Squid spermiogenesis: molecular characterization of testis-specific pro-protamines

Danièle WOUTERS-TYR,*, Annie MARTIN-PONTHIEU,*, Nathalie LEDOUX-ANDULA,* Mostafa KOUACH, Michel JAQUINOD,† Juan A. SUBIRANA§ and Pierre SAUTIÈRE*†

*URA 1309 CNRS, Institut Pasteur de Lille, Chimie des Biomolécules, 1 rue du Professeur Calmette, 59019 Lille, France, †Service Commun de Spectrométrie de Masse, Faculté de Médecine, Place de Verdun, 59045 Lille Cedex, France, §Laboratoire de Spectrométrie de Masse Bio-organique, Institut de Chimie, Université Louis Pasteur, 67008 Strasbourg, France, and ‡Département d’Ingénierie Chimique, ETSEIB Universitat Politècnica de Catalunya, Diagonal 647, Barcelona E-08028, Spain

Cuttlefish spermiogenesis is characterized by a two-step nuclear protein transition: histones → spermatid-specific protein (protein T) → sperm protamine (protein Sp). A similar situation can be observed in another Cephalopod species, the squid Loligo pealeii. The protein T from Loligo consists of two structural variants, T1 and T2 (molecular masses: 10788 and 10791 Da respectively), phosphorylated to different degrees (2–6 phosphate groups). The primary structures of these two variants and of the protamine variant Sp2 were established from sequence analysis and mass spectrometric data of the proteins and their fragments. T1 and T2 are closely related proteins of 79 residues. The complete structural identity of the C-terminal domain (residues 22–79) of protein T2 with the sperm protamine Sp2 (molecular mass 8562 Da, 58 residues) strongly suggests that the testis-specific protein T2 is indeed the precursor of the protamine. The transition between the precursor protein T and protein Sp results from a hydrolytic cleavage similar to that found in many proteins that are synthesized as precursors. The processing mechanism involves the specific cleavage of a Gly–Arg bond in the sequence Met/Leu–Lys–Gly–Arg–Arg. Furthermore, the study provides molecular evidence on the taxonomic relationship between Loligo and Sepia.

INTRODUCTION

The transformation of precursor proteins into active proteins by specific hydrolysis is known to occur in many cases, such as proenzymes, prohormones, viral proteins, etc. In sperm proteins, a similar transformation has been found to take place in one of the mammalian protamine components (P2). This mammalian P2 protein is usually a minor component of the protamine complement and in many cases it is even absent from the final protamine complement [1].

In the cuttlefish Sepia officinalis, it has been previously shown that the protamine components are synthesized as precursors [2]. Indeed, spermiogenesis of the cuttlefish is characterized by a two-step nuclear basic protein transition [3]. The somatic-type and/or testis-specific histones are first replaced in round and elongating spermatids by a more basic and highly phosphorylated protein called protein T. In elongated spermatids, protein T is replaced by a true protamine (protein Sp) which contains 77% arginine and up to six phosphate groups. During the epididymal transit, the phosphorylated protamine is dephosphorylated [4]. In the mature spermatozoa of cuttlefish, this protamine, which is the most basic protein ever described, is practically the only basic component of chromatin [3].

The total identity of the C-terminal sequence of the spermatidal protein T with protein Sp has led us to postulate that protein T must be considered as a pro-protamine [2].

A two-step transition has also been observed in another Cephalopod decapod species, the squid Illex argentinus [5]. Two proteins, Illexine I1 and Illexine I2, were partially characterized: Illexine I1 replaces histones at the early stages of spermiogenesis, whereas Illexine I2 was found in mature spermatozoa. These proteins were shown to be involved in rearrangements of chromatin structure which appears highly packed in mature spermatozoa [6]. This paper deals with the isolation and molecular characterization of the testis-specific protein variants T1 and T2 and of the protamine variant Sp2 from ripe spermatozoa of the squid Loligo pealeii. As in cuttlefish, the testis-specific protein T from Loligo has to be considered as the precursor of the protamine Sp.

MATERIALS AND METHODS

Materials

Squids, L. pealeii, were obtained from the Marine Biological Laboratory, Woods Hole, MA, U.S.A.

Isolation of protein T

Protein T was extracted from squid testis chromatin. The gonad was homogenized in a Waring Blender with 0.15 M NaCl and the sediment extracted overnight with 5% acetic acid. The extract was discarded and the sediment was extracted again with 0.25 M HCl for 14 h. The extract, which contains the T proteins, was precipitated with six volumes of acetone and the precipitate dried under vacuum.

Isolation of protein Sp

Spermatozoa were suspended in sea water and were broken with the help of tweezers in order to liberate the spermatozoa.

Abbreviations used: ES-MS, electrospray MS; FAB-MS, fast atom bombardment MS; TFA, trifluoroacetic acid.
†To whom correspondence should be addressed.
The spermatophore sheaths were removed by filtration with cheese-cloth. Sperm were recovered by centrifugation and extracted overnight with 5% acetic acid. The extract was discarded and the sediment was extracted again with 0.25 M HCl for 14 h. The Sp protein was recovered from the extract by precipitation with six volumes of acetone and dried under vacuum.

**Fractionation of protein T variants**

Squid testis protein variants T1 and T2 were purified by reverse-phase HPLC on a Superspher 100 RP-18 endcapped column (250 × 4 mm) from Merck, equilibrated in 0.1% trifluoroacetic acid (TFA) and eluted for 40 min with a linear gradient of acetonitrile (0–50%) in 0.1% TFA.

**Analytical electrophoresis**

Protein samples were analysed by polyacrylamide slab gel electrophoresis at pH 3.2 in the presence of 6.25 M urea using a 17% acrylamide concentration [3]. The purity of the peptides was assessed as described in [7].

**Amino acid analyses and characterization of phosphoamino acids**

Protein T variants, protamine Sp2 and their peptides were hydrolysed under vacuum in 6 M HCl containing 50 μl/ml of 5% phenol at 110 °C for 24 h. Amino acid analyses were performed on a Beckman 6300 amino acid analyser.

The presence of phosphorylated residues was investigated by amino acid analysis after a 2 h hydrolysis in 6 M HCl at 110 °C under vacuum.

**Enzymic hydrolyses**

Endoproteinase Lys-C

Squid proteins T1 (17 nmol, 190 μg) and T2 (35 nmol, 385 μg) were digested with endoproteinase Lys-C (lysyl endopeptidase, EC 3.4.21.50; Boehringer) at 30 °C for 1 h in 0.2 ml of 25 mM Tris/HCl buffer, pH 7.7, containing 1 mM EDTA, with an enzyme-to-substrate mass ratio of 1:25.

Chymotrypsin

Squid proteins T1 (24 nmol, 265 μg), T2 (48 nmol, 530 μg) and protamine Sp2 (26 nmol, 220 μg) were hydrolysed with tosyl-lysylchloromethane-treated chymotrypsin (EC 3.4.21.1; Merck) at 30 °C for 4 h in 0.2 ml of 0.1 M ammonium acetate buffer, pH 5.0, using an enzyme-to-substrate mass ratio of 1:200. The C-terminal peptides generated by cleavage of protein T variants with endoproteinase Lys-C were submitted to the action of chymotrypsin under the same conditions.

Endoproteinase from *Astacus fluviatilis*

Proteins T1 (17 nmol, 190 μg), T2 (34 nmol, 375 μg), protamine Sp2 (20 nmol, 170 μg) and the C-terminal peptides obtained by hydrolysis of the variants T1 and T2 with endoproteinase Lys-C were hydrolysed with *Astacus* proteinase (EC 3.4.24.21; Serva) in 0.2 ml of 0.2 M ammonium bicarbonate, pH 8.1, for 20 min at 30 °C, using an enzyme-to-substrate mass ratio of 1:50.

Enzymic hydrolyses were stopped by lowering the pH to 3.0 with 70% formic acid and the hydrolysates were evaporated to dryness before fractionation by reverse-phase HPLC, as described in [7].

**Sequence analysis**

Sequencing of whole protein T, variants T1 and T2, protamine Sp2 and their peptides was performed on a gas-phase sequencer (Applied Biosystems 470A) using the 03CArg program [8]. Phenylthiohydantoin derivatives of amino acids were identified on-line as described previously [9].

**MS**

The molecular masses of the variants T1 and T2 of the squid testis-specific protein T, of the protamine Sp2 and of some peptides were determined by electrospray MS (ES-MS) either on a VG BIO-Q quadrupole mass spectrometer as described in [10] or on a VG Biotech BioQ mass spectrometer (VG Biotech, Altrincham, U.K.) with a mass range of 4000 Da. Scanning was usually performed from m/z = 400 to m/z = 1400 in 10 s with the resolution adjusted so that the m/z = 998 peak from horse heart myoglobin was 1.5–2.0 Da wide at the base.

Data were acquired by operating the data system as a multi-channel analyser and several scans were summed to obtain the final spectrum. Calibration was performed using the multiply charged ions from a separate analysis of horse heart myoglobin (16951.4 Da).

The electrospray ion source was operated at atmospheric pressure with an extraction cone voltage value of 55 V. Each molecular species produced a series of multiply charged protonated molecular ions from which the molecular mass was determined by a simple data-system routine. Samples (15–20 pmol/μl) were introduced into the ion source by syringe pumps (Applied Biosystems Inc. model 140A, Foster City, CA, U.S.A.) at a flow rate of 4 μl/min.

The mass of some peptides was measured by fast atom bombardment-MS (FAB-MS) on a CONCEPT II HH (Kratos Analytical, Manchester, U.K.) four-sector tandem mass spectrometer. Peptide sample (500 pmol in 1 μl of de-ionized water) was loaded on to the target and 1 μl of matrix was added. The matrix was constituted of glycerol/water/TFA (10:88:2, by vol.). Caesium iodide was used as standard compound for mass calibration of samples.

**RESULTS AND DISCUSSION**

The crude *Loligo* testis-specific protein T (Figure 1, lane 2) was purified by reverse-phase HPLC and recovered into two well-resolved fractions, corresponding to variants T1 and T2 respectively (Figure 2). The major variant T2 represents about 70% of the whole fraction, as estimated from amino acid analysis.

The amino acid compositions of the two *Loligo* protein T variants are mainly characterized by a high amount of arginine (about 60%) and a relative abundance of serine (7–8%). Based on the amino acid compositions, the major *Loligo* protein variant T2 is more closely related to the minor cuttlefish protein variant T2 and the minor *Loligo* protein variant T1 looks like the major cuttlefish protein variant T1 [2]. Interestingly, *Loligo* protein T exhibits more similarity with cuttlefish spermatid-specific protein T than with protein T1 isolated from immature testes of another squid *I. argentinus* [5]. Protein T1 differs from *Loligo* and *Sepia* proteins T by having a much lower amount of arginine (45%), a higher level of lysine and serine, and by the presence of at least two histidine residues.

Figure 3 shows the electrospray mass spectrum of the major variant T2 from *L. pealeii*. Several series of multiply charged ions were detected: they correspond to the calculated molecular mass with 2–6 additional phosphate groups (the mass of a phosphate
3.2, pH destained mature sperm histone from protamine.

Electrophoresis

Crude protein T (1 mg) was dissolved in 0.01 M HCl and loaded on to a Superspher 100 RP-18 endcapped column (250 x 4 mm) equilibrated in 0.1% TFA. Elution was performed with a linear gradient of acetonitrile (0–50%) in 0.1% TFA. Elution of the fractions was monitored at 220 nm. Fractions of 0.5 ml were collected at a flow rate of 1 ml/min and pooled as indicated by solid bars.

Figure 1: Electrophoretic analysis of crude testis-specific protein T isolated from squid testis chromatin and of protamine Sp extracted from squid mature sperm nuclei.

Electrophoresis was performed as described in [3] on slab gel (160 x 160 x 0.75 mm) at pH 3.2, in 0.9 M acetic acid/6.25 M urea, using a 17% acrylamide concentration. Samples (3 µg), dissolved in 0.01 M HCl/8 M urea/0.5 M 2-mercaptopropanol, were run at 22 mA for 2.5 h at 10 °C after a pre-electrophoresis of 1.5 h. The gel was stained overnight with 0.1% Coomassie blue R-250 in the mixture water/methanol/acetic acid (5/1/5, by vol.), then destained by diffusion in water/methanol/acetic acid (35/20/28, by vol.). Lanes 1 and 4, whole histone from calf thymus as reference; lane 2, crude squid protein T; lane 3, crude squid protamine Sp.

Figure 2: Separation by reverse-phase HPLC of squid testis-specific protein T variants

Crude protein T (1 mg) was dissolved in 0.01 M HCl and loaded on to a Superspher 100 RP-18 endcapped column (250 x 4 mm) equilibrated in 0.1% TFA. Elution was performed with a linear gradient of acetonitrile (0–50%) in 0.1% TFA. Elution of the fractions was monitored at 220 nm. Fractions of 0.5 ml were collected at a flow rate of 1 ml/min and pooled as indicated by solid bars.

group is 80 Da). The two major forms are tri- and tetra-phosphorylated (molecular masses 11031 and 11113 Da respectively). A similar pattern is obtained from the minor variant T1 with very close masses, 11026 and 11105 Da, for the tri- and tetra-phosphorylated molecules respectively.

Each protein variant was submitted to automated Edman degradation. Their N-terminal sequences were determined up to residue 39 for T1 and 42 for T2. Unlike the similar cuttlefish protein, all molecular species of Loligo protein T contain a methionine residue at the N-terminal position.

To establish the sequence of the C-terminal region, Loligo proteins T1 and T2 were first digested by endoproteinase Lys-C in order to generate the C-terminal peptide K-5 (residues 20–79). Chymotrypsin and Astacus protease were used to obtain the overlapping peptides necessary to extend the amino acid sequence up to the C-terminus. Interestingly, the particular pH conditions (pH 5.0) used with chymotrypsin [11] led to the cleavage of only the tyrosyl bonds, avoiding the cleavage of the seven seryl bonds present in the C-terminal part of the variants (residues 22–79).

The amino acid compositions and the sequences of the peptides used for the elucidation of the complete sequences of Loligo protein T variants were established and, in most cases, substantiated by MS analysis of the peptides either in FAB or electrospray mode. The peptides T1 K-5 and T1 C-3 (Figure 4) exhibit molecular masses which differ from the calculated masses. These differences correspond to the presence of additional phosphate residues. Thus, K-5 appears as a mixture of tri- and tetra-phosphorylated species whereas C-3 is a mixture of non- and mono-phosphorylated forms.

The amino acid sequences of Loligo protein variants T1 and T2 (79 residues) are presented in Figures 4 and 5. The alignment of chymotryptic peptides was established with the peptides generated by Astacus protease digestion of fragment K-5. These two variants are structurally related proteins which differ only by a few conservative changes in the N-terminal region (residues 1–21), the distribution of serine residues in the central region (residues 33–59) and the substitution of a tyrosine residue by an arginine residue at position 50 in variant T2. They are also structurally related to Sepia protein variants T1 and T2 respectively: the variant T1, minor in Loligo, is major in Sepia whereas the variant T2, major in Loligo, is minor in Sepia [2]. The T1 proteins present in both species differ only in the presence of an additional arginine in Loligo compared with Sepia. A few other changes are found in T2.

Like similar cuttlefish proteins, Loligo testis-specific proteins T display two well-defined domains: a relatively hydrophobic N-terminal domain (residues 1–21), which could play a vital role in mediating the folding of the molecule into a bioactive conformation allowing correct positioning on DNA fibres, and a C-terminal domain (residues 22–79) with polyarginine clusters separated by serine and/or tyrosine residues (Figure 6). Furthermore, the sequence SRSPYRR (residues 42–48) and the palindromic sequence RRSSYSRR (residues 65–73) are strictly conserved in protein T from both species. The clustering of neutral amino acids in such a highly charged protein is surprising and suggests a very precise structural role for these uncharged groups of amino acids. The former sequence is similar to the sequence SPKK which is involved in the preferential linkage of testis-specific histones H1 and H2B to AT-rich sites in the minor groove of DNA [12]. Both sequences might have a critical role in DNA–protein T interaction which could be modulated by the specific phosphorylation of serine residues at positions 42 and 68. These residues were found to be phosphorylated by Edman degradation or amino acid analysis of a 2 h hydrolysate of the relevant peptides. Serine residues 44 and 70 have not been found to be phosphorylated.

Furthermore, the sequence of the Loligo major protamine variant Sp2 isolated from mature sperm cells has been deduced...
Figure 3  Electrospray mass spectrum of testis-specific protein T2 from the squid L. pealeii

Five series (A-E) of multicharged ions with 11–15 charges were detected. The first series (A: 10951 ± 3, 2P) corresponds to the expected mass for a diphosphorylated species. The major series (B: 11021 ± 1, 3P and C: 11113 ± 5, 4P) correspond to tri- and tetra-phosphorylated species respectively. The series D (11190 ± 2, 5P) and E (11271 ± 4, 6P) correspond to the addition of 5 and 6 phosphate groups respectively.

![Mass Spectrum Diagram]

Figure 4  Complete amino acid sequence of testis-specific protein T1 from the squid L. pealeii

Methods used for the determination of the sequence are indicated as follows: bold-faced type, automated Edman degradation of the isolated variant T1. Automated Edman degradation of peptides, K- and C- indicate peptides obtained by cleavage of protein T1 with endoproteinase Lys-C and chymotrypsin respectively. A- indicates peptides obtained by cleavage of peptide K-5 with Astacus proteinase. The molecular masses (m) expressed in Da are indicated below the peptides and were measured by ES-MS for peptides K-5 and C-3 and by FAB-MS for the others. Peptide K-5 was identified as tri- and tetra-phosphorylated forms and peptide C-3 as non- and mono-phosphorylated forms.

![Amino Acid Sequence Diagram]

Figure 5  Complete amino acid sequences of testis-specific protein T2 and of protamine Sp2 from the squid L. pealeii

Methods used for the determination of the sequence are indicated in the legend to Figure 4. K-, C- and A- indicate peptides obtained by cleavage of protein T2 with endoproteinase Lys-C, chymotrypsin and Astacus proteinase respectively. Peptides obtained from protamine Sp2 by cleavage with chymotrypsin and Astacus proteinase are designated by c- and a- respectively. The molecular masses (m) expressed in Da are indicated below the peptides and were measured by FAB-MS.

![Amino Acid Sequence Diagram]
The amino acid site of the interaction.

Figure 6  Comparison of the amino acid sequences of proteins T1 and T2 from the squid *L. pealeii* and the cuttlefish *S. officinalis*

The amino acid sequences of cuttlefish protamines Sp1 and Sp2 are added to show up the structural identity with the C-terminal domain of cuttlefish T1 and T2 and to point out the cleavage site of the pro-protamine. Boxes indicate strict identity between the six proteins. Shaded residues indicate the strictly conserved sequences that are likely to be involved in the DNA-protein T interaction.

Figure 7  Electrospray mass spectrum of protamine Sp2 from squid *L. pealeii*

One series of multi-charged ions with 8–13 charges was detected. It corresponds to a molecular mass of 8563.5 ± 0.9 Da.

non-phosphorylated protamine molecule. As observed in cuttlefish spermiogenesis, the phosphorylated protamine present in elongated spermatids is likely to be dephosphorylated at the early stage of epididymal transit [4]. The total identity between the C-terminal sequence (residues 22–79) of *Loligo* testis-specific protein T2 and protamine Sp2 strongly suggests that protein T2 is indeed the precursor of protamine Sp2.

An identical situation has already been observed in another Cephalopod, the cuttlefish *S. officinalis* [2]. The transition pro-protamine → protamine which, in both species, occurs during the
late phase of spermatogenesis, probably results from the same
processing mechanism. This mechanism, involving the specific
 cleavage of a Gly–Arg bond in the sequence Met/Leu\textsuperscript{14}–Lys-Gly-
Gly-Arg-Arg\textsuperscript{19} (Figure 6), remains to be elucidated. It could
imply an N-arginine dibasic convertase analogous to that
recently characterized in rat brain cortex and testis [13,14]. By
immunofluorescence, the enzyme was shown to be distributed in
the cytoplasm of round and elongating spermatids. It must be
emphasized that cuttlefish pro-protamine appears in round
spermatid nuclei, is abundant in elongating spermatid nuclei and
is replaced by protamine in elongated spermatid nuclei [15].

The one-step transition pro-protamine $\rightarrow$ protamine in Loligo
and Sepia relies upon a processing mechanism quite different from
that involved in the multi-step transition pro-protamine $\rightarrow$
protamine P2 in mammals, such as man [9,16,17] and mouse
[18,19]. In this case, the lack of any consensus sequence would
imply either the intervention of several processing enzymes of
different specificities or the cleavage with a protase of broad
specificity at exposed bonds of the folded structure of the pro-
protamine P2, as suggested by Retief and Dixon [1].

Interestingly, the study of the nuclear basic proteins which
replace histones during spermiogenesis of Cephalopods, the most
evolutionarily organized molluscs, is linked to the phylogenetic
relationships among decapod Cephalopods. Indeed, two major
morphological criteria are used to establish the phylogenetic tree
of decapod Cephalopods: the nature of the shell and the eye
type. Whatever the classification, the squid I. argentinus and the
cuttlefish S. officinalis are clearly separated. In contrast, the
position of Loligo remains ambiguous: Loligo is either placed
close to Illex according to the shell type or placed close to Sepia
according to the eye type [20]. Our study indicates that Loligo
and Sepia are very similar in the protamines found in their
spermatozoa, a feature which is in agreement with the classi-

fication of Decapoda suggested by Berthold and Engeser [21].

The authors are grateful to Dr. J. Ausio who provided some of the squid material and
to A. Hénez and J. Herno for their skilful technical assistance. They thank M.J.
Dupire for amino acid analyses and O. Hamedouche for sequencing. This work was
supported by grants from the Centre National de la Recherche Scientifique, from the
Université de Lille II, from the Commission of the European Communities (grant
CT91-0619) and from DGICYT (grant PB 93-1067).

REFERENCES

2 Wouters-Tyrou, D., Chartier-Harlin, M. C., Martin-Ponthieu, A., Boutilier, C., Van
Eur. J. Biochem. 220, 463–468
609–613
7 Martin-Ponthieu, A., Wouters-Tyrou, D., Bélaïche, D., Sautière, P., Schindler, P. and
Chem. 263, 11059–11062
10 Schindler, P., Bilotch, F., Klarakov, K., Roepstorff, P., Briand, G., Wouters-Tyrou, D.,
(1994) J. Biol. Chem. 269, 2056–2061
14 Pierrotti, A. R., Prat, A., Chesneau, V., Gaudoux, F., Leseney, A. M., Foulon, T. and
15 Rousseaux-Prestot, R., Engelhardt, B. P., Rousseaux, J., Wouters-Tyrou, D. and
Biochem. 181, 449–451
17 Alimi, E., Martinage, A., Arkhis, A., Belaïche, D., Sautière, P. and Chevallier, P.
J. Biochem. 204, 759–764
40, 84–90
Vie 317, 581–588

Received 7 December 1994/20 February 1995; accepted 7 March 1995