Post-translational Cleavage of a Histone H1-like Protein in the Sperm of Mytilus*

Steve Carlos‡, Donald F. Hunt‡, Corinne Rocchini§, David P. Arnott§, and Juan Ausio‡¶

From the ‡Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8W 3P6, Canada and the §Department of Chemistry, University of Virginia, Charlottesville, Virginia 22901

Starting with total RNA from spermatogenic cells of Mytilus trossulus and using random priming, we have cloned and sequenced the c-DNAs corresponding to two variants of the sperm-specific protein PLII* (φ2B). DNA sequencing in conjunction with mass spectrometry and protein sequence data have allowed us to establish that of the three sperm-specific proteins present in the sperm of Mytilus (PL-II* (φ2B), PL-III (φ1), PL-IV (φ3)), the first and the last one are the result of post-translational cleavage of a common precursor. This common precursor is a member of the histone H1 family, and it exhibits inter- and intraspecific microheterogeneity.

The nuclei of the sperm of Mytilus contain three major sperm-specific proteins: PL-II* (φ2B), PL-III (φ1), and PL-IV (φ3) (Subirana et al., 1973; Ausio and Subirana, 1982a; Ausio, 1986). These proteins coexist with approximately 10–15% of the somatic-like histones in the mature sperm (Ausio, 1986).

For many years, the relation existing among these proteins as well as their individual role in chromatin condensation during spermatogenesis has remained a mystery. In an attempt to elucidate this relationship, Bloch (1966) conducted a series of elegant experiments designed to follow the sequential appearance of these proteins during the differentiation process of spermatogenesis. In these experiments C14 radiolabeled amino acids were injected into the gonadal tissues, and their incorporation into the different sperm-specific proteins was monitored at different time intervals. A sequential labeling of proteins was observed in the order PL-II*, PL-IV, and finally PL-III. It was concluded that the most obvious explanation was that of a precursor-product relationship in which Protein PL-II* was converted into PL-IV, and this in turn converted to PL-III. However, the mechanism involved in the conversion could not be ascertained.

The possible relationship between PL-II* and PL-IV is intriguing, especially if one considers that both proteins have been linked to the family of H1 histones. Protein PL-IV has been related to the COOH region of histone H1 (Phelan et al., 1974). Recently, protein PL-II* has also been identified as a sperm-specific histone H1-like protein (Jutglar et al., 1991).

In the present paper we confirm the existence of a precursor-product relationship between these two proteins.

MATERIALS AND METHODS

Living Organisms—Specimens of Mytilus trossulus and Mytilus californianus were collected at Esquimalt Lagoon and Point No Point (Sooke) on Vancouver Island.

Preparation of PolyA+ RNA—PolyA+ RNA was isolated from immature male gonads that had been previously frozen with liquid nitrogen, following the method described by Ausubel et al. (1987).

cDNA was prepared from a mixture of poly(A)+/poly(A)- mRNA by random priming (Koike et al., 1987) using a Rib0 Clone cDNA system from Promega. The cDNA fragments thus obtained were cloned into a λ-ZapII vector (Stratagene) following the method suggested by the manufacturer. Immunoscreening of the cDNA library was carried out with the aid of a pico-Blue immunoscreening kit from Stratagene following the method described by Sambrook et al. (1989), and using polyclonal antibodies elicited against protein PL-II* (see accompanying paper by Carlos et al. (1993)).

DNA sequencing was carried out by the dideoxynucleotide method (Sanger et al., 1977) using a Sequenase version 2.0 kit from Sequenase.

In Vitro Translation—The poly(A)+ and poly(A)- mRNA fractions were expressed in a wheat-germ cell-free system purchased from Promega. Translation was carried out according to the recommendations of the manufacturer. Fractionation and Purification of Poly(A)+ mRNA—Fractionation of total RNA was carried out with oligo(dT) cellulose using a mRNA purification kit from Pharmacia LKB Biotechnology Inc.

HPLC Fractionation—Reversed-phase HPLC1 was carried out as described in the accompanying paper (Carlos et al., 1993).

1 Gel Electrophoresis—Acetic acid-urea-Triton X-100 polyacrylamids were prepared essentially as described by Bonner et al. (1980) with some minor modifications. The concentration of acrylamide in the stacking gel was 5% (acrylamide, N,N'-methylenebisacrylamide: 20:1). The stacking gel did not contain Triton X-100. The separating gel was 15% acrylamide (acrylamide, N,N'-methylenebisacrylamide: 30:1) and contained 6 mM Triton X-100. In addition, both gels contained 5% acetic acid, 3.8 mM urea, 0.135 mM NH4OH, 0.5% TEMED, and riboflavin (0.0004% stacking; 0.0025% separating). Polymerization was carried out under UV light.

Mass Spectrometry—Electrospray ionization to determine molecular masses was carried out as described by Hunt et al. (1991).

Amino Acid Analysis—Amino acid analysis was carried out on an ABI Model 420A derivatizer-analyzer system as described in the accompanying paper (Carlos et al., 1993).

Enzymatic Digestions—The main protein component of PL-IV from M. trossulus was digested using endoproteinase Arg-C from Boehringer under the same conditions described previously (Ausio and McParland, 1989). The enzyme-to-sulfate ratio used was 1:100.

1 The abbreviations used are: HPLC, high-performance liquid chromatography; PL, protamine-like proteins; TEMED, N,N,N',N'-tetramethylethylenediamine.
Immediately after digestion the sample was loaded onto an HPLC column without any further treatment.

Amino Acid Sequencing Techniques—Peptide sequencing was performed as described in the accompanying paper (Carlos et al., 1993).

RESULTS AND DISCUSSION

The Protein Precursor of PL-IP—We have isolated the total RNA from a cell suspension of immature gonadal tissue from *M. trossulus*. The starting cell mixture consisted of cells at different stages of spermatogenesis, mainly secondary spermatocytes, spermatids and some fully differentiated sperm cells. We then prepared cDNA from the whole RNA mixture using random priming on the assumption that large amounts of a specific mRNA for the nuclear sperm proteins must be present in these cells. After cloning the cDNA fragments thus obtained into a λ-ZapII vector, we were able to isolate two positive clones by immunoscreening with polyclonal antibodies elicited against protein PL-IP*. Fig. 1 shows the partial nucleotide sequence of the coding region of these two clones, as well as the respective amino acid sequence predicted by the open reading frame. These coding regions were identified with the help of the protein sequence obtained by Edman degradation of the main component of protein PL-IP* from *M. californianus* (see Carlos et al. (1993), accompanying paper).

As can be seen in Fig. 1, the two positive clones isolated from *M. trossulus* consist of two protein variants. This can be ascertained from the microheterogeneity observed at the amino acid level and to a larger extent at the nucleotide level. Twenty-two differences can be found at the codon level within the region coding for PL-IP* (see Fig. 1). In terms of the number of substitutions per number of nucleotides present in the coding region, the heterogeneity observed at the nucleotide level is similar to that of the protamine variants of trout (Dixon et al., 1986). At the protein level the extent of microheterogeneity found between the two variants of *M. trossulus* is similar to the microheterogeneity that exists between each of them and the major protein component of PL-IP* from *M. californianus* (see Fig. 2). Indeed, inter- and intraspecific protein microheterogeneity of the PL proteins of the sperm of *Mytilus* had been previously documented (Mogensen et al., 1983).

However, the most surprising result did not come from the microheterogeneity observed for PL-IP* but from the fact that the coding region extended far beyond the sequence established for this protein using Edman degradation (Carlos et al. (1993), the accompanying paper).

Protein PL-IP Is the Result of Post-translational Cleavage—

The first hint of the possible nature of the puzzle described at the end of the preceding section was obtained from comparison of the amino acid sequence predicted by the nucleotide sequence of the prolonged region and the amino acid sequence of PL-IV from *M. californianus* (Ausio and McParland, 1989). As it can be seen in Fig. 2, the extent of similarity is remarkable.

However, in trying to identify the two sequences with each other, a major problem was encountered in that the amino acid sequence predicted for the region extending beyond the COOH-terminal region of PL-IP* was larger than the amino acid sequence established for PL-IV (Ausio and McParland, 1989). If the extended protein region of PL-IP* were related to PL-IV, one would expect them to have a very similar molecular mass, especially since all variants of PL-IV from *M. trossulus* and *M. californianus* exhibit very similar electrophoretic mobility (see Fig. 3). This prompted us to re-evaluate the molecular mass of PL-IV and look into the possibility that the sequence previously published for this protein was not complete. In order to measure the molecular mass of PL-
The discrepancy observed between the molecular sequence of the fraction component of the accompanying paper and that predicted from its sequence (Ausio and McParland, 1989) is that the sequences for the three variants of protein P1-II* from M. californianus represent partial rather than complete sequences. The reason for this is the improper assignment of the COOH-terminal end of the molecule. Because of the repetitive nature and high lysine content of this region carboxypeptidase analysis does not provide a clear cut answer, and it is then necessary to rely on indirect methods such as the determination of the molecular mass to determine the end of the primary structure of the protein.

That the sequence extending beyond NKSNN of P1-II* corresponds to P1-IV is further corroborated by the amino acid sequence deduced from the cDNA sequence of the major protein fraction of P1-IV from M. trossulus (see Figs. 1 and 5). The sequence thus established has a molecular mass of 6451 Da which is in excellent agreement with the molecular mass determined by mass spectrometry. This sequence has great similarity to the amino acid sequence of the 3'-coding sequence also of the cDNA clone shown in Figs. 1 and 2. Only two differences are observed, of which, one of them (R → K in position 8 from the NH2 terminus) is a conserved replacement. Furthermore, the partial sequence determined from an incomplete cDNA clone (see M.t. variant 2 in Fig. 2), also has K in position 8 as in the case of the sequence determined by Edman degradation (Fig. 2, M.t. Edman). Thus, M.t. variant 1 shown in Fig. 2 of molecular mass 6510 Da, is most likely one of the minor components of P1-IV. It probably corresponds to the one with an assigned molecular mass of 6493 Da in Fig. 4B.

It is therefore apparent from all these results that the P1-II* and P1-IV proteins from the sperm of Mytilus are the result of post-translational cleavage of a common protein precursor. The sequence NKSNN found at the COOH terminus of P1-II* of M. trossulus is unusual for either a histone or a protamine. Thus this sequence may operate as the signal being recognized by the processing enzyme.

**Poly(A')/Poly(A') Nature of the P1-II*-P1-IV Precursor**—As has been mentioned at the beginning, cDNA was initially prepared from unfractionated RNA consisting of a mixture of poly(A')/poly(A') mRNA. The only information available in that regard on bivalve mollusks is that of an early characterization of the mRNA coding for the sperm-specific proteins of a member of the family Mytilidae (Zalensky and Chelomina, 1986). Using Cremonomytilus grayanus, which has a nuclear protein composition almost identical to that found in the genus Mytilus, Zalensky and Chelomina (1986) concluded that the mRNA of the PL proteins does not contain sequences of poly(A).
tein presence of an electrophoretic band with lower mobility than the sperm of M. trossulus. However, this result is not surprising considering preliminary results it looks likely that both forms of mRNA with endoproteinase Arg-C.

Early study, Bloch was already able to electrophoretically amino acids containing tritiated lysine. Further. To that purpose we fractionated the total RNA into a poly(A+) and a poly(A-) fraction using oligo(dT)-cellulose. Poly(A-) forms have also been observed in rainbow trout and the relatedness of these proteins to protamines. Poly(A+) and PL-11* that was present in both the poly(A+) and the poly(A-) used at that time. In these kinds of gels both protein β and its precursor would have run very close together. The β (PL-II*) precursor would then be processed almost immediately into the β (PL-II*) and δ (PL-IV) proteins that appeared to be labeled 12 days later.

The origin of PL-III (γ) which is the main PL protein of the nuclei in the mature sperm of Mytilus (Ausio and Subirana, 1982b) remains as yet unknown and may be of a completely different nature. The fact that PL-II* and PL-III share a common NH2-terminal domain (see Carlos et al., 1993; accompanying paper) would suggest that the relationship between these two proteins possibly occurs at the level of transcription perhaps through the sharing of some common exon sequence(s).

CONCLUSION

In addition to what has been discussed in the preceding section, our results provide further support for the microheterogeneous nature of the PL proteins of Mytilus (Mogensen et al., 1991). The extent of microheterogeneity observed at the amino acid and at the nucleotide levels, are very similar to those found in the trout protamines (Dixon, 1986). Microheterogeneity of chromatin-associated proteins has been well documented in the case of histone H1 in somatic tissues (Cole, 1987) and in the case of protamines in the male germinal tissues (Subirana, 1983).

The fact that the protamine-like proteins (PL) of the bivalve mollusks also exhibit microheterogeneity is not surprising and adds further support to the suggestion of the existence of a close evolutionary relationship between histone H1 and protamines (Ausio, 1992). The excision of the COOH-terminal domain of a histone H1-like PL-I (PL-II* + PL-IV) precursor could be viewed as one of the many different mechanisms involved in this evolutionary process.

Acknowledgment.—We are very thankful to Cheryl Gonnason for careful typing of the manuscript.

REFERENCES


We therefore decided to investigate this point a little bit further. To that purpose we fractionated the total RNA into a poly(A*) and a poly(A-) fraction using oligo(dT)-cellulose columns. The fractions thus obtained were expressed in a cell-free wheat germ translation system using a mixture of amino acids containing tritiated lysine.

Autoradiographic analysis (results not shown) showed the presence of an electrophoretic band with lower mobility than PL-II* that was present in both the poly(A*) and the poly(A-) fractions. No other bands could be detected. From these preliminary results it looks likely that both forms of mRNA are present in the case of the protamine-like PL-II*-PL-IV precursor. However, this result is not surprising considering the relatedness of these proteins to protamines. Poly(A*) and poly(A-) forms have also been observed in rainbow trout protamines (Iatrou and Dixon, 1977).

Confirming Bloch's Hypothesis—The results presented in this paper confirm the conclusions put forward by David Bloch (1966) in that there is a precursor-product relationship among the sperm-specific nuclear proteins of Mytilus. In this early study, Bloch was already able to electrophoretically distinguish four proteins which he called δ (PL-IV), γ (PL-III), β (PL-II*), and a diffuse band, α, which most likely corresponds to the remnant histone fraction which is also present in the mature sperm (Ausio and McParland, 1988). Upon injection of the gonadal tissues with 14C-labeled amino acids, it was found that after 1 day protein β was already labeled. In the next sample analyzed 12 days after injection, both β and δ were found to be labeled, but no trace of labeled γ could be detected. Only several weeks after injection, had protein γ (the major nuclear protein of the mature sperm, PL-III) incorporated some of the labeled amino acids. To explain this sequential labeling Bloch proposed several interpretations, the most obvious being that of a precursor-product relationship in which β (PL-II*) was converted into δ (PL-IV) in which turn would be converted into γ (PL-III) in the mature sperm. It was suggested that the phenomenon involved either a complete break down with the products of hydrolysis being utilized for de novo synthesis or the conversion of large molecules (Bloch, 1966).

Our data fully support the precursor-derived origin of PL-II* (β) and PL-IV (δ). This larger precursor could tentatively be identified with the electrophoretic band that in Bloch's experiment appeared to be labeled immediately upon injection of the radiolabeled amino acids. Although in that experiment this band was identified with β (PL-II*), we think that was due to the poor resolution achieved with the starch gels used at that time. In these kinds of gels both protein β and its precursor would have run very close together. The β (PL-II*) precursor would then be processed almost immediately into the β (PL-II*) and δ (PL-IV) proteins that appeared to be labeled 12 days later.

The origin of PL-III (γ) which is the main PL protein of the nuclei in the mature sperm of Mytilus (Ausio and Subirana, 1982b) remains as yet unknown and may be of a completely different nature. The fact that PL-II* and PL-III share a common NH2-terminal domain (see Carlos et al., 1993; accompanying paper) would suggest that the relationship between these two proteins possibly occurs at the level of transcription perhaps through the sharing of some common exon sequence(s).

Fig. 4. A, electrospray ionization spectrum of unfractionated protein PL-IV from the sperm of M. californianus. B, matrix-assisted desorption mass spectrometry of unfractionated protein PL-IV from the sperm of M. trossulus.

Fig. 5. Amino acid sequence obtained by Edman degradation of the main PL-IV protein component of M. trossulus. WP, whole protein; R-E, peptides obtained by enzymatic digestion with endoprotease Arg-C.

We therefore decided to investigate this point a little bit further. To that purpose we fractionated the total RNA into a poly(A*) and a poly(A-) fraction using oligo(dT)-cellulose columns. The fractions thus obtained were expressed in a cell-free wheat germ translation system using a mixture of amino acids containing tritiated lysine.

Autoradiographic analysis (results not shown) showed the presence of an electrophoretic band with lower mobility than PL-II* that was present in both the poly(A*) and the poly(A-) fractions. No other bands could be detected. From these preliminary results it looks likely that both forms of mRNA are present in the case of the protamine-like PL-II*-PL-IV precursor. However, this result is not surprising considering the relatedness of these proteins to protamines. Poly(A*) and poly(A-) forms have also been observed in rainbow trout protamines (Iatrou and Dixon, 1977).

Confirming Bloch's Hypothesis—The results presented in this paper confirm the conclusions put forward by David Bloch (1966) in that there is a precursor-product relationship among the sperm-specific nuclear proteins of Mytilus. In this early study, Bloch was already able to electrophoretically distinguish four proteins which he called δ (PL-IV), γ (PL-III), β (PL-II*), and a diffuse band, α, which most likely corresponds to the remnant histone fraction which is also present in the mature sperm (Ausio and McParland, 1988). Upon injection of the gonadal tissues with 14C-labeled amino acids, it was found that after 1 day protein β was already labeled. In the next sample analyzed 12 days after injection, both β and δ were found to be labeled, but no trace of labeled γ could be detected. Only several weeks after injection, had protein γ (the major nuclear protein of the mature sperm, PL-III) incorporated some of the labeled amino acids. To explain this sequential labeling Bloch proposed several interpretations, the most obvious being that of a precursor-product relationship in which β (PL-II*) was converted into δ (PL-IV) in which turn would be converted into γ (PL-III) in the mature sperm. It was suggested that the phenomenon involved either a complete break down with the products of hydrolysis being utilized for de novo synthesis or the conversion of large molecules (Bloch, 1966).

Our data fully support the precursor-derived origin of PL-II* (β) and PL-IV (δ). This larger precursor could tentatively be identified with the electrophoretic band that in Bloch's experiment appeared to be labeled immediately upon injection of the radiolabeled amino acids. Although in that experiment this band was identified with β (PL-II*), we think that was due to the poor resolution achieved with the starch gels used at that time. In these kinds of gels both protein β and its precursor would have run very close together. The β (PL-II*) precursor would then be processed almost immediately into the β (PL-II*) and δ (PL-IV) proteins that appeared to be labeled 12 days later.

The origin of PL-III (γ) which is the main PL protein of the nuclei in the mature sperm of Mytilus (Ausio and Subirana, 1982b) remains as yet unknown and may be of a completely different nature. The fact that PL-II* and PL-III share a common NH2-terminal domain (see Carlos et al., 1993; accompanying paper) would suggest that the relationship between these two proteins possibly occurs at the level of transcription perhaps through the sharing of some common exon sequence(s).
Post-translational Cleavage of a Histone H1-like Protein

Bloom, D. P. (1966) Protoplasmatologia 6, 1-36