BBA 36278

THE COMPLETE AMINO ACID SEQUENCE OF THE BASIC NUCLEAR PROTEIN OF BULL SPERMATOZOA

JAN P. COELINGH[a], CORNELIS H. MONFOORT[a], THOMAS H. ROZIJN[a], JAN A. GEVERS LEUVEN[a], ROELOF SCHIPHOF[a], ELIZABETH P. STEYN-PARVÉ[a], GERHARD BRAUNITZER[b], BARBARA SCHRANK[b] AND ANNETTE RUHFUS[b]

[a]*Laboratory for Physiological Chemistry, State University of Utrecht (The Netherlands) and* [b]*Max-Planck-Institute for Biochemistry, Munich (Germany)*

(Received August 9th, 1972)

SUMMARY

The complete amino acid sequence of the basic nuclear protein of bull spermatozoa has been established. The sequence was partially deduced by characterization of peptides isolated from thermolysine and chymotryptic digests of the reduced and S-aminoethylated protein. The complete sequence of the first 37 residues was determined by Edman–Begg degradation (Sequenator) of the protein. The amino acid sequence of the last 8 residues was determined by manual Edman degradation of an octapeptide obtained on cleavage of the reduced and S-methylated protein with *N*-bromosuccinimide.

The basic nuclear protein of bull spermatozoa contains 47 amino acid residues with alanine at the amino terminus and glutamine at the carboxyl terminus. The protein contains 24 arginine residues and 6 half-cystine residues. 19 of the arginine residues are in the middle of the molecule, arranged in three clusters of, respectively, 7, 6 and 6 residues. Of the 6 half-cystine residues 2 are present in the amino-terminal region and 2 in the carboxyl-terminal region of the molecule. The half-cystine residues are responsible for the cross-linking of the deoxyribonucleoprotein molecules which gives the mature sperm nucleus an unusual resistance towards chemical and mechanical damage.

The complete amino acid sequence is: H_2N–Ala–Arg–Tyr–Arg–$(Cys)_2$–Leu–Thr–His–Ser–Gly–Ser–Arg–Cys–$(Arg)_7$–Cys–$(Arg)_6$–Phe–Gly–$(Arg)_6$–Val–Cys–Tyr–Thr–Val–Ile–Arg–Cys–Thr–Arg–Gln.

INTRODUCTION

The nuclei of mammalian spermatozoa contain basic proteins that are very rich in arginine. They differ from the protamines found in sperm nuclei of some fishes and birds in the difficulty with which they are extracted. The observation that mammalian

spermatozoa are relatively rich in cystine led some authors[1,2] to the assumption that the sperm heads are surrounded by a keratinous membrane which impedes the extraction of nuclear proteins. Bril-Petersen and Westenbrink[3] demonstrated that the nuclear protein of mammalian spermatozoa is not only rich in arginine but also relatively rich in cystine (see also Henricks and Mayer[4,5]). On the basis of their results they rejected the idea of an impermeable keratinous membrane, and instead proposed that in the nuclei of mammalian spermatozoa the DNA is associated with a sponge-like network of arginine-rich keratinoid protein molecules which are cross-linked by numerous disulfide bonds[3]. Recent ultrastructural studies of sperm nuclei from a variety of eutherian mammals also indicate the presence of a thiol-rich protein whose cross-linking is responsible for the firmness of mature sperm nuclei[6,7]. The disulfide cross-linking appears to increase during the maturation of the spermatozoa[6].

Coelingh *et al.*[8] developed a method for the isolation and purification of the arginine-rich keratinoid protein from mammalian sperm heads, based on reductive cleavage of the disulfide bonds. They found that nuclei of bull spermatozoa contain only a single type of basic protein. This could be characterized as a small protein (47 residues, molecular weight approx. 6200), rich in arginine (24 residues) and half-cystine (6 residues).

Knowledge of the amino acid sequence is required to understand how the molecules are cross-linked to form a coherent network and how they are associated with the DNA. For this reason we have undertaken the analysis of the amino acid sequence of the basic nuclear protein of bull spermatozoa, which is reported in this paper.

METHODS

Spermatozoa

The bull spermatozoa used in this study were supplied by Dr S. W. J. van Dieten from the station for artificial insemination "De Kempen", Oerle, The Netherlands.

Isolation and purification of the basic nuclear protein from bull spermatozoa

The basic nuclear protein was isolated and purified as described previously[8], with some modifications.

S-aminoethylated protein was prepared as follows. Isolated spermatozoan heads (150 mg) were dissolved in 100 ml of a guanidine·HCl solution (5 M guanidine·HCl, 0.5 M Tris–HCl, 5 mM EDTA, pH 8.5) containing 0.1 M 2-mercaptoethanol. The mixture was stirred under a stream of N_2 for 60 min at 37 °C. Ethyleneimine was then added to a concentration of 0.25 M. After an additional incubation at 37 °C for 90 min, the solution was dialyzed against 0.25 M HCl as described previously[8] to precipitate the DNA.

S-methylated protein was prepared using the procedure described by Heinrikson[9]. Isolated spermatozoan heads (150 mg) were dissolved in 100 ml of a denaturing solvent (6 M guanidine·HCl, 0.25 M Tris–HCl, 3.3 mM EDTA, 25% (v/v) of acetonitrile, pH 8.6) containing 0.015 M 2-mercaptoethanol. The mixture was incubated under N_2 in a closed vial at 50 °C for 1 h. A solution of methyl-*p*-nitrobenzene sulfonate in acetonitrile was then added to give a concentration of 0.032 M. After an additional incubation at 37 °C for 90 min the solution was dialyzed against 0.25 M HCl to

precipitate the DNA. The S-methylated basic nuclear protein is completely soluble in 0.25 M HCl and in water.

The S-aminoethylated and the S-methylated basic nuclear proteins were purified by gelfiltration and ion exchange chromatography as described previously[8].

Thermolysine digestion of the basic nuclear protein

The purified basic protein was dissolved to a concentration of 0.4 μmole per ml in 0.2 M ammonium acetate buffer, pH 8.5, containing 3 mM $CaCl_2$. Thermolysine (Biocal, Grade B) was added to give an enzyme: substrate weight ratio of 1:200. Digestion was allowed to proceed for 1 h at 37 °C. The reaction mixture was then acidified to pH 3.5 by addition of glacial acetic acid and dried *in vacuo*.

Fractionation of thermolysine digest by ion-exchange chromatography

The dried thermolysine digest of the basic nuclear protein (approx. 20 mg) was dissolved in 3 ml 0.1% (v/v) formic acid and then applied to a column (0.9 cm $\times$ 30 cm) of Amberlite CG-50 II (H^+) equilibrated with 0.1% formic acid. Elution was accomplished with linear gradients of formic acid. A flow rate of 21 to 22 ml/h was maintained using a Technicon peristaltic pump. Fractions of 1.8 ml were collected. The effluent was monitored by ninhydrin reaction after alkaline hydrolysis[10].

Column fractions containing peptides were analyzed by descending paper chromatography in *n*-butanol–pyridine–acetic acid–water (15:12:3:10 by vol.) on Whatman 3 MM paper. The peptides were located by specific staining for arginine[11] and by ninhydrin spraying. On the basis of the position and the colour intensity of the spots column fractions were pooled as indicated in Fig. 1.

Pooled fractions containing two or more peptides were fractionated by preparative chromatography on Whatman 3 MM paper with *n*-butanol–pyridine–acetic acid–water (15:12:3:10, by vol.) as solvent. The peptides were eluted from the paper with 1 M acetic acid at 30 to 40 °C and subsequently dried *in vacuo*.

Chymotrypsin digestion of the basic nuclear protein

The purified basic protein was dissolved in water to a concentration of 0.4 μmole/ml and the solution was adjusted to pH 8 with 0.01 M NaOH. α-Chymotrypsin (Merck, tosyl-L-leucine chloromethyl ketone treated) was added to give an enzyme: substrate weight ratio of 1:50. Throughout digestion the mixture was maintained at pH 8 with 0.01 M NaOH using an autotitrator. Digestion was carried out for 20 min at 37 °C. An additional amount of chymotrypsin, equivalent to 1/100 (w/w) of the protein, was then added and digestion was allowed to proceed for another 20 min. The reaction mixture was then acidified to pH 3.5 by addition of glacial acetic acid and dried *in vacuo*.

Fractionation of chymotryptic digest by gelfiltration

The dried chymotryptic digest of the basic nuclear protein (approximately 6 mg) was dissolved in 1 ml of 0.01 M HCl and then applied to a column (1.1 cm $\times$ 120 cm) of Sephadex G-50 equilibrated with 0.01 M HCl. The column was developed with the same solvent at a flow rate of 6 ml per hour. Fractions (1 ml) were monitored by ninhydrin reaction after alkaline hydrolysis[10]. Fractions containing peptides were analyzed by descending paper chromatography on Whatman 3 MM paper with *n*-

butanol–pyridine–acetic acid–water (15:12:3:10 by vol.) as solvent. Appropriate fractions were pooled and evaporated to dryness *in vacuo*. The pooled fraction (fraction II, Fig. 2) that contained a mixture of small peptides was fractionated by gelfiltration on Sephadex G-25. The dried peptide material was dissolved in 1 ml of 0.01 M HCl and then applied to a column (1.1 cm × 120 cm) of Sephadex G-25 equilibrated with 0.01 M HCl. The column was developed with the same solvent at a flow rate of 6 ml/h. Fractions of 1.5 ml were collected and monitored by ninhydrin reaction after alkaline hydrolysis[10]. Fractions were pooled as indicated in Fig. 3. The pooled fractions were purified further by preparative paper chromatography on Whatman 3 MM paper with *n*-butanol–pyridine–acetic acid–water (15:12:3:10, by vol.) as solvent.

N-Bromosuccinimide cleavage of the basic nuclear protein

N-Bromosuccinimide cleavage was performed in 50% (v/v) acetic acid[12] at a protein concentration of 0.4 μmole/ml and a *N*-bromosuccinimide concentration of 8 μmoles/ml. After 4 h at 25 °C the mixture was dried *in vacuo*.

Separation of N-bromosuccinimide cleavage fragments

The cleavage products from 3.3 μmoles of protein were dissolved in 1 ml of 0.02 M HCl and applied to a 1.1 cm × 100 cm column of Sephadex G-50 (superfine) equilibrated with 0.02 M HCl. The column was developed with the same solvent at 4.2 ml/h and 1.5-ml fractions were collected. The effluent was monitored at 230 nm to detect peptide-containing fractions, and at 260 nm to detect fractions which contained peptides with a carboxyl-terminal dienone spirolactone. Three major peaks were eluted. Only the second peak, which contained an octapeptide derived from the carboxyl-terminal part of the protein molecule, was used for further studies. This octapeptide was submitted to a final purification by ion-exchange chromatography. After evaporation to dryness *in vacuo* the peptide material was dissolved in 1 ml 0.001 M HCl and applied to a column (0.7 cm × 10 cm) of Amberlite CG-50 II (H^+). A small amount of contaminating material was washed from the column with 15 ml 0.001 M HCl. The octapeptide was then eluted with approx. 20 ml of 0.01 M HCl and subsequently dried *in vacuo*.

Amino-terminal analysis

Quantitative amino-terminal amino acid determinations were performed according to the method of Stark and Smyth[13,14]. The basic nuclear protein (0.8 μmole) was carbamylated for 16 h at 50 °C in 1 ml of 4 M guanidine·HCl containing 0.6 M potassium cyanate and 1 M *N*-ethylmorpholine acetate, pH 8.0. Excess cyanate was destroyed by dialysis against 50% (v/v) acetic acid. The remaining procedures were as described by Stark and Smyth[13,14].

Carboxyl-terminal analysis

1 mg of carboxypeptidase A (Sigma, DFP treated) was dissolved in 0.2 ml of 10% LiCl and then diluted with water to a concentration of 0.6 mg/ml. Digestion was carried out in reaction mixtures (0.2 ml) containing 86 nmoles of basic protein and 0.7 nmole of carboxypeptidase A in 0.2 M *N*-ethylmorpholine acetate, pH 7.6. The mixtures were incubated at 37 °C for 2, 4 and 6 h. After incubation 0.1 ml of 3 M acetic acid was added and the reaction mixture was then placed for 1 min in a boiling water

bath. To separate the released amino acids from the remaining protein, the digests were chromatographed on a column (1 cm × 4 cm) of Amberlite CG-50 II (H^+). The released amino acids were eluted with 35 ml of 0.2 M acetic acid, the protein with 15 ml of 0.05 M HCl. The amino acid fraction was analyzed on a Technicon amino acid analyzer. To discriminate between glutamine and threonine, which run together in the Technicon analytical system, a portion of the amino acid fraction was analyzed after hydrolysis for 1 h in 1 M HCl at 100 °C.

Degradation according to Edman–Begg[15]

1 μmole of the S-aminoethylated protein was subjected to the automatic degradation in the sequenator according to Edman–Begg[15] (sequencer: Beckman Instruments, Palo Alto, Calif., U.S.A.) and using the Quadrol double cleavage protein program. The phenylthiohydantoin derivatives were identified by thin-layer chromatography on DC-Fertigplatten Kieselgel F 254 (Merck) as described by Braunitzer *et al.*[16]. The phenylthiohydantoin derivatives of arginine and histidine were identified by the phenantroquinone method of Yamada and Itano[11] and by the Pauly reagent[17], respectively. After step 37 it was no longer possible to identify unambiguously the phenylthiohydantoin amino acid derivatives that were removed.

For the sequence determination of the carboxyl-terminal octapeptide (residue 40 through 47) obtained by means of *N*-bromosuccinimide cleavage, the subtractive method was used as described by Elzinga[18]. Aliquots of the octapeptide containing 0.1 μmole were used for each step. As the ε-phenylthiocarbamyl group of *S*-aminoethylcysteine is only partially removed on hydrolysis with 6 M HCl at 110 °C for 24 h (see also Konigsberg[19]), an octapeptide was used that was prepared from S-methylated protein instead of S-aminoethylated protein.

Amino acid analysis

Samples of 0.5 mg of protein or peptide were hydrolyzed with 2 ml of 5.7 M HCl containing 0.05% thioglycollic acid in evacuated and sealed tubes at 110 °C for 24 h (and for 48 and 72 h if indicated). After hydrolysis the acid was removed by rotary evaporation. The samples were then dissolved in 1 ml of 0.01 M HCl. The amino acid compositions were determined on an automatic amino acid analyzer (Technicon Auto Analyzer).

Nomenclature

Thermolysine peptides are indicated by the letter Th and chymotrypsin peptides by the letter C. The peptides have been numbered arbitrarily according to their order of elution from a chromatographic column.

RESULTS

Amino acid composition and terminal residues of the basic nuclear protein

The results of the amino acid analysis and of the end group determinations are summarized in Table I. From the amino acid composition a minimum molecular weight of 6190 can be deduced. Using this molecular weight, quantitative amino-terminal determination indicated 0.89 residue of alanine at the amino terminus. Similarly, the digestion with carboxypeptidase A indicated 0.88 residue of glutamine

TABLE I

AMINO ACID COMPOSITION OF THE S-AMINOETHYLATED BASIC NUCLEAR PROTEIN

The *S*-aminoethylprotein was hydrolyzed for 24, 48 and 72 h. The values for threonine and serine were extrapolated to zero-time hydrolysis. For valine and isoleucine the 72 h values were taken. The value for *S*-aminoethylcysteine has not been corrected for the hydrolytic loss that can occur in some instances (up to 10%, see also Bartelt and Greene[31]). Tryptophan was determined according to Goodwin and Morton[32].

Amino acid	*Molar ratios*	*Residues/6190 mol. wt*
Arginine	24.3	24
S-Aminoethylcysteine	5.7	6
Threonine	3.0	3
Serine	2.0	2
Glycine	2.1	2
Valine	2.0	2
Tyrosine	1.9	2
Glutamic acid	1.0	1
Leucine	1.0	1
Alanine	1.0	1
Isoleucine	1.0	1
Phenylalanine	1.0	1
Histidine	1.0	1
Aspartic acid	0	
Proline	0	
Lysine	0	
Methionine	0	
Tryptophan	0	
Total number of residues		47
NH_2-terminal residue	Ala (0.89)	1
COOH-terminal residue	Gln (0.88)	1

at the carboxyl terminus. Thus it can be concluded that all of the alanine in the protein is at the amino terminus and all of the glutamic acid is present as carboxyl-terminal glutamine. The amino acid composition together with the data of the end group analyses establish the following composition of the protein: $Ala(Arg_{24},Cys_6, Thr_3,Tyr_2,Ser_2,Gly_2,Val_2,Leu_1,Ile_1,Phe_1,His_1)Gln$. The digestion with carboxypeptidase A released only glutamine and no other amino acids, even after prolonged digestion. This suggests the presence of either arginine, *S*-aminoethylcysteine or proline as the penultimate amino acid residue. The protein does not contain proline. The penultimate residue will therefore probably be arginine or *S*-aminoethylcysteine.

Isolation and characterization of thermolysine peptides

A thermolysine digest of the basic nuclear protein was fractionated on a column of Amberlite CG-50 II (H^+) as shown in Fig. 1. Of the seven fractions obtained, only Fraction I (Th-1) and Fraction VII (Th-7) contained a single pure peptide. The Fractions II through VI were impure. The peptides in these fractions were purified by paper chromatography. By this means five additional pure peptides were isolated: Th-2 from Fraction II, Th-3 and Th-4 from Fraction III, Th-5 from Fraction IV and Th-6 from Fraction V. Fraction VI contained the same peptide (Th-7) as Fraction VII.

The composition of the thermolysine peptides is given in Table II. It can be seen that the sum of the amino acid compositions of the peptides is in good agreement with the composition of the *S*-aminoethylprotein, except for a shortage of one valine

TABLE II

AMINO ACID COMPOSITION OF THERMOLYSINE PEPTIDES

Values are given in molar ratios. The numbers in parentheses are the assumed numbers of residues.

Amino acid	*Th-1*	*Th-2*	*Th-3*	*Th-4*	*Th-5*	*Th-6*	*Th-7*	*Sum of peptides** Th-1, Th-4, Th-5, Th-6 and Th-7*	*Composition of nuclear protein*
Arg		1.0 (1)	1.0 (1)	2.6 (2–3)	2.1 (2)	6.0 (6)	14.0 (14)	24–25	24
Cys (aminoethyl)			1.0 (1)	2.3 (2)	1.1 (1)		2.0 (2)	5	6
Thr	1.0 (1)	1.0 (1)	0.1	0.2	1.0 (1)		1.0 (1)	3	3
Ser							1.5 (2)	2	2
Gly						1.1 (1)	1.2 (1)	2	2
Val*			0.6 (1)	0.2	0.6 (1)		0.2	1	2
Tyr	1.0 (1)			1.0 (1)	0.1		0.2	2	2
Glu		1.0 (1)		0.1	1.0 (1)			1	1
Leu							0.9 (1)	1	1
Ala				0.9 (1)	0.1			1	1
Ile*			0.6 (1)		0.6 (1)			1	1
Phe				0.1		0.9 (1)	0.2	1	1
His							0.9 (1)	1	1
Total number of residues	2	3	4	6–7	7	8	22	45–46	47

* 24 h value, not corrected for imcomplete hydrolysis.
** Peptides Th-2 + Th-3 = Th-5.

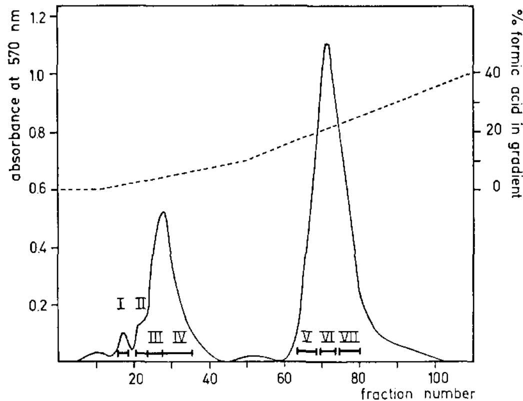
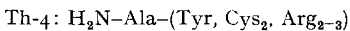


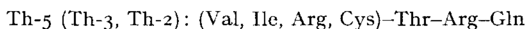
Fig. 1. Chromatographic separation of the thermolysine peptides of the basic nuclear protein. A thermolysine digest of approx. 20 mg of protein was applied to a 0.9 cm × 30 cm column of Amberlite CG-50 II (H^+). The column was developed by linear gradient elution with formic acid. The effluent was monitored as described in the text. Fractions were pooled into I, II, III, IV, V, VI and VII. ———, absorbance at 570 nm; — — —, % formic acid in gradient.

and one *S*-aminoethylcysteine residue. It could be that we have overlooked a dipeptide Val–Cys (aminoethyl) which can be formed by the thermolysine digestion. Dipeptides with an amino-terminal valine are known to give only a very faint colour with ninhydrin[20].

As all alanine residues in the protein are amino-terminal, it is clear that Th-4 must contain the first 6 or 7 residues from the amino terminus.


Th-4: H_2N–Ala–(Tyr, Cys_2, Arg_{2-3})

Similarly, as all glutamic acid residues in the protein occur as carboxyl-terminal glutamine, Th-5 must contain the last seven residues at the carboxyl terminus. The peptides Th-2 and Th-3 are apparently derived from Th-5. The results of the carboxypeptidase A digestion indicated arginine or *S*-aminoethylcysteine as the penultimate residue. As the carboxyl-terminal peptide Th-2 does not contain *S*-aminoethylcysteine, it follows that the carboxyl-terminal sequence (residue 41 through 47) is probably:


Th-5 (Th-3, Th-2): (Val, Ile, Arg, Cys)–Thr–Arg–Gln

Isolation and characterization of chymotryptic peptides

A chymotryptic digest of the basic nuclear protein was fractionated by means of gel filtration on Sephadex G-50 to give the pattern shown in Fig. 2. Two pooled fractions were collected. Fraction I contained a fairly pure peptide, C-1. Fraction II which contained a mixture of small peptides was separated by gelfiltration on Sephadex G-25 into four fractions: C-2, C-3, C-4 and C-5,6 (Fig. 3). Each of these peptide fractions was purified further by preparative paper chromatography.

The composition of the chymotryptic peptides is given in Table III. The sum of the amino acid compositions of the chymotryptic peptides agrees well with the composition of the *S*-aminoethylprotein.

The alanine-containing peptide C-5 must represent the first 3 residues from the

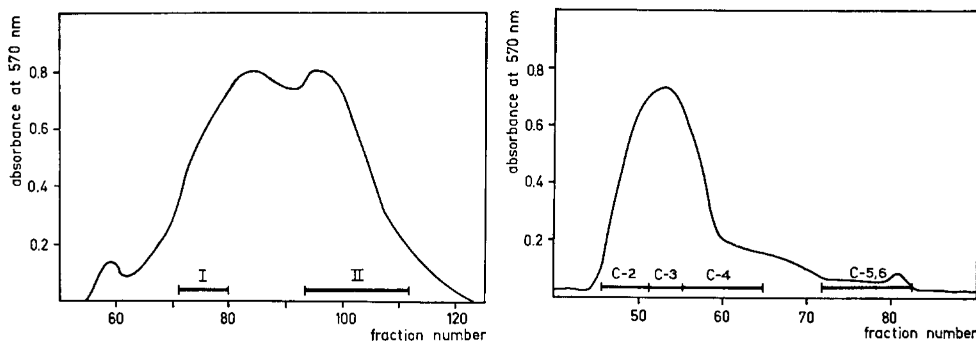
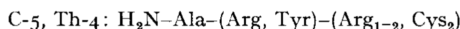


Fig. 2. Separation of the chymotryptic peptides of the basic nuclear protein by gel filtration. A chymotryptic digest of approx. 6 mg of protein was applied to a Sephadex G-50 column (1.1 cm × 120 cm) equilibrated with 0.01 M HCl, and eluted with 0.01 M HCl. The effluent was monitored as described in the text. The Roman numerals indicate pooled fractions.

Fig. 3. Gel filtration on Sephadex G-25 of chymotryptic fraction II. Fraction II from Fig. 2 was rechromatographed on a 1.1 cm × 120 cm column of Sephadex G-25 equilibrated with 0.01 M HCl. The column was developed with the same solvent. The effluent was monitored as described in the text. The bars indicate the pooled fractions.

amino terminus, because all alanine residues in the protein are amino-terminal. From the compositions of peptides C-5 and Th-4 the following partial sequence of the first 6 or 7 residues can be deduced.


C-5, Th-4: H_2N-Ala-(Arg, Tyr)-(Arg_{1-2}, Cys_2)

The glutamic acid-containing peptide C-3 must represent the last 8 residues at

TABLE III

AMINO ACID COMPOSITION OF CHYMOTRYPTIC PEPTIDES

Values are given in molar ratios. The numbers in parentheses are the assumed numbers of residues.

Amino acid	*C-1*	*C-2*	*C-3*	*C-4*	*C-5*	*C-6*	*Sum of peptides*	*Composition of nuclear protein*
Arg	14.0 (14)	5.9 (6)	2.1 (2)	1.4 (1)	1.0 (1)		24	24
Cys (aminoethyl)*	2.0 (2)	0.9 (1)	0.9 (1)	1.5 (2)			6	6
Thr	0.9 (1)	0.2	2.0 (2)	0.3			3	3
Ser	2.2 (2)	0.3	0.2	0.3			2	2
Gly	1.0 (1)	1.0 (1)	0.2	0.3			2	2
Val**		0.9 (1)	0.9 (1)				2	2
Tyr	0.2			0.2	1.0 (1)	1.0 (1)	2	2
Glu		0.2	1.1 (1)	0.3			1	1
Leu			0.2	1.0 (1)			1	1
Ala				0.3	1.0 (1)		1	1
Ile**			0.8 (1)				1	1
Phe	1.0 (1)						1	1
His	0.7 (1)	0.2					1	1
Total number of residues	22	9	8	4	3	1	47	47

* Not corrected for hydrolytic loss (see Table I).
** 24 h value, not corrected for incomplete hydrolysis.

the carboxyl terminus, because all glutamic acid residues in the protein occur as carboxyl-terminal glutamine. From the compositions of the peptides C-3, Th-2, Th-3 and Th-5 the following partial sequence of the last 8 residues at the carboxyl terminus can be deduced.

C-3, Th-5 (Th-3, Th-2): Thr–(Val, Ile, Arg ,Cys)–Thr–Arg–Gln

Isolation and composition of a N-bromosuccinimide fragment

The nuclear basic protein contains two tyrosine residues, so that *N*-bromosuccinimide might be expected to cleave the protein into three pieces. Cleavage of the peptide bond next to histidine is not likely to occur under the conditions used[21]. The cleavage fragments to be expected are a very small amino-terminal fragment (compare composition of peptide C-5) and a fragment derived from the middle region of the molecule, both containing a carboxyl-terminal dienone spirolactone[21], and a third fragment which is derived from the carboxyl-terminal part of the molecule. To separate the fragments from each other and from the reagents, the reaction mixture was subjected to gelfiltration on Sephadex G-50. The first peak that was eluted had a high absorbance at 260 nm. Apparently it contained a large fragment, probably the middle region of the molecule, with a carboxyl-terminal dienone spirolactone. The absorbance at 260 nm of the second peak was negligible. This suggested that the material in this peak was derived from the carboxyl-terminal part of the molecule. Finally some small peptide material and salts both with a high absorbance at 260 nm were eluted. This peptide material was probably derived from the amino-terminal part of the molecule. Only the second peak was investigated further. Amino acid analysis indicated that this peak contained an octapeptide that was identical to the carboxyl-terminal peptide C-3. This octapeptide was purified by chromatography on Amberlite CG-50 II (H^+), and was obtained in a yield of 34%.

As *N*-bromosuccinimide cleaves the peptide bond next to the tyrosine carboxyl group, the partial sequence of residues 39 through 47 must be (compare compositions of C-3, Th-2, Th-3 and Th-5):

Tyr–Thr–(Val, Ile, Arg, Cys)–Thr–Arg–Gln

To determine the complete sequence of the last eight residues at the carboxyl terminus, the *N*-bromosuccinimide octapeptide was subjected to the subtractive Edman degradation.

Sequence analysis

Residues 1 to 38. Automatic sequential Edman degradation of the basic nuclear protein provided the sequence of the first 37 residues and permitted the ordering of the peptides Th-4, Th-6, Th-7, C-1, C-2, C-4 and C-5. The series of sequential degradations was performed twice and the same result was obtained for each residue on each occasion. The amino acid residue (cysteine) at position 38 could be inferred from the composition of peptide C-2*. The results are summarized in Table IV.

* Peptide C-2 is apparently formed by an uncommon, tryptic-like, cleavage between *S*-aminoethylcysteine and tyrosine (residue 39). This also explains the finding of free tyrosine (C-6) in the chymotryptic digest. Similar unusual cleavages have been repeatedly observed by others (see, *e.g.* ref. 22).

TABLE IV

SEQUENCE OF AMINO ACID RESIDUES 1–38

	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
Sequence	*H_2N–Ala–Arg–Tyr–Arg–Cys–Cys–Leu–Thr–His–Ser–Gly–Ser–Arg–Cys–Arg–Arg–Arg–Arg–Arg–Arg–Arg–Cys–Arg–Arg–Arg–Arg–Arg–Arg–Phe–Gly–Arg–Arg–Arg–Arg–Arg–Arg–Val–Cys–*
Se-quenator	H_2N–Ala–Arg–Tyr–Arg–Cys–Cys–Leu–Thr–His–Ser–Gly–Ser–Arg–Cys–Arg–Arg–Arg–Arg–Arg–Arg–Arg–Cys–Arg–Arg–Arg–Arg–Arg–Arg–Phe–Gly–Arg–Arg–Arg–Arg–Arg–Arg–Val–
C-5	H_2N–Ala (Arg, Tyr)
Th-4	H_2N–Ala (Arg, Tyr, Arg, Cys, Cys)
C-4	(Arg, Cys, Cys, Leu)
Th-7	(Leu, Thr, His, Ser, Gly, Ser, Arg, Cys, Arg, Arg, Arg, Arg, Arg, Arg, Arg, Cys, Arg, Arg, Arg, Arg, Arg, Arg)
C-1	(Thr, His, Ser, Gly, Ser, Arg, Cys, Arg, Arg, Arg, Arg, Arg, Arg, Arg, Cys, Arg, Arg, Arg, Arg, Arg, Arg, Phe)
Th-6	(Phe, Gly, Arg, Arg, Arg, Arg, Arg, Arg)
C-2	(Gly, Arg, Arg, Arg, Arg, Arg, Arg, Val, Cys)

TABLE V

SUBTRACTIVE SEQUENCE ANALYSIS OF THE *N*-BROMOSUCCINIMIDE OCTAPEPTIDE (RESIDUE 40–47)

	*Composition**					
	Thr	*Val*	*Ile*	*Cys (Met)***	*Arg*	*Glu*
Peptide	2.0	0.9	0.9	0.6	2.2	1.0
Edman degradation						
Step 1	*0.9*	0.9	0.8	0.6	2.3	1.0
Step 2	0.9	*0.1*	0.9	0.6	2.1	1.0
Step 3	0.9	0.1	*0.2*	0.6	2.1	1.0
Step 4	0.9	0.1	0.1	0.6	*1.4*	1.0
Step 5	0.9	0.1	0.1	*0.3*	1.4	1.0
Step 6	*0.4*	0.1	0.1	0.0	1.2	1.0

* Values based on 48 h hydrolysis, expressed as molar ratios.
** *S*-Methylcysteine is probably partially oxidized by *N*-bromosuccinimide[33,34] (compare composition of peptide C-3).

Residues 39 to 47. The amino acid sequence of the *N*-bromosuccinimide octapeptide (residue 40–47) was determined by manual Edman degradation up to and including residue 45. The data presented in Table V refer to the amino acid compositions (expressed as molar ratios) of the intact peptide and of the residue remaining after each step of Edman degradation. The amino acid that was apparently removed at each step is indicated in italics. The octapeptide, which is identical to peptide C-3, represents the last 8 residues in the protein because it contains the carboxyl-terminal glutamine (position 47). By exclusion, residue 46 can only be arginine. This is in accordance with the composition of peptide Th-2 (see p. 8) and the results of carboxypeptidase A digestion.

The sequence data are summarized in Table VI together with the amino acid compositions of the peptides Th-1, Th-2, Th-3 and Th-5. The assignment of tyrosine to position 39 is based on the cleavage specificity of *N*-bromosuccinimide and on the composition of peptide Th-1.

TABLE VI

SEQUENCE OF AMINO ACID RESIDUES 39–47

Sequence	39 –*Tyr* –	40 *Thr* –	41 *Val* –	42 *Ile* –	43 *Arg* –	44 *Cys* –	45 *Thr* –	46 *Arg* –	47 *Gln*
N-Bromosuccinimide peptide, C-3		Thr –	Val –	Ile –	Arg –	Cys –	Thr –	Arg –	Gln
Th-1	(Tyr,	Thr)							
Th-5			(Val,	Ile,	Arg,	Cys,	Thr,	Arg)	Gln
Th-3			(Val,	Ile,	Arg,	Cys)			
Th-2							(Thr,	Arg)	Gln

DISCUSSION

The basic nuclear proteins (histones) of spermatozoa have been studied mainly in some fishes and birds, where very basic proteins called protamines are found. Protamines are small proteins (30–35 amino acid residues) which are very rich in arginine (20–22 residues). However, protamines are not ubiquitous sperm histones. On the

contrary, the basic proteins of sperm nuclei vary widely among different organisms. A review on the different types of sperm histones was recently published by Bloch[23].

The histone of bull sperm is very rich in arginine and, unlike the protamines, also relatively rich in half-cysteine residues. In addition to the histone of bull spermatozoa we have isolated and partially characterized the sperm histones from four other mammals, namely boar, ram, stallion and man. They vary in amino acid composition and carboxyl-terminal residue, but all are rich in arginine and half-cystine. A report on this work will be published later. Calvin and Bedford[6] concluded from their electron microscopic observations that cross-linked, thiol-rich proteins are present in the sperm nuclei of a variety of eutherian mammals, such as bull, boar, rabbit, rat, mouse, guinea pig, cat, ferret, shrew and man. Sperm nuclei of metatherian mammals (*e.g.* marsupials) on the other hand seem to be devoid of such thiol-rich proteins (see also Bloch[23]). It thus seems unquestionable that the sperm histones of eutherian mammals belong to a separate class distinct from protamines such as the clupeines[24–26], salmines[27], iridines[27], thynnines[28,29] and gallines[30]. The most striking difference is the presence of a relatively large number of half-cystine residues, which apparently participate in the cross-linking of the deoxyribonucleoprotein molecules in the sperm nuclei of eutherian mammals.

Until recently[8] no reliable data on the composition of mammalian sperm histones were available. We now present the first example of a complete amino acid sequence of a mammalian sperm histone. The amino acid sequence of the bull sperm histone shows a remarkable irregular distribution of basic and other residues. 20 out of 24 arginine residues are located in a 24 residues long middle-region (residues 13–36) of the molecule. 19 of these arginine residues are arranged in three clusters of 7, 6 and 6 residues respectively. These clusters are separated by only one or two monoamino acid residues.

13 … 31–36
–Arg–Cys–Arg$_7$–Cys–Arg$_6$–Phe–Gly–Arg$_6$–

A clustering of arginine residues is a common feature of protamines. However, the arginine clusters in the bull sperm histone are on the average larger and also less varied in length than the arginine clusters in the protamines of which the sequence is known[24–27,29]. On the other hand, the number of arginine clusters in the bull sperm histone (3) is smaller than the number usually encountered in the protamines (4–6). In addition, the confinement of the arginine clusters to the middle region of the molecule is not found in protamines.

Of the six cysteine residues two are present in the amino-terminal region (residue 1–12) and two in the carboxyl-terminal region (residue 37–47). Two are present in the middle region, each inserted between two arginine residues.

1 … 12
H_2N–Ala–Arg–Tyr–Arg–Cys–Cys–Leu–Thr–His–Ser–Gly–Ser–

37 … 47
–Val–Cys–Tyr–Thr–Val–Ile–Cys–Arg–Thr–Arg–Gln

Of the seven hydrophobic amino acids (tyrosine, phenylalanine, leucine, isoleucine and valine) two are present in the amino-terminal region and four in the carboxyl-terminal region.

The middle region is very positively charged as compared with the amino-

terminal and carboxyl-terminal regions. It is therefore likely that the middle region is the primary site for binding to DNA, whereas the amino-terminal and carboxyl-terminal regions might be responsible for interaction with neighbouring protein molecules to form a keratinoid protein network.

ACKNOWLEDGEMENTS

The present investigations were supported in part by the Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.), and by a short term fellowship from the European Molecular Biology Organization (E.M.B.O.) to Dr C. H. Monfoort.

REFERENCES

1 W. W. Green, *Anat. Rec.*, 76 (1940) 455.
2 C. A. Zittle and R. A. O'Dell, *J. Biol. Chem.*, 140 (1941) 899.
3 E. Bril-Petersen and H. G. K. Westenbrink, *Biochim. Biophys. Acta*, 76 (1963) 152.
4 D. M. Henricks and D. T. Mayer, *Exp. Cell Res.*, 40 (1965) 402.
5 D. M. Henricks and D. T. Mayer, *Proc. Soc. Exp. Biol. Med.*, 119 (1965) 769.
6 H. I. Calvin and J. M. Bedford, *J. Reprod. Fertilit.*, Suppl. 13 (1971) 65.
7 B. Lung, *J. Cell Biol.*, 52 (1972) 179.
8 J. P. Coelingh, T. H. Rozijn and C. H. Monfoort, *Biochim. Biophys. Acta*, 188 (1969) 353.
9 R. L. Heinrikson, *J. Biol. Chem.*, 246 (1971) 4090.
10 C. H. W. Hirs, *Methods Enzymol.*, 11 (1967) 325.
11 S. Yamada and H. A. Itano, *Biochim. Biophys. Acta*, 130 (1966) 538.
12 M. Bustin and R. D. Cole, *J. Biol. Chem.*, 244 (1969) 5291.
13 G. R. Stark and D. G. Smyth, *J. Biol. Chem.*, 238 (1963) 214.
14 G. R. Stark, *Methods Enzymol.*, 11 (1967) 125.
15 P. Edman and G. Begg, *Eur. J. Biochem.*, 1 (1967) 80.
16 G. Braunitzer, R. Chen, B. Schrank and A. Stangl, *Hoppe-Seyler's Z. Physiol. Chem.*, 353 (1972) 832.
17 C. N. Easly, *Biochim. Biophys. Acta*, 107 (1965) 386.
18 M. Elzinga, *Biochemistry*, 9 (1970) 1365.
19 W. Konigsberg, *Methods Enzymol.*, 11 (1967) 461.
20 P. X. Callahan, J. A. Shepard, T. J. Reilly, J. K. McDonald and S. Ellis, *Anal. Biochem.*, 38 (1970) 330.
21 L. K. Ramachandran and B. Witkop, *Methods Enzymol.*, 11 (1967) 283.
22 M. Sokolovsky and M. Moldovan, *Biochemistry*, 11 (1972) 145.
23 D. P. Bloch, *Genetics*, 61 (1969), suppl. 1, 93.
24 T. Ando, K. Iwai, S. Ishii, M. Azegami and C. Nakahara, *Biochim. Biophys. Acta*, 56 (1962) 628.
25 T. Ando and K. Suzuki, *Biochim. Biophys. Acta*, 121 (1966) 427.
26 T. Ando and K. Suzuki, *Biochim. Biophys. Acta*, 140 (1967) 375.
27 T. Ando and S. Watanabe, *Intern. J. Protein Res.*, 1 (1969) 221.
28 G. Bretzel, *Hoppe-Seyler's Z. Physiol. Chem.*, 352 (1971) 1025.
29 G. Bretzel, *Hoppe-Zeyler's Z. Physiol. Chem.*, 353 (1972) 933.
30 M. Nakano, T. Tobita and T. Ando, *Biochim. Biophys. Acta*, 207 (1970) 553.
31 D. C. Bartelt and L. J. Greene, *J. Biol. Chem.*, 246 (1971) 2218.
32 T. W. Goodwin and R. A. Morton, *Biochem. J.*, 40 (1946) 628.
33 S. Sasakawa, *J. Biochem. Tokyo*, 53 (1963) 188.
34 T. F. Spandle and B. Witkop, *Methods Enzymol.*, 11 (1967) 498.