Nucleotide Sequence of Bovine Prolactin Messenger RNA
EVIDENCE FOR SEQUENCE POLYMORPHISM*

(Received for publication, May 28, 1981)
Nancy L. Sasavage‡, John H. Nilson‡, Sarah Horowitz‡, and Fritz M. Rottman‡
From the Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824

Hybrid molecules containing DNA sequences complementary to bovine pituitary mRNA were constructed in the Pst I site of pBR322 by the dC-dG tailing technique. Recombinant plasmids containing bovine prolactin (bPRL) sequences were amplified in bacteria and identified by hybridization to purified [32P]bPRL cDNA sequences. Nucleotide sequence analysis was performed on the inserts from two of the positive clones. One clone, pBPRL72, contained a 982-base pair insert that included 67 nucleotides of the 5'-untranslated region, the complete coding region of the preprolactin protein (690 nucleotides), and the entire 3'-untranslated region (150 nucleotides) of bPRL mRNA. The nucleotide sequence analysis of clone pBPRL72 predicted the sequence of a 30-amino acid signal peptide and confirmed the published amino acid sequence of the protein with one exception. A comparison of the pBPRL72 cDNA sequence with a second bPRL clone, pBPRL4, revealed four silent nucleotide differences. Three of the base changes occurred in the third position of amino acid codons, and one occurred in the 3'-noncoding region. The sequence polymorphism suggests the existence of alleles or multiple loci for bPRL that do not alter the protein structure.

The hormone prolactin is a major protein product of the anterior pituitary gland of mammals and the prominence of this protein is reflected in the composition of total cellular mRNA. Bovine prolactin mRNA constitutes approximately 60% of polysomal mRNA from the bovine anterior pituitary gland (1) and is readily purified on sucrose density gradients (2). In a previous report, we described the construction and analysis of a cDNA clone containing a portion of the bPRL 5'-untranslated region (3). In order to examine the structure and expression of the bPRL gene, a full length clone was necessary. The present study describes the cloning and sequence analysis of a full length bPRL cDNA. Comparison of several bPRL cDNA clones shows that the nucleotide sequence of bPRL mRNA contains heterogeneities in the coding and noncoding portions of the message. The sequence polymorphisms are silent and therefore do not affect the amino acid sequence of the bPRL protein.

MATERIALS AND METHODS

Materials. DNA-DNMTs (Boehringer Mannheim) were obtained from New England Nuclear. The T7 DNA polymerase (200 U/mg) was from Boehringer Mannheim. All restriction enzymes, polymersases, and nucleotides were from Boehringer Mannihem or Bethesda Research Laboratories. The sequence analysis was performed on the inserts from two of the positive clones. One clone, pBPRL72, contained a 982-base pair insert that included 67 nucleotides of the 5'-untranslated region, the complete coding region of the preprolactin protein (690 nucleotides), and the entire 3'-untranslated region (150 nucleotides) of bPRL mRNA. The nucleotide sequence analysis of clone pBPRL72 predicted the sequence of a 30-amino acid signal peptide and confirmed the published amino acid sequence of the protein with one exception. A comparison of the pBPRL72 cDNA sequence with a second bPRL clone, pBPRL4, revealed four silent nucleotide differences. Three of the base changes occurred in the third position of amino acid codons, and one occurred in the 3'-noncoding region. The sequence polymorphism suggests the existence of alleles or multiple loci for bPRL that do not alter the protein structure.

The hormone prolactin is a major protein product of the anterior pituitary gland of mammals and the prominence of this protein is reflected in the composition of total cellular mRNA. Bovine prolactin mRNA constitutes approximately 60% of polysomal mRNA from the bovine anterior pituitary gland (1) and is readily purified on sucrose density gradients (2). In a previous report, we described the construction and analysis of a cDNA clone containing a portion of the bPRL 5'-untranslated region (3). In order to examine the structure and expression of the bPRL gene, a full length clone was necessary. The present study describes the cloning and sequence analysis of a full length bPRL cDNA. Comparison of several bPRL cDNA clones shows that the nucleotide sequence of bPRL mRNA contains heterogeneities in the coding and noncoding portions of the message. The sequence polymorphisms are silent and therefore do not affect the amino acid sequence of the bPRL protein.

*This work was supported by Public Health Service Research Grant CA 10175 (R. F.) from the National Cancer Institute. This is Michigan Agriculture Experiment Station Journal No. 9981. 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.
‡Present address, Department of Pharmacology, Case Western Reserve University, Cleveland, OH 44106.
§Present address, Department of Microbiology, Case Western Reserve University, Cleveland, OH 44106.

The abbreviations used are: (b, r) bPRL (bovine, rat, human) prolactin; GH, growth hormone; ds-cDNA, double-stranded complement DNA; dNTP, 2',3'-dideoxy nucleoside triphosphate; bp, base pair.

The sequence polymorphism suggests the existence of alleles or multiple loci for bPRL that do not alter the protein structure.

The hormone prolactin is a major protein product of the anterior pituitary gland of mammals and the prominence of this protein is reflected in the composition of total cellular mRNA. Bovine prolactin mRNA constitutes approximately 60% of polysomal mRNA from the bovine anterior pituitary gland (1) and is readily purified on sucrose density gradients (2). In a previous report, we described the construction and analysis of a cDNA clone containing a portion of the bPRL 5'-untranslated region (3). In order to examine the structure and expression of the bPRL gene, a full length clone was necessary. The present study describes the cloning and sequence analysis of a full length bPRL cDNA. Comparison of several bPRL cDNA clones shows that the nucleotide sequence of bPRL mRNA contains heterogeneities in the coding and noncoding portions of the message. The sequence polymorphisms are silent and therefore do not affect the amino acid sequence of the bPRL protein.

*This work was supported by Public Health Service Research Grant CA 10175 (R. F.) from the National Cancer Institute. This is Michigan Agriculture Experiment Station Journal No. 9981. 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.
‡Present address, Department of Pharmacology, Case Western Reserve University, Cleveland, OH 44106.
§Present address, Department of Microbiology, Case Western Reserve University, Cleveland, OH 44106.

The abbreviations used are: (b, r) bPRL (bovine, rat, human) prolactin; GH, growth hormone; ds-cDNA, double-stranded complement DNA; dNTP, 2',3'-dideoxy nucleoside triphosphate; bp, base pair.
RESULTS AND DISCUSSION

Detection of Clones with Large Prolactin Inserts—We had earlier reported the cloning and sequence analysis of a 225-base pair insert coding for amino acids 119-192 of the bPRL protein (199 amino acids total) (3). In the present study, mild S(+)-treatment and subsequent sizing of ds-cDNA synthesized from bovine pituitary poly(A)-containing RNA were employed in order to obtain full length PRL sequences. A total of 165 PRL-positive colonies were identified by hybridization to purified [3P]bPRL cDNA from approximately 600 transformants. Recombinant plasmids from selected colonies were prepared by the minipreparation technique of Birnboim and Doly (7), digested with Psi I and electrophoresed in agarose gels to determine the size of the PRL inserts. The majority of the clones contained inserts of approximately 600 bp (data not shown). Two of our upstream plasmid sequences, insert lengths of approximately 750 and 950 bp were selected for sequence analysis. These plasmids were designated pBPRL4 and pBPRL72 (plasmid bovine prolactin, 4th and 2nd colonies), respectively.

Sequence Analysis of Prolactin Inserts—The insert from pBPRL72 was excised with Psi I and separated on a preparative polyacrylamide gel. The Psi I digestion resulted in a single insert fragment indicating the absence of a Psi I site in the bPRL sequence. The purified insert was analyzed with a series of restriction enzymes. Digestion with the enzymes Alu I, Dde I, Hae III, Hinf I, and Hpa II resulted in fragments (data not shown) of appropriate size (20-150 bp) for use as primers in the dideoxy DNA sequencing method (8,9). Several of these fragments were employed for DNA sequence analysis of the inserts from pBPRL4 and pBPRL72 according to Smith (9). The chemical method for DNA sequencing (12) was used to determine the length of the homopolymer tails resulting from the dG-dC tailing technique. Fig. 1 shows a summary of the sequencing strategy.

The 982-bp insert sequence of the larger clone, pBPRL72, was determined using both methods for DNA sequencing (Fig. 2). Most areas of the insert were sequenced two times and for 70% of the insert length both complementary strands were analyzed (Fig. 1). The cloned insert contained a total of 907 nucleotides corresponding to the bPRL mRNA sequence as well as 21 A residues from the poly(A) portion of the message. The homopolymer tails consisted of 26 and 28 bases at the 5' and 3' ends of the insert, respectively. The amino acid sequence predicted from one reading frame (Thr, +1, Fig. 2) agrees with that of the bovine prolactin protein (13) with one minor exception. The DNA sequence predicts an aspartic acid instead of an asparagine at amino acid position 31. Confirmation of the assignment was obtained from both DNA strands in this region and therefore suggests an alternate bPRL protein sequence. Assignments from the DNA sequence at the indeterminate positions (Glx and Asx) of the amino acid sequence show glutamic acid for amino acids 70, 78, 119, 120, 121, 122, 143, and 145, glutamine for 71, 73, and 74, asparagine for 10 and 92, and aspartic acid for 93.

The clone pBPRL72 also contained the entire sequence corresponding to the signal peptide of the bPRL protein. Lingappa et al. (14) reported the length of this signal sequence to be 30 amino acids, but the complete amino acid sequence was not determined. The nucleotide sequence of pBPRL72 also predicts a signal peptide of this length. A single AUG codon (Met, +30, Fig. 2) is found in the sequence preceding the codon for the NH2-terminal threonine of the authentic bovine prolactin protein. In addition, the location of eight leucine codons in the hydrophobic center of the signal sequence reported by Doolan and Blobel (15) are found in the DNA sequence of the pBPRL72 insert. The amino acid sequence deduced from the reading frame following the AUG codon represents the first complete protein sequence for the signal peptide of bPRL. This signal peptide sequence is similar in nature to other known signal peptides and is the longest known signal sequence of a secreted protein (15).

The start of the signal peptide sequence in pBPRL72 is preceded by 67 nucleotides corresponding to the 5'-untranslated portion of bPRL mRNA. By electrophoresis in denaturing gels, we have previously estimated the size of bPRL cytoplasmic mRNA to be approximately 1900 nucleotides, including the 3'-poly(A) segment (1). If the poly(A) region of bPRL mRNA is 190-150 nucleotides long, the 907-bp sequence of pBPRL72 may represent a nearly complete copy of the mRNA. Furthermore, electron microscopic analysis of cytoplasmic bPRL mRNA and pBPRL72 duplexes does not show an unhybridized 5'-terminal extension of the mRNA. This observation indicates that any additional sequence for this region of the mRNA is likely to be less than 50 nucleotides. Together these data indicate that the 5'-noncoding region of bPRL mRNA represented in the cloned pBPRL72 insert is nearly complete.

The complete 3'-noncoding portion of bPRL mRNA is also included in clone pBPRL72. The appearance of a poly(A) sequence in this clone establishes the exact length of the 3'-untranslated region to be 150 nucleotides. The common AUAAA sequence found in the 2'-noncoding regions of eucaryotic mRNAs is located 28 bases upstream of the poly(A) addition site. This sequence has been postulated to play a role in the addition of poly(A) segments to mRNAs and generally occurs approximately 20 bases prior to the poly(A) junction (16).

Comparison of Cloned Prolactin Sequences—Previous analysis of two partial bPRL cDNA clones from our laboratory (3) and another laboratory (17) has revealed several nucleotide differences between the cloned sequences. Sequence data from two laboratories on the rat PRL mRNA sequence (18, 19) also suggests that nucleotide sequence heterogeneities occur in cDNA clones obtained from different inbred strains of rats. Therefore, we decided to examine another PRL cDNA clone for potential sequence polymorphism. The present analysis further documents the extent of such nucleotide heterogeneities in the bPRL mRNA. Clone pBPRL4, which contains an...
Bovine Prolactin Sequence

Fig. 2. Nucleotide sequence and predicted amino acid sequence of the mRNA coding for bPRL. The amino acid sequence of bPRL is shown with the mRNA sequence deduced from clone 680 Bovine Prolactin Sequence. The amino acid sequence is inferred from the mRNA sequence, and the significance of this trend is unknown at this time.

The changes are summarized in Table I. In the third position of codons, it seems unlikely that the differences result from random errors by reverse transcriptase or random mutations in the bPRL gene. The origin of the sequence heterogeneities in these bPRL cDNA clones may result from several possible sources. Although errors in the initial copying of the mRNA template by reverse transcriptase have not proven to be a significant problem in cDNA cloning, the possibility of introducing random base changes is feasible. Furthermore, growth and amplification of the recombinant plasmids in Escherichia coli may introduce mutations in the cloned sequences. However, since we only detect heterogeneities in the third positions of codons, it seems unlikely that the differences result from random errors by reverse transcriptase or random mutations of the hybrid plasmids in E. coli. A third possibility is that

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Position Number</th>
<th>mRNA</th>
<th>bPRL72</th>
<th>bPRL4</th>
<th>bPRL6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>95</td>
<td>CGG</td>
<td>CGG</td>
<td>CGG</td>
<td></td>
</tr>
<tr>
<td>Valine</td>
<td>103</td>
<td>GGA</td>
<td>GAG</td>
<td>GAG</td>
<td></td>
</tr>
<tr>
<td>Valine</td>
<td>137</td>
<td>GUA</td>
<td>GUA</td>
<td>GUA</td>
<td></td>
</tr>
<tr>
<td>Serine</td>
<td>151</td>
<td>CUC</td>
<td>CUC</td>
<td>n.d.</td>
<td></td>
</tr>
<tr>
<td>Glycine</td>
<td>152</td>
<td>GGU</td>
<td>GGA</td>
<td>n.d.</td>
<td></td>
</tr>
<tr>
<td>Leucine</td>
<td>156</td>
<td>CUC</td>
<td>CUC</td>
<td>n.d.</td>
<td></td>
</tr>
<tr>
<td>Serine</td>
<td>166</td>
<td>GUC</td>
<td>GUC</td>
<td>n.d.</td>
<td></td>
</tr>
</tbody>
</table>

The location and identity of the amino acids from bPRL with differences in the third position of their codons is listed. The total number of amino acids in the protein is 199. The bPRL cDNA clones contain the nucleotide sequence corresponding to the following amino acids: pPRL, Val to Val (17); bPRL72, -30 to 199; bPRL4, 26 to 199; and bPRL6, 99 to 199 (Miller et al., 1985). The codons that have not been sequenced are indicated as n.d.

TABLE I

Summary of Sequence Polymorphisms in Bovine Prolactin mRNA

Met (30) and terminating with the ochre UAA codon, labeled OC. The beginning of processed PRL protein is labeled Thr (+1). The bases marked by * are nucleotides that differ in the insert sequence of clone bPRL 4. The changes are summarized in Table I.
the sequence heterogeneities result from the existence of multiple bPRL mRNA sequences. Multiple sequences may indicate a number of alleles for PRL in the gene pool of cattle or multiple loci within each animal. The poly(A)-containing RNA used to generate the cDNA clones described in this study was obtained from several animals. We have therefore initiated sequencing studies of bPRL mRNA from single animals by extension of DNA restriction fragment primers hybridized to the mRNA template. Preliminary results suggest the presence of two nucleotides at the same position on the sequencing gel corresponding to the third nucleotide of some amino acid codons (data not shown). Animals with this type of sequence polymorphism in the cytoplasmic PRL mRNA would presumably be heterozygous at the PRL gene locus or have duplicated genes. Confirmation of these results will require cDNA cloning from one anterior pituitary gland.

The existence of multiple alleles for bPRL may be further examined at the genomic level. Jeffreys (20) has recently described sequence polymorphism in the human globin genes of normal individuals detected by Southern blot hybridization. All three variant restriction enzyme cleavage sites occurred in the intervening sequence of the genes, similar to the results of Lai et al. (21) with the chicken ovalbumin gene. No variants were detected in the coding regions of the globin genes examined in this study. However, restriction enzyme analysis of this type is limited to sequence changes that occur within specific enzyme recognition sites. Based on Jeffreys' analysis, an average of at least 1 in 100 bp may be expected to vary polymorphically throughout the human genome. Comparison of cloned DNA sequences is more extensive and has allowed us to detect polymorphism in the coding region of bPRL. We may be able to investigate polymorphism in the coding sequence of bPRL at the genomic level due to the occurrence of restriction enzyme sites containing the variant sequences described here, as well as polymorphism of the intervening sequences. Such information may be useful in determining the number of PRL sequences per genome.

**Sequence Homology of Prolactin mRNAs**—The primary structure of bPRL mRNA is also of interest for purposes of comparison with PRL mRNA sequences from other species. Substantial stretches of homology are apparent between the rat (18, 19), human (22), and bovine PRL sequences at both the amino acid and nucleotide level (data not shown). The overall homology for the protein sequences of bPRL and rPRL including the signal peptide is 57%, while the total base homology including the untranslated portions is 68%. Similarly, the protein sequences of bPRL and hPRL contain 73% identical amino acids, while the nucleotide sequences are 80% homologous. The bPRL mRNA sequence reported here contains more extensive sequence information for the 5'- and 3'-noncoding regions of the mRNA than the rPRL or hPRL sequences. Significant homologous stretches are apparent in the untranslated regions of the three sequences by introducing gaps for comparisons. The apparent sequence homology among these mRNAs may provide information on specific mRNA secondary structures that are functionally conserved in evolution.

PRL is structurally related to growth hormone which is also produced in the pituitary (23). In comparison with bGH mRNA, which has a 59% G + C content, bPRL mRNA contains only 49% G + C. This difference is also reflected in the nonrandom codon usage in the translated portions of these two mRNAs. The bPRL message utilizes G or C in 61% of its codons, and the bGH message favors G or C in 82% of its codons. This difference in the G + C content and codon usage of PRL and GH mRNAs is also present in the rat (19) and human (22). The per cent homology between the two bovine sequences also parallels the homology for the rat PRL and GH sequences. Both sets of nucleotide sequences are approximately 38% homologous while the proteins are approximately 22% homologous. The areas of homology for the bovine related sequences are rather short. The longest exact amino acid match is three, and the longest nucleotide overlap is nine. It is difficult to make comparisons of the untranslated regions for these mRNAs where the sequences have apparently diverged extensively.

The availability of the full length bPRL cDNA clone described here will allow us to examine the expression of the hormone PRL in the bovine system. Such studies will be of interest for both comparative and mechanistic purposes.

**Acknowledgments**—We are indebted to Dr. R. Roberts in whose laboratory some of the sequence analysis was performed. We would also like to thank Drs. R. Blumenthal and T. Gingeras for assistance in the sequencing methodology and computer analysis, K. Friderici and R. Woychik for help throughout the course of this work and in the preparation of the manuscript, R. Swift for assistance in preparing our computer programs, Dr. R. Gilmore for discussions concerning the signal peptide sequence, Drs. E. Fritsch and J. Dodgson for critical reading of the manuscript, and S. Uselton for typing the manuscript.

**REFERENCES**

Nucleotide sequence of bovine prolactin messenger RNA. Evidence for sequence polymorphism.
N L Sasavage, J H Nilson, S Horowitz and F M Rottman


Access the most updated version of this article at http://www.jbc.org/content/257/2/678

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 0 references, 0 of which can be accessed free at http://www.jbc.org/content/257/2/678.full.html#ref-list-1