# Structural characterization of human and bovine lung surfactant protein D

Rikke LETH-LARSEN*, Uffe HOLMSKOV† and Peter HØJRUP*[1]

*Institute of Molecular Biology, University of Southern Denmark, Odense University, DK-5000 Odense, Denmark, and
†Department of Immunology and Microbiology, Institute of Medical Biology, University of Southern Denmark, Odense University, DK-5000 Odense, Denmark

Human and bovine surfactant proteins D (SP-D) were purified from late amniotic fluid and bronchioalveolar lavage on the basis of its $Ca^{2+}$-dependent affinity for maltose. The molecular mass of a trimeric subunit was determined by matrix-assisted laser desorption ionization MS to lie in the range 115–125 kDa for human SP-D and 110–123 kDa for bovine SP-D. A single polypeptide chain was determined at 37–41 and 36–40 kDa for the human and bovine species respectively. The major parts of the primary structures of both SP-D molecules were determined by a combination of MS and Edman degradation. The heterogeneity in SP-D was caused mainly by a high number of post-translational modifications in the collagen-like region. Proline and lysine residues were partly hydroxylated and lysine residues were further O-glycosylated with the disaccharide galactose-glucose. A partly occupied N-linked glycosylation site was characterized in human SP-D. The carbohydrate was determined as a complex type bi-antennary structure, with a small content of mono-antennary and tri-antennary structures. No sialic acid residues were present on the glycan, but some had an attached fucose and/or an *N*-acetylglucosamine residue linked to the core. Bovine SP-D was determined as having a similar structure.

Key words: mass spectrometry, N-glycosylation, O-glycosylation.

## INTRODUCTION

Lung surfactant protein D (SP-D) belongs to a group of carbohydrate-binding proteins called collectins [1,2]. The proteins are oligomers of trimeric subunits organized into different quaternary structures. Each polypeptide in the subunit is composed of a short N-terminal region, a collagen-like sequence of repeating Gly-Xaa-Yaa tripeptides, a short $\alpha$-helical neck region and a C-terminal C-type carbohydrate recognition domain (CRD). Six members of the group have been characterized: conglutinin, mannan-binding lectin (MBL) and collectin 43 (CL-43), which are serum proteins produced by the liver; collectin, liver 1 ('CL-L1'), which is the newest characterized collectin found mainly in liver as a cytosolic protein [3]; lung surfactant protein A (SP-A) and SP-D, which are produced mainly by alveolar type II cells and Clara cells in the lung, but both SP-A and SP-D have been found in other mucosa-associated tissues, including the digestive and genital tracts. As with the rest of the collectins, SP-D is important in innate immunity by binding to carbohydrate structures on the surface of pathogenic micro-organisms. The binding promotes effector mechanisms such as the hindrance of infection, aggregation, activation of phagocytes and opsonization for phagocytosis.

The primary structures of SP-D have been characterized by cDNA cloning in human [4,5], rat [6], mouse [7] and cow [8]. Only one gene product per species has been identified, and it is believed that the SP-D subunit is a homotrimer. This has been substantiated by the finding that the expression of rat SP-D in Chinese-hamster ovary cells produced functional SP-D that was virtually indistinguishable from native rat SP-D [9].

The collectins are all modified post-translationally. The modifications include cleavage of signal peptides, partial hydroxylation of proline and lysine residues in the collagen-like region (except for SP-A) and assembly into subunits and higher oligomers through disulphide linkages and non-covalent interactions [9–15]. Except for SP-A, all the collectins show glycosyl-galactosyl O-linked glycosylation of hydroxylated lysine residues. The hydroxylation of proline or lysine residues followed by glycosylation is a common phenomenon in collagens, in which hydroxylation stabilizes the triple helix [16]. Human SP-D and SP-A are N-glycosylated within the collagen-like region and the CRD respectively [9,17]. None of the serum collectins are N-glycosylated.

Electron micrographs of SP-D show dodecamers in a non-randomly arrayed structure (four trimeric subunits linked at the N-terminal region) as the dominant form of native SP-D, but single subunits, as well as multimers of up to eight dodecamers, have been observed for rat SP-D. The subunits emanate from a poorly defined central core in two opposite pairs that are parallel to each other for the first 10 nm [13]. Each subunit of approx. 46 nm is terminated by a globular region 8–9 nm in diameter [9,13,18,19]. The main part of a subunit is the collagen-like region in which each chain forms a left-handed helical structure in the right-handed triple helix, which is stabilized by inter-chain hydrogen bonds. The neck region and the globular C-terminal regions are considered to be fundamental in the recognition and association of the three polypeptide chains, resulting in the nucleation of the triple-helix formation. The joining of three polypeptide chains proceeds from the neck towards the N-terminus of the protein in a zipper-like fashion [20]. Two cysteine residues (residues 15 and 20) of the N-terminal region (residues 1–25) are believed to stabilize the assembly of SP-D into dodecamers or higher-order oligomers [19]. However, the N-terminal disulphide linkages have not been identified in SP-D or any other collectin, except for the two single-subunit collectins bovine CL-43 [15] and rat MBL-C [21].

Abbreviations used: BAL, bronchioalveolar lavage; CL-43, collectin 43; CRD, carbohydrate recognition domain; DTT, dithiothreitol; MALDI-MS, matrix-assisted laser desorption ionization MS; MBL, mannan-binding lectin; PNGase, peptide N-glycosidase; RP-HPLC, reverse-phase HPLC; SP-A, surfactant protein A; SP-D, surfactant protein D; TBS, Tris-buffered saline; TFA, trifluoroacetic acid.

[1] To whom correspondence should be addressed (e-mail php@pr-group.sdu.dk).

Here we describe the structural characterization of human and bovine lung SP-D, purified from amniotic fluid and bronchioalveolar lavage (BAL).

## EXPERIMENTAL

### Purification of SP-D

Human SP-D was purified from late amniotic fluid obtained from Caesarean operations, filtered and stored at −20 °C. Bovine SP-D was purified from BAL obtained from lungs from the local slaughterhouse. The lungs were lavaged with Tris-buffered saline [(TBS; 140 mM NaCl/10 mM Tris/0.02 % (w/v) $NaN_3$ (pH 7.4)] including protease inhibitors {5 mM iodoacetamide, 5 mM cyclocapron [*trans*-4-(aminomethyl)-cyclohexacarboxylic acid (Kabi Pharmacia AB, Uppsala, Sweden)], 5 mM EDTA and 10 i.u./ml tracylol (aprotinin; Bayer, Leverkusen, Germany)}. The BAL was centrifuged at 2000 ***g*** for 10 min at 4 °C for clarification. The supernatant was kept at −20 °C.

The purification of SP-D was performed on a computer-monitored FPLC system (FPLCdirector Version 1.3; Pharmacia). The human amniotic fluid and bovine BAL were thawed at 4 °C while 10 mM EDTA was added with stirring. The solutions (0.5 litre per purification) were recalcified to a 5 mM excess of $Ca^{2+}$ over the amount of EDTA, pH adjusted to 7.4 and applied to a 15 ml maltose–agarose affinity column that was washed with 5 mM $CaCl_2$/1 M NaCl in TBS. SP-D was eluted with 100 mM $MnCl_2$ in TBS and the column was stripped with 10 mM EDTA in TBS. The $MnCl_2$ eluate was dialysed against TBS for 12 h (1:500 dilution of $Mn^{2+}$) at 4 °C before further analysis.

In some experiments SP-A was detected in the purified solutions of human SP-D, and a further purification step was performed with a CNBr-activated Sepharose 4B column (Pharmacia Biotech), containing immobilized monoclonal anti-(human SP-A). Annexin was often detected in bovine SP-D preparations, which were then further purified by gel filtration and anion-exchange chromatography.

The amount of SP-D was estimated by $A_{280}$ to be 1 mg/ml, assuming an $\epsilon_{280}$ of 1.0.

### SDS/PAGE and blotting for N-terminal sequencing

Electrophoresis was performed on 4–20 % (w/v) polyacrylamide gradient gels with a discontinuous buffer system [22]. Samples were reduced by being heated at 100 °C for 1 min in sample buffer [1.5 % (w/v) SDS/5 % (v/v) glycerol/0.2 % (v/v) Bromophenol Blue/0.1 M Tris (pH 8.0)] containing 60 mM dithiothreitol (DTT) and alkylated by adding 140 mM iodoacetamide. Unreduced samples were heated at 100 °C for 1 min in sample buffer and alkylated by the addition of iodoacetamide (140 mM).

Protein bands were detected by staining with silver [23].

Before being blotted for N-terminal sequencing, proteins were concentrated by mixing and incubation with 0.1 vol. of StrataClean Resin Pearls (Stratagene Cloning Systems, La Jolla, CA, U.S.A.). Proteins were separated by SDS/PAGE as described above and blotted on PVDF membranes (Immobilon P; Millipore, Bedford, MA, U.S.A.) with a semi-dry blotter system (Novablot; Pharmacia). After blotting, the membrane was blocked in 0.1 % (v/v) Tween 20 (polyoxyethylene sorbitan monolaurate; Merck–Schuchardt, Hohenbrunn, Germany) for 10 min. To decrease the concentration of glycine, the membranes were soaked in methanol followed by an extensive wash in water. The membrane was stained by Colloidal Coomassie Stain (Novex, San Diego, CA, U.S.A.).

### Reduction and alkylation of SP-D

Native SP-D in TBS was denatured by the addition of 8 M urea to a final concentration of 1 M, then incubated at room temperature for 30 min. DTT (45 mM, 10 $\mu$l) was added to 500 $\mu$l of SP-D (less than 1 pmol/$\mu$l) and the solution was incubated at 37 °C for a further 30 min. Iodoacetamide (100 mM, 10 $\mu$l) was added and left for 2 h in the dark at 50 °C.

Freeze-dried proteins and peptides for matrix-assisted laser desorption ionization MS (MALDI-MS) were redissolved in a reducing buffer [0.03 M DTT/100 mM $NH_4HCO_3$ (pH 8.5)] and incubated for 5–15 min at 37 °C. The reduced proteins and peptides were then prepared for MALDI-MS and analysed immediately.

Samples already on the MALDI-MS target were redissolved in 1 $\mu$l of 30 % (v/v) acetonitrile and freeze-dried. DTT (0.03 M, 1 $\mu$l) in 100 mM $NH_4HCO_3$, pH 8.5, was added and incubated at 37 °C for 15 min. The reduced proteins and peptides were then prepared for MALDI-MS and analysed immediately.

### Digestion with trypsin and collagenase

Native SP-D in TBS was denatured by dilution with 8 M urea (to a final concentration of 1 M) and digested overnight with trypsin (3 %, w/w) at 37 °C or denatured, reduced, alkylated and digested with trypsin (3 %, w/w) (modified pig, EC 3.4.21.4; Promega, Madison, WI, U.S.A.) overnight at 37 °C.

Freeze-dried protein was redissolved in a buffer containing 1 M urea, 50 mM $NH_4HCO_3$, pH 7.8, and trypsin (3 %, w/w) and incubated overnight at 37 °C. Native SP-D in TBS/10 mM $CaCl_2$ was digested with collagenase (0.1 unit of collagenase/$\mu$l of solution; clostridiopeptidase A from *Clostridium histolyticum* type VII, EC 3.4.24.3; Sigma Chemical Co., St. Louis, MO, U.S.A.) for 24–48 h at 37 °C. The peptides were separated by reverse-phase HPLC (RP-HPLC) and characterized by MALDI-MS.

### RP-HPLC

The HPLC system consisted of two LKB 2150 HPLC pumps (Pharmacia AB, Uppsala, Sweden), a variable-wavelength detector operated at 214 nm and a thermostatically controlled column oven operated at 45 °C. Two columns were used for the separation of peptides and of proteins and large peptides respectively: A 4.0 mm × 250 mm Nucleosil $C_{18}$ (7 $\mu$m particles, pore size 300 Å) and a 2.1 m × 100 mm prepacked Poros R1 column (10 $\mu$m particles). The HPLC system was controlled by a computer with software developed in-house.

Solvents in all experiments were as follows: A, 0.1 % (v/v) trifluoroacetic acid (TFA); B, 90 % (v/v) acetonitrile/0.08 % (v/v) TFA.

### MALDI-MS

The molecular-mass values of peptides and proteins were recorded either on a PerSeptive Voyager Elite mass spectrometer (PerSeptive Biosystems, Framingham, MA, U.S.A.) or on a Bruker Reflex mass spectrometer (Bruker–Franzen Analytik GmbH, Bremen, Germany). All spectra were acquired in positive-ion linear or positive-ion reflector mode. Typically, 50–250 single-shot spectra were accumulated. For data acquisition and processing on the Voyager instrument, the PerSeptive Grams/386 software (Galactic Industries Corporation) was used. The LaserOne software package (EMBL, Heidelberg, Germany) was used for the Bruker instrument.

Samples were prepared for MS with different matrices depending on the sample type: peptides, $\alpha$-cyano-4-hydroxy-

cinnamic acid (Sigma) [20 $\mu$g/$\mu$l in acetonitrile/0.1 % (v/v) TFA (70:30, v/v)]; proteins, sinapic acid (Fluka Chemie AG, Buchs, Switzerland) [20 $\mu$g/$\mu$l in acetonitrile/0.1 % (v/v) TFA (40:60, v/v)]; N-glycosylation, 2,5-dihydroxybenzoic acid (Hewlett–Packard, Palo Alto, CA, U.S.A.) [freeze-dried and redissolved in an equal volume of acetonitrile/0.1 % (v/v) TFA (30:70, v/v)] or 2,5-dihydroxybenzoic acid (Aldrich, Steinheim, Germany) [saturated solution of 50 % (v/v) acetonitrile and 2.5 % (v/v) formic acid].

Approx. 0.5 $\mu$l each of TFA (2 %, v/v), sample and matrix solution were placed on the target, mixed and left to dry. The sample was sometimes washed three times by placing 0.1 % (v/v) TFA on top of the dried sample for several seconds, followed by removal.

### N-terminal sequencing of peptides by Edman degradation

N-terminal sequencing was performed on an HP G1005A protein sequence system (Hewlett–Packard, Palo Alto, CA). Protein/peptide samples were retained on HP biphasic sequencing or membrane-compatible columns by using samples electroblotted to PVDF membrane. The sequencer was run in accordance with the manufacturer's directions.

### Sequential characterization of N-linked glycans

The N-linked glycan was characterized by sequential digestion by glycosidases immobilized on magnetic porous glass, coated with long-chain alkylamines from CPG (Lincoln Park, NJ, U.S.A.). Portions (20 $\mu$l) of magnetic particles were immobilized with 300 m-units of neuraminidase (*Arthrobacter ureafaciens*, EC 3.2.1.18; Boehringer Mannheim, Mannheim, Germany), 60 m-units of $\beta$-galactosidase (*Diplococcus pneumoniae*, EC 3.2.1.23; Boehringer Mannheim) and 60 m-units of *N*-acetylglucosaminidase (jack bean, EC 3.2.1.30; Oxford Glycosystems, Bedford, MA, U.S.A.). Before use, the magnetic particles were washed thoroughly with digestion buffer (50 mM ammonium acetate, pH 5.0). Approx. 20 pmol of peptide was freeze-dried and redissolved in 10 $\mu$l of digestion buffer. The peptide was added to one of the immobilized glycosidases and incubated for 1 h at room temperature. The magnetic particles were separated with a magnet, and the solvent was removed before analysis by MALDI-MS and further digestion with the next glycosidase.

### Digestion with peptide N-glycosidase F (PNGase F)

Freeze-dried peptides were dissolved in 50 mM $Na_2HPO_4$, pH 8.0, and incubated for 24 h at 37 °C with 0.01 unit of PNGase F/pmol of peptide (N-glycosidase F, *Flavobacterium meningosepticum*, EC 3.2.2.18/3.5.1.52; Boehringer Mannheim).

### Microscale sample purification

Microscale sample purification, with GELoader Tips (Eppendorf, Hamburg, Germany) and Poros 50 R2 material (PerSeptive Biosystems) [24], was used for the characterization of N-glycosylation in combination with MALDI-MS and 2.5-dihydroxybenzoic acid (Aldrich) as the matrix solution. For each sample to be purified, a micropurification column was prepared from a GELoader tip squeezed at the lower end to avoid the loss of stationary-phase material. The tip was filled with a few microlitres of Poros material suspended in 80 % (v/v) acetonitrile/0.1 % (v/v) TFA, so that the tip contained 0.5 cm of stationary-phase material when packed. The column was washed with 20 $\mu$l of 80 % (v/v) acetonitrile/0.1 % (v/v) TFA and equilibrated with 20 $\mu$l of 0.1 % (v/v) TFA or sample buffer. The sample was loaded on the column and washed with 20 $\mu$l of 0.1 % (v/v) TFA to remove salts and other contaminants. The proteins and peptides were eluted with 1 $\mu$l of a saturated solution of 2,5-dihydroxybenzoic acid in 2.5 % (v/v) formic acid/50 % (v/v) acetonitrile directly on to the MALDI-MS target.

### Calculation of molecular mass

The GPMAW software package (Lighthouse data, Odense, Denmark) was used for all calculations of molecular mass.

## RESULTS

### Molecular mass of native and reduced human and bovine SP-D obtained by MALDI-MS

MALDI-MS of human and bovine SP-D identified the mass range of a subunit and, after reduction, the mass range of a polypeptide chain (Table 1). A considerable amount of purified bovine SP-D was truncated after amino acid residue 20, as verified by MALDI-MS and N-terminal sequencing from protein blots. Human SP-D did not show any truncation. The broad mass range observed in the MALDI-MS spectra is due to the heterogeneity of post-translational modifications within the collagen-like region of the protein. No sequence data were obtained from N-terminal Edman sequencing of the full-length polypeptide chain of either SP-D, indicating the presence of a blocked N-terminus. This is in accordance with experiments by Lu et al. [5] and Crouch et al. [13]. However, Shimizu et al. [6] and Mason et al. [25] succeeded in obtaining N-terminal sequences of SP-D in rat and human SP-D, indicating that alanine is likely to be the N-terminal residue in SP-D of all species.

### Primary structure of human SP-D

The primary structure of human SP-D was characterized, on the basis of the known cDNA sequence predicting 355 amino acid residues (Swiss-Prot P35247) [4,5]. Purified SP-D was digested with trypsin and the resulting peptides were separated by RP-HPLC. The masses of the peptides were determined by MALDI-MS; peptides from reduced and non-reduced tryptic digests were compared. All peptides whose observed mass could not readily be identified were sequenced by Edman degradation. Heterogeneity of the primary structure caused by post-translational modifications resulted in a complicated interpretation of the mass spectra. Most RP-HPLC separated fractions were heterogeneous owing to co-eluting peptides originating either from

**Table 1 Native and reduced masses of human and bovine SP-D characterized by MALDI-MS**

The mass range was obtained at 10 % peak height and is compared with the mass obtained by SDS/PAGE.

| SP-D species | Observed molecular mass (kDa) | | |
|---|---|---|---|
| | MALDI-MS | Centroid | SDS/PAGE |
| Human SP-D, non-reduced | 115–125 | 119.9 | 150 |
| Human SP-D, reduced | 37–41 | 39.8 | 43 |
| Bovine SP-D, non-reduced | 110–123 | 116.7 | 130 |
| Bovine SP-D, reduced | 36–40 | 39.0 | 41 |
| Truncated bovine SP-D, non-reduced | 33–35 | 33.8 | 38 |
| Truncated bovine SP-D, reduced | 33–35 | 33.8 | 38 |

**Table 2 Identification by MALDI-MS and Edman sequencing of RP-HPLC fractions of tryptic peptides of reduced and alkylated human SP-D**

T1 to T20 are the numbers assigned to consecutive fractions of the separation by RP-HPLC. Proline and lysine residues in the collagen-like sequence were partly or fully modified to hydroxyproline (+16 Da). Lysine residues were hydroxylated and further O-glycosylated with galactose-glucose (+340.3 Da). T7 might contain a modified amino acid; (!) indicates chymotryptic-like cleavage; an asterisk indicates a peptide containing alkylated cysteine. $M_{measured}$ are average masses of peptides measured by MALDI-MS. Mass values in parentheses were not observed by MALDI-MS, but the identities of the peptides were verified by Edman sequencing.

| Fraction | $M_{measured}$ (Da) | Fragment (residues) | $M_{calculated}$ (Da) | Deviation (%) |
|---|---|---|---|---|
| T1 | (589.61) | 340–343 | 589.61 | |
| T2 | (406.53) | 244–246 | 406.53 | |
| | (457.49) | 37–40 | 457.49 | |
| | 1224.46 | 140–148 (1 × 340 Da) | 1224.28 | 0.01 |
| T3 | 1836.15 | 140–154 (1 Hyp + 1 × 340 Da) | 1835.94 | 0.01 |
| T4 | 2057.05 | 155–178 (2 Hyp) | 2057.19 | 0.01 |
| T5 | 2026.31 | 155–178 | 2025.19 | 0.05 |
| T6 | 3838.89 | 106–139 (1 Hyp + 2 × 340 Da) | 3840.05 | 0.03 |
| | 3856.03 | 106–139 (2 Hyp + 2 × 340 Da) | 3856.05 | 0.00 |
| T7 | 3898.30 | 106–139 (modifications?) | | |
| T8* | 1291.85 | 340–349 | 1291.41 | 0.03 |
| T9 | 6383.05 | 41–105 (3 × 340 Da) | 6782.14 | 0.01 |
| | 6799.35 (N-glycosylated) | 41–105 (1 Hyp + 3 × 340 Da) | 6798.14 | 0.02 |
| T10 | 1531.21 | 230–243 | 1531.73 | 0.03 |
| | 1900.55 | 119–139 (1 Hyp) | 1900.08 | 0.02 |
| T11 | 1403.05 | 231–243 | 1403.55 | 0.04 |
| | 1900.46 | 119–139 | 1900.08 | 0.02 |
| T12 | 1313.45 | 288–299 | 1313.45 | 0.00 |
| | 2721.17 | 188–208 (1 Hyp + 2 × 340 Da) | 2720.82 | 0.01 |
| | 3587.31 | 179–208 (2 Hyp + 2 × 340 Da) | 3586.81 | 0.01 |
| | 3603.03 | 179–208 (3 Hyp + 2 × 340 Da) | 3602.81 | 0.01 |
| | 3927.09 | 179–208 (2 Hyp + 3 × 340 Da) | 3927.43 | 0.02 |
| | 3943.69 | 179–208 (3 Hyp + 3 × 340 Da) | 3943.09 | 0.02 |
| T13 | 2381.07 | 188–208 (3 Hyp + 1 × 340 Da) | 2380.54 | 0.02 |
| | 3587.27 | 179–208 (2 Hyp + 2 × 340 Da) | 3586.81 | 0.01 |
| | 3603.38 | 179–208 (3 Hyp + 2 × 340 Da) | 3602.81 | 0.02 |
| T14 | 1512.98 | 273–287 | 1512.73 | 0.02 |
| T15 | 2704.88 | 188–208 (2 × 340 Da) | 2704.82 | 0.00 |
| | 3229.88 | 5–33(1 Hyp) | 3229.65 | 0.01 |
| T16* | 2045.90 | 213–230 (!) | 2045.33 | 0.03 |
| | 2257.10 | 10–30 (1 Hyp) | 2256.63 | 0.02 |
| | 2903.04 | 5–30 (1 Hyp) | 2901.32 | 0.06 |
| T17 | 2401.14 | 209–229 | 2401.66 | 0.02 |
| T18 | 2529.11 | 209–230 | 2529.84 | 0.03 |
| T19* | 2749.54 | 247–272 | 2749.14 | 0.02 |
| T20* | 3954.37 | 304–339 | 3954.17 | 0.01 |

different parts of the protein or from the same part of the protein but with varying degrees of post-translational modification (Table 2). Although MALDI-MS is not a quantitative method, it can provide an estimate of the ratio of peptides with various degrees of modification if the ionization efficiency can be considered equal.

Three small segments of amino acid residues were not identified owing to tryptic cleavage sites resulting in peptides of less than 500 Da: residues 1–4, 34–36 and 300–303. No substitutions of amino acid residues compared with the predicted sequence of cDNA were found. However, one peptide, T7 (residues 106–139) was recovered with a mass that did not fit the theoretical value (Table 2). We did not succeed in determining the reason for this discrepancy; the possibility of an amino acid substitution cannot be ruled out.

The collagen-like region of SP-D contains a number of post-translationally modified residues. These residues were characterized by MALDI-MS and N-terminal sequencing to be either always modified or partly modified. The designation 'partly modified' indicates that the particular amino acid residue was observed in both modified and unmodified forms.

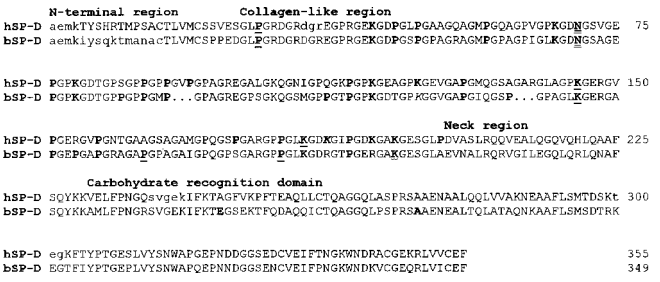


**Figure 1 Characterized amino acid sequences of human SP-D (hSP-D) and bovine SP-D (bSP-D)**

Residues in lower-case letters indicate residues predicted from the cDNA sequencing but not characterized at the protein level. Gaps within the sequence of bovine SP-D have been introduced to align the sequence of human SP-D. Residues in bold in the collagen-like region are partly modified. Bold and underlined residues are always modified. Lys-187 in human SP-D (bold italic) is unmodified or partly O-glycosylated. Lys-190 in bovine SP-D (wavy underline) is partly hydroxylated, and partly hydroxylated and glycosylated. Asn-70 (doubly underlined) is partly N-glycosylated. Residues in bold in the CRD indicate substitutions relative to the sequence predicted from cDNA.

All the proline residues located in the Yaa position relative to glycine in the Gly-Xaa-Yaa repeat of the collagen-like region were found to be modified to hydroxyproline. MALDI-MS of peptides containing one or several hydroxyproline residues showed a characteristic pattern of peaks differing by 16 Da. Human SP-D contained 16 modified proline residues; one was only observed modified, whereas the remaining fifteen were partly modified (Figure 1). The hydroxylation of each partly modified proline residue varied somewhat; on the basis of MS and Edman sequence results, this was estimated to be 20–50 %. Proline residues located in the Xaa position relative to glycine were never found to be hydroxylated; neither were proline residues outside the collagen-like region.

Lysine residues in the Yaa position relative to glycine were usually modified post-translationally to yield a mass of 340.3 Da. The mass change was identical with the modification found on lysine residues in CL-43 [15] and characterized as hydroxylysine glycosylated with a disaccharide, lysine-*O*-galactose-glucose (average mass 468.3 Da). No phenylthiohydantoin derivative was observed during the Edman degradation of modified lysine residues. In tryptic digests, no cleavage sites C-terminal to the modified lysine residues were found, but a mass difference of 340.3 Da (compared with the calculated mass of a tryptic peptide from the cDNA-derived sequence) was obtained in all cases of modification. Human SP-D contained nine modified lysine residues, of which two were always O-glycosylated and seven were partly O-glycosylated hydroxylysine residues. Lys-187 was only observed unmodified, but the possibility of partial O-glycosylation cannot be excluded (Figure 1).

A partial N-glycosylation was characterized at Asn-70, as described below.

### Primary structure of bovine SP-D

The primary structure of bovine SP-D was characterized in the same way as human SP-D on the basis of the known cDNA sequence, predicting 349 residues (Swiss-Prot P35246) [8] (results not shown). The difference in the number of residues in human

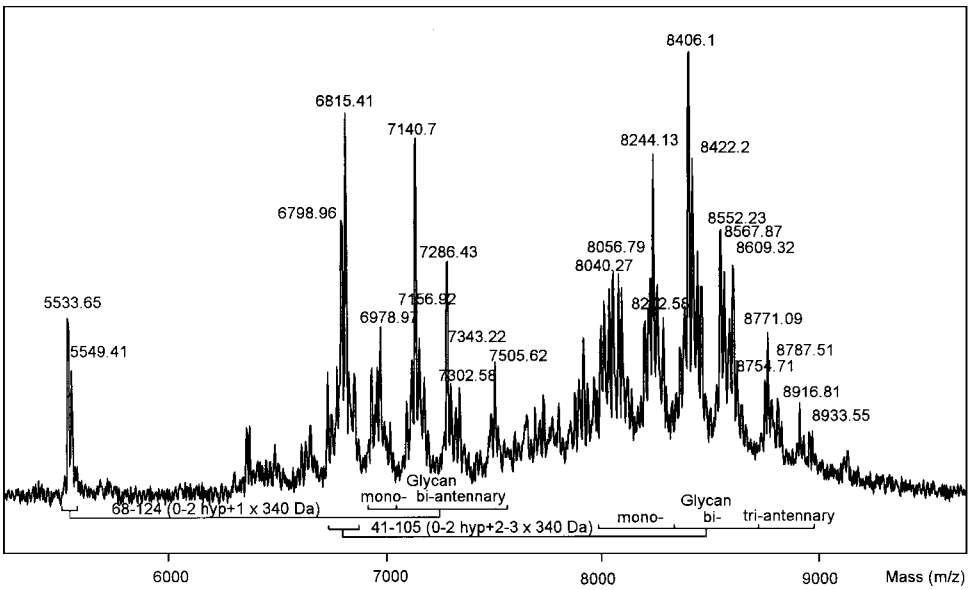


**Figure 2 MALDI-MS of residues 68–124 and 41–105 from human SP-D with partly linked N-glycosylation to Asn-70**

The dominant glycans are complex type bi-antennary structures, but traces of mono- and tri-antennary structures are also observed. Fucose and *N*-acetylglucosamine are linked to the core of some of the glycans. The closely spaced peaks in each group are the result of partial hydroxylations.

and bovine SP-D is within the collagen-like sequence (59 and 57 Gly-Xaa-Yaa repeats respectively). The complete amino acid sequence of bovine SP-D was characterized by MS and protein sequencing, except for residues 1–20 and 233–237. The N-terminal 20 residues were not characterized owing to truncation of the major fraction of the protein. The peptide containing residues 233–237 was not found in the HPLC separation, most probably because it was too small for identification with MALDI-MS. A few amino acid substitutions were found in comparison with the cDNA-derived amino acid sequence: bovine SP-D contained Glu-242 instead of Val-242; residue 268 was identified as an alanine as well as the predicted glycine.

As in human SP-D, all proline residues located in the Yaa position relative to glycine in the collagen-like region were observed as being modified to hydroxyproline. Bovine SP-D contained 18 modified proline residues; three of them were always modified to hydroxyproline and fifteen were partly modified. Seven lysine residues were modified; one was always hydroxylated and O-glycosylated, and six were partly hydroxylated and O-glycosylated. The last lysine residue in the collagen-like region (Lys-190) was seen as both unmodified and hydroxylated, as well as hydroxylated and O-glycosylated (Figure 1).

## Characterization of N-glycosylation

Both forms of SP-D contained a single site of N-glycosylation (Asn-70), which was recovered both modified and unmodified from human SP-D. The level of occupancy could not be established owing to the small amount of sample.

N-terminal sequencing of T9 from the RP-HPLC separation of trypsin-digested human SP-D characterized the peptide as residues 41–105, whereas MALDI-MS of the fraction showed a heterogeneous pattern of peptides (Figure 2). Treatment of the peptides with PNGase F, followed by MALDI-MS, verified the presence of an N-linked glycan. The tryptic peptide fraction used for characterization of the N-glycosylation was modified with two or three O-glycosylated hydroxylysine residues and zero to two hydroxyproline residues, as determined by MALDI-MS. A minor fraction of a peptide, residues 68–124 (5176.65 Da), with one O-glycosylation and one or two hydroxyproline residues was also present. T9 could not be separated efficiently by RP-HPLC and was collected in three fractions. The most hydrophilic part contained mainly the N-glycosylated peptide; the middle part contained equal amounts of N-glycosylated and unglycosylated (but O-glycosylated) peptide; the more hydrophobic part contained mostly the peptide without the N-linked glycan. The middle fraction, with the peptide with and without N-glycosylation, was used for the analysis of the glycan. The N-glycosylation was analysed by sequential digestion with glycosidases immobilized on magnetic beads, followed by MALDI-MS [26]. Because the unglycosylated peptide remained unchanged during the sequential digestion steps of the monosaccharides, the interpretation of the mass spectra of the digested glycan became less difficult.

Digestion with neuraminidase as the first step did not show any changes in the glycan, indicating that no sialic acids were present. After treatment with $\beta$-galactosidase, the molecular mass of the glycan was decreased by 324 Da, corresponding to the loss of two galactose residues. Minor fractions of the peptide showed the loss of one and three galactose residues. Digestion of the intact peptide with $\beta$-galactosidase, without prior treatment with neuraminidase, showed the same pattern, confirming that no sialic acids were present on the glycan. On continuing digestion with *N*-acetylglucosaminidase, the same pattern was observed, with the loss of one to three *N*-acetylglucosamine residues. PNGase F treatment in solution removed the core of the glycan. These results showed that the glycan consisted mainly of the complex type bi-antennary structure, but minor fractions of mono- and tri-antennary structures were also present. Further, the digestion showed that some structures had a fucose and/or *N*-acetylglucosamine residue attached to the core (Figure 2).

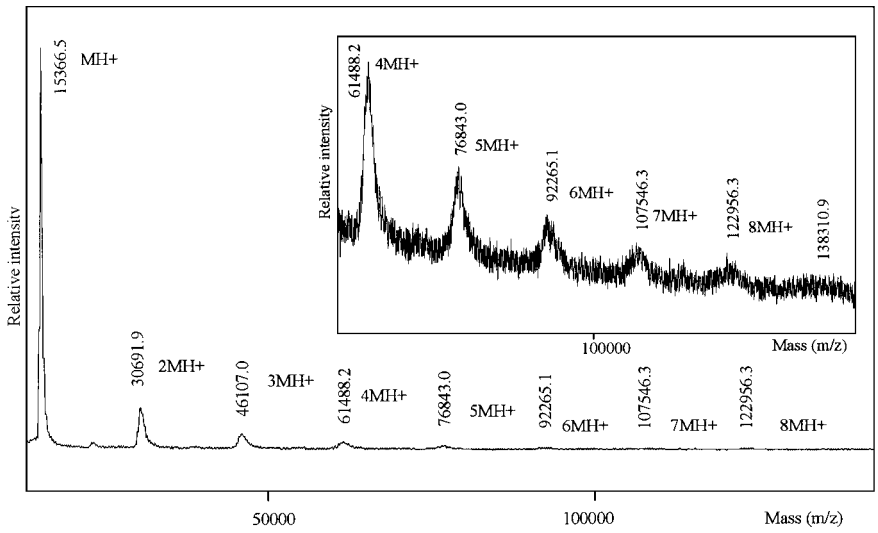


**Figure 3 MALDI-MS of the N-terminal region after digestion with collagenase followed by RP-HPLC**

A high-mass-range mass spectrum of the N-terminal region also showing multimers of up to 8-fold the 15–16 kDa peptide. The inset shows an expanded view of the 60–150 kDa range.

The N-glycosylation was not characterized in bovine SP-D. However, on the basis of the results from human SP-D, it was possible to recognize the same peptide mass pattern in the MALDI-MS analysis of peptide residues 55–102, although in a yield too low for structural studies. Edman sequencing of the peptide showed no phenylthiohydantoin-asparagine at the position of residue 70, also indicating the presence of a modified residue.

### Intra-chain disulphide linking in the C-terminal part

As in other C-type lectin CRDs, the four conserved cysteine residues were expected to be intra-chain bonded as cysteines 1 to 4 and 2 to 3. This was confirmed by MALDI-MS and Edman sequencing of non-reduced tryptic peptides, followed by reduction and re-analysis. In human SP-D the linkage was Cys-261–Cys-353 and Cys-331–Cys-345. In bovine SP-D the linkage was Cys-255–Cys-347 and Cys-325–Cys-339.

### Characterization of the N-terminal region

Human SP-D was digested with collagenase in an attempt to characterize the disulphide-bonding pattern of the N-terminal region of the protein. The separation of collagenase-digested SP-D on RP-HPLC resulted in several peptides, which were characterized by MALDI-MS and Edman sequencing. However, the peptides were heterogeneous owing to different collagenase cleavage sites. A number of peptides with a mass of less than 5 kDa could be identified as originating from the collagen-like region, whereas a heterogeneous peptide of approx. 17 kDa was identified as the C-terminal domain. The N-terminal region was identified as a heterogeneous peptide with a mass of 15–16 kDa that also showed multimers of up to 8-fold the 15–16 kDa peptide (Figure 3). After reduction, several minor peptides of 2–4 kDa were observed. On the basis of the calculated mass of a single peptide chain containing residues 1–25 (2.8 kDa), the 15–16 kDa peptide (2–4 kDa reduced) corresponded to the mass of six polypeptide chains (the N-terminal part) of various lengths, connected as two subunits through disulphide bonds. Multimers of the 15–16 kDa peptide could denote oligomerization of the paired subunits by inter-subunit disulphide bonds.

## DISCUSSION

All collectins analysed so far are extensively modified post-translationally [9,11,12,15]. Mass-spectrometric analysis of intact human and bovine SP-D confirms this, because the mass obtained is considerably higher than predicted from the amino acid sequence. Furthermore, the mass spectra show very wide peaks, both of subunits and reduced polypeptide chains, indicating the heterogeneous nature of the post-translational modifications. The spectra of bovine SP-D also show the presence of a truncated version. The site of truncation was characterized by Edman sequencing and peptide mapping to be C-terminal to Cys-20. There are no published reports of a truncated version of SP-D, but truncation has also been found in conglutinin [27,28] and in bovine CL-43 [15]. Despite the high resemblance of conglutinin and SP-D in primary and quaternary structures, there are no indications of a naturally occurring truncation in human SP-D. The dominant form of human SP-D is clearly the 37–41 kDa polypeptide (43 kDa by SDS/PAGE), however, in a recent report, Mason et al. [25] demonstrated the presence of a single-subunit form of human SP-D with a reduced mass of 50 kDa (SDS/PAGE). The major difference from the form described here seems to be the glycosylation of three threonine residues in the N-terminal part of the sequence. In the present investigation, no indication of glycosylated threonines was observed (Table 2). This could have been a result of our purification procedure, which specifically targeted the purification of the high-molecular-mass species.

The minor difference in mass between human and bovine SP-D (Table 1) is caused by a difference in length of the collagen-like region (177 and 171 residues respectively). The length of the collagen-like region does not seem to be critical for the function of SP-D. The heterogeneity of the SP-D species (Tables 1 and 2) is caused mainly by the post-translational modifications in the collagen-like region. The potentially modified residues have been characterized by MALDI-MS and Edman sequencing to be always modified, partly modified or unmodified. The

modifications are very similar in human and bovine SP-D; similar patterns should be expected for SP-D in other species. The characteristics of the modified residues are the hydroxylation of proline and lysine residues in the Yaa position of Gly-Xaa-Yaa repeats. Lysines are in most cases further O-glycosylated with galactose-glucose (glycosyl-galactosyl-hydroxylysine). On the basis of the intensity of the different hydroxylation/glycosylation variants in MALDI-MS and during sequencing of the peptides, the partly modified residues were judged to be 20–50 % hydroxylated/glycosylated, with only slight variation between preparations. No correlation between the modifications of different sites could be found. Although the purpose of the hydroxylated prolines is to stabilize the collagen triple helix [29], the purpose of the hydroxylation and glycosylation of lysine residues seems less clear.

The glycosylated lysine residues of both human and bovine SP-D were found in segments of -Gly-Xaa-Lys-Gly-Asp/Glu-. Two unmodified lysine residues were found in the collagen-like region positioned in segments of -Gly-Xaa-Lys-Gly-Gly/Ile-. This could indicate that the presence of an acidic residue in the second position to lysine has a role in recognition by the hydroxylation enzyme. The same segments around potentially modified lysine residues are found in conglutinin, MBL, CL-43 and SP-A. However, although produced by alveolar type II cells as SP-D, no hydroxylation (and further O-glycosylation) of lysine residues has been found in SP-A [30].

SP-D has previously been characterized as a glycoprotein, with the N-linked glycosylation located at position 70 [4,6,9]. However, like most of the hydroxylations, the glycosylation of Asn-70 is only partial. We have characterized the main form of attached glycan in human SP-D to be a complex type bi-antennary structure, containing minor fractions of mono- and tri-antennary structures. Fucose and/or *N*-acetylglucosamine are linked to the core structure of some of the glycans. Surprisingly, no sialic acids were detected on the glycan. All known SP-D sequences (human, bovine, rat and mouse) contain a potential glycosylation site at position 70 that is expected to be partly glycosylated. Evolutionary conservation of the site of Asn-70 within the collagen-like region of SP-D suggests that the glycosylation, although only partial, has an important role in the biosynthesis and/or function of SP-D. However, the carbohydrate is not required for the alignment of subunits or subsequent dodecamer secretion [19]. No N-linked glycosylations have been found in conglutinin, CL-43 or MBL, whereas a complex antennary type is present in the CRD of SP-A [31]. The interaction between SP-A and micro-organisms, and the subsequent opsonization, is critically dependent on the presence of the N-linked glycan [32,33].

The mass of a single polypeptide chain can be calculated on the basis of the above information for post-translational modifications. The theoretical mass range for human SP-D would, for a random distribution of the partial modifications, range from 36 163 Da (polypeptide plus permanent hydroxylation and two O-glycosylations) to a full complement of 40 758 Da (16 hydroxyproline residues, 9 O-glycosylations and one bi-antennary N-glycosylation including core fucose and *N*-acetylglucosamine residues). These values are in accordance with the measured mass range of 37–41 kDa as determined by MALDI-MS (Table 1). Similar theoretical values for bovine SP-D are 35 541–39 796 Da (assuming a glycosylation similar to human SP-D), which fits with the determined range of 36–40 kDa.

A thorough analysis of tryptic digests from both human and bovine native non-reduced SP-D failed to reveal the N-terminal peptide containing Cys-15 and Cys-20. However, digestion of SP-D with collagenase resulted in the recovery of a peptide with a mass of 15–16 kDa. Multimers of this peptide with a mass of up to 123 kDa were observed. Reduction of the peptide resulted in several N-terminal peptides that, owing to non-specific cleavage by collagenase, ranged from 2 to 4 kDa. The 15–16 kDa peptide therefore corresponds to six N-terminal peptides connected as two subunits through disulphide bonds. Multimers of the 15–16 kDa peptide might denote oligomerization of the paired subunits by inter-subunit disulphide bonds. Crouch et al. [13] reported that electron micrographs of SP-D revealed a highly homogeneous population of molecules characterized by four or more relatively rigid-appearing subunits, non-randomly arrayed about a central core. The core was only poorly defined as the centre of interaction for the N-termini of the subunits. The subunits seemed to emanate from opposite sides of the core in pairs closely apposed over a distance of 10 nm from the centre of the core. The results for paired subunits presented here correspond very well to the structure of SP-D observed in electron micrographs.

Site-directed mutagenesis of Cys-15 and Cys-20 has shown the necessity for N-terminal cysteine residues to stabilize oligomerization, but the cysteine residues are not necessary for trimer formation [34]. However, other results [9] provide strong evidence for the existence of at least one intra-subunit disulphide bond within the N-terminal region after limited peptic digestion.

It is possible that apposing pairs of SP-D subunits are associated in an overlapping anti-parallel arrangement restricting the spatial distribution of the subunits around the central core. Another hypothesis is the association of paired subunits end to end. Two subunits are disulphide-bonded between and within subunits, whereas the two (or more) opposing pairs of subunits are associated through an undefined molecule. It has not been possible to exclude the possibility of the contribution of other molecules in the assembly of several SP-D subunits. It is also likely that cross-link formation between two or four subunits involves previously unpaired cysteine residues, or that a process of rearrangement converts a specific subset of intra-subunit disulphide bonds to inter-subunit bonds.

We thank Ida Tornoe and Inger Christiansen for valuable assistance. This work was supported by the Novo Nordisk Foundation and the Benzon Foundation.

## REFERENCES

1 Weis, W. I., Taylor, M. E. and Drickamer, K. (1998) Immunol. Rev. **163**, 19–34
2 Hansen, S. and Holmskov, U. (1998) Immunobiology **199**, 165–189
3 Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Yamazaki, H., Shimada, T., Keshi, H., Sakai, Y., Fukuoh, A. et al. (1999) J. Biol. Chem. **274**, 13681–13689
4 Rust, K., Grosso, L., Zhang, V., Chang, D., Persson, A., Longmore, W., Cai, G. Z. and Crouch, E. (1991) Arch. Biochem. Biophys. **290**, 116–126
5 Lu, J., Willis, A. C. and Reid, K. B. (1992) Biochem. J. **284**, 795–802
6 Shimizu, H., Fisher, J. H., Papst, P., Benson, B., Lau, K., Mason, R. J. and Voelker, D. R. (1992) J. Biol. Chem. **267**, 1853–1857
7 Motwani, M., White, R. A., Guo, N., Dowler, L. L., Tauber, A. I. and Sastry, K. N. (1995) J. Immunol. **155**, 5671–5677
8 Lim, B. L., Lu, J. and Reid, K. B. (1993) Immunology **78**, 159–165
9 Crouch, E., Chang, D., Rust, K., Persson, A. and Heuser, J. (1994) J. Biol. Chem. **269**, 15808–15813
10 Colley, K. J. and Baenziger, J. U. (1987) J. Biol. Chem. **262**, 10296–10303
11 Lee, Y. M., Leiby, K. R., Allar, J., Paris, K., Lerch, B. and Okarma, T. B. (1991) J. Biol. Chem. **266**, 2715–2723
12 McCormack, F. X., Calvert, H. M., Watson, P. A., Smith, D. L., Mason, R. J. and Voelker, D. R. (1994) J. Biol. Chem. **269**, 5833–5841
13 Crouch, E., Persson, A., Chang, D. and Heuser, J. (1994) J. Biol. Chem. **269**, 17311–17319
14 Holmskov, U., Laursen, S. B., Malhotra, R., Wiedemann, H., Timpl, R., Stuart, G. R., Tornoe, I., Madsen, P. S., Reid, K. B. and Jensenius, J. C. (1995) Biochem. J. **305**, 889–896
15 Rothmann, A. B., Mortensen, H. D., Holmskov, U. and Højrup, P. (1997) Eur. J. Biochem. **243**, 630–635

16 Kivirikko, K. I. and Myllyla, R. (1985) Ann. N. Y. Acad. Sci. **460**, 187–201
17 Phelps, D. S., Floros, J. and Taeusch, H. W. (1986) Biochem. J. **237**, 373–377
18 Lu, J., Wiedemann, H., Holmskov, U., Thiel, S., Timpl, R. and Reid, K. B. (1993) Eur. J. Biochem. **215**, 793–799
19 Brown-Augsburger, P., Chang, D., Rust, K. and Crouch, E. (1996) J. Biol. Chem. **271**, 18912–18919
20 Hoppe, H. J., Barlow, P. N. and Reid, K. B. (1994) FEBS Lett. **344**, 191–195
21 Wallis, R. and Drickamer, K. (1997) Biochem. J. **325**, 391–400
22 Laemmli, U. K. (1970) Nature (London) **227**, 680–685
23 Merril, C. R., Goldman, D. and Van-Keuren, M. L. (1983) Methods Enzymol. **96**, 230–239
24 Kussmann, M., Nordhoff, E., Rahbek-Nielsen, H., Háebel, S., Rossel-Larsen, M., Jakobsen, L., Gobom, J., Mirgorodskaya, E., Kroll-Kristensen, A., Palm, L. and Roepstorff, P. (1997) J. Mass. Spectrom. **32**, 593–601
25 Mason, R. J., Nielsen, L. D., Kuroki, Y., Matsuura, E., Freed, J. H. and Shannon, J. M. (1998) Eur. Respir. J. **12**, 1147–1155
26 Krogh, T. N., Berg, T. and Højrup, P. (1999) Anal. Biochem., **274**, in the press
27 Kawasaki, N., Yokota, Y. and Kawasaki, T. (1993) Arch. Biochem. Biophys. **305**, 533–540
28 Malhotra, R., Laursen, S. B., Willis, A. C. and Sim, R. B. (1993) Biochem. J. **293**, 15–19
29 Segal, D. M. (1969) J. Mol. Biol. **43**, 497–517
30 Hawgood, S (1991) Annu. Rev. Physiol. **53**, 375–394
31 Munukata, H., Nimberg, R. B., Snider, G. L., Robins, A. G., Van Halbeek, H., Vliegenthart, J. F. G. and Schmid, K. (1982) Biochem. Biophys. Res. Commun. **108**, 1401–1405
32 Van Iwaarden, J. F., van Strijp, J. A., Visser, H., Haagsman, H. P., Verhoef, J. and van Golde, L. M. (1992) J. Biol. Chem. **267**, 25039–25043
33 Benne, C. A., Benaissa-Trouw, B., van Strijp, J. A., Kraaijeveld, C. A. and van Iwaarden, J. F. (1997) Eur J. Immunol. **27**, 886–890
34 Brown-Augsburger, P., Hartshorn, K., Chang, D., Rust, K., Fliszar, C., Welgus, H. G. and Crouch, E. (1996) J. Biol. Chem. **271**, 13724–13730

Received 7 June 1999/22 July 1999; accepted 2 September 1999