Amino acid sequence analysis of the glycopeptides from human complement component C3

Karen G. Welinder and Allan Svendsen

Institute of Biochemical Genetics, University of Copenhagen, Ø. Farimagsgade 2A, DK-1353 Copenhagen K, Denmark

Received 15 April 1986

Human complement component C3 has been cleaved completely by trypsin in the presence of 2-propanol. The hydrolysate was fully solubilized and fractionated by reversed-phase HPLC. Two peptides only contained glucosamine, Unambiguous sequence analyses identified Asn-63 of the β-chain and Asn-268 of the α-chain as the sites of carbohydrate attachment. A third potential Asn-Xaa-Thr/Ser glycosylation site, Asn-946 of the α-chain, is not modified. The different states of glycosylation of the three sites cannot be explained by differences in exposure or secondary structure. All three are predicted reverse turn.

Glycosylation *Amino acid sequence* *Complement component C3* *Trypsin* *2-Propanol*

1. INTRODUCTION

Complement component C3 is a central effector molecule in the complement cascade [1]. Human C3 is a glycoprotein of M_r 187500. The biologically active molecule is composed of two disulfide-linked chains. The β-chain comprises 645 residues and the α-chain 992 residues. The complete cDNA sequence coding for human prepro-C3 was recently determined by De Bruijn and Fey [2]. The molecule is post-translationally modified by removal of the 22-residue signal peptide that precedes the β-chain, and the four arginine residues that connect the β-chain and α-chain. Thirteen disulfide bonds and a functionally important thioester bond are formed (review [2]). Furthermore, two high mannose-type oligosaccharides are attached, one to each chain [3]. The cDNA derived amino acid sequence predicts a maximum of three *N*-glycosylation sites [2], whereas electrophoretic separation of C3 fragments followed by staining for carbohydrate restricts glycosylation sites to a 17 kDa fragment of the β-chain [4] and a 26 kDa fragment of the α-chain [4–6].

Here we show by sequence analysis that Asn-63 of the β-chain and Asn-268 of the α-chain are the only sites of *N*-glycosylation in human complement component C3. We have used a 1:1 mixture of the two common genetic variants C3 S and C3 F in the experiments and have no indication that they differ in this respect. The carbohydrate chains of human C3 play no known physiological role.

2. MATERIALS AND METHODS

Active human complement component C3 S and C3 F in a 1:1 mixture was kindly donated by N. Behrendt. It was purified and characterized as described [7], and stored at −80°C as pellets containing 2.5% C3 in 20 mM KH_2PO_4, 150 mM NaCl, pH 7.4. Digestions of 1% C3 solutions were brought about at 37°C under nitrogen or argon.

Analytical tryptic digestions of C3 samples denatured to different extents were carried out overnight with E:S ratios (enzyme:substrate) of 1:20 mol/mol. Samples in buffer, in 0.2% SDS, in 15% and in 22.5% 2-propanol were preincubated at 60°C for 10 min in one series of experiments and untreated in a parallel series before the addition of trypsin. For preparative purposes 90 nmol C3 was titrated to pH 8.0 with 0.25 M NaOH and preincubated in 15% 2-propanol at 60°C for

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

10 min. The sample became milky. Trypsin (trypsin-TPCK, Worthington) in 10 mM $CaCl_2$ was added twice, at zero time and after 6 h, giving a final E:S = 1:10 mol/mol. The clear but highly viscous digest was centrifuged after 22 h. The gel-like precipitate 'melted' on addition of 50 μl 10% SDS. The supernatant and dissolved gel were fractionated separately by reversed phase-high performance liquid chromatography (RP-HPLC). 3% of each fraction was subjected to amino acid analysis. Peptic digestion of C3 was carried out at E:S = 1:30 in 5% HCOOH for 24 h. Thermolytic digestion of methylamine-inactivated C3 at E:S = 1:40 was stopped after 4 h. The completeness of all digestions was evaluated by SDS-PAGE (polyacrylamide gel electrophoresis) in 15% gels [8].

All peptide separations were accomplished on a 7.8 × 180 mm column of Nucleosil 5 C18 (Macherey-Nagel), except that the SDS solubilized, gel-like, tryptic peptides were fractionated on a 4.6 × 125 mm column of Nucleosil 300-5 C4. Columns were packed in the laboratory. A 2 ml sample loop was used.

Amino acid analyses were performed on HPLC equipment (Waters) with post-column derivatization by *o*-phthalaldehyde in non-halide buffers. Glucosamine appears as a broad peak after phenylalanine. Hydrolysis was accomplished in the vapor of 6 M HCl at 108°C, for 6 h to detect glucosamine, and for 20 h to determine amino acid compositions. Amino acid sequence analyses were carried out on a gas-phase sequencer (model 470A, Applied Biosystems) using the program O2NRUN. PTH-amino acids generated by the sequencer were analysed by HPLC equipment (Waters) with 254 and 313 nm detectors on a 4.6 × 125 mm column of Spherisorb S5-ODS 2 (Phase Sep) [9].

3. RESULTS

3.1. *Tryptic digestion*

SDS-PAGE evaluation of the analytical tryptic digestions of human complement component C3 showed that samples preincubated at 60°C were the most degraded, but that only the preincubated, 2-propanol containing samples were completely digested. In the preparative experiment neither the supernatant nor the gel-like precipitate showed bands of high molecular mass, only diffuse

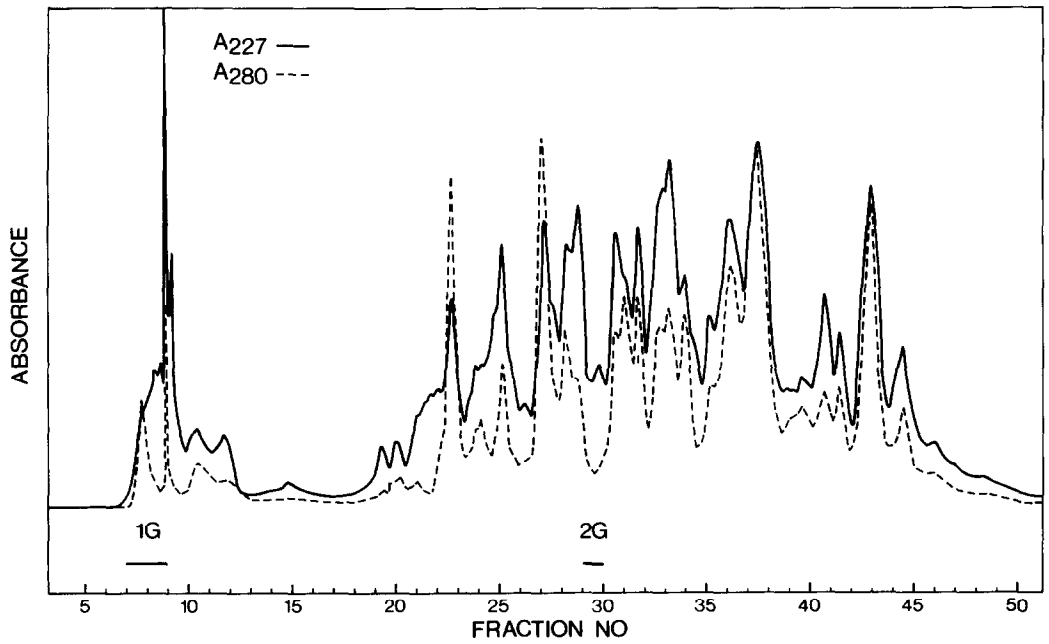

Fig.1. Isolation of tryptic glycopeptides 1G and 2G by RP-HPLC. A 7.8 × 180 mm column of Nucleosil C18 was eluted by a 1 h linear gradient of 5–70% 2-propanol in 0.1% TFA at a flow rate of 1 ml/min. The fraction size was 0.5 ml.

material below 14.3 kDa. The tryptic supernatant (fig.1) and precipitate (not shown) were initially fractionated in 5–70% 2-propanol gradients on C18 and C4 columns, respectively. Complete sequence analyses on selected, rechromatographed fractions (tryptic peptides corresponding to residues 282–289, 291–304, 387–408, 557–566, 567–573, 611–615, 616–622, 723–735, 1052–1060 of prepro-C3 were present in the supernatant, and tryptic peptides 344–359, 409–425, 1073–1084, 1140–1155, 1186–1201 were present in the gel-like precipitate) support that our tryptic digestion was complete and specific and also verify the cDNA derived amino acid sequence [2] of these residues.

3.2. *Glycopeptides 1G and 2G*

Two fractions only, 1G and 2G, both of the supernatant (fig.1), contained glucosamine. Fraction 1G was rechromatographed twice on the same column, firstly in a 5–70% methanol gradient, secondly in a 1–70% methanol gradient which gave pure 1G. Fraction 2G was rechromatographed in a 5–70% ethanol gradient. The amino acid compositions and yields of 1G and 2G are shown in table 1. The high yields suggest that 1G and 2G are equally present in the two genetic variants C3 S and C3 F. Fig.2 shows the unambiguous, complete sequence analyses of the glycopeptides from table 1. It is a general experience that glycosylated residues are removed normally by Edman degradation, but that the glycosylated hydantoin derivative is nearly insoluble in the solvent used to extract the normal hydantoins. Therefore, sequence analysis of a glycopeptide shows a gap at the position of the glycosylated residue.

```
1G          Met-Asx(GlcN)-Lys
1Gth   Ile-Arg-Met-Asx(GlcN)-Lys-Thr

2G     Thr-Val-Leu-Thr-Pro-Ala-Thr-Asn-His-Met-Gly-
       Asx(GlcN)-Val-Thr-Phe-Thr-Ile-Pro-Ala-Asn-Arg
```

Fig.2. Sequence analyses of glycopeptides from human complement component C3. Sequence analyses on 1 nmol aliquots were carried out once (1G), twice (1Gth) and three times (2G) with average repetitive yields of 90–95%. Residues identified unambiguously by sequence analysis are underlined; gaps in the sequences are assigned Asx(GlcN) in agreement with the amino acid compositions shown in table 1.

Table 1

Amino acid composition of glycopeptides

	1G	1Gth	2G
Asp + Asn	1.0 1	1.1 1	2.9 3
Thr		1.0 1	4.9 5
Ser		0.2	0.3
Glu + Gln		0.1	0.8
Pro			2.2 2
Gly		0.1	1.3 1
Ala			2.2 2
Cys			
Val			2.1 2
Met	0.9 1	1.0 1	1.1 1
Ile		0.7 1	1.1 1
Leu			1.5 1
Tyr			0.4
Phe			1.0 1
His			1.0 1
Lys	1.0 1	1.0 1	0.2
Trp			
Arg		0.8 1	1.1 1
GlcN	1.9	1.6	1.2
Total	3	6	21
Yield (%)	53		56

Residues per molecule determined by amino acid and sequence analysis, respectively

3.3. *Thermolytic and peptic digestions*

Initial digestions of C3 with pepsin and with thermolysin were also rather efficient. Both, however, left insoluble material which made it difficult to ascertain the recovery of all glycosylated peptides. Glycopeptides different from the 1G and 2G sequences were not observed. Glycosylations were confirmed by the thermolytic peptide 1Gth (table 1 and fig.2) and by two large glucosamine-containing peptic peptides, one beginning at residue 239 of the α-chain as shown by sequence analysis, the other giving rise to a glycopeptide identical to 2G on further cleavage by trypsin.

4. DISCUSSION

Human complement component C3 is cleaved in vivo to a number of well-characterized fragments which can be mimicked in vitro by incubation with either trypsin, plasmin, elastase or plasma. This suggests that the huge C3 molecule is composed of

a number of stable structural domains which also in our experiments prove to be very resistant to complete proteolytic digestion. Here we show that preincubation at 60°C for 10 min followed by tryptic digestion of C3 in 15% 2-propanol gives complete and specific cleavage judging from SDS-PAGE and sequence analyses.

Analysis of the carbohydrate composition in human complement components showed that only *N*-acetylglucosamine and mannose are present in C3 [10]. Hase et al. [3] have solved the covalent structure of the carbohydrate side chains of human C3 after release by hydrazinolysis. They found that the α-chain contains mainly $Man_8(GlcNAc)_2$ and $Man_9(GlcNAc)_2$, while the β-chain contains mainly $Man_5(GlcNAc)_2$ and $Man_6(GlcNAc)_2$. These compositions were recently confirmed by Hirani et al. [11] by compositional analysis of sugar chains released by endo-β-*N*-acetylglucosamidase H.

We have isolated and sequenced all glucosamine-containing tryptic peptides from human complement component C3. Our results fully confirm the recent cDNA derived amino acid sequence of human C3 [2]. The sequence of the tryptic glycopeptide 2G is identical to residues 74–94 of prepro-C3, and 1G to residues 938–940. Asn-85, Asn-939 and Asn-1617 are potential *N*-glycosylation sites in prepro-C3. We prove that the residues corresponding to Asn-85 and Asn-939 are indeed glycosylated in C3, and we find only these two sites. These results are in agreement with the results of carbohydrate staining of the C3 fragments mentioned in section 1.

The observed pattern of glycosylation in human complement component C3 raises the question why Asn-268 of the α-chain carries more mannose residues than Asn-63 of the β-chain, and why Asn-946 of the α-chain is unmodified. Aubert et al. [12] and Beeley [13] showed that *N*-glycosylated sites are nearly always exposed turns or loops and that most sites are predicted reverse turns. Secondary structure prediction for C3 by the method of Chou and Fasman [14] predicts that all three asparagine residues are part of reverse turns. In the sequence of post translational *N*-glycosylation events (review [15]) attachment of oligosaccharide is followed by sequential removal of the three glucose residues and the four α-1,2 linked mannose residues. Therefore, we may expect that Asn-63 of the β-chain, which is the most processed as it binds 5–6 mannose residues only, is also the most exposed, and that the unprocessed α-chain Asn-946 is the least exposed among the three. Evaluation of the exposure of the three asparagine residues by the hydropathic index [16] averaged over the seven-residue peptide with asparagine centrally suggests the exactly opposite order of exposure and comes to -0.4 for Asn-946 and -0.3 for Asn-268 of the α-chain, and $+0.2$ for Asn-63 of the β-chain. Consequently, the answer might await the appearance of the spatial structure of C3 or might be sought in an unrecognized glycosylation requirement.

ACKNOWLEDGEMENTS

This work was supported by grants 12-5029 and 12-5206 from the Danish Medical Science Research Council. Sequence analysis of peptic and thermolytic peptides were carried out in collaboration with F. Hansen, Rigshospitalet.

REFERENCES

[1] Reid, K.B.M. and Porter, R.R. (1981) Annu. Rev. Biochem. 50, 433–464.
[2] De Bruijn, M.H.L. and Fey, G.H. (1985) Proc. Natl. Acad. Sci. USA 82, 708–712.
[3] Hase, S., Kikuchi, N., Ikenaka, T. and Inoue, K. (1985) J. Biochem. 98, 863–874.
[4] Chiu, F.J. and Atkin, C.L. (1983) J. Biol. Chem. 258, 7200–7207.
[5] Taylor, J.C., Crawford, I.P. and Hugli, T.E. (1977) Biochemistry 16, 3390–3396.
[6] Hirani, S., Lambris, J.D. and Müller-Eberhard, H.J. (1985) J. Immunol. 134, 1105–1109.
[7] Behrendt, N. (1985) Mol. Immunol. 22, 1005–1008.
[8] Laemmli, U.K. (1970) Nature 227, 680–685.
[9] Foltmann, B. and Jensen, A.L. (1982) Eur. J. Biochem. 128, 63–70.
[10] Tomana, M., Niemann, M., Garner, C. and Volanakis, E. (1985) Mol. Immunol. 22, 107–111.
[11] Hirani, S., Lambris, J.D. and Müller-Eberhard, H.J. (1986) Biochem. J. 233, 613–616.
[12] Aubert, J.-P., Biserte, G. and Loucheux-Lefebvre, M.-H. (1976) Arch. Biochem. Biophys. 175, 410–418.
[13] Beeley, J.G. (1977) Biochem. Biophys. 76, 1051–1055.
[14] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.
[15] Hubbard, S.C. and Ivatt, R.J. (1981) Annu. Rev. Biochem. 50, 555–583.
[16] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.