Amino acid sequence of a polymorphic segment from fragment C4d of human complement component C4

D.N. Chakravarti, R.D. Campbell and J. Gagnon*

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

Received 4 February 1983

The amino acid sequence of a segment of 106 residues of C4d has been determined by automated sequence analysis of fragments obtained by CNBr cleavage and enzymic digestion with trypsin. Polymorphism has been detected at 3 positions. Residues 9 and 12 are either valine and leucine or alanine and arginine, respectively. Residue 102 is either valine or arginine. When comparing the protein sequence with the nucleic, acid sequence [Carroll, M. and Porter, R.R. (1983) Proc. Natl. Acad. Sci. USA 80, in press] alanine or serine are found at position 98. These results may in part help to explain the inherited variants of human C4 seen on gel electrophoresis.

Human C4d *Sequence* *Polymorphism*

1. INTRODUCTION

In man, complement component C4 is a highly polymorphic glycoprotein with a M_r of about 200 000 [2]. It consists of 3 polypeptide chains linked by disulphide bonds, the α-, β- and γ-chains, with M_r-values of about 93 000, 75 000 and 33 000, respectively [3,4]. The inherited variability of C4 has been studied by electrophoresis on agarose gels followed by immunofixation [5,6]. The various haplotypes seen, have been grouped under fast moving and slow moving components now called A and B. Population genetic studies have suggested that the two loci are closely linked to each other and are located between the HLA-B and HLA-D loci on human chromosome 6 [7]. Two-dimensional gel electrophoresis has shown that the electrophoretic polymorphism in C4 is due to a charge difference in the α-chain [8]. These differences are also reflected in the C4d fragment which comes from the central portion of the α-chain [13].

Red blood cell antigens Chido and Rogers, which are known to be related to the C4 molecule [5] have been shown to be carried on the C4d fragment of the α-chain [9,10]. C4A and C4B molecules react, respectively, with anti-Rogers and anti-Chido sera. More recently, it has been demonstrated [11] that polymorphism in the α-chain of C4 is not only related to charge, but is also due to a difference in M_r. The α-chain of C4A has a M_r of 96 000 and is anti-Rg positive while the α-chain of C4B has a M_r of 93 000 and is anti-Ch positive.

Here we present evidence that polymorphism in C4 is in part due to differences in the amino acid sequence of fragment C4d.

2. MATERIALS AND METHODS

2.1. *Isolation of C4 and the fragment C4d*

Human C4 was prepared by sequential chromatography of the pseudo-globulin fraction of pooled human serum through DEAE-Sephadex A-50, DEAE-Sepharose CL-6B and Sephacryl S-200 (superfine grade) as in [4,12]. The preparation of C4b from C4, and of C4d from C4b was carried out as in [13], with the exception that factor I was prepared from human plasma by

* To whom all correspondence should be addressed

Published by Elsevier Science Publishers
00145793/83/0000–0000/$3.00

monoclonal-antibody affinity chromatography [14].

2.2. *Purification of CNBr and tryptic fragments of C4d*

CNBr fragments of C4d were prepared by incubating the reduced and alkylated protein with CNBr in 70% (v/v) formic acid in the dark for 24 h at 4°C. The CNBr peptides were purified from the digest by gel filtration on Sephadex G-50 (superfine) as described in [15].

Further purification of the peptides was achieved by high pressure liquid chromatography (HPLC) using a μ-Bondapak C_{18} column and one of two different solvent systems as in [16]. In system A, 0.1% trifluoroacetic acid and a solvent mixture of acetonitrile–methanol–propan-2-ol (1:1:1, by vol.) was used, and in system B 0.1% NH_4HCO_3 and acetonitrile.

Trypsin digestions of reduced, alkylated and succinylated C4d, and of CNBr peptide CN-3 were carried out using conditions previously reported [14]. The succinylated tryptic peptides were purified by gel filtration on Sephadex G-50 (superfine) and HPLC using system B. In one instance (peptide ST3/F/DE-C) ion-exchange chromatography using a column of DEAE-Sephacel as described in [16] was also performed. The trypsin digest of peptide CN-3 was directly separated by HPLC using system B.

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 Beckman 890c sequencer as described in [16]. The phenylthiohydantoin derivatives were identified and quantitated by HPLC [16].

3. RESULTS AND DISCUSSION

Previous work has established the amino acid sequence of 90 residues at the N-terminus of C4d [15]. Polymorphism was not detected in that sequence. The sequence of another 106 residues from within the C4d fragment has now been determined by automated sequence analysis of fragments obtained by CNBr cleavage and enzymic digestion with trypsin. This segment contains a number of polymorphic positions and represents the first evidence at the protein sequence level of the well-established polymorphism of C4.

CNBr digestion of C4d followed by gel filtration on Sephadex G-50 produced 5 major pools (see fig. 2 in [15]). Subsequent analysis of pool D by HPLC using system B gave 3 major fragments, CN-5, CN-6a and CN-6b. CN-5 was previously sequenced and shown to belong to the N-terminal region of C4d [15]. Cn-6a and CN-6b, 34 and 27 amino acids long, respectively, were subjected to automated se-

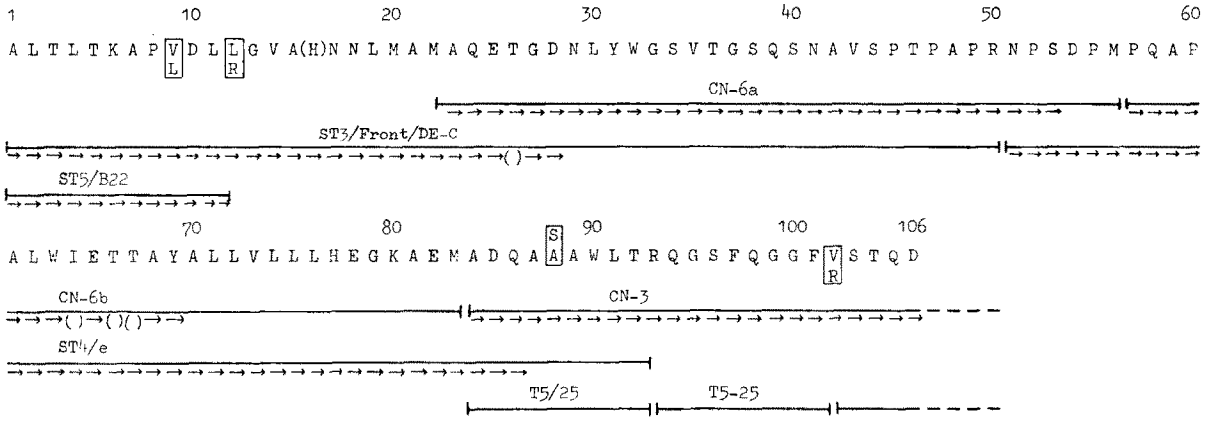

Fig. 1. Automated sequence analysis and alignment of CNBr peptides CN-3, CN-6a and CN-6b, and the succinylated tryptic peptides ST3/F/DE-C, ST5/B22 and ST4/e of C4d. Details of the automated Edman degradation and identification of the PTH derivatives are given in [16]. (→) denotes that positive identification of a particular residue was made. Amino acid residues differing between the peptides are boxed in.

quence analysis. Thirty residues of CN-6a and 11 residues of CN-6b were determined (fig. 1). A third CNBr peptide CN-3 was purified from pool A by HPLC using system A. CN-3, assigned as the C-terminal peptide due to the lack of Hse, is composed of about 65 residues; 23 residues were identified by sequence analysis (fig. 1).

The alignment of the 3 CNBr peptides and the extension of the sequence was made possible by the isolation of two peptides, ST4/e and ST3/F/DE-C, after trypsin cleavage of succinylated C4d. Peptide ST4/e contains 43 amino acid residues. The first 36 residues were determined and were found to overlap CN-6a, CN-6b and CN-3, aligning them in that order (fig. 1). The second peptide, ST3/F/DE-C was analysed for 28 cycles and gave an overlap of 6 residues with CN-6a, providing an N-terminal extension to the 3 contiguous CNBr fragments (fig. 1).

In total this segment represents a continuous sequence of 106 residues. The cDNA sequence corresponding to this region of C4d has been determined [1]. Both sequences are identical, with the exception of position 88. The cDNA codes for a Ser at this position, whereas Ala was determined by amino acid sequence analysis. This discrepancy is likely to be due to the polymorphism of C4.

A better documented instance of polymorphism was obtained from another succinylated tryptic peptide, ST5/B22. This peptide of 12 residues was sequenced in its entirety (fig. 1). It is homologous to the N-terminal sequence of the segment except for positions 9 and 12 where Ala and Arg were found in contrast to the Val and Leu residues, respectively, found in ST3/F/DE-C. Comparison of the amino acid sequence with that derived from the cDNA sequence [1] suggests that the tryptic peptides ST3/F/DE-C and ST5/B22 are derived from a chymotrypsin-like cleavage at the carboxyl side of a Tyr residue prior to Ala at position 1.

Another polymorphic location in this 106-residue segment of C4d occurs at residue 102. At this position a Val residue was determined by sequence analysis of CN-3 (fig. 1). Sequence analysis of a pool from a HPLC separation of a tryptic digest of CN-3 gave an unambiguous although double sequence, related to the N-terminus of CN-3, namely Ala-Asp-Gln-Ala-Ala-Ala-Trp-Leu-Thr-Arg (positions 84 to 93) and Gln-Gly-Ser-Phe-Gln-Gly-Gly-Phe-Thr-Arg (positions 94 to 102). This result indicated that both Val and Arg could be found at position 102 (fig. 1). A tryptic peptide starting with the sequence Ser-Thr-Gln-Asp confirmed this.

In summary, polymorphism at 4 positions within this 106 residue segment of C4d have been detected. These are Val/Ala at position 9, Leu/Arg at position 12, Ser/Ala at position 88 and finally Val/Arg at position 102. As suggested from the cDNA sequence, Val-9, Leu-12 and Ser-88 are associated with one form of C4d. This is consistent with peptide ST3/F/DE-C containing Val-9 and Leu-12. At present it is not possible to decide whether the Val-9, Leu-12, Ser-88 form of C4d has either Val or Arg at position 102.

Since the amino acid sequence described is that of a protein isolated from pooled serum, it cannot be related directly to a specific form of C4. Recently, fragments resembling C4d (Tyr-C4d) have been isolated from trypsin-digested C4 [17]. The major fragments of M_r 30 000 and 28 000 were found to carry the Rogers and Chido determinants, respectively. Sequence analysis showed that the difference was at the C-terminus of the fragments. It is possible that the Leu/Arg polymorphism at position 12 of this study contributed to the different sizes of these Tyr-C4d fragments and suggests that the Val-9, Leu-12, Ser-88 form of C4d might be associated with the larger Rogers (C4A) variant, and conversely the Ala-9, Arg-12, Ala-88 form with the smaller Chido (C4B) variant. Further work at the genomic level will undoubtedly provide the answer.

ACKNOWLEDGEMENTS

We wish to thank Professor R.R. Porter for his continued advice and encouragement during the course of this work. We thank Mr T. Gascoyne for performing the amino acid analysis, Mr A. Willis for sequence analysis and Miss N. Janjua for excellent technical assistance. D.N.C. held a State Scholarship from the Government of West Bengal, India.

REFERENCES

[1] Carroll, M. and Porter, R.R. (1983) Proc. Natl. Acad. Sci. USA, in press.
[2] Rosenfeld, S.I., Ruddy, S. and Austen, K.F. (1969) J. Clin. Invest. 48, 2283–2292.
[3] Schrieber, R.D. and Müller-Eberhard, H.J. (1974) J. Exp. Med. 140, 1324–1335.
[4] Gigli, I., von Zabern, I. and Porter, R.R. (1977) Biochem. J. 165, 439–446.
[5] O'Neill, G.J., Yang, S.Y., Tegoli, J., Berger, R. and Dupont, B. (1978) Nature 273, 668–670.
[6] Awdeh, Z.L. and Alper, C.A. (1980) Proc. Natl. Acad. Sci. USA 77, 8576–8580.
[7] Weitkamp, L.R. and Lamm, L.U. (1982) Cytogenet. Cell Genet. 32, 130–143.
[8] Mevag, B., Olaisen, B. and Teisberg, P. (1981) Scand. J. Immunol. 14, 303–307.
[9] Tilley, C.A., Romans, D.G. and Crookston, M.C. (1978) Nature 276, 713–715.
[10] Chu, V.F.H., Marsh, W.L. and Gigli, I. (1982) J. Immunol. 128, 181–185.
[11] Roos, M.H., Mollenhauer, E., Demant, P. and Rittner, C. (1982) Nature 298, 854–856.
[12] Campbell, R.D., Dodds, A.W. and Porter, R.R. (1980) Biochem. J. 189, 67–80.
[13] Press, E.M. and Gagnon, J. (1981) Biochem. J. 199, 351–357.
[14] Hsiung, L.-M., Barclay, A.N., Brandon, M.R., Sim, E. and Porter, R.R. (1982) Biochem. J. 203, 293–298.
[15] Campbell, R.D., Gagnon, J. and Porter, R.R. (1981) Biochem. J. 199, 359–370.
[16] Christie, D. and Gagnon, J. (1982) Biochem. J. 201, 555–567.
[17] Lundwall, A., Hellman, U., Eggertsen, G. and Sjöquist, J. (1982) Molec. Immunol. 19, 1655–1665.