Quail (Coturnix japonica) Protamine, Full-length cDNA Sequence, and the Function and Evolution of Vertebrate Protamines*

(Received for publication, July 11, 1989)
Rafael Oliva, Robyn Goren, and Gordon H. Dixon
From the Department of Medical Biochemistry, University of Calgary, Calgary, Alberta T2N 4N1, Canada

Using the chicken protamine gene as a probe, we have isolated and sequenced several positive clones from a quail testis cDNA library which reveal the complete sequence for the quail protamine cDNA. The predicted amino acid sequence for the quail protamine contains the N-terminal tetrapeptide ARYR present in the N-terminal region of the mammalian protamines as well as several conserved motifs and arginine clusters. In addition the size of the quail protamine (56 amino acids) is closer to that of mammals (50 amino acids) than that of the chicken (61 amino acids). Altogether this data strongly suggests the existence of an avian-mammalian protamine gene line during evolution. Southern blot analysis suggests a small number of copies (2) per haploid genome (similar to that of chicken). The reported quail protamine cDNA sequence is the second avian protamine for which the amino acid sequence is available so far and provides new insights into vertebrate protamine function and evolution.

Protamines are small highly basic proteins which act by compacting the DNA in the sperm nuclei of many species (Bloch, 1969; Dixon, 1972; Subirana, 1975; Mezquita and Teng, 1977a, 1977b; Oliva et al., 1982; Oliva and Mezquita, 1982, 1986; Christensen and Dixon, 1982; Oliva et al., 1987). Because of the great variability of sequence in these proteins, the determination of the structure of the nucleoprotamine, the pathways of vertebrate protamine evolution, and the function of protamines have remained elusive (Coelingh et al., 1972; Dixon et al., 1975; Warrant and Kim, 1978; Mezquita, 1986a, 1985b; Poccia, 1986; Risley, 1988). It is, therefore, necessary to compare closely related sequences or species in order to answer many of these questions (Dixon et al., 1985; Kasinsky et al., 1987; Kasinsky, 1989). In this paper, the determination of the sequence of quail protamine cDNA coupled with knowledge of the domestic rooster protamine gene suggests the existence of an avian-mammalian protamine gene line in evolution.

* This work was supported by a term operating grant from the Medical Research Council of Canada (to G. H. D.) and an Alberta Heritage Foundation for Medical Research Post-Doctoral Fellowship (to R. O.). 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.

MATERIALS AND METHODS

Quails (Coturnix japonica) were obtained from Bailey's Game Birds, Alberta, Canada. RNA and DNA were isolated as described by Oliva and Dixon (1989). Poly(A)+ RNA was selected as described by Aviv and Leder (1972). cDNA was prepared as described by Okayama and Berg (1982). Transformations were performed by electroporation (Fiedler and Wirth, 1988; Taketo, 1988) using Bio-Rad equipment. Primary and secondary screenings, template DNA preparation, sequencing, double-stranded plasmid preparation, exonuclease III/mung bean nuclease directional deletions, and Southern blots were performed as described by Oliva and Dixon (1989). Northern blots were performed as described by Oliva et al. (1988). Dot matrix analysis was performed using the Seqaid program (Rodas and Roufa, 1987), and protamine gene and amino acid alignments were performed with the aid of the SEQALIGN and GENALIGN (Needleman-Wunsch) programs from BIONET (1989).

RESULTS

Northern blot analysis of quail testes RNA using the chicken protamine cDNA probe revealed a sharp band in the 0.4 kilobase range (Fig. 1). This was the basis for quail testis cDNA library construction, screening with the chicken protamine cDNA probe, and sequencing the positive clones to obtain the quail protamine cDNA sequence. One colony out of every 240 turned out to be positive. Three independent positive clones were sequenced, and their sequence revealed the complete cDNA sequence for the quail protamine (Fig. 2). The identification of the sequenced cDNAs as corresponding to the quail protamine is based on: 1) the very high homology of these clones with the chicken protamine cDNA and amino acid sequence (Figs. 3–5); 2) the known properties of protamines (Bloch, 1969; Subirana, 1982); and 3) the consistency of the expected size for the quail protamine, as determined on acid gels (Chiva et al., 1987, 1988), with the size predicted from the cDNA (Fig. 2). The expression of this gene is clearly

---

\[ \text{FIG. 1. Northern blot analysis of testis RNA from different species using a chicken protamine probe.} \]
Quail cDNA Sequence

GGCGTCCAC CCACCGGCGG C -1
ATG GCC CCC TAC CCG GCG ACG AGG ACC CGC GGC CCC CCC GCG TCC CCC +60
R A R Y R R T R T R S R S R S R S R S R

GGC CCC CCC ACC TCC CGA CCC CCC CCC TAC GGG AGG TCG CCC CCC TCC TAC CCC TCC TCG OTC +120
G R R R S S R Y G R R R R R Y R Y Y Y

GGG AGA CGG CCC CCC CCC TAC GGG AGA AGA CCC CCC CCC CCC CCC CCC CCC TAC TGG +174
G R R R S S R Y G R R R R R Y R Y Y Y

GGGCACCG ACAAGGGCAC CGGCCGCGAC ACCGGCACCG CCACCGCCAC CCACGCACCC CCGGCCACCC +244

GGGCGGGCG GCTGCGGCTC CGGCAATCTAA CTGGCCTGGC ACCGGAAAA AAAA

FIG. 3. Dot matrix analysis of all the protamine gene sequences presently available. 20 matches or more in a 32-nucleotide window are indicated as a dot. The arrow at the x or y axis indicates the scale size of the transcriptional unit. The shaded area indicates the coding region. The positions of known introns have not been included in the comparison. The quail protamine sequence is from Fig. 5, in which the homologous nature of this group of protamines is apparent.

FIG. 4. Detailed alignments of the protamine cDNA sequence from quail with that from chicken and their corresponding predicted amino acid sequences. Nucleotide positions are indicated for the quail cDNA using the ATG start codon as a reference. Neutral mismatches are indicated by an asterisk, whereas the changes leading to a different amino acid sequence are indicated by a black dot. a and b designate two alternative alignments for the region shown in brackets. Amino acid differences between quail and chicken are underlined.

Southern blot analysis indicates a reduced complexity of the quail protamine genes, comparable to the chicken genes (Oliva and Dixon, 1980) and suggests a low copy number, perhaps of two, based on the existence of two bands on the Southern blots (Fig. 6) and on a preliminary intensity comparison with known standards (not shown).

FIG. 2. Quail protamine cDNA sequence. Translation start codon (ATG), termination codon (TGA), polyadenylation signal (AATAAA), and poly(A) tail site are underlined. The differences between the chicken and quail sequences responding to the chicken (Fig. 6).

Established from its poly(A) mRNA origin from quail testis. The differences between the chicken and quail sequences (Figs. 4 and 5) rule out the possibility of artifactual (by contamination) subcloning and resequencing the cDNA corresponding to the chicken (Fig. 6).

Discussion

We have cloned and sequenced the cDNA corresponding to the quail protamine, the second avian protamine for which
the nucleotide and amino acid sequence is now known (Fig. 2). The predicted amino acid sequence for the quail protamine shows 11 differences and is shorter by 5 amino acids than chicken protamine (Figs. 2, 4, and 5), resulting in a total length of 56 amino acids. The homologous relationship between the quail and the chicken protamine is even more marked when the corresponding nucleotide sequences are compared (Figs. 3 and 4). These similarities between chicken and quail protamines clarify to some extent the origin of the differences between the amino acid sequence for the chicken protamine determined by Nakano et al. (1976) and the predicted amino acid sequence from the genome or the reetermined amino acid sequence for galline (Oliva and Dixon, 1989). These differences raised the question of whether different protamine allelic variants existed among chickens, with the possibility that Nakano et al. (1976) had sequenced one of those, and Oliva and Dixon (1989) a different one. The fact that the unusual amino acids (such as threonine and valine) are present in the same position in the quail and chicken protamines (two different avian species) as well as in many mammals (Fig. 5), only if the recently reetermined sequence (Oliva and Dixon, 1989) is considered, makes the hypothesis of the differences in sequence corresponding to different chicken allelic variants much more unlikely.

Dot matrix analysis of all protamine gene sequences available so far clearly indicates their relationships (Fig. 3). This is seen as a diagonal succession of dots at the intersection of the corresponding genes. The similarities between the two avian protamine genes, among the four mammalian P1 protamine family genes, and within the two mammalian P2 family genes are clearly apparent (Fig. 3). Although less clear at this level of nucleotide comparison, the avian protamine genes also display significant similarity (in their coding regions) to the mammalian P1 protamine family genes (indicated as several dots in the dot matrix corresponding to the coding regions). The similarities among avian and mammalian P1 protamines are also clearly evident by a comparison of their amino acid sequences (Fig. 5). Such similarities are sufficient to postulate a homologous relationship and hence a common avian-mammalian protamine gene line during evolution. Particularly conserved are the N-terminal tetrapeptides ARYR and SRSR followed by cluster of 5-7 arginines (Fig. 5). These sequences are also present in the marsupials opossum and wallaby (Balhorn et al., 1989). In addition, in this region at position 8, either a serine or a theonine occurs, both polar uncharged amino acids being susceptible to phosphorylation. The similarities at the C terminus are less marked than at the N terminus (Fig. 5); nevertheless, the arginine clustering, the presence of the valine at position 44 (allowing for gaps in the alignment), the tyrosines at position 52, and the C-terminal tyrosine follow a clearly conserved pattern.

Among the differences between the avian and mammalian protamines, the main one is the absence of cysteines in bird protamines (Fig. 5). This implies that either cysteines were lost in birds or gained in mammals. Several mechanisms for the appearance or loss of cysteines are possible. If cysteine codons appeared in mammalian protamine genes, a potential mechanism would be their origin by mutation of arginine codons CGT/C to a TGT/C cysteine codon. This pathway might be greatly facilitated as a result of a previous cytosine methylation in the arginine codon, since 5mC is thought to be evolutionarily unstable and tends to mutate to T by deamination at position 5 (Coulondre et al., 1978).

Despite many conserved motifs between mammalian and bird protamines, the presence of several cysteines in mammals leads to an essentially different mechanism for condensation of the nucleoprotamine through the formation of disulfide bonds in the sperm nuclei (Balhorn, 1982; Balhorn et al., 1984). The fact that cysteines are not conserved while other motifs are (such as the N-terminal tetrapeptide ARYR) may indicate a relatively less essential function of the cysteines as compared to the conserved amino acid clusters. Therefore, we think that the consensus sequences between avian and mammalian protamines (Fig. 5) will provide very useful insights.
in understanding the key elements in nucleoprotamine formation, structure, and function (Oliva et al., 1987; Nakano et al., 1988; Tobita et al., 1988). In a similar way, comparisons of the conserved nucleotide sequences at the 3’ or 5’ of the quail and chicken genes should contribute to an understanding of the important regulatory and structural elements controlling proteamine gene expression. For example, there is strong conservation of the 5’ region immediately preceding the initiation codon as well as the 3’ region from position +240 to the polyadenylation site at +290 (only two mismatches). In contrast, the region immediately 3’ to the TGA termination codon (+174 to +240) is much less conserved (13 mismatches as well as 2 deletions in the chicken). It was noted (Oliva et al., 1988; Oliva and Dixon, 1989) that this region in the chicken represented an in-frame duplication of the C-terminal coding region; however, such a region is absent from the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken 5’ region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quail cDNA. One explanation would be that the chicken has undergone a partial duplication of its C-terminal coding region with the possible evolutionary result of a lengthening of the quai...