Amino acid sequence of human factor D of the complement system

Similarity in sequence between factor D and proteases of non-plasma origin

D.M.A. Johnson, J. Gagnon and K.B.M. Reid*

MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, England

Received 31 October 1983

The amino acid sequence of human factor D is proposed from the analysis of the peptides produced by treatment of the factor D with cyanogen bromide, iodosobenzoic acid, trypsin and V-8 protease. Comparison of the proposed sequence with the sequences of other serine esterases indicated that factor D, although it is a plasma serine esterase, is more closely related to certain proteases not found in the plasma than to other plasma serine esterases of the complement system. For example, 36% and 32% identity in amino acid sequence was found on comparison of factor D with elastase and group-specific protease, respectively. Whereas only 27% and 18% identity was observed between factor D and the other complement serine esterases, Clr and factor B, respectively.

Human factor D *Sequence* *Esterase* *Homology*

1. INTRODUCTION

Factor D, of the alternative pathway of human complement, has been shown to be a typical serine protease both by its irreversible inhibition by di-isopropyl phosphorofluoridate [1] and by amino acid sequence analysis of the N-terminal, and active site, regions of the molecule [2–4]. The N-terminal amino acid sequence of factor D is more similar to that of rat group-specific protease [5] and porcine elastase [6] than to some of the serine esterases of the complement and clotting systems [3]. In this respect it is of interest that like rat group-specific protease, there is no strong evidence for the existence of a zymogen form of the enzyme. The physiological role of factor D is as an activator of the complex protease C3b, B which is a Mg^{2+} dependent complex of factor B and activated C3[7]. The splitting of a single Arg-Lys polypeptide bond in the factor B results in the formation of an active enzyme C3b,Bb which can split intact C3 the central component of the alternative, and classical, pathways of complement.

2. MATERIALS AND METHODS

2.1. *Isolation of human factor D*

Human factor D was isolated from outdated plasma by a procedure [8] involving euglobulin precipitation, ion-exchange chromatography on CM-Sephadex C-50 and CM-cellulose-32, gel-filtration on Sephadex G-75, and affinity chromatography.

2.2. *Preparation and isolation of peptides used in the sequencing studies*

The cyanogen bromide (CNBr)-derived peptides were prepared and purified as in [3]. Treatment of

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V.
001457593/84/$3.00

factor D with iodosobenzoic acid (IBA) was performed as follows: reduced and alkylated factor D(100 nmol) and IBA (1 mg) were dissolved in 1.0 ml of a 38.2% (w/v) guanidine hydrochloride in 80% (v/v) acetic acid, the mixture was incubated in a sealed flask, in the dark, for 24 h at 25°C [9] then applied to a column (1.5 cm × 100 cm) of Sephadex G-50 superfine equilibrated with 5% (v/v) formic acid. The large IBA-peptides were further purified on Sephadex G-75 superfine equilibrated with 10% (v/v) formic acid and the low M_r peptides were desalted on Sephadex G-10 equilibrated with 0.1 M NH_4HCO_3.

2.3. *Trypsin treatment of factor D*

Succinylated, reduced and alkylated, factor D (90 nmol) in 1.45 ml of 0.1 M NH_4HCO_3 (pH 8.0) was digested with TPCK-treated trypsin (40 μg) at 37°C for 4 h. The digest was stopped by the addition of di-isopropyl phosphorofluoridate (30-fold excess over the trypsin present). The digest was spun at 45 000 × *g* for 20 min and the supernatant applied to a column (1.5 cm × 100 cm) of Sephadex G-75 superfine equilibrated with 0.1 M NH_4HCO_3. The major peptides were further purified by HPLC on a μ-Bondapak C-18 column (0.39 cm × 30 cm) equilibrated with 0.1% (w/v) NH_4HCO_3 containing 5% (v/v) CH_3CN. The peptides were eluted over a period of 1 h in a linear gradient from 5% to 65% (v/v) CH_3CN in 0.1% (w/v) NH_4HCO_3 at a rate of 1.0 ml/min.

2.4. *Cleavage by V-8 protease*

Peptide CNBr-1 (100 nmol) was digested with V-8 protease from *S. aureus* (0.08 mg) in 1.2 ml of 0.1 M NH_4HCO_3 [10] for 8 h at 37°C. The digest was freeze-dried then the peptides were purified by HPLC, as described above, except that a gradient from 5% to 55% (v/v) CH_3CN in 0.1% (w/v) NH_4HCO_3 was used.

2.5. *Digestion with carboxypeptidase Y*

The method in [11] was used: protein, or peptide (5 nmol), and carboxypeptidase Y (5–10 μg) in acetic acid–pyridine (100 μl, pH 5.5) was incubated at 25°C for times between 0 and 5 h. The samples were boiled for 5 min to stop enzyme action and freeze-dried prior to amino acid analysis.

2.3. *Amino acid analysis and sequence determination*

Amino acid analysis was carried out as in [12]. Automated amino acid sequence determination was performed on a Beckmann 890C Sequencer and the phenylthiolhydantoin derivatives were identified and quantitated by HPLC [12].

3. RESULTS AND DISCUSSION

It can be seen (table 1) that the number of residues calculated by amino acid analysis, for fac-

Table 1

Amino acid compositions of factor D and its cyanogen bromide-derived peptides[a]

Amino acid	Residues/mol			
	Factor D	CNBr-1	CNBr-2	CNBr-3
½Cys[b]	8.4 (8)	4.2 (4)	4.0 (4)	0.1 (0)
Asx[c]	18.6 (18)	12.6 (13)	5.2 (5)	0.3 (0)
Thr	8.2 (7)	5.2 (5)	2.1 (2)	0.1 (0)
Ser	14.4 (12)	6.3 (6)	5.6 (6)	0.1 (0)
Glx[c]	18.6 (16)	12.9 (12)	2.6 (2)	2.1 (2)
Pro	14.4 (13)	9.5 (10)	2.6 (2)	1.1 (1)
Gly	21.7 (23)	9.5 (12)	7.4 (9)	2.2 (2)
Ala	21.7 (22)	12.6 (15)	4.8 (4)	2.7 (3)
Val	20.0 (23)	13.3 (16)	5.8 (7)	0.3 (0)
Met[d]	2.2 (2)	0.9 (1)	0.2 (0)	0.7 (1)
Ile	5.2 (5)	3.1 (2)	2.1 (2)	0.8 (1)
Leu	23.7 (28)	21.7 (24)	2.8 (3)	1.0 (1)
Tyr	4.1 (4)	1.0 (1)	1.7 (2)	0.9 (1)
Phe	1.4 (1)	1.2 (1)	0.4 (0)	– (0)
His	8.2 (6)	6.9 (5)	0.4 (0)	1.0 (1)
Lys	7.2 (7)	4.2 (4)	3.3 (3)	– (0)
Arg	16.5 (17)	10.3 (10)	5.4 (5)	1.6 (2)
Trp[e]	3.4 (4)	NE (3)	NE (1)	– (0)
X	– (2)	– (2)	– (0)	– (0)
	(218)	(146)	(57)	(15)

[a]Comparison of values from acid hydrolysates with those obtained by sequence analysis. The figures in brackets indicate the results of sequence analysis

[b]½Cys was determined as *S*-carboxymethylcysteine

[c]Asx = 14 Asp + 4 Asn; Glx = 10 Glu + 6 Gln

[d]Met was determined as methionine in factor D and as homoserine in the CNBr-derived peptides

[e]Trp was determined after hydrolysis of factor D in 4 N methanesulphonic acid

tor D and its CNBr-peptides, agrees closely with those predicted from the amino acid sequence analysis (table 1; fig.1). This indicates that, despite the lack of 'overlap' peptides in certain positions (fig.1), no large stretch of amino acid sequence could have been omitted from the proposed sequence. From the amino acid composition data the only region where it is considered possible that there may be residues not accounted for is in the region 68–69 where the insertion of 2, or 3 residues cannot be excluded at present. However, the sequence presented in fig. 1 should allow for the preparation of suitable oligonucleotide probes which could be used to screen cDNA libraries for the provision of the cDNA sequence which would then allow the confirmation, or correction, of the proposed amino acid sequence.

Virtually all the residues which are invariant in other serine proteases are also found in equivalent positions in factor D. An exception is residue 184 (the chymotrypsinogen numbering is used throughout the Discussion) which is usually Gly

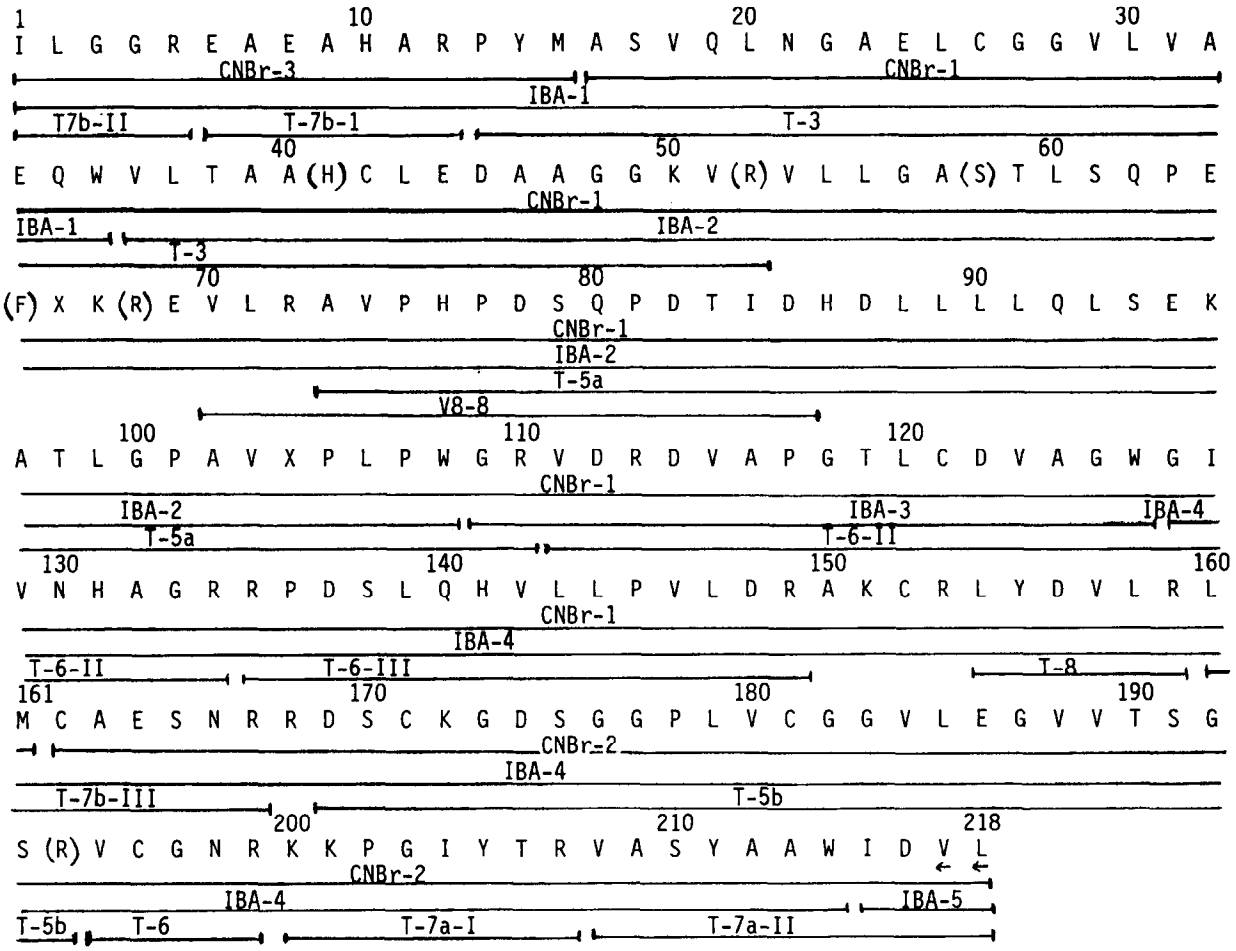

Fig. 1. Proposed amino acid sequence of human factor D. Peptides prefixed by CNBr-, IBA-, T- and V8- were derived from cyanogen bromide, idosobenzoic acid, trypsin and V8-protease digests respectively. The N-terminal sequences of the 3 CNBr-derived peptides have been published in [3]. Extensive N-terminal sequence analysis was performed on all the peptides shown and positive identifications of residues were obtained at all positions except where the residue is enclosed in brackets, denoting tentative identification, or given as X, denoting an unidentified residue. Sequence determination from the C-terminus by use of carboxypeptidase Y is shown by ←. No overlap in amino acid sequence was obtained at positions 69, 108, 149, 153, 159, 194, 199 and 207.

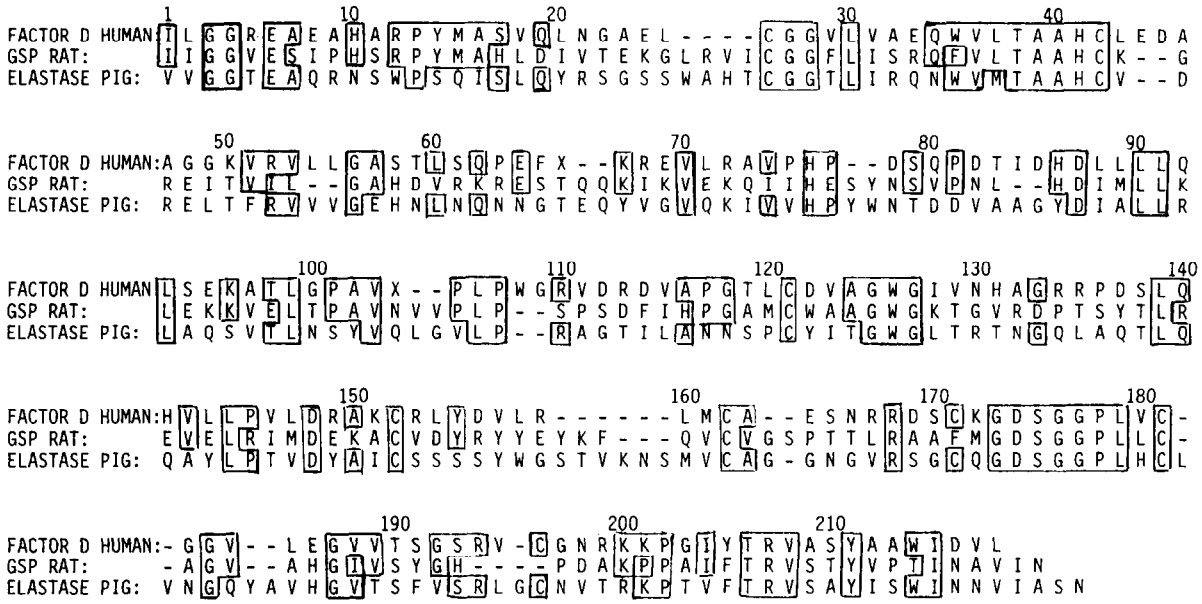

Fig. 2. Comparison of the amino acid sequence of factor D with the sequences of group-specific protease and elastase. The rat GSP and pig elastase amino acid sequences are taken from [5] and [6], respectively. Residues common to GSP and factor D, or elastase and factor D, are boxed. X denotes an unidentified residue; — denotes gap introduced to maximise homology.

but in factor D is replaced by Glu. The Asp-194 residue which forms an ion pair with Ile-16 is present as are the active site residues His-57, Asp-102 and Ser-195 which form part of the charge relay system. Of the residues which are highly conserved in serine proteases many were also found in factor D. The specificity residue at 189 is Asp, as is expected for an arginine-specific serine protease. An interesting substitution is found at residue 192 which in other serine proteases is deeply buried in the zymogen form, and which, on activation, shifts to the surface of the enzyme and forms a flexible 'lid' to the binding pocket [13]. Serine proteases with an Arg/Lys specificity tend to have Glu in this position (although factor B and thrombin are exceptions) whereas Lys was found in factor D.

Previously [3] the similarity of the N-terminal 20 amino acids of factor D to the equivalent regions in rat group specific protease and elastase was noted. When similar comparisons, using the statistical method in [14] were made for the regions of the active site serine (chymotrypsinogen residues 180–201), and histidine (chymotrypsinogen residues 42–58) the probability of a random relationship between factor D and other sequences was $<<0.01\%$. The statistic for the region of the active site aspartic acid (chymotrypsinogen residues 102–112) was more variable, with the probability of a random relationship ranging from 0.01% (groups specific protease), 0.1% (elastase) to 1% (most other serine esterases) except for factor B (5%), thrombin (10%) and factor X (10–50%).

The apparent similarity of factor D to serine proteases of a non-plasma origin is further emphasized by the positioning of its 8 half-cystine residues which appear to occur in positions analogous to those of elastase (fig.2). When aligned with various serine esterases, factor D suprisingly shows greater overall homology with group-specific protease and elastase giving 36% and 32% homology, respectively (fig.2), while comparison with the B chain of the complement enzyme C1r [15] yielded 27% homology, and comparison with the Bb portion of the complement enzyme factor B yielded only about 18% homology [16].

REFERENCES

[1] Fearon, D.T., Austen, K.F. and Ruddy, S. (1974) J. Exp. Med. 139, 355–366.

[2] Davis, A.E. (1980) Proc. Natl. Acad. Sci. USA 77, 4938-4942.
[3] Johnson, D.M.A., Gagnon, J. and Reid, K.B.M. (1980) Biochem. J. 187, 863-874.
[4] Volanakis, J.E., Bhown, A.S., Bennett, J.C. and Mole, J.E. (1980) Proc. Natl. Acad. Sci. USA 77, 1116-1119.
[5] Woodbury, R.G., Katunuma, N., Kobayashi, K., Titani, K. and Neurath, H. (1978) Biochemistry 17, 811-819.
[6] Dayhoff, M.O. (ed) (1978) Atlas of Protein Sequence and Structure, vol. 5, Suppl. 3, National Biomedical Research Foundations, Silver Springs.
[7] Reid, K.B.M. and Porter, R.R. (1981) Annu. Rev. Biochem. 50, 433-464.
[8] Reid, K.B.M., Johnson, D.M.A., Gagnon, J. and Prohaska, R. (1981) Methods. Enzymol. 80, 134-143.
[9] Mahoney, W.C. and Hermodson, M.A. (1979) Biochemistry 18, 3810-3819.
[10] Drapeau, G.R., Boily, Y. and Houmard, J. (1972) J. Biol. Chem. 247, 6720-6728.
[11] Hayashi, R. (1977) Methods Enzymol. 47, 84-88.
[12] Christie, D. and Gagnon, J. (1982) Biochem. J. 201, 555-567.
[13] Krieger, M., Kay, L.M. and Stroud, R.M. (1974) J. Mol. Biol. 83, 209-220.
[14] Moore, G.W. and Goodman, M. (1977) J. Mol. Evol. 9, 121-130.
[15] Arlaud, G.J. and Gagnon, J. (1983) Biochemistry 22, 1758-1764.
[16] Christie, D.L. and Gagnon, J. (1983) Biochem. J. 209, 61-70.