Human Prolactin  
cDNA STRUCTURAL ANALYSIS AND EVOLUTIONARY COMPARISONS*

(Received for publication, September 8, 1980, and in revised form, December 1, 1980)

Nancy E. Cooke, Doris Coit, John Shine, 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, and chorionic somatomammotropin form a set (the "Prl set") of hormones which is thought to have evolved from a common ancestral gene. This assumption is based on several lines of evidence: overlap in their biological and immunological properties, similarities in their amino acid sequences, and homologies in the nucleic acid sequences of their structural genes. In the current study we report the cloning, amplification in bacteria, and sequence analysis of DNA complementary to Prl mRNA isolated from human pituitary Prl-secreting adenomas. The cloned DNA contains 914 bases, which includes the entire coding sequence of human prePrl as well as portions of the 5' and 3'-untranslated regions of the mRNA. The amino acid sequence predicted by our data differs from a previously reported amino acid sequence in 8 positions. With the results of this study we can now compare in one species the nucleotide sequences of the structural gene coding for each of the hormones of the Prl set. The sequence divergence at replacement sites may differ from those for GH. While prolactin is mainly from normal pituitaries,2 this cloned hPrl cDNA should be used as a probe to map and isolate the corresponding hPrl genomic sequences, to determine the chromosomal location of the hPrl gene (Overbach et al., 1980b), to synthesize hPrl in bacteria, and to study the regulation of Prl gene expression. The latter is of particular interest, since the mechanisms controlling Prl gene expression may differ from those for GH. Whereas prolactin is primarily under negative dopaminergic regulation (Takahara et al., 1974; MacLeod, 1976; Ben-Jonathan et al., 1977), GH may be predominantly under positive control (Kulich et al., 1968; Frohman et al., 1971).

EXPERIMENTAL PROCEDURES

mRNA Preparation—Human pituitary Prl-secreting adenomas were collected by transphenoidal hypophysectomy, frozen immediately, and stored in liquid nitrogen. When six tumors had been collected, RNA was prepared from each according to the method of Chirgwin et al. (1979) with modification as described previously (Cooke et al., 1980). The RNA purification included enrichment by chromatography on oligo(dT)-cellulose (P-L Biochemicals, Type 7).

mRNA Translation and Immunoprecipitation—The polyadenylated RNA was translated in an in vitro rabbit reticulocyte lysate translation system in the presence of 35Smethionine (Amersham, 600–1300 Ci/mmol) as previously described (Delham and Jackson, 1967). Human prePrl synthesized in the cell-free system was immunoprecipitated with rabbit antiserum to hPrl (a gift of H. Friesen, Department of Physiology, University of Manitoba). Control immunoprecipitation using rabbit antiserum to hGH (Antibodies, Inc., Davis, CA) or normal rabbit serum was also carried out. Immune complexes were precipitated with protein A of the Cowan strain of Staphylococcus aureus (Kessler, 1975; Martial et al., 1977). 35S-labeled proteins were electrophoresed on sodium dodecyl sulfate-10% polyacrylamide slab gels (Laemmli, 1970) at 20 mA/gel, fixed, washed, vacuum dried, and exposed to x-ray film (Kodak, NS-2T) by standard methods (Cooke et al., 1980).

cDNA Synthesis and Construction of Recombinant Plasmids—The mRNA from four of the tumors (about 19 mg) was pooled and used as template for synthesis of cDNA as described previously (Seeburg et al., 1977). Avian myeloblastosis reverse transcriptase was received from Research Resources, National Cancer Institute.

M. John, H. Friesen, and H. Niall, personal communication.

Manuscript in preparation.
Human Prolactin cDNA

Bethesda, MD, and S-1 nuclease from Miles Laboratories. The double-stranded cDNA was tagged with dCMP and the Pst I-cleaved pBR322 was tagged with dGMP (Rochehodhury et al., 1976) using deoxyribonucleotidyltransferase (P-L Biochemicals). The annealing of the tagged cDNA and plasmid were as described (Cooke et al., 1980).

Restriction enzymes were obtained from New England Biolabs, Beverly, MA, and Bethesda Research Laboratories, Bethesda, MD.

Transformation and Colony Screening. An aliquot (20%) of the double-stranded cDNA was annealed to the Pst I-cleaved and dGMP-tailed pBR322. Transformation of Escherichia coli x1776 was carried out in compliance with NIH guidelines. There were 232 colonies, of which 69 were ampicillin-resistant. All of the colonies were replica-plated onto Whatman 540 paper (Grunstein and Hognes, 1975) and screened with a 32P-labeled cDNA probe (Cooke et al., 1980), made using mRNA from tumor 5 as template. Sixteen prominently hybridizing colonies were selected, plasmids were isolated from them, cleaved with Pst I, and analyzed by agarose gel electrophoresis and Southern hybridization (Southern, 1975) using the same cDNA probe. The colony that contained the largest hybridizing insert and that was selected for nucleotide sequence analysis was both tetracycline- and ampicillin-resistant. After positive identification of the clone by sequence analysis, the hPrl-containing plasmid was isolated and transformed into E. coli RR1.

DNA Sequence Determination—After restriction enzyme mapping of the hPrl cDNA clone, appropriate restriction fragments were selected and labeled at their 5' ends with T-4 polynucleotide kinase (P-L Biochemicals) and [γ-32P]ATP (ICN, 4200-4500 Ci/mmole) after removal of the 5'-phosphate groups with bacterial alkaline phosphatase (Enzo Biochemicals, Inc., New York, NY), or at their 3'-ends with the Klenow fragment of DNA polymerase I (New England Biolabs) and the appropriate [α-32P]dNTP (Amersham, ~5000 Ci/mmole). The chemical cleavage technique of Maxam and Gilbert (1977) was used with one change. The adenine modification reaction contained 25 μM of formic acid (Mallinckrodt) and 10 μM of 32P-labeled DNA restriction fragments. Reaction was for 10 min at 19°C. The reaction was stopped by addition of 200 μl of 0.3 M sodium acetate, 0.1 mM EDTA, and 25 μg/ml of T1 DNA. DNA fragments were separated on polyacrylamide thin gels (8 and 15%) containing 7 M urea (Sanger et al., 1977).

RESULTS

Isolation and Identification of hPrl-enriched mRNA—RNA was isolated from six human pituitary Prl-secreting adenomas; 100 to 270 μg of RNA were obtained from each tumor. The total RNA from each tumor was passed through oligo(dT)-cellulose to enrich for polyadenylated species. This yield was not quantitated in order to minimize losses. The relative abundance of hPrl mRNA from each tumor was analyzed by cell-free translation, polyacrylamide gel electrophoresis, and autoradiography of the protein products. A prominent band at about 26,000 daltons was seen in each translation that contained tumor mRNA (Fig. 1A). Other prominent bands synthesized in the translation were endogenous to the reticulocyte lysate system. The band at 26,000 daltons was specifically immunoprecipitated by antibody to hPrl (Fig. 1B). This band was not precipitated by antisera to hGH or by normal rabbit serum, indicating that the predominant protein produced by the polyadenylated RNA from these pituitary tumors was human prePrl.

Double-stranded cDNA Synthesis and Cloning—The mRNA from four of the six tumors (1 through 4, Fig. 1A) was pooled and used as template for the synthesis of a single-stranded DNA. Polyacrylamide gel electrophoresis of the DNA indicated that a high molecular weight species had been produced (Fig. 2A). When the single-stranded cDNA was cleaved with Hae III as described previously (Seeburg et al., 1977b), three prominent bands were produced (Fig. 2A). This suggested that the synthesized single-stranded cDNA reflected an enrichment for a unique mRNA species in the tumors, presumably hPrl.

The second strand of DNA was synthesized using reverse transcriptase. Following this, the resulting hairpin loop was digested by S-1 nuclease. Fig. 2B shows the Hae III digestion of this double-stranded DNA. Comparison of lanes 1 and 2 (before and after S-1 nuclease digestion) confirms that most of the DNA was double-stranded. The single-stranded background decreased with S-1 digestion, but the major bands were resistant to S-1 digestion and their sizes are consistent with the Hae III restriction fragments of the final cloned cDNA (Fig. 3B).
A strand of dCMPs was added to the 3'-ends of the cDNA, and it was then annealed to plasmid DNA, which had been previously digested with Pst I and tailed with dGMPs. When a portion of this recombinant plasmid was transformed into the E. coli strain λ776, 232 tetracycline-resistant colonies were produced. The 16 colonies hybridized prominently with a cDNA probe made from the mRNA of tumor 5 (Fig. 1A). Plasmid DNA was isolated from each of these colonies, cleaved with Pst I, and analyzed by gel electrophoresis. The largest insert, which contained 959 base pairs (including synthetic dGMP/dCMP tails), was selected for nucleotide sequence analysis.

**Nucleotide Sequence of hPrl cDNA**—The sequencing strategy used is shown in Fig. 3A. Both message and anti-message untranslated regions of the corresponding Prl mRNA. B, the restriction sites predicted by the DNA sequence were checked by restriction enzyme digestion followed by sizing on polyacrylamide gels. In some instances restriction sites indicated in A were not copied onto B to minimize crowding, although the digestions were performed. Both A and B are drawn to scale as indicated by the calibration in base pairs at the bottom of the figure.

**Fig. 3. Sequencing strategy and restriction map of hPrl cDNA.** A, restriction fragments were labeled at the site of the solid circles and DNA sequencing proceeded in the direction of the arrowheads for the distance indicated by the length of the arrow. The 5'-end of the cDNA was sequenced in the 5' to 3' direction after labeling at the Alu I site in pBR322 (indicated by dashed line). The dotted bar indicates the coding region of the secreted hormone, the white bar indicates the signal peptide, and the solid lines indicate the untranslated regions of the corresponding Prl mRNA. B, the restriction sites indicated on the DNA sequence were checked by restriction enzyme digestion followed by sizing on polyacrylamide gels. In some instances restriction sites indicated in A were not copied onto B to minimize crowding, although the digestions were performed. Both A and B are drawn to scale as indicated by the calibration in base pairs at the bottom of the figure.

**Fig. 4. Autoradiographs of 32P-labeled DNA fragments generated by Maxam-Gilbert cleavage reactions.** A, two representative 15% polyacrylamide sequencing gels are shown in which the DNA was 3' labeled at the Ava II site that occurs between amino acids-19 and -18 of the signal peptide. The bottom band of the left sequence begins with the third nucleotide of amino acid-19 and proceeds in the 5'-direction on the coding strand, while the right sequence begins at the second nucleotide of amino acid-18 and proceeds in the 3'-direction on the noncoding strand. B, a segment of an 8% polyacrylamide sequencing gel generated from DNA labeled at the Pvu II site shows the first base of the codon for amino acid 79 and continues through amino acid 95.
The hPrl mRNA sequence and the derived amino acid sequence are shown in Fig. 6. If this assignment of amino acids is correct and if translation begins at the Met codon at position –28, then the primary translation product of hPrl mRNA is a protein of 25,880.03 daltons. The sequence of 227 amino acids includes the 28 amino acid signal peptide of hPrl and an additional amino acid, Ser, at position 86, which was not reported previously (Shome and Parlow, 1977). The sequence includes all of the 3′-untranslated region, the AAUAAA seen in most eukaryotic mRNAs, and a part of the poly(dAMP) tail. Our DNA sequence, however, is lacking most of the 5′-untranslated region, possibly due to incomplete synthesis of the first strand or to digestion of the cDNA corresponding to this region with S-1 nuclease following preparation of the second strand. By comparison, the 5′-untranslated region in the analogous rPrl mRNA is at least 51 base pairs in length (Cooke et al., 1980). The codon selection for hPrl is nonrandom, as it is in rPrl. In contrast to rPrl, however, a 63% preference is shown for dGMP or dCMP in the third position of the codons.

The orientation of hPrl in the plasmid was determined by cleavage of the plasmid with HindIII. There is one HindIII site in the insert and one in the plasmid. The fragments predicted for an orientation of 3 to 5′ (3801 and 1529 base pairs) were observed (data not shown). This places the hPrl cDNA within and in the same transcriptional orientation as the β-lactamase gene.

Comparisons of Prl, GH, and CS—The sequences of the related human polypeptide hormones Prl, GH, and CS were aligned to emphasize their homologies. As shown in Fig. 7, arbitrary gaps were introduced into the sequences to maximize their nucleotide and amino acid identity and also the number of conservative amino acid replacements (Dayhoff, 1978). Since CS and GH are highly homologous, only those positions at which CS differs from GH are indicated. No attempt was made to introduce gaps in the untranslated areas. Fig. 8 shows a similar alignment of hPrl and rPrl in which identical amino acids have been shaded to emphasize homologous areas. Since rat prePrl is only 225 amino acids in length, two amino acids shorter than human prePrl, a gap was introduced after amino acids 28 and 29.

![Fig. 5. Restriction digests of hPrl cDNA displayed on a 5% polyacrylamide gel stained with ethidium bromide. A, the plasmid containing the hPrl cDNA cleaved with Pst I (1) is displayed next to the molecular size marker pBR322 cleaved with Hpa II (2). From top to bottom the marker fragments are: 622, 527, 403, 309, 242, 238, 217, 201, 190, 180, 160, 147, 122, and 110 base pairs. B, the plasmid containing hPrl cDNA was cleaved with Pst I and the two hPrl cDNA fragments were isolated from a 5% polyacrylamide preparative gel.](image-url)

**Fig. 6.** Nucleotide sequence and predicted amino acid sequence of the mRNA coding for hPrl. The amino acid sequence is deduced from the mRNA sequence using the genetic code. OC, the ochre terminator.
Human Prolactin cDNA

Human Prolactin
Human Growth Hormone
Human Chorionic Somatomammotropin

Human Prolactin

[Translated text follows, containing detailed information about the nucleotide and amino acid sequences of prolactin, growth hormone, and chorionic somatomammotropin (CS). The text includes discussions on the alignments, homologies, and evolutionary speculations between the sequences of these hormones.]

**DISCUSSION**

We have presented the nucleotide sequence of a 914 base pair-cloned DNA fragment containing the hPrl structural gene sequence. The nucleotide sequence of this cDNA enabled us to predict the amino acid sequence of the human preprolactin signal peptide and to revise the known sequence of the secreted protein. Comparison of the hPrl sequence to the other related human hormones, GH and CS, has permitted us to make some evolutionary speculations concerning the protein-coding areas of these three genes.

**RNA Isolation, cDNA Cloning, and Nucleotide Sequence**

The mRNA sequence for the human prolactin gene was isolated from poly(A) RNA by oligo(dT) cellulose chromatography. The cDNA was synthesized from the human prolactin mRNA using the reverse transcriptase from E. coli. The cDNA sequence was determined by the dideoxy method. The sequence of the cDNA was analyzed using the genetic code and the amino acid sequences were deduced from the mRNA sequences using the genetic code. The hCS sequence is indicated only where it differs from the hGH sequence. The ochre terminator is indicated by UAG. The gaps were arbitrarily introduced to maximize homology.
Human Prolactin cDNA

**Analysis**—Polyadenylated RNA was isolated from hPrl-secreting pituitary adenomas. The cell-free translation products of this mRNA were highly enriched in prePrl. This enrichment of hPrl mRNA in pituitary tumors was remarkable in that it appeared to exceed that which we were able to produce previously in rats by estrogen implantation and hypothalamic ablation (Cooke et al., 1980). This assumes, of course, that both Prl mRNAs were translated with equal efficiency in the cell-free system. In the translation of mRNA from human pituitary tumors, no hGH was seen either in the total translation or by immunoprecipitation using hGH antiserum. Although it is possible that the presence of some preGH is obscured by the width of the prePrl signal or by the limited exposure of the gel to x-ray film, these data may indicate that...
the tumor is not synthesizing GH. In earlier work the mRNA from a Prl-secreting tumor was transcribed and did contain a small amount of detectable GH mRNA, although it was not determined whether this GH mRNA originated in the tumor tissue or from adjacent normal pituitary tissue (Martial et al., 1979).

The mRNA from these human tumors, naturally enriched in Prl mRNA, yielded a cDNA highly enriched in Prl sequences. Initial attempts to screen the bacterial colonies containing recombinant plasmids with a probe made from cloned rPrl CDNA (Cooke et al., 1980) were unsuccessful. Even under conditions of low stringency of hybridization, rPrl CDNA did not cross-hybridize with hPrl sequences. In retrospect, this result was unexpected, since the nucleotide sequence homology of rat and human Prl is 73%.

The accuracy of the sequence was documented by analyzing both strands of the hPrl cDNA completely and by verifying the presence of most of the predicted restriction sites. Errors in reverse transcriptase synthesis of the first DNA strand (Gopinathan et al., 1979; Fagan et al., 1980) cannot be excluded.

The clone containing hPrl CDNA that was ultimately studied was both tetracycline- and ampicillin-resistant (Villa-Komaroff et al., 1978). Resistance to both antibiotics was retained after retransformation of the hPrl-containing plasmid into the RRI strain of E. coli. Since only one plasmid species was visible on gel electrophoresis, these results suggest that the ampicillin resistance was not due to co-transformation with a wild type plasmid carrying an intact β-lactamase gene. Restriction endonuclease mapping showed that the CDNA was inserted in the transcriptional direction of β-lactamase. Attempts to sequence from the Alu I site just 5' to the Pst I site of the β-lactamase gene into the inserted cDNA suggested that the hPrl CDNA was not in the β-lactamase reading frame. However, determination of the exact length of the long 5'-dGMP tail was difficult due to base stacking and compressions on the gels. Immunoprecipitation with hPrl antiserum of a 45,000-dalton protein from the extracts of E. coli infected with the hPrl recombinant plasmid suggests that a hybrid of β-lactamase and hPrl is produced by this clone. Thus, the retention of ampicillin resistance by the hPrl recombinant plasmid might be explained if the fusion protein retained β-lactamase activity.

The signal peptide sequence of hPrl has not been previously reported. Our sequencing data (Fig. 4A) has shown that it is 28 amino acids in length. It is identical with the signal peptide of rPrl in 16 positions (Fig. 8), including a hydrophobic block of 4 Leu residues. The secreted protein contains 199 amino acids, one more than previously reported by amino acid sequencing techniques (Shome and Parlow, 1977). In addition, these sequences disagree at eight positions. At positions 82, 83, 85, 86, 105, 144, 162, and 163, we report Ser, Leu, Val, Ser (the additional amino acid) (Fig. 4B), Met, Asp, Glu, and Ser, while Shome and Parlow (1977) report Val, Ser, Leu, no amino acid, Asx, Asp, Ser, and Glu, respectively. Partial revision of the amino acid sequence of hPrl is in complete agreement with our results at positions 105, 144, 162, and 163; however the chromatographic peptide at 81–89 was not isolated or resequenced, so the discrepancies at positions 82–86 were not resolved.

The reported nucleotide sequence represents a single hPrl molecule produced by a Prl-secreting pituitary adenoma, whereas the amino acid sequence represents the predominant Prl sequence in pools of normal human pituitaries. Consequently, the degree of identity between these sequences is striking and might suggest that both adenomatous and normal pituitaries can produce identical or nearly identical Prl molecules. We cannot determine, however, if the molecule that we sequenced was produced by the adenoma or by the surrounding normal pituitary tissue, although adenomatous tissue predominated in the samples used.

The codons used in hPrl mRNA are nonrandom (data not shown) as found in all eukaryotic mRNAs sequenced to date. The codon choices of hGH (Martial et al., 1979) and hCS (Shine et al., 1977) mRNAs are more similar to each other than either of them is to hPrl mRNA. It seems unlikely that differences in codon choice reflect evolutionary pressures to adapt to different tRNA populations. First, almost all of the codons are used in all of the genes of the Prl set. Second, GH and Prl genes are expressed in the same tissue in vivo and can be expressed by the same cell (the rat GH cell line; Dannies and Tashjian, 1973). The differences in codon choice between GH and Prl contrast with the similarities in the codon choices of GH and CS, which are expressed by different tissues. An analogous situation exists in the globin gene family. All of the globins have a nonrandom and unique distribution of codon choices, yet are expressed either sequentially (α, β) or simultaneously (α, β) in the same tissue. Consequently, no relationship between expression and tRNA availability has been shown in this family (Wilson et al., 1980). In the Prl set of genes there is a spectrum of preference for guanosine plus cytosine in the third position of the codons: bovine GH, 82% (Miller et al., 1980a); hCS, 80%; hGH, 76%; rGH, 74%; hPrl, 63%; and rPrl, 50% (Cooke et al., 1980). Codons that end with dUMP or dCMP and codons that end with dAMP or dGMP code for the same amino acid (Met and Trp are exceptions). Consequently, a bias in the use of dGMP and dCMP in the third position of codons cannot be explained on the basis of evolutionary constraints against amino acid substitution. This suggests that synonymous codons might be expected to be neutral in terms of evolution, and as discussed, this appears to be the case.

**Molecular Evolution of the Prolactin Set of mRNAs—** A direct comparison of the sequences of the Prl family of genes (Table 1), aligned as shown in Figs. 7 and 8, shows a greater nucleotide than amino acid sequence homology in each case. In addition, very similar amino acid and nucleotide homologies were obtained in the interspecies comparisons of Prl and GH (i.e. hPrl with rPrl, and hGH with rGH). The same parallelism in homologies was observed in the intraspecies comparisons of Prl and GH (i.e. rPrl with rGH, and hPrl with hGH). There is a very high degree of homology between hGH and hCS, as has been noted previously (Martial et al., 1979).

There can be two types of nucleotide differences between genes when they are compared codon by codon. The first is a difference that results in an amino acid in nonrandom (replacement substitutions); the second results in a synonymous codon (silent substitutions). The percentage of silent substitutions in each comparison in Table I (two identical nucleotides/codon) is greater than the 25% that would be expected when comparing random sequences (Jukes and King, 1979). This may reflect the evolution of these genes from a common ancestor.

Surprisingly, the hGH and hCS cDNAs, which have the highest nucleotide sequence homology of the set (92%), also have the lowest ratio of silent to expressed single base substitutions, 0.45 (31%/69%, Table I). In contrast, the hPrl and rPrl genes with only 73% nucleotide homology have a silent to expressed ratio of 1.2 (54%/46%, Table I). These results are consistent with the hypothesis that positive selective influences have caused rapid fixation of replacement substitutions in the genes of these recently diverged hormones (hGH and hCS).

To improve the estimates of divergence among the members of the Prl family of genes, the methods of Ferler et al. (1980)
Amino acid and nucleotide sequence comparison within the Prl family

The results are derived from the sequence alignments in Figs. 7 and 8. Codons corresponding to gaps were excluded from the calculations, as were the termination codons. Only coding sequences have been compared.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Amino acids</th>
<th>Nucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Identical</td>
<td>Conservative differences</td>
</tr>
<tr>
<td></td>
<td>% No.</td>
<td>% No.</td>
</tr>
<tr>
<td>hPrl/rPrl</td>
<td>63 (38)</td>
<td>73 (38)</td>
</tr>
<tr>
<td>hPrl/rGH</td>
<td>65 (38)</td>
<td>77 (38)</td>
</tr>
<tr>
<td>rPrl/rGH</td>
<td>26 (38)</td>
<td>54 (38)</td>
</tr>
<tr>
<td>hPrl/hCS</td>
<td>27 (38)</td>
<td>54 (38)</td>
</tr>
<tr>
<td>hGH/hCS</td>
<td>85 (38)</td>
<td>72 (38)</td>
</tr>
</tbody>
</table>

*Conservative amino acid substitutions were calculated according to Dayhoff (1978) and include only the most common two categories of substitutions.

**Results** are in agreement with data from Perler et al., 1980.

and Efstratiadis et al. (1980) were used. This technique assumes that nucleotide substitutions occur in a Poisson distribution and that transitions (pyrimidine to pyrimidine or purine to purine substitutions) and transversions (purine to pyrimidine or pyrimidine to purine) are equally probable. It attempts to determine the mutation rate more accurately by considering all of the possible ways that codon differences could have arisen and assigns a probability to each of these. The possibility of multiple events at a single site enables the divergence to be greater than 100%. The results of this correction on both silent and replacement sites are displayed in Table II.

To establish an evolutionary clock for the Prl family of genes, the corrected divergence at replacement sites for each pair of genes has been plotted against divergence time estimated from the fossil record (Fig. 9). Silent site diversions were similarly calculated and plotted. The divergence between human and rat hormones was assumed to occur during the mamalian radiation about 85 to 100 million years ago (Romero-Herrera, et al., 1973; McKenna, 1969). The divergence between Prl- and GH-like genes (including CS) was assumed to occur about 400 million years ago, when fish and tetrapods diverged, because the existence of distinct Prl-like and GH-like hormones has been dated to this time (Acher, 1976). The straight line (R) drawn through the origin and the replacement sites serves as the molecular clock for the Prl gene family. Its slope gives the rate of fixation of replacement substitution in this family. The reciprocal of this slope gives the rate of fixation of replacement substitution in a human GH and CS comparison, we arrive at a divergence time of 10 million years ago. This is very different from previous estimates of 56 million years ago (Martial et al., 1979), and would date the divergence long after the mammalian radiation began, and long after the first presumed requirement for specific placental hormones. This paradox might be explained as follows. A mechanism for correcting one gene against another during evolution has been postulated based on a structural analysis of the human fetal y-globin genes (Slightom et al., 1986). A homologous but unequal cross-over may occur between two such mutually correcting genes ("concerted evolution," Zimmer et al., 1980), and a calculated divergence time will date this correction event instead of the original gene duplication. Since the human GH and CS genes reside on chromosome 17 (Owerbach et al., 1980a), and since a human GH gene has been shown to be linked to a CS gene on a single DNA fragment isolated from a human DNA library (Goodman et al., 1980), it might be postulated that the 10 million year divergence time dates an exchange of DNA between the GH and CS loci on sister chromatids during meiosis rather than the original gene duplication. Dating of the original gene duplication by comparison of these two sequences would be more difficult and perhaps impossible if this is true.

Since hPrl and hGH are located on separate chromosomes (6 and 17, respectively), they are less likely to exchange DNA in this manner. Therefore, the 392 million year divergence time between these two genes may more accurately reflect their initial duplication than does the 10 million year divergence between hGH and hCS.

The rate of silent site divergence in the Prl family may be described by a biphasic curve (Fig. 9) as has been reported for
subset of rapidly diverging sites is not known. It should be resulting in an amino acid change with no accompanying substitutions compared to silent substitutions in the hGH and hCS divergence may be an example of such an acceleration of substitutions compared to silent substitutions in the hGH and hCS (see text).


It is clear that the silent sites diverge rapidly at first. They may then continue to diverge at a slower rate. The nature of the subset of rapidly diverging sites is not known. It should be borne in mind that the accuracy of both replacement and silent replacements. These examples suggest that positive selection can rapidly fix replacement mutations without being accompanied by random silent or neutral mutations. Finally, in an unequal cross-over between two correcting genes, dating based on sequence homology will also reflect the extent of the correction. The derived divergence data may therefore be a composite of the dates of divergence and of last correction of the various portions of the genes.

The 3'-untranslated area of hPrl, when compared without gaps to the 3'-untranslated area of rPrl, is identical in 43 of the 94 bases compared (54.3% divergence). The area containing the sequence AAUAAA found in most eukaryotic mRNAs and its surrounding sequences is not available in the rPrl sequence for comparison. Homology in the 3'-end of the adult β-globin family of genes (Efstratiadis et al., 1980) has been highest near this signal, so our 3'-end divergence may be an underestimate. One might speculate that some constraint on drift within this area exists for the Prl molecule. However, in general, no such constraints have been found (Efstratiadis et al., 1980). Comparing hGH and rGH directly without gaps, a 61.8% divergence is found. It is clear by inspection that the homology can be increased by creating appropriate gaps. Human Prl and GH have diverged 87.5%, hPrl and hCS 87.2%, while hGH and hCS only 6.4% in the 3'-untranslated area. It is impossible to compare these results to the silent site replacements within the coding portions of the molecule, since the correction for multiple events within each site cannot be performed. However, compared to the nucleotide identity in Table I, it would appear that the 3'-untranslated areas in each case (except in the comparison of hGH and hCS) have diverged more than the coding areas. This is consistent with results found in comparisons of other eukaryotic genes (Heilig et al., 1980; van den Berg et al., 1978; Bell et al., 1980).

REFERENCES


Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294-5299


Dickerson, R. E. (1971) J. Mol. Biol. 61, 36-45


Fig. 9. Curves for divergence of coding sequences among the Prl set of genes. The corrected percentage of divergences at silent (S) and replacement (R) sites for comparisons between hPrl and hGH (●), hPrl and rGH (△), hGH and rGH (□), and hGH and hCS (○) are taken from Table II. These are plotted against divergence times in millions of years (MY) which are minimal estimates derived from the fossil record (see text for discussion and references). The solid lines (S and R) were drawn through the points. Line R was passed through the origin. The reciprocal slope of R gives an estimate of the unit evolutionary period (UEP, the time in millions of years for a 1% sequence divergence to arise; Dickerson, 1971; Wilson et al., 1977) for replacement substitutions within the Prl set of genes. Line S does not extrapolate through the origin. In analogy with the preproinsulin (Perler et al., 1980) and globin (Efstratiadis et al., 1980) gene families, we presume that the silent sites fixed mutations with a higher rate (dashed line) for approximately 85 to 100 millions of years. The arrows indicate the derived time of last correction between hGH and hCS (see text).
Human Prolactin cDNA


MacLeod, R. M. (1975) Nature 255, 7521-7524


