Complete localization of the disulfide bridges and glycosylation sites in boar sperm acrosin

E. Töpfer-Petersen[1], J. Calvete[2,3], W. Schäfer[3] and A. Henschen[3,4]

[1] *Department of Dermatology, Andrology Unit, University of Munich, Frauenlobstr. 9/11, 8000 Munich 2, FRG,* [2] *Institute of Physical Chemistry C.S.I.C., Madrid, Spain,* [3] *Max-Planck Institute of Biochemistry, 8033 Martinsried, FRG* and [4] *Institute of Molecular Biology and Biochemistry, University of California, Irvine, USA*

Received 28 August 1990; revised version received 4 October 1990

Acrosin is a disulfide-bonded two-chain glycoprotein, which belongs to the serine proteinase family and which plays a central role in mammalian fertilization. The amino acid sequence of acrosin from different species has been recently derived by cDNA analysis. Boar sperm acrosin contains twelve cysteine residues forming two interchain and 4 intrachain disulfide bonds. Protein-chemical and mass-spectroscopic analyses of fragments and subfragments obtained by proteolytic and chemical degradation of the isolated protein allowed the unambiguous localization of all disulfide bridges and glycosylation points in boar acrosin. The 12 cysteines and the glycosylated asparagines in the porcine enzyme are absolutely conserved in number and position within all known acrosin sequences. Thus, the disulfide bond and glycosylation patterns outlined here are conserved during evolution and may be important for enzyme function.

Pro/acrosin; Disulfide bridge; Glycosylation; Amino acid sequence; Boar spermatozoa; Sequence homology

1. INTRODUCTION

Acrosin (EC 3.4.21.10) is a serine proteinase of trypsin-like specificity [1,2] which has been involved in the complex events of sperm-egg interactions by its carbohydrate binding sites [3–5] and in the penetration process through the oocyte zona pellucida by specific and limited proteolysis of the glycoprotein matrix [6,7]. Acrosin is synthesized in early round spermatids in a single-chain polypeptide form, proacrosin, and stored in the acrosomal vesicle overlaying the anterior part of the sperm head [8–10]. It is believed that proacrosin is activated to the mature enzyme by limited autoproteolysis [11] at the time of the acrosome reaction [12] and discharged from the acrosome in the near proximity of the investing egg [13]. The mature enzyme consists of a 23-amino acid residue light chain covalently linked by two disulfide bridges [11] to the heavy chain (M_r 37000) containing the active center of the serine proteinase family [14–17]. The primary structure of mouse, human, and boar acrosin have been deduced by nucleotide sequencing of cDNA clones [14–17]. The boar proacrosin cDNA encodes a polypeptide of about 400 amino acids including both, the light and the heavy chains [15,16]. Alignment of the acrosin heavy chain sequence comprising its first 250 residues with the sequence of other serine proteinases like trypsin [18], chymotrypsin [19], elastase [20], kallikrein [21] and prothrombin [22,23] shows a sequence similarity of about 55% including the location of the catalytic triad essential for proteolytic activity. The light chain on the other hand shows sequence similarity with the activation peptides of other serine proteinases [11]. A feature of acrosin not observed in other serine proteinases is a 125-residue C-terminal extension containing a stretch of 23 consecutive proline residues. Recently, Baba et al. [14,15] have postulated an activation and maturation mechanism for the boar acrosin zymogen including both a proteolytic processing in the N-terminal region by cleavage of the Arg-23–Val-24, leading to the formation of the light and heavy chain [24] and liberation of certain polypeptide segments from the C-terminus of the heavy chain including the proline-rich segment [14,15,25].

In this communication protein-chemical studies of the isolated active boar acrosin are presented to establish its chemical identity such as its polypeptide length and the post-translational modifications, i.e. glycosylation sites and disulfide bridges which cannot be deduced from the cDNA analysis.

Correspondence address: E. Töpfer-Petersen, Department of Dermatology, Andrology Unit, University of Munich, Frauenlobstr. 9/11, D-8000 Munich 2, FRG

Abbreviations: HPLC, high-performance liquid chromatography; SDS-PAGE, sodium dodecylsulfate polyacrylamide gel electrophoresis; CNBr, cyanogen bromide; BAPA, N^{α}-benzoyl-L-arginine *p*-nitroanilid

2. MATERIALS AND METHODS

2.1 *Materials*

TPCK-trypsin and porcine pancreas elastase were purchased from Sigma and Boehringer, respectively. CNBr was from Merck.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

2.2. Methods

High-molecular mass acrosin (55–53 kDa) was purified from acidic boar sperm extracts as recently described [4]. Amidolytic activity was determined according to [26] as the hydrolysis of the chromogenic substrate BAPA. SDS-PAGE was performed on 7.5–20% polyacrylamide gradient slab gels (0.75 × 80 mm) as previously described [4]. Amino acid and amino sugar analysis were carried out on a Biotronik analyzer after hydrolysis at 110°C in 6 N HCl for 24 h or in 4 N HCl for 4 h, respectively.

Boar acrosin (1 mg/ml) in 50 mM ammoniumhydrogen carbonate, pH 8.0, 1 M guanidine hydrochloride was digested with TPCK-trypsin (Sigma) at an enzyme/protein ratio of 1:25 (w/w) for 18 h at 37°C. The reaction was stopped by freeze-drying. The cleavage mixture was either analyzed by reverse-phase HPLC or further cleaved with CNBr (100 mg/ml) in 70% formic acid under nitrogen and in the dark. After 4 h at room temperature the mixture was diluted with water and freeze-dried. Both tryptic and CNBr degradation products were separated by reverse-phase HPLC on a Lichrosphere RP-100 column (5 μm, 0.4 × 25 cm; Merck) using a linear H_2O/acetonitrile gradient in 0.1% (v/v) trifluoroacetic acid. Subdigestion of isolated peptides was performed with elastase in 100 mM ammoniumhydrogen carbonate, pH 8.0, 1 mM $CaCl_2$ at a peptide concentration of 0.2 μg/μl and an enzyme/peptide ratio of 1:20 (w/w) for 18 h at 37°C. After acidification with formic acid, peptides were isolated by reverse-phase chromatography as described above.

N-terminal sequence determination was performed using a Beckman 890C spinning cup sequenator. Phenylhydantoin derivatives were identified following [27]. Mass spectra were recorded with a mass spectrometer MAT 900 (Finnigan MAT) equipped liquid SIMS ionization.

3. RESULTS AND DISCUSSION

Boar sperm acrosin was purified to homogeneity as previously described [4]. This preparation shows a single broad band of apparent molecular mass of 55000 by SDS-PAGE under non-reducing conditions, exhibits amidolytic activity (6 U/mg protein) measured as the hydrolysis of the chromogenic substrate BAPA [26], and contains in an equimolar ratio the N-terminal sequences corresponding to the light and the heavy chain [4].

This two-chain acrosin preparation was digested with TPCK-trypsin. The resulting tryptic peptides were isolated by reverse-phase HPLC (Fig. 1) and characterized in terms of N-terminal sequence and amino acid and amino sugar composition. Among the peptides characterized (Fig. 2), peptide 18 contains the proline-rich domain and peptides 9, 4 and 7 correspond to the polypeptide stretches within the very C-terminal region of the acrosin cDNA-encoded message [15]. Our results indicate that the two-chain full-length acrosin molecule is an active enzyme. This conclusion is in contrast to the hypothesis of Baba et al. [15] that activation of proacrosin is achieved by both the cleavage of the Arg-23–Val-24 bond yielding the two-chain molecule, and by the stepwise loss of the C-terminal polypeptide stretches including the proline-rich segment. Our current hypothesis is that proacrosin forms an active high-molecular intermediate containing the full-length heavy chain, which then undergoes a maturation process leading to the stable form of the enzyme (β-acrosin).

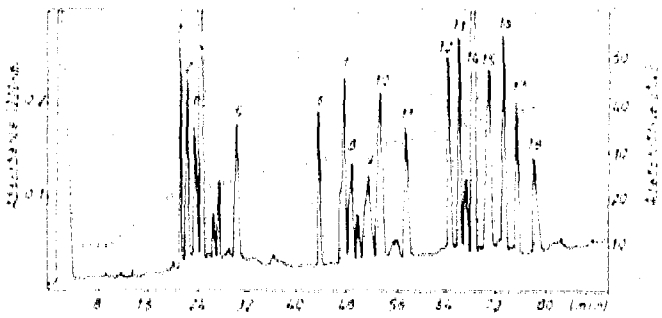

Fig. 1. Separation of the tryptic digest of high-mass acrosin by reverse-phase chromatography on a Lichrosphere RP 100 column (5 μm, 0.4 × 25 cm) in 0.1% trifluoroacetic acid, 0–60% acetonitrile.

This C-terminal processing may have an important biological function in fertilization.

Recently, two different groups have sequenced cDNA clones from boar testis cDNA libraries encoding the complete protein sequence of preproacosin [15,16]. Their results, however, are not identical. In this respect our protein sequence analyses are in agreement with the sequence reported by Baba et al. [15].

Boar acrosin [15,16], as well as acrosin from mouse [17] and human [14] contains 12 cysteine residues forming part of two interchain [11] and four intrachain disulfide bonds. The cysteines at positions 55, 71, 159, 191, 207, 218, 228, and 248 of the chain are conserved among the serine proteinase family members, while the

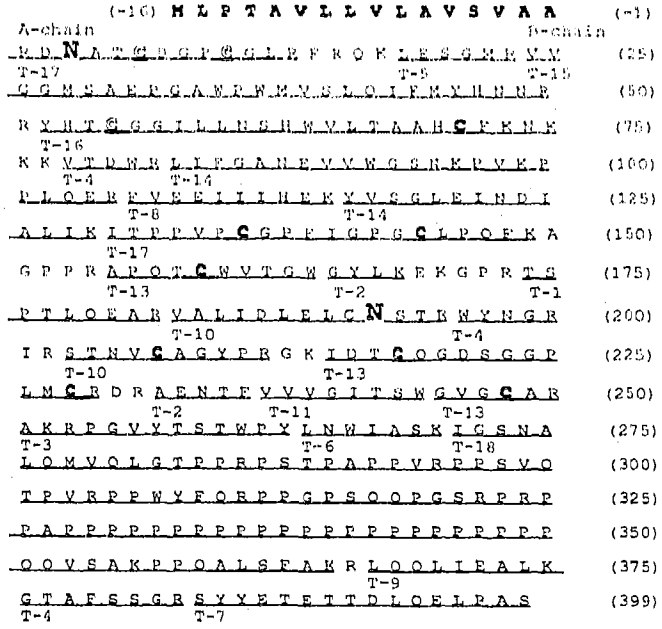

Fig. 2. Amino acid sequence of boar preproacrosin deduced by cDNA analysis [15] showing underlined the position of the tryptic peptides isolated in Fig. 1. The signal sequence is outlined, cysteine residues are shadowed, and the glycosylated asparagine residues are shown in boldface. The N-terminals of the A- and B-chains are indicated.

cysteine residues at position 10 in the light chain and 144 in the heavy chain are also present in chymotrypsin [18] and thrombin [22,23]. The remaining two cysteines, Cys-6 and Cys-136, in the light and heavy chain, respectively, have only been found in acrosin. Although it has been suggested that the known cysteine pairing in other serine proteinases will most probably be conserved in acrosin [15], the precise arrangement of these disulfide bonds have not been further investigated.

In order to identify the disulfide bond pattern, boar sperm acrosin was digested with trypsin. The tryptic peptides were isolated by reverse-phase HPLC (Fig. 1), and characterized in terms of N-terminal sequence and amino acid and amino sugar analysis. In this way, single or groups of two disulfide bridges could be isolated: HPLC fraction 10 (Figs 1,2) contained two peptides crosslinked by a single bridge between Cys-191 and Cys-207; fraction 16 corresponded to the polypeptide stretch 52-75 containing a loop between Cys-55 and Cys-71; fraction 13 showed three sequences in equimolar ratio indicating that Cys-159, Cys-218, Cys-228 and Cys-248 are linked together by two disulfide bonds. When acrosin was digested with trypsin and then degraded with CNBr, the peptides 155-168 and 215-227 disulfide-bonded to the peptides 228-229 and 237-250, respectively, were found (data not shown). This clearly indicates that Cys-159 is linked to Cys-228 and that Cys-218 forms a bridge with Cys-248. Fraction 17 corresponded to the first 13 residues of the light chain covalently linked by two disulfide bonds to fragment 130-149, localized within the heavy chain. To unambiguously locate the position of the two interchain disulfide bonds, the isolated fraction 17 was subdigested with elastase, an enzyme of broad specificity.

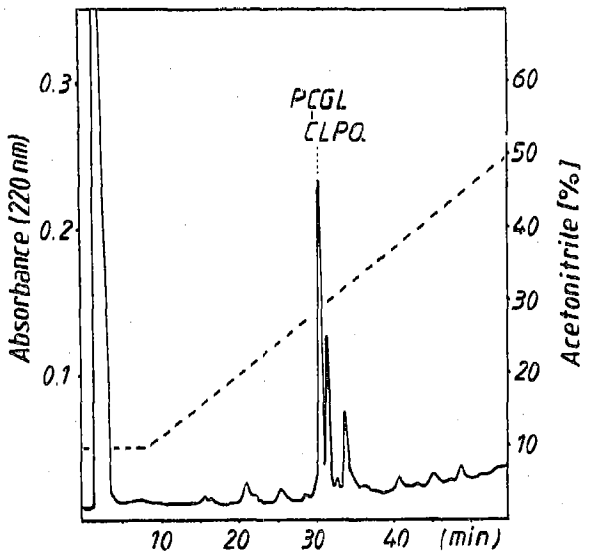

Fig. 3. Separation of the elastase digest of fraction 17 (from Fig. 1) by reverse-phase chromatography. Chromatographic conditions as in Fig. 1.

After HPLC separation of the digestion products (Fig. 3) a peptide containing the polypeptide stretch 9-12 single disulfide-bonded to the stretch 144-147 was characterized. This establishes, that Cys-10 and Cys-144 actually form a bridge, and thus, the remaining interchain disulfide bond, between Cys-6 and Cys-136, can be deduced by exclusion.

Our results establish the complete arrangement of the 6 disulfide bonds of boar acrosin (Fig. 4) and, presumably also in its mouse and human counterparts, since the 12 cysteine residues are absolutely conserved. As expected the disulfide bridge pairing follows the classic pattern found within the members of the serine proteinase family pointing to its essentiality for maintaining the appropriate conformation of its active site. It is worth noticing, that in human thrombin which through X-ray studies has been recently resolved [28] the residues occupying analogous sequence positions as the acrosin Cys-6 and Cys-136 are 9.0 Å apart. This distance may allow disulfide connection (disulfide Cα–Cα distance ~5.3 Å) taking into account that the N-terminal region of human thrombin is located in a highly disordered part of the molecule (W. Bode, personal communication). This is the same case as in chymotrypsin [29] and elastase [30], where the two disulfide bridges present in trypsin but absent in

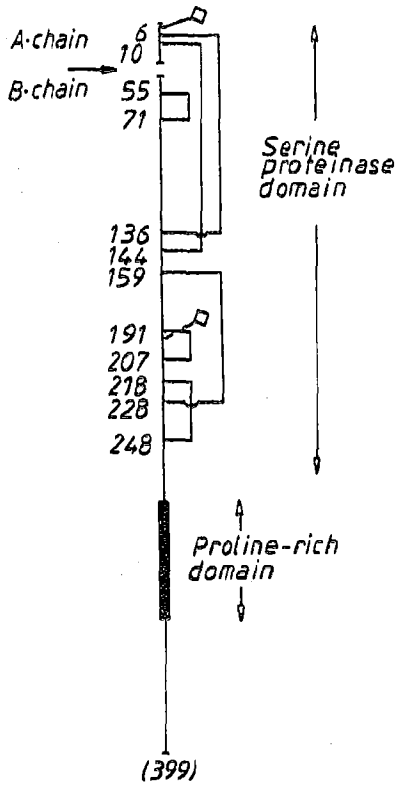

Fig. 4. Schematic outline of the arrangement of the disulfide bridge pattern and *N*-glycosylation sites (—◇) in boar acrosin. Numbers at the left correspond to the cysteine positions in proacrosin. The arrow indicates the cleavage site between Arg-23-Val-24.

chymotrypsin and elastase can be built into the atomic model of the latter enzymes with little or no distortion of the polypeptide chain.

It has been reported that boar acrosin contains only one carbohydrate side chain attached to Asn-3 in the light chain [11]. Our data, however, clearly showed that peptide 10 of the heavy chain (Fig. 2), comprising the typical sequon Asn-Xaa-Thr, contained 3 mol glucosamine/mol peptide. Moreover, when peptide-10 was analyzed by fast atom bombardment mass spectrometry a $M+H^+$: 3596.2 was found (data not shown), but no signal at $M+H^+$: 2511.2, which would conform to the non-glycosylated peptide. This difference in molecular mass could correspond to an incomplete biantennary oligosaccharide unit- like GlcN-GlcN-$(Man)_3$-GlcN. The exact nature of the oligosaccharide chains in boar acrosin needs further investigation. As the two possible *N*-glycosylation points in boar acrosin are also conserved in the mouse [17] and human [14] enzymes, we predict that these asparagines will be also glycosylated in the latter molecules.

Acknowledgements: This work was supported by the Deutsche Forschungsgemeinschaft (To114/1-1) and BMFT (PTB 038824).

REFERENCES

[1] Schleuning, W.-D. and Fritz, H. (1974) Hoppe-Seyler's Z. Physiol. Chem. 355, 125–130.
[2] Polakoski, K.L. and McRorie, R.A. (1973) J. Biol. Chem. 248, 8183–8188.
[3] Töpfer-Petersen, E. and Henschen, A. (1987) FEBS Lett. 226, 38–42.
[4] Töpfer-Petersen, E. and Henschen, A. (1988) Biol. Chem. Hoppe-Seyler 369, 69–76.
[5] Jones, R., Brown, C.R. and Lancaster, R.T. (1988) Development 102, 781–792.
[6] Dunbar, B.S., Dudkiewicz, A.B. and Bundman, D.S. (1985) Biol. Reprod. 32, 619–630.
[7] Urch, U.A., Wardrip, N.J. and Hedrick, J.L. (1985) J. Exp. Zool. 233, 479–483.
[8] Phi-van, L., Muller-Esterl, W., Florke, S., Schmid, M. and Engel, W. (1983) Biol. Reprod. 29, 479–486.
[9] Florke, S., Phi-van, L., Muller-Esterl, W., Scheuber, H.P. and Engel, W. (1983) Differentiation 24, 250–256.
[10] Müller-Esterl, W. and Fritz, H. (1981) Methods Enzymol. 80, 621–632.
[11] Fock-Nuzel, R., Lottspeich, F., Henschen A. and Muller-Esterl, W. (1984) Eur. J. Biochem. 14, 441–446.
[12] Polakoski, K.L. and Parrish, R.F. (1977) J. Biol. Chem. 252, 1888–1894.
[13] Green, D.P.L. (1978) J. Cell Sci. 32, 153–164.
[14] Baba, T., Michikawa, Y., Kawakura, K. and Arai, Y. (1989) FEBS Lett. 244, 132–136.
[15] Baba, T., Kishiwabara, S.-I., Watanabe, K., Michikawa, Y., Kimura, K., Takada, M., Fukamizu, A. and Arai, Y. (1989) J. Biol. Chem. 264, 11920–11927.
[16] Adham, I.M., Klemm, U., Maier, W.-M., Hoyer-Fender, S., Tsaousidou, S. and Engel, W. (1989) Eur. J. Biochem. 182, 563–568.
[17] Klemm, U., Maier, W.-M., Tsaousidou, S., Adham, I.M., Willison, K. and Engel, W. (1990) Differentiation 42, 160–166.
[18] Mikes, O., Holeysovsky, V., Tomasek, V. and Sorm, F. (1966) Biochem. Biophys. Res. Commun. 24, 346–352.
[19] Brown, J.R. and Hartley, B.S. (1966) Biochem. J. 101, 214–228.
[20] Shotton, D.M. and Hartley, B.S. (1973) Biochem. J. 131, 643–675.
[21] Tschesche, H., Mair, G., Godec, G., Friedler, F., Ehret, W., Hirschauer, C., Lemon, M., Fritz, H., Schmidt-Kastner and Kutzbach, C. (1979) Adv. Exp. Med. Biol. 120, 245–260.
[22] MacGillivray, R.T.A and Davie, E.M. (1984) Biochemistry 23, 1626–1634.
[23] Degen, S.J.F., MacGillivray, R.T.A. and Davie, E.W. (1983) Biochemistry 22, 2087–2097.
[24] Cechova, D., Töpfer-Petersen, E. and Henschen, A. (1988) FEBS Lett. 241, 136–140.
[25] Zelezna, B., Cechova, D. and Henschen, A. (1989) Biol. Chem. Hoppe Seyler 370, 323–327.
[26] Schleuning, W.-D. and Fritz, H. (1976) Methods Enzymol. 45, 89–95.
[27] Lottspeich, F. (1980) Hoppe-Seyler's Z. Physiol. Chem. 361, 1829–1834.
[28] Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S.R. and Hofsteenge, J. (1989) EMBO J. 8, 3467–3475.
[29] Sigler, P.B., Blow, D.M., Matthews, B.W. and Henderson, R. (1968) J. Mol. Biol. 35, 143–164.
[30] Shotton, D.M. and Watson, H.C. (1970) Nature 225, 811–816.