Volume 291, number 1, 33–36 FEBS 10221 October 1991

ADONIS 001457939100931V

The amino acid sequence of AQN-3, a carbohydrate-binding protein isolated from boar sperm

Location of disulphide bridges

Libia Sanz[1], Juan J. Calvete[2], Karlheinz Mann[2], Wolfram Schäfer[2], Erich R. Schmid[3] and Edda Töpfer-Petersen[1]

[1]*Department of Dermatology, Andrology Unit, University of München, München, Germany,* [2]*Max-Planck-Institut für Biochemie, W-8033 Martinsried, Germany* and [3]*Institut für Analytische Chemie, Wien, Austria*

Received 22 July 1991

Gamete recognition and adhesion are essential steps in fertilization. Among others, carbohydrate-binding proteins on the sperm surface have been recognized to play a central role in the initial interaction of the male gamete with components of the zona pellucida of the homologous investing oocyte. We have isolated several members of a carbohydrate- and zona pellucida-binding protein family from ejaculated sperm. Here we report the biological origin and structural characterization of AQN-3, a component of this carbohydrate-binding family. The molecular weight of purified AQN-3 was determined by plasma desorption mass spectrometry. The protein was chemically and enzymatically degraded, the proteolytic fragments isolated and characterized by N-terminal sequencing and fast atom bombardment mass spectrometry. In this manner we established the complete amino acid sequence of AQN-3 and the location of its two disulphide bonds. No analogous protein sequence could be found in the MIPS protein sequence data bank, indicating that AQN-3 may belong to a novel mammalian carbohydrate-binding protein family.

Boar sperm; Zona pellucida-binding protein; AQN-3; Mass spectrometry

1. INTRODUCTION

The fusion of sperm and egg during fertilization, which produces the zygote, is preceded by gamete recognition and attachment. First postulated in 1913 by Lillie [1], it is now broadly recognized that specific adhesion molecules located on the surfaces of both gametes are involved in these initial events (reviewed in [2]). In mouse, one of the most studied model species in fertilization, certain oligosaccharide chains attached to the polypeptide core of the zona pellucida glycoprotein 3 (ZP3) provide the primary anchor points for carbohydrate-binding counter-receptor(s) located on the anterior part of the acrosome-intact sperm surface (reviewed in [3]). On the other hand, Bindin, a 24 kDa major protein associated with the inner acrosomal membrane of the sea urchin sperm, is the only yet characterized sperm receptor-binding protein (reviewed in [4]). Bindin has a strong affinity for sulfated, fucose-containing polysacharides of the egg vitelline layer [5], arginine residues being critical for the binding [6]. In other species, putative high- and low-molecular weight sperm proteins possessing zona pellucida-binding site(s) have been identified ([7,8] and references cited therein). In a previous work we have described the isolation and N-terminal amino acid sequence determination of three memberns of a low-molecular weight zona pellucida-binding boar sperm protein family [9]. Since all three proteins begin with the amino acid sequence alanine–glutamine– asparagine [9], we have denominated them 'the AQN family'. Here we report the biological origin, the complete amino acid sequence, and the location of disulphide bridges in AQN-3, using a combination of protein-chemical and mass spectrometric methods. When the AQN-3 amino acid sequence was compared with the protein sequence data bank, no analogous entry was found. Thus, AQN-3 may belong to a new carbohydrate-binding protein family involved in fertilization. Our data may provide clues for further understanding the species-specific mechanism underlaying protein-carbohydrate complementarity in gamete recognition, an aspect of species evolution.

Correspondence address: E. Töpfer-Petersen, Dermatologische Klinik, Department of Dermatology, Andrology Unit, University of Munich, Frauenlobstraße 9/11, W-8000 München 2, Germany.

2. MATERIALS AND METHODS

Ejaculated boar spermatozoa were collected, washed and extracted as previously described [10]. Isolation of AQN-3 from ejaculated boar spermatozoa was done following [9].

Zona pellucida-binding activity measurement, using biotinylated zona pellucida as ligand, and inhibition experiments with fucoidan were performed by immobilizing the isolated protein either on ELISA plates, or onto nitrocellulose sheets by electroblotting, as described [9].

Epididymal fluid was a generous gift from Prof. A.E. Friess (Bern, Switzerland).

N-Terminal sequence analyses were effected in an Applied Biosystems gas-phase sequencer model 473A, following the manufacturer's instructions.

Reduction and alkylation was done as described in [11].

SDS-gel electrophoresis was done according to Laemmli [12].

Enzymatic degradation of the isolated protein (5 mg/ml in 100 mM Tris-HCl, 1 M guanidine-HCl, pH 7.6) was done with endoproteinase Lys-C (Boehringer Mannheim) overnight at 37°C, at an enzyme/substrate ratio of 1:50 (w/w). Alternatively, the reduced and ethypyridylated AQN-3 was degraded with TPCK-trypsin (Sigma) using the same digestion conditions.

Chemical degradation of the protein (10 mg/ml in 70% (v/v) formic acid) after methionine residues was done at room temperature for 4 h using a 100 mg/ml final concentration of cyanogen bromide.

Isolation of peptides was done by reverse-phase HPLC on a Lichrospher RP-100 (Merck, Darmstadt) C18 column (25 × 0.4 cm, 5 μm particle size) eluting at 1 ml/min with a linear stepwise gradient of 0.1% TFA in (A) water and (B) acetonitrile.

Fast bombardment mass spectra were recorded with a mass spectrometer MAT 900 (Finnigan MAT, Bremen) equipped with liquid secondary ion ionization system. Time-of-flight plasma desorption mass spectrometry was done using a Biomol (Uppsala, Sweden) spectrometer with a ^{252}Cf ionization source.

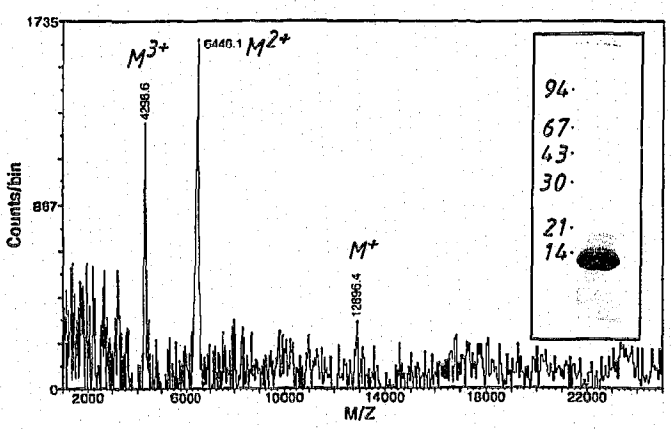

Fig. 1. Determination of the molecular weight of AQN-3 by time-of-flight plasma desorption mass spectrometry using ^{252}Cf ionization. Species corresponding to the molecular ion (M^+), and the di- and tri-protonated (M^{2+} and M^{3+}, respectively) quasi-molecular ions are observed. The spectrum was recorded at an acceleration voltage of 17 kV and corresponded to the average of 20×10^6 primary ions. (Inset) Analysis of purified AQN-3 by SDS-polyacrylamide (10–25%) gel electrophoresis; at the left, molecular weight markers in kDa.

3. RESULTS AND DISCUSSION

3.1. *Isolation, origin, and biological characterization of AQN-3*

AQN-3 was isolated from acid-extracts of ejaculated washed boar spermatozoa as previously described [9]. It consisted of a single polypeptide chain of 12 894.8 ± 1.6 Da, as determined by plasma desorption mass spectrometry (Fig. 1). The isolated protein bound biotinylated zona pellucida. This interaction was inhibited by fucoidan [9], a sulphated heteropolysaccharide which blocks fertilization in a number of mammalian species [2]. It is believed that the interaction of fucoidan with spermatozoa mimics that of the natural substrate (i.e. the zone pellucida) and that the mechanism of binding is similar [13]. Interaction with either zona pellucida or fucoidan was not affected by disulphide bridge reduction indicating that, most probably, a linear sequence of AQN-3 is responsible for its carbohydrate-binding ability. AQN-3 was localized by immunofluorescence microscopy on the sperm head, and the isolated protein was effective in blocking sperm–oocyte binding in vitro (Sanz et al., manuscript in preparation). Recently, Hanqing et al. [14] isolated a group of low molecular weight (14–18 kDa) boar sperm zona pellucida-binding proteins, and reported identical results. Although they did not characterize these proteins, they will include, most probably, the one we describe here. Altogether, these data reinforce our conclusion [9] that the AQN proteins mediate the binding of boar spermatozoa to component(s) of the egg's zona pellucida by a carbohydrate-recognizing mechanism. Their binding site on the egg's exterior coat remains, however, to be identified.

It has been well established that the plasma membrane of mammalian spermatozoa undergoes remodelling by uptake, modification, or loss of macromolecules during epididymal maturation, mixing with seminal vesicle secretions at ejaculation, and during capacitation in the female genital tract [15]. The presence of AQN-3 in male gonadal fluids was investigated using the extraction and isolation procedure described in [9]. All three members of the AQN-family [9] were found in seminal plasma, but not in epididymal fluid, indicating that the members of the AQN-family are not sperm-specific proteins, but secretory components of the male accessory glands being coated to the sperm surface at the time of ejaculation. Since seminal plasma results from admixture of secretions from the various accessory glands (i.e. epididymis, seminal vesicles, and prostate) during seminal emission, the exact biological origin of the proteins of the AQN family deserves further detailed investigation.

3.2. *The primary structure of AQN-3*

AQN-3 was digested with endoproteinase Lys-C (Fig. 2), trypsin and/or cyanogen bromide, the peptides isolated by reverse-phase HPLC and characterized by N-terminal sequencing and Fab mass spectrometry. They provided the complete primary structure of AQN-3 (Fig. 3): N-terminal sequence analysis of reduced and pyridylethylated AQN-3 yielded its first 42 residues (Fig. 3). Two endo-Lys-C peptides (K2, M^+H^+ = 1281.0; and K3) and a CNBr degradations product (CNBr-1) were aligned within this region. A second CNBr-derived fragment (CNBr-2) included peptide K1 ($M+H^+$ = 2050.3) and overlapped with the last 6 residues of the N-terminal sequence and with the first 5 residues of a peptide in K5 (Fig. 3). Finally, the tryptic

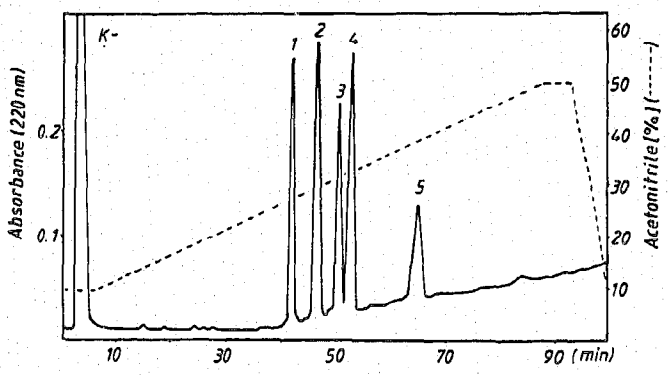

Fig. 2. Isolation of endoproteinase Lys-C proteolytic peptides by reverse-phase HPLC. The column was eluted isocratically with 10% B for 5 min, followed by a linear gradient up to 50% B in 80 min.

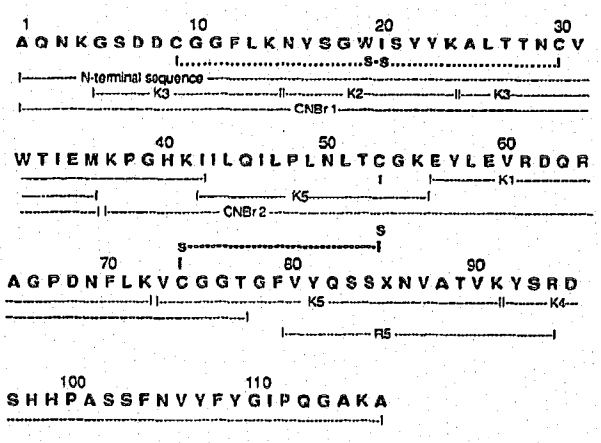

Fig. 3. The complete primary structure of AQN-3, showing the alignment of the sequences obtained by N-terminal sequence analysis of the whole protein, endoproteinase Lys-C peptides (K-), tryptic fragments (R-), and cyanogen bromide degradation products (CNBr-). The unknown residue at position 85 (*X*) has been tentatively assigned as *N*-methylhistidine. s---s indicates the position of disulphide bridges.

peptide R5 overlapped peptides K4 and K5 providing the position of K4 ($M+H^+$ = 2800.4) within the structure of AQN-3, and thus its complete sequence.

Peptide K3 showed two sequences in a 1:1 molar ratio, suggesting the presence of a disulphide bond. Fast atom bombardment mass spectrometry showed that K3 had an $M+H^+$ of 2133.4, in full agreement with its assigned sequence and confirming the presence of a disulphide bridge between cysteines 9 and 30. A second disulphide bridge was found in K5, a peptide showing two equimolar sequences which could only be separated after reduction. Both disulphide bridges are conserved in AWN, an analogous boar sperm zona-binding protein [16].

It may be noticed that in a minor tryptic peptide threonine replaced glycine at position 78. Thus, although the major structure of AQN-3 corresponds to the one shown in Fig. 3, some polymorphism can be anticipated. At position 85 no PTH-derivative of any known amino acid was found. The difference between the molecular weight determined for AQN-3 ($M_{average}$observed = 12 894.8 ± 1.6; Fig. 1) and the calculated for its sequence shown in Fig. 3 ($M_{average}$calculated = 12 743.44 + *X*) indicated that the unknown amino acid (*X*) had a molecular mass of 151.36 ± 1.6. For this molecular mass the only modified amino acid residue that we could find is *N*-methylhistidine. Its presence in AQN-3 and, if so, its possible function, is currently under investigation.

When the sequence of AQN-3 was compared with all others in the MIPS data bank, no analogous entry was found. The structure of AQN-3 is, thus, not related to any other carbohydrate-binding protein characterized so far. We had previously noticed a weak analogy between the N-terminal sequence of the members of the AQN-family and the N-terminal region of the sea urchin sperm bindin [9], however. Whether this analogy is a result of divergent of convergent evolution requires further study.

Carbohydrate-binding sites are usually rich in polar planar side chain- and hydroxyl-containing amino acids, involved in the formation of an intricate hydrogen bond network with the carbohydrate moiety [17]. AQN-3 contains a polypeptide stretch which fulfills these requirements located at its C-terminal region, between G^{76} and G^{109}. The availability of its primary structure may be useful for the characterizartion of the binding site(s) for carbohydrate(s). The identification and characterization of the zona pellucida counter-receptor carrying the carbohydrate structure(s) recognized by AQN-3 is another challenge for future investigation.

Acknowledgements: The provision of a Max-Planck-Gesellschaft fellowship to J.J.C. is gratefully acknowledged. This work was supported by the Deutsche Forschungsgemeinschaft (To 114/1-2 grant to E.T.-P.) and Grant BMFT 0318824A (W.S.).

REFERENCES

[1] Lillie, F.R. (1913) Science 38, 524–528.
[2] Yanagimachi, R. (1988) in: The Physiology of Reproduction (Knobil, E. and Neill, J.D. eds.) pp. 135–185, Raven, New York.
[3] Wassarman, P.M. (1990) Development 108, 1–17.
[4] Minor, J.E., Gao, B. and Davidson, E.H. (1989) in: The Molecular Biology of Fertilization (Schatten, H. and Schatten, G. eds.) pp. 73–88, Academic Press, San Diego.
[5] DeAngelis, P.L. and Gable, C.G. (1987) J. Biol. Chem. 262, 13946–13952.
[6] DeAngelis, P.L. and Gable, C.G. (1988) Biochemistry 27, 8189–8194.
[7] Macek, M.B. and Shur, B.D. (1988) Gamete Res. 20, 93–109.
[8] O'Rand, M.G. (1988) Gamete Res. 19, 315–328.
[9] Jonáková, V., Sanz, L., Calvete, J.J., Henschen, A., Cèchova, D. and Töpfer-Petersen, E. (1991) FEBS Lett. 280, 183–186.
[10] Cèchova, D., Töpfer-Petersen, E. and Henschen, A. (1988) FEBS Lett. 241, 136–140.
[11] Henschen, A. (1986) in: Advanced Methods in Protein Sequence Analysis (Wittmann-Liebold, B., Salnikov, J. and Erdmann, A.V. eds.) pp. 244–255, Springer-Verlag, Berlin.

[12] Laemmli, U.K. (1970) Nature 227, 680–685.
[13] Jones, R. and Williams, R.M. (1990) Development 109, 41–50.
[14] Hanking, M., Tai-Hing, Y. and Zhao-Wen, S. (1991) Mol. Reprod. Dev. 28, 124–130.
[15] Russell, L.D., Peterson, R.N., Hunt, W. and Strack, L.E. (1984) Biol. Reprod. 30, 959–978.
[16] Sanz, L., Calvete, J.J., Schäfer, W., Mann, K. and Töpfer-Petersen, E. (1991) Biochim. Biophys. Acta (submitted).
[17] Quiocho, F.A. (1988) Curr. Top. Microbiol. Immunol. 139, 135–148.