THE PRIMARY STRUCTURE OF THE HUMAN RETINOL-BINDING PROTEIN

Lars RASK, Helena ANUNDI and Per A. PETERSON
Department of Cell Research, The Wallenberg Laboratory, University of Uppsala, Box 562, S-751 22 Uppsala, Sweden

Received 22 May 1979

1. Introduction

Vitamin A is transported in plasma by a specific carrier protein, the retinol-binding protein (RBP), which consists of a single polypeptide chain with mol. wt 21 000 [1,2]. Under physiological conditions RBP forms a complex with thyroxine-binding prealbumin [3,4]. Both the primary and the tertiary structure of prealbumin have been determined [5,6]. For RBP only limited sequence information in amino-terminal region has been published [7,8]. The primary structure of RBP is presented here.

2. Materials and methods

RBP was isolated as in [9]. Fragmentation of the protein was achieved with CNBr [10] and partial acid hydrolysis [11]. Reduction and [^{14}C]carboxymethylation of the intact protein and various fragments have been described [12]. The intact protein and the various CNBr and acid hydrolysis fragments were digested with trypsin, chymotrypsin and thermolysin, all products of Worthington. Enzymatic digestions were also carried out with clostripain (Boehringer) and *Staphylococcus aureus* protease [13], a kind gift of Dr J. Fohlman.

CNBr and acid hydrolysis fragments were isolated by gel chromatography on Sephadex G-100 and G-50 columns equilibrated with 0.05 M sodium acetate buffer (pH 5.5) containing 6 M guanidine hydrochloride. After desalting on columns of Sephadex G-25 equilibrated with 0.1% ammonia–10% propanol in water the fragments were lyophilized.

Peptides generated on enzymatic digestion were chromatographed on Sephadex G-50 and G-25 columns equilibrated with 0.1% ammonia–10% propanol in water. Further purification was achieved by ion-exchange chromatography on sulfonated polystyrene resin with use of a modified Jeol 6AH amino acid analyzer [14]. Large peptides present in clostripain digests were isolated by a procedure involving gel chromatography on Sephadex G-100 columns equilibrated with 0.05 M sodium acetate buffer (pH 5.5) containing 6 M guanidine hydrochloride and subsequent chromatography on DEAE–Sephadex columns at pH 8.0 in 0.02 M NH_4HCO_3. Elution was performed with gradients of 0.02–0.5 M NH_4HCO_3. Amino acid analysis was performed on a Beckman 121M amino acid analyzer. Tryptophane was quantitated after hydrolysis in methanesulfonic acid [15]. Amino acid sequence determinations were either carried out manually by the dansyl-Edman technique [16,17] or automatically in a Beckman 890C liquid phase sequencer with use of polybrene to retain the peptide. Phenylthiohydantoin derivatives were identified by high-pressure liquid chromatography [12]. Digestions with carboxypeptidases A and B were performed as in [18] with the notable exceptions that quantities of both the substrates and the enzymes were reduced 10-fold, and lowered to pH 7 to reduce unspecific tryptic-like activity. The digestions were carried out at room temperature.

3. Results and discussion

CNBr fragmentation of unreduced RBP give rise to 3 distinct peaks termed A–C, on gel chromatography. Amino acid analyses showed that A but not B and C contained cysteine. Reduction and carboxymethylation of fragment A followed by gel chromatography

resolved the material into 3 peaks A1, A2 and A3. Whereas A1 and A2 represented single fragments, A3 was a mixture of 2 fragments. The latter material was therefore subjected to gel chromatography on a Sephadex G-50 column which resolved the fragments into 2 peaks, A3a and A3b. As A3a and A3b are of similar size it seems reasonable to suggest that the retarded elution position of A3b is dependent on its interaction with the gel matrix.

The amino acid compositions of the isolated CNBr fragments (table 1) showed that all fragments contained homoserine except A1. This information together with carboxypeptidase digestions of A1 and intact RBP firmly established that A1 is the carboxy-terminal fragment (fig.1). NH_2-terminal amino acid sequence determinations of intact RBP and the various CNBr fragments showed that CNBr fragment A3a is the NH_2-terminal one and that it is succeeded by fragment B. The amino acid composition of fragment A2 was approximately identical to the sum of the amino acid compositions of fragments B and A3b (table 1). This information was further substantiated by amino acid sequencing of fragment A2 which established that it represented fragment B succeeded by fragment C (see fig.1). Moreover, of the 3 HCOOH-fragments isolated amino acid sequence analysis of H-1 corroborated the order of CNBr-fragments B and A3b and demonstrated in addition that fragment C appeared next in the sequence. One clostripain peptide (Cl-1) also provided an overlap between A3b and C. Acid cleavage fragment H-2 clearly demonstrated that C preceeded A1. As shown in fig.1 these data allowed the unambiguous ordering of the CNBr fragments.

The amino acid sequence of A3a, B, A3b and C were almost fully elucidated by analyses in the automatic sequencer. Confirmation and identification of missing residues were obtained by automatic sequencing of fragments H-1 and H-2 and the clostripain peptide C1-1. Fragment A1, representing > 50% of the RBP chain, was sequenced by the automatic procedure which provided 50 consecutive

Table 1
Amino acid compositions[a] of cyanogen bromide fragments of RBP. The integral values in parentheses are based on the sequence

	CNBr-A1	CNBr-A2	CNBr-A3a	CNBr-A3b	CNBr-B	CNBr-C	RBP[d]
Lysine	3.41 (3)	2.75 (3)	2.00 (2)	1.00 (1)	1.86 (2)	1.82 (2)	10.19 (10)
Histidine	2.00 (2)	0.26					2.00 (2)
Arginine	7.88 (8)	2.36 (2)	3.68 (4)	2.09 (2)	0.30		14.02 (14)
CM-Cysteine	4.12 (4)	1.33 (1)	1.21 (1)	0.72 (1)			4.91 (6)
Aspartic acid	14.01 (14)	6.90 (8)	3.36 (3)	3.54 (4)	4.00 (4)	2.00 (2)	26.03 (27)
Threonine[b]	3.45 (3)	1.90 (2)	1.13 (1)	0.95 (1)	1.25 (1)	2.60 (3)	9.35 (9)
Serine[b]	5.67 (6)	2.65 (2)	3.15 (3)	1.23 (1)	1.35 (1)	0.38	10.89 (11)
Homoserine	0.05	0.56 (2)	0.39 (1)	0.42 (1)	0.40 (1)	0.48 (1)	
Glutamic acid	10.30 (10)	5.94 (5)	2.30 (2)	0.43	4.38 (5)	1.41 (1)	18.28 (18)
Proline	3.36 (3)	1.36 (1)			1.10 (1)	1.18 (1)	6.84 (5)
Glycine	6.23 (6)	2.87 (3)	1.17 (1)	1.20 (1)	2.00 (2)	1.41 (1)	11.21 (11)
Alanine	5.35 (5)	4.08 (5)	2.36 (2)	2.57 (3)	2.01 (2)	1.12 (1)	13.90 (13)
Valine	5.95 (6)	3.69 (4)	2.05 (2)	1.97 (2)	1.89 (2)	1.00 (1)	12.20 (13)
Methionine							3.84 (4)
Isoleucine[c]	3.07 (3)	0.78 (1)			0.68 (1)		3.96 (4)
Leucine[c]	9.00 (9)	3.89 (4)		1.65 (2)	1.95 (2)	0.36	12.78 (13)
Tyrosine	6.75 (7)	0.41	0.93 (1)		0.37		8.06 (8)
Phenylalanine	3.29 (3)	2.24 (2)	2.64 (3)	0.34	1.81 (2)	1.80 (2)	10.14 (10)
Tryptophane	1.48 (2)	0.61 (1)	0.67 (1)	0.71 (1)			5.22[e] (4)

[a] Except where noted all figures are average values of one 24 h and one 72 h hydrolysis
[b] Values were obtained by extrapolation to 0 h hydrolysis
[c] 72 h hydrolysis value
[d] From [18]
[e] Determined spectrophotometrically

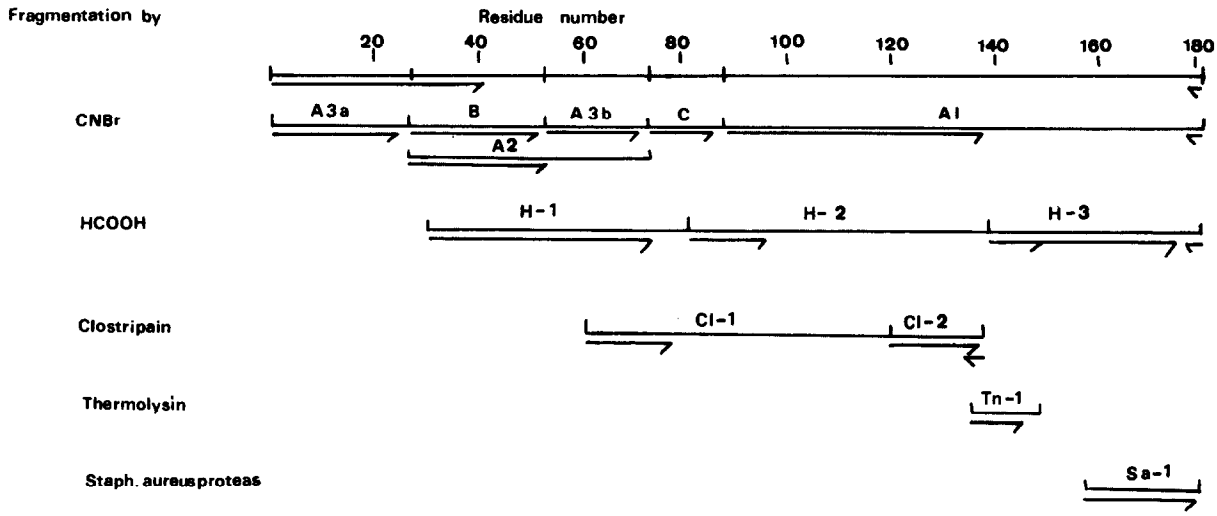

Fig.1. Schematic outline of the various fragments and peptides used to establish the amino acid sequence of RBP. Arrows denote amino acid sequences determined by automatic sequencing (directed to the right) and by carboxypeptidase digestions (directed to the left).

residues, clostripain peptide C-2 and the thermolysin peptide Tn-1 connected the first 52 amino acids of fragment A1 with the following 36 residues determined by automatic sequencing of acid hydrolysis fragment H-3. The remaining amino acid residues of fragment A1 were determined by automatic sequencing of a *Staph. aureus* protease peptide (Sa-1) in conjunction with carboxypeptidase digestions of intact RBP, fragment A1 and peptide Sa-1. This information was sufficient to establish the primary structure of RBP. However, tryptic, chymotryptic and thermolysin peptides, sequenced by the manual dansyl-Edman technique, provided corroborative data for the entire sequence.

The complete amino acid sequence of RBP is outlined in fig.2. The first 50 amino acids are in complete agreement with the limited information available for this region of the molecule [7,8]. This stretch includes the first methionyl residue (position 27) and the first aspartyl–prolyl bond (residues 31–32) which is cleaved by acid. The second methionine (position 53) is succeeded by serine which explains why CNBr gave incomplete cleavage at this point (see [19]). Following the third methionine in

```
 1                5                10               15                20                25                30
Glu-Arg-Asp-Cys-Arg-Val-Ser-Ser-Phe-Arg-Val-Lys-Glu-Asn-Phe-Asp-Lys-Ala-Arg-Phe-Ser-Gly-Thr-Trp-Tyr-Ala-Met-Ala-Lys-Lys-

 31               35                40               45                50                55                60
-Asp-Pro-Glu-Gly-Leu-Phe-Leu-Gln-Asp-Asn-Ile-Val-Ala-Glu-Phe-Ser-Val-Asp-Glu-Thr-Gly-Gln-Met-Ser-Ala-Thr-Ala-Lys-Gly-Arg-

 61               65                70               75                80                85                90
-Val-Arg-Leu-Leu-Asn-Asn-Trp-Asp-Val-Cys-Ala-Asp-Met-Val-Gly-Thr-Phe-Thr-Asp-Thr-Glu-Asp-Pro-Ala-Lys-Phe-Lys-Met-Lys-Tyr-

 91               95                100              105               110               115               120
-Trp-Gly-Val-Ala-Ser-Phe-Leu-Gln-Lys-Gly-Asn-Asp-Asp-His-Trp-Ile-Val-Asp-Thr-Asp-Tyr-Asp-Thr-Tyr-Ala-Val-Gln-Tyr-Ser-Cys-

 121              125               130              135               140               145               150
-Arg-Leu-Leu-Asn-Leu-Asp-Gly-Thr-Cys-Ala-Asp-Ser-Tyr-Ser-Phe-Val-Phe-Ser-Arg-Asp-Pro-Asn-Gly-Leu-Pro-Pro-Gln-Ala-Gln-Lys-

 151              155               160              165               170               175               180
-Ile-Val-Arg-Gln-Arg-Gln-Glu-Glu-Leu-Cys-Leu-Ala-Arg-Gln-Tyr-Arg-Leu-Ile-Val-His-Asn-Gly-Tyr-Cys-Asp-Gly-Arg-Ser-Glu-Arg-

 181
-Asn-Leu.
```

Fig.2. Primary structure of the human retinol-binding protein.

position 73 the second aspartyl–prolyl bond occurs (residues 82–83) only 6 residues before the fourth methionine (position 88). The third aspartyl–prolyl bond in the sequence occurs at positions 140–141. It it obvious that all 3 aspartyl–prolyl bonds are highly susceptible to limited acid hydrolysis. The two clostripain peptides Cl-1 and Cl-2 have arginine in the COOH-terminus consistent with the specificity of the enzyme. Likewise, the *Staph. aureus* peptide was generated on cleavage of a glutamyl–leucyl bond confirming the specificity of this enzyme.

Several laboratories including our own have presented data regarding the carboxy-terminal amino acid sequence of RBP [18–21]. As these determinations have been at variance, particular attention was paid to this region. The data presented here were obtained by carboxypeptidase-digestions, sequencing of a large peptide (Sa-1) in the automatic sequencer, and isolation and manual sequence determination of tryptic peptides. The sequence now established differs from all previously published. Our earlier data [18], demonstrating arginine and lysine as COOH-terminal residues of RBP, were most probably due to tryptic activity in our carboxypeptidase preparations. It has been suggested [21] that the sequence is Arg-Leu. It is likely that they overlooked the asparaginyl residue due to its low ninhydrin colour yield. Moreover, in most amino acid analysis systems asparagine occurs in about the same position as serine.

The amino acid compositions of the individual CNBr fragments were in good agreement with the sequence established. Furthermore, the amino acid composition of intact RBP was in excellent agreement with that calculated from the primary structure information and was very similar to the compositions [18,22] with the obvious exception of [21]. These authors found markedly lower contents of amide and isoleucine and higher values of methionine and tryptophane.

The knowledge of the primary structure of RBP is a prerequisite for determinations of its tertiary structure. Such an analysis is now feasible and may be very rewarding as the RBP molecule is engaged in at least 3 molecular interactions, i.e., with prealbumin, with retinol and with a cell surface receptor [23].

Acknowledgements

Expert assistance has been provided by Mr Kjell Andersson, Mr Jörgen Ericson, Mrs Inga Sjöquist and Mrs Yvonne Tillman. Drs Hans Bennich and Jan Fohlman are gratfully acknowledged for invaluable discussions throughout this work. This study was supported by grants from the Swedish Medical Research Council and from the National Institute of Health (project no. 1Ro 1EYo2417–01).

References

[1] Kanai, M., Raz, A. and Goodman, D. S. (1968) J. Clin. Invest. 47, 2025–2044.
[2] Peterson, P. A. (1969) Abstr. Uppsala Diss. Med. 75, 1–12.
[3] Peterson, P. A. (1971) J. Biol. Chem. 246, 34–43.
[4] Raz, A., Shiratori, T. and Goodman, D. S. (1970) J. Biol. Chem. 245, 1903–1912.
[5] Kanda, Y., Goodman, D. S., Canfield, R. E. and Morgan, F. J. (1974) J. Biol. Chem. 249, 6796–6805.
[6] Blake, C. C. F., Swan, I. D. A., Rerat, C., Berthou, I., Laurent, A. and Rerat, B. (1971) J. Mol. Biol. 61, 217–224.
[7] Morgan, F. J., Canfield, R. E. and Goodman, D. S. (1971) Biochim. Biophys. Acta 236, 798–801.
[8] Poulik, M. D., Farrah, D., Malek, G. H., Shinnick, C. J. and Smithies, O. (1975) Biochim. Biophys. Acta 412, 326–334.
[9] Peterson, P. A. and Berggård, I. (1971) J. Biol. Chem. 246, 25–33.
[10] Gross, E. and Witkop, K. (1962) J. Biol. Chem. 237, 1856–1862.
[11] Landon, M. (1977) Methods Enzymol. 47, 145–149.
[12] Wiman, K., Trägårdh, L., Rask, L. and Peterson, P. A. (1979) Eur. J. Biochem. 95, 265–273.
[13] Rydén, A. C., Rydén, L. and Philipson, L. (1974) Eur. J. Biochem. 44, 105–114.
[14] Hartley, R. W. (1972) Anal. Biochem. 46, 676–680.
[15] Moore, S. (1972) in: Chemistry and Biology of Peptides (Meienhofer, J. ed) pp. 629–653, Ann Arbor, Michigan.
[16] Gray, W. R. (1967) Methods Enzymol. 11, 469–475.
[17] Woods, K. R. and Wang, K.-T. (1967) Biochim. Biophys. Acta 133, 369–370.
[18] Rask, L., Vahlquist, A. and Peterson, P. A. (1971) J. Biol. Chem. 246, 6638–6646.
[19] Schroeder, W. A., Shelton, J. B. and Shelton, J. R. (1969) Arch. Biochem. Biophys 130, 531–556.
[20] White, G. H., Western, S. M. and Glover, J. (1972) FEBS Lett. 27, 107–110.
[21] Fex, G. and Hansson, B. (1979) Eur. J. Biochem. 94, 307–313.
[22] Muto, Y. and Goodman, D. S. (1972) J. Biol. Chem. 247, 2533–2541.
[23] Rask, L. and Peterson, P. A. (1976) J. Biol. Chem. 251, 6360–6366.