Structure of Cloned DNA Complementary to Rat Prolactin Messenger RNA*

Nancy E. Cooke, Doris Coit, Richard I. Weiner; John D. Baxter, and Joseph A. Martial

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

Prolactin (Prl), growth hormone (GH), and chorionic somatomammotropin (CS) form a set (the "Prl set") of mammalian hormones which is thought to have evolved from a common ancestral gene. Until now this assumption was based on similarities in their amino acid sequences, on their often overlapping biological and immunological properties, and on nucleic acid sequence data for two members of the set (GH and CS). In the current study, we report the amplification in bacteria and the sequence analysis of DNA complementary to rat Prl mRNA. Two independent clones were obtained, one of which contained 823 bases including the entire sequence coding for rat prolactin and its signal peptide and parts of the 5'- and 3'-untranslated regions of the mRNA. The nucleotide sequences of the two cloned DNAs differ by three nucleotides. In addition, both predicted amino acid sequences also differ by about 14% from the published amino acid sequence of the protein. Although technical problems cannot be excluded for each of these discrepancies, the data suggest a revised structure for the protein and the possible existence of distinct nonallelic Prl genes. The results of this study have allowed us to compare the nucleotide sequences of the structural genes coding for each of the hormones of the Prl set. The codons for rPrl mRNA do not show the marked preference for GMP and CMP in the third position observed in rGH, hGH, and hCS mRNAs. rGH and rPrl mRNAs show 39% nucleic acid homology in the coding portion, but little homology in the untranslated regions. This differs from the comparison between hGH and hCS where the same degree of homology (about 93%) was found in both the coding and noncoding regions. Ten per cent of the codons of rPrl and rGH are identical, and 26% differ by only one base. Of the latter, only 58% result in amino acid changes, which most often are conservative. These results strongly support the hypothesis that the Prl and GH genes derive from a common precursor. The data allow us to suggest a date for this gene duplication event around 380 million years ago.

The polypeptide hormones prolactin (Prl), growth hormone (GH), and chorionic somatomammotropin (CS) are closely related (Catt et al., 1967; Li et al., 1969; Sherwood, 1967; Niall et al., 1971). Their amino acid sequence homologies range from 20 to 85%, depending on the species compared, possibly explaining the observed overlap in some of their biological and immunological properties (Niall et al., 1973). DNA complementary to rat GH1 (Seeburg et al., 1977), human GH (Martial et al., 1979), and human CS (Shine et al., 1977) mRNAs have been cloned and sequenced recently. More homology was found between the nucleic acid sequences of hGH and hCS, than between those of hGH and rGH (Martial et al., 1979). These studies supported the hypothesis that the genes coding for these hormones originated from a common precursor by gene duplication and that the CS-GH gene duplication occurred at some time after the separation of the human and rat species (Bewley et al., 1972; Niall et al., 1971). Although a fragment of DNA complementary to a portion of rPrl mRNA has been cloned and sequenced (Gubbins et al., 1979), the complete nucleotide sequence of a Prl structural gene has been missing from these comparisons.

A cloned Prl cDNA would be useful along with the GH and CS cDNAs for studying the differential expression of these related genes. Both Prl and GH genes are expressed predominantly in the pituitary, whereas the CS gene is expressed predominantly in the placenta. The Prl and GH genes are also expressed in several lines of cultured rat pituitary tumor cells (Dannies and Tashjian, 1973) which can be manipulated by a variety of agents to favor their differential expression (Martin and Tashjian, 1977; Vale et al., 1973). Thus, cloned cDNA to Prl mRNA would be a useful probe for studying the molecular mechanism of this regulation. In addition, this cDNA constitutes the probe needed to isolate the genomic Prl sequences from chromosomal DNA.

We report here the cloning in bacteria and the complete nucleotide sequence of the structural gene coding for rPrl. These data enable us to propose a revised amino acid sequence for rPrl, including its signal peptide. While the data demonstrate some homology between rPrl and rGH supporting the evolutionary relatedness of the two genes, they also document major differences in the structures and codon choices. The data are consistent with the possibility that Prl is the most primitive molecule in this "Prl set" of related polypeptide hormones.

EXPERIMENTAL PROCEDURES

Animals—Female Sprague-Dawley (Wistar) rats were anesthetized with Equithesen (pentobarbital 1%, ethanol 11.5%, chloral hydrate 4%, MgSO4, 2%, propylene glycol 44% in sterile water) and the medial basal hypothalami destroyed using a modified Halasz knife as previously described (Cheung and Weiner, 1976). At the same time, 40-mm silastic capsules filled with 17β-estradiol were implanted subcutaneously. Twenty-three days later the animals were killed by decapitation, the pituitaries removed, and trunk blood collected for

* This work was supported by National Institutes of Health Grant GM 25549 to J.A.M. 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.

† Department of Obstetrics, Gynecology and Reproductive Sciences.

1 The abbreviations used are: GH, growth hormone; CS, chorionic somatomammotropin; Prl, prolactin.
radioimmunoassay of Prl. Prl concentrations were measured using rabbit anti-rPrl provided by Dr. Albert F. Parlow through the National Institute of Arthritis, Metabolism and Digestive Diseases rat pituitary hormone distribution program.

**Cultured Pituitary Cells—**GH, and GC cells were grown at 37°C in monolayer as previously described (Martial et al., 1977), in Ham F-10 and MEM Joklik medium, respectively. Both media were supplemented with 15% fetal calf serum. For hormonal stimulation, GH cells received fresh medium containing thyrotropin-releasing hormone (TRH, 30 nM, Sigma) and estradiol (100 nM, Sigma) 48 h prior to harvest. Cells were harvested with 0.05% trypsin and 0.02% EDTA.

**Preparation and Characterization—**The frozen pituitaries or cell pellets were added to a sterile solution containing 5% guanidine thiocyanate, 50 mM Tris-HCl, pH 7.5, 2 mM EDTA, and 5% mercaptoethanol (Chirgwin et al., 1979). After homogenization, Polytron of the tissue, N-lauryl sarcosine was added to 2%, CsCl to 0.35 M, and the mixture was heated for 2 min at 65°C. This solution was layered over a cushion of 5.7 M CsCl, 1 mM EDTA, and centrifuged in an SW 40 rotor at 32,000 rpm for 24 h at 20°C. The resultant pellets were dissolved in 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, and 1% N-lauryl sarcosine, extracted with phenol and chloroform and precipitated with ethanol as previously described (Seeburg et al., 1979).

Polyadenylated RNA was isolated by affinity chromatography on oligo(dT) cellulose (P-L Biochemicals, type 7) and translated in a wheat germ cell-free system in the presence of [35S]methionine (Amersham, 600 to 1300 Ci/mmole) as previously reported (Martial et al., 1977). Preprolactin synthesized in the cell-free system was immunoprecipitated with rabbit antiserum to rPrl and the Cowan strain of Staphylococcus aureus as described earlier (Martial et al., 1977). 

**Hybridization Probes—**cDNA probes were prepared in 100 µl reactions containing 60 mM NaCl, 10 mM diithiothreitol, 50 mM Tris-HCl, pH 8.3, 6 mM MgCl2, 1.5 µg of oligo(dT)19, 1.5 μg of mRNA, 40 units of reverse transcriptase, 400 mM each dATP, dTTP, dGTP, and 11 µM [α-32P]dCTP dried under N2. Incubation was at 37°C for 60 min. Reactions were stopped with addition of EDTA to 10 mM followed by phenol and chloroform extraction and fractionation over Sephadex G-50 as described above. The ethanol-precipitated void volume peak was taken up in 0.4 M NaOH, 5 mM EDTA, heated to 37°C for 60 min, and then neutralized with HCl. Incorporation of 10cpm/µg of DNA was obtained routinely.

Nick-translated probes were prepared as described (Maniatis et al., 1975); 10cpm/µg of DNA were obtained routinely.

**DNA Gel Electrophoresis—**Electrophoresis on 4.5 to 10% polyacrylamide slab gels was performed in Tris/borate/EDTA buffer (Peacock and Dingman, 1967) at 30 mA/gel. Electrophoresis on 0.8 to 2% agarose horizontal gels was performed in Tris/acetate/NaCl/EDTA buffer (Peacock and Dingman, 1968) at 50 V/gel.

**DNA Sequence Analysis—**3’-Ends were labeled with polynucleotide kinase (Boehringer-Mannheim) and [γ-32P]ATP (Amerham, 2400 Ci/mmole) after previous dephosphorylation of the DNA with alkaline phosphatase (Enzo Biochemicals). 3’-Ends were labeled by E. coli DNA polymerase I, Klunglow fragment (Boehringer-Mannheim) and the appropriate [α-32P]deoxynucleoside triphosphate. The chemical cleavage technique (Maxam and Gilbert, 1977), with one modification, was used on DNA restriction fragments. For chemical modification at the adenine residues, 25 µl of formic acid (Mallinkrodt) were added to 10 µl of [32P]dDNA and allowed to react for 10 min at 20°C. Reaction was stopped with addition of 200 µl of 0.3 mM sodium acetate, 0.1 mM EDTA, and 25 µg/ml of RNA. The thin gel system was used (Sanger et al., 1977).

**RESULTS**

**Stimulation of rPrl Secretion and Characterization of the mRNA—**Polyadenylated RNA was prepared as described under "Experimental Procedures" from rat pituitaries and from various cultured GH cells (Dannies and Tashjian, 1973). Analysis of the relative abundance of rPrl mRNA in the individual preparations was performed by cell-free translation and, in some cases, immunoprecipitation followed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Fig. 1). Preprolactin and pregrowth hormone co-migrated in our gel electrophoresis system. Pregrowth hormone is the major translation product of the mRNA extracted from GC cells (Lanes c), whereas immunoprecipitated proprolactin is shown in Lanes g and i. In the cultured GH cells, treatment for 48 h with estradiol and thyrotropin-releasing hormone (TRH) resulted in a severalfold enrichment in rPrl mRNA (Fig. 1, Lanes b and a, control and immunoprecipitated, respectively). However, the RNA extracted from normal rat pituitaries (Fig. 1, Lanes e and f, female and male, respectively) contained a greater amount of Prl or GH mRNA, or both, than the cultured cells. The normal female rat pituitary is more enriched in Prl mRNA than normal male rat pituitary as shown by the immunoprecipitation of the translation products with antisem to rPrl (Fig. 1, Lanes g and k). The greatest amount of Prl mRNA was found in the pituitaries of the female rats

J. Shine, personal communication.
treated with chronic estrogen stimulation and ablation of the medial basal hypothalamus (Fig. 1, Lanes d and i; total cell-free translation and immunoprecipitation, respectively). The serum concentrations of Prl in these animals were found by radioimmunoassay to be elevated to 1200 ng/ml, a 40-fold increase over normal. Based on the cell-free translation data, the Prl mRNA represents about 80% of the total pituitary mRNA after these treatments.

Double-stranded cDNA Synthesis and Cloning—Rat Prl cDNA was synthesized and cloned in two separate experiments. In the first experiment, double-stranded cDNA was prepared from the Prl-enriched pituitary polyadenylated RNA template using reverse transcriptase from avian myeloblastosis virus (AMV) to synthesize both strands followed by S1 nuclease digestion of the resulting hairpin loop. Aliquots of this material digested with several restriction endonucleases generated a series of discrete DNA fragments separable by polyacrylamide gel electrophoresis (data not shown). The presence of prominent bands suggested that this cDNA was quite enriched in one or a few specific DNA species. The DNA complementary to rPrl mRNA was expected to be among these species based on the cell-free translation data.

Chemically synthesized 32P-labeled double-stranded DNA decamers containing the restriction site for HindIII (5'-C-C-A-A-G-C-T-T-G-G-3') were then ligated to each end of the double-stranded cDNA to rat pituitary mRNA. After digestion of this product with HindIII to remove the extra HindIII concatamers, the double-stranded cDNA with a single HindIII cohesive terminus at each end was purified on a preparative polyacrylamide gel. Double-stranded cDNA with a length of about 450 to 1000 base pairs was electrophoresed from the gel and ligated into the HindIII site of pBR322. This plasmid was then used to transform E. coli strain χ1776 as described under...
Fig. 3. Detection of the cloned plasmid containing the 823-base pair rPrl cDNA sequences. Tet" amp" colonies were grown on Whatman 540 paper and analyzed by hybridization with a [32P]-cDNA probe complementary to the initial pituitary mRNA template (A). The plasmid DNA from the colonies that hybridized were digested with Pst I, electrophoresed on a 1.5% agarose gel, and stained with ethidium bromide (B). DNA from pM2 markers (Lane M) and the sizes in kilobases are indicated on the right. DNA from the gel (see B) was transferred to a nitrocellulose filter and hybridized with the 486-base pair rPrl insert labeled by nick-translation (Maniatis et al., 1975) (C). The lane on the right contained the cloned 823-base pair rPrl cDNA.

### Table II

Summary of nucleotide usage in the third position of the codon in some eukaryotic hormone mRNAs

| Hormone                        | G + C % |
|--------------------------------|---------|
| hCS                           | 80      |
| hGH                           | 76      |
| rGH#                          | 74      |
| rPrl                          | 50      |
| Rat Insulin I                  | 75      |
| Rat Insulin II                 | 75      |
| Bovine corticotropin-β-lipotropin# | 65    |
| Human chorionic gonadotropin (α subunit)# | 56    |
| Bovine preproparathyroid hormone# | 43    |

# Experimental Procedures.

Of 25 amp" tet" colonies, eight contained a plasmid with an insert. Only one of these eight plasmids hybridized (Southern, 1975) with a [32P]-cDNA probe complementary to the initial mRNA template (Fig. 2). This plasmid was amplified in bacteria and purified. It contained a 486-base pair DNA insert, the nucleotide sequence of which was consistent with that predicted from the known amino acid sequence of rPrl (Shome and Parlow, 1977).

In the second cloning experiment, single-stranded cDNA was prepared from the same pituitary RNA preparation. Next, an oligopolymeric tract (tail) of dCMPs was added to its 3' end as described under "Experimental Procedures." An estimated 80 dCMP residues were added/molecule of cDNA. Therefore, the synthesis of the second strand could be primed on this poly(dC) tail with the use of an oligo(dG) primer and reverse transcriptase. The double-stranded product was then 3' tailed with about 15 dCMP residues/end as before. This product was annealed to pBR322 previously digested with Pst I and 3' tailed with dGMP and the recombinant plasmid transformed into the E. coli strain x177fi. As the Pst I site of pBR322 interrupts the 8-lactamase gene near its 3'-end, insertion of foreign DNA in this site may not abolish ampicillin resistance. Nevertheless, 210 tet" amp" colonies were obtained with 50 ng of cDNA. Twenty-five of these colonies were detected by colony hybridization to a cDNA probe prepared from the pituitary mRNA. An autoradiograph of a represent-
FIG. 5. Comparison of the nucleotide sequences and predicted amino acid sequences of the mRNAs coding for rPrl and rGH. The mRNA sequences are taken from Fig. 4 and from Seeburg et al., 1977. The amino acid sequences were deduced from the mRNA sequences using the genetic code. Shaded rectangles indicate identical amino acids and dashed rectangles indicate amino acids related through favored substitutions (Dayhoff, 1978).

The rPrl mRNA sequence and the derived amino acid sequence are shown in Fig. 5. If this assignment of amino acids is correct and if translation begins with the methionine codon at position -28 as indicated, then the primary translation product of rPrl mRNA is a protein of 25,655.65 daltons. Preliminary amino acid sequences for rPrl (Shome and Parlow, 1977) and its NH-terminal signal peptide (McKean and Maurer, 1978) have been reported. Our predicted amino acid sequence agrees with the published sequences in 172 of the 197 amino acids of rPrl and in 27 of 28 amino acids in the signal peptide. Our sequence does not contain the alanine residue reported at position -20 of the signal peptide, suggesting that the actual signal peptide contains 28 rather than 29 amino acids. Similarly, our data indicate that the mature hormone contains 197 rather than 198 amino acids as previously reported. Of the remaining discrepancies, 15 are omissions or additions, 9 are differences in the reported amino acid, and 4 result from two inversions of two amino acids. Three of our changes convert the amino acid to the one found in the corresponding position in hPrl (Shome and Parlow,
Prolactin cDNA Structure

6/1UU/Ph e 3/1UU/Sec 5/1AU/Tyr 1/OGU/Cys
2/1UC/Ph e 5/1UC/Sec 2/1UC/Tyr 6/1OC/Cys
0/1UA/Leu 4/1UC/Sec 1/1UA/OC 0/1OG/OP
4/1OG/Leu 0/1OG/Sec 0/1UNG/NM 2/1OG/Tyr

4/1UU/Leu 6/1CU/Pro 3/1CN/Thr 2/1CGA/Arg
7/1UC/Leu 1/1CC/Pro 3/1CN/Thr 1/1CC/Arg
3/1CA/Leu 2/1CC/Pro 8/1CA/Gln 0/1CGA/Arg
11/1CU/Leu 6/1CC/Pro 6/1CA/Gln 3/1CGA/Arg

6/1AU/1le 3/1CU/Thr 5/1AN/Asn 1/1AGA/Arg
7/1UC/1le 2/1CC/Thr 8/1AC/Asn 1/1AGA/Arg
3/1NA/1le 3/1CC/Thr 7/1NA/Lys 1/1AGA/Arg
5/1UG/Met 0/1CA/Pro 6/1AG/Lys 2/1AGA/Arg

4/1GU/Val 4/1CU/Ala 6/1GNA/Asp 2/1GCU/Gly
5/1GC/Val 7/1CC/Ala 7/1GCA/Asp 2/1GCA/Gly
6/1CA/Val 11/1GAA/Glu 4/1GAA/Glu
4/1GU/Val 0/1GCA/Glu 5/1GAA/Glu 2/1GAA/Gly

**Fig. 6. Codon usage in rPrl mRNA.** The numbers indicate the frequency with which the codons are used in the coding region of rPrl mRNA. OC, OP, and AM designate the terminators ochre, opal, and amber, respectively.

1977). A cluster of 18 discrepancies occurs between amino acids 70 and 104, an area whose amino acid sequence was apparently determined with chymotryptic digests (our conclusion by analogy with the strategy used for hPrl, Shome and Parlow, 1977). Since proper alignment is often difficult with such digests, the nucleic acid sequence may be more reliable in this case. Also, shown in Fig. 6 are the amino acid and predicted amino acid sequences for rGH (Seeburg et al., 1977). The amino acid sequence of rGH and rPrl were aligned by introducing arbitrary gaps in order to maximize their identity. Table I summarizes conclusions based on this comparison of the molecules.

The codon selection in rPrl mRNA is nonrandom (Fig. 6), as in rGH, hGH, and lCS. However, in contrast to these other members of the Prl set, rPrl has no preference for GMP and CMP over AMP and UMP in the third position of the codons (Table II).

**DISCUSSION**

In the studies reported here, analysis of two different cloned DNA fragments containing structural rPrl gene sequences is presented. One of the fragments, containing 486 base pairs, was inserted into a plasmid after addition of Hinfl linker DNA to blunt-ended double-stranded cDNA. The other fragment which contains almost all of the bases complementary to rPrl mRNA was inserted into a plasmid with the use of dGMP-dCMP tailing. The nucleotide sequence of these DNAs, although generally agreeing with that expected from the amino acid sequence data (Shome and Parlow, 1977), has enabled us to propose a substantial revision in the preprolactin sequence. The nucleic acid sequence was also compared to the nucleic acid sequence of these DNAs, although generally agreeing with that expected from the amino acid sequence data (Shome and Parlow, 1977). Since proper alignment is often difficult with such digests, the nucleic acid sequence may be more reliable in this case. Also, shown in Fig. 6 are the amino acid and predicted amino acid sequences for rGH (Seeburg et al., 1977). The amino acid sequence of rGH and rPrl were aligned by introducing arbitrary gaps in order to maximize their identity. Table I summarizes conclusions based on this comparison of the molecules.

**RNA Isolation and cDNA Cloning**—Three unusual approaches were applied to the isolation of rPrl mRNA, its conversion to double-stranded cDNA and subsequent cloning. First, the production of mRNA highly enriched in Prl sequences was stimulated by subcutaneous 17β-estradiol implantation and by destruction of the medial basal hypothalamus in a group of female rats. Prl secretion from mammalian anterior pituitary is mainly under inhibitory hypothalamic control (Takahara et al., 1974; MacLeod, 1976; Ben-Jonathan et al., 1977). Destruction of the medial basal hypothalamus increases circulating Prl concentrations while inhibiting the secretion of other anterior pituitary hormones (Bishop et al., 1972). In addition, the number of mammmotrophs increases 3-fold in animals with lesions.1 Estrogen treatment also stimulates prolactin synthesis (Lieberman et al., 1978; Seeburg et al., 1979) and causes hyperthrophy of mammotrophs (Gersten and Baker, 1970). In the present study, combining these two treatments caused a greater than 2-fold increase in the weight of the anterior pituitary gland (16 to 38 mg) and a 40-fold increase in serum Prl (30 to 1200 ng/ml). Analysis by cell-free translation of the RNA isolated from the pituitaries of these treated rats showed that the level for Prl mRNA was substantially greater than that seen in normal rats or in cultured pituitary tumor cells (GH, cell line) that were stimulated with estrogens and thyrotropin-releasing hormone. This obviated the need for further mRNA purification and thus enabled us to synthesize directly cDNA highly enriched in Prl gene sequences. Secondly, for synthesis of the 823-base pair fragment, we devised a technique to prevent the loss of DNA sequences complementary to the 5'-end of the mRNA. This loss inevitably occurs when 5' nucleic acid is used to digest the hairpin loop joining the two strands of DNA resulting from the self-primation of the second strand on the first one. In this technique, an oligoplymeric tract of dCMPs was added to the 3'-end of the single-stranded cDNA. This tract used as a template to prime the synthesis of the second cDNA strand using a commercial oligo(dG) primer. Sequencing of the final double-stranded cDNA molecule which had been dCMP-tailed again at both 3'-ends, revealed tracts of 14 and 16 dCMPs at each end, not the predominance of dCMPs on the 5'-end of the message strand expected from the initial tailing of the single-stranded DNA. We subsequently found that our terminal deoxynucleotidyl transferase was contaminated with exonuclease, and this may have caused loss of the 3'-poly(A) tract and adjacent sequences in our cloned molecule (including the AAUAAA found in most other eukaryotic mRNAs). The loss was probably minimal based on estimates of full length rPrl mRNA as 840 ± 40 bases (Evans and Rosenfeld, 1979). Despite this problem the technique resulted in a longer 5'-untranslated region than was previously possible in our hands. Third, the time of the first and second strand synthesis with reverse transcriptase was extended from the usual 15 min (Seeburg et al., 1977; Shire et al., 1977) to 75 to 90 min. This was done after control experiments with our enzyme indicated that [3H]dCMP continued to accumulate in specific restriction fragments of the product for up to 90 min. This result is in agreement with the results of Buell et al. (1978).

**Comparison of the Closed rPrl cDNAs**—The 486-base pair fragment coding for amino acids 33 to 194 of the protein, the 823-base pair fragment coding for the full length protein and untranslated regions, and the 387-base pair fragment coding for amino acids 1-145 previously reported (Gubbin et al., 1979) have nearly identical nucleotide sequences. The 486-base pair fragment differs from the 823-base pair fragment by only three nucleotides: adenine instead of cytosine in the second base of the codon for amino acid 98 and thymine instead of cytosine for the third base of the codons for amino acids 125 and 128. The latter two differences are silent, but the first converts asparagine in the 486-base pair clone to threonine in the 823-base pair clone. Both of these amino acids are members of the same chemical subfamily. Such substitutions which commonly occur in related proteins may have little effect on the secondary or tertiary structure or function of the hormone (Dayhoff, 1978).

* M. Cronin, personal communication.
Likewise, comparison of the full length clone with the 387-base pair sequence of Gubbins et al. (1979), reveals differences in three nucleotides. These are: CTT (leucine) instead of CTC (leucine), ATT (isoleucine) instead of ATC (isoleucine), and AGC (serine) instead of GCC (glycine) at amino acids 125, 128, and 133, respectively. Among these differences, which all represent a single nucleotide change per codon, two are silent. The third one leads to a serine to glycine change which is a commonly observed substitution and which may not affect biological function significantly (Dayhoff, 1978). The change in our sequence at amino acid 133 results in the addition of a second Hae III site five bases in the 5′ direction from the single Hae III site previously reported. This change also results in the addition of an Eco RII site which is confirmed by restriction enzyme analysis. Our rPrl mRNA came from Sprague-Dawley rats while the RNA of Gubbins et al. was from Fischer rats. Hence, these differences may represent allelic polymorphism in the Prl molecule as has been observed in bovine GH (Fellows and Rogol, 1969), or the cloned molecules may represent transcripts from nonallelic genes as has been shown for rat preproinsulin (Ullrich et al., 1977). Errors in reverse transcription (Gopinathan et al., 1979) or sequencing may cannot be excluded. Of interest in this regard, two-dimensional gels of rPrl mRNA translation products have revealed two molecular forms of rPrl (Evans et al., 1978) and hybridization with rPrl cDNA to the genomic DNA of GH cells has given a calculated gene copy number of 2.4 per haploid genome (Evans and Rosenfeld, 1979). Whether the various rPrl cDNA molecules will be assigned to one or more chromosomal loci will depend upon the isolation, cloning and characterization of all genomic Prl sequences from the normal rat.

Comparison of rPrl and rGH—When the amino acid sequences of rPrl and rGH (Seeburg et al., 1977) are aligned with arbitrary gaps to maximize their identity, 51 of 209 amino acid residues are identical (Table I) and 118 more represent common substitutions of related residues (Dayhoff, 1978). This degree of identity and the prevalence of conservative differences suggest that the genes coding for the two hormones arose by duplication of a common ancestor (Bewley et al., 1972; Niall et al., 1971). Since the pituitaries of birds, reptiles, and amphibians all contain distinct Prl-like and GH-like molecules, the Prl-GH gene duplication has been estimated to have occurred before the divergence of fish and tetrapods, 300 to 400 million years ago (Acher, 1976). Another estimate of this time of divergence can be made by applying the “evolutionary clock” hypothesis. This assumes that neutral changes accumulate in a given protein at a constant rate expressed as unit evolutionary period. One unit evolutionary period is the average time in millions of years for a 1% difference in amino acid sequence to arise between lineages; for Prl, the unit evolutionary period is 5 (Wilson et al., 1977a). From the 76% identity between rPrl and rGH (Fig. 5), the Prl-GH divergence time is estimated to have occurred about 380 million years ago, in good agreement with the taxonomic estimate. This significantly predates the divergence time that we calculate for hGH and hCS at 56 million years ago, using the GH unit evolutionary period of 4 (Wilson et al., 1977a) and homology data from Martial et al. (1979).

Using the same alignment (Fig. 5), direct comparison of the nucleic acid sequences indicate a 39% incidence of identical nucleotides. As previously noted when comparing hGH to rGH and hCS (Martial et al., 1979), the nucleotide homology is greater than the amino acid homology. The drift of silent and presumably neutral mutations in the comparison of the two sequences reveals that 24 of 52 single nucleotide changes within codons (Table I) are not reflected as amino acid changes. This silent mutation frequency is 46% whereas, on a random basis, one would expect a rate of only 25%. Similar high rates of silent mutations have been reported between homologous nucleotide sequences from different organisms (Jukes and King, 1979). Such high rates of silent mutations may reflect evolutionary pressure to conserve certain important structural elements and further support the common precursor hypothesis previously derived from amino acid sequence data.

Our cloned DNA contains 52 base pairs corresponding to the 5′-untranslated region of rPrl mRNA. The number of nucleotides still missing from the complete structure is unknown. There is no striking homology with rGH mRNA in this region, and only limited complementarity with the 3′-end of eukaryotic ribosomal RNA (Hagenbichle et al., 1978). In the 5′-untranslated region of rPrl, the sequence UUCGUGU is repeated three times between 47, 18, and 9 nucleotides from the AUG initiation codon. GUGG is repeated twice in rGH mRNA, 19 and 2 nucleotides from AUG, once in hGH mRNA, 19 nucleotides from AUG, and once in hCS mRNA, 5′ nucleotides from AUG.5 GUGG is absent from the 5′-untranslated region of mRNAs for ovalbumin (McReynolds, 1978), human α-chorionic gonadotropin (Fiddes and Goodman, 1979), bovine corticotropin-β-lipotropin (bACTH-LPH; Nakanishi et al., 1979), bovine preproparathyroid hormone (Kronenberg et al., 1979), and rat preproinsulin I and II (Lommedico et al., 1979). Fig. 7 shows a possible secondary structure for this region of the rPrl mRNA which leaves the GUGGU unpaired in a hairpin loop. It is not known whether the preservation of this structure at a fixed distance from the initiation codon in these three related polypeptide hormones is significant. The short palindrome (5′-UUCUCUU-3′) in the 5′-untranslated region of rPrl mRNA is not present in rGH mRNA, but less of the 5′-untranslated sequence of the latter is known.

The 3′-untranslated regions of rPrl and rGH mRNAs do not show significant homology. In this area, rPrl lacks the regions rich in guanine and cytosine which are present in rGH, hGH, and hCS, and has a guanine and cytosine content of 36%, similar to the 42% of total rat DNA (Sueoka, 1961). The sequence 5′-UUCUUAAAGUGCUAUUUCUU-3′ located in the 3′-untranslated region of rPrl may correspond to the

5 P. Seeburg, personal communication; only 15 nucleotides of the 5′-untranslated region of hCS mRNA have been sequenced.
palindromic sequences found at a similar position in rGH, hGH, and hCS mRNAs. Symmetrical sequences of this type have been identified in other eukaryotic mRNAs (Wilson et al., 1977b; Proudfoot, 1977) but their function, if any, remains unknown.

The limited homology between both 5'- and 3'-untranslated regions of rPrl and rGH indicates that these regions have undergone evolutionary divergence faster than do the coding regions. Human CS and hGH which, as discussed above, may have diverged 56 million years ago, have 93% homology in both coding and noncoding regions; hGH and rGH which diverged about 132 million years ago have 75% homology in the coding region, 73% in the 5'- and 38% in the 3'-noncoding regions (Martial et al., 1979); and rPrl and rGH which diverged 380 million years ago have 39% homology in the coding regions and about 30% in the noncoding regions. These data are consistent with the hypothesis of Nishioka and Leder (1979), that two modes of nucleotide sequence evolution can be distinguished, a slow mode preserving sequence homology and a faster mode in areas where there are no RNA or protein sequence constraints.

Codon usage in rPrl is nonrandom, as has been found in other eukaryotic mRNAs. No consistent pattern has emerged (Martial et al., 1979), although one striking difference has been observed within the Prl set. In total rat DNA as in all vertebrate DNA, guanine and cytosine constitute 42% of the residues (Sueoka, 1961). However, in the third position of the codons for rGH, hGH, and hCS there is a strong preference for guanine and cytosine (74%, 76%, and 80%, respectively, Table II) whereas in rPrl the third codon choice is random (50%, Table II). Table II summarizes the usage of guanine and cytosine in the third position of the codons for several mRNAs of eukaryotic hormones. Only bovine preproparathyroid hormone mRNA has a lower guanine plus cytosine percentage than rPrl mRNA (Kronenberg et al., 1979). The guanine plus cytosine preference is more marked in those hormones of the Prl set thought to be of more recent origin; however, the extent to which this trend is generalized and significant is unknown.

With the results of the current study, coding sequence data for at least one gene from each of the entire Prl set is now available, and there is clear evidence for the evolutionary relatedness of Prl, GH, and CS. Although nucleic acid sequence data for all three hormone genes from one species are not yet available, the rPrl and rGH genes are clearly much more different than are the hGH and hCS genes. The biological activities of these hormones show considerable interspecies variation. Thus, it will be of interest to apply the same analysis to this set of genes in a variety of species to determine the extent of intra- and interspecies similarities in these genes.

Acknowledgments—We wish to thank Ms. Ann Forsham and Drs. Stephen Liebhaber, Walter Miller, Peter Seeburg, John Shire, and Stephen Spindler for help and advice during the course of these studies and Ms. Sonja Bock and Dr. Hugo Martinez for use of computer programs in analyzing DNA sequences.

REFERENCES

Acher, R. (1976) Ciba Found. Symp. 41, 31-59
Ben-Jonathan, N., Oliver, C., Weiner, H. J., Mical, R. S., and Porter, J. C. (1977) Endocrinology 100, 452-458
Bewley, T. A., Dixon, J. S., and Li, C. H. (1972) Int. J. Peptide Protein Res. 4, 281-287

*This number could be raised by creating gaps in the compared sequences. We prefer not to do so because of the arbitrary nature of such manipulation without the guidance of the amino acid alignment.
Prolactin cDNA Structure

H. M. (1977) Nature 270, 486–494
Seo, H., Refetoff, S., Vassart, G., and Brocas, H. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 824–828
Sherwood, L. M. (1967) Proc. Natl. Acad. Sci. U. S. A. 58, 2307–2314
Shine, J., Seiberg, P. H., Martial, J. A., Baxter, J. D., and Goodman, H. M. (1977) Nature 270, 494–499
Shone, B., and Farlow, A. F. (1977) J. Clin. Endocrinol. Metab. 45, 1112–1115
Southern, E. M. (1975) J. Mol. Biol. 98, 503–517
Sueoka, N. (1961) J. Mol. Biol. 3, 31–40
Takahara, J., Arimura, A., and Schally, A. V. (1974) Endocrinology 95, 462–465
Tinoco, I., Borer, P. N., Dengler, B., Levine, M., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. (1973) Nature New Biol. 246, 40–41
Ulirsch, A., Shine, J., Churgin, J., Pictet, R., Tischer, E., Rutter, W. J., and Goodman, H. M. (1977) Science 196, 1313–1319
Vale, W., Grant, G., and Guillemin, R. (1973) in Frontiers in Neuroendocrinology (Ganong W. F., and Martini, L., eds) pp. 375–413, Oxford Press, New York
Wilson, A. C., Carlson, S. S., and White, T. J. (1977a) Annu. Rev. Biochem. 46, 573–639
Wilson, J. T., deRiel, J. K., Forget, B. G., Marotta, C. A., and Weissman, S. M. (1977b) Nucleic Acids Res. 4, 2353–2368