Human free secretory component is composed of the first 585 amino acid residues of the polymeric immunoglobulin receptor

Graham J. Hughes[a], Séverine Frutiger[a], Luc-Alain Savoy[b], Andrew J. Reason[c], Howard R. Morris[d], Jean-Claude Jaton[a,*]

[a] *Dept. of Medical Biochemistry, Faculty of Medicine, University of Geneva, Medical Center, rue Michel-Servet 1, CH-1211 Geneva 4, Switzerland*
[b] *M-Scan SA, 12 chemin des Aulx, CH-1228 Geneva 4, Switzerland*
[c] *M-Scan Ltd, Silwood Park, Sunninghill, Berkshire SL5 7PZ, UK*
[d] *Department of Biochemistry, Imperial College, London SW7 2AZ, UK*

Received 25 April 1997

Abstract The main objective of this work was to unequivocally determine the C-terminal sequence of human milk free secretory component (SC). It was found to end at arginine-585, i.e. 33 amino acids downstream from the major heterogeneous C-terminal residue previously identified for colostrum SC. In contrast, our data showed that the C-terminal end of SC was found to be homogeneous. Conflicting assignments, Asp/Gln, a missing Asn-211, Asp/Asn, Glu/Gln were corrected and found to agree with the cDNA sequence. An Ala/Val substitution at position 562 (domain VI) was identified. Its genetic significance is uncertain at present.

Key words: Secretory component; Amino acid sequence; Polymeric immunoglobulin receptor; Carboxy-terminal region

1. Introduction

Secretory component (SC) is produced by cleavage of the polymeric Ig receptor during the transport of dimeric and polymeric IgAs into the secretions of the gastrointestinal, respiratory and genitourinary tracts. Dimeric and polymeric IgAs are synthesized by plasma cells present in mucosa-associated lymphoid tissues and thus, need to be transported across the epithelia in order to perform their protective role against environmental antigens [1–3]. This is effected by binding of these IgAs to the poly-Ig receptor at the surface facing the mucosal tissues. After internalization, the complex is transported through the cell and inserted in the apical membrane of the epithelium [4]. At that site, a still uncharacterized protease cleaves the poly-Ig receptor and releases SC bound to IgA [5–8]. This complex is known as sIgA. The transcytosis of the poly-Ig receptor can occur independently of ligand binding and thus, free SC is also found in secretions [9].

Since it was discovered that SC was derived from the poly-Ig receptor, efforts have been almost entirely focused on the biosynthesis and the transcytosis of the receptor [4]. Alignments of cDNA deduced amino acid sequences have revealed a high degree of homology between rabbit, human, rat, bovine and mouse poly-Ig receptors [8,10–15]. The extracellular portion is composed of five Ig-like domains of 104–114 amino acids. The sixth, less related domain, contains the membrane spanning segment while the seventh domain is cytoplasmic. Ambiguities about the structure of SC itself, such as the location of the exact extracellular cleavage site in the poly-Ig receptor, its primary sequence, and the type of oligosaccharides present on the molecule, are still unresolved. From carboxypeptidase Y digestion, it has been reported that the C-terminal of rabbit SC was Ala-Glu [16]. By aligning sequences of domain VI from all known poly-Ig receptors, a highly conserved turn region followed by a stretch of 5 to 14 amino acids forming an alpha helix was disclosed [13]. A putative cleavage site was suggested in the middle of domain VI at the carboxyl side of Glu-589 in the human sequence [13]. In contrast, the primary structure of human SC disclosed that the C-terminal was ragged and that the molecule was composed of a maximum of 558 amino acid residues [17,18].

Conflicts in the primary structure of human SC, not involving deamidation of predicted asparagine or glutamine residues, are still to be resolved (SWISS-PROT database accession number: PO1833). The availability of free SC from breast milk obtained from a single individual (M.W.) allowed us to unequivocally identify the C-terminal sequence of free SC, to correct 6 discrepancies between DNA and protein sequencing [11,12,17,18] and to suggest a mutation of Ala to Val at position 562 in the human sequence.

2. Materials and methods

2.1. Materials

Sequencing-grade trypsin (EC 3.4.21.4), endoproteinase Glu-C (EC 3.4.21.19) and peptide-*N*-glycosidase F (EC 3.2.2.18) were purchased from Boehringer-Mannheim (Germany). Solvents for HPLC, dithiothreitol and iodoacetic acid were from Merck, Darmstadt (Germany). Solvents used for sequencing were from Fluka, Buchs (Switzerland) and were further purified as described [19]. Water used was obtained from a Milli-Q water system (Millipore Corp., Bedford, MA). All other chemicals were of the best grade available.

2.2. Purification of human SC

Secretory component was purified from breast milk obtained from a single individual by a combination of gel filtration and ion-exchange chromatography as previously described [20].

2.3. Peptide mapping of SC tryptic and endoproteinase Glu-C peptides

SC (1 mg) was fully reduced with a five-fold excess of DTT over cysteine content for 2 h at room temperature in 75 mM Tris-HCl, 6 M guanidine-HCl, pH 8.6. After reaction with a two-fold molar excess of iodoacetic acid (over DTT) for 30 min in the dark, the protein was dialyzed extensively against 0.5% ammonium bicarbonate and concentrated by vacuum centrifugation. Aliquots (0.2–0.4 mg) were incu-

*Corresponding author. Fax: (41-22) 702 55 02.

Abbreviations: DTT, dithiothreitol; ES-MS, electrospray mass spectrometry; FAB-MS, fast atom bombardment mass spectrometry; PNGase F, peptide-*N*-glycosidase F; poly-Ig receptor, polymeric Ig receptor; RFLP, restriction fragment length polymorphism; SC, secretory component; sIgA, secretory immunoglobulin A; TFA, trifluoroacetic acid; TPCK, tosyl-L-phenylalanine chloromethyl ketone

PII S0014-5793(97)00629-7

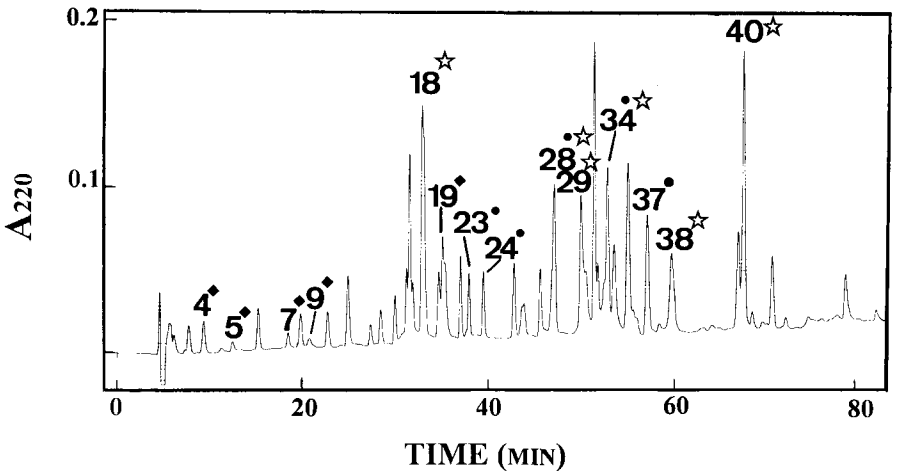

Fig. 1. Elution profile of a tryptic digest of human SC from whey of a single individual (M.W.). Tryptic peptides of fully reduced and carboxymethylated SC were applied to a RP300 reversed phase column (4.6×250 mm) equilibrated with aqueous 0.1% (w/v) TFA containing 3% (v/v) acetonitrile. A linear gradient of increasing acetonitrile concentration was applied at a flow rate of 0.7 ml/min and the effluent was monitored at 220 nm. Tryptic peptides of interest only were numbered according to their elution order. ♦, peptides encompassing the carboxy-terminal region of SC. ☆, glycopeptides present in SC; tryptic peptide T_{29} contains 2 glycosylation sites. Full structural elucidation of the oligosaccharide moieties will be presented elsewhere. ●, discrepancies between predicted protein sequence from cDNA [11] and Eiffert et al. [17] protein sequence data. See also Table 1.

bated at room temperature for 16 h with trypsin or endoproteinase Glu-C with an enzyme to protein ratio of 1 : 50 (w/w). Peptides were separated on a Brownlee RP300 reversed phase column (4.6×250 mm) at a flow rate of 0.7 ml/min and were detected at 220 nm. Solvent A was 0.1% (w/v) TFA containing 3% (v/v) acetonitrile and solvent B was 0.08% (w/v) TFA in acetonitrile containing 8% (v/v) water. A linear gradient of 0 to 50% solvent B was applied during 90 min. If necessary, peptides of interest were further purified at a flow rate of 150 μl/min on a 2.1×100 mm HAIsil 300 C18 column (Higgins Analytical Inc.) equilibrated with a buffer consisting of 20 mM methylphosphonic acid titrated to pH 6.9 with 12.5% trimethylamine. Linear gradients of either 0 to 30% or 0 to 40% of increasing acetonitrile concentration were applied during 30 min. Fractions from such chromatographic runs can be directly analyzed in an automated sequencer.

2.4. Peptide-N-glycosidase F digestion

PNGase F (0.5 mU) was added to individual purified glycopeptides and digestions were performed for 18 h at 37°C. Resulting products were purified using a C18 Sep-Pak cartridge: carbohydrates were eluted with 5% acetic acid and peptides were then released from the cartridge with 20% and 60% isopropanol in 5% acetic acid.

2.5. Peptide characterization

N-terminal sequence determination was carried out with a pulse-liquid phase microsequencer, model 477A (Applied Biosystems, Foster City, CA) using standard programs. ES-MS was carried out using a VG Bio-Q mass spectrometer. FAB-MS was performed using a VG AutoSpect E mass spectrometer equipped with a Caesium ion gun operating at 30 kV at an emission of 1 μA [21].

3. Results

A re-examination of the C-terminal structure of SC derived from human whey of a single individual (M.W.) was carried out. Fully reduced and carboxymethylated SC was digested with TPCK-treated trypsin, the digest was analysed by reversed phase HPLC (Fig. 1) and selected peptides were characterised by partial N-terminal micro-sequencing and mass spectrometry, as detailed in Table 1.

Three informative data are noteworthy from Table 1:

(a) Peptides T_5 (T_7), T_{19B}, T_{4A}, and T_9 (or E2) constitute

Table 1
Analysis of tryptic and endoproteinase glu-C derived peptides from human SC by Edman degradation and mass spectrometry

HPLC fraction no.	Calculated peptide mass	ES-MS mass observed	Residue no.	Proposed sequence ([11,12] and this work)	Sequence from [17,18]
T34A	2176.1^{mono}	2176.3^{FAB}	100–120	GLSFDVSLEVSQGPGLLND^{118}TK	Q-118
T23	2044.1^{av}	2043.8	194–212	LSDAGQYLCQAGDDSNSN^{211}K	N-211 missing
T37	2241.5^{av}	2241.8	214–232	NAD^{216}LQVLKPE^{223}PEPELVYEDLR	N-216, Q-223
T28B	1858.1^{av}	1858.0	233–250	GSVTFHCALGPE^{244}VANVAK	Q-244
T24	1707.8^{av}	1707.6	255–270	QSSGENCD^{262}VVVNTLGK	N-262
T5	815.4^{mono}	815.5	560–567	ADA^{562}APDEK	Not found
T7	843.4^{mono}	843.6	560–567	ADV^{562}APDEK	Not found
T19B	792.4^{mono}	792.2	568–574	VLDSGFR	Not found
T4A	631.3^{mono}	631.2	575–579	EIENK	Not found
T9	698.4^{mono}	698.2	580–585	AIQDPR	Not found
E2	940.5^{mono}	940.6	578–585	NKAIQDPR	Not found

Peptide designation is according to Fig. 1. T and E refer to tryptic and endoproteinase glu-C derived peptides, respectively. A or B following the tryptic peptide designation refers to peaks obtained from a reversed phase chromatographic run of the tryptic peptide fraction using a methylphosphonic acid/trimethylamine buffer at pH 7.0. Residues underlined were determined by Edman sequencing and cysteine was observed as the carboxymethyl derivative. Calculated masses of peptides with M_r under 1000 are monoisotopic (mono) and those for larger peptides studied with ES-MS are average values (av). The glycopeptide T34A was subjected to FAB-MS after treatment with PNGase F. Its monoisotopic mass was calculated taking into account the hydrolysis of the carbohydrate linked Asn at position 117 to an Asp residue.

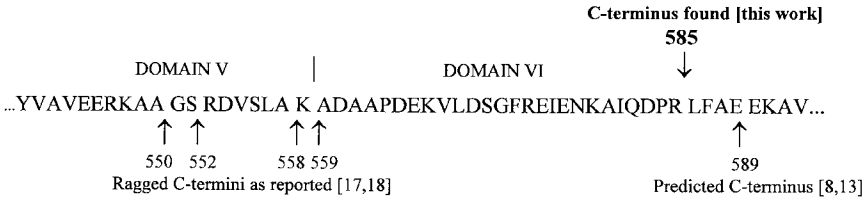

Fig. 2. The carboxy-terminal sequence of human free secretory component. The sequence shown is that deduced from cDNA [11,12] encoding the human poly-Ig receptor. Carboxy-termini were reported to be at positions 550, 552, 558 and 559 with 552 being the major species [17]. Position 589 was predicted from an alignment of all known poly-Ig receptor sequences [8,13].

the carboxy-terminal region of secreted free human SC. Tryptic peptide T_9, Ala-Ile-Gln-Asp-Pro-Arg, ends with Arg-585 and peptide E2, generated upon Glu-C endoproteinase digestion of SC, also terminates at position 585 (numbering of human SC according to SWISS-PROT database: accession number PO1833). Thus, the C-terminal end of SC is located four residues upstream from the putative extracellular cleavage site at Glu-589 [8,13]. The data is at variance with that of Eiffert and collaborators who proposed ragged C-terminal tails at positions 559, 558, 552 and 550 [17,18].

(b) Our data agree with the cDNA nucleotide sequence [11,12] and we correct conflicting assignments reported [17,18] at positions Asp-118, Asn-211, Asp-216, Glu-223, Glu-244 and Asp-262 in peptides T_{34A}, T_{23}, T_{37}, T_{28B} and T_{24}, respectively.

(c) Both T_5 and T_7 peptides span positions 560–567; they differ from each other by an Ala/Val interchange at position 562 and they were recovered in equal yield from the digest. Val-562 has not been detected at the cDNA level [12].

4. Discussion

The physiological cleavage of the extracellular part of the poly-Ig receptor following transcytosis across epithelial cells leads to the release of SC, either free or bond to polymeric Igs. To date, the precise cleavage site is still ambiguous, as human SC was found to have a ragged C-terminus, varying from Ala-550 to Lys-559, with Ser-552 as the dominant C-terminal residue [17,18]. On the basis of sequence homology among various animal species, a putative extracellular cleavage site was postulated to occur at the carboxyl side of Glu-589 [8,13], in part, because of an unlocalizable Ala-Glu C-terminal sequence identified by carboxypeptidase Y digestion in rabbit SC [16].

Re-examination of the C-terminal structure of human SC from breast milk of a single individual unequivocally identifies Arg-585 as the C-terminus of secreted SC (Table 1), i.e. 33 amino acids downstream from the major C-terminal residue determined for colostrum SC [17,18]. These data are summarized in Fig. 2.

Our data show that most if not all human SC molecules terminate at Arg-585, thus four residues upstream from the predicted putative physiological cleavage site of the poly-Ig receptor [8,10,13]. Whether the observed C-terminal Arg-585 could have resulted from further exopeptidase processing, e.g. from Ala-588–Glu-589 after the release of free extracellular SC, cannot be excluded, but it appears unlikely for the following reason: cleavage of the membrane associated poly-Ig receptor to soluble free SC was shown to be, at least partially, inhibited by leupeptin, known to bind to serine/cysteine protease active sites like trypsin [5,6]. Arg-585 appears, then, to be a compatible candidate as the true C-terminus of free human SC. In addition, Arg-585 is located in a predicted turn region, i.e. a section susceptible to proteolysis, in contrast to Ala-588–Glu-589 sequence, located within a hypothetical α-helical structure [8], known to be generally resistant to cleavage unless unfolded [22]. The consensus putative cleavage site at Glu-589, based on sequence similarities among species may not be the universal physiological site, since experimental data have so far failed to prove that this in vivo site is correct.

C-terminal sequences of SCs have been determined on rabbit, first [16], on human [17,18], and re-examined on human SC (this work). Rabbit and human SC C-terminal sequences differ [16] suggesting that distinct enzymatic specificities may be required for generating SCs from different species. Experiments designed to test this hypothesis would be awaited with interest.

Ala/Val interchange at position 562 is of interest because of possible allelic polymorphism detected in SC gene [12]. As the first polymorphic restriction site reported for *Pvu*II was localized in the third intron [23], it is irrelevant for the Ala/Val interchange cited above. A second *Rsa*I RFLP fragment was described and could represent another polymorphism. However, none of the 10 theoretical *Rsa*I fragments could account for Ala/Val change. Alternatively, a somatic point mutation (C for T) in Ala/Val codon cannot be ruled out.

Acknowledgements: This work was supported in part by Grants 5002-038919 and 31-43373.95 from the Swiss National Science Foundation.

References

[1] Krahenbuhl, J.-P. and Neutra, M.N. (1992) Physiol. Rev. 72, 853–879.
[2] Mazanec, M.B., Nedrud, J.G., Kaetzel, C.S. and Lamm, M.E. (1993) Immunol. Today 14, 430–435.
[3] Brandtzaeg, P. (1995) APMIS 103, 1–19.
[4] Mostov, K.E. (1994) Annu. Rev. Immunol. 12, 63–84.
[5] Musil, L.S. and Baenziger, J.U. (1987) J. Cell. Biol. 104, 1725–1733.
[6] Musil, L.S. and Baenziger, J.U. (1988) J. Biol. Chem. 263, 15799–15808.
[7] Solari, R., Schaerer, E., Tallichet, C., Braiterman, L.T., Hubbard, A.L. and Kraehenbuhl, J.-P. (1989) Biochem. J. 257, 759–768.
[8] Piskurich, J.F., Blanchard, M.H., Youngman, K.R., France, J.A. and Kaetzel, C.S. (1995) J. Immunol. 154, 1735–1747.
[9] Mullock, B.M., Jones, R.S. and Hinton, R.H. (1980) FEBS Lett. 113, 201–205.
[10] Mostov, K.E., Friedlander, M. and Blobel, G. (1984) Nature 308, 37–43.
[11] Krajci, P., Solberg, R., Sandberg, M., Oyen, O., Jahnsen, T. and Brandtzaeg, P. (1989) Biochem. Biophys. Res. Commun. 158, 783–789.
[12] Krajci, P., Grzeschik, K.H., Geurts van Kessel, A.H.M., Olaisen, B. and Brandtzaeg, P. (1991) Hum. Genet. 87, 642–648.

[13] Banting, G., Brake, B., Braguetta, P., Luzio, J.P. and Stanley, K.K. (1989) FEBS Lett. 254, 177–183.
[14] Kulseth, M.A., Krajci, P., Myklebost, O. and Rogne, S. (1995) DNA Cell. Biol. 14, 251–256.
[15] Verbeet, M.P., Vermeer, H., Warmerdam, G.C.M., de Boer, H.A. and He Lee, S. (1995) Gene 164, 329–333.
[16] Kühn, L.C., Kocher, H.P., Hanly, W.C., Cook, L., Jaton, J.-C. and Kraehenbuhl, J.-P. (1983) J. Biol. Chem. 258, 6653–6659.
[17] Eiffert, H., Quentin, E., Decker, J., Hillemeir, S., Hufschmidt, M., Klingmüller, D., Weber, M.H. and Hilschmann, N. (1984) Hoppe-Seyler's Z. Physiol. Chem. 365, 1489–1495.
[18] Eiffert, H., Quentin, E., Wiederhold, M., Hillemeir, S., Decker, J., Weber, M. and Hilschmann, N. (1991) Hoppe-Seyler's Z. Physiol. Chem. 372, 119–128.
[19] Hunkapiller, M.W., Huwick, R.M., Dreyer, W.J., Hood, L.E., Methods Enzymol. 91 (1983) 399–413.
[20] Klingmüller, D. and Hilschmann, N. (1979) Hoppe-Seyler's Z. Physiol. Chem. 360, 1895–1902.
[21] Dell, A., Reason, A.J., Khoo, K.-H., Panico, M., McDowell, R.A. and Morris, H.R. (1994) Methods Enzymol. 230, 108–132.
[22] Resnick, N.M., Lee Maloy, W., Guy, H.R. and Zasloff, M. (1991) Cell 66, 541–544.
[23] Krajci, P., Kvale, D., Tasken, K. and Brandtzaeg, P. (1992) Eur. J. Immunol. 22, 2309–2315.