THE PRIMARY STRUCTURE OF HUMAN PLASMINOGEN: CHARACTERIZATION AND ALIGNMENT OF THE CYANOGEN BROMIDE PEPTIDES

How-Ming LEE and Richard A. LAURSEN
Department of Chemistry, Boston University, Boston, Massachusetts, 02215 USA

Received 1 June 1976

1. Introduction

Human plasminogen, a single-chain polypeptide containing about 810 amino acid residues, is converted on activation to plasmin, which possesses a heavy and a light chain containing about 510 and 230 residues, respectively. Recent amino acid sequence studies [1–6] have demonstrated the existence of a high degree of internal homology within the heavy chain of plasminogen. In fact Magnusson and coworkers [1,2] have shown that about 80% of the heavy chain can be accounted for by five mutually homologous regions ('kringles') that are characterized by six invariant half-cystine residues and are also homologous with two analogous regions in prothrombin.

The heavy chain sequence proposed by Magnusson and co-workers [1,2] * is based on overlapping sets of tryptic and chymotryptic peptides. In the present paper we present the results of our studies with the cyanogen bromide peptides of plasminogen, which, for the most part, confirm the assignments of Magnusson, but which also suggest that the process of gene duplication in plasminogen occurred with less regularity than previously thought [1,2]. In addition we demonstrate the utility of a selective solid-phase peptide coupling technique for aligning cyanogen bromide peptides.

* A revised version of the sequence presented in [1] is given in [2].

2. Materials and methods

Human plasminogen (lysine form [7]) was prepared from Cohn fraction III or $III_{2,3}$ by the method of McClintock et al. [8].

Plasminogen was cleaved with cyanogen bromide as described earlier [3] and then chromatographed on Sephadex G-50 in 10% acetic acid to separate non-cystine peptides from the cystine core. The core was then reduced and carboxymethylated and further fractionated by gel filtration in 2.5% acetic acid and chromatography on DEAE cellulose and DEAE BioGel using NH_4HCO_3 gradient elution. In some cases, peptides were further cleaved with trypsin after maleylation.

Overlapping methionine peptides were obtained by digestion of reduced, carboxymethylated plasminogen with trypsin, followed by separation of the peptides on Sephadex G-25, DEAE BioGel and Dowex-50 columns.

Sequencing was accomplished by solid-phase Edman degradation [9,10] using a Sequemat Model 12 sequencer. Cyanogen bromide peptides were coupled to resins by the homoserine lactone method [11], and lysine peptides by the diisothiocyanate method [12]. In order to sequence the C-terminal portion of cyanogen bromide peptides, the peptide was digested with trypsin and the C-terminal fragment was selectively removed from the mixture by coupling the homoserine moiety to a sequencing resin [11,13]. Large peptides were attached to aminopropyl glass [14]. Phenylthiohydantoins were identified by thin-layer chromatography and HI hydrolysis to amino acids.

3. Results and discussion

Cyanogen bromide cleavage of Lys-plasminogen gave nine peptides, of which eight were characterized in this work (tables 1 and 2). The ninth, CB-VIII, an approximately 200-residue peptide which comprises most of the light chain, precipitated in 2.5% acetic acid and did not elute from the Sephadex G-50 column.

The strategy used for alignment of the peptides involved sequencing of the N- and C-terminal portions of the peptides (or in some cases, the entire peptide), and then overlapping with the methionine-containing tryptic peptides. C-terminal sequencing was greatly facilitated by solid-phase methods, by which the C-terminal fragments of large cyanogen bromide peptides could be attached selectively to sequencing resins at homoserine after tryptic digestion [13]. This work represents the first example of the use of this strategy in aligning the peptides of an entire protein.

Table 1
Comparison of amino acid compositions of cyanogen bromide peptides CB-II and CB-VI from plasminogen. Nearest integral values are given for convenience

	CB-II			CB-VI		
	CBII[a]	D_b[b]	c	CBVI[a]	D_a[b]	c
CMCys	8	8	6	7	7	5
Asp	12	12	8	12	11	8
Thr	7	8	6	6	6	3
Ser	8	8	6	7	6	4
Glu	15	14	10	8	8	5
Pro	12	11	9	14	12	9
Gly	5	4	3	3	3	2
Ala	2	1	1	5	4	3
Val	0	0	0	8	9	6
Ile	4	3	2	0	0	0
Leu	4	3	2	5	5	3
Tyr	7	6	4	4	4	2
Phe	2	2	1	2	2	1
Trp	n.d.	3	2	n.d.	3	2
His	3	2	2	0	0	0
Lys	5	5	4	4	4	5
Arg	5	5	4	3	3	2
Hse	(1)	(1)	(1)	(1)	(1)	(1)
Total	100	96	73	89	88	61

[a] This work.
[b] The corresponding peptides reported by Wiman and Wallen [5].
[c] From sequence data of Magnusson et al. [2].

Tables 2 and 3 list sequence data for cyanogen bromide peptides and tryptic peptides containing methionine, respectively. The alignment of these peptides is shown in fig.1.

In general our results confirm those of Magnusson et al. [1,2], who found the heavy chain to be comprised primarily of five homologous segments (referred to as 'kringles' by Magnusson) of about 80 residues each. These regions are also homologous with two similar regions in prothrombin (fig.2), and are characterized by six invariant half-cystine residues which are believed to form three disulfide loops in each kringle [2]. Some homology (analogy) is also seen in the non-kringle regions (fig.2).

The main discrepancy between our data and that of Magnusson et al. [1,2] lies in peptide CB-VI, where we find a 22-residue segment with the sequence Val–Val–Ser–Gly–Pro–Glu–Trp–Val–Val–Leu–Lys–Pro–Asp–Glu–Gly–Thr–Leu–Leu–Glu–Asp–Asp–Lys, whereas Magnusson finds Lys–Lys, Lys–Lys (table 2). Thus according to our results, kringle D (fig.2) is approximately 20 residues longer than the others. Furthermore, this 20-residue segment (D′ in fig.2) appears to have arisen by duplication of part of the sequence immediately following it. Wiman and Wallen [5] have also isolated CB-VI and report an amino acid composition and mol. wt. (9000) consistent with our findings (table 1).

A second difference is seen in peptide CB-II which both we and Wiman and Wallen find to have a mol. wt. of at least 11 000 and an amino acid composition of around 100 residues, whereas Magnusson et al. [1,2] find only 72 residues. Although our sequence analysis of this peptide is not complete, we find some peptides derived from CB-VI that do not fit Magnusson's sequence, indicating that it, and therefore kringle A, may be larger than indicated in fig.2.

Despite a few discrepancies that remain to be resolved, our results provide an independent confirmation of Magnusson's sequence, which shows that the heavy chain of plasminogen is composed of five homologous regions resulting from multiple gene duplication. At this time we do not know whether any of the differences are due to errors in sequencing or to isolation of different microheterogeneous forms [7,16] of plasminogen.

Table 2

Sequence data for cyanogen bromide peptides. Peptides are numbered in order, starting with the N-terminal peptide. Unless otherwise noted, the sequences were determined in our laboratory. Underlined amino acids indicate residues confirmed by Wiman and Wallen [4–6]. Amino acids given in capitals are residues found by Magnusson et al. [1,2] which differ from our assignments; otherwise the two sets of data are in agreement.

Peptide	Sequence
CB-I	NH_2-Lys–Val–Tyr–Leu–Ser–Glu–Cys–Lys–Thr–Gly-XXX-Gly-XXX-Asn–Tyr–Arg–Gly–Thr–Hse ASP LYS
CB-II	Ser–Lys–Thr–Lys–Asn–Gly–Ile-(ca. 80 residues)-Arg–Tyr–Asp–Tyr–Cys–Asp–Ile–Leu–Glu–Cys–Glu– GLX Glu–Glu–Cys–Hse
CB-III	His–Cys–Ser–Gly–Glu–Asn–Tyr–Asp–Gly–Lys–Ile–Ser–Lys–Thr–Hse GLX ASX
CB-IV	Ser–Gly–Leu–Glu–Asn–Gln–Ala–Lys–Ser-(ca. 185 residues)-XXX-Gln–Ser–Trp-XXX-XXX-Hse CYS TRP CYS SER SER
CB-V	Thr–Pro–His–Arg–His–Gln–Lys–Thr–Pro–Glu–Asn–Tyr–Pro–Asn–Ala–Gly–Leu–Thr–Hse GLX GLX ASX
CB-VI	Asn–Tyr–Cys–Arg–Asn–Pro–Asp–Ala–Asp–Gly–Gly–Pro–Trp–Cys–Phe–Val–Thr–Asp-XXX-XXX-XXX-Arg– ASX ASX ASX THR Arg Val Trp–Glu–Tyr–Cys–Asn–Leu–Lys–Val–Val–Ser–Gly–Pro–Glu–Trp–Val–Val–Leu–Lys–Pro–Asp–Glu–Gly– LYS-LYS,LYS-LYS – – – – – – – – – – – – Thr–Leu–Leu–Glu–Asp–Asp–Lys–Cys–Ser–Gly–Thr–Glu–Ala–Ser–Val–Val–Ala–Pro–Pro–Pro–Val–Val–Leu– – – – – – – – GLX (PRO,PRO,VAL,LEU, Leu–Pro–Asn–Val–Glu–Pro–Pro–Ser–Glu–Glu–Asp–Cys–Hse LEU,PRO,ASX,VAL,GLX,PRO,PRO,SER,GLX,GLX,ASX,CYS,MET)
CB-VII[a]	Phe–Gly–Asn–Gly–Lys-(113 residues)-Arg–Phe–Gly–Hse
CB-VIII[b]	His–Phe–Cys–Gly–Gly–Thr–Leu–Ile–Ser–Pro–Gly–Trp–Val–Leu–Ser–Ala–Ala-(ca. 185 residues)-Hse
CB-IX	Arg–Asn–Asn-COOH

[a] The complete sequence of this peptide has been reported by Wiman and Wallen [4].

[b] The N-terminal sequence of this peptide was determined by Robbins et al. [15].

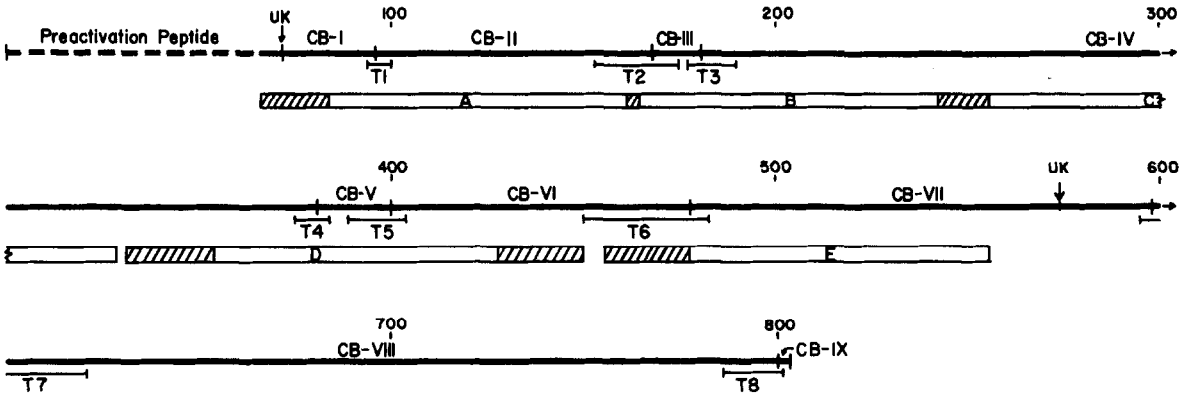

Fig.1. Alignment of cyanogen bromide peptides in plasminogen. The heavy-chain kringle regions (A–E) are indicated by open bars, and the inter-kringle regions by cross-hatching.

Table 3
Sequence data for tryptic peptides containing methionine

T1	Gly–Thr–Met–Ser–Lys
T2	Tyr–Asp–Tyr–Cys–Asp–Ile–Leu–Glu–Cys–Glu–Glu–Glu–Cys–Met–His–Cys–Ser–Gly–Glu–Asn–Tyr–Asp–Gly–Lys
T3	Thr–Met–Ser–Gly–Leu–Glu–Asn–Gln–Ala–Lys
T4	Cys–Gln–Ser–Trp–Ser–Ser–Met–Thr–Pro–His–Arg
T5	Thr–Pro–Glu–Asn–Tyr–Pro–Asn–Ala–Gly–Leu–Thr–Met–Asn–Tyr–Cys–Arg
T6[a]	Cys–Ser–Gly–Thr–Glu–Ala–Ser–Val–Val–Ala–Pro–Pro–Pro–Val–Val–Leu–Leu–Pro–Asn–Val–Glu–Pro–Pro-(Ser,Glx,Glx,Asx,Cys)-Met–Phe–Gly–Asn–Gly–Lys
T7[b]	Phe–Gly–Met–His–Phe–Cys–Gly–Gly–Thr–Leu–Ile–Ser–Pro-(Glx,Trp,Val,Leu,Ser,Ala,His,Cys,Leu)-Lys
T8	(Cys_1 Asx, Thr_3 Glx_2 Pro_2 Gly_1 Ala_2 Val_1 His_1)-Met–Arg

[a] Not sequenced entirely. The portion in parentheses was deduced from the compositions and known sequences of CB-VI and CB-VII.

[b] Not sequenced entirely. Sequence in parentheses corresponds to that reported by Robbins et al. [15].

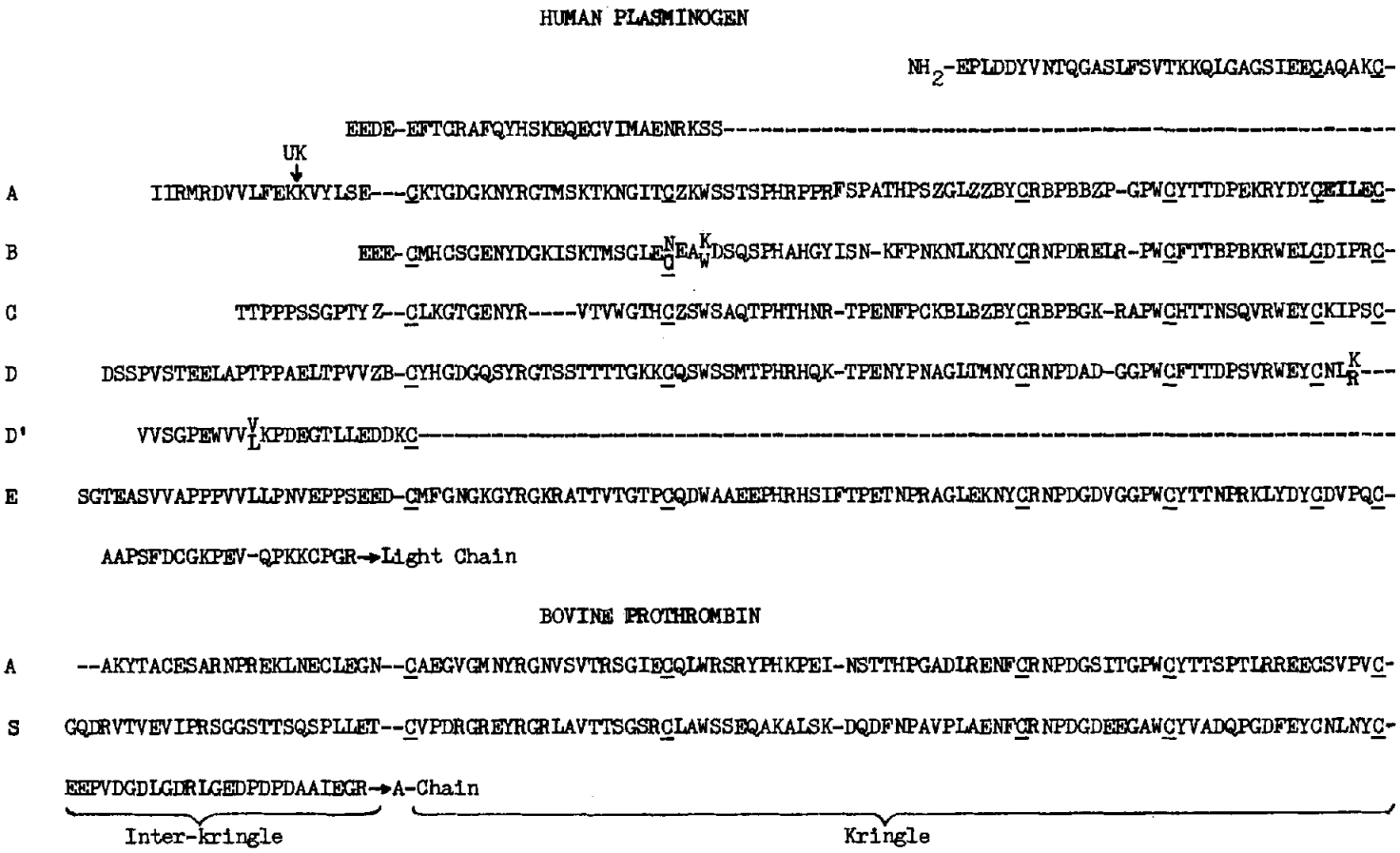

Fig.2. Tentative sequence of human plasminogen heavy chain based on data from several laboratories [1–6] and this work, and showing the five regions (A–E) of homology. The analogous regions in prothrombin are given in the lower lines. Inter-kringle regions are arranged to show maximum homology. The placement of the top two lines, which comprise most of the pre-activation peptide, is speculative, and the observed analogous sequences may be the result of coincidence rather than any sort of gene duplication process. Segment D′ is a part of kringle D and has a sequence similar to the inter-kringle region following it. Positions for which there is evidence for more than one amino acid are indicated by two symbols. The symbols B and Z are Asx and Glx, respectively.

Acknowledgements

We wish to thank Dr S. Magnusson for providing us with a preprint of his sequence data and Mrs Mei-Chu Lee for excellent technical assistance. This work was supported by a grant (HL-15535) from the National Institutes of Health. R.A.L. was the recipient of a USPHS Research Career Development Award and an Alfred P. Sloan Fellowship.

References

[1] Claeys, H., Sottrup-Jensen, L., Zajdel, M., Petersen, T. E. and Magnusson, S. (1975) FEBS Lett. 61, 20–24.
[2] Magnusson, S., Sottrup-Jensen, L., Petersen, T. E., Dudek-Wojciechowska, G. and Claeys, H. (1976) in: Proceedings of the 8th Miami Winter Symposium, January 12–16, 1976 (K. Brew and D. W. Ribbons, eds.) Academic Press, New York.
[3] Laursen, R. A. and Lee, H. M. (1975) FEBS Lett. 56, 70–72.
[4] Wiman, B. and Wallen, P. (1975) Eur. J. Biochem. 58, 539–547.
[5] Wiman, B. and Wallen, P. (1975) Eur. J. Biochem. 57, 387–394.
[6] Wiman, B. and Wallen, P. (1975) Eur. J. Biochem. 50, 489–494.
[7] Summaria, L., Arzadon, L., Bernabe, P. and Robbins, K. C. (1973) J. Biol. Chem. 248, 2984–2991.
[8] McClintock, D. K., Englert, M. E., Dziobkowski, C., Snedeker, E. H. and Bell, P. H. (1974) Biochemistry 13, 5334–5344.
[9] Laursen, R. A. (1971) Eur. J. Biochem. 20, 89–102.
[10] Laursen, R. A. (1975) in: Immobilized Enzymes, Antigens, Antibodies and Peptides, pp. 567–634. (H. H. Weetall, ed.) Marcel Dekker, New York.
[11] Horn, M. J. and Laursen, R. A. (1973) FEBS Lett. 36, 285–288.
[12] Laursen, R. A., Horn, M. J. and Bonner, A. G. (1972) FEBS Lett. 21, 67–70.
[13] Horn, M. J. (1975) Anal. Biochem. 69, 583–589.
[14] Wachter, E., Hofner, H. and Machleidt, W. (1975) in: Solid-Phase Methods in Protein Sequence Analysis, pp. 31–46. (R. A. Laursen, ed.) Pierce Chemical Co., Rockford, Ill.
[15] Robbins, K. C., Bernabe, P., Arzadon, L. and Summeria, L. (1973) J. Biol. Chem. 248, 1631–1633.
[16] Wallen, P. and Wiman, B. (1972) Biochim. Biophys. Acta 257, 122–134.