Nuclear transition protein 1 from ram elongating spermatids. Mass spectrometric characterization primary culture and phosphorylation sites of two variants

F. Chirat, A. Martinage, Gwenola Briand, M. Kouach, Alain van Dorsselaer, Maurice Loir, Pierre Sautière

▶ **To cite this version:**

F. Chirat, A. Martinage, Gwenola Briand, M. Kouach, Alain van Dorsselaer, et al.. Nuclear transition protein 1 from ram elongating spermatids. Mass spectrometric characterization primary culture and phosphorylation sites of two variants. European Journal of Biochemistry, 1991, 198, pp.12-20. 10.1111/j.1432-1033.1991.tb15980.x . hal-02716129

HAL Id: hal-02716129
https://hal.inrae.fr/hal-02716129

Submitted on 1 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Eur. J. Biochem. *198*, 13–20 (1991)
© FEBS 1991
001429569100303Q

Nuclear transition protein 1 from ram elongating spermatids

Mass spectrometric characterization, primary structure and phosphorylation sites of two variants

Frédéric CHIRAT[1], Arlette MARTINAGE[1], Gilbert BRIAND[1], Mostafa KOUACH[1], Alain VAN DORSSELAER[2], Maurice LOIR[3] and Pierre SAUTIÈRE[1]

[1] Unité de Recherche Associée n° 409 au Centre National de la Recherche Scientifique, Université de Lille II, Institut de Recherches sur le Cancer, Lille, France
[2] Laboratoire de Chimie Organique des Substances Naturelles, Unité de Recherche Associée n° 31 au Centre National de la Recherche Scientifique, Strasbourg, France
[3] Laboratoire de Physiologie des Poissons, Institut National de la Recherche Agronomique, Campus de Beaulieu, Rennes, France

The ram transition protein 1 (TP1) is present in spermatid cell nuclei in the nonphosphorylated, monophosphorylated and diphosphorylated forms. Its primary structure was determined by automated Edman degradation of S-carboxamidomethylated protein and of peptides generated by cleavage with thermolysin and endoproteinase Lys-C.

The ram TP1 is a small basic protein of 54 residues and structurally very close to other mammalian TP1. The mass spectrometric data obtained from the protein and its fragments reveal that ram TP1 is indeed a mixture (approximately 5:1) of two structural variants (M_r 6346 and 6300). These variants differ only by the nature of the residue at position 27 (Cys in the major variant and Gly in the minor variant). The study of phosphorylation sites has shown that four different serine residues could be phosphorylated in the monophosphorylated TP1, at positions 8, 35, 36 or 39.

From previous physical studies, it has been postulated that the Tyr32 surrounded by two highly conserved basic clusters was responsible for the destabilization of chromatin by intercalation of its phenol ring between the bases of double-stranded DNA.

The presence of three phosphorylatable serine residues in the very conserved sequence 29–42 is another argument for the involvement of this region in the interaction with DNA.

Spermatogenesis involves a progressive differentiation of the germ cells towards a terminal stage represented by the mature spermatozoa.

In mammals, the first part of spermiogenesis where the spermatid development takes place, is quiescent relative to changes in chromatin structure, while the second part of this stage is characterized by a progressive condensation of the diffuse chromatin to a highly compact and transcriptionally inactive structure [1–6].

This condensation is correlated with a double-protein transition. The first transition corresponds to the complete removal and replacement of histones (somatic and testis-specific variants) by several spermatid-specific basic proteins [7–9] which are themselves eliminated and replaced by one [10–16] or two protamines [17–23] during the second transition.

Correspondence to P. Sautière, URA 409 CNRS, Université de Lille II, Institut de Recherches sur le Cancer, Place de Verdun, F-59045 Lille Cédex, France

Abbreviations. ESMS, electrospray mass spectrometry; FAB-MS, fast-atom-bombardment mass spectrometry; PDMS, plasma-desorption mass spectrometry; TP, transition protein.

Enzymes. Carboxypeptidase A (EC 3.4.17.1); endoproteinase Lys-C (EC 3.4.21.50); thermolysin (EC 3.4.24.4).

While histones and protamines have been extensively studied, little is known about spermatid-specific proteins also called transition proteins.

In rat and mouse four transition proteins [7, 8] named TP1, TP2, TP3 and TP4 have been identified. Among them, only TP1 and TP2 have been fully characterized.

Rat and mouse TP1 are low molecular mass basic proteins (54 residues), rich in arginine (20%), lysine (19%) and serine (14%) and devoid of cysteine [24–27].

TP2 is larger than TP1 (116 and 117 residues in rat and mouse, respectively) [28, 29]. This protein is characterized by a high amount of serine (22%) and proline (19%) and by the presence of cysteine (about 5%).

Little is known about transition proteins in other mammalian species. In man, bull and boar, a transition protein related to TP1 has been identified and sequenced [30, 31]. Like rat and mouse TP1, human TP1 lacks of cysteine, while boar and bull TP1 contain one cysteine residue.

In ram, four transition proteins called P1, P3, P7 and T have been purified and characterized [9]. The amino acid composition and the C-terminal sequence (27 residues) of ram protein P1 (F. Chirat, unpublished results) clearly indicate a close structural relationship between this protein and rat and mouse TP2, whilst the amino acid composition of ram protein

T [9] is very similar to those of the other mammalian TP1. Moreover, this protein was shown to be monophosphorylated and diphosphorylated to a small extent [9].

In this paper, we report the complete primary structure of the ram protein T and the different sites of phosphorylation identified in the monophosphorylated form of the protein. Since the sequence analysis of ram protein T has shown a close structural relationship with known mammalian TP1 proteins [31], the ram protein T will be named from now, for the sake of clarity, ram transition protein 1 or ram TP1.

MATERIALS AND METHODS

See Supplementary section.

RESULTS

Ram TP1 is present in elongating spermatid nuclei in three forms: nonphosphorylated, monophosphorylated and diphosphorylated [9]. These forms were separated by fractionation of the 0.75-M-perchloric-acid-soluble cysteine-containing proteins extracted from EDTA-resistant spermatid nuclei, on a CM-Sephadex C25 column. Prior to ion-exchange chromatography, the spermatidal proteins were reduced and alkylated with iodoacetamide.

The nonphosphorylated and monophosphorylated forms of TP1 were obtained in a highly pure state, while the diphosphorylated form coeluted with the spermatidal protein 7.

Amino acid sequence of ram TP1

The strategy and sequence data which allowed us to elucidate the complete sequence of ram TP1 are presented in Fig. 1. The elution diagrams of the peptides derived from the protein and the table providing sequence data are given in the Supplementary section (Figs S1 and S2, Table S1).

The amino-terminal sequence of TP1 was determined up to residue 38.

The digestion of the S-carboxamidomethylated protein with carboxypeptidase A for 15 min released 0.9 nmol leucine/nmol protein and 0.5 nmol histidine/nmol protein.

The remainder of the sequence was unambiguously established by sequencing peptides derived from hydrolysis of the S-carboxamidomethylated protein with thermolysin and endoproteinase Lys-C.

Cleavage of the protein with thermolysin generated six peptides (Th1 to Th6) which altogether account for the total number of amino acids of the TP1 molecule (54 residues) (Table 1).

Among the peptides derived from the cleavage of the protein with endoproteinase Lys-C, only three peptides K1, K2 and K3 were useful in providing the necessary overlaps for the alignment of the thermolysin peptides Th4, Th5 and Th6 (Table S1). The peptides K1 and K2 were obtained as a mixture in equal amounts as shown by the amino acid composition. The correct assignment of the Edman data to each peptide was possible only because the sequence was known at that point.

Mass spectrometry of the S-carboxamidomethylated protein

Electrospray mass spectrometry (ESMS) of the nonphosphorylated form of TP1 yields two fragments having

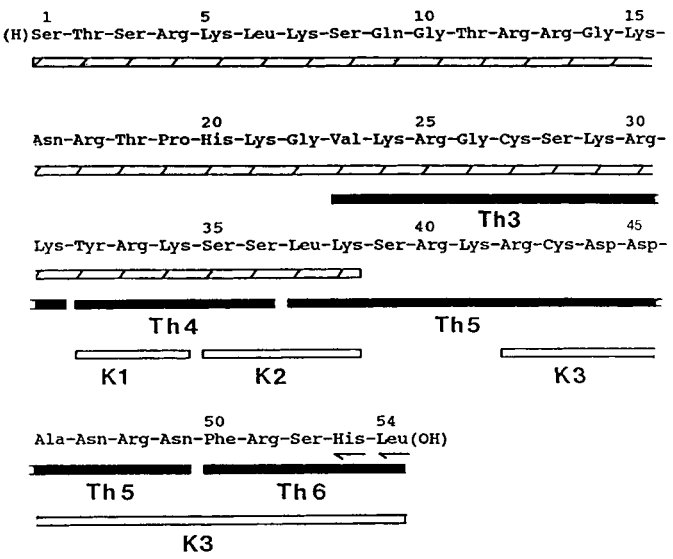

Fig. 1. *Complete amino acid sequence of ram TP1 protein.* (Th) and (K) indicate peptides derived from cleavage with thermolysin and endoproteinase Lys-C, respectively. (▨) Automated Edman degradation of the S-carboxamidomethylated ram TP1; (■) automated Edman degradation of thermolysin peptides (Th); (□) automated Edman degradation of endoproteinase Lys-C peptides (K); (—) carboxypeptidase A digestion

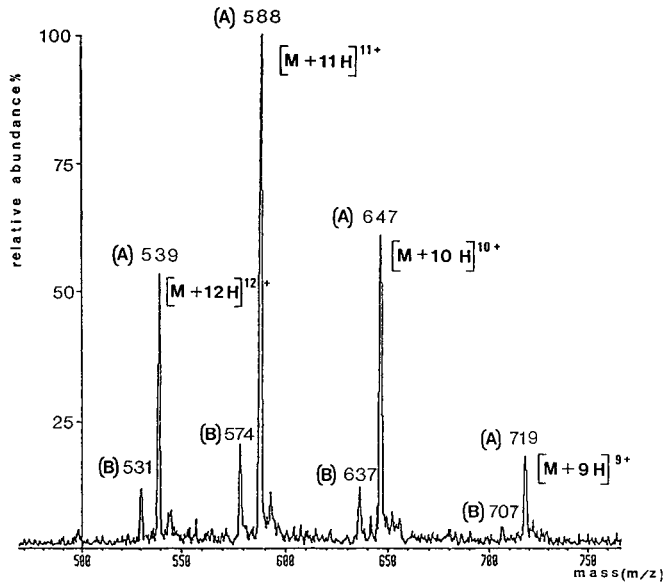

Fig. 2. *Section of the ESMS spectrum of ram TP1.* Two series (A) and (B) of multicharged ions with 9 to 12 charges were detected. The major series (A) corresponds to the major variant (6460 Da), and the minor series (B) corresponds to the minor variant (6357 Da)

molecular masses of 6460 Da and 6357 Da (Fig. 2 and Table 1). The fragment of molecular mass 6460 Da corresponds to the presence of two S-carboxamidomethylated cysteine residues. The other fragment, with a molecular mass of 6357 Da, is in agreement with the replacement of one cysteine residue by a glycine residue.

From these data, the two components were estimated to be present in a 5:1 ratio.

Table 1. *Amino acid composition of ram TP1 and peptides generated by cleavage with thermolysin (Th) and with endoproteinase Lys-C (K)* Values in parentheses are the number of residues/molecule of protein or peptide derived from the sequence. CmCys, *S*-carboxymethyl cysteine; n.d., not determined

Amino acid	TP1	Amino acid content of TP1 fragment (residue nos)							
		Th1 (1–5)	Th2 (6–22)	Th3 (23–31)	Th4 (32–36)	Th5 (37–49)	Th6 (50–54)	K1/2 (32–34/ 35–38)	K3 (42–54)
	mol/mol								
CmCys	2.1(2)	–	–	0.8(1)	–	0.9(1)	–	–	0.8(1)
Asx	5.0(5)	–	1.2(1)	–	–	4.0(4)	–	–	4.1(4)
Thr	2.9(3)	1.0(1)	2.0(2)	–	–	–	–	–	–
Ser	7.6(8)	2.0(2)	1.5(1)	1.4(1)	2.1(2)	1.5(1)	1.3(1)	1.6(2)	1.5(1)
Glx	1.3(1)	–	1.1(1)	–	–	–	–	–	–
Pro	1.2(1)	–	1.0(1)	–	–	–	–	–	–
Gly	4.3(4)	–	3.0(3)	1.3(1)	–	–	–	–	–
Ala	1.3(1)	–	–	–	–	1.1(1)	–	–	1.2(1)
Val	1.1(1)	–	–	1.0(1)	–	–	–	–	–
Met	–	–	–	–	–	–	–	–	–
Ile	–	–	–	–	–	–	–	–	–
Leu	3.0(3)	–	1.0(1)	–	–	1.0(1)	1.0(1)	1.0(1)	1.2(1)
Tyr	1.1(1)	–	–	–	0.9(1)	–	–	0.7(1)	–
Phe	1.0(1)	–	–	–	–	–	1.0(1)	–	1.0(1)
His	1.9(2)	–	1.0(1)	–	–	–	1.1(1)	–	1.0(1)
Lys	10.3(10)	1.0(1)	3.1(3)	3.0(3)	1.0(1)	2.1(2)	–	2.0(2)	–
Arg	11.3(11)	0.9(1)	3.2(3)	2.0(2)	1.0(1)	3.2(3)	0.9(1)	1.1(1)	3.0(3)
Mass[a]	6462	–	–	1117.6	–	1632.8	–	–	–
Mass[b]	6460 6357	n.d.	n.d.	1117.6 1014.6	n.d.	1633.8	n.d.	n.d.	n.d.
Total[c]	54	5	17	9	5	13	5	3/4	13

[a] Calculated masses in Da.
[b] Masses in Da, measured by mass spectrometry (ESMS for TP1, PDMS for Th5 and FAB-MS for Th3).
[c] Total number of amino acid residues.

Mass spectrometry of S-carboxamidomethylated thermolysin peptides Th3 and Th5

When subjected to fast-atom-bombardment mass spectrometry (FAB-MS), peptide Th3 (residues 23–31) yielded two fragments having molecular masses of 1117.6 Da and 1014.6 Da (Fig. 3 and Table 1).

The difference (103 Da) between these two masses is identical to that observed between the two molecular masses determined for the S-carboxamidomethylated protein TP1.

The analysis of peptide Th5 (residues 37–49) by ^{252}Cf plasma-desorption mass spectrometry (PDMS) gave a single mass of 1633.8 Da which corroborates the sequence determined by the automated Edman degradation. It must be emphasized that FAB-MS failed to give the mass of this peptide.

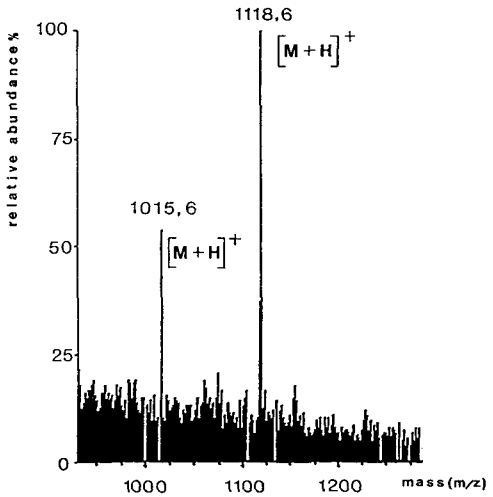

Fig. 3. *Section of the FAB-MS spectrum of peptides Th3 derived from cleavage of S-carboxamidomethylated ram TP1 with thermolysin*

Identification of the phosphorylation sites in ram TP1

The phosphorylation sites were identified in the monophosphorylated form of ram TP1. Due to the low amount of protein available, no mass determination of phosphorylated TP1 by ESMS was carried out.

The S-carboxamidomethylated derivative of the monophosphorylated protein was cleaved by thermolysin. The elution pattern of the digest presented in Fig. S1B shows an additional peptide, Thx, by comparison with the elution profil of thermolysin peptides generated from the nonphosphorylated protein (Fig. S1A). Among the seven peptides obtained, only peptides Th2 and Thx were found to contain

phosphoamino acid(s) upon amino acid analysis after a 2-h hydrolysis in 6 M HCl at 110 °C.

These two peptides were treated according to the procedure described by Meyer et al. [33] in order to convert the phosphoserine residues into S-ethyl cysteine residues.

The sequence data presented in Table S2 led us to conclude that serine residues at positions 8, 35, 36 and 39 are partially phosphorylated in the monophosphorylated form of the protein. The data also indicate that peptides Th2 and Thx were cross contaminated and that peptide Thx corresponds to the sequence 32–49 of ram TP1.

In the course of the sequence analysis, we have not observed the two peaks characteristic of the phenylthiohydantoin derivative of the carboxamidomethylated cysteine expected at position 12 (Cys43 in TP1), but another peak with a retention time identical to that of the phenylthiohydantoin derivative of S-ethyl cysteine. To check the possibility that the procedure of Meyer et al. [33], could also affect the S-carboxamidomethylated cysteine, we submitted peptide Th3, VKRGCSKRK (Table 1 in Supplementary section) (residues 23–31) generated from the non-phosphorylated TP1, which contains one S-carboxamidomethylated cysteine at position 27, to an automated Edman degradation. This peptide was sequenced twice, before and after derivatization according to the procedure of Meyer et al. [33]. The sequence data, presented in the Fig. S3 obviously show that the S-carboxamidomethylated cysteine is also transformed by this procedure.

A complete study of this transformation, the mechanism of which does not appear to be a β-elimination, will be published elsewhere.

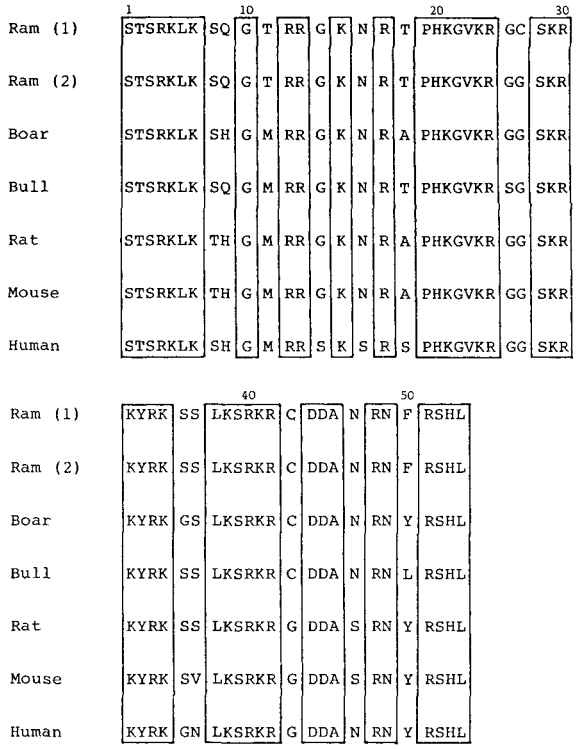

Fig. 4. *Comparison of the amino acid sequence of TP1 from ram, bull [31], boar [31], rat [25, 26], mouse [27] and man [28].* The boxes correspond to sequence identity between the different TP1

DISCUSSION

From mass spectrometric data, TP1 was found to be a mixture of two structural variants which only differ by a single change: the cysteine residue at position 27 in the major variant is replaced by a glycine residue in the minor one.

The contribution of mass spectrometry was essential to reveal the existence of the minor variant. Indeed, during the course of the automated Edman degradation of the protein, the identification of glycine at position 27 was hindered, both the low amount of this variant and the carry-over of glycine from the previous cycle.

With 43% of basic residues, ram TP1 has a basicity intermediate between that of histone H4 (26%) [35] and that of ram protamine (56%) [13]. The basic amino acids are distributed all along the polypeptide chain, most of them arranged in clusters of two or three residues. This protein is also characterized by a relative abundancy of hydroxylated residues (22%) which accumulate in the first 75% of the molecule. The occurrence of two basic clusters interspersed with one tyrosine residue at position 32 and three serine residues at positions 35, 36 and 39, constitutes the most interesting feature of ram TP1.

These residues of serine were found to be phosphorylated in the monophosphorylated form of TP1. Two of them (Ser35 and Ser36) are present in a sequence Arg/Lys-Xaa-Ser specifically recognized by the cyclic AMP-dependent protein kinase, where Xaa represents any amino acid except proline [36, 37]. The two other sites of phosphorylation identified in ram TP1, are located in the sequence LKS, which occurs twice in the polypeptide chain, at positions 6–8 and 37–39 respectively. Thus, the presence of four phosphorylation sites in the monophosphorylated form of ram TP1 indicates a lack of specificity in the phosphorylation of this protein.

All the known mammalian TP1 exhibit a high degree of similarity (Fig. 4) particularly in the very basic region 29–42 which contains one residue of tyrosine at position 32 and at least one phosphorylatable residue of serine at position 35, 36 or 39. In this region the tyrosine residue, likely to be involved in interactions with DNA, could be essential for the destabilization of the chromatin through its intercalation between the bases of DNA as postulated by Singh and Rao [38]. Moreover, the phosphorylation of serine residues which is known to modulate interaction of basic proteins (i.e. histones and protamines) with DNA [39–42] would allow the correct positioning of TP1 on DNA. Later on, the dephosphorylation of TP1 would bring together the region 29–42 and DNA, with as a consequence, the intercalation of Tyr32 between the DNA bases, thus inducing the destabilization of the chromatin necessary to the complete removal of histones before the deposition of protamines.

We are indebted to K. Klarskov and P. Roepstorff (University of Odense, Denmark) for the PDMS analysis of peptide Th5. We thank M. J. Dupire for amino acid analysis, the *Service commun de séquence peptidique (URA 409 Centre National de la Recherche Scientifique, U 16 Institut National de la Santé et de la Recherche Médicale*) and T. Ernout for the preparation of this manuscript. This work was supported by grants from the *Centre National de la Recherche Scientifique*, from the *Université de Lille II* and from the *Fondation pour la Recherche Médicale.*

REFERENCES

1. Kierszenbaum, A. L. & Tres, L. L. (1975) *J. Cell Biol. 65*, 258–270.

2. Meistrich, M. L., Reid, B. O. & Barcellona, W. J. (1976) *Exp. Cell Res. 99*, 72–78.
3. Loir, M. & Lanneau, M. (1978) *Exp. Cell Res. 115*, 231–243.
4. Meistrich, M. L., Trostle, P. K. & Brock, W. A. (1981) in *Bioregulators of development* (Jagiello, G. & Vogel, H. J., eds) pp. 151–166, Academic Press, New York.
5. Loir, M. & Lanneau, M. (1984) *J. Ultrastruct. Res. 86*, 262–276.
6. Loir, M., Bouvier, D., Fornells, M., Lanneau, M. & Subirana, J. A. (1985) *Chromosoma (Berl.) 92*, 304–312.
7. Meistrich, M. L., Brock, W. A., Grimes, S. R., Platz, R. D. & Hnilica, L. S. (1978) *Fed. Proc., FASEB. 37*, 2522–2525.
8. Balhorn, R., Weston, S., Thomas, C. & Wyrobeck, A. J. (1984) *Exp. Cell Res. 150*, 298–308.
9. Dupressoir, T., Sautière, P., Lanneau, M. & Loir, M. (1985) *Exp. Cell Res. 161*, 63–74.
10. Coelingh, J. P., Monfoort, C. H., Rozijn, T. H., Gevers Leuven, J. A., Schiphof, R., Steyn-Parvé, E. P., Braunitzer, G., Schrank, B. & Ruhfus, A. (1972) *Biochim. Biophys. Acta 285*, 1–14.
11. Calvin, H. I. (1976) *Biochim. Biophys. Acta 434*, 377–389.
12. Tobita, T., Tsutsumi, H., Kato, A., Suzuki, H., Nomoto, M., Nakano, M. & Ando, T. (1983) *Biochim. Biophys. Acta 744*, 141–146.
13. Sautière, P., Bélaïche, D., Martinage, A. & Loir, M. (1984) *Eur. J. Biochem. 144*, 121–125.
14. Mazrimas, J. A., Corzett, M., Campos, C. & Balhorn, R. (1986) *Biochim. Biophys. Acta 872*, 11–15.
15. Ammer, H. & Henschen, A. (1988) *FEBS Lett. 242*, 111–116.
16. Ammer, H. & Henschen, A. (1988) *Biol. Chem. Hoppe-Seyler 369*, 1301–1306.
17. McKay, D. J., Renaux, B. S. & Dixon, G. H. (1985) *Biosci. Rep. 5*, 383–391.
18. Ammer, H., Henschen, A. & Lee, C. H. (1986) *Biol. Chem. Hoppe-Seyler 367*, 515–522.
19. Gusse, M., Sautière, P., Bélaïche, D., Martinage, A., Roux, C., Dadoune, J. P. & Chevaillier, Ph. (1986) *Biochim. Biophys. Acta 884*, 124–134.
20. McKay, D. J., Renaux, B. S. & Dixon, G. H. (1986) *Eur. J. Biochem. 156*, 5–8.
21. Bélaïche, D., Loir, M., Kruggle, W. & Sautière, P. (1987) *Biochim. Biophys. Acta 913*, 145–149.
22. Bellve, A. R. (1988) *Biochemistry 27*, 2890–2897.
23. Pirhonen, A., Linnala-Kankkunen, A. & Mäenpää, P. H. (1989) *FEBS Lett. 244*, 199–202.
24. Kistler, W. S., Geroch, M. E. & Williams-Ashman, H. G. (1973) *J. Biol. Chem. 248*, 4532–4543.
25. Kistler, W. S., Noyes, C., Hsu, R., Heinrikson, R. L. (1975) *J. Biol. Chem. 250*, 1847–1853.
26. Heidaran, M. A., Kozak, C. A. & Kistler, W. S. (1989) *Gene 75*, 39–46.
27. Kleene, K. C., Borzorgzadeh, A., Flynn, J. F., Yelick, P. C. & Hecht, N. B. (1988) *Biochim. Biophys. Acta 950*, 215–220.
28. Luerssen, H., Maier, W. M., Hoyer-Fender, S. & Engel, W. (1989) *Nucleic Acids Res. 17*, 3585.
29. Kleene, K. C. & Flynn, J. F. (1987) *J. Biol. Chem. 262*, 17272–17277.
30. Luerssen, H., Hoyer-Fender, S. & Engel, W. (1988) *Nucleic Acids Res. 16*, 7723.
31. Kremling, H., Luerssen, H., Adham, I. M., Klemm, U., Tsaousidou, S. & Engel, W. (1989) *Differentiation 40*, 184–190.
32. Sundqvist, B., Kamensky, I., Hakansson, P., Kjellberg, J., Salehpour, M., Widdiyasekera, S., Fohlman, J., Petersson, P. A. & Roepstorff, P. (1984) *Biomed. Mass Spectrom. 11*, 242–257.
33. Meyer, H. E., Hoffmann-Posorske, E., Korte, H. & Heilmeyer, L. M. G. (1986) *FEBS Lett. 204*, 61–66.
34. Swiderek, K., Jaquet, K., Meyer, H. E. & Heilmeyer, L. M. G. (1988) *Eur. J. Biochem. 176*, 335–342.
35. DeLange, R. J., Fambrough, D. M., Smith, E. L. & Bonner, J. (1969) *J. Biol. Chem. 244*, 319–334.
36. Shlyapnikov, S. V., Arutyunyan, A. A., Kurochkin, S. N., Memelova, L. V., Nesterova, M. V., Sashchenko, L. P. & Severin, E. S. (1975) *FEBS Lett. 53*, 316–319.
37. Martinage, A., Mangeat, P., Laine, B., Couppez, M., Sautière, P., Marchis-Mouren, G. & Biserte, G. (1980) *FEBS Lett. 118*, 323–329.
38. Singh, J. & Rao, M. R. S. (1987) *J. Biol. Chem. 262*, 734–740.
39. Marushige, Y. & Marushige, K. (1975) *J. Biol. Chem. 250*, 39–45.
40. Balhorn, R. (1982) *J. Cell Biol. 93*, 298–305.
41. Marushige, Y. & Marushige, K. (1978) *Biochim. Biophys. Acta 518*, 440–449.
42. Pruslin, F. H., Imesh, E., Winston, R. & Rodman, T. C. (1987) *Gamete Res. 18*, 179–190.

Supplementary material to

Nuclear transition protein 1 from ram elongating spermatids

Mass spectrometric characterization, primary structure and phosphorylation sites of two variants

Frédéric CHIRAT, Arlette MARTINAGE, Gilbert BRIAND, Mostafa KOUACH, Alain Van DORSSELAER, Maurice LOIR and Pierre SAUTIÈRE

EXPERIMENTAL PROCEDURES

Materials

Testes of sexually mature rams (*Ovis aries*) were collected at the *Station de la Fertilité mâle de Nouzilly* (France) and were used immediately or after freezing in liquid nitrogen and storage at −80 °C. Ram spermatidal nuclear proteins were extracted and purified as described in [9].

Carboxypeptidase A treated with phenylmethylsulfonyl fluoride was from Sigma. Thermolysin was from Merck. Endoproteinase Lys-C was purchased from Boehringer. Acetonitrile for reverse-phase HPLC was obtained from Rathburn. All solvents and reagents for gas-phase sequencing were from Applied Biosystems. All other reagents were of the highest purity available.

Carboxypeptidase A digestion

Ram TP1 (1 nmol) was digested with carboxypeptidase A in 0.1 M N-methylmorpholine acetate buffer, pH 8.0, with an enzyme/substrate ratio of 1:50 (by mass) for 15 min at 37 °C. The digestion was stopped by lowering the pH with glacial acetic acid. The amino acids released were identified on the amino acid analyzer.

Cleavage with thermolysin

10 nmol ram TP1, dissolved in 0.4 ml 0.1 M ammonium bicarbonate buffer, pH 8.0, was digested with thermolysin at 37 °C for 1 h, using an enzyme/substrate ratio of 1:50 (by mass). The digestion was stopped by lowering the pH with formic acid and the hydrolysate was freeze-dried prior to fractionation.

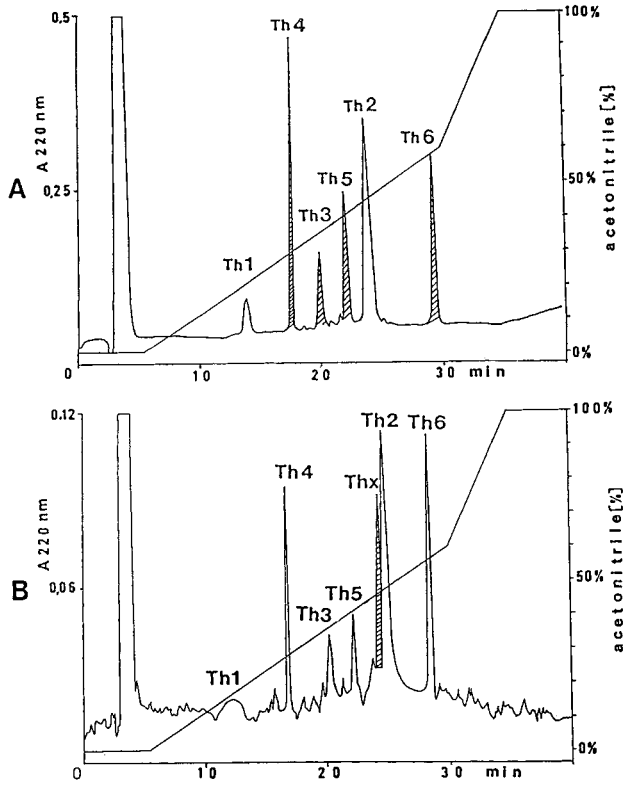

Fig. S1. *Fractionation by reverse-phase HPLC of peptides generated from thermolysin cleavage of the nonphosphorylated ram TP1 (A) and monophosphorylated ram TP1 (B).* The peptides (10 nmol) were separated on a C_{18} column, pore size 5 µm (Brownlee 220 × 4.6 mm). The elution was performed with a 0–60% linear gradient of 50% acetonitrile in 0.1% trifluoroacetic acid, at a flow rate of 1 ml/min. The absorbance was measured at 220 nm. The hatched peaks correspond to peptides submitted to automated Edman degradation. The peptides were numbered according to their position in the sequence of the protein

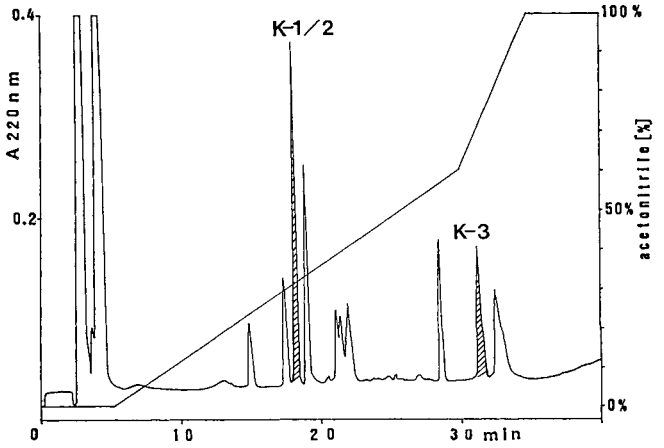

Fig. S2. *Fractionation by reverse-phase HPLC of peptides generated from endoproteinase Lys-C cleavage of the nonphosphorylated ram TP1.* The peptides (10 nmol) were separated and numbered as indicated in legend to Fig. S1. The hatched peaks correspond to peptides used to establish the complete amino acid sequence of ram TP1

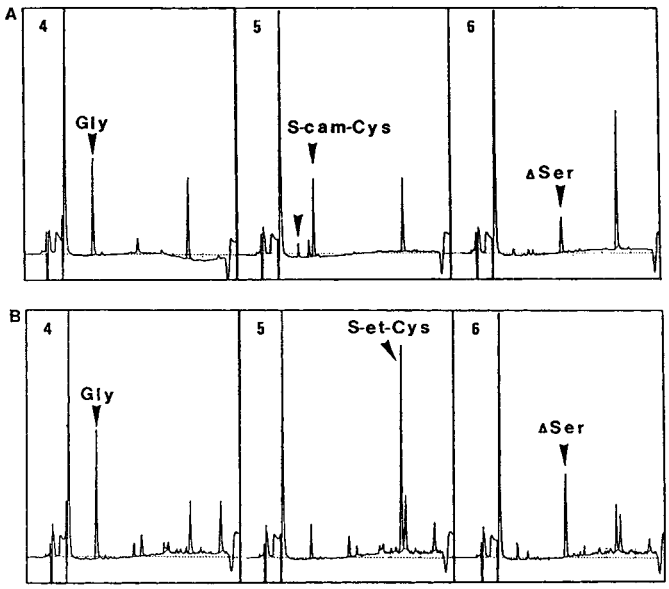

Fig. S3. *Automated Edman degradation of peptide Th3 derived from the cleavage of ram TP1 with thermolysin.* Only cycles 4, 5 and 6 are presented. Peptide Th3 (1 nmol) was sequenced (A) before and (B) after the treatment described by Meyer et al. [33] for the transformation of P-Ser in *S*-ethyl cysteine. At cycle 5, the phenylthiohydantoin derivative of *S*-carboxamidomethyl cysteine (*S*-cam-Cys) observed before treatment (A) was no longer observed after treatment (B), however, a phenylthiohydantoin derivative with a retention time identical to that of the phenylthiohydantoin of *S*-ethyl cysteine (*S*-et-Cys) was present in its place

Cleavage with endoproteinase Lys-C

10 nmol ram TP1 dissolved in 0.4 ml 0.1 M ammonium bicarbonate buffer, pH 8.0, was digested with endoproteinase Lys-C at 37°C for 2 h, using an enzyme/substrate ratio of 1:100 (by mass). The digestion was stopped by lowering the pH with formic acid and the hydrolysate was freeze-dried prior to fractionation.

Fractionation of peptides

Peptides generated by enzymatic cleavage were separated by reverse-phase HPLC on a C_{18} Brownlee column (220 × 4.6 mm) (particle size 5 µm) using the gradient systems indicated in the figure legends.

Amino acid analyses

Peptide samples were hydrolyzed in evacuated sealed tubes at 110°C for 24 h in 6 M HCl. One drop of 1% phenol was added to prevent excessive degradation of tyrosine.

Sequence determination

Microsequencing was performed on a gas-phase sequencer, Applied Biosystems 470A, using the 03 RPTH Standard Program. The phenylthiohydantoin derivatives of amino acids were identified on an on-line phenylthiohydantoin analyzer, Applied Biosystems 120A, equipped with an integrator-calculator, Shimadzu C-R4A. The amount of protein and of each peptide submitted to Edman degradation was approximately 1 nmol.

Mass spectrometry

Positive FAB-MS was carried out on a ZAB-HF double-focusing mass spectrometer (mass range 3200 Da at 8 keV) and recorded on

Table S1. *Amino acid sequence of S-carboxamidomethylated ram TP1 and of peptides derived from cleavage with thermolysin (Th) and endoproteinase Lys-C (K)*
Phenylthiohydantoin derivatives were identified and quantified by reverse-phase HPLC. Peptides K1 and K2 could not be separated and were sequenced together. n.q., not quantified

Cycle	TP1		Protein fragment of TP1													
			Th3		Th4		Th5		Th6		K1/2				K3	
	residue	amount	residue	amount	residue	amount	residue	amount	residue	amount	residue	amount	residue	amount	residue	amount
		pmol		pmol		pmol		pmol		pmol		pmol		pmol		pmol
1	S	339	V	295	Y	68	L	270	F	157	S	n.q.	Y	23	R	21
2	T	267	K	88	R	28	K	318	R	n.q.	S	n.q.	R	15	C	n.q.
3	S	217	R	30	K	22	S	66	S	83	L	13	K	12	D	46
4	R	83	G	116	S	n.q.	R	56	H	8	K	10			D	48
5	K	341	C	n.q.	S	n.q.	K	171	L	23					A	46
6	L	294	S	n.q.			R	77							N	40
7	K	344	K	33			C	76							R	21
8	S	108	R	29			D	45							N	30
9	Q	194	K	n.q.			D	61							F	38
10	G	175					A	92							R	20
11	T	59					N	92							S	n.q.
12	R	38					R	49							H	12
13	R	53					N	51							L	n.q.
14	G	125														
15	K	168														
16	N	112														
17	R	34														
18	T	39														
19	P	67														
20	H	20														
21	K	96														
22	G	62														
23	V	50														
24	K	97														
25	R	23														
26	G	58														
27	C	n.q.														
28	S	21														
29	K	58														
30	R	17														
31	K	59														
32	Y	18														
33	R	18														
34	K	47														
35	S	15														
36	S	n.q.														
37	L	9														
38	K	34														

Table S2. *Edman degradation of thermolysin peptides Th2 and Thx from monophosphorylated ram TP1, performed on the gas-phase sequencer* These peptides were derivatized according to Meyer et al. [33]. Phenylthiohydantoin derivatives were identified and quantified by reverse-phase HPLC. n.q. not quantified

Cycle	Peptide			
	Th2		Thx	
	residue	amount	residue	amount
		pmol		pmol
1	Leu	433.2	Tyr	215.3
2	Lys	297.9	Arg	85.3
3	Ser/*S*-ethyl Cys	n.q.	Lys	n.q.
4	Gln	151.4	Ser/*S*-ethyl Cys	n.q.
5	Gly	140.9	Ser/*S*-ethyl Cys	n.q.
6	Thr	29.3	Leu	44.0
7	Arg	86.4	Lys	130.6
8	Arg	135.6	Ser/*S*-ethyl Cys	n.q.
9	Gly	58.5	Arg	n.q.
10	Lys	n.q.	Lys	n.q.
11	Asn	46.0	Arg	50.0
12	Arg	n.q.	*S*-ethyl Cys[a]	n.q.
13	Thr	n.q.	Asp	24.4
14	Pro	23.4	Asp	25.3
15	His	11.8	Ala	10.8

[a] At cycle 12, a phenylthiohydantoin derivative with a retention time identical to that of phenylthiohydantoin of *S*-ethyl cysteine was observed in place of phenylthiohydantoin of *S*-carboxamidomethyl cysteine.

a VG 11/250 data system (VG Analytical Ltd, Manchester, UK). Ionisation of the sample was performed with 1 mA of a 8-keV energy Xe atom beam. The peptides were dissolved in desionized water at a concentration of 1 nmol/μl. The matrix was thioglycerol/TFA 1%. 1 μl matrix was deposited on a stainless steel target and 1 μl peptide solution was added.

The wide-scan spectra were generated by classical exponential magnetic scanning at a resolution of 1000 and at an accelerating voltage of 8 keV.

ESMS. ESMS was performed on a VG BIO-Q quadrupole mass spectrometer with a mass range of 3000 Da. The mass spectrometer was scanned from $m/z = 100$ – $m/z = 1000$ in 15 s at unit resolution.

Calibration was performed using charged ions from poly(ethyleneglycol) 800 obtained from a separate introduction.

The electrostatic spray ion source was operated at atmospheric pressure with an extraction cone voltage (V_c) value of 150 volts. The protein sample was first dissolved in water containing 5% acetic acid then an equivalent volume of methanol was added. The concentration used was between 30–100 pmol/μl. This solution (2–10 μl) was introduced into the ion source at a flow rate of 2 μl/min.

PDMS. The ^{252}Cf PDMS of peptide Th5 was obtained on a BIO-ION BIN 10 K Instrument (BIO-ION AB, Box 15045, Uppsala, Sweden). The spectrum was acquired for 10^6 fission events (15–20 min) or for 4×10^6 events). Data-handling procedures have been described by Sundqvist et al. [32]. The sample was dissolved (2.5 nmol/μl) in a water/methanol solution (4:1, by vol.) containing 0.1% TFA. The sample (2 μl) was deposited on the matrix and left to dry. Finally, the dry target was inserted into the mass spectrometer.

Nomenclature of peptides

Peptides obtained from cleavage with thermolysin and with endoproteinase Lys-C were designated as Th and K, respectively. Peptides were numbered according to their position in the amino acid sequence of the protein.

Conversion of phosphoserine to S-ethyl cysteine

The phosphoserine-containing thermolysin peptides were derivatized according to Meyer et al. [33] modified by Swiderek et al. [34]. 1 nmol peptide placed in a small glass tube was dissolved in 25 μl derivatization mixture consisting of 0.4 ml water, 0.1 ml absolute ethanol, 65 μl 5 M sodium hydroxyde and 60 μl ethanethiol.

The tube was flushed with argon, tightly closed and incubated at 50 °C for 1 h. After incubation the solution was evaporated to dryness. The derivatized peptide was then dissolved in 80 μl water and applied directly on a polybrene-treated glass-fibre disk for sequence determination.

The use of a small volume of derivatization mixture avoids clogging of the glass fibre disk by salts.