Protamines of Reptiles*

Janice G. Hunt‡, Harold E. Kasinsky§, Ruth M. Elsey¶, Cynthia L. Wright†‡, Philip Rice‡, Jason E. Bell‡, Daniel J. Sharp‡, Andor J. Kiss‡, Donald F. Hunt**, David P. Arnott**, Mark M. Russ**, Jeffrey Shabanowitz**, and Juan Ausió‡ ‡

From the ‡Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8W 3P6, Canada, the §Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada, the †Department of Wildlife and Fisheries, Rockefeller Wildlife Refuge, Fur and Refuge Division, Grand Chenier, Louisiana 70643, and the **Department of Chemistry, University of Virginia, Charlottesville, Virginia 22901

We have characterized for the first time the complete primary structure of the main protamine components of the sperm from four reptiles: Chrysemys picta (turtle), Elaphe obsoleta (snake), Anolis carolinensis (lizard), and Alligator mississippiensis (crocodilian). These species were chosen to represent one of each of the main phylogenetic branches of this taxonomic group. Comparison of these protamine sequences with those already available from other vertebrate groups allows us to define properly the chemical consensus composition of protamines and provides a unique insight into their molecular evolution and classification.

The word protamine was coined by Friederich Miescher (Miescher, 1874) to account for the nitrogenous base that he found complexed through salt-like bridges to nuclein (now DNA) from the sperm nucleus of Rhine salmon. It was not, however, until 20 years later that the true protein nature of protamines was established by Kossel working with sperm nuclei from sturgeon (Kossel, 1896a, 1896b). The discovery of protamines in fish was almost coincidental. Miescher’s interest in the chemical composition of the cell nucleus led him to search for cells consisting mainly of nuclei. Such is the case of the sperm he obtained from salmon, which presumably was easily available at the time. Because of this historical component and also because of the easy availability of the fish system, fish protamines (particularly in salmonids) have been the most extensively characterized of all protamines to date (Dixon et al., 1985). Unfortunately, all this has inadvertently led to the use of the misleading term “typical protamines” to refer to this particular group of protamines. Indeed, protamines are widespread in other vertebrate groups (Oliva and Dixon, 1991), and they have also been found in the sperm of different groups of invertebrates (Bloch, 1969, 1976; Subirana, 1983; Kasinsky, 1989; Ausió, 1995). Furthermore, the primary structures of some of the invertebrate protamines strongly resemble those of the vertebrate counterparts (Daban et al., 1995).

In the case of the vertebrate protamines (see Oliva and Dixon (1991) for a comprehensive review), the mammalian protamines were the next to be widely characterized after the fish protamines. The use of polymerase chain reaction techniques has led to an outburst in the amount of sequence information within this group of vertebrates (Oliva, 1995). In recent years, a whole bounty of information has also been obtained regarding the primary structure of the protamines from monotremes and marsupials (Winkfein et al., 1993; Retief et al., 1993, 1995a, 1995b). In contrast to fish and mammals, there is very little information regarding the primary structure of the protamines from birds (Nakano et al., 1976; Oliva and Dixon, 1989; Oliva et al., 1990) and amphibians (Takamune et al., 1991; Hiyoshi et al., 1991; Ariyoshi et al., 1994; Mita et al., 1995). In the case of reptiles, the only information available to date regarding these proteins was that from PAGE1 analysis and several amino acid compositions (Kasinsky et al., 1987; Chiva et al., 1989).

In this paper, we characterize the primary structure of the protamines from several representative species corresponding to different groups of reptiles. The major scope of this work has been to fill in the gap that presently exists regarding the primary structure of vertebrate protamines. All of this is directed toward gaining a better understanding of the chemical identity and constraints of protamines and their changes during the course of vertebrate evolution.

MATERIALS AND METHODS

Living Organisms—Crocodilians (Alligator mississippiensis, family Alligatoridae) were caught in the wild. Testis epididymis/ductus deferens were collected by Louisiana Department of Wildlife and Fisheries biologists at the Rockefeller Wildlife Refuge from nuisance/research alligators. The testes were kept in 90% ethanol (Lee et al., 1991) at 4 °C, shipped to Canada by air with an appropriate CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora Export) permit, and processed for protein extraction in British Columbia, Canada. Turtles (Chrysemys picta (painted turtle) and Pseudemys scripta (pond slider), both family Emydidae), snakes (Elaphe obsoleta quadrivittata (black rat snake), family Colubridae), and lizards (Anolis carolinensis (green anole) and Sceloporus magister (spiny lizard), both family Iguanidae) either were caught in the wild in the southeastern United States or were purchased from commercial dealers throughout North America. Testis and vas deferens were collected from these organisms and either were kept in ethanol as described above or were immediately frozen and stored at −70 °C until further processing. Another snake, Thamnophis scalaris (garter snake) (also family Colubridae), was captured in the wild near San Cristobal de las Casas, Chiape, Mexico. Classifications are based on Zug (1993).

Extraction and Isolation of SNBP—Chromosomal sperm proteins were extracted and isolated as described (Ausió, 1986) with a few minor modifications. Octanol (0.2%, v/v) was added to the buffer used in the homogenization of the starting tissue (testis/vas deferens) to prevent

* This work was supported in part by Natural Sciences and Engineering Research Council of Canada Grant OGP 0046399 (to J. A.), United States Public Health Service Grant GM-37537 (to D. F. H.), and Grant 585854 (1970–1995) (to H. E. K.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported by the Department of Zoology, University of British Columbia and Kasuray Educational.

Page 23547

The Journal of Biological Chemistry
© 1996 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 271, No. 38, Issue of September 20, pp. 23547–23557, 1996
Printed in U.S.A.

¶ To whom correspondence should be addressed. Tel.: 604-721-8863; Fax: 604-721-8855.

†† To whom correspondence should be addressed. Tel.: 604-721-8863; Fax: 604-721-8855.

†‡ To whom correspondence should be addressed. Tel.: 604-721-8863; Fax: 604-721-8855.

23547
foaming. Also, in addition to 0.2 mM phenylmethylsulfonyl fluoride, all solutions contained N\textsuperscript{\textalpha}-p-tosyl-1-lysine chloromethyl ketone (20 μg/ml).

**Ion Exchange Chromatography**—Ion exchange chromatography was performed on 18 × 1.5-cm (analytical) or 40 × 2.5-cm (preparative) columns using Bio-Rex 70 resin (Bio-Rad). The proteins were eluted using a guanidinium chloride gradient in 0.1 M phosphate buffer, pH 6.8, as described by Cole (1989).

**HPLC Fractionation**—HPLC was routinely carried out on 5-μm Vydac C\textsubscript{4} or C\textsubscript{8} columns (25 × 0.46 cm (analytical) or 25 × 1 cm (preparative)) with 0.1% trifluoroacetic acid as eluant with varying acetonitrile gradients (Ausió, 1988).

**Urea (2.5 M)/acetic acid (5%, v/v)-PAGE** was used as described (Jutglar et al., 1991). The peptides produced by proteolytic cleavage of protamines were analyzed electrophoretically on 30% polyacrylamide gels as described (Ausió and McParland, 1989).

**Enzymatic Digestion of SNBPs**—Endoproteinase Glu-C (protease V8; EC 3.4.21.19; Boehringer Mannheim) digestion of the turtle protamine fractions was carried out at a 1:20 (w/w) E/S ratio for 60 min at 37°C. Chymotrypsin (EC 3.4.21.1; Boehringer Mannheim) digestion of the alligator protamine fractions was carried out at a 1:50 (w/w) E/S ratio for 60 min at room temperature. *Astacus fluviatilis* protease (EC 3.4.99.6; Serva) digestion of alligator and lizard protamine samples was performed at a 1:50 (w/w) E/S ratio for 30 min at room temperature. The protein concentration in all instances was ~1 mg/ml, and the buffer was 0.1 M ammonium bicarbonate, pH 8.0. Immediately after incubation for the appropriate digestion time, as established from a time course experiment, the samples were injected directly onto a reverse-phase HPLC C\textsubscript{4} column and eluted with different acetonitrile gradients in 0.1% trifluoroacetic acid.

**Alkaline Phosphatase Treatment**—Turtle (*C. picta*) protamine fractions were dissolved to a final concentration of ~0.2 mg/ml in 100 mM Tris-HCl, pH 8.0, 0.2 mM MgCl\textsubscript{2}, 0.2 mM ZnCl\textsubscript{2}. Next, calf intestinal alkaline phosphatase (EC 3.1.3.1; New England Biolabs Inc.) was added to an E/S ratio of ~300 units/1 mg, and the solution was incubated for 1 h at 37°C. After incubation, the sample was exhaustively dialyzed against double-distilled water and lyophilized. A protein sample treated exactly under the same conditions but in the absence of alkaline phosphatase was used as a control.

**Amino Acid Analysis and Protein Sequencing**—Amino acid analysis and protein sequencing were performed as described (Jutglar et al., 1991).

**Sequence Analysis**—Protein alignments and similarities were determined using the CLUSTAL V multiple sequence alignment program (Higgins and Sharp, 1989) and the BLITZ server at EMBL, which uses the best local similarity algorithm (Smith and Waterman, 1981). A phylogenetic tree was generated using the Phylogeny Inference package (PHYLIB) (Felsenstein, 1989).

**Mass Spectrometry**—Mass spectrometry analysis was carried out according to Hunt et al. (1991) as described (Carlos et al., 1993). The monoisotopic and average masses (MH\textsuperscript{+}) from the protein sequences were calculated using Mac ProMass Program Version 1.05 (prepared by Dr. Terry Lee, Beckman Instruments).

**RESULTS**

**Protamine Fractionation and Purification**—The major aim of this work was the characterization of the protamines from organisms belonging to the main phylogenetic branches of reptiles (Fig. 1A). The protamine composition of several representative organisms of each such group is shown in Fig. 1B. One of the major experimental efforts in the protamine characterization was the fractionation and purification of the protein samples. The problem was compounded by the small size of the sperm samples, usually obtained from vas deferens, and also by the restricted access and seasonal availability of these organs, especially in those animals (such as alligators and snakes) that had to be collected from the wild. Also, it was not always possible to obtain organs from individuals at the same stage of gonadal maturity, and thus, the protein yields varied enormously between different individuals. On top of all this, several of the protamines from the representative reptiles studied here exhibit a large extent of protein microheterogeneity, such as in the case of the turtle *C. picta* and the crocodilian *A. mississippiensis* (Fig. 2). Although we observed that the microheterogeneity patterns were extremely conserved in all the individuals studied, the relative amounts of some of the different fractions varied slightly between individuals and also seasonally (data not shown).

Because of all this, the need to purify the fractions of interest to a high extent was deemed crucial for their sequencing analysis. The monotonous amino acid composition of some of these proteins with only a few different amino acid residues present (for example, only four in the lizard *A. carolinensis*) made it exceedingly difficult in these cases to use the conventional sequencing approach that consists of fractionating the protein into small pieces with proteases, followed by reconstitution of...
the sequence from the partial sequences of the fragments. In these instances, the proteins were thoroughly purified through multiple rounds of HPLC fractionation until extremely homogeneous fractions were obtained, as monitored by mass spectrometry. Under these circumstances, it was found possible to sequence the whole protein from beginning to end in only one sequencing run without any ambiguity, especially if the starting amount of sample analyzed was \( \geq 2 \) nmol.

**Major Protamine Components of the Sperm from the Turtle C. picta**—Fig. 2 (A, trace \( \text{Tt} \); and B, panel \( \text{Ti} \)) shows the extent of protein microheterogeneity of the protamines from the turtle \( \text{C. picta} \). Despite the resolution observed in the chromatogram shown in Fig. 2A (trace \( \text{Tt} \)) and the apparent electrophoretic purity of the fractions from the corresponding peaks (Fig. 2B, panel \( \text{Ti} \)), mass spectrometry analysis indicates that each of the peaks consists of a mixture of different protein fractions (data not shown). Part of this complexity arises from the fact that some of the protamine fractions present in the different HPLC peaks shown in Fig. 2A (trace \( \text{Tt} \)) are phosphorylated (Fig. 3B). Thus, the proteins from each of the peaks recovered from the HPLC elution have to be treated with alkaline phosphatase and refractionated by HPLC (data not shown).

We noticed that although the microheterogeneity pattern (as envisaged in Fig. 3A) remained fairly constant from year to year (between different batches of turtles), the extent of phosphorylation of each of the individual fractions was very variable (data not shown). This was due presumably to differences in gonadal maturity of the different batches. To sequence the different protamine fractions, we took advantage of a unique glutamic residue in their amino acid composition. This allowed us to split the molecule into two well defined (highly homogeneous) peptides (Fig. 3C) consisting of the N- and C-terminal domains. The peptides thus obtained were fractionated by HPLC (Fig. 3C, panel 2). After determination of their respective masses (Table I) using mass spectrometry, their sequences were determined by conventional automated Edman degradation microsequencing.

The sequences of the different protamine fractions analyzed are shown in Fig. 4. The sequences of the protamine fractions from the fastest moving band (band II) on the gels (Fig. 3B) are extremely similar, with only one to three amino acid variations among all of them, thus revealing the true microheterogeneity nature of their variability. The protamine fractions running with the fast electrophoretic mobility band can be grouped into two major groups (Tt-II-1–3 and Tt-II-4 and -5), which basically differ in the hydrophobic nature of their last C-terminal amino acids. This confers a difference in the overall hydrophobicity of the molecules, which results in the high extent of resolution observed by HPLC (Fig. 3A).

We also sequenced one of the protamine components from the band of lower electrophoretic mobility (group I) (Fig. 3B). This band is only present in very low amounts (Fig. 3A), and it is in itself microheterogeneous. In fact, although we refer to the protamine fractions of this group as Tt-I-1 and Tt-I-2, each of these HPLC peaks consists of more than one fraction. The sequence of one such fraction (Tt-I-1) is shown in Fig. 4. The amino acid residues in parentheses indicate the sites of microheterogeneity. As can be seen in Fig. 4, the sequences of the protamine fractions from electrophoretic groups I and II are extremely alike except for the presence of an additional 13 amino acids in the N-terminal region of the fractions of group I (see also Fig. 9). The additional sequence is ARYRNR(S)\(_3\) in the case of the protamine fractions from Tt-I-1 and Tt-I-2, which is similar to the ARYR-\( -(BS)\)_3 sequence (\(B = \) basic amino acid), which is commonly found in the protamines from the sperm of birds and mammals (Oliva and Dixon, 1991; Oliva, 1995).
Fig. 3. A, schematic representation of the HPLC elution profile of *C. picta* (turtle) SNBPs (see Fig. 2A, trace Tt) to indicate the nomenclature followed to designate the different fractions sequenced (see Fig. 4). B, urea (2.5 M)/acetic acid (5%, v/v)-PAGE of some of the protamine fractions shown in A, before (−) or after (+) treatment with alkaline phosphatase. *C. picta* protamines (lane Tt) ran in these gels in two well defined groups, I and II. Thus, fraction I-x depicts the fractions x (x = 1–5) from A that ran in group I of PAGE, and fraction II-y denotes the fractions y (y = 1–2) from A (shaded peaks) that ran in group II of PAGE. C, electrophoretic analysis of the peptides resulting from the digestion of different *C. picta* protamine fractions with protease V8 (panel 1) and elution profile (reverse-phase HPLC) of the peptides resulting from digestion of the turtle protamine fractions II-1 and II-4 with protease V8 (panel 2). The direction of electrophoresis for all the gels shown is from top (anode) to bottom (cathode).

### Table I

**Molecular mass of the turtle protamines**

The numbers in parentheses indicate protein components present in lower amounts.

| Protein fraction | Molecular mass | Mass spectrometry |
|------------------|----------------|-------------------|
|                  | Whole protein  | N-terminal^b       | C-terminal^b       |
|                  | MF AV         | MI AV              | MI AV              |
| Tt-I             | 7616.5 7620.7 | 1657 ± 2 1965 ± 2 | 4208 ± 2 278 ± 2 |
| 1                | 6151.7 6155.2 | 4205.5 4207.9     | 1965.2 1966.3     |
| 2                | 6192.6 6196.2 | 4136.5 4138.8     | 2075.2 2076.5     |
| 3                | 6011.6 6015.0 | 4110.6 4112.8     | 1856.2 204 ± 2    |
| 4                | 6289.8 6293.4 | 4205.5 4207.9     | 2104.2 2104.5     |
| 5                | 6289.8 6293.4 | 4205.5 4207.9     | 2104.2 2104.5     |

^a The nomenclature of the protein fractions is as described for Fig. 3.

^b The N- and C-terminal regions resulted from cleavage with staphylococcal protease V8 at a unique glutamic residue present in these proteins.

^c MI, monoisotopic mass; AV, average mass.
Protamines of Reptiles

Protamines of Squamata (Lizards and Snakes)—Fig. 1B shows an electrophoretic analysis of the SNBP composition of two snakes (lanes 3 and 4) and two lizards (lanes 5 and 6). The electrophoretic pattern corresponding to the protamine region appears extremely conserved, not only within, but also between, each group. The different extent of histones observed in each case is due to the different degree of contamination by immature spermatogenic cells, with those showing the lower amounts of histones present (i.e., Fig. 1B, lane 3) corresponding to the less contaminated samples.

Reverse-phase HPLC analysis of the SNBPs and the corresponding electrophoretic analysis of the fractions from each group are shown in Fig. 2 (A, traces Sn and Lz; and B, panels Sn and Lz). As can be seen, the levels of microheterogeneity exhibited by protamines from both the snake E. obsoleta quadrivittata and the lizard A. carolinensis are much lower than in the case of the turtle C. picta (Fig. 2, A, trace Tt; and B, panel Tt).

Upon reverse-phase HPLC refractionation of each of the individual peaks shown in Fig. 2 (A, traces Sn and Lz), mass spectrometry analysis (Figs. 5B and 6B) reveals the presence of highly purified monodisperse protamine fractions. It was possible, in both instances, to obtain the complete amino acid sequence of the different fractions by direct sequencing of the molecules involved (Figs. 5A and 6A). The average molecule masses determined from the primary structure were 5337.2 for Sn-I, 5190.0 for Sn-II, and 5068.9 for Lz, which are in excellent agreement with the masses experimentally determined by mass spectrometry. The sequences of the protamines from these organisms are very similar to the sequences of the group II protamines from the turtle C. picta (Fig. 4). There is an increase, however, in the presence of histidine and glycine as well as in the incorporation of lysine, which was absent in the turtle protamines. From the compositional point of view, the primary structure of the main protamine component from the lizard is one of the most simple protamine sequences, with only four amino acids present: arginine, glycine, lysine, and histidine.

Protamines of a Crocodilian—The apparent decrease of microheterogeneity in protamines of Squamata contrasts with the high extent of microheterogeneity found in A. mississippiensis (Fig. 2, A, trace Al; and B, panel Al). It is important to note that SNBP samples from at least 10 different alligators were analyzed in the course of this research. In each instance, the chromatographic and electrophoretic patterns observed were extremely constant, with only minor variations in the relative amounts of the different fractions, most likely resulting from the unavoidably different extent of sexual maturity. The electrophoretic pattern of the starting SNBP samples was also very similar to that reported earlier (Kasinsky et al., 1987).

One interesting feature of this system was that despite the apparently high resolutions of the HPLC fractionation (Fig. 2A, trace Al), none of the chromatographic peaks appeared to be electrophoretically pure (Fig. 2B, panel Al). The only way to obtain pure fractions, both at the electrophoretic level (Fig. 7B) and according to mass spectrometry criteria, was to fractionate the starting SNBPs extracts by ion exchange chromatography (Fig. 7A) in the presence of guanidinium chloride (Cole, 1989), followed by several rounds of reverse-phase HPLC. Thus, it looks as if the protamines from alligators consist of a complex microheterogeneous mixture of proteins. Following the aforementioned fractionation approach, we have been able to purify and sequence three of these protamine fractions (Figs. 7B and 8). The purified fractions belong to each of the three major electrophoretic bands in the starting sample (Fig. 7B, lane Al).

The sequences in this case were established from the overlap of the partial sequences of HPLC-purified peptides (Fig. 7C) obtained by digestion with different proteases.

In Table II, the molecular masses of these three protamine fractions (Al-I, Al-II, and Al-III), as determined by mass spectrometry, are shown in comparison to those established from their respective amino acid sequences. One remarkable struc-

---

**Fig. 4.** Amino acid sequences of the major protamine components of the sperm from the turtle C. picta. The proteins were digested with staphylococcal protease V8 (only the C-terminal peptide is underlined) upon HPLC fractionation of the resulting peptides (see Fig. 3C, panel 2). The mass of the purified peptides was determined by mass spectrometry (see Table I), and the sequences were determined using conventional Edman degradation. The nomenclature of the protamine components is as described for Fig. 3A. The asterisks designate the phosphorylated serines (see Fig. 3B) as detected by mass spectrometry.

Protamines of Squamata (Lizards and Snakes)—Fig. 1B shows an electrophoretic analysis of the SNBP composition of two snakes (lanes 3 and 4) and two lizards (lanes 5 and 6). The electrophoretic pattern corresponding to the protamine region appears extremely conserved, not only within, but also between, each group. The different extent of histones observed in each case is due to the different degree of contamination by immature spermatogenic cells, with those showing the lower amounts of histones present (i.e., Fig. 1B, lane 3) corresponding to the less contaminated samples.

Reverse-phase HPLC analysis of the SNBPs and the corresponding electrophoretic analysis of the fractions from each group are shown in Fig. 2 (A, traces Sn and Lz; and B, panels Sn and Lz). As can be seen, the levels of microheterogeneity exhibited by protamines from both the snake E. obsoleta quadrivittata and the lizard A. carolinensis are much lower than in the case of the turtle C. picta (Fig. 2, A, trace Tt; and B, panel Tt).

Upon reverse-phase HPLC refractionation of each of the individual peaks shown in Fig. 2 (A, traces Sn and Lz), mass spectrometry analysis (Figs. 5B and 6B) reveals the presence of highly purified monodisperse protamine fractions. It was possible, in both instances, to obtain the complete amino acid sequence of the different fractions by direct sequencing of the molecules involved (Figs. 5A and 6A). The average molecule masses determined from the primary structure were 5337.2 for Sn-I, 5190.0 for Sn-II, and 5068.9 for Lz, which are in excellent agreement with the masses experimentally determined by mass spectrometry. The sequences of the protamines from these organisms are very similar to the sequences of the group II protamines from the turtle C. picta (Fig. 4). There is an increase, however, in the presence of histidine and glycine as well as in the incorporation of lysine, which was absent in the turtle protamines. From the compositional point of view, the primary structure of the main protamine component from the lizard is one of the most simple protamine sequences, with only four amino acids present: arginine, glycine, lysine, and histidine.

Protamines of a Crocodilian—The apparent decrease of microheterogeneity in protamines of Squamata contrasts with the high extent of microheterogeneity found in A. mississippiensis (Fig. 2, A, trace Al; and B, panel Al). It is important to note that SNBP samples from at least 10 different alligators were analyzed in the course of this research. In each instance, the chromatographic and electrophoretic patterns observed were extremely constant, with only minor variations in the relative amounts of the different fractions, most likely resulting from the unavoidably different extent of sexual maturity. The electrophoretic pattern of the starting SNBP samples was also very similar to that reported earlier (Kasinsky et al., 1987).

One interesting feature of this system was that despite the apparently high resolutions of the HPLC fractionation (Fig. 2A, trace Al), none of the chromatographic peaks appeared to be electrophoretically pure (Fig. 2B, panel Al). The only way to obtain pure fractions, both at the electrophoretic level (Fig. 7B) and according to mass spectrometry criteria, was to fractionate the starting SNBP extracts by ion exchange chromatography (Fig. 7A) in the presence of guanidinium chloride (Cole, 1989), followed by several rounds of reverse-phase HPLC. Thus, it looks as if the protamines from alligators consist of a complex microheterogeneous mixture of proteins. Following the aforementioned fractionation approach, we have been able to purify and sequence three of these protamine fractions (Figs. 7B and 8). The purified fractions belong to each of the three major electrophoretic bands in the starting sample (Fig. 7B, lane Al).

The sequences in this case were established from the overlap of the partial sequences of HPLC-purified peptides (Fig. 7C) obtained by digestion with different proteases.

In Table II, the molecular masses of these three protamine fractions (Al-I, Al-II, and Al-III), as determined by mass spectrometry, are shown in comparison to those established from their respective amino acid sequences. One remarkable struc-
tural feature of these sequences is the presence of tryptophan in Al-I. This amino acid is seldom found in chromosomal proteins. As in the case of the protamines from Squamata, the protamine fractions from alligator are also histidine-rich. In contrast, however, lysine is not present. Instead, several tyrosine residues are present in all of them. Another important structural feature of the alligator protamines is the incorporation of the sequence ARYR-(BS)n motif (B = basic amino acid) in the protamine fractions from electrophoretic bands I and II. These two electrophoretic bands can amount to 60% of the protamine components found in the sperm. Such a major presence of this motif in alligator protamines represents an important departure from the rest of the reptilian protamines, in which the ARYR-(BS)n protamine fractions were only present in very low amounts.

**DISCUSSION**

*The Protamines of Reptiles: General Considerations*—The sequence information presented in the preceding section provides a better understanding of the chemical nature of vertebrate protamines as well as of the chemical transitions undergone by these proteins during the course of vertebrate evolution. It also provides a unique insight into the general chemical composition of the protamine class of proteins. In the sections that follow, we are going to address these points specifically. Fig. 9A shows the grouping (alignment) of the protamine sequences from a representative species of each of the main phylogenetic branches of vertebrates using parsimony analysis. Several general considerations can be drawn from this sequence comparison. The first is that all vertebrate protamines seem to have a protein “core” consisting of several arginine clusters. The size of the protamine molecules has experienced a gradual but significant increase in the course of the evolution of vertebrates. This may reflect the marked physiological advantage of protamine size during the processes of displacement of the histones from the early spermatogenic stages (Oliva and Mezquita, 1986). As shown in Fig. 9A, this increase in size has occurred as a result of increasing the N- and C-terminal domains of the molecule.

Chelonian SNBPs consist of several protamine variants showing an almost identical core/C-terminal region, but differing in their N-terminal domains. This provides some indirect evidence for the possible involvement of either alternative splicing or crossing-over mechanisms leading to the acquisition of the motif ARYR(X)y(BS)n (where $B = \text{basic residue}$ and $X$ is...
a variable amino acid, but usually R, N, H, or C, and \( n = 2 - 3 \), which is very conserved in the higher phylogenetic groups. The great extent of protein microheterogeneity exhibited by turtle protamines in particular and by amniote protamines in general cannot, in most instances, be accounted for by single point mutation events occurring in a precursor gene. This provides an indication of the rapid divergence of these proteins and their encoding genes (Oliva and Dixon, 1991; Oliva, 1995).

**FIG. 7.** A, ion exchange chromatography (Bio-Rex 70) of the SNBPs from *A. mississippiensis*. \( H \) and \( P \) depict the regions of the guanidinium chloride gradient corresponding to the elution of histones and protamines, respectively. The inset shows the electrophoretic analysis (on urea (2.5 M)/acetic acid (5%, v/v)-polyacrylamide gel) of some of the different fractions. *Lane Al*, starting unfraccionated protein sample. *B*, urea/acetic acid-PAGE of the alligator protamine fractions (I–III) obtained upon further HPLC fractionation in comparison with whole SNBPs obtained from alligator vas deferens. *C*, reverse-phase HPLC fractionation of alligator fractions I–III upon digestion with chymotrypsin (Ic–IIIc). The inset shows an electrophoretic analysis of these fractions and their corresponding digestion products. The arrowheads identify the peptide fractions used in the sequencing analysis (see Fig. 8).
Defining a Protamine: The Chemical Nature of Protamines—
The availability, for the first time, of amino acid sequence information from at least one organism within each of the different vertebrate taxa allows for a better definition of SN-BPs of the protamine type (Ausió, 1995). Thus, the definition proposed by Subirana (1983) (Lys + Arg = 45–80 mol % and Ser + Thr = 10–25 mol %) can now be revised in light of this new information. With the presence in reptiles of significant large amounts of histidine and glycine as well as the presence, to a lesser extent, of lysine, we propose the following compositional definition of protamines: Arg = 30 mol %, His + Lys + Arg = 45–80 mol %, and Ser + Thr + Gly = 10–25 mol %. This definition also takes into account the bounty of information gained in recent years on SN-BPs from other invertebrate groups (Ausió, 1995) that also contain protamines (Daban et al., 1995; Chiva et al., 1995).

A significant departure in this definition of the consensual protamine composition is the stress on the fact that protamines are always arginine-rich proteins regardless of their additional contents of histidine and lysine. It is important to point out that most of the lysine-rich proteins that had been previously included in this definition (Subirana, 1983) have now been shown to be related to proteins of the histone H1 family (Ausió, 1995). The rest of the amino acids that appear in the amino acid composition do not seem to exhibit any major restrictions, as can be clearly seen from a glance at the sequences shown in Fig. 9A. It is also obvious from an inspection of Fig. 9A that a new amino acid, cysteine, has been incorporated into the protamine sequence during the transition from birds to mammals (Oliva, 1995), in a similar fashion to the acquisition of histidine and glycine during the transition from fish protamines to those of amphibians and reptiles. Therefore, there appears to be no reason to classify the cysteine-rich protamines from mammals within a different group of proteins. In fact, as we will discuss next, we will see that in the course of the evolution of these proteins, the appearance of amino acids other than arginine has not occurred randomly.

Sequence Alignment Analysis—Fig. 9C shows a tree obtained from sequence alignment analysis of the protamines shown in Fig. 9A. The incomplete resemblance of the tree shown in Fig. 9C to the cladogram based on the fossil record shown in Fig. 1A emphasizes the limitation of parsimony analysis (Stewart, 1993). This may arise from different causes: mainly, the reduced number of sequences available as well as their short length (30–60 amino acids) and their arginine-rich nature (≥30 mol %), which are the most limiting factors of this kind of analysis (Doolittle, 1986). Despite all this, the tree in Fig. 9C does provide useful information at the molecular level about the evolutionary trend of protamines in vertebrates. As in the case of the protamine P1 fraction from mammals (Oliva, 1995), the existence of a clear trend along the main phylogenetic lines can be taken as indicative of the mechanism of vertical evolution of the vertebrate protamines and their genes, rather than horizontal transmission (see Oliva and Dixon (1991)). Also, the existence of such a trend implies that in the course of evolution, the nature of the non-arginine residues that are present in protamines is not random and has been carefully selected. The analysis in Fig. 9C shows that protamines from reptiles, particularly crocodilians, are closely related to those from birds and mammals, as will be discussed next.

The ARYR Protamine Motif and the Evolution of Protamines in Amniotes—The availability, for the first time, of protamine amino acid sequences from reptiles fills an important gap in the understanding of the molecular evolution of vertebrate protamines. The presence of a common ARYR-...-BS motif in bird, mammal, and reptile protamine sequences is also indicative of the mechanism of vertical evolution operating in the vertebrate protamines and their genes, rather than a horizontal transmission (see Oliva and Dixon (1991)). Also, the existence of such a trend implies that in the course of evolution, the nature of the non-arginine residues that are present in protamines is not random and has been carefully selected. The analysis in Fig. 9C shows that protamines from reptiles, particularly crocodilians, are closely related to those from birds and mammals, as will be discussed next.

Table II

| Protein fraction | Molecular mass | Sequence | Mass spectrometry |
|------------------|----------------|----------|-------------------|
|                  | Monoisotopic   | Average  |                   |
| Al-I             | 8204.6         | 8209.3   | 8204 ± 4 (8287 ± 7, 8327 ± 5) |
| Al-II            | 7534.3         | 7538.6   | 7540 ± 2          |
| Al-III           | 5637.2         | 5640.4   | 5658 ± 2 (5793 ± 3) |
|                  | (5786.3)       |          |                   |

a Calculated from the Al-III sequence with His in position 10 and Arg in position 13 (see Fig. 8, Al-III).
**FIG. 9.** A, comparison of the sequences of protamines from vertebrates. The sequences were aligned following the method described under "Materials and Methods." The nomenclature followed for the different protamine fractions is the same as described for Figs. 2–7. The sequences for protamines other than those obtained in this work were as follows: salmon, SL-I–IV (salmon protamine from *O. keta* (Hoffmann et al., 1990); protamines Bf1 and Bf2 from toad (*B. japonicus*) (Takamune et al., 1991); galline, Gll (fowl protamine from *Gallus domesticus*) (Nakano et al., 1976); protamine from adasyurid marsupial, Sm (*Sminthopsiscrassicaudata*) (Retief et al., 1995a); protamines P1 and P2 from boar Brp1 and Brp2 (*S. scrofa*) (Tobita et al., 1983; Maier et al., 1990). The arrowheads point to those protamine components whose sequences start with ARYR.

C, Bootstrapped protein parsimony tree generated from the vertebrate protamine sequences shown in A. The numbers at the nodes indicate the number of times the group consisting of the species, which are to the right of the fork, occurred among the trees out of 100 trees. The analysis was carried out using the PHYLIP package (Felsenstein, 1989).
sent in fish and amphibians, indicates that the protamines in these three groups are closely related. This can be clearly seen in Fig. 9C. Thus, the incorporation of the ARYR sequence, which is usually followed by a short stretch of alternating (BS) (B = basic residue and S = seryl residue) (Fig. 9A), must have taken place in the transition from amphibians to the monophyletic amniote group by the Upper Carboniferous (Carroll, 1988). The acquisition of this sequence (through possible mechanisms already discussed) and/or the appearance of this protamine variant seems to have occurred gradually. Whereas in turtles, snakes, and lizards, this protein appears only as a minor component (Fig. 9B, arrowheads), in the alligator, it amounts to almost 60% of the SNBPs (Al-I and Al-II). In birds, the complete replacement by the ARYR-BS-containing protein seems to have already taken place, and in mammals, this also represents the major protamine component. In the latter case, the protamine P1 component in some species coexists in variable amounts with a histidine-containing P2 component (Hecht, 1989). This protein transition affects not only the N-terminal region, but also the core region of the protamine molecule, which becomes increasingly rich in glycine (turtles, snakes, and lizards) as well as in histidine (crocodilians). In the transition to bird protamines, most of the histidines disappear. An important compositional change in this region takes place in mammals (which also affects the N- and C-terminal domains of the molecule) during the transition from Metatheria to Eutheria (Oliva, 1991) with the incorporation of cysteine. The appearance of cysteine also seems to have taken place gradually. Thus, while no cysteine can be detected in protamines from monotremes and in most marsupials (Oliva, 1991), a cysteine-containing protamine has been detected in at least one genus of shrew-like dasyurid marsupials (Planigales) (Retief et al., 1995a). This presence of cysteine in Planigales has been ascribed to a case of evolutionary convergence toward the cysteine-rich protamine P1 of eutherian mammals (Retief et al., 1995a). Yet the point mutation mechanism involved in the transition from non-cysteine- to cysteine-containing protamines in marsupials might have been similar to the ones that possibly occurred also in the transition from the non-cysteine-containing protamine precursor to the protamines in eutherian mammals.

The different amino acid transitions observed during vertebrate protamine evolution most likely reflect the different strategies followed in the packing of chromatin and in the stabilization of the sperm nuclei. The recent acquisition of multiple cysteines (with the possibility of inter- and intramolecular bond formation) in the protamines of eutherian mammals represents an obvious advantage in this respect.

An interesting feature that comes out from the cladogram shown in Fig. 9C is that the minor histidine-containing protamine P2 that is present in some eutherian mammals appears to be more primitive than protamine P1. However, this could simply reflect the slower rate of divergence of the P2 genes when compared with P1 (Retief and Dixon, 1993). As has already been pointed out earlier, it is hard to draw phylogenetic conclusions from the protein cladogram shown in Fig. 9C. Nevertheless, the significant increase in the presence of the ARYR protamine in the sperm of alligators when compared with other reptiles is fully consistent with a closer phylogenetic relation between crocodilians and birds. From looking at Fig. 1A, it remains unclear, however, what was the common selective pressure that led to the complete selection of this protamine in both birds and mammals.

Protamine Microheterogeneity—One of the striking features of protamines regarding their chemical and structural composition is their microheterogeneity (Subirana, 1983). This extent of microheterogeneity is extremely variable within and between different taxa, and hence, its origin and/or significance is still very puzzling. In salmonid fish (order Salmoniformes), most of this microheterogeneity can be accounted for by the presence of substantially polymorphic multicopy genes that encode multiple copies of protamines of restricted sequence variation (Oliva and Dixon, 1991). However, whereas in this case the protamine genes constitute a multigene family with a minimum of 15–20 members (Oliva and Dixon, 1991), in the yellow perch (order Perciformes), only one protein species seems to be present (Chao and Davies, 1992).

In amphibians, different extents of SNBP microheterogeneity have been observed in the toad Bufo japonicus (Takamune et al., 1991) and in species in the genus Xenopus (Mann et al., 1982). In birds, a copy number of only two protamine genes has been found in the two species that have been studied so far (Oliva and Dixon, 1991). In mammals, a single copy of P1 and P2 per haploid genome has been detected (Oliva and Dixon, 1991). In reptiles, as has already been pointed out, the extent of microheterogeneity also varies significantly from one group to another (Fig. 2A), with turtle (C. picta) and alligator exhibiting the maximum extent of microheterogeneity (Figs. 1B, lane 7; and 2A, traces T1 and Al). This high extent of microheterogeneity in chelonians and crocodilians contrasts with the lower microheterogeneity observed in the phylogenetically close avian group. As noted by Subirana (1983), this broad variation in the extent of protamine microheterogeneity between different taxa points to a low physiological relevance of this phenomenon. It may, on the other hand, have significant evolutionary implications.

From a careful inspection of Fig. 9A, it appears that during the evolutionary burst leading to the appearance of amniotes, protamines underwent an important departure from the chemical and structural organization in previous vertebrate groups. While the arginine-rich nature was maintained, an important reshuffling of the non-basic residues connecting the arginine clusters as well as an increase in the overall size of the molecule occurred at this point. In this respect, turtles (Fig. 1A) represent the earliest group to undergo the transition that led to the initial reptilian protamine pattern, a pattern that already appears to be established in snakes and lizards. However, the incorporation of the ARYR-BS-containing protamine as a major SNBP component did not occur until the appearance of crocodilians, and it appears to have become fully established in birds. Therefore, it looks as if the extent of microheterogeneity increases in those organisms belonging to phylogenetic groups in which the main protamine transitions have taken place. It is thus quite tempting to ascribe protamine microheterogeneity to the onset of gene duplication processes involved in the evolution of protamine genes (see Oliva and Dixon (1991). Although the precise mechanisms involved remain to be established, the sequence data presented in this paper represent an important step in this direction. With this information, it should now be possible to design DNA primers that, in combination with the manifold molecular biology techniques presently available, should allow us to obtain the information that is ultimately required to understand the evolution of protamine genes in the vertebrates.

Acknowledgments—We are very indebted to Monica Salmerón, Drs. Wayne Vogel and E. William Byrd, Jr., Dave Kulak, and Ellen Rosenberg for providing some of the specimens used in this work. We thank Ann Gillespie, John Caveness, and Dr. Horacio De La Cueva for helping in transporting biological material. We also thank Jenny Colom for an initial experiment on snake SNBPs. All the sequencing work reported in this paper was carried out at the Protein Microchemistry Center of the University of Victoria (British Columbia, Canada). We are very thankful to Sandy Kielland of this center for providing skillful protein sequencing. We are also very grateful to Sharon Belton and Stacey Gavora for skillful typing of the manuscript.
Protamines of Reptiles
Janice G. Hunt, Harold E. Kasinsky, Ruth M. Elsey, Cynthia L. Wright, Philip Rice, Jason E. Bell, Daniel J. Sharp, Andor J. Kiss, Donald F. Hunt, David P. Arnott, Mark M. Russ, Jeffrey Shabanowitz and Ausió Juan

J. Biol. Chem. 1996, 271,23547-23557.
doi: 10.1074/jbc.271.38.23547

Access the most updated version of this article at http://www.jbc.org/content/271/38/23547

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 41 references, 5 of which can be accessed free at http://www.jbc.org/content/271/38/23547.full.html#ref-list-1