Molecular Cloning of DNA Complementary to Bovine Growth Hormone mRNA

(Received for publication, May 19, 1980)

Walter L. Miller,‡ Joseph A. Martial,‡ and John D. Baxter‡

From the ‡Department of Pediatrics and the ‡Departments of Medicine and Biochemistry and Biophysics, The Howard Hughes Medical Institute Laboratories, and the Metabolic Research Unit, University of California, San Francisco, California 94143

We have cloned DNA complementary to mRNA coding for bovine growth hormone (bGH). Double-stranded DNA complementary to bovine pituitary mRNA was inserted into the Pst I site of plasmid pBR322 by the dC-dG tailing technique and amplified in E. coli. A recombinant plasmid containing bGH cDNA was identified by hybridization to cloned rat growth hormone cDNA. It contains the entire coding and 3'-untranslated regions and 31 bases in the 5'-untranslated region. Nucleotide sequence analysis determined the amino acid sequence of the secreted hormone at all but 2 residues. Codon usage is not random, with 81.7% of the codons ending in G or C. The nucleotide sequence of bGH mRNA is 83.9% homologous with rat GH mRNA and 76.5% homologous with human GH mRNA, while the respective amino acid sequence homologies are 83.5% and 66.8%.

MATERIALS AND METHODS

Preparation of mRNA—Female bovine pituitaries were collected shortly after killing at a local abattoir and were frozen immediately in liquid N₂. Total RNA was prepared by extraction with guanidine thiocyanate, centrifugation through 5.7 M CaCl₂ (6, 7), phenol extraction, and ethanol precipitation. Polyadenylated RNA (poly(A)⁺ RNA) was prepared by oligo(dT)-cellulose chromatography using a modification (8) of the method of Aviv and Leder (9). Integrity of the poly(A)⁺ RNA was assessed by cell-free translation using a rabbit reticulocyte system (8, 10). Bovine growth pre-hormone was immunoprecipitated using an heterologous anti-ovine GH antiserum and prepared by adsorption to formalin-fixed S. aureus Cowan strain I, as described previously (10, 11). Cell-free translation products and immunoprecipitates were displayed on NaDodSO₄-acrylamide gels (12).

Preparation of Double-stranded cDNA for Cloning—Polyadenylated RNA was reverse-transcribed into single-stranded complementary DNA using reverse transcriptase by a modification (8) of the method of Monahan et al. (13). After alkaline digestion of the RNA, restriction extraction, Sephadex G-50 chromatography, and ethanol precipitation, the single-stranded cDNA was used to prime the synthesis by reverse transcribe of a complementary DNA strand. The single-stranded "hairpin loop" at the 5' end of the first cDNA strand, as well as any other single-stranded DNA regions, was digested briefly with S, nuclease (14) and the double-stranded cDNA was again purified by phenol extraction, Sephadex G-50 chromatography, and ethanol precipitation. Single-stranded tracts of homopolymeric deoxycytidine were then added to the 3' ends of the cDNA using terminal transferase (Enzo Biochemicals) and the potassium cacodylate buffer system of Roychoudhury et al. (15). Plasmid pBR322 (16) was cut with Pst I and similarly "tailed" with deoxyguanosine. Twenty nanograms of tailed cDNA was hybridized to 50 ng of tailed pBR322 in 50-µl reactions as described previously (17). This mixture was then used directly to transform bacteria under P₂. EK 2 conditions, in accordance with NIH Guidelines. E. coli J776 were rendered permeable to DNA by incubation in 75 mM CaCl₂, 5 mM MgCl₂, 10 mM Tris, pH 7.5, for 20 min at 4°C, then 2 min at 41°C, mixed with 10 volumes of nutrient broth, incubated 2 to 3 h at 37°C, and plated onto agar containing 15 µg of tetracycline/ml.

Identification and Preparation of pBP348—E. coli harboring recombinant plasmids were identified by antibiotic sensitivities and screening for the presence of cloned DNA by colony hybridization to freshly prepared bovine pituitary [³²P]cDNA (18). Plasmid DNA was prepared from selected colonies, cut with Pst I, electrophoresed on 1% agarose, stained with ethidium bromide, and photographed, then transferred to nitrocellulose paper (19). The presence of GH gene sequences in the transferred DNA was assessed by hybridization to cloned, full-length rat GH cDNA (20) labeled by nick translation to ~10⁷ cpm/µg of DNA (21). One clone, pBP348 (plasmid bovine pituitary, 348th clone obtained), which hybridized with the nick-translated rGH cDNA, was chosen for further study.

DNA Sequence Analysis—Plasmid pBP348 was cut with Pst I, and the phosphatase on the 5' ends of the DNA fragments was removed with alkaline phosphatase and replaced with [³²P]phosphate using polynucleotide kinase. Subsequent cutting with a variety of other enzymes, polynucleoside gel electrophoresis, and staining and autoradiography of the bands of DNA provided a restriction map of the cloned DNA. A large batch of pBP348 was then prepared and cut with Pst I, Pvu II, or Sau 3A, labeled with [³²P]-dATP and polynucleotide kinase, and then cut with other enzymes to yield DNA fragments labeled at a single end. These fragments were prepared by elution from polycrylamide gel and sequenced by the modified (17) method of Maxam and Gilbert (22).

RESULTS

Preparation and Translation of RNA—Pituitary tissue weighing 8.9 g yielded 7.0 mg of total RNA, of which 430 µg was polyadenylated. This poly(A)⁺ RNA supported efficient incorporation of [³⁵S]methionine into proteins in the reticulocyte system, as shown in Fig. 1. The clarity of the high

7521
mRNA. Reticulocyte lysate, prepared from rabbits made anemic with acetylphenylhydrazine, was treated with micrococal nuclease to digest endogenous mRNA. Lysate, poly(A)+ RNA and amino acids including [35S]methionine were incubated, and the synthesized proteins were displayed by electrophoresis on NaDodSO4 gels containing 12.5% polyacrylamide. Gels were fixed, stained, dried, and autoradiographed. Lane A, lysate incubated without added RNA. Lane B, lysate incubated with bovine pituitary poly(A)+ RNA. Lanes C and G: molecular weight markers (C-proteins, Amersham), myosin (200,000), phosphorylase b (92,500), bovine serum albumin (69,000), ovalbumin (45,000), carbonic anhydrase (30,000), and lysozyme (14,300). Lane D, proteins immunoprecipitated from the total lysate by antisera to ovine GH. Lane E, immunoprecipitate with antiserum to ovine PRL. Lane F, immunoprecipitate with nonimmune serum.

molecular weight bands testifies to the integrity of the poly(A)+ RNA preparation. The two prominent bands having molecular weights of about 25,000 and 24,000 are pre-prolactin and growth pre-hormone, respectively, of the total lysate, mRNAs for GH and PRL represent 12.6 and 14.3% of the total lysate, respectively.

**Synthesis and Cloning of cDNA**—Total pituitary poly(A)+ RNA was used as template for the synthesis of cDNA with reverse transcriptase using oligo(dT) primers for first strand synthesis and self-priming for the second strand. Efficiencies of reverse transcription were generally 6 to 10% for the first strand and 75 to 100% for the second. When aliquots of double-stranded cDNA were digested with a variety of restriction endonucleases and displayed on acrylamide gels, discrete bands were seen suggesting that a small number of cDNA species predominated (20, 23). This cDNA was then trimmed with S1, tailed with dCMP, and ligated into the Pst I site of pBR322 as described under "Materials and Methods." Transformation efficiencies with such recombinant plasmids were about 5 colonies/ng of pBR322 used, whereas supercoiled pBR322 gave efficiencies of 1000 to 4000 colonies/ng.

**Nucleotide Sequence of bGH cDNA**—A colony containing bGH gene sequences was identified by hybridization to cloned rat GH cDNA as described under "Materials and Methods." Restriction enzyme analysis of the DNA cloned in pBP348 showed that it contained two internal Pst I sites and two Pvu II sites, which divided the cDNA into pieces of convenient size for sequencing. The two Pvu II sites subdent a DNA fragment of 489 base pairs; a Pvu II fragment of this size was seen in the analytical digests of the original cDNA. The cloned cDNA contains 831 base pairs including 31 base pairs in the 5'-untranslated region, the entire region coding for the pre-hormone (551 base pairs), the entire 3'-untranslated region (104 base pairs), a brief stretch of poly(A), and the dC:dG tails (Fig. 2). The amino acid sequence predicted by the nucleotide sequence differs from the published amino acid sequence of bGH (5, 24) only at positions 47 and 66, where aspartic acid and glutamic acid are replaced by their respective amides. Bovine growth hormone has previously been noted to contain either valine or leucine at position 127 (24, 25); the individual mRNA molecule leading to pBP348 coded for leucine at this position. The amino acid sequence of the signal peptide was deduced from the nucleotide sequence preceding the codon for amino acid 1. We have assigned this position to alanine, although bGH isolated from bovine pituitaries may have either alanine or phenylalanine at its NH2 terminus (5). Two methionine codons, the first one presumably being the initiation site for translation, are found in the sequencing frame at positions 25 and 26. The positions of the methionines and 9 other amino acids in the pre-sequence have been determined.
Amino acid and nucleotide homologies among bovine, human, and rat growth pre-hormones

The entire coding regions, including the "pre-sequences," as well as the termination codons, were compared. Gaps in the amino acid sequence were introduced as shown in Fig. 3; where gaps were not needed in a paired comparison, they were eliminated, accounting for the differences in the lengths of the compared sequences. Times of evolutionary divergence, in millions of years before the present, were calculated using a unit evolutionary period = 4 (30).

| Amino acid sequence | bGH/rGH | bGH/bGH | rGH/hGH |
|---------------------|---------|---------|---------|
| Identical           | 182/218 | 83.5    | 147/229 | 66.8    | 141/219 | 64.4    |
| Conservative differences | 27/36  | 75.0    | 43/73   | 58.9    | 50/78   | 64.1    |
| Nucleotide sequence |         |         |         |         |         |         |
| Identical nucleotides | 549/654 | 83.9    | 505/660 | 76.5    | 492/657 | 74.9    |
| Identical codons     | 133/218 | 61.0    | 107/220 | 48.6    | 92/219  | 42.0    |
| Two identical nucleotides per codon | 67/218  | 30.7    | 81/220  | 36.8    | 96/219  | 43.8    |
| Not expressed        | 47/67   | 70.1    | 38/81   | 48.1    | 48/98   | 50.0    |

Divergence (yrs before present, \( \times 10^6 \))

\begin{tabular}{c|c|c|c|c|c|c}
 & bGH & rGH & rGH/hGH & bGH/\(d\)GH & rGH/\(d\)GH & bGH/\(i\)GH \\
\hline
Amino acid sequence & & & & & & \\
Identical & 182 & 218 & 83.5 & 147 & 229 & 66.8 \\
Conservative differences & 27 & 36 & 75.0 & 43 & 73 & 58.9 \\
Nucleotide sequence & & & & & & \\
Identical nucleotides & 549 & 654 & 83.9 & 505 & 660 & 76.5 \\
Identical codons & 133 & 218 & 61.0 & 107 & 220 & 48.6 \\
Two identical nucleotides per codon & 67 & 218 & 30.7 & 81 & 220 & 36.8 \\
Not expressed & 47 & 67 & 70.1 & 38 & 81 & 48.1 \\
\hline
Divergence (yrs before present, \( \times 10^6 \)) & 96.0 & 132.8 & 142.4 & & & \\
\end{tabular}

Fig. 3. Homologies among bovine, rat, and human growth pre-hormones. Gaps were introduced at the amino acid residues shown to permit matching of regions of obvious homology. The arrangement was used to calculate the data in Table I.

Discussion

Although non-primate growth hormones are inactive in man (28, 29), they exhibit considerable amino acid homology. We have compared the nucleotide sequences of the coding regions from the mRNAs for bovine, rat (20), and human (23) growth pre-hormones (Table I). One or two single codon gaps were introduced in each sequence as shown in Fig. 3 to permit alignment of long sequences with obvious homology. The coding regions of the bovine and rat mRNAs are very similar, having 83.5% homology, whereas the coding regions of human and bovine or human and rat GH mRNAs are only 76.5 and 74.9% homologous, respectively. The homology in the amino acid sequences of bovine and rat growth pre-hormones is 83.5%, similar to the nucleotide homology; this contrasts with the other two paired comparisons, where the amino acid homology is about 10% less than the nucleotide homology. Of the 36 amino acids differing in bovine and rat GHS, 27 (75.0%) are replaced by "favorable" substitutions, i.e. those which occur more frequently than random chance would predict (5), whereas only 59 to 64% of the substitutions are favorable in the other two paired comparisons.

These findings might suggest that cattle and rats are evolutionarily more closely related than human beings are with either of the other species. Wilson et al. have described a parameter called the unit evolutionary period, defined as the length of time in millions of years of evolution required for 1% non-homology to arise in the same protein in two different species (30). Using a unit evolutionary period of 4 for growth hormone (30), we calculate that the evolutionary precursors to cattle and rats diverged about 68 million years ago, while the evolutionary precursors to men diverged from those to rats and cattle about 140 million years ago. As Wilson et al. calculated units of evolutionary period for proteins rather than their pre-proteins, it could be argued that the pre-sequence should be eliminated from these calculations. This would reduce the time of divergence of rats and cattle to 48 million years ago, but would not change the time of divergence of human precursors from the other two species.

The 5'-untranslated region of the cloned bGH cDNA contains 31 nucleotides without apparent pallindromes or inverted repeats to suggest secondary structures. The 3'-untranslated region contains three long pallindromic regions (Fig. 2), the first of which (5'-U-G-C-C-C-C-U-C-C-C-C-C-G-3') is remarkably similar to pallindromic sequences rich in cytosine in the 3'-untranslated regions of mRNAs for bGH (23), rGH (20), and hCS (27). The significance of such conserved nontranslated pallindromes remains unknown.

Although non-primate growth hormones are inactive in man (28, 29), they exhibit considerable amino acid homology. We have compared the nucleotide sequences of the coding regions from the mRNAs for bovine, rat (20), and human (23) growth pre-hormones (Table I). One or two single codon gaps were introduced in each sequence as shown in Fig. 3 to permit alignment of long sequences with obvious homology. The coding regions of the bovine and rat mRNAs are very similar, having 83.5% homology, whereas the coding regions of human and bovine or human and rat GH mRNAs are only 76.5 and 74.9% homologous, respectively. The homology in the amino acid sequences of bovine and rat growth pre-hormones is 83.5%, similar to the nucleotide homology; this contrasts with the other two paired comparisons, where the amino acid homology is about 10% less than the nucleotide homology. Of the 36 amino acids differing in bovine and rat GHS, 27 (75.0%) are replaced by "favorable" substitutions, i.e. those which occur more frequently than random chance would predict (5), whereas only 59 to 64% of the substitutions are favorable in the other two paired comparisons.

These findings might suggest that cattle and rats are evolutionarily more closely related than human beings are with either of the other species. Wilson et al. have described a parameter called the unit evolutionary period, defined as the length of time in millions of years of evolution required for 1% non-homology to arise in the same protein in two different species (30). Using a unit evolutionary period of 4 for growth hormone (30), we calculate that the evolutionary precursors to cattle and rats diverged about 68 million years ago, while the evolutionary precursors to men diverged from those to rats and cattle about 140 million years ago. As Wilson et al. calculated units of evolutionary period for proteins rather than their pre-proteins, it could be argued that the pre-sequence should be eliminated from these calculations. This would reduce the time of divergence of rats and cattle to 48 million years ago, but would not change the time of divergence of human precursors from the other two species.

The 5'-untranslated region of the cloned bGH mRNA is similar to the 5'-untranslated regions of mRNAs coding for bGH and rGH. We have not attempted to calculate homologies as it is unclear where to introduce the required gaps in the sequences without the additional guidance of an amino acid sequence. However, the relatedness of these sequences can be seen in Fig. 4. The homologies are clustered in three regions, of which the first two contain long pallindromes. The third homologous region includes precisely the same 12-nucleotide sequence 5' U-C-C-C-C-C-C-U-C-C-C-C-3' in all three species. This last homologous region contains the 5' A-A-U-U-A-A-A-A-L 3' found in the 3'-untranslated region of most eukaryotic mRNAs (31).

Excluding the region of poly(A) and the "tails" of enzymatically added dC or dG, the cloned cDNA contains 787 nucleotides, of which 468 (59.5%) are either C or G. This contrasts sharply with 42% G + C content of vertebrate DNAs in general (32). This is due to non-random codon selection in bGH mRNA; 81.7% of which end in G or C. G and C are highly favored in the third position of the codons for rat (20) and human (23) growth pre-hormones and human chorionic somatomammotropin (27), but are only slightly favored in rat (17) and bovine PRL. The reasons for this preference for G and C in the third position of codons in mRNAs for growth hormone is unknown.

\footnote{W. L. Miller, unpublished observations.}
Acknowledgments—We thank Mrs. Doris Coit and Dr. Nancy Cooke for skillful assistance and advice, and Ms. Sonja Bock and Dr. Hugo Martinez for use of computer programs in analyzing DNA sequences.

REFERENCES
1. Catt, K. J., Moffat, B., and Niall, H. D. (1967) Science 157, 321
2. Sherwood, L. M. (1967) Proc. Natl. Acad. Sci. U. S. A. 68, 2307–2314
3. Li, C. H., Dixon, J. S., Lo, T. B., Pankov, Y. M., and Schmidt, K. D. (1967) Nature 224, 695
4. Niall, H. D., Hogan, M. L., Sayer, R., Rosenblum, I. Y., and Greenwood, F. C. (1971) Proc. Natl. Acad. Sci. U. S. A. 68, 898–899
5. Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. (1978) In Atlas of Protein Sequence and Structure (Dayhoff, M. O., ed) Vol. 5, Suppl. 3, pp. 345–352, National Biomedical Research Foundation, Washington, D.C.
6. Ulrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. J., and Goodman, H. M. (1977) Science 196, 1313–1319
7. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry, 18, 5294–5299
8. Miller, W. L., and McCarthy, B. J. (1979) J. Biol. Chem. 254, 742–748
9. Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1408–1412
10. Martial, J. A., Baxter, J. D., Goodman, H. M., and Seeburg, P. H. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 1816–1820
11. Miller, W. L., Baxter, J. D., and Roberts, J. L. (1980) Proc. Natl. Acad. Sci. U. S. A. in press
12. Laemmli, U. K. (1970) Nature 227, 680–685
13. Monahan, J. J., Harris, S. E., Woo, S. L. C., Robertson, D. L., and O’Malley, B. W. (1976) Biochemistry 15, 223–233
14. Leong, J. A., Garapic, A. C., Jackson, N., Fanshier, L., Levinson, W., and Bishop, J. M. (1972) J. Virol. 9, 891–902
15. Roycehodhury, R., Jay, E., and Wu, R. (1976) Nucleic Acids Res. 3, 863–877
16. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Cross, J. H., and Falkow, S. (1977) Gene 2, 95–113
17. Cooke, N. E., Coit, D., Weiner, R. I., Baxter, J. D., and Martial, J. A. (1980) J. Biol. Chem. 255, 6502–6510
18. Grunstein, M., and Horsness, D. S. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 3961–3965
19. Southern, E. M. (1975) J. Mol. Biol. 98, 503–517
20. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D., and Goodman, H. M. (1977) Nature 270, 486–494
21. Maniatis, R., Jeffrey, A., and Kleid, D. G. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 1184–1188
22. Maxam, A. M., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 560–564
23. Martial, J. A., Hallevoy, R. A., Baxter, J. D., and Goodman, H. M. (1979) Science 205, 652–657
24. Wallis, M. (1973) FEBS Lett. 35, 11–14
25. Seavey, B. K., Singh, R. N. P., Lewis, U. J., and Geschwind, I. I. (1971) Biochem. Biophys. Res. Commun. 43, 189–195
26. Lingappa, V. R., Devillers-Thiery, A., and Blobel, G. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 2432–2436
27. Shine, J., Seeburg, P. H., Martial, J. A., Baxter, J. D., and Goodman, H. M. (1977) Nature 267, 494–499
28. Raben, M. S. (1962) N. Engl. J. Med. 266, 31–35
29. Li, C. H., and Kuo, W. (1964) Experientia 20, 169–178
30. Wilson, A. C., Carlson, S. S., and White, J. T. (1977) Annu. Rev. Biochem. 46, 573–639
31. Proudfoot, N. J., and Brownlee, G. G. (1976) Nature 263, 211–214
32. Sueoka, N. (1961) J. Mol. Biol. 3, 31–40

Fig. 4. The 3'-untranslated regions of the mRNAs for bovine GH (top sequence), rat GH (20) (middle sequence), and human GH (23) (lower sequence). Homologous regions found in all three mRNAs are connected by lines.