Volume 222, number 1, 129–134 FEB 05134 September 1987

Identification of *erythro*-β-hydroxyasparagine in the EGF-like domain of human C1r

Gérard J. Arlaud, Alain Van Dorsselaer+, Alex Bell*, Michael Mancini*, Catherine Aude and Jean Gagnon*

Département de Recherches Fondamentales, Laboratoire d'Immunochimie, Unité INSERM 238, CEN-Grenoble 85X, 38041 Grenoble Cédex, +Centre de Neurochimie, 5 rue Blaise Pascal, 67084 Strasbourg Cédex, France and **Institut de Recherches en Biotechnologie, Montréal, Québec H4P 2R2, Canada*

Received 3 August 1987

Previous studies [(1987) Biochem. J. 241, 711–720] have shown that position 150 of human C1r is occupied by a modified amino acid that, after acid hydrolysis, yields *erythro*-β-hydroxyaspartic acid. In view of further investigations on the nature of this residue, peptide CN1a T8/T9 TL8 (positions 147–155) was isolated from $C\bar{1}r$ A chain by CNBr cleavage followed by enzymatic cleavages by trypsin and thermolysin. Amino acid analysis, sequential Edman degradation and FAB-MS of this peptide indicate that the residue at position 150 is an *erythro*-β-hydroxyasparagine resulting from post-translational hydroxylation of asparagine.

β-Hydroxyasparagine; Post-translational modification; EGF-like domain; C1r; Complement system

1. INTRODUCTION

The first component of the classical pathway of complement, C1, is a complex protease comprising a recognition unit, C1q, and C1s-C1r-C1r-C1s, a Ca^{2+}-dependent tetrameric association of two serine proteases, C1r and C1s, which are sequentially activated through limited proteolysis during C1 activation [1,2]. The complete amino acid sequence of C1r, the protease responsible for C1 self-activation, has been determined by protein [3,4] and cDNA sequencing [5,6]. The proenzyme is a single-chain glycoprotein (688 residues), which upon activation is split, through cleavage of a single Arg-Ile bond [7] into two disulphide-linked chains, A (N-terminal) and B (C-terminal). The B chain (242 residues) contains the active site serine, whereas the A chain (446 residues) is divided into five structural units [4,5,8], including two pairs of internal repeats and one EGF-like domain. Recent studies have shown that residue 150 of $C\bar{1}r$ A chain, which is located in the EGF-like domain, is coded as asparagine [5,6] whereas, after acid hydrolysis of the corresponding peptides, it yields *erythro*-β-hydroxyaspartic acid [4]. With a view to further investigations on the nature of this amino acid, a nonapeptide (positions 147–155) was isolated from $C\bar{1}r$ A chain and analyzed by sequential Edman degradation and FAB-MS.

Correspondence address: G.J. Arlaud, Département de Recherches Fondamentales, Laboratoire d'Immunochimie, Unité INSERM 238, CEN-Grenoble 85X, 38041 Grenoble Cédex, France

Abbreviations: EGF, epidermal growth factor; FAB, fast atom bombardment; DADI-MIKES, direct analysis of daughter-ion mass analyzed ion kinetic energy spectrum; CAD-DADI-MIKES, collision-activated DADI-MIKES

2. MATERIALS AND METHODS

2.1. *Materials*

Iodo-[2-^{3}H]acetic acid (106 mCi/mmol) was from Amersham. Trypsin (treated with 1-chloro-4-phenyl-3-L-tosyl amidobutan-2-one) and ther-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

molysin (protease type X) were obtained from Worthington and Sigma, respectively. *erythro-β*-Hydroxyaspartic acid was made starting from fumaric acid according to Okai et al. [9], and further derivatized to *erythro-β*-hydroxyasparagine following the procedure described by Singerman and Liwschitz [10].

2.2. *Tryptic cleavage of peptide CN1a*

Purification of $C\bar{1}r$ from pooled human plasma, reduction of $C\bar{1}r$, alkylation by iodo-[2-^{3}H]acetic acid and separation of $C\bar{1}r$ A and B chains by high-pressure gel-permeation chromatography were performed as described in [11,12]. Reduced and *S*-[^{3}H]carboxymethylated $C\bar{1}r$ A chain was cleaved by CNBr and the CNBr-cleavage peptide CN1a (positions 113–351) was purified as in [4]. CN1a (200 nmol) was incubated in 0.1 M NH_4HCO_3 with trypsin (75 μg) for 2 h at 37°C, and the incubation was repeated after addition of another 75 μg of enzyme. Peptides CN1aT8 and CN1aT9 (positions 135–169) were purified by successive fractionation of the tryptic digest by reversed-phase HPLC on a μBondapak C18 column using solvent systems 1 and 2 described in [13].

2.3. *Cleavage of tryptic peptides CN1aT8 and CN1aT9 by thermolysin*

Peptides CN1aT8 (38 nmol) and CN1aT9 (15 nmol) were incubated in 0.1 M NH_4HCO_3 with thermolysin (3 and 1.2 μg, respectively) for 2 h at 37°C, and each incubation was repeated after addition of the same amounts of thermolysin. Each digest was fractionated by reversed-phase HPLC on a μBondapak C18 column using solvent system 1 described in [13].

2.4. *Amino acid analysis and sequence determination*

Amino acid analyses were performed on acid hydrolysates with a Beckman 7300 amino acid analyser using ninhydrin for detection. Automated Edman degradation of the peptides was performed as in [4].

2.5. *Mass spectrometry*

Spectra were obtained on a VG Analytical ZAB-SE double-focusing mass spectrometer. The accelerating voltage was 10 kV and the fast xenon atom beam was operated with an emission current of 0.1 mA at 8 kV. Mass spectra were recorded with the acquisition data system, and calibration was performed with CsI. Peptide CN1aT8/T9 TL8 (10 nmol) was dissolved in 10 μl of 5% (v/v) acetic acid and 1 μl of this solution was added to the thioglycerol matrix. DADI-MIKES [14] was performed on the pseudo-molecular ion with the acquisition data system from 1000 eV to 8000 eV. A precision of 0.5 amu was generally obtained for the mass of the daughter ion. CAD-DADI-MIKES was obtained by addition of helium to the collision chamber of the second field free zone. The stream of helium was adjusted in order to divide the intensity of the molecular ion by a factor of three.

3. RESULTS

3.1. *Isolation of peptides CN1aT8 and CN1aT9*

Previous sequence studies [4] have shown that two different amino acid residues, serine and leucine, occur at position 135 of human $C\bar{1}r$. As expected from this heterogeneity, full tryptic cleavage of peptide CN1a (positions 113–351) yielded two different peptides covering region 135–169, both arising from cleavage of the lysyl bond at position 134 and of the Arg-His bond at position 169. These were separated by reversed-phase HPLC and identified by amino acid analysis and a short N-terminal sequence analysis. The N-terminal sequence of the major peptide, CN1aT8, was Ser-Gly-Glu-Glu-Asp-..., whereas that of CN1aT9 was Leu-Gly-Glu-Glu-Asp-.

3.2. *Isolation, amino acid analysis and sequence of peptide CN1aT8/T9 TL8*

Digestion of peptides CN1aT8 and CN1aT9 by thermolysin led to partial or complete cleavage of peptide bonds at position 146 (His-Leu), 151 (Tyr-Val), 155 (Tyr-Phe) and 164 (Glu-Leu) (see fig.2), in agreement with the known specificity of thermolysin [15]. Peptides generated from these cleavages were separated by reversed-phase HPLC (fig.1) and identified by amino acid analysis. In both cases of CN1aT8 and CN1aT9, cleavage by thermolysin yielded peptide CN1aT8/T9 TL8, resulting from cleavage of peptide bonds at positions 146 and 155 (figs 1 and 2). Amino acid analysis of an acid hydrolysate of this peptide indicated the presence of *erythro-β*-hydroxyaspartic acid (table 1), estimated as 0.9 mol/mol of pep-

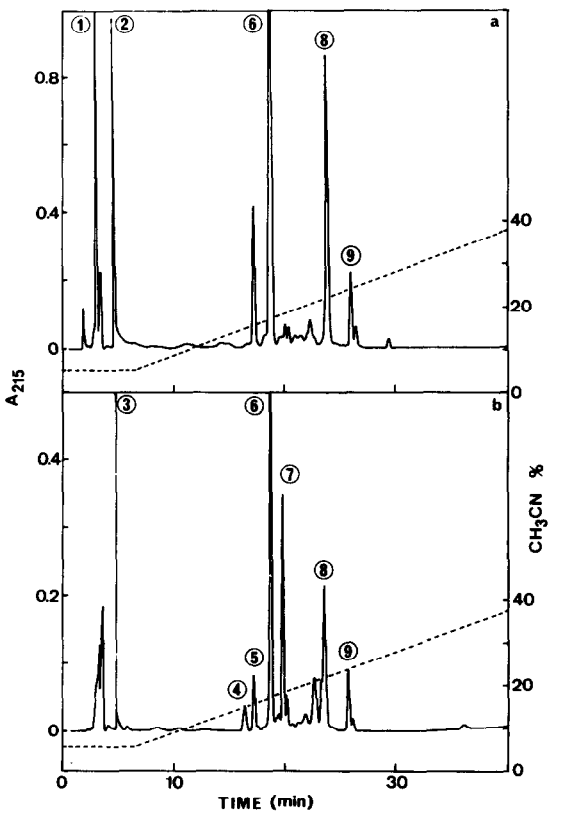

Fig.1. Separation by reversed-phase HPLC of peptides from the digests of CN1aT8 (a) and CN1aT9 (b) by thermolysin. Each digest was dissolved in 0.1% (v/v) NH_4HCO_3 and separated on a μBondapak C18 column using solvent system 1 as described in [11].

tide, thus confirming previous analyses of peptides αST3b (positions 133–169) and CN1aSP4 (positions 129–164) [4]. N-terminal sequence analysis was performed on 1.0 nmol of peptide CN1aT8/T9 TL8 and the sequence Leu-Cmc-His-(OH-Asn)-Tyr-Val-Gly- was obtained. The tentative assignment of the β-hydroxyasparagine was made on the basis of comparable retention times on HPLC of the phenylthiohydantoin-derivative with that of a synthetic standard.

Cleavage of CN1aT9 by thermolysin yielded a minor peptide, CN1aT9 TL4, resulting from partial cleavage of the Tyr-Val bond at position 151 (figs 1 and 2). Amino acid analysis indicated the absence of *erythro*-β-hydroxyaspartic acid, and the presence of 2 Asx residues (table 1). This result was confirmed by N-terminal sequence analysis of

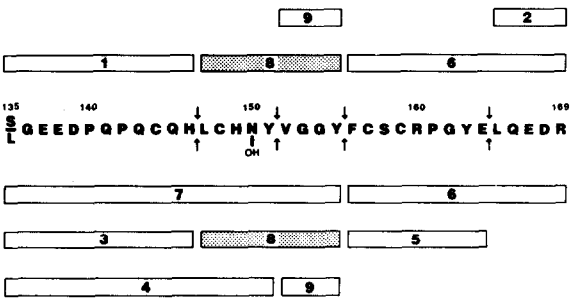

Fig.2. Diagrammatic representation of the major fragments generated by digestion of peptides CN1aT8 (upper part) and CN1aT9 (lower part) by thermolysin. Fragments are numbered according to their elution position on HPLC (see fig.1). Numbering of amino acid residues refers to their position in the sequence of C1r.

CN1aT9 TL4, which allowed unambiguous identification of amino acid residue 150 as asparagine.

3.3. *Analysis of CN1aT8/T9 TL8 by mass spectrometry*

FAB-MS is now a well established method for molecular mass measurement of peptides. The FAB-mass spectrum of peptide CN1aT8/T9 TL8 showed a large peak at m/z = 1099.4 (fig.3). Such a peak is usually interpreted as a pseudo-molecular peak corresponding to a protonated (MH^+) or cationised (MNa^+) form of the peptide. It could also be interpreted as a molecular ion if the peptide is already ionised [16]. As shown in table 2, these different hypotheses lead to the following molecular masses for peptide CN1aT8/T9 TL8: 1098.4, 1076.4 and 1099.4 Da. The eight amino acids unambiguously identified from Edman degradation and amino acid analysis of peptide CN1aT8/T9 TL8 correspond to a total mass of 968.4 amu. The three possible mass values for amino acid at position 150 are given in table 2. Among these values, the only one which is consistent with the amino acid analysis is 130.0 (hydroxyasparagine), which corresponds to a mass of 1098.4 for the whole peptide. A hydroxyaspartic acid residue would give a pseudo-molecular ion of m/z = 1100.4 (MH^+). Sequence analysis was performed by DADI-MIKES on the MH^+ ion of peptide CN1aT8/T9 TL8 (fig.4). Sequences from both the N- and C-terminal ends could be interpreted from the spectrum: an N-terminal sequence Leu-

Table 1

Amino acid compositions of selected peptides covering position 150 of human C1r[a]

Amino acid	Residues/mol		
Peptide:	CN1aT8	CN1aT8/T9 TL8	CN1aT9 TL4
Position in sequence:	135–169	147–155	135–151
β-OH-Asx	ND[c] (1)	0.9 (1)	–
Asx	2.1 (2)	–	2.2 (2)
Thr	–	–	–
Ser	1.6 (2)	–	–
Glx	7.8 (8)	–	4.4 (5)
Pro	2.5 (3)	–	ND[c] (2)
Gly	3.7 (4)	2.3 (2)	1.0 (1)
Ala	–	–	–
Val	1.0 (1)	1.0 (1)	–
1/2Cys[b]	4.6[d] (4)	0.8 (1)	2.2 (2)
Ile	–	–	–
Leu	2.0 (2)	1.0 (1)	2.1 (2)
Tyr	2.9 (3)	1.8 (2)	1.0 (1)
Phe	1.0 (1)	–	–
His	2.0 (2)	0.8 (1)	2.0 (2)
Lys	–	–	–
Arg	2.0 (2)	–	–

[a] Figures in brackets are obtained from sequence
[b] Estimated as *S*-(carboxymethyl)cysteine
[c] Not determined
[d] The overestimate is due to β-OH-Asx, not resolved from *S*-(carboxymethyl)cysteine

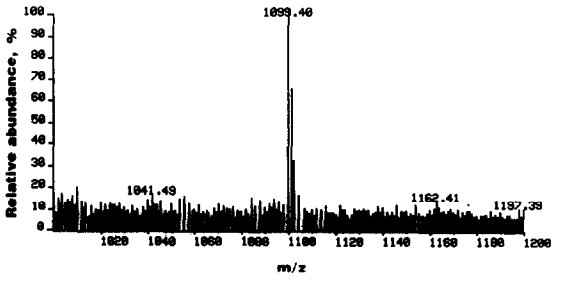

Fig.3. FAB-mass spectrum of peptide CN1aT8/T9 TL8.

Cmc-, and a C-terminal sequence -Val-Gly-Gly-Tyr were clearly visible. On enlarged spectra with CAD, the loss of a His residue from an N-terminal ion (m/z = 688) and the loss of a Tyr residue from a C-terminal ion (m/z = 542) were visible. The sequence which can be deduced from the DADI-MIKES studies is therefore:

Leu-Cmc-[His, OH-Asn, Tyr]-Val-Gly-Gly-Tyr

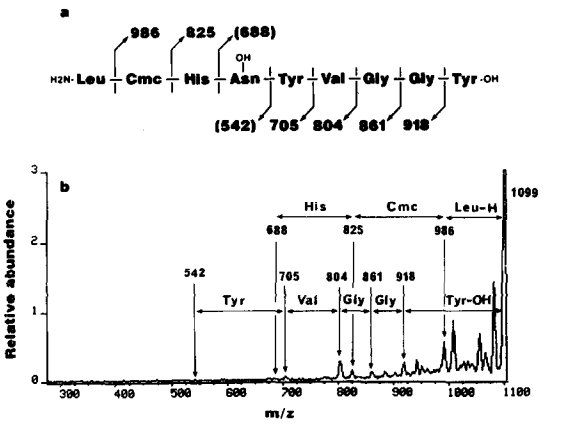

Fig.4. DADI-MIKES analysis of the pseudo-molecular ion (MH^+) of peptide CN1aT8/T9 TL8. (a) Diagrammatic representation of the fragment ions observed in the analysis. (b) DADI-MIKES analysis. Peaks at m/z = 688 and m/z = 542 were only visible on enlarged spectra with CAD.

Table 2

Different hypotheses for the calculation of the molecular mass of amino acid residue at position 150 of C1r

	Type of ion 1099.4		
	M+H	M+Na	M
Peptide mass (*M*)	1098.4	1076.4	1099.4
Mass of unidentified amino acid (*M* − 968.4)	130.0	108.0	131.1

4. DISCUSSION

Amino acid analysis and Edman degradation of peptide CN1aT8/T9 TL8 led to the tentative identification of residue 150 of C1r as *erythro-β*-hydroxyasparagine, in agreement with previous analysis of larger peptides covering this position [4]. Analysis by FAB-MS of peptide CN1aT8/T9 TL8 rules out the hypothesis of a hydroxyaspartic acid and is compatible with the presence of a hydroxyasparagine. From the combination of these data, it can therefore be concluded that residue 150 of human C1r is an *erythro-β*-hydroxyasparagine. This conclusion is consistent with: (i) analysis of the minor peptide CN1aT9 TL4, which was found to contain asparagine at position 150; this partial lack of hydroxylation of the asparagine could be the result either of an incomplete hydroxylation of the asparagine, or of a slight chemical de-hydroxylation during peptide handling; (ii) the finding that residue 150 does not react with dicyclohexyl carbodiimide or 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, two reactants known to be specific for acidic amino acids (unpublished); (iii) published sequence analyses of C1r cDNA [5,6], which show that residue 150 is coded as asparagine.

It can therefore be concluded that the *erythro-β*-hydroxyasparagine at position 150 of human C1r results from post-translational hydroxylation of an asparagine. So far, the only other reported case of a protein containing hydroxylated asparagine is bovine protein S [17]. This vitamin K-dependent protein contains 3 *β*-hydroxyasparagine residues which, like residues 150 of human C1r [4], are all located in sequences homologous to the EGF precursor. Several cases of *β*-hydroxyaspartic acid residues have also been described in a number of other vitamin K-dependent proteins, including protein C [18,19], factors VII, IX and X [20,23] and again, these residues all occur in EGF-like sequences. A consensus sequence representing the structural requirements of the enzyme(s) responsible for the hydroxylation of these aspartic acid or asparagine residues has recently been proposed by Stenflo et al. [17]. It should be mentioned that the amino acid sequence surrounding residue 150 of human C1r is compatible with the proposed consensus.

This work confirms the existence of a serine/leucine polymorphism at position 135 of C1r, as previously shown by protein [4] and cDNA sequence studies [5,6]. As this represents the only sequence heterogeneity found in whole human C1r, it provides a structural basis for the existence of the two common C1r alleles detected in human plasma by isoelectric focusing [24]. As C1r molecules containing either serine or leucine at position 135 are both hydroxylated at position 150, it can be concluded that residue 135 is not essential for the specific recognition of the EGF-like domain of human C1r by the hydroxylating enzyme.

ACKNOWLEDGEMENT

We thank VG Analytical for the use of mass spectrometry facilities.

REFERENCES

[1] Cooper, N.R. (1985) Adv. Immunol. 37, 151–216.
[2] Sim, R.B. (1981) Methods Enzymol. 80, 6–16.
[3] Arlaud, G.J. and Gagnon, J. (1983) Biochemistry 22, 1758–1764.
[4] Arlaud, G.J., Willis, A.C. and Gagnon, J. (1987) Biochem. J. 241, 711–720.
[5] Leytus, S.P., Kurachi, K., Sakariassen, K.S. and Davie, E.W. (1986) Biochemistry 25, 4855–4863.
[6] Journet, A. and Tosi, M. (1986) Biochem. J. 240, 783–787.
[7] Arlaud, G.J. and Gagnon, J. (1985) FEBS Lett. 180, 234–238.
[8] Arlaud, G.J., Colomb, M.G. and Gagnon, J. (1987) Immunol. Today 8, 106–111.
[9] Okai, A., Imamura, N. and Izumiya, N. (1967) Bull. Chem. Soc. Japan 40, 2154–2159.
[10] Singerman, A. and Liwschitz, Y. (1968) Tetrahedron Lett. 46, 4733–4734.

[11] Arlaud, G.J., Sim, R.B., Duplaa, A.-M. and Colomb, M.G. (1979) Mol. Immunol. 16, 445–450.
[12] Arlaud, G.J., Gagnon, J. and Porter, R.R. (1982) Biochem. J. 201, 49–59.
[13] Gagnon, J. and Arlaud, G.J. (1985) Biochem. J. 225, 135–142.
[14] Heerma, W., Kamerling, J.P., Slotboom, A.J., Van Scharrenburg, G.J.M., Green, B.N. and Lewis, I.A.S. (1983) Biomed. Mass Spectrom. 10, 13–16.
[15] Heinrikson, R.L. (1977) Methods Enzymol. 47, 175–189.
[16] Martinage, A., Briand, G., Van Dorsselaer, A., Turner, C.H. and Sautière, P. (1985) Eur. J. Biochem. 147, 351–359.
[17] Stenflo, J., Lundwall, A. and Dahlbäck, B. (1987) Proc. Natl. Acad. Sci. USA 84, 368–372.
[18] Drakenberg, T., Fernlund, P., Roepstorff, P. and Stenflo, J. (1983) Proc. Natl. Acad. Sci. USA 80, 1802–1806.
[19] Foster, D.C., Yoshitake, S. and Davie, E.W. (1985) Proc. Natl. Acad. Sci. USA 82, 4673–4677.
[20] McMullen, B.A., Fujikawa, K., Kisiel, W., Sasagawa, T., Howald, W.N., Kwa, E.Y. and Weinstein, B. (1983) Biochemistry 22, 2875–2884.
[21] Fernlund, P. and Stenflo, J. (1983) J. Biol. Chem. 258, 12509–12512.
[22] Sugo, T., Fernlund, P. and Stenflo, J. (1984) FEBS Lett. 165, 102–106.
[23] McMullen, B.A., Fujikawa, K. and Kisiel, W. (1983) Biochem. Biophys. Res. Commun. 115, 8–14.
[24] Kamboh, M.I. and Ferrel, R.E. (1986) Am. J. Hum. Genet. 39, 826–831.