Lindahl, P. E. (1947) *Nature (London) 161*, 648–649.
Ohkuma, S., & Poole, B. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 3327–3331.
Petersen, C. M., Ejlersen, E., Hansen, P. W., & Gliemann, J. (1987) *Scand. J. Clin. Lab. Invest. 47*, 55–61.
Sanderson, R. J., Bird, K. E., Palmer, N. F., & Brenman, J. (1976) *Anal. Biochem. 71*, 615–622.
Scatchard, G. (1949) *Ann. N.Y. Acad. Sci. 51*, 660–672.
Sottrup-Jensen, L., Petersen, T. E., & Magnusson, S. (1980) *FEBS Lett. 121*, 275–279.
Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Wierzbicki, D. M., Jones, C. M., Londblad, P. B., Magnusson, S., & Petersen, T. E. (1984) *J. Biol. Chem. 259*, 8318–8327.
Straight, D. L., & McKee, P. A. (1982) *Biochemistry 21*, 4550–4556.
Swensen, R. P., & Howard, J. B. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4313–4316.
Van Leuven, F., Cassiman, J. J., & Van Den Berghe, H. (1978) *Exp. Cell Res. 117*, 273–282.
Williams, L. T., Snyderman, R., Pike, M. C., & Lefkowitz, R. J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 1204–1208.
Willingham, M. C., Maxfield, F. R., & Pastan, I. H. (1979) *J. Cell Biol. 82*, 614–625.

Purification and Characterization of Mouse Protamines P1 and P2. Amino Acid Sequence of P2†

Anthony R. Bellvé*
Department of Physiology and Biophysics and Laboratory of Human Reproduction and Reproductive Biology, Harvard Medical School, Boston, Massachusetts 02115
Donald J. McKay, Bernard S. Renaux, and Gordon H. Dixon
Department of Medical Biochemistry, University of Calgary, Calgary, Alberta, Canada T2N 1N4
Received September 2, 1987; Revised Manuscript Received December 16, 1987

ABSTRACT: Two mouse protamines, denoted as P1 and P2, have been purified directly from mature sperm nuclei and characterized as distinct polypeptide species. The complete primary structure of P2 was determined by peptide sequencing analyses. P1 and P2 were purified by a sequence of cation-exchange chromatography on Bio-Rex 70 and permeation chromatography on Bio-Gel P10, both in the presence of guanidine hydrochloride. Biochemical analyses demonstrate P1 has a molecular weight of 7400 and is characterized by the presence of arginine, cysteine, lysine, and tyrosine. By contrast, P2 is unusual in containing an abundance of arginine, histidine, lysine, and cysteine, but no tyrosine. The primary structure of P2 was determined from the sequencing of overlapping, high-pressure liquid chromatography purified peptides generated by thermolysin and endoproteinase Lys-C digestions and by chemical cleavage at each of four serine residues. Sequence analyses have demonstrated that P2, with a molecular weight of 8841, contains 62 amino acids, in the sequence NH_2-Arg-Gly-His-His-His-His-Arg-His-Arg-Arg-Cys-Ser-Arg-Lys-Arg-Leu-His-Arg-Ile-His-Lys-Arg-Arg-Arg-Ser-Cys-Arg-Arg-Arg-Arg-Arg-His-Ser-Cys-Arg-His-Arg-Arg-Arg-His-Arg-Arg-Gly-Cys-Arg-Arg-Ser-Arg-Arg-Arg-Arg-Arg-Cys-Arg-Cys-Arg-Lys-Cys-Arg-Arg-His-His-COOH. Thus, the primary structure includes six clusters of arginine and histidine, distributed throughout the polypeptide, each ranging from five to eight amino acids in length. Sequence comparisons of mouse and human protamines by the Dayhoff program have revealed greater homology exists between human P2 and mouse P2 than within the P1 family from the two mammalian species.

During mammalian spermiogenesis, the germ cell nucleus undergoes a marked transition as many chromosomal proteins are removed and then replaced by an array of novel polypeptides (Bellvé et al., 1975; O'Brien & Bellvé, 1980a,b). In this process of nuclear transformation, the histones are replaced initially by a set of testis-specific proteins (TSP)[1] (Grimes et al., 1977) and then by the protamines (Bellvé et al., 1975; Balhorn et al., 1984), the predominant proteins comprising the mature sperm nucleus. On the basis of current evidence, the low molecular weight protamines exist in two classes: a ubiquitous P1 family rich in arginine, lysine, cysteine, and tyrosine and a P2 species rich in arginine, histidine, and lysine but lacking tyrosine. Expression of the latter protamine has been detected in the mouse (Bellvé et al., 1975; Balhorn et al., 1977; Bellvé & Carraway, 1978) and human (Kolk & Samuel, 1975). Complete amino acid sequences for members of the P1 family have been determined for bulls (Coelingh et al., 1972; Mazrimas et al., 1986), boars (Tobita et al., 1983), rams (Sautiere et al., 1984), and humans (McKay et al., 1985;

† This research was funded by National Institutes of Child Health and Human Development Research Grant RO1 HD 08270 to A.R.B. and by the Medical Research Council of Canada through an operating grant to G.H.D. and the Alberta Heritage Foundation for Medical Research which provided the sequencing facilities.
* Address correspondence to this author at the Department of Anatomy and Cell Biology, College of Physicians and Surgeons of Columbia University, 650 West 168th St., New York, NY 10032.

[1] Abbreviations: M_r, relative molecular weight; HPLC, high-pressure liquid chromatography; Gdn·HCl, ultrapure guanidine hydrochloride; PMSF, phenylmethanesulfonyl fluoride; DTT, dithiothreitol; TSP, testis-specific protein(s); SDS, sodium dodecyl sulfate; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; EDTA, ethylenediaminetetraacetic acid; PAGE, polyacrylamide gel electrophoresis; PTH, phenylthiohydantoin.

0006-2960/88/0427-2890$01.50/0

Ammer et al., 1986) and, based on cDNA sequencing, have been predicted for mouse P1 (Kleene et al., 1985). Primary sequences for members of the P2 family have been determined for human by protein sequencing (McKay et al., 1986; Ammer et al., 1986) and for mouse by cDNA sequencing (Yelick et al., 1987).

The P1's from these mammalian species have proven to be comparable in length (50 amino acids), each consisting of a central core of 3 arginine clusters and the less basic amino- and carboxyl-terminal regions. The conserved domains of arginine are considered to interact tightly with DNA to yield highly condensed nucleoprotamine [see review by Bellvé and O'Brien (1983)], with the cysteine groups stabilizing the complex through the formation of covalent disulfide bonds (Bedford & Calvin, 1974). Presumably, given the degree of sequence homology among the P1 species, it should be feasible to reach some concensus concerning the various models through which P1 interacts with sperm DNA.

The P2 species, by contrast, are limited in expression to the mouse and human and are unusual in containing high levels of both arginine and histidine, but lacking in tyrosine (Bellvé & Carraway, 1978). Moreover, recent evidence suggests that this particular protamine is synthesized as a larger precursor that is processed to the mature form of the polypeptide within the condensing spermatid nucleus (Yelick et al., 1987). Due to the unusual characteristics of mouse P2, it was considered important to undertake a complete primary sequence analysis and thereby allow a direct comparison of this protein with the known sequences of human P2a and P2b. The ensuing data have provided a biochemical characterization of the mouse sperm P1 and P2 species. Moreover, determination of the complete primary structure of P2 has demonstrated a 62 amino acid polypeptide containing 6 major domains each comprised exclusively of arginine, histidine, and lysine. Primary sequencing demonstrates P2 terminates with the carboxyl sequence of -Cys-Arg-Arg-His-His.

Materials and Methods

Materials. Adult male, CD-1 mice, 10–20-weeks-old, were obtained from Charles River Breeding Laboratories (Wilmington, MA). Ultrapure urea and Gdn·HCl were purchased from Schwarz/Mann (Orangeburg, NY); ethyleneimine and sequanal-grade glacial acetic acid and HCl were from Pierce Chemical Co. (Rockford, IL); PMSF, DTT, β-lactoglobulin, and thermolysin (EC 3.4.24.4) were from Sigma Chemical Co. (St. Louis, MO); endoproteinase Lys-C (EC 3.421.50) was from Boehringer-Mannheim. Spectrapor 3 dialysis tubing with a 3500 molecular weight cutoff was obtained from Fisher Scientific Co. (Waltham, MA), and L-[^{14}C(U)]arginine (specific activity 300 mCi/mmol) and L-[2,3-^{3}H]arginine (specific activity 25 Ci/mmol) were from New England Nuclear Corp. (Boston, MA).

Isolation and Dissociation of Sperm Nuclei. Spermatozoa were recovered from the caudae epididymides and vasa deferentia of 450 mice, washed 3 times in 75 mM NaCl and 24 mM EDTA (S-EDTA), and decapitated by a 2-min homogenization in 1% SDS in 50 mM Tris-HCl (pH 7.2) (Bellvé et al., 1975; O'Brien & Bellvé, 1980a). After filtration through 80-μm Nitex cloth, nuclei were purified (>99%) by centrifugation through 1.6 M sucrose at 5000*g* and 20 °C for 60 min, by using a Beckman J2-21B centrifuge equipped with a 7.5S rotor. All sperm tails were aspirated from the sucrose interface. Nuclei (~8×10^9) in the pellets were suspended and then washed 3 times with S-EDTA to remove residual SDS. The final preparation was suspended in 6.0 M Gdn·HCl, 50 mM DTT, 5 mM EDTA, and 1 mM PMSF in 0.5 M Tris-HCl (pH 8.6) and incubated with 50 mM DTT, at 33 °C, under N_2 for 60 min. Reduced cysteinyl groups were alkylated by addition of ethyleneimine to 250 mM, with constant stirring (Coelingh et al., 1969). After 60 min, the reaction was terminated by adding excess DTT. DNA was sedimented by centrifugation at 200000*g* for 12–14 h at 4 °C in a Beckman L2-65B centrifuge equipped with a Ti 75 rotor. The supernatant was dialyzed (3500 molecular weight cutoff) to a final concentration of 8% Gdn·HCl and 1 mM PMSF in 0.1 M sodium phosphate buffer (pH 6.8).

Purification of Protamines. Dialyzed, alkylated proteins were fractionated by ion-exchange chromatography on Bio-Rex 70 (1.5 × 35 cm), previously equilibrated with 8% Gdn·HCl in 100 mM sodium phosphate buffer (pH 6.8). After the column was washed with the same buffer to remove nonbasic constituents, bound moderately basic proteins were eluted with 150 mL of 18% Gdn·HCl (pH 6.8). Protamines then were eluted with a linear gradient of 20–50% Gdn·HCl in 100 mM sodium phosphate buffer (pH 6.8). Those 10-mL fractions containing protein peaks, at ~18%, 24%, and 45% Gdn·HCl, respectively, were pooled, dialyzed againt 250 volumes of 1 mM HCl followed by 50 volumes of deionized water, and then lyophilized.

Protamine peaks resolved by chromatography on Bio-Rex 70 were subjected to permeation chromatography on Bio-Gel P10 (2.5 × 100 cm), in the presence of 2.5 M Gdn·HCl in 200 mM Tris-HCl (pH 6.8). Each sample was dissolved in 2 mL of 2.5 M Gdn·HCl and 50 mM DTT in 200 mM Tris-HCl (pH 7.6), and ethyleneimine was added to 200 mM for 30 min. The sample was applied to the column, and eluant fractions containing protein were pooled, dialyzed against 10 volumes of 100 mM, sequanal-grade, acetic acid, and concentrated by adsorption to Amberlite IRC-50. After being extensively washed, absorbed protamines were eluted with 1.0 M, sequanal-grade, acetic acid and stored in 10 nM aliquots at −70 °C.

Molecular Weight Determinations. Estimates of molecular weights were obtained by permeation chromatography on 1.5M agarose in 6 M Gdn·HCl/0.2 M Tris-HCl (pH 7.2), by applying protamines and standard polypeptides of known mass. Protamines 1 and 2, [^{14}C]arginine or [^{3}H]arginine labeled, were prepared by injecting 25 μCi of the respective precursor in 20 μL of enriched Krebs–Ringer bicarbonate medium (Romrell et al., 1976) into the testes of 2 groups of 10 adult mice. Isotopically labeled spermatozoa were recovered 14 days later, from the caudae epididymides and vasa deferentia (Bellvé et al., 1975). After isolation and dissociation of the nuclei, the protamines, ^{14}C or ^{3}H labeled, were aminoethylated and purified as described above. Standard proteins were carboxymethylated and used to calibrate the column; these included β-lactoglobulin (M_r 18 433), cytochrome *c* (M_r 12 433), and the CNBr-cleaved peptides of cytochrome *c*, denoted as C_I (M_r 7678), C_{II} (M_r 2530), and C_{III} (M_r 1540) (Fish et al., 1969). These peptides were generated and purified by the procedure of Steers et al. (1965). Blue dextran and dansyl-L-alanine were used to determine void (V_o) and total (V_i) column volumes, respectively. The molecular weights of P1 and P2 were estimated from the volume determinations of four chromatography profiles. These data were averaged and analyzed by the procedure of Porath (1963) for flexible polymer analysis, where $K_d^{1/3}$ is plotted against $M_r^{0.555}$, by using the relationship of $K_d = (V_e - V_o)/(V_i - V_o)$ in which the respective volumes are substituted by weight (Fish et al., 1969).

Assessment of Protamine Purity by PAGE. Homogeneity of both P1 and P2 was assessed by acetic acid–urea/PAGE

(Panyim & Chalkley, 1969; Bellvé et al., 1975), with modifications to enhance resolution. Protamine samples were dissolved in 8 M urea/100 mM acetic acid and electrophoresed on 25 × 12 × 0.3 cm slab gels of 15% polyacrylamide, 0.9 M acetic acid, and 2.5 M urea. Following electrophoresis, proteins were stained with 0.1% amido black in 10% acetic acid and 25% methanol and then destained electrophoretically and stored in 10% acetic acid and 25% methanol. Absorbance profiles were quantified by using an Ortec Systems gel scanner.

Amino Acid Analyses. Aminoethylated derivatives of P1 and P2 were subjected to amino acid analyses after acid hydrolyses. Purified protamine was dissolved in sequanal-grade 6 N HCl, flushed with N_2, and hydrolyzed in vacuo at 140 °C for 4, 8, 16, and 24 h. Composition was determined by using a Beckman 121M amino acid analyzer equipped with a 25-cm physiological column to facilitate resolution of (aminoethyl)cysteine. Estimates for serine, threonine, and tyrosine were derived by extrapolation to zero time.

Peptide Sequencing Procedures. Amino acid analyses undertaken for the sequencing of P2 were performed on a Beckman 6300 amino acid analyzer. Lysine and (aminoethyl)cysteine were resolved separately by reducing the column temperature from 60 to 40 °C after the elution of phenylalanine and then by returning to 60 °C after the elution of ammonia. Peptide sequencing was performed by using an Applied Biosystems 470A protein sequencer (Hunkapillar et al., 1983; McKay et al., 1985). The PTH-amino acids produced were identified by using reverse-phase HPLC, equipped with an Alltech Econosphere C18 cartridge column (5 μm × 4.6 mm × 250 mm). The elution program consisted of a 21-min sodium acetate (pH 5.2, 15–50 mM)/acetonitrile gradient, run at 1.0 mL/min and 50 °C. In these conditions, PTH-(aminoethyl)cysteine eluted just after PTH-lysine.

Aminoethylated P2 was digested with thermolysin by dissolving 20 nmol of protein in 0.2 mL of 0.5% ammonium bicarbonate, adding 2 μg of thermolysin in ammonium bicarbonate, and incubating at 37 °C for 20 h. Endoproteinase Lys-C digestions were accomplished by dissolving 10 nmol of aminoethylated P2 in 0.1 mL of 0.05 M Tris-HCl (pH 9.0), adding 0.1 unit of enzyme, and incubating at 37 °C for 20 h.

Chemical cleavage on the N-terminal side of the serine residues of P2 was performed by using a modification of a procedure previously applied to herring protamines (Iwai & Ando, 1967; Iwai et al., 1971). Aminoethylated P2 (10 nmol) was dissolved in 0.1 mL of concentrated sulfuric acid and incubated at 22 °C for 96 h. After addition of 0.5 mL of H_2O and then 0.6 mL of 6 N HCl at 10 °C, the incubation was continued at 22 °C for another 30 h. The digest was cooled to 5 °C, partially neutralized with 0.1 mL of concentrated ammonium hydroxide, and concentrated to 0.2 mL in a vacuum centrifuge. The samples were not dried completely, as salts started to precipitate as the volume decreased below 0.2 mL.

The peptides in each digest were purified by reverse-phase HPLC on a Vydac C18 column, by eluting with an aqueous trifluoroacetic acid/acetonitrile solvent system (Mahoney & Hermodson, 1980; McKay et al., 1985).

RESULTS

Purification of Protamines. Native protamines from mouse sperm nuclei gave a single polypeptide band of high mobility on acetic acid–urea/PAGE, but after aminoethylation, two protein species were resolved. Both modified protamines expressed enhanced electrophoretic mobilities relative to the native protein, as reported previously (Bellvé et al., 1975; Balhorn et al., 1977; cf. Figure 3). Together, this evidence suggests mouse sperm, unlike those from most other mammals, contains two protamine species that differ substantially in their molar content of cysteinyl groups (Bellvé, 1979; Bellvé & O'Brien, 1983). Direct evidence to support this contention is provided by the following data.

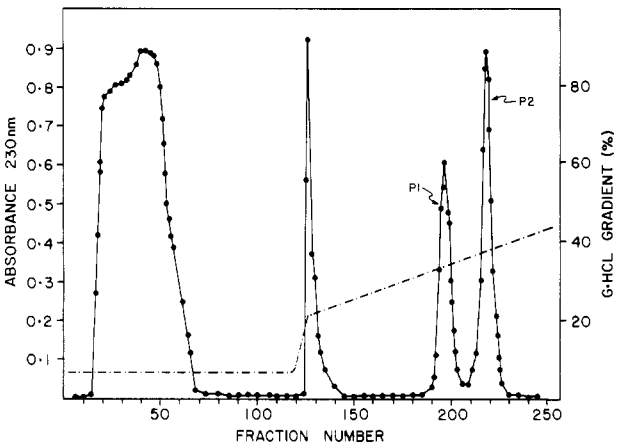

FIGURE 1: Separation of mouse protamines by ion-exchange chromatography on Bio-Rex-70. Sperm nuclear proteins, in 8% Gdn·HCl and 0.1 M sodium phosphate buffer (pH 6.8), were applied to the column (1.5 × 35 cm). Acidic and moderately basic proteins were eluted in the first two peaks. Two protamine species, denoted as P1 and P2, then were separated by applying a linear 20–50% gradient of Gdn·HCl (·–·). Absorbance at 230 nm (●).

Proteins dissociated from highly purified sperm nuclei were resolved into three major classes—acidic, basic and highly basic constituents—after being aminoethylated and resolved by ion-exchange chromatography on Bio-Rex 70 (Figure 1). The first two classes have been shown to be comprised of a heterogeneous array of proteins (O'Brien & Bellvé, 1980a). The last class consists of two symmetrical peaks of protamine, designated in order of elution as P1 at ~25% and P2 at ~45% Gdn·HCl. The difference in elution volume from the ion-exchange column suggests that P1 and P2 differ in their net basic charge. When the two protein peaks were reaminoethylated and rechromatographed separately, each eluted again as a single symmetrical peak at the original Gdn·HCl concentration. In neither case was protein detected at the elution position of the other protamine, thereby negating the possibility that the apparent charge difference was due to incomplete aminoethylation of the cysteinyl groups or due to dimerization of a single protamine species. Moreover, only two protein peaks were resolved when the chromatography conditions were buffered to pH 5.5 or 6.0 (data not shown).

The two protamines were purified further by exclusion chromatography on Bio-Gel P10, in the presence of 2.5 M Gdn·HCl, to enhance resolution and recovery (Figure 2). In these conditions, P1 and P2 each chromatographed as a major symmetrical peak and with a marginal difference in elution volumes. Only minor amounts of material eluted at higher molecular weights. Identity of purified P1 and P2 was ascertained by coelectrophoresis with native and aminoethylated mouse protamines on acetic acid–urea/PAGE (Figure 3). By this procedure, purified P1 comigrated with the faster polypeptide band that presumably contains a higher molar content of (aminoethyl)cysteine, whereas P2 comigrated with the band expressing the lower electrophoretic mobility.

Estimation of Molecular Weights. Existing evidence suggests P1 and P2 differ in relative mass. First, on each of three separate chromatographs on Bio-Gel P10, the two proteins were recovered at measurably different elution volumes (Figure 2). Second, *native* P1 and P2 had comigrated on acetic

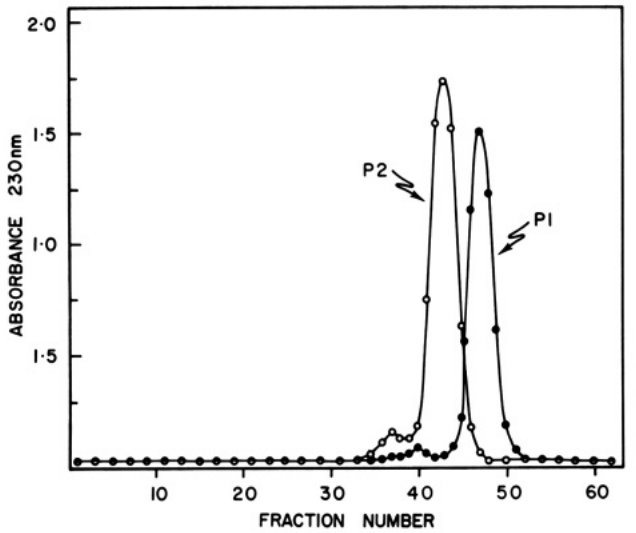

FIGURE 2: Purification of mouse protamines by permeation chromatography on Bio-Gel-P10 (100 × 2.5 cm) in the presence of 2.5 M Gdn·HCl. P1 and P2, after being resolved by Bio-Rex-70 ion-exchange chromatography (see Figure 1), were applied separately to the column. Aside from some material of higher molecular weight the two protamines were recovered as major symmetrical peaks, with a measurable difference (~10 mL) in elution volumes.

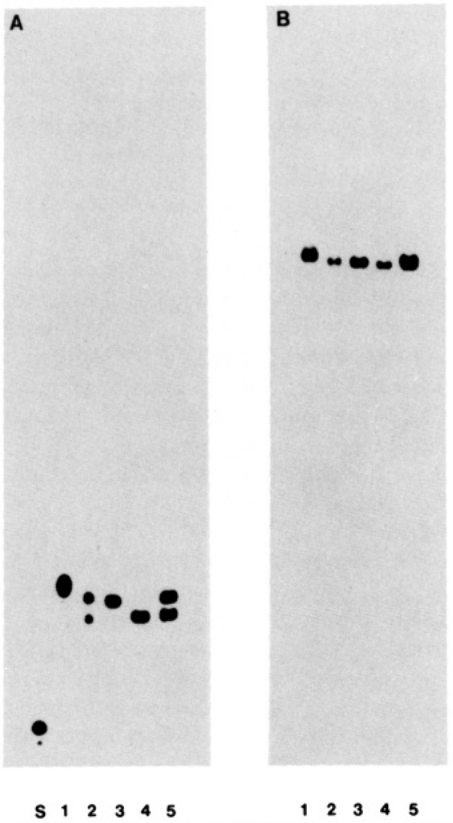

FIGURE 3: Purity of mouse sperm protamines assessed by high-resolution acetic acid–urea/PAGE, at pH 2.7 (A) and pH 4.5 (B). (Panel A) Native protamines recovered directly from sperm nuclei migrated as a single band, but after aminoethylation, P1 and P2 were resolved as two distinct bands, each showing enhanced electrophoretic mobility. Following purification by a sequence of ion-exchange and gel permeation chromatography, aminoethylated P1 and P2 again both migrated as single and apparently homogeneous species. (Panel B) Mobility and resolution of both proteins were reduced markedly at pH 4.5. Samples included (1) native protamines, (2) aminoethylated protamines, (3) purified P2, (4) purified P1, (5) recombined P1 and P2, and (S) salmon protamines.

acid–urea/PAGE (Figure 3), even though P2 exhibited a greater net basic charge by ion-exchange chromatography

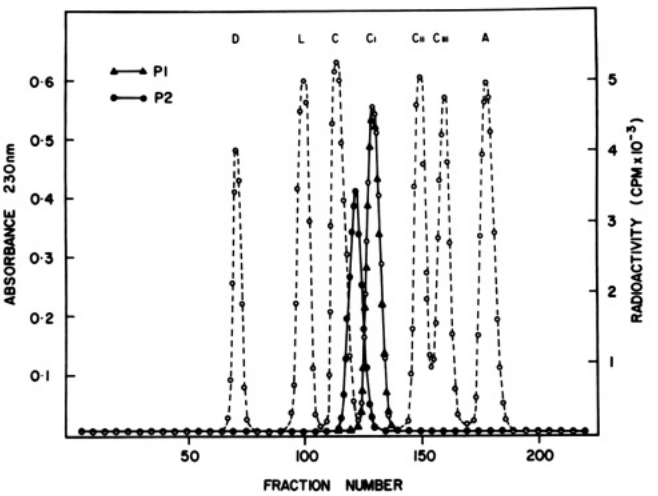

FIGURE 4: Molecular weight determination of mouse protamines by applying chromatography on agarose 1.5M (100 × 2 cm) in the presence of 6.0 M Gdn·HCl. P1 and P2, after being labeled in vivo with [^{3}H]arginine and [^{14}C]arginine, respectively, were aminoethylated, purified, and then applied simultaneously to the column, in a 500-μL, sample volume. The elution profiles of P1 (▲) and P2 (●) were determined by quantitating the respective isotopes. Carboxymethylated polypeptide standards were applied simultaneously. These included the following: (L) β-lactoglobulin (M_r 18 433); (C) cytochrome c (M_r 12 428); and the CNBr-generated cytochrome c peptides C_I (M_r 7678), C_{II} (M_r 2530), and C_{III} (M_r 1540). (D) Blue dextran and (A) dansyl-L-α-alanine were used to define V_o and V_i, respectively. Note than P1 coeluted with the cytochrome C_I fragment.

Table I: Amino Acid Compositions[a]

amino acid	MP1[b]	MP2	S1	S2
Thr	1 (1)			
Ser	3 (4)	3 (4)	1.0	1.0 (1)
Gly		2 (2)	1.0	0.7 (0)
Ala	1 (1)			
Cys	9 (9)	6 (7)	*c*	*c* (3)
Ile	1 (1)	1 (1)		
Leu		2 (1)		
Tyr	4 (3)			
His		12 (12)	2.3	2.7 (2)
Lys	3 (3)	3 (3)	3.8[c]	4.7[c] (1)
Arg	33 (28)	35 (32)	7.7	7.0 (9)
total	55 (50)	63 (62)		16.1 (16)

[a] Data in parentheses are derived from the amino acid sequences. [b] Mouse P1 composition is from cDNA of Kleene et al. (1985). [c] (Aminoethyl)cysteine and lysine were determined as lysine.

(Figure 1). The difference in molecular weight was demonstrated definitively by cochromatography of [^{3}H]arginine-labeled P1 and [^{14}C]arginine-labeled P2 on agarose 1.5 M equilibrated with 6.0 M Gdn·HCl, according to the procedure of Fish et al. (1969). Comparable analyses also were undertaken by using [^{14}C]arginine-labeled P1 and [^{3}H]arginine-labeled P2. By use of methods for dual-isotope quantitation, along with cochromatography of standard proteins, the relative molecular weights of P1 and P2 were determined accurately (Figure 4). Analysis of the data by applying Porath plots, using the relationship of $K_d^{1/3}$ versus $M_r^{0.555}$, gave molecular weight estimates for aminoethylated P1 of ~7500 and P2 ~8800. The molecular weight estimate for P1 was consistent with its comigration with the CNBr-generated C_I peptide (M_r ~7678).

Amino Acid Analyses of P1 and P2. Both protamines are comprised predominantly of arginine (55–60%), whereas P2 also contains an unusually high molar content of histidine (Table I). Overall, P2 is highly basic with almost 80% of its residues consisting collectively of arginine, histidine, and lysine. This marked difference in net basicity of P1 and P2 is con-

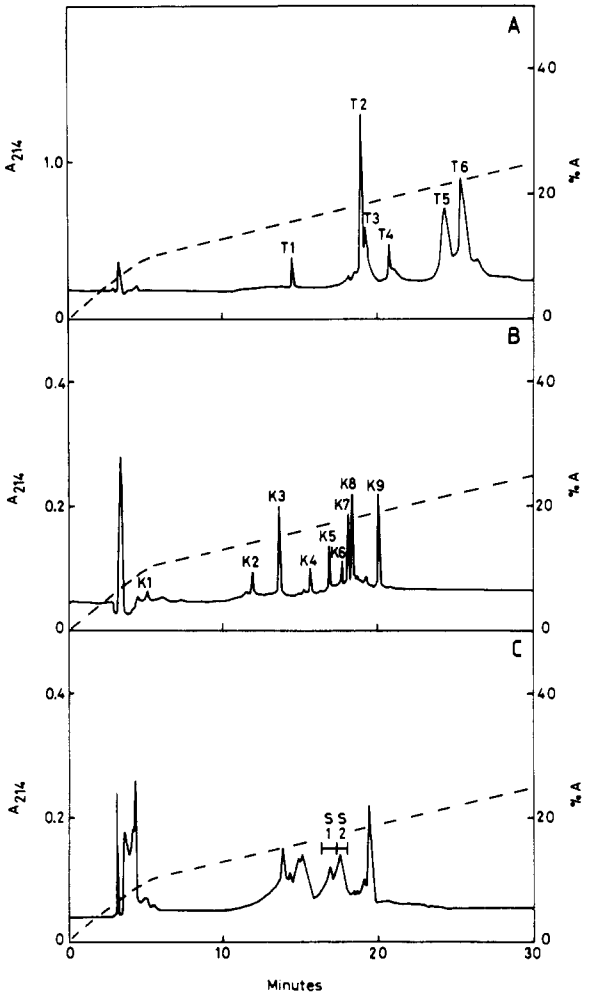

FIGURE 5: Reverse-phase HPLC separation of mouse protamine P2 fragments. Column, Vydac C18; temperature, 30 °C; solvent A, 0.09% trifluoroacetic acid in acetonitrile; solvent B, 0.1% trifluoroacetic acid in water; flow rate, 1.0 mL/min. (A) Thermolysin digest of P2; (B) endoproteinase Lys-C digest of P2; (C) P2 fragments from chemical cleavage at serine residues.

sistent with their apparent difference in net charge on Bio-Rex-70 chromatography. In addition, both protamines contain cysteine, ranging from 6% to 12%, a feature shared with other mammalian protamines. The two polypeptides also differ in the presence or absence of the amino acids threonine, tyrosine, alanine, glycine, and leucine (Table I).

Amino Acid Sequence of Protamine P2. Direct N-terminal sequencing for 20 cycles of Edman degradation on 5 nmol of aminoethylated P2 demonstrated the presence of a second shorter sequence in the P2 preparation. Arginine and serine were detected as the major amino acids in cycle 1, and the subsequent data from cycles 2–10 were difficult to interpret. However, the results from cycles 11–20 were clearer and consistent with the final sequence in this region, when carry-over was taken into account. Leucine first appeared in cycle 16, and isoleucine first appeared in cycle 19, indicating that these two possible sites for thermolysin cleavage were located in similar positions to those in human P2 (McKay et al., 1986; Ammer et al., 1986).

Consequently, the strategy of undertaking separate digestions with thermolysin and endoproteinase Lys-C was employed, as used previously for sequencing human P2 (McKay et al., 1986). Digestions were undertaken with each enzyme on 5-nmol aliquots of aminoethylated P2 for 4 h and overnight, and then the samples were subjected to preparative, reverse-phase HPLC. The purified fragments were hydrolyzed and

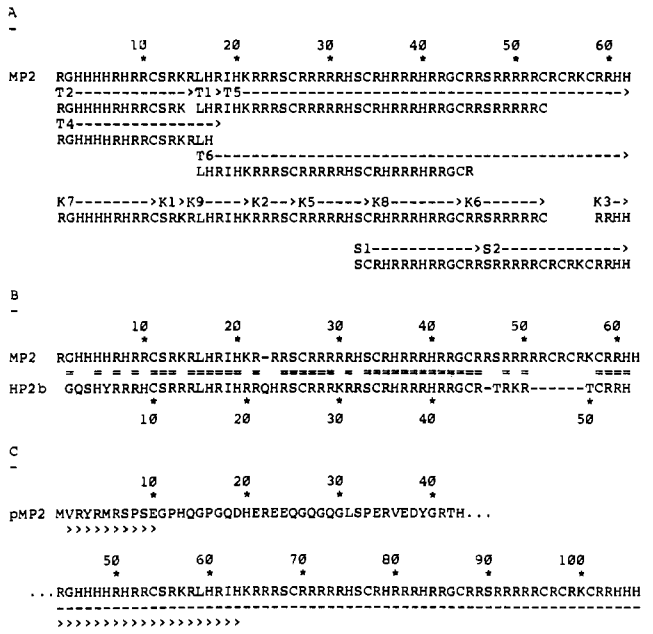

FIGURE 6: P2 amino acid sequence. (A) Summary of the sequencing results for mouse P2. T, thermolysin peptides; K, endoproteinase Lys-C peptides; S, fragments from chemical cleavage at serine residues; (→), peptide length. (B) Optimal alignment of mouse P2 and human P2b amino acid sequences. (=) Identical residues; (–) gap introduced into sequence to optimize alignment. (C) Amino acid sequence of mouse P2 precursor, pMP2, determined by Yelick et al. (1987). The complete sequence was predicted from the cDNA sequence; (>) the N-terminal sequence of the precursor starting at residue 7 and that of mature P2 starting at residue 44 were determined by protein sequencing; (--) predicted amino acid sequence of mature MP2.

analyzed for amino acids. Overnight digestion gave the better result in each case.

Six peptide peaks labeled T1–T6 were obtained by using reverse-phase HPLC from a thermolysin digestion of 20 nmol of aminoethylated P2 (Figure 5A). The entire amount of each fraction was applied to the gas-phase sequencer and processed as far as possible. Each fraction gave a single sequence. These results provided confirmatory evidence for residues 1–45 and a tentative sequence for residues 46–53 (Figure 6A). Fraction T3 (sequence not shown) yielded only 10 residues corresponding to residues 1–10.

Endoproteinase Lys-C was used to digest 10 nmol of aminoethylated P2, and the fragments produced were purified by reverse-phase HPLC to yield nine fractions labeled K1–K9 (Figure 5B). All of each fraction, except for K3, was used for sequencing on the gas-phase sequencer. A single complete sequence was obtained in each case (Figure 6A). Fraction K3 was suspected as being the C-terminal fragment because amino acid analyses from one of the preliminary digests had indicated that this fragment probably contained neither lysine nor (aminoethyl)cysteine. One-fifth of fraction K3 when subjected to amino acid analyses gave a definitive result (0.9 nmol of histidine, 1.0 nmol of arginine). The remainder of fraction K3 was used for sequencing and produced the sequence R-R-H-H (Figure 6; Table II), which was placed tentatively at the C-terminus of the P2 sequence. Fraction K4 gave a new seven-residue sequence, S-R-R-H-(R)-(R)-K. This was placed tentatively between residue 53 and the K3 fragment to yield a total of 64 amino acid residues for P2. However, this latter assignment proved to be incorrect based on the subsequent sequencing of S1 and S2 peptides generated by chemical cleavage (see below).

The sequence of P2 was completed by using a chemical method to cleave at the N-terminal side of serine. Ten na-

Table II: Sequencing Yields for the P2 C-Terminal Fragments S2 and K3[a]

peptide S2			peptide K3		
residue	amino acid	yield (pmol)	residue	amino acid	yield (pmol)
1	S	1061	1	R	352
2	R	1181	2	R	416
3	R	1175	3	H	205H, 135R
4	R	1076	4	H	214H
5	R	1054	5		62H[b]
6	R	1072			
7	C	1011			
8	R	717			
9	C	638			
10	R	652			
11	K	579			
12	C	536			
13	R	447			
14	R	275			
15	H	224H, 148R			
16	H	200H			
17		88H[b]			

[a] Yields calculated with respect to the internal standard PTH-norleucine. [b] Residual histidine represents carryover from the previous degradation cycle.

nomoles of aminoethylated P2 was cleaved, and the resulting fragments were fractionated by using reverse-phase HPLC (Figure 5C). Eight pooled fractions were taken, and one-fifth of each was used for amino acid analyses. Two fractions, S1 and S2 (Figure 5C), had higher contents of lysine + (aminoethyl)cysteine (Table II) than expected for any region of the firmly established sequence (residues 1–53). Therefore, the peptides in both fractions were sequenced completely. Fraction S1 contained two sequences. The major sequence corresponded to residues 33–46, while the minor fragment, obtained by subtraction of this known sequence, contained 16 residues starting at position 47, and ended with peptide K3 at its C-terminal end. The minor fragment of S1 included a new five-residue sequence instead of the seven-residue sequence of peptide K4. Fraction S2 gave a single sequence that confirmed the minor sequence of fraction S1 corresponding to residues 47–62 (Figure 6A). This completed the primary sequence of P2.

The sequence analyses gave a molecular weight of 8841 for P2, which compares favorably with the estimated molecular weight of 8800 obtained by gel filtration chromatography on agarose 1.5M in the presence of 6.0 M Gdn·HCl.

Sequence Homology of Mouse and Human P2. A computer analysis based on the ALIGN program of Dayhoff et al. (1983) was used to obtain optimal alignments of various pairs of the mouse and human protamine sequences. Homologies of these optimal alignments were assessed by both omitting and including the contribution of the amino acid compositions to the sequence. An optimal alignment of mouse P2 and human P2b demonstrated that these proteins are highly homologous (Figure 6B). Quantitative homology estimates for this sequence comparison plus those for three other sequence pairs are shown in Table III. The similarity values, which measure the number of exact amino acid matches and omit the contribution of the amino acid compositions to the degree of homology, demonstrated that all pairs have considerable homology and that the P1 and P2 sequence pairs are more homologous than are the P1 and P2 pairs in the same species. By contrast, the alignment scores, which eliminate the contribution of the amino acid compositions to the sequence homologies, indicated a much lower degree of homology among all sequence pairs. These latter values may be questioned, however, because the significant deviations of the protamine amino acid compositions from those of average proteins make the applicability of the method uncertain (Dayhoff et al., 1983).

Discussion

Two mouse protamines, P1 and P2, have been purified and characterized by relative molecular weights and amino acid compositions. Moreover, the primary structure of the predominant P2 species has been sequenced to completion and shown to contain six clusters of arginine and histidine distributed throughout the protein. Purification of P1 and P2 was facilitated by the prior isolation of pure populations (99%) of sperm nuclei, free of acrosomal and tail elements (Bellvé et al., 1975; O'Brien & Bellvé, 1980a). The nuclei, being free of extraneous protein, could be processed directly to yield an enriched preparation of the native protamines. Also, the sequence of Bio-Rex-70 ion-exchange chromatography and Bio-Gel-P10 permeation chromatography, both undertaken in the presence of Gdn·HCl, consistently gave high yield and resolution in isolating the P1 and P2 species. The possible existence of more than two forms of P2 in mouse sperm nuclei, as is the case for human sperm (McKay et al., 1986; Ammer et al., 1986), was negated by three criteria. Only the two mouse proteins, P1 and P2, were present on the basis of ion-exchange chromatography applied over a pH range from 5.5 to 6.8, on peptide homogeneity by high-resolution acetic acid–urea/PAGE at pH 2.7, 3.0, and 4.5 and on the subsequent primary sequencing of the P2 species. The apparent heterogeneity encountered on the initial N-terminal sequencing of mouse P2 proved to be due to a contaminating 10 amino acid fragment (see below).

Determination of the molecular weights of mouse P1 and P2 by permeation chromatography on agarose 1.5M in the presence of 6.0 M Gdn·HCl, as developed by Fish et al. (1969), proved to be a highly accurate method for determining the relative mass of these unusual basic proteins. The combination of prior alkylation of all cysteinyl groups and the presence of Gdn·HCl ensured that the proteins behaved as linear random coils (Tanford, 1968). Further, the direct relationship between the molecular radii of randomly coiled proteins and the number of constituent amino acids (M_r) provided a basis for the Porath plot for treatment of flexible polymers (Porath, 1963). This procedure, combined with cochromatography of isotopically labeled protamines, provided an accurate method for quantitating the molecular weights of P1 and P2. The protocol therefore gave reliable estimates of the molecular weights of mouse protamines and circumvented the problem associated with the insolubility of these proteins in SDS, as required for standard SDS–PAGE analyses. The molecular weight estimates for mouse P1 and P2 of 7500 and 8800, respectively, compare favorably with a molecular weight of 6810 predicted for P1 by the cDNA sequence of Kleene et al. (1985) and a molecular weight of 8841 for P2 calculated directly from the protein sequence.

The present study has focused on the peptide sequencing of mouse P2 to provide a valuable comparison for the known human P2 sequences, P2a and P2b (McKay et al., 1986; Ammer et al., 1986). Generally, mammalian protamines contain less than 63 amino acid residues and have unblocked N-termini. On this basis, N-terminal sequencing on an intact protein can often yield a substantial portion of the sequence and reveal the best fragmentation method (human P1; McKay et al., 1985). However, when this strategy was applied to mouse P2, the data were confounded by the presence of more than one sequence in the P2 preparation. Consequently, to

Table III: Sequence Alignment Data for Mouse and Human Protamines

protocol	protamines			
	MP2-HP2b	MP1-HP1	MP1-MP2	HP1-HP2b
similarity scores[a] (%)	75.5	68.8	56.0	56.3
gaps[b]	3	3	8	5
alignment scores[c]	4.7	4.4	1.0	3.1
probabilities[d]	3.4E-6	3.2E-5	1.6E-1	1.4E-3

[a] Similarity: 100[matches/(matches + mismatches)]; an amino acid versus a gap was not counted as a match or as a mismatch; the percent similarity increases with the number of gaps allowed. [b] Gaps: total number of gaps in both sequences. [c] Alignment score: standard deviations from the mean of the scores of 30 randomized pairs of the two sequences. [d] Probability: probability of alignment occurring by chance.

enhance resolution, a two-enzyme protocol was applied by using thermolysin and endoproteinase Lys-C. This strategy previously had led to the successful sequencing of human P2a and P2b. However, when used for sequencing mouse P2, the protocol did not generate suitable fragments for determining the primary sequence beyond residue 53. Fortunately, the less conventional method of chemical cleavage at the serine residues, involving an acyl N → O shift followed by ester hydrolysis, proved successful. Two of the fractions from this cleavage method were sequenced, and the amino acid analyses of the other fractions confirmed that the fragmentation had proceeded as predicted (Iwai & Ando, 1967; Iwai et al., 1971). All residues in mouse P2 were sequenced in at least two different fragments except for residues 54–58, which were sequenced twice in the same fragment (residues 47–62).

The peptide sequence derived for mouse P2 is consistent with that predicted from a recent cDNA sequence analyses of two overlapping clones of a larger 106 amino acid precursor molecule (Yelick et al., 1987). The amino-terminal sequence of P2 obtained from primary sequencing confirms the final cleavage site of the precursor molecule as occurring between the -H–R- bond at positions 43 and 44 of the larger precursor protein, as reasoned by Yelick et al. (1987). Moreover, the remainder of the P2 primary structure is in agreement with that predicted from the cDNA sequence, except for the carboxy terminus. On the basis of direct peptide sequencing, mouse P2 is composed of 62 amino acids with a -C-R-R-H-H carboxy terminus (Figure 6A), rather than the -C-R-R-H-H-H terminating at residue 63 (Figure 6C), as predicted by cDNA analysis (Yelick et al., 1987). Three independent experimental results are consistent with histidine-62 being the C-terminal residue of mouse P2. The amino acid analysis of fragment K3 gave a definitive His:Arg ratio of 0.9:1.0. The sequencing of peptide K3 gave an unambiguous R-R-H-H peptide, and the sequencing of peptide S2 gave the same unequivocal sequence at its C-terminus (Table II). By comparison, human P2, although recovered in the P2a and P2b forms of 57 and 54 amino acids, respectively, shares a common carboxy-terminal sequence of -C-R-R-H (McKay et al., 1986; Ammer et al., 1986).

The seven-residue K4 peptide does not belong in the P2 sequence. Reexamination of the data from the initial N-terminal sequencing experiment had indicated that the K4 fragment came from the 10-residue peptide, S-(R)-K-S-R-R-H-(R)-(R)-K, that was present as a contaminant in the purified P2 preparation. The presence of this peptide plus the repetition of residues in the P2 sequence had caused the equivocal results during the N-terminal sequencing experiment. It was not due to the presence of multiple forms of mouse P2 as had been found during the sequencing of human P2 (McKay et al., 1986; Ammer et al., 1986). The existence of the K4 peptide in the P2 preparation was surprising considering that the final purification step involved gel permeation chromatography in the presence of 50 mM DDT and 2.5 M Gdn·HCl. Moreover, this peptide was not detected during acetic acid–urea/PAGE, either because it migrated with the dye front or because it comigrated with P2 during this procedure as well. Regardless, its presence made the sequencing of P2 more difficult and influenced the strategy employed.

Assessing the degree of sequence homology between the mouse and human protamines presented difficulties. The simplest method of measuring the number of exact amino acid matches in optimal alignments indicated considerable homology exists among all four protamines, as reflected in the similarity scores shown in Table III. A problem with this method is that the considerable arginine contents of the protamines make contributions toward the estimates of homology. The standard method for coping with this difficulty is to express the alignment score in terms of the standard deviation of the alignment scores of a large number of randomized pairs of the two sequences and then to calculate the probability of the alignment occurring by chance. However, this method becomes less applicable as the amino acid compositions deviate significantly from the average protein amino acid composition (Dayhoff et al., 1983), as do protamines with their substantial arginine and histidine content and complete absence of several amino acids. Consequently, for these particular proteins, the calculated alignment scores and their associated probability of occurrence by chance are difficult to interpret (Table III). This leaves the similarity scores as the best means for estimating the sequence homology of these proteins. In this regard, the high values of the similarity scores are consistent with the concept that the P1 and P2 sequences can be assigned to two separate families and that the two families are sufficiently related as to belong to a common superfamily.

The amino acid sequences of mouse P2 and human P2 are more homologous to each other than to the corresponding P1 sequences of the two species (Table III). This interspecies P2 homology is not consistent with the hypothesis of Calvin (1976), in which it is contended that the type 2 protamines arose independently by duplication of the P1 gene, followed then by subsequent divergent evolution. However, the close P2 homology does provide additional support for the hypothesis of McKay et al. (1986) that all eutherian mammals have inherited separate genes for P1 and P2 from a *common ancestral gene* and that many species and/or orders subsequently have lost the ability to synthesize a functional P2 protein. This latter hypothesis predicts that the P1 and P2 sequences within a species should be homologous, but less so than either the P1 or the P2 sequences between species. The mouse P2 sequence and the data in Table II support this prediction.

It is feasible to estimate the rate of acceptance of amino acid substitutions for the mammalian protamines, given the differences between pairs of sequences in the same family and the amino acid PAM (accepted point mutations per 100 amino acid residues) scale of Dayhoff (1978b), and the time of mammalian radiation of 75 million years ago. By this method, the sequence differences between human P1 (McKay et al., 1986; Ammer et al., 1986) and the individual P1's of bull [Coelingh et al., 1972; Mazrimus et al., 1986; Krawetz et al., 1987 (cDNA based)], ram (Sautiere et al., 1984), boar (Tobita et al., 1983), and mouse [Kleene et al., 1985 (cDNA based)] yield a rate of mutation acceptance of 42.6 ± 6.6 PAMs per 100 million years. This rate of mutation acceptance ranks the P1 family among the most rapidly evolving of proteins. By

comparison, the usual rates of mutation acceptance of 60 other protein families have ranged from 0.0 to 37 PAMs per 100 million years with a mean value of 10.7 (Dayhoff, 1978a). The mutation rate for P2 was calculated to be 19.4 PAMs per 100 million years based on the alignment shown in Figure 6B. This estimate probably is low because insertions, deletions, and length differences were not taken into account in the calculations. Regardless, P2 still is evolving at approximately twice the average rate of most other proteins. The evolutionary distance of the P1 and P2 families was estimated at 65 PAMs by using the MP1-MP2 and HP1-HP2b similarity scores from Table III and the amino acid PAM scale. If the average of the rates of mutation acceptance of P1 and P2 (31 PAMs per 100 million years) is used, then the divergence time for the P1 and P2 genes is estimated to be 104 million years ago, i.e., from a common ancestral protamine gene of all eutherian mammals, as predicted previously (McKay et al., 1986).

Mammalian protamines display a remarkable amount of variability on several different levels. All mammals investigated, including 10 species within 5 orders, express P1. However, spermatozoa of only two species—mouse (Bellvé et al., 1975; Calvin, 1976; Balhorn et al., 1977; Bellvé & Carraway, 1978) and man (Kolk & Samuel, 1975; Gaastra et al., 1978)—have been shown to contain P2. The 5 known P1 sequences all contain 50 amino acid residues, suggesting some restrictions on length, but these proteins possess one of the highest known rates of mutation acceptance. By contrast, although P2 is accepting mutations more slowly than P1, both its length and the number of mature forms can vary. Mouse P2 occurs in 1 form containing 62 amino acid residues, whereas human P2 exists in 2 forms containing 54 and 57 amino acid residues (McKay et al., 1986; Ammer et al., 1986). The most unusual type of P2 variability is its complete absence from the mature spermatozoa of most other mammals. Yet, this range of protamine variability has not had an obvious effect on nuclear conformation or on the relative viability of mature spermatozoa in these different mammalian species.

The considerable amount of variability observed for protamines is in direct contrast to the very limited changes observed among the somatic histones. According to Dayhoff (1978a), the rates of acceptance of amino acid mutations in H2a, H2b, H3, and H4 are each <0.1 PAM per 100 million years, thereby placing histones among the most slowly evolving of proteins. Clearly, the structural requirements for protamines in complexing with DNA must be considerably less stringent than those for somatic histones.

ACKNOWLEDGMENTS

We thank Diane Renz for typing the manuscript.

REFERENCES

Ammer, H., Henschen, A., & Lee, C.-H. (1986) *Biol. Chem. Hoppe-Seyler 367*, 515–522.

Balhorn, R., Gledhill, B. L., & Wyrobek, A. J. (1977) *Biochemistry 16*, 4074–4080.

Balhorn, R., Weston, S., Thomas, C., & Wyrobek, A. J. (1984) *Exp. Cell Res. 150*, 298–308.

Bedford, J. M., & Calvin, H. I. (1974) *J. Exp. Zool. 188*, 137–156.

Bellvé, A. R. (1979) in *Oxford Reviews of Reproductive Biology* (Finn, C. A., Ed.) Vol. 1, pp 159–261, Oxford University Press, Oxford and New York.

Bellvé, A. R., & Carraway, T. (1978) *J. Cell Biol. 79*, 177a.

Bellvé, A. R., & O'Brien, D. A. (1983) in *Mechanisms and Control of Animal Fertilization* (Hartman, J. F., Ed.) pp 55–137, Academic, New York.

Bellvé, A. R., Anderson, E., & Hanley-Bowdoin, L. (1975) *Dev. Biol. 47*, 349–365.

Calvin, H. I. (1976) *Biochim. Biophys. Acta 434*, 377–389.

Coelingh, J. P., Rozijn, T. H., & Monfoort, C. H. (1969) *Biochim. Biophys. Acta 188*, 353–356.

Coelingh, J. P., Monfoort, C. H., Rozijn, T. H., Leuvin, J. A. G., Schiphoff, R., Steyn-Parvé, E. P., Braunitzer, G., Schrank, B., & Ruhfus, A. (1972) *Biochim. Biophys. Acta 285*, 1–14.

Dayhoff, M. O. (1978a) in *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, p 3, National Biomedical Research Foundation, Washington, DC.

Dayhoff, M. O. (1978b) *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, p 375, National Biomedical Research Foundation, Washington, DC.

Dayhoff, M. O., Barker, W. C., & Hunt, L. C. (1983) *Methods Enzymol. 91*, 524–545.

Fish, W. W., Mann, K. G., & Tanford, C. (1969) *J. Biol. Chem. 244*, 4989–4994.

Gaastra, W., Hofstra, J. L., & Kolk, A. H. J. (1978) *Biochem. Genet. 16*, 525–529.

Grimes, S. R., Jr., Meistrich, M. L., Platz, R. D., & Hnilica, L. S. (1977) *Exp. Cell. Res. 110*, 31–39.

Hunkapillar, M. W., Hewick, R. M., Dreyer, W. J., & Hood, L. E. (1983) *Methods Enzymol. 91*, 399–413.

Iwai, K., & Ando, T. (1967) *Methods Enzymol. 11*, 263–282.

Iwai, K., Nakahara, C., & Ando, T. (1971) *J. Biochem. (Tokyo) 69*, 493–509.

Kleene, K. C., Distel, R. J., & Hecht, N. B. (1985) *Biochemistry 24*, 719–722.

Kolk, A. H. J., & Samuel, T. (1975) *Biochim. Biophys. Acta 393*, 307–319.

Krawetz, S. A., Conner, W., & Dixon, G. H. (1987) *DNA 6*, 47–57.

Lam, D. M. K., & Bruce, W. R. (1971) *J. Cell. Physiol. 78*, 13–24.

Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem. 255*, 11199–11203.

Mazrimas, J. A., Corzett, M., Campos, C., & Balhorn, R. (1986) *Biochim. Biophys. Acta 872*, 11–15.

McKay, D. J., Renaux, B. S., & Dixon, G. H. (1985) *Biosci. Rep. 5*, 383–391.

McKay, D. J., Renaux, B. S., & Dixon, G. H. (1986) *Eur. J. Biochem. 156*, 5–8.

O'Brien, D. A., & Bellvé, A. R. (1980a) *Dev. Biol. 75*, 386–404.

O'Brien, D. A., & Bellvé, A. R. (1980b) *Dev. Biol. 75*, 405–418.

Panyim, S., & Chalkley, R. (1969) *Biochemistry 8*, 3972–3979.

Porath, J. (1963) *Pure Appl. Chem. 6*, 233–244.

Romrell, L. J., Bellvé, A. R., & Fawcett, D. W. (1976) *Dev. Biol. 49*, 119–131.

Sautiere, P., Belaiche, D., Martinage, A., & Loir, M. (1984) *Eur. J. Biochem. 144*, 121–125.

Steers, E., Jr., Craven, G. R., & Anfinsen, C. B. (1965) *J. Biol. Chem. 240*, 2478–2484.

Tanford, C. (1968) *Adv. Protein Chem. 23*, 121–282.

Tobita, T., Tsutsumi, H., Kato, A., Suzuki, H., Nomoto, M., Nakano, M., & Ando, T. (1983) *Biochim. Biophys. Acta 744*, 141–146.

Yelick, P. C., Balhorn, R., Johnson, P. A., Corzett, M., Mazrimas, J. A., Kleene, K. C., & Hecht, N. B. (1987) *Mol. Cell. Biol. 7*, 2173–2179.