Nucleotide Sequence of Bovine Prolactin Messenger RNA: Evidence for Sequence Polymorphism

Nancy L. Sasavage, John H. Nilson, Sarah Horowitz, and Fritz M. Rottmann

From the Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824

Hybrid molecules containing DNA sequences complementary to bovine pituitary mRNAs 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 [3P]pBR322 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 sequence (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

Sequences were determined by the Maxam and Gilbert (4) and Sanger and co-workers (5) methods. The bPRL sequence data were analyzed by computer using the FlexDNA software program (Bethesda Research Laboratories, Bethesda, MD). The sequence data have been submitted to the DNA Data Bank of Japan (DDBJ) and the EMBL Data Library (EMBL) with the following accession numbers: D84803 (pBPRL4), D84804 (pBPRL72), and D84805 (pBPRL1). The GenBank accession numbers for these sequences are L19537, L19538, and L19539, respectively.

Construction and Analysis of cDNA Clones. Poly(A)-containing RNA from bovine anterior pituitary gland was used as a template for the synthesis of ds-cDNA with reverse transcriptase, as previously described (3). The ds-cDNA was ligated into pBR322 (Pharmacia) and the DNA insert was determined by Pst I digestion of the plasmid. The complete coding region of the preprolactin mRNA was confirmed by hybridization with the 32P-labeled bPRL insert (150 nucleotides) of pBPRL72 containing a Pst I site. The Pst I fragments of pBPRL72 were separated by electrophoresis on 1.5% neutral agarose gels, visualized by staining with ethidium bromide, and excised out of the gel. The Pst I fragments were ligated into pBR322 which had been digested with Pst I. Recombinant plasmids containing bovine prolactin sequences were identified by hybridization with the 32P-labeled bPRL insert. Transformation of Escherichia coli (DH5α) with the recombinant plasmids was performed as described by Sambrook and Russell (6). Transformants were detected by X-gal and IPTG color reactions, and subsequent analysis of restriction endonuclease digestion patterns confirmed the presence of recombinant DNA molecules. The correct orientation of the DNA sequences in the recombinant plasmids was determined by digestion with restriction enzymes. The sequences of the regions containing the bPRL sequences were determined using the Maxam and Gilbert protocol for DNA sequencing. The sequence data were analyzed by computer using the FlexDNA software program (Bethesda Research Laboratories, Bethesda, MD). The sequence data have been submitted to the DNA Data Bank of Japan (DDBJ) and the EMBL Data Library (EMBL) with the following accession numbers: D84803 (pBPRL4), D84804 (pBPRL72), and D84805 (pBPRL1). The GenBank accession numbers for these sequences are L19537, L19538, and L19539, respectively.

The plasmid pBPRL72 was restricted with Pst I and similarly tagged with deoxyguanosine residues, but the reaction was at 37°C for 5 min. Hybridization of the insert DNA and the plasmid was performed at a high-stringency condition of 0.1% sodium dodecyl sulfate (SDS) and 50% formamide at 65°C for 30 min, followed by washing the filter at high-stringency conditions. The probe was a 200-bp Pst I fragment containing the bPRL insert (5). The hybridization was performed at 65°C for 30 min. The probe was excised from the gel, purified by electroelution, and 10 ng of the probe was used for hybridization. The filters were then hybridized at higher stringency conditions, and the hybridization signal was detected by autoradiography. The nucleotide sequence data have been deposited in the DDBJ, EMBL, and GenBank databases.

The sequence data have been submitted to the DNA Data Bank of Japan (DDBJ) and the EMBL Data Library (EMBL) with the following accession numbers: D84803 (pBPRL4), D84804 (pBPRL72), and D84805 (pBPRL1). The GenBank accession numbers for these sequences are L19537, L19538, and L19539, respectively.
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 S1-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 ['"P]bPRL cDNA from approximately 600 transformants. Recombinant plasmids from selected colonies were prepared by the miniscreen technique of Birnboim and Doly (7), digested with Pst 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 assignment plasmids containing inserts of approximately 750 and 590 bp were selected for sequence analysis. These plasmids were designated pBPRL4 and pBPRL72 (plasmid bovine prolactin, 4th and 72nd colonies), respectively.

Sequence Analysis of Prolactin Inserts—The insert from pBPRL72 was excised with Pst I and separated on a preparative polyacrylamide gel. The Pst I digestion resulted in a single insert fragment indicating the absence of a Pst 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, Hind III, 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 this 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 were not determined. The nucleotide sequence of pBPRL72 also predicts a signal peptide of this length. A single AUG codon (Met, +90, 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 Jackson and Blobel (15) is correctly predicted from 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 1000 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 AAUAAA 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

S. Camper, personal communication.
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 pBPRL72. The length corresponding to the coding portion of bovine preprolactin is 690 base pairs, starting from the AUG codon labeled insert of approximately 750 bp, was analyzed simultaneously with pBPRL72. Approximately 50% of pBPRL4 has been sequenced, and it contains the nucleotide sequence corresponding to amino acid number 26 through the poly(A) portion of the mRNA (Fig. 2).

Comparison of the insert DNA sequences from pBPRL72 and pBPRL4 indicates four positions where a change in the nucleotide assignments was obtained from both complementary strands. All the nucleotide substitutions are silent with respect to the amino acid sequence of bPRL. Three of the changes occur in the third position of codons, while the fourth difference occurs in the 3' noncoding region, nucleotides from the poly(A) addition labeled +1. The bases marked by * are nucleotides that differ in the insert sequence of clone pBPRL4. The changes are summarized in Table I.

| Amino Acid | Position | pBPRL4 | pBPRL72 | pBPRL4 | pBPRL72 |
|------------|----------|--------|---------|--------|---------|
| leucine    | 95       | GUC    | GUC     | GUC    | GUC     |
| valine     | 103      | GAG    | GAG     | GAG    | GAG     |
| valine     | 127      | GUG    | GUG     | GUG    | GUG     |
| serine     | 151      | UCU    | UCA     | n.d.   | UCA     |
| glycine    | 152      | GCG    | GCA     | n.d.   | GCA     |
| leucine    | 186      | GUC    | GUC     | n.d.   | UCA     |
| serine     | 186      | GUC    | GUC     | n.d.   | UCA     |

The location and identity of the amino acids from bPRL with differences in the third base of their codons is listed. The total number of amino acids in the protein is 199. The pBPRL4 clone contains the nucleotide sequence corresponding to the following amino acids: pBPRL1 to 121 (17); pBPRL72, 122 to 199; pBPRL4, 26 to 199; and pBPRL4, 99 to 199 (Miller et al., 19). The codons that have not been sequenced are indicated as n.d.

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
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. Novel 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 sequence 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 mRNA 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 for the protein sequences of bPRL 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 26% homologous at the third nucleotide. 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. Fritz and J. Dodgson for critical reading of the manuscript, and S. Ushet for typing the manuscript.

REFERENCES

1. Nilson, J. H., Convey, E. M., and Rottman, F. M. (1979) J. Biol. Chem. 254, 1516-1520.
2. Nilson, J. H., Barringer, K. J., Convey, E. M., Friderici, K., and Rottman, F. M. (1980) J. Biol. Chem. 255, 5871-5878.
3. Nilson, J. H., Thomason, A. R., Horowitz, S., Sasavage, N. L., Blenis, J., Albers, R., Saber, W., and Rottman, F. M. (1980) Nucleic Acids Res. 8, 1561-1573.
4. Chang, A. C. Y., Nunberg, J. H., Kaufman, R. J., Erlich, H. A., Schinke, R. T., and Cohen, S. N. (1978) Nature 275, 617-629.
5. Goeddel, D. V., Heyneker, H. L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D. G., Ross, M. J., Mossa, G., Creau, R., and Seeburg, P. H. (1979) Nature 281, 544-548.
6. Grunstein, M., and Hogness, D. S. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 3961-3965.
7. Birnboim, H. C., and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
8. Sanger, F., Nickelson, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467.
9. Smith, A. J. H. (1979) Nucleic Acids Res. 6, 831-848.
10. Zain, B. S., and Roberts, R. J. (1979) J. Mol. Biol. 131, 341-352.
11. Sanger, F., and Coulson, A. R. (1978) FEBS Lett. 77, 107-110.
12. Maxam, A., and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
13. Dayhoff, M. O. (1978) Atlas of Protein Sequence, Vol. 5, Suppl. 2, National Biomedical Research Foundation, Washington, D. C.
14. Lingappa, R. V., Devillers-Thiry, A., and Blobel, G. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 2432-2436.
15. Jackson, R. C., and Blobel, G. (1980) Ann. N. Y. Acad. Sci. 343, 391-404.
16. Proudfoot, N. J., and Brownlee, G. G. (1979) Nature 281, 211-214.
17. Miller, W. L., Thirion, J. P., and Martial, J. A. (1980) Endocrinology 107, 851-854.
18. Gubbins, E. J., Maurer, R. A., Lagrimini, M., Erwin, C. R., and Donelson, J. E. (1980) J. Biol. Chem. 255, 8655-8662.
19. Cooke, N. E., Coit, D., Weiner, R. L., Baxter, J. D., and Martial, J. A. (1980) J. Biol. Chem. 255, 6592-6610.
20. Jeffreys, A. J. (1979) Cell 18, 1-70.
21. Lai, E. C., Woo, S. L. C., Daguyczyk, A., and O'Malley, B. W. (1979) Cell 16, 291-311.
22. Cooke, N. E., Coit, D., Shine, J., Baxter, J. D., and Martial, J. A. (1981) J. Biol. Chem. 256, 4007-4016.
23. Sharp, F., Nickelson, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 68, 868-869.
Nucleotide sequence of bovine prolactin messenger RNA. Evidence for sequence polymorphism.
N L Sasavage, J H Nilson, S Horowitz and F M Rottman

J. Biol. Chem. 1982, 257:678-681.

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