Protein Science (1999), *8*:1551–1556. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Determination of the complete covalent structure of the major glycoform of DQH sperm surface protein, a novel trypsin-resistant boar seminal plasma O-glycoprotein related to pB1 protein

KAREL BEZOUŠKA,[1] JAN SKLENÁŘ,[1] PETR NOVÁK,[1] PETR HALADA,[2] VLADIMÍR HAVLÍČEK,[2] MAREK KRAUS,[3] MARIE TICHÁ,[1] AND VĚRA JONÁKOVÁ[3]

[1]Department of Biochemistry, Faculty of Science, Charles University Prague, Hlavova 8, CZ-12840 Praha 2, Czech Republic
[2]Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, CZ-14220 Praha 2, Czech Republic
[3]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, CZ-16637 Praha 6, Czech Republic

(RECEIVED December 15, 1998; ACCEPTED March 19, 1999)

Abstract: The complete covalent structure of a novel boar DQH sperm surface protein resistant to many classical procedures of enzymatic fragmentation was determined. The relative molecular mass of the major form of this protein determined by ESI-MS and MALDI-MS was 13,065.2 ± 1.0 and 13,065.1, respectively. However, additional peaks differing by 162 Da (i.e., minus hexose), 365 Da (i.e., minus hexose and N-acetylhexosamine), 146 Da (i.e., plus deoxyhexose), and 291 Da (i.e., plus sialic acid) indicated the heterogeneity due to differences in glycosylation. The complete covalent structure of the protein was determined using automated Edman degradation, MALDI-MS, and post-source decay (PSD) MALDI-MS, and shown to consist of N-terminal O-glycosylated peptide followed by two fibronectin type II repeats. The carbohydrates are O-glycosidically linked to threonine 10, as confirmed by PSD MALDI-MS of the isolated N-terminal glycopeptide. Eight cysteine residues of the protein form four disulfide bridges, the positions of which were assigned from MALDI-MS and Edman degradation data. We conclude that mass spectral techniques provide an indispensable tool for the detailed analysis of the covalent structure of proteins, especially those that are refractory to standard approaches of protein chemistry.

Keywords: boar seminal plasma; DQH sperm surface protein; ESI-MS; fibronectin type II repeat; MALDI-MS; O-glycosylation; post-source decay MALDI-MS

Reprint requests to: Karel Bezouška, Department of Biochemistry, Faculty of Science, Charles University Prague, Hlavova 8, CZ-12840 Praha 2, Czech Republic; e-mail: bezouska@biomed.cas.cz.

Abbreviations: BNPS-skatol, 3-bromo-3-methyl-2-(2-nitrophenylmercapto)-3H-indole; calc., calculated molecular mass; Deoxyhex, deoxyhexose; ESI, electrospray ionization; Hex, hexose; HexNAc, N-acetylhexosamine; HPLC, high-performance liquid chromatography; MALDI, matrix-assisted laser desorption/ionization; meas., measured molecular mass; MS, mass spectrometry; PSD, post-source decay; SDS, sodium dodecyl sulfate; TFA, trifluoroacetic acid.

Spermadhesins are proteins located on the boar sperm surface, which play roles in the complex chain of events leading to sperm capacitation, gamete recognition, and binding to the ovum (Calvete et al., 1994). The amino acid sequence of several spermadhesins belonging to AQN and AWN families (AQN 1, AQN 3, AWN, PSP I, PSP II, and bovine aSFP) have already been reviewed (Töpfer-Petersen et al., 1995); if glycosylated, N-glycosylation occurs at asparagine 50. Amino acid sequence pattern analysis (Bork & Beckmann, 1993; Calvete et al., 1994) revealed that spermadhesins may represent a prototype of an immunoglobulin-like antiparallel β-barrel (a structure module called CUB domain), which is widely distributed among developmentally regulated proteins.

It has been shown that boar seminal plasma contains proteins that are not members of AQN and AWN families (Fernlund et al., 1994; Hadjisavas et al., 1994; Jonáková et al., 1998). One of them, the phosphorylcholine-binding protein pB1, has been recently characterized (Calvete et al., 1997). This protein exhibits sequence homology with major proteins of bovine and stallion seminal plasma sequenced previously (Seidah et al., 1987; Calvete et al., 1995). This paper describes covalent amino acid structure and oligosaccharide chain characterization of a novel boar DQH sperm surface protein that differs in its structure and binding properties from those of boar spermadhesins of AQN and AWN family and shows sequence homology with pB1.

Results and discussion: As a part of the project in which proteins in the heparin-binding fraction of boar seminal plasma were investigated, DQH sperm surface protein with high binding affinity to acidic polysaccharides has been obtained (Tichá et al., 1998). Its precise molecular mass was determined by electrospray and MALDI-MS, both of which revealed its microheterogeneity (Table 1). The predominant protein peak in MALDI spectrum (Fig. 1) at *m/z* 13,065 is accompanied by minor peaks at *m/z* 12,903, 13,212, and 13,358. Similar molecular pattern was observed in ESI-MS. Both

Table 1. *Determination of molecular weight of several glycoforms of native DQH protein by mass spectrometry (m/z)*

	$[M+H]^+$ by MALDI-MS	$[M+H]^+$ by ESI-MS
Deglycosylated	Not found	12,698.8 ± 1.4
T_{10}-HexNAc	12,903.2	12,900.0 ± 1.0
T_{10}-HexNAC-Hex	13,065.1	13,065.2 ± 1.0
T_{10}-HexNAc-Hex–(Deoxyhex)	13,212.3	Not found
T_{10}-HexNAc-Hex–NeuAc	13,358.3	13,355.2 ± 1.7

profiles of the protonated molecules indicate the microheterogeneity in glycosylation. The differences of 162, 201, 147, and 293 Da represent Hex, HexNAc, Deoxyhex, and N-acetylneuraminic acid additions, respectively. Quantitative carbohydrate analysis performed by capillary zone electrophoresis revealed hexose to be galactose (Gal), N-acetylhexosamine to be N-acetylgalactosamine (GalNAc), and deoxyhexose to be fucose (Fuc). Thus, the carbohydrate moiety seems to consist of a mixture of GalNAc and the disaccharide Gal-GalNAc; the portion of the latter is further substituted with Fuc or NeuAc. It is difficult to exclude some losses of sialic acids during the preparation of DQH protein in acidic environments (0.1% TFA), and thus the fully sialylated oligosaccharide may possibly be more prevalent in the native protein. Indeed, the abundance of the peak corresponding to the sialylated glycoform in the MALDI-MS spectrum varied in different batches of DQH protein. On the other hand, the occurrence of fucose, the glycosidic bond of which is much more stable in acidic environment, may be a subject of genuine variations in different DQH protein preparations. The results of the detailed analysis of oligosaccharide components of various DQH sperm surface protein preparations will be published elsewhere.

When DQH protein was subjected to N-terminal sequencing (Jonáková et al., 1998), there was a blank cycle in the position of amino acid 10 indicating it as the possible site for carbohydrate attachment. Therefore, we aimed at the characterization of the N-terminal (amino acids 1–21) pepsin peptide 1 (Table 2) in its native as well as chemically deglycosylated (by HF treatment) forms. When the mass of this peptide was determined by MALDI-MS in linear mode (Fig. 1), the heterogeneity recorded matched that observed with the whole protein with differences by 146 and 291 Da between the major and minor peaks. In addition, the accurate monoisotopic masses of the corresponding protonated molecules were measured in reflectron mode, in which sialic acids are generally lost due to post-source fragmentation. In this mode, therefore, only two glycoforms at *m/z* 2,796.5 (2,796.4 calc. for Hex-HexNAc peptide) and *m/z* 2,942.6 (2,942.4 calc. for Deoxyhex-Hex-HexNAc peptide) were detected. After the chemical deglycosylation of peptide 1, only two major peaks at *m/z* 2,431.2 and 2,634.3 were observed. Interestingly, while the molecular mass of the former peak corresponds exactly to the 21 N-terminal amino acids of the protein (Jonáková et al., 1998) with threonine in position 10 (calc. 2,431.2), the latter form of the peptide represents the same N-terminal peptide with HexNAc attached (calc. 2,634.4). This suggestion has been further confirmed by post-source decay MALDI-MS analysis of the peptide 2 (Table 3): y_7 and y_{10}–y_{13} fragments all possess the 365 Da increment corresponding to the Hex-HexNAc substitution. Thus, DQH sperm surface protein is an O-glycoprotein in which the threonine-10 represents the site for the attachment of carbohydrates.

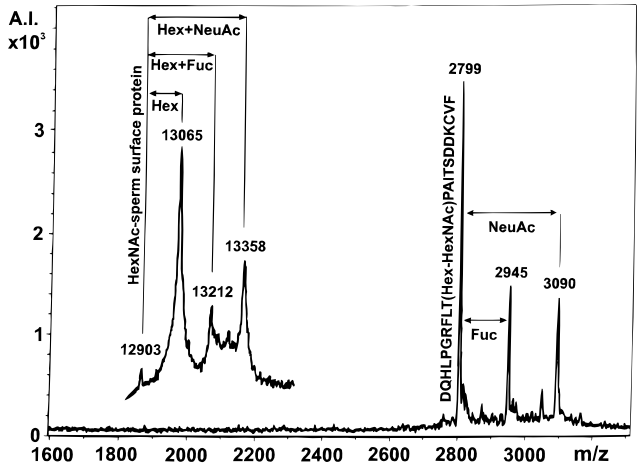

Fig. 1. Linear mode MALDI-MS spectrum of N-terminal peptide 1 from DQH sperm surface protein revealing the microheterogeneity in glycosylation. The inset represents molecular profile of the entire protein showing its individual glycoforms.

The complete amino acid sequence of DQH protein was determined by automated Edman degradation of the entire protein, and of peptides obtained by chemical cleavages or enzymatic digestions. The protein has been sequenced several times from its N-terminus, and from the longest stretch of sequence we could call the amino acids in the first 45 cycles (Fig. 2). Since this sequence had a tryptophan residue at position 42, the DQH protein was cleaved by BNPS-skatol, and one of the three peptides thus obtained allowed us to extend the sequence up to another tryptophane at position 53. The latter peptide contained a methionine residue at position 50, and since the amino acid analysis indicated the presence of a single methionyl residue in DQH protein, the cleavage of this protein with CNBr promised a feasible way to obtain more of the internal sequence. Indeed, CNBr treatment of DQH sperm surface protein resulted in the generation of two large peptides with the relative molecular mass of ~6,000 Da. One of these peptides yielded a sequence identical with the N-terminal sequence, but the sequencing of the second large CNBr peptide allowed us to obtain a considerable length of additional internal sequence of DQH protein, residues 51–84. The generation of additional peptides from the DQH protein by enzymatic digestion posed significant problems because of the extremely low solubility of the protein at neutral or slightly alkaline pH, and its considerable resistance to digestions with many proteinases. Thus, the only enzymatic fragmentation that resulted in reasonable quantities of peptide fragments was the treatment with V8 protease (using the ability of this enzyme to cleave at the second, acidic pH optimum in acetate buffer at pH 4.6) and pepsin. The cleavage of DQH protein with V8 protease resulted in the generation of three peptides. The largest of these peptides corresponded to the large region of the protein already sequenced (residues 1–74), while the smallest peptide also encompassed the already known sequence (residues 75–81). However, the third V8 peptide contained the entire C-terminal portion of the protein, and its sequencing allowed us to complete the sequence of DQH protein, residues 82–105.

Table 2. *Results of MALDI-MS measurements of m/z for the most relevant peptides from DQH sperm surface protein*

Number	Position	Sequence	Modification	$[M+H]^+_{calc.}$	$[M+H]^+_{meas.}$	Enzyme
1	1–21	DQHLPGRFLTPAITSDDKCVF	Gal-GalNAc at T^{10} Acrylamido-C^{19}	2,796.34	2,796.47	Pepsin
2	1–14	DQHLPGRFLTPAIT	Gal-GalNAc at T^{10}	1,930.98	1,931.06	Pepsin
3	16–31	DDKCVFPFIYKGNLYF	Acrylamido-C^{19}	2,039.99	2,039.96	Asp-N
4	30–40	YFDCTLHDSTY	Acrylamido-C^{33}	1,435.58	1,435.55	Pepsin
5	49–61	YMKRWRYCRSTDY	Acrylamido-C^{56}	1,898.87	1,998.92	Pepsin
6	64–71	CALPFIFR	Acrylamido-C^{64}	1,037.55	1,037.50	Trypsin
7	71–85	RGKEYDSCIKEGSVF	Unmodified	1,717.82	1,717.83	Pepsin
8	82–99	GSVFSKYWCPVTPNYDQD	Carboxymethyl-C^{90}	2,163.93	2,163.87	Glu-C
9	100–105	RAWRYC	Carboxymethyl-C^{105}	912.41	912.47	Glu-C
10	[15,23]–[40,47]	SDDKCVFPF–YYWCSVTT	Unmodified Nonreduced	2,076.87	2,077.02	Pepsin
11	[15,25]–[42,49]	SDDKCVFPFIY–WCSVTTYY	Unmodified Nonreduced	2,353.02	2,353.09	Pepsin
12	[32,40]–[56,61]	DCTLHDSTY–CRSTDY	Unmodified Nonreduced	1,795.69	1,795.78	Pepsin
13	[26,34]–[51,59]	KGNLYFDCT–KRWRYCRST	Unmodified Nonreduced	2,313.10	2,313.04	Pepsin
14	[62,70]–[86,96]	ARCALPFIF–SKYWCPVTPNY	Unmodified Nonreduced	2,392.16	2,392.23	Pepsin
15	[64,71]–[88,100]	CALPFIFR–YWCPVTPNYDQDR	Unmodified Nonreduced	2,620.19	2,620.22	Trypsin
16	[60,70]–[86,96]	DYARCALPFIF–SKYWCPVTPNY	Unmodified Nonreduced	2,670.25	2,670.22	Pepsin
17	[71,89]–[103,105]	RGKEYDSCIKEGSVFSKYW–RYC	Unmodified Nonreduced	2,704.27	2,704.19	Pepsin

The suggested sequence has been extensively verified by advanced mass spectral techniques. For this purpose, the exact masses of several peptides prepared by digestion with pepsin were measured (Table 2). Moreover, the entire DQH molecule was also reduced and carboxymethylated or reduced and modified with acrylamide to improve slightly the solubility of the protein in buffers with neutral pH. This procedure allowed the generation of a few additional peptides by digestion with trypsin, endoproteinase Asp-N, or Lys-C. In this way an extensive spectrum of DQH peptide masses could be recorded (Table 2), and the vast majority of the masses obtained corresponded well to those predicted from the sequence suggested by Edman degradation. The only discrepancy was found in the case of peptide 8, the measured mass of which was 2 Da lower than expected from the sequence suggested by Edman degradation. To clarify this discrepancy and get an additional confirmation of the suggested sequence, post-source decay MALDI-MS measurements were performed (Fig. 3). Data of exceptionally good quality were obtained even with the limited amount of the peptide fragments available. In particular, by subtracting the masses of b_{13} minus b_{12} fragments generated from peptide 8 (Table 3), it became apparent that the amino acid in position 94 had to be proline, and not valine as originally suggested (since the mass of proline is 2 Da lower, this reassignment would also account for the above-mentioned discrepancy in the total mass of peptide 8).

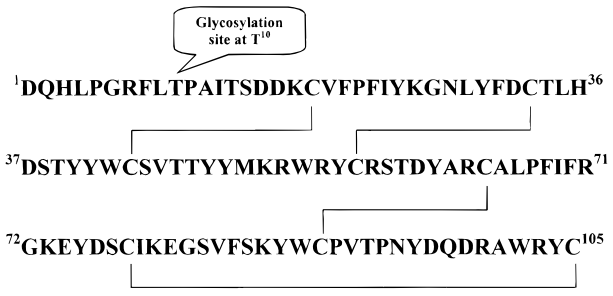

Fig. 2. Complete covalent structure of the major glycoform of DQH sperm surface protein. Disulfide bonds are formed between the following cysteines: 19–43, 33–56, 64–90, and 78–105. Threonine-10 is substituted with a disaccharide: Gal-GalNAc-O-Thr. Additional (minor) glycoforms of this glycoprotein contain GalNAc-O-Thr, Fuc-(Gal-GalNAc)-O-Thr, or NeuAc-(Gal-GalNAc)-O-Thr substitutions.

The theoretical molecular mass of DQH protein devoid of carbohydrates was calculated to be 12,707.0, while the experimental value obtained from ESI-MS measurements was 12,698.8 (Table 1). Since no other substitutions could be found in the protein by any of the techniques employed for its analysis, the difference of 8.2 Da between the two values indicates that all eight cysteines present in DQH sequence may be oxidized, thus forming four disulfide bonds. To establish the pattern of disulfide bond

Table 3. *Post-source decay MALDI-MS data of the most important peptides isolated from DQH sperm surface protein (for peptide numbering and sequence, see Table 2)*

Number	b_1 y_{18}	b_2 y_{17}	b_3 y_{16}	b_4 y_{15}	b_5 y_{14}	b_6 y_{13}	b_7 y_{12}	b_8 y_{11}	b_9 y_{10}	b_{10} y_9	b_{11} y_8	b_{12} y_7	b_{13} y_6	b_{14} y_5	b_{15} y_4	b_{16} y_3	b_{17} y_2	b_{18} y_1
2	n.o.[a]	n.o.	380	493	n.o.	647	804	951	1,065	n.o.	n.o.	n.o.	n.o.	n.o.	—	—	—	—
	—	—	—	—	—	1,817	1,689	1,551	1,438	n.o.	n.o.	1,127	n.o.	n.o.	400	n.o.	233	n.o.
3	n.o.	n.o.	359	532	633	780	877	1,024	1,137	n.o.	1,428	1,485	1,600	1,713	1,876	2,024	—	—
	—	—	—	n.o.	1,811	n.o.	n.o.	1,409	1,262	n.o.	1,018	905	741	n.o.	556	442	328	n.o.
4	n.o.	311	426	600	701	814	951	1,066	1,154	n.o.	n.o.	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	n.o.	1,125	1,010	836	735	622	n.o.	n.o.	283	n.o.
5	n.o.	295	423	579	n.o.	921	1,085	1,260	1,416	1,503	n.o.	1,719	n.o.	—	—	—	—	—
	—	—	—	—	—	—	1,736	1,605	1,477	1,321	n.o.	979	816	641	n.o.	398	297	n.o.
6	175	246	359	456	603	n.o.	864	n.o.	—	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—	—	—	864	n.o.	680	n.o.	435	322	175
7	n.o.	215	342	471	635	750	n.o.	940	1,052	1,181	1,309	n.o.	n.o.	1,553	n.o.	—	—	—
	—	—	—	—	n.o.	1,505	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.	666	n.o.	n.o.	n.o.	n.o.	n.o.
8	n.o.	n.o.	243	390	477	606	769	956	1,117	1,213	1,313	1,414	1,511	1,625	1,788	1,903	2,031	n.o.
	—	n.o.	n.o.	1,921	1,774	1,688	1,560	1,396	1,210	1,049	n.o.	852	751	n.o.	539	n.o.	262	n.o.
9	157	228	414	570	734	895	—	—	—	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—	—	—	—	—	755	686	500	343	n.o.

[a] n.o., Not observed.

pairing, the native (nonreduced) DQH protein was digested by pepsin under acidic conditions that would favor the conservation of disulfide bonds (Schrohenloher & Bennett, 1986). The resulting cystic peptides were purified by HPLC, reduced in sequencer (Brune, 1992). The double sequences were recorded and are shown in the lower part of Table 2. Altogether the sequencing of four of these peptides (peptides 10, 12, 14, and 17) provided an unambiguous pattern of disulfide bond pairing shown in Figure 2. This covalent structure has been further confirmed by MALDI-MS measurements of exact masses of several isolated, disulfide-linked peptides (Table 2). As previously described, conditions may be established for post-source decay MALDI-MS analysis of cystic peptides that would cause minimal fission of the interchain disulfide bonds (Gorman et al., 1997). Using this approach, fragmentation data further confirming our assignments could be recorded. As an example of such data, an extensive series of fragments generated by decomposition of peptide 15 (Table 2) was recorded. These included peaks at m/z 2,447 (loss of R), 2,204 (loss of QDR), 2,187 (loss of IFR), 2,038 (loss of FIFR), 1,924 (loss of YDQDR), and 1,714 (loss of PNYDQDR) fragments followed by those generated by cystine fission (m/z 1,657 and 966). Finally, further fragmentation of both cysteine peptides (extensive assignment of *b* and *y* ions) occurred.

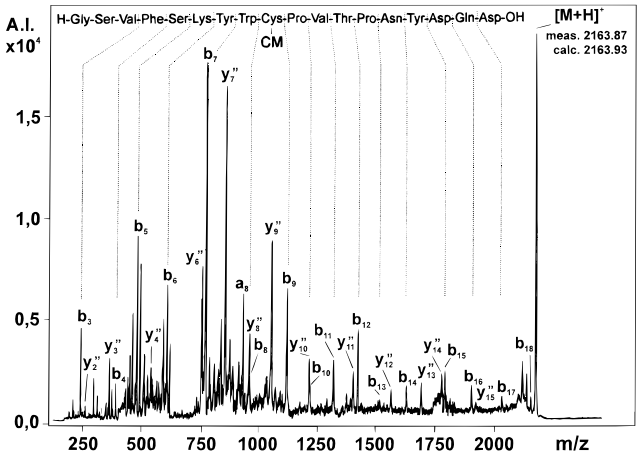

Fig. 3. PSD/MALDI-MS analysis of C-terminal peptide 8 from DQH sperm surface protein (CM = carboxymethylated cysteine).

Search in sequence databases using BLAST algorithm (Altschul et al., 1997) revealed that the DQH protein sequenced here is related to the recently described protein pB1 (Calvete et al., 1997). However, the two sequences clearly represent two different proteins, as evident from differences at five positions throughout the entire sequence (positions 53, 59, 91, 98, and 103). We reason that these differences might reflect the existence of two different, albeit closely related, proteins (pB1 and DQH), allelic polymorphism in the gene coding the porcine pB1 sperm surface protein, or may have occurred due to errors in the original sequence of pB1 protein. The recent deposition into the sequence database of cDNA encoding porcine seminal plasma protein precursor (Plucienniczak et al., 1998) seems to favor the former possibility. Translation of the recently deposited, corrected version of the above cDNA clone into amino acid sequence, and removal of the putative 25 amino acids long N-terminal signal peptide, now provides an identity with the sequence presented here (correction of the sequence done by the original depositors of this cDNA clone consisted of a single change Phe68 → Ser68 (Plucienniczak et al., 1998)). Therefore, until any further confirmation of the original sequence of Calvete et al. (1997) is provided (e.g., by isolating and sequencing the corresponding cDNA clone), we have to assume that the protein sequence reported here (which has been extensively verified by mass spectral techniques) and the cDNA sequence reported by Plucienniczak et al. (1998) both define a novel protein. This new protein is distinguished from pB1 protein by its unique sequence, and by distinct physicochemical and biochemical properties (solubility, resistance to proteases) described in this paper. In accor-

dance with the practice in the field, we suggest to call this novel protein DQH sperm surface protein.

Moreover, we have found that the DQH sperm surface protein may be identical with the cell–cell adhesion inducing protein (pAIF-1) of Hadjisavas et al. (1994), isolated from boar seminal vesicle fluid. N-terminal sequence of DQH protein matches exactly the determined N-terminal sequence (32 amino acid residues) of pAIF-1 (Hadjisavas et al., 1994; Jonáková et al., 1998).

On the other hand, it is obvious that the DQH protein described here and the pB1 protein described previously belong to the same family of seminal plasma proteins as bovine BSP A1, BSP A3 (Seidah et al., 1987), and stallion HSP-1 (Calvete et al., 1995). These proteins are all members of the large family of cell and matrix adhesion proteins, which includes seminal plasma proteins, colagenases, fibronectins, blood coagulation factors, and large cell surface receptors such as the macrophage mannose receptor, and the cation independent mannose-6-phosphate receptors (Hadjisavas et al., 1994). Unlike in seminal plasma proteins, where the tandemly arranged fibronectin type II repeats encompass most of the protein sequence, in the other proteins of this family, this repeat represents only a small portion of the entire protein sequence. The ways in which these simple protein modules participate in the formation of the mosaic structure of large gene product, and the ways in which they may contribute to the functional properties of large receptor proteins, such as the ability to bind various soluble and cellular ligands, will be the subject of the future research. In our laboratory, we have just started the investigation of the structural dispositions of the two fibronectin type II repeats in DQH sperm surface protein using protein modeling and crystallography.

Materials and methods: DQH sperm surface protein was isolated from boar seminal plasma. Boar semen was collected from healthy large white pigs (Insemination Station, Nové Mlýny, Czech Republic), seminal plasma was isolated by centrifugation at $600 \times g$ for 20 min at 5 °C, and stored at −20 °C until used. Plasma was thawed, centrifuged 10 min at 10,000 × *g*, diluted 1:1 with PBS (20 mM phosphate buffer pH 7.3, containing 150 mM NaCl), and 50 mL applied on the heparin polyacrylamide column (24 × 70 mm) pre-equilibrated with the same buffer. Nonadsorbed proteins were washed off with the buffer until the absorbance at 280 nm reached baseline level, and the bound proteins were eluted with sodium chloride (3 M) in the same buffer. Fractions containing heparin-binding proteins were pooled, desalted on Sephadex G-25 (Amersham Pharmacia Biotech, Uppsala, Sweden), eluted in 0.2% acetic acid, and lyophilized. Purification of heparin-binding proteins was achieved by HPLC using Biocompatible Quaternary Gradient System (Waters, Milford, Connecticut). Reversed-phase Vydac C-18 column (218TP 54, 4.6 × 250 mm, 5 μm particle size) was from the separations group (Hesperia, California). Sample of 1 mg in 1 mL of 0.05% trifluoroacetic acid was used, and the separated proteins were eluted at 1 mL/min with a linear gradient of 20–60% acetonitrile in 60 min. Fractions were lyophilized, and each fraction was subjected to SDS gel electrophoresis (Laemmli, 1970).

For chemical deglycosylation, (glyco)peptide 1 encompassing the N-terminal 21 amino acids (300 pmol) was treated with anhydrous HF (incubation for 1 h at 0 °C (Sojar & Bahl, 1987)).

For BNPS-skatol digestion of the reduced and acrylamide-modified (Brune, 1992) DQH sperm surface protein, lyophilized protein (0.4 mg) was dissolved in 100 μL of 80% acetic acid saturated with BNPS-skatol, sealed under nitrogen, and incubated at room temperature for 48 h in the dark. Sample was diluted with Milli-Q water to ~600 μL, excess reagent removed by extraction (three times) with ethylacetate (discarding the organic layer at each extraction), and the extracted water phase evaporated to dryness.

For CNBr cleavage the lyophilized, reduced, and acrylamide modified (Brune, 1992) protein (0.2 mg) was dissolved in 100 μL of 70% formic acid containing 400 μg of CNBr, sealed under nitrogen, and incubated for 24 h at room temperature in the dark. The sample was evaporated, dissolved in 60 μL of 70% formic acid, and redried.

For *Staphylococcus aureus* V8 protease digestion, DQH sperm surface protein (0.4 mg) was reduced, modified with acrylamide (Brune, 1992), and digested in 100 μL of 0.1 M ammonium acetate buffer pH 4.6 with 4 μg (2.8 U) of protease V8 (Type XVII-B) (Sigma, St. Louis, Missouri). The reaction was terminated by lyophilization.

For pepsine digestion 1 mg of native (nonreduced) DQH sperm surface protein was dissolved in 1% acetic acid, and 10 μg of pepsin (Sigma) was added. After 2 h incubation at 37 °C, the reaction was terminated by drying.

Lys-C, Asp-N, and trypsin digestions were performed with reduced and carboxymethylated (Ruegg & Rudinger, 1977) or acrylamide modified (Brune, 1992) DQH sperm surface protein. For Lys-C or Asp-N digestions, 0.4 mg of protein was dissolved in 0.1 M Tris-HCl pH 8.0 containing 0.01% SDS, and 4 μg of the respective sequencing grade protease (Boehringer-Mannheim, Germany) was added. For trypsin digestion, 0.4 mg of protein was dissolved in 0.1 M ammonium bicarbonate pH 8.0 containing 0.01% (w/v) SDS, and 4 μg modified sequencing grade trypsin (Boehringer-Mannheim) was added. Both digestions proceeded for 18 h at 37 °C and were terminated by drying.

After each individual digestion, peptides were separated by reversed-phase high performance liquid on Vydac C-18 column (4.6 × 250 mm, 5 μm particle size) connected to Protein Purification System BioSys500 (Beckman Instruments, Fullerton, California). Peptide mixtures were dissolved in 100 μL of 0.1% trifluoroacetic acid, and the separated peptides were eluted at 1 mL/min with a linear gradient of 0–80% acetonitrile in 60 min. Separated peptides were collected manually and dried before the subsequent analyses.

Amino acid analysis was done after hydrolysis of the protein in 6 N constant-boiling HCl in vacuo. Quantitative carbohydrate analysis was performed with 1 mg of DQH sperm surface protein that was hydrolyzed in 2 N TFA at 120 °C for 2 h, and the hydrolyzate passed through Dowex 50W in H^+ cycle. Evaporated residue was derivatized with 2-aminopyridine (Honda et al., 1989), and separated by capillary electrophoresis (P/ACE 2100) (Beckman Instruments) using 50 μm capillary at 29 kV. Sample was applied with 10 s pressure pulse, separation was performed in 0.2 M borate buffer pH 10.5 at 25 °C, and the separated compounds detected spectrophotometrically at 254 nm. D-Galactose, D-mannose, L-fucose, N-acetyl-D-glucosamine, and N-acetyl-D-galactosamine were used as standards.

Protein and peptide sequencing was performed on Protein Sequencer LF3600D (Beckman Instruments) according to the manufacturer's manual. Cysteine residues were modified in sequencer (Brune, 1992).

Positive ion MALDI mass spectra were measured on a Bruker BIFLEX time-of-flight mass spectrometer (Bruker-Franzen, Bremen, Germany). Somatostatin and cytochrome *c* were used to calibrate the mass range of the instrument. The PSD spectra were

typically recorded in 10–14 segments, each successive segment representing a 20% reduction in reflector voltage. The corresponding precursor ions were selected by FAST^{TM} deflecting pulses. About 200 shots were averaged per segment, and the segments were pasted, calibrated, and smoothed under the computer control of Bruker XTOF 3.0 software.

Electrospray mass spectrum was recorded on double-focusing instrument Finnigan MAT 95 of BE geometry (Finnigan MAT, Bremen, Germany). The sample was dissolved in pure water containing 1% acetic acid and infused into the ion source by syringe pump at a flow rate of 40 μL min^{-1}. The raw acquired data were deconvoluted by the BIOMASS software supplied by the instrument manufacturer.

Note added in proof: The unpublished cDNA sequence of Plucienniczak et al. (1998) is now published: Plucienniczak G, Jagiello A, Plucienniczak A, Holody D, Strzezek J. 1999. Cloning of complementary DNA encoding the pB1 component of the 54-kilodalton glycoprotein of boar seminal plasma. *Mol Reprod Dev 52*:303–309.

Acknowledgments: This work was supported in part by a Program Grant from the Ministry of Education of the Czech Republic (program Renaissance of Science at Czech Universities, grant VS 96141), and by grants No. 303/99/0357 and 524/96/K162 from the Grant Agency of the Czech Republic.

The pB1 protein sequence has been deposited to PIR of NBRF under the accession number A58837.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: New generation of protein database search programs. *Nucleic Acid Res 25*:3389–3402. (Available at www.ncbi.nlm.nig.gov.)

Bork P, Beckmann G. 1993. The CUB domain. A widespread module in developmentally regulated proteins. *J Mol Biol 231*:539–545.

Brune DC. 1992. Alkylation of cysteine with acrylamide for protein sequence analysis. *Anal Biochem 207*:285–290.

Calvete JJ, Mann K, Schäfer W, Sanz L, Reienert M, Nessau S, Raida M, Töpfer-Petersen E. 1995. Amino acid sequence of HSP-1, a major protein of stallion seminal plasma: Effect of glycosylation on its heparin- and gelatin-binding capabilities. *Biochem J 310*:615–620.

Calvete JJ, Raida M, Gentzel M, Urbanke C, Sanz L, Töpfer-Petersen E. 1997. Isolation and characterization of heparin- and phosphorylcholine-binding proteins of boar and stallion seminal plasma. Primary structure of porcine pB1. *FEBS Lett 407*:201–206.

Calvete JJ, Sanz L, Töpfer-Petersen E. 1994. Spermadhesins: Structure-function relationship. *Ass Repr Tech Androl 6*:316–330.

Fernlund P, Granberg LB, Roepstorff P. 1994. Amino acid sequence of β-microseminoprotein from porcine seminal plasma. *Arch Biochem Biophys 309*:70–76.

Gorman JJ, Ferguson BL, Speelman D, Mills J. 1997. Determination of the disulfide bond arrangement of human respiratory syncytial virus attachment (G) protein by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Protein Sci 6*:1308–1315.

Hadjisavas M, Armstrong DT, Seamark RF. 1994. Purification of a cell–cell adhesion regulator from porcine seminal vesicle fluid. *Biochem Biophys Res Commun 205*:1206–1216.

Honda S, Iwase S, Makino A, Fujiwara S. 1989. Simultaneous determination of reducing monosaccharides by capillary zone electrophoresis as the borate complexes of N-2-pyridylglycamines. *Anal Biochem 176*:72–77.

Jonáková V, Kraus M, Veselský L, Čechová D, Bezouška K, Tichá M. 1998. Spermadhesins of AQN and AWN family, DQH sperm surface protein, and HNK protein in the heparin binding fraction of boar seminal plasma. *J Reprod Fertil 114*:25–34.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*:680–685.

Plucienniczak G, Jagiello A, Plucienniczak A, Holody D, Strzezek J. 1998. Unpublished cDNA encoding a porcine seminal plasma protein pB1. Genbank accession AF047026. Deposited on 7th February 1998, corrected version deposited on 15th September 1998.

Ruegg UT, Rudinger J. 1977. Reductive cleavage of cystine disulfides with tributylphosphine. *Methods Enzymol 47*:111–116.

Schrohenloher R, Bennett JC. 1986. Disulfide bonds. In: Darbre A, ed. *Practical protein chemistry*. Chichester, United Kingdom: John Wiley & Sons. pp 149–163.

Seidah NK, Manjunath P, Rochemont J, Sairam MR, Chretien M. 1987. Complete amino acid sequence of BSP-A3 from bovine seminal plasma. Homology to PDC-109 and to the collagen-binding domain of fibronectin. *Biochem J 243*:195–203.

Sojar HT, Bahl OMP. 1987. Chemical deglycosylation of glycoproteins. *Methods Enzymol 138*:341–350.

Tichá M, Kraus M, Čechová D, Jonáková V. 1998. Saccharide-binding properties of boar AQN spermadhesins and DQH sperm surface protein. *Folia Biologica (Praha) 44*:15–21.

Töpfer-Petersen E, Calvete JJ, Sanz L, Sinowatz F. 1995. Carbohydrate- and heparin-binding proteins in mammalian fertilization. *Andrologia 27*:303–324.