structure. In 1970 we began in this Laboratory systematic studies of human serum albumin aimed at the determination of the complete amino acid sequence of the protein. The primary structure of human serum albumin was concurrently studied also at the University of Buffalo [6,8], at Craighton University Medical School [9], and at the University of Texas [10].

This paper reports on the complete amino acid sequence of human serum albumin, including the positions of all amides; the results presented are confronted with sequential data reported from other laboratories.

2. Experimental

The monomer of human serum albumin, prepared from a commercial preparation of Imuna, Šarišské Michalany by the method of Pedersen [11], served as starting material. A preparation made at the Institute of Sera and Vaccines, Prague, was used during later stages of work. The amino acid analysis of human serum albumin [12] showed that the molecule of this protein contains approximately 575 amino acid residues, of their number 6 methionines. We cleaved the molecule of human serum albumin with intact disulfide bonds by cyanogen bromide, in analogy to the procedure used for bovine serum albumin by King and Spencer [13]. Three cyanogen bromide fragments [12,14] were isolated and marked by symbols N,M, and C [15] with respect to their position in the molecule (fig.1). After the scission of the disulfide bonds [14,16], seven fragments were obtained and characterized by amino acid analyses and determination of their N- and C-terminal amino acid sequences. The order of these fragments was determined after the analysis of methionine peptides from the tryptic digest of *S*-sulfo-albumin and provided a basis of a rational nomenclature (fig.1) of the cyanogen bromide fragments [17].

At the first stage of our sequential studies we determined the amino acid sequence of the 37-residue fragment CB7(Asp) [14,18] of the 31-residue fragment CB4(Pro) [14,19], and of the 36-residue fragment CB2(Ala) [20]. Subsequently, the structure of the N-terminal fragment CB1(Asp), containing 87 residues [15,21], was established. Later, we revised [22] the data concerning positions 44, 59, 61, 63, and 73–75. The amino acid sequence of fragment CB6(Pro), containing 102 residues, was determined from the analysis of its tryptic and chymotryptic digest [23], complemented by the analysis [24,25] of peptides obtained from its maleyl derivative. When studying the sequence of the 117-residue fragment CB5(Phe), we determined first its N-terminal sequence of 40 residues in the sequencer and analyzed the tryptic digest of the fragment [26]. Its complete sequence was derived from these data and from the analysis

North-Holland Publishing Company – Amsterdam

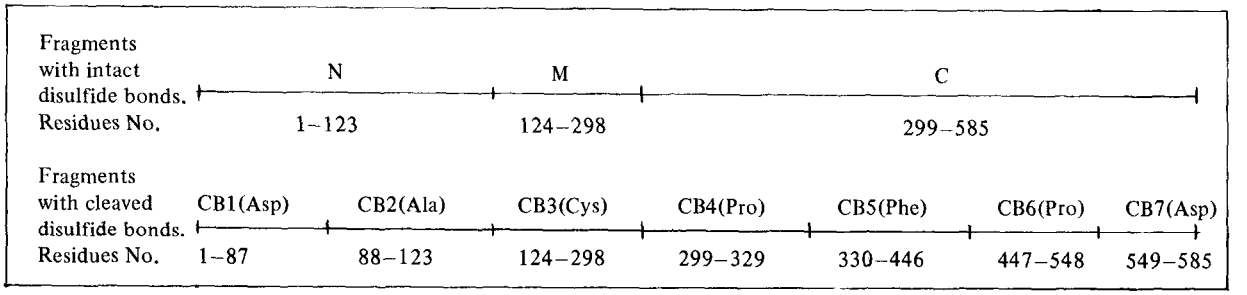

Fig.1. Nomenclature of cyanogen bromide fragments of human serum albumin.

of its chymotryptic digest and the analysis of the tryptic digest of its maleyl derivative [27]. Fragment CB3(Cys), containing 175 amino acid residues, was analyzed as the last one. Its N-terminal 40-residue sequence determined in the sequencer together with the analysis of the tryptic digest of the fragment permitted us to propose [28] a partial structure of the fragment containing 170 amino acid residues. The necessary sequential overlaps provided the analysis of the chymotryptic digest of the fragment and of the tryptic digest of its maleyl derivative [22].

3. Results and discussion

The results of our sequential studies are summarized in fig.2 which shows the complete amino acid sequence of human serum albumin. The molecule of this protein consists of a single polypeptide chain and contains 585 amino acid residues: Lys_{59}, His_{16}, Arg_{24}, Asp_{53}, Thr_{28}, Ser_{24}, Glu_{82}, Pro_{24}, Gly_{12}, Ala_{62}, Cys_{35}, Val_{41}, Met_{6}, Ile_{8}, Leu_{61}, Tyr_{18}, Phe_{31}, Trp. The mol. wt of the protein calculated from its amino acid composition is 66 500. The number of amino acid residues derived from the sequence data is in good agreement with our amino acid analysis of the protein [12] and with the analysis performed by Saifer and Palo [29]. These authors reported the content of 37 amides in the molecule of albumin. The sequence presented by us involves 17 asparagines and 21 glutamines. The data on the cyanogen bromide fragments of human serum albumin published by McMenamy et al. [7] are in good agreement with our results [12,14,16], as well as the data of Lapresle and Bellon [30] on the order of two of the cyanogen bromide fragments. The information of Babin and Goose [9] on the sequence of residues 299–329 of our structure (fig.2) also corresponds to our data. The so far most complete sequence data on human serum albumin have been published in a preliminary report by Behrens et al. [10]. When the results of these authors are compared with our complete amino acid sequence of human serum albumin numerous differences emerge. We determined the acid/amide state of amino acid residues 86, 109, 167, 170, 266, 267, 268, 269, 295, 382, 390, 391 and 397 which have not been identified by Behrens et al. [10]. We differ from these authors in the assignment of amides at positions 33, 111, 130, 131, 393, 501, 520, and 550. The sequence of residues at positions 329–348, 352–353, and 517–520, not determined by the authors, has been established in our work. There are also differences in regions sequentially determined in full, namely at positions 37, 38, 159, 160, 199, 270, 279, 281, 282, 344, 345, 352, 364, 365, 367, 368, 370, 455, 464, 465, 470, 471, and 521. At the site of phenylalanine 157, placed at this position by us and also by Gambhir and McMenamy [8], a deletion exists in the structure of Behrens et al. [10]. As a result of this fact the total number of amino acid residues reported by these authors [10] is lower by one residue.

The distribution of half-cystine residues in serum albumin, which markedly reflects its internal structure, deserves special interest. The preliminary results of our investigation of the disulfide bonds [31] lead us to postulate that the albumin chain is built up of subunits in which distances between half-cystine residues are almost equal. This hypothesis is supported by the existence of an internal homology in the amino acid sequence of other residues (fig.3). Our observations suggest that the evolution of the albumin

```
                5                  10                  15
  1  Asp-Ala-His-Lys-Ser-Glu-Val-Ala-His-Arg-Phe-Lys-Asp-Leu-Gly-
 16  Glu-Glu-Asn-Phe-Lys-Ala-Leu-Val-Leu-Ile-Ala-Phe-Ala-Gln-Tyr-
 31  Leu-Gln-Gln-Cys-Pro-Phe-Glu-Asp-His-Val-Lys-Leu-Val-Asn-Glu-
 46  Val-Thr-Glu-Phe-Ala-Lys-Thr-Cys-Val-Ala-Asp-Glu-Ser-Ala-Glu-
 61  Asn-Cys-Asp-Lys-Ser-Leu-His-Thr-Leu-Phe-Gly-Asp-Lys-Leu-Cys-
 76  Thr-Val-Ala-Thr-Leu-Arg-Glu-Thr-Tyr-Gly-Glu-Met-Ala-Asp-Cys-
 91  Cys-Ala-Lys-Glu-Gln-Pro-Glu-Arg-Asn-Glu-Cys-Phe-Leu-Gln-His-
106  Lys-Asp-Asp-Asn-Pro-Asn-Leu-Pro-Arg-Leu-Val-Arg-Pro-Glu-Val-
121  Asp-Val-Met-Cys-Thr-Ala-Phe-His-Asp-Asn-Glu-Glu-Thr-Phe-Leu-
136  Lys-Lys-Tyr-Leu-Tyr-Glu-Ile-Ala-Arg-Arg-His-Pro-Tyr-Phe-Tyr-
151  Ala Pro Glu-Leu-Leu-Phe-Phe-Ala-Lys-Arg-Tyr-Lys-Ala-Ala-Phe-
166  Thr-Glu-Cys-Cys-Glu-Ala-Ala-Asp-Lys-Ala-Ala-Cys-Leu-Leu-Pro-
181  Lys-Leu-Asp-Glu-Leu-Arg-Asp-Glu-Gly-Lys-Ala-Ser-Ser-Ala-Lys-
196  Gln-Arg-Leu-Lys-Cys-Ala-Ser-Leu-Gln-Lys-Phe-Gly-Glu-Arg-Ala-
211  Phe-Lys-Ala-Trp-Ala-Val-Ala-Arg-Leu-Ser-Gln-Arg-Phe-Pro-Lys-
226  Ala-Glu-Phe-Ala-Glu-Val-Ser-Lys-Leu-Val-Thr-Asp-Leu-Thr-Lys-
241  Val-His-Thr-Glu-Cys-Cys-His-Gly-Asp-Leu-Leu-Glu-Cys-Ala-Asp-
256  Asp-Arg-Ala-Asp-Leu-Ala-Lys-Tyr-Ile-Cys-Glu-Asn-Gln-Asp-Ser-
271  Ile-Ser-Ser-Lys-Leu-Lys-Glu-Cys-Cys-Glu-Lys-Pro-Leu-Leu-Glu-
286  Lys-Ser-His-Cys-Ile-Ala-Glu-Val-Glu-Asn-Asp-Glu-Met-Pro-Ala-
301  Asp-Leu-Pro-Ser-Leu-Ala-Ala-Asp-Phe-Val-Glu-Ser-Lys-Asp-Val-
316  Cys-Lys-Asn-Tyr-Ala-Glu-Ala-Lys-Asp-Val-Phe-Leu-Gly-Met-Phe-
331  Leu-Tyr-Glu-Tyr-Ala-Arg-Arg-His-Pro-Asp-Tyr-Ser-Val-Val-Leu-
346  Leu-Leu-Arg-Leu-Ala-Lys-Thr-Tyr-Glu-Thr-Thr-Leu-Glu-Lys-Cys-
361  Cys-Ala-Ala-His-Asp-Pro-Tyr-Glu-Cys-Ala-Ala-Lys-Val-Phe-Asp-
376  Glu-Phe-Lys-Pro-Leu-Val-Glu-Glu-Pro-Gln-Asn-Leu-Ile-Lys-Gln-
391  Asn-Cys-Glu-Leu-Phe-Glu-Gln-Leu-Gly-Glu-Tyr-Lys-Phe-Gln-Asn-
406  Ala-Leu-Leu-Val-Arg-Tyr-Thr-Lys-Lys-Val-Pro-Gln-Val-Ser-Thr-
421  Pro-Thr-Leu-Val-Glu-Val-Ser-Arg-Asn-Leu-Gly-Lys-Val-Gly-Ser-
436  Lys-Cys-Cys-Lys-His-Pro-Glu-Ala-Lys-Arg-Met-Pro-Cys-Ala-Glu-
451  Asp-Tyr-Leu-Ser-Val-Val-Leu-Asn-Gln-Leu-Cys-Val-Leu-Glu-His-
466  Lys-Thr-Pro-Val-Ser-Asp-Arg-Val-Thr-Lys-Cys-Cys-Thr-Glu-Ser-
481  Leu-Val-Asn-Arg-Arg-Pro-Cys-Phe-Ser-Ala-Leu-Glu-Val-Asp-Glu-
496  Thr-Tyr-Val-Pro-Lys-Gln-Phe-Asn-Ala-Glu-Thr-Phe-Thr-Phe-His-
511  Ala-Asp-Ile-Cys-Thr-Leu-Ser-Glu-Lys-Glu-Arg-Gln-Ile-Lys-Lys-
526  Gln-Thr-Ala-Leu-Val-Glu-Leu-Val-Lys-His-Lys-Pro-Lys-Ala-Thr-
541  Lys-Glu-Gln-Leu-Lys-Ala-Val-Met-Asp-Asp-Phe-Ala-Ala-Phe-Val-
556  Glu-Lys-Cys-Cys-Lys-Ala-Asp-Asp-Lys-Glu-Thr-Cys-Phe-Ala-Glu-
571  Glu-Gly-Lys-Lys-Leu-Val-Ala-Ala-Ser-Gln-Ala-Ala-Leu-Gly-Leu
```

Fig.2. Complete amino acid sequence of human serum albumin. Cysteine residue 34 bears the only free SH-group of the albumin.

Fig.3. Internal homology in sequence of amino acid residues in human serum albumin.

molecule – similarly to certain other proteins – also started from a simpler precursor of lower molecular weight.

Acknowledgements

We are indebted to Mr K. Grüner and Miss V. Špindlerová for their valuable help with the Edman degradation. We thank Mrs A. Kulhánková, Miss J. Chundelová and Mrs E. Bulantová for skillful technical assistance. We are indebted to Mr J. Zbrožek, Miss V. Himrová, and Mrs E. Dršková for the amino acid analyses.

References

[1] Atlas of Protein Sequence and Structure (Dayhoff, M. O., ed.) National Biomedical Research Foundation, Silver Spring, Md. USA.
[2] Bradshaw, R. A. and Peters, T. Jr. (1969) J. Biol. Chem. 244, 5582–5589.
[3] Liu, C. S., Shih, T. B., Hisn, M. H. and Blackwell, R. Q. (1971) J. Chinese Chem. Soc. 18, 137–144.
[4] Janatová, J., Mikeš, O. and Šponar, J. (1968) Collection Czech. Chem. Commun. 33, 788–800.
[5] Janatová, J. (1974) J. Med. 5, 149–216.
[6] Brown, J. R. (1975) Fed. Proc. 34, 591, Abstr. No. 2105.
[7] McMenamy, R. H., Dintzis, H. M. and Watson, F. (1971) J. Biol. Chem. 246, 4744–4750.
[8] Gambhir, K. K. and McMenamy, R. H. (1973) Fed. Proc. 32, 457, Abstr. No. 1317.
[9] Babin, D. R. and Goos, S. M. (1973) Eur. J. Biochem. 34, 409–414.
[10] Behrens, P. Q., Spiekerman, A. M. and Brown, J. R. (1975) Fed. Proc. 34, 591, Abstr. No. 2106.
[11] Pedersen, K. O. (1962) Arch. Biochem. Biophys., suppl. 1, 157–168.
[12] Meloun, B. and Kušnír, J. (1972) Collection Czech. Chem. Commun. 37, 2812–2816.
[13] King, T. P. and Spencer, M. (1970) J. Biol. Chem. 245, 6134–6148.
[14] Meloun, B. and Kušnír, J. (1972) Abstr. Commun. Meet. Fed. Eur. Biochem. Soc. Amsterdam, Abstr. No. 454.
[15] Meloun, B., Saber, M. A. and Kušnír, J. (1975) Biochim. Biophys. Acta 393, 505–519.
[16] Kušnír, J. and Meloun, B. (1973) Collection Czechoslov. Chem. Commun. 38, 143–154.
[17] Meloun, B., Morávek, L. and Kostka, V. (1975) Collection Czech. Chem. Commun. 40, 2195–2207.
[18] Meloun, B. and Kušnír, J. (1972) FEBS Lett. 27, 121–124.
[19] Kušnír, J., Kluh, I. and Meloun, B. (1973) Collection Czech. Chem. Commun. 38, 155–158.
[20] Kušnír, J. and Meloun, B. (1973) Biochim. Biophys. Acta 310, 124–129.
[21] Meloun, B., Saber, M. A. and Kušnír, J. (1974) 9th FEBS Meeting, Budapest, Abstr. No. fld6.
[22] Meloun, B. and Morávek, L., unpublished results.
[23] Morávek, L., Kostka, V., Saber, M. A. and Meloun, B. (1975) Collection Czech. Chem. Commun. 40, 1103–1111.
[24] Kostka, V., Saber, M. A., Morávek, L. and Meloun, B. (1975) 10th FEBS Meeting, Paris, Abstr. No. 547.
[25] Kostka, V., Saber, M. A., Morávek, L. and Meloun, B., Collection Czech. Chem. Commun., in the press.
[26] Kostka, V., Morávek, L., Rosenberg, I. and Meloun, B. (1975) Collection Czech. Chem. Commun. 40, 2544–2549.
[27] Morávek, L., Kostka, V., Rosenberg, I. and Meloun, B., Collection Czech. Chem. Commun., in the press.
[28] Morávek, L. and Meloun, B., Collection Czech. Chem. Commun., in the press.
[29] Saifer, A. and Palo, J. (1969) Analyt. Biochem. 27, 1–14.
[30] Lapresle, C. and Bellon, F. (1973) FEBS Lett. 34, 74–76.
[31] Saber, M. A., Stöckbauer, P., Morávek, L. and Meloun, B., unpublished results.