COMPLETE AMINO ACID SEQUENCE OF BOAR PROTAMINE

TOHRU TOBITA, HIROYUKI TSUTSUMI, AKEMI KATO, HIROMASA SUZUKI, MINORU NAMOTO, MINORU NAKANO and TOSHIO ANDO

Department of Chemistry, Faculty of Science, Chiba University, Chiba City 260 (Japan)

(Received September 30th, 1982)
(Revised manuscript received January 4th, 1983)

Key words: Protamine; Amino acid sequence; Sequence homology; (Boar sperm nucleus)

The complete amino acid sequence of boar protamine was determined. In order to specify cleaving points for enzymatic fragmentation, two S-alkyl derivatives of the protamine were used. 1. The sequences other than arginine stretches were determined by analysis of tryptic peptides from the S-carboxymethyl-protamine. The sequence of the central arginine-clustered region was determined by stepwise degradation from the C-terminus of the thermolysin core-fragment with carboxypeptidase A and B, and then with acid carboxypeptidase. 2. S-Methylation of cysteine residues of the protamine was found to introduce a new point for thermolysin and chymotryptic cleavage. Thus, thermolysin cleavage of the chymotryptic core-fragment resulted in oligo-arginine peptides with methylcysteine at the N-terminus, by which the sequence of arginine-clusters was confirmed. 3. The complete sequence was deduced by overlap of these sequenced peptides, which is: Ala-Arg-Tyr-Arg-Cys₂-Arg-Ser-His-Ser-Arg-Ser-Arg-Cys-Arg-Pro-Arg₄-Cys-Arg₆-Cys₂-Pro-Arg₅-Ala-Val-Cys₂-Arg₂-Tyr-Thr-Val-Ile-Arg-Cys-Arg₂-Cys. Boar protamine, as well as bull protamine, is composed of the less basic amino- and carboxy-terminal domains and the central arginine-clustered domain, and 80% homology was found between boar and bull protamines.

Introduction

Protamine has been found in the sperm nuclei of many but not all fish and domestic fowl [1]. These protamines are very rich in arginine, which amounts to about 70% of total amino acid residues, but are devoid of cysteine. During the final stage of spermatogenesis, these protamines displace directly somatic histones in the nucleoprotein complex [2,3], thus forming a more condensed structure than that of somatic chromatin [4,5].

In mammalian species, cysteine-rich protamine has been found in bull, boar, ram, rat, stallion, guinea pig, rabbit and human [6-15]. During spermatogenesis these protamines associate ultimately with DNA through transition of several nuclear basic proteins including somatic histones, testis-specific histones and spermatidal basic proteins [16,17]. Thereafter, the cysteine residues in the nucleoprotamine complex are completely cross-linked to form a tight network structure during transit of spermatozoa in epididymis [18-20].
Among the mammalian protamines, the bull protamine is the only one of which the complete primary structure has been determined [7]. More sequential studies are required for a better understanding of characteristics of mammalian protamine and for elucidation of its mode of interaction with DNA.

We report the complete amino acid sequence of boar protamine, which is compared with that of bull and the available partial sequences of some other mammalian protamines [7-10]. As a whole, boar and bull protamines show 80% homology, and a general feature of the mammalian protamines together with those of fish and fowl protamines is discussed.

Materials and Methods

Protamines. S-Carboxymethyl-protamine was prepared from sperm nuclei from boar semen (Yorkshire) by our methods reported previously [21]. S-Methyl-protamine was also prepared according to the same procedure used for S-carboxymethyl-protamine [21] except for reduction and alkylation of the nuclei. Sperm nuclei were reduced with dithiothreitol (100 mol/mol half-cystine in the protein), and liberated sulfhydryl groups were methylated with methyl-p-nitrobenzenesulfonate (a 1.2-fold molar ratio to the dithiothreitol) in the presence of 50% (v/v) acetonitrile.

Enzymes. Trypsin (EC 3.4.21.4) and chymotrypsin A (EC 3.4.21.1) were products of Worthington. Thermolysin (EC 3.4.24.4) was purchased from Seikagaku-Kogyo. Carboxypeptidase A (EC 3.4.17.1) and B (3.4.17.2) were isolated from porcine pancreas glands according to a modification of the methods of Folk and Schirmwer [22]. Acid carboxypeptidase (peptidyl-L-amino-acid hydrolase, EC 3.4.16.1) was kindly donated by Professor M. Yamasaki (College of General Education, University of Tokyo) [23,24].

Other chemicals. CM-32 and Bio-Rex 70 were purchased from Whatman and Bio-Rad, respectively. Other reagents used were of special grade.

Fragmentation of the protein and separation of the peptides

Tryptic peptides of S-carboxymethyl-protamine. Protein (2.0 μmol) was hydrolyzed with trypsin (4.1 nmol) in 0.05 M sodium borate buffer (pH 8.0, 4.5 ml) for 24 h at 25°C. The reaction mixture was chromatographed on Bio-Rex 70 at pH 7.5. The passed-through peptides were further separated on a Hitachi amino acid analyzer (KLA 3B) at pH 4.25.

Thermolysin-peptides of S-carboxymethyl-protamine. Protein (4.0 μmol) was digested with thermolysin (20 nmol) in 5.2 mM CaCl₂/0.025 M Tris-HCl (pH 8.0, 10 ml) for 40 h at 25°C. The digest was brought to pH 5.2 with 0.1 M HCl and chromatographed on CM-32 at pH 5.3. Fractions containing two or more peptides were further separated by chromatography on Bio-Rex 70 under the same conditions as with the tryptic peptides.

Carboxypeptidase A and B digestion of the thermolysin core-fragment. The core-fragment (1.8 μmol) resulting from the thermolysin digestion was hydrolyzed with both carboxypeptidase A (8.2 nmol) and B (9.0 nmol) in 0.5 M NaCl/0.05 M Tris-HCl (pH 8.0, 3 ml) for 18 h at 25°C. After incubation for 2, 7 and 18 h, an aliquot of the reaction mixture was withdrawn, and the released amino acids were determined by amino acid analysis. The final digest was applied to a column (1 × 4 cm) of Bio-Rex 70 equilibrated with 0.1 M sodium acetate buffer (pH 5.2). The released amino acids were eluted with 0.2 M acetic acid and the remaining core-fragment with 0.1 M HCl, step-wise. The latter fraction was lyophilyzed for acid carboxypeptidase digestion.

Partial hydrolysis of the core-fragment with acid carboxypeptidase. The core-fragment (1.6 μmol) obtained from carboxypeptidase digestion was further hydrolyzed with acid carboxypeptidase (6.0 nmol) in 0.1 M NaCl/0.05 M sodium acetate buffer (pH 5.0, 2 ml) for 30 min at 25°C. The digest was boiled to inactivate the enzyme and chromatographed on CM-32 at pH 5.3.

Limited chymotryptic hydrolysis of S-methyl-protamine. Protein (5.0 μmol) was limitedly hydrolyzed with chymotrypsin A (2.5 nmol) in 0.02 M sodium phosphate buffer (pH 8.0, 5 ml) for 15 min. After digestion was stopped by boiling of the digest for 5 min, the reaction mixture was brought...
to pH 5.0 with HCl and chromatographed on CM-32 at pH 5.5.

**Thermolysin peptides from the chymotryptic core-fragment.** The core-fragment (1.4 μmol) from the S-methyl-protamine was digested with thermolysin (20 nmol) in 5 mM CaCl₂/0.025 M Tris-HCl (pH 8.0, 6 ml) for 24 h at 25°C. After completion of digestion, the reaction mixture was brought to pH 6.0 with HCl and chromatographed on CM-32 at pH 6.0.

**Peptide sequencing.** All the purified peptides obtained from the enzymatic digestion were sequenced as follows. The amino acid compositions of the peptides were determined on an amino acid analyzer after hydrolysis of the peptide in 6 M HCl at 110°C for 24 h in vacuo. The N-terminal amino acids were determined by the dansyl and dinitrophenyl methods [25,26]. The sequences of most of the peptides were determined by the alternate carboxypeptidase A and B digestion from the C-terminus, which was described in the previous paper [21]. The N-terminal sequence of some peptides from the S-methyl-protamine was determined by the dansyl-Edman method [27]. Proline residue at the C-terminus was confirmed by hydrazinolysis [21].

**Results and Discussion**

For the enzymatic fragmentation of the molecule, two S-alkyl derivatives of the protamine, S-carboxymethyl- and S-methyl-protamines were used.

**S-Carboxymethyl-protamine**

The peptide bonds involving carboxymethylcysteine are resistant to tryptic and thermolysin hydrolysis, and so the fragments resulting from their digestion preserved the sequence around half-cystine residues.

The sequences of tryptic peptides of S-carboxymethyl-protamine correspond to those of non-basic amino acids between arginines or arginine-clusters, including half-cystine and proline residues. These sequences are shown in part A of Fig. 1. *See also supplementary data (see footnote on p. 141).*

Cleavage of S-carboxymethyl-protamine with thermolysin gave rise to a large fragment containing the centered arginine-clusters (residues 10–35) from the middle region and small fragments from the N-terminal and the C-terminal regions. The large fragment was completely degraded with carboxypeptidases A and B. Arginines (5.0 mol/mol) were released step-wise from the C-terminus and the core-fragment (residues 10–30) with proline at the C-terminus remained. This core-fragment was further subjected to partial hydrolysis of acid carboxypeptidase that releases the C-terminus proline and other common amino acids. Proline (0.88 mol/mol), carboxymethyl-cysteine (2.92 mol/mol) and arginine (8.20 mol/mol) were released and three core-peptides, of which the N-terminus began with Ser-Arg-Ser-Arg- (residues 10–17 and 10–21) or Ser-Arg- (residues 12–17), were left behind. The N-terminal difference of these peptides is ascribed to partial thermolysin cleavage at alternative peptide bonds of the two serine residues (residues 10 or 12). These results are compiled in part B of Fig. 1.

**S-Methyl-protamine**

In contrast to S-carboxymethyl-protamine, S-methyl-protamine was found to be hydrolyzed at methylcysteine residues with thermolysin and chymotrypsin A as a result of methylation of cysteine residues, which was very useful for obtaining fragments comprising an arginine-cluster.

In confirmation of the sequence of the arginine-clusters, S-methyl-protamine was limitedly digested with chymotrypsin A. The large fragment (residues 4–38), Ala-Arg-Tyr (residues 1–3) from the N-terminus, Thr-Val-Ile-Arg-Cys(Me)-Arg₂-Cys(Me) (residues 43–50) and Thr-Val-Ile-Arg-Cys(Me)-Arg₂ (residues 43–49) from the C-terminal region, and Cys(Me)-Arg₂-Tyr (residues 39–42) were released. The last peptide arose from the large fragment ending with tyrosine (residues 4–42) by preferential cleavage at Cys(Me)-Cys(Me) (residues 38–39) bond with chymotrypsin A under the present limiting conditions. These results are consistent with those of chymotryptic peptides from S-carboxymethyl-protamine reported previously [21].

Thereafter, the chymotryptic core-fragment (residues 4–38) was completely digested with thermo-
lysine, yielding oligo-arginine fragments, Cys(Me)-Arg-Pro-Arg₄ (residues 14–20), Cys(Me)-Arg₆ (residues 21–27), Cys(Me)$_2$-Pro-Arg₅-Ala (residues 28–36) and Cys(Me)$_2$-Pro-Arg₃ (residues 28–35), and other small fragments containing serine, histidine or valine at the N-terminus. These results proved that thermolysin cleaved at the amino-side of methylcysteine residues as well as those of naturally occurring amino acids with hydrophobic side-chains. Chymotrypsin A also hydrolyzed the carbonyl bond at methylcysteine residues more slowly than that of tyrosine residues (results of complete hydrolysis are not shown). The results for S-methyl-protamine are summarized in part C of Fig. 1.*

The complete amino acid sequence of boar protamine was reconstructed by overlap of the sequenced peptide fragments, which is given at the top in Fig. 1.

We reported the minor component of boar protamine to be deficient in the C-terminal half-cystine [21], which was also confirmed in this experiment by identifying the two C-terminal peptides with

* See also supplementary data.
Fig. 2. Comparison of the sequences of mammalian protamines, and the sequences of fish and fowl protamines. The sequences of mammalian protamines are aligned for maximum homology, in which the relevant residues are underlined, and the numbering refers to the sequence position of boar protamine. Partial sequence of human protamine is showed only for one component, because the component II shows no similarity to other mammalian protamines [10]. In fish and fowl protamines, arginine-clusters are represented in stippled boxes, where the numerals following Arg denote the numbers of arginine residues. Fish protamine shows a common sequence among homologous proteins and the residues on the two lines indicate alternative occurrences at the corresponding position.

and without half-cystine at the C-terminus.

The complete amino acid sequence of boar and bull protamines, together with the available partial sequences of other mammalian protamines, are shown in Fig. 2. Comparison of these sequences reveals that the mammalian protamine is composed of three domains: the N-terminal and the C-terminal less-basic regions and the central highly basic region containing three arginine-clusters, each corresponding to one-third of the molecule. The half-cystine residues characteristic of mammalian protamines are interspersed throughout the whole molecule, and are ultimately cross-linked through disulfide bonds in nucleoprotein complexes [18–20].

On closer inspection, mammalian protamines have the common tetrapeptide Ala-Arg-Tyr-Arg at the N-terminus and the sequences of the first 15 residues are well conserved among bull, boar and rat (residue 9: His/Lys) and three substitutions between boar and bull (residues 7, 8 and 11: Arg/Leu, Ser/Thr and Arg/Gly, respectively) were found. The three half-cystine residues (residues 5, 6 and 14) locate at the corresponding positions, and the three potential sites for phosphorylation (residues 8, 10 and 12) are also conserved even in human [19,28].

The central highly basic region contains 60–70% of the total arginines and is composed of the three clusters of 5–7 arginine residues which are separated by half-cystine or non-basic amino acids. Boar protamine comprised two proline residues in this region, thereby the peptide chain bends to allow the arginine stretches to interact with DNA.

The C-terminal sequence of mammalian protamine is thought to be variable between species [8–10]. The sequence from residue 39 to the end of boar protamine differs from that of bull protamine by two substitutions (residues 48 and 50: Arg/Thr
and Cys/Gln, respectively) and the insertion of Cys-Arg₂ (residues 39–41), indicating that, with regard to boar and bull protamines, the C-terminal regions show a considerable similarity. As a whole, boar and bull protamines show about 80% homology.

Fig. 2 also shows the sequences of fish and fowl protamines. Fish protamines consisting of about 30 amino acid residues show rather symmetrical distribution of basic arginine-clusters over the whole molecule: stretches of 4–6 arginines separated by non-basic amino acids with short side-chains. On the other hand, the fowl protamine (galline) [29] consisting of 65 amino acid residues exhibits an N-terminal arginine-dispersed region (residues 1–14) with Ala-Arg-Tyr-Arg at the N-terminus identical to that of mammalian protamine, and the remainder (residues 15–65) contains six arginine-clusters dispersed along the whole region in a similar way to the fish arginine-clusters. Consequently, the fowl protamine seems to be an intermediate between fish and mammalian protamines.

Acknowledgement

This study was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

References