Characterization of a human cDNA encoding a widely expressed and highly conserved cysteine-rich protein with an unusual zinc-finger motif

Stephen A.Liebhaber[1,2,3]*, John G.Emery[1,2,3], Margrit Urbanek[2], Xinkang Wang[3,4] and Nancy E.Cooke[2,3]

[1]Howard Hughes Medical Institute, Departments of [2]Human Genetics, [3]Medicine and [4]Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

Received March 27, 1990; Revised and Accepted June 5, 1990 Genbank accession no. M33146

ABSTRACT

A human term placental cDNA library was screened at low stringency with a human prolactin cDNA probe. One of the cDNAs isolated hybridizes to a 1.8 kb mRNA present in all four tissues of the placenta as well as to every nucleated tissue and cell line tested. The sequence of the full-length cDNA was determined. An extended open reading frame predicted an encoded protein product of 20.5 kDa. This was directly confirmed by the *in vitro* translation of a synthetic mRNA transcript. Based upon the characteristic placement of cysteine (C) and histidine (H) residues in the predicted protein structure, this molecule contains four putative zinc fingers. The first and third fingers are of the C_4 class while the second and fourth are of the C_2HC class. Based upon sequence similarities between the first two and last two zinc fingers and sequence similarities to a related rodent protein, cysteine-rich intestinal protein (CRIP), these four finger domains appear to have evolved by duplication of a pre-existing two finger unit. Southern blot analyses indicate that this human cysteine-rich protein (hCRP) gene has been highly conserved over the span of evolution from yeast to man. The characteristics of this protein suggest that it serves a fundamental role in cellular function.

INTRODUCTION

A wide variety of DNA and RNA binding proteins have now been structurally characterized. Comparisons of their sequences have suggested certain shared features in their structural and functional domains (1). One prominent structural motif results from the coordination of a zinc ion by sets of cysteine and/or histidine residues. This structure organizes a short segment of the protein into a discrete structural domain referred to as a zinc-finger (2). This structural domain was first identified and characterized in Xenopus laevis transcription factor IIIA (TFIIIA) (3,4). Protein segments closely related to the structure of the TFIIIA zinc-fingers have been noted in a number of transcriptional control proteins (1). In addition, variations of this structural motif have now been described in a variety of nuclear proteins (5,6). While the structural importance of zinc in each of these proteins may not be the same as in TFIIIA (7) and the presence of a zinc-coordinated finger in these domains can only be inferred (2,3,8), the structural similarities suggest shared functions through nucleic acid binding. Furthermore, since certain zinc-finger motifs may be conserved in proteins with similar functions, the identification of a previously reported zinc-finger motif in an anonymous protein suggests similar potential functions for that protein (5).

We now report the isolation and characterization of a cDNA encoding a human cysteine-rich protein which contains four putative zinc-fingers. Several features of this protein's primary structure, similarity to previously reported proteins, generalized tissue distribution, and marked conservation from yeast to man suggest that it may play a fundamental role in cellular function.

MATERIALS AND METHODS

DNA Modification, Labeling, and Sequencing

Restriction and modification enzymes were purchased from New England Biolabs (NEBL, Beverly, MA) and Bethesda Research Laboratories (BRL, Gaithersburg, MD) and were used according to the manufacturer's specifications. The cloned DNAs used as probes in these studies include human prolactin (hPrl) cDNA (9), an 18S ribosomal genomic probe pB (10), a kind gift of J Sylvester (Hahnemann University), and the A4 and 5F cDNAs (present report, see Fig. 1). Each cloned insert was released from its plasmid vector by digestion with the appropriate restriction enzyme: hPrl by PstI, the 18S rRNA, A4, and 5F by EcoRI. The inserts were gel purified prior to labeling. Probes were labeled using DNA polymerase by priming with random hexamer oligonucleotides (Boehringer Mannheim (BMB), Indianapolis, IN) in the presence of [α-^{32}P]ATP to a specific activity of about

* To whom correspondence should be addressed at Howard Hughes Medical Institute, Department of Human Genetics, University of Pennsylvania School of Medicine, 422 Curie Boulevard, Philadelphia, PA 19104-6145, USA

1×10^8 cpm/ug. Oligonucleotides were synthesized by the DNA synthesis core facility of the Cancer Center at the University of Pennsylvania. Each oligonucleotide was sequenced prior to use. For primer extension analysis and the polymerase chain reaction, oligonucleotides were labeled at their 5'-ends with [γ^{32}P]ATP using T4 polynucleotide kinase (NEBL) followed by purification from unincorporated label by passage through a Sephadex G-25 spin column (BMB). DNA sequencing was carried out by the chemical degradation method (11) on double-stranded restriction fragments that had been [^{32}P]-labeled at either 5'- or 3'-ends. Alternatively, the full-length cDNAs as well as selected restriction fragments were subcloned into M13mp19 or M13mp18 (12) followed by dideoxy DNA sequencing (13) using either universal oligonucleotide primers or specific oligonucleotide primers that hybridized to internal regions of the cDNAs. All sequences were determined on both strands.

Library Screening

A human placental cDNA library in λgt11, constructed from the total RNA of a full-term gestation human placenta, was the kind gift of M. Weiss (University of Pennsylvania) (14). This library was screened by hybridization to [^{32}P]-hPrl cDNA under low and high stringency wash conditions. Hybridizations were carried out at 65°C in standard hybridization buffer lacking formamide (15). Low stringency washes were done in 2× SSC for 30 min at room temperature three times, while high stringency washes were done in 0.1× SSC for 30 minutes at 65°C three times. Plaques positive after low stringency washing but negative after high stringency washing were identified by *in situ* hybridization (16). Selected positives plaques were purified by sequential plating and rehybridization to the hPrl probe at low density until every plaque was positive. DNA was isolated from confluent plate lysates of plaque-purified phage by standard methodology (17). cDNA inserts were released from λgt11 by EcoRI digestion and were sized on an 0.8% agarose gel. These inserts were subcloned into M13mp18/19 for dideoxy DNA sequence analysis and into pGEM3 in both orientations (PB, Promega Biotech, Madison, WI) for further reconstruction, transcription, and Maxam-Gilbert sequencing.

Primer extension

Polyadenylated RNA was isolated from the chorionic layer of a human term-gestation placenta by oligo(dT)-cellulose chromatography (18). 500 ng of antisense oligonucleotide complementary to bases 95−116 of hCRP (Fig. 1 and 2A) (5'-CTTCGCACTGAACCTCTTCG-3') were labeled at their 5'-ends with [γ^{32}P]ATP. The labeled oligonucleotide was mixed with 5 ug of the poly A+ mRNA and avian myeloblastosis virus reverse transcriptase (Life Sciences) at 45°C for 1 hour under the conditions previously described (19). The reverse transcriptase reaction was terminated by phenol/chloroform extraction buffered by 50 mM Tris pH 8.6, followed by ethanol precipitation. The size of the extended primer was determined by electrophoresis through an 8 M urea, 6% polyacrylamide sequencing gel. A dideoxy sequencing ladder of a characterized cDNA was loaded as size marker in an adjacent lane. The gel was dried and autoradiographed at −70°C with a Cronex Lightning-Plus intensifying screen.

Northern and Southern Analyses

Total genomic DNA was isolated from *Saccharomyces cerevisiae*, *Schistosoma mansoni*, *Drosophila melanogaster*, chicken primary myoblasts (respectively gifts from J. Kuncio, G. Guild, and J. Choi, University of Pennsylvania), mouse A9 cells, and human peripheral leukocytes (15). 10 ug DNA from each organism were digested with EcoRI, resolved on an 0.8% agarose gel, and transferred to nylon membranes (Zetabind, AMF, Meriden CT). Southern hybridizations were carried out at 68°C overnight in 0.5 M Na_2HPO_4, 7% SDS, 1% BSA, 1 mM $(Na)_2EDTA$, 100 mg/ml boiled salmon sperm DNA with 2×10^6 cpm/ml denatured probe. High stringency washes were in 0.1× SSC, 0.1% SDS at 60°C for one hour followed by a single repeat. Low stringency washes were identical except that the wash temperature was 45°C.

The sources of the human RNAs were adult skin fibroblasts (ATCC #GM08429), HeLa cells, peripheral blood reticulocytes, myeloid cell line K562, (ATCC #CCL243), T-cell lymphoma-derived cell line SupT1 (gift of J. Hoxie, University of Pennsylvania), hepatoma-derived cell lines Hep 3B and Hep G2 (20), a term placenta dissected into amnion, chorion, villi, and decidual layers, brain, primary B lymphocytes, and adult kidney (a gift of E. Neilson, University of Pennsylvania). RNA was isolated from these cell lines and tissues by one of two methods (21,22). 10 ug of total RNA were denatured in 6.6% formamide at 65°C for 5 min and electrophoresed through 1.5% agarose/6.5% formaldehyde submerged slab gels (23), transferred to GeneScreen Plus (New England Nuclear Research Products, Boston, MA), and hybridized overnight with 0.5×10^6 cpm/ml of probe at 42°C in 5× SSPE, 50% formamide, 5× Denhardt's solution, 10% dextran sulfate, 2% SDS, and 200 mg/ml boiled salmon sperm DNA. Each of two successive washes were carried out in 2× SSPE, 2% SDS at 65°C for 30 min.

Reverse Transcription/Polymerase Chain Reaction

To resolve the discrepancy at nucleotide 335 (a C in clone 5F that was absent in clone A4, see Fig. 2A), placental mRNA was reverse transcribed from antisense oligonucleotide 5FD, complementary to bases 429 through 412 (5'-CTTCTCCGCAGCATAGAC-3') of hCRP (Fig. 1 and 2A). This synthesized cDNA was amplified (24) between oligonucleotide 5FD and sense oligonucleotide 5FE, identical to bases 178 through 197 (5'-ACTGTGGCCGTGCATGGTGA-3') followed by direct sequence analysis (see below). To confirm the termination codon, placental RNA was reversed transcribed from antisense oligonucleotide 5FF, complementary to bases 685 through 665 (5'-AGCTGCTGGGAATGGAATGGC-3') and the cDNA was then amplified between this oligonucleotide and sense oligonucleotide 5FC, identical to bases 508 through 528 (5'-CTGGCAGACAAGGATGGCGAG-3') followed by direct sequence analysis (see below).

In the tissue survey for hCRP transcripts using the polymerase chain reaction (PCR) the sources of the RNA were: placental cytotrophoblast cell lines JEG-3 (ATCC # HTB 36) and BeWo (ATC # CCL 98), the B-lymphoblast-derived cell line GM1500 (gift of J. Haddad, University of Pennsylvania), the Sup T1 cell line, kidney, HeLa cells, fibroblasts, reticulocytes, and rat liver. 5 μg of each total tissue RNA were reverse transcribed with 0.15 μg of antisense oligonucleotide 5FF, or an actin antisense oligonucleotide complementary to bases 1091 to 1112 (5' CAGGTCCAGACGCAGGATGGC-3'; (25). These cDNAs were then phenol/chloroform extracted and ethanol precipitated. Next these cDNAs were amplified between an additional 0.15 μg of the antisense primer and 0.15 μg of sense oligonucleotide 5FC, or an actin sense oligonucleotide identical to bases 403 to 424

(5′ CTACAATGAGCTGCGTGTGG-3′). In each case the sense primer was [^{32}P]-labeled at its 5′-end. 25 cycles of amplification were carried out in a thermocycler (Perkin-Elmer-Cetus, Norwalk, CT) under conditions previously detailed (26) with the following specifications: initial denaturation, 3 min at 93°C; initial reannealing, 1 min at 54°C; and initial extension, 3 min at 72°C. Subsequent cycles were: 30 seconds at 95°C, 15 sec at 54°C, and 1 min at 72°C respectively. PCR was terminated with a 10 min extension cycle at 72°C. The final products were phenol/chloroform extracted and ethanol precipitated. For the tissue survey, the PCR products were analyzed on 8 M urea, 8% polyacrylamide gels which were dried and autoradiographed. Other amplified fragments were directly sequenced by the dideoxy reaction using oligonucleotides 5FD or 5FE as primers to clarify the sequence at position 335 or using oligonucleotides 5FF or 5FC to confirm the termination codon.

In vitro transcription and translation

To generate a template for synthesis of an hCRP mRNA that would encode the entire protein, the inserts of 5F cDNA, containing the nearly full-length 5′-region, and A4 cDNA, containing the complete 3′-region, were ligated at a common NsiI site contained in both clones (Fig. 3A). The A4 insert was first ligated into the EcoRI site of the pGEM3 vector in the transcriptional orientation determined by the T7 promotor (PB). A fragment of A4 was excised which extended from NsiI at base 358 of the insert to the HindIII site in the polylinker. The 5F cDNA was subcloned into the EcoRI site of the pGEM4 vector in the SP6 orientation. A 3′-fragment of the 5F insert was removed by another NsiI and HindIII digestion. The remainder of this plasmid containing the 5′-end of the cDNA along with the pGEM-4 vector on an NsiI and HindIII fragment was ligated to the full-length 3′-end of the cDNA released as an NsiI/HindIII fragment from the A4 subclone. The resultant plasmid, pGEM4-hCRP, contains a near full-length hCRP cDNA in the transcriptional orientation of the SP6 promotor (Fig. 3A). Prior to *in vitro* transcription, 500 ng of pGEM4-hCRP were linearized by digestion with AvaI 3′ of the cDNA insert in the polylinker. The linearized cDNA was then transcribed in the presence of SP6 RNA polymerase at 40°C for 1 hr. The transcription products were subsequently purified on a G-50 Sephadex spin column (BMB) which had been prewashed with sterile water. The transcription reaction was brought to a final volume of 50 ul and a 2 ul aliquot was analyzed on an 8 M urea, 6% polyacrylamide gel to confirm that the transcript was intact and of the appropriate size. 1, 2, and 5 ul of the transcription reaction were translated *in vitro* in a micrococcal nuclease-treated rabbit reticulocyte lysate (27) in the presence of [^{3}H]-leucine at 30°C for 40 min as previously described (28). Translation products were analyzed on a 15% polyacrylamide-SDS gel. The gel was treated with Resolution Enhancer (EM Corp., Chestnut Hill, MA) dried, and exposed to X-ray film (Kodak AR-5) for 96 hours at −70°C.

RESULTS

Isolation of the hCRP cDNA

The original goal of this study was to identify transcripts in a human placental cDNA library that are structurally related to hPrl mRNA. To avoid isolating authentic hPrl cDNA from the placental cDNA library (Prl is expressed in the decidual layer (29)), all plaques that hybridized after full stringency washes were excluded from consideration, and of those remaining, only plaques detected consistently after low stringency washes were isolated. Using these criteria, a total of four separate recombinant phage were chosen and plaque purified. Each of the cloned cDNA inserts was sized. The largest, A4, containing an insert of about 1.8 kb was selected for further study.

Analysis of the cDNA sequence

The A4 cDNA is 1778 base pairs (bp) in length. It's terminus is marked by the arrow in figure 2A. The 5′ to 3′ orientation of this cDNA was established by the position of the 68 base polyadenosine (poly A) tail. Its predicted amino acid sequence begins with GTT (Val) at codon 17 of hCRP (Fig. 2A). The transcription start site of the native mRNA was determined by primer extension on placental RNA using an antisense primer corresponding to bases 12−32 of the A4 cDNA (see Methods). A single extension product of 134 nucleotides (nt) was observed (data not shown) indicating that the A4 clone was missing the first 102 nt of the mRNA. To obtain the full 5′ end, the library was rescreened at high stringency with a 190 bp EcoRI-BanI restriction fragment isolated from the 5′ terminus of A4 cDNA. cDNA inserts from 12 positive isolates were mapped and the cDNA fragment with the furthest 5′ extent, 5F, was fully sequenced and aligned to the A4 sequence (Figs. 1 and 2A).

The 5F cDNA contains 84 5′-bases not present in the A4 cDNA, therefore being 19 nt shorter than the native mRNA. The

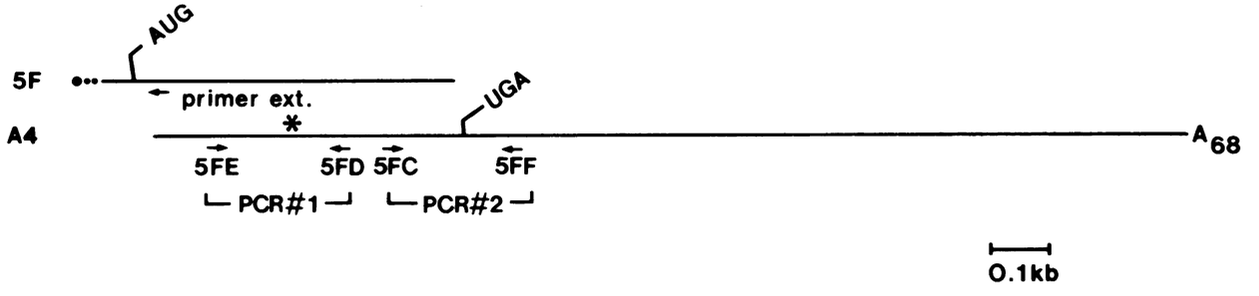

Figure 1. Structure and analysis of the hCRP cDNAs. The structures of the two characterized hCRP cDNA clones, 5F and A4, are shown. The full 5′ extent of hCRP mRNA as determined by primer extension is shown by the dotted extension of the 5F cDNA. The location of the primer used in this analysis is indicated by the arrow beneath 5F. The two regions of hCRP that were studied by reverse transcriptase/PCR analysis are bracketed by the relevant primers shown as arrows and labeled PCR #1 and PCR #2. The sequence of each of the labeled primers and the nature of the discrepant base (*) are described in the text. The positions of the initiation codon (AUG) termination codon (UGA) and poly A tail (A_{68}) are labeled.

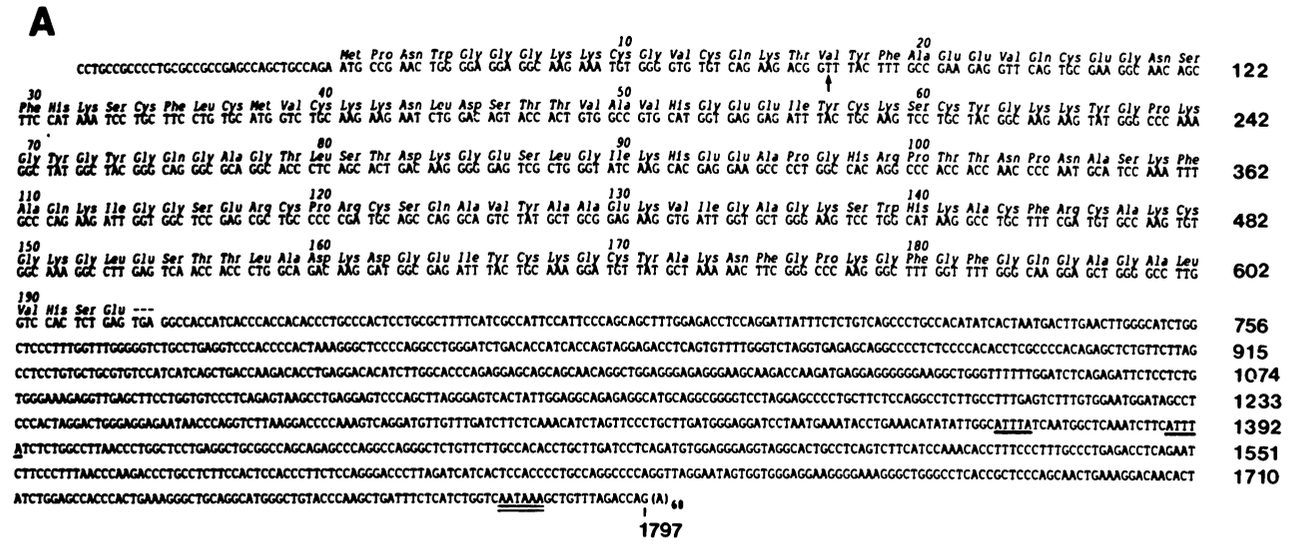

Figure 2. Structural analysis of the hCRP mRNA and its encoded protein. **A.** Primary sequence of hCRP cDNA and its predicted protein product. The sequence shown is a combination of that derived from sequencing both A4 and 5F cDNA clones. The numbers above the sequence refer to amino acid positions and the numbers to the right of each line refer to the nucleotide position. The upward arrow at base 85 marks the 5′ extent of the A4 cDNA with all bases to the left of the arrow determined from the 5F clone. The sequenced 5′-nontranslated region begins 19 bases 3′ of the mRNA cap as determined by primer-extension mapping. The position of the two AUUUA motifs (32) in the 3′-nontranslated region are underlined and the position of the polyadenylation signal, AAUAAA (31) is underlined twice. **B.** Alignment and domain organization of hCRP and r/mCRIP. The sequence of hCRP is shown in the single letter amino acid code with the residue positions numbered. The sequence of r/mCRIP (35) is aligned with hCRP. Gaps (−) have been inserted to maximize alignment. Identical residues are boxed. A consensus sequence between the two sets of hCRP putative finger domains and the single set of r/mCRIP putative finger domains along with the associated glycine-rich region is shown. The class of each individual putative finger domain is indicated by C_2HC or C_4. The symbol Ø indicates an aromatic amino acid residue. **C.** Schematic representation of hCRP demonstrating the duplicated domains each containing two putative zinc fingers and a glycine-rich repeat. Amino acid residue positions are noted below the diagram. The highly conserved cysteines (C) and histidines (H) at the base of each of the putative zinc fingers, and the sequence in the glycine-rich domains following each double finger structure are shown. The presence of the indicated secondary structure and Zn coordination are strictly hypothetical (see text). A summary of the class, size, and spacing of the putative fingers is noted at the bottom of the diagram.

5F contains an AUG at nt 36 followed by an extended open reading frame. The A4 and 5F cDNA clones overlap by 501 nt (Figs. 1 and 2A). This region of overlap contains a single discrepant base; three C's at positions 333–335 of 5F were represented by only two C's at the corresponding position in A4. The A4 sequence did not appear to be a compression as it was clearly read on both strands by both the chemical degradation and dideoxy sequencing techniques. To clarify this discrepancy, this region of the mRNA was reverse transcribed and then subjected to amplification using the PCR technique (PCR #1 in Fig. 1, see Methods). Direct sequencing of the amplified fragment demonstrated the presence of three C's at the discrepant position in native mRNA (not shown) suggesting that the A4 clone contained a one base deletion.

The consensus sequence of the two cDNA clones along with the primer extension data define an mRNA containing 1816 nt exclusive of its poly A tail. We have named this mRNA human cysteine-rich protein (hCRP) as explained below. The hCRP mRNA contains a 54 nucleotide 5′ nontranslated region; 35 of these bases were assigned by DNA sequence analysis of 5F cDNA and the existance of the remaining 19 bases was inferred from primer extension analysis. The reading frame begins with an AUG surrounded by a fair consensus (30) which includes an A in the −3 position and a C in the +4 position. The open reading frame contains 193 codons. The coding region is followed by a very long (1180 nt) 3′-nontranslated region. The position of the termination codon was confirmed by reverse transcription of placental mRNA, followed by PCR (see PCR #2 in Fig. 1 and Methods). An AAUAAA polyadenlyation signal (31) precedes the poly A tail by 14 bases. There are also two adjacent AUUUA motifs at nucleotides 1365–1369 and 1388–1392 (underlined in Fig. 2A). Similar motifs have been reported to regulate mRNA stability (32).

Analysis of the protein structure

The predicted amino acid sequence of hCRP is shown above the mRNA sequence in figure 2A. This 193 amino acid protein has a calculated molecular weight of 20,547 daltons. At residues 105–107 there is the sequence Asn-Ala-Ser, a predictive signal for N-linked glycosylation. Hydropathy plotting failed to demonstrate regions consistent with a signal peptide or a membrane-spanning domain.

The most striking features of hCRP are its highly basic composition and the number and distribution of cysteine residues. The estimated pI is 10.38 reflecting the presence of 33 basic residues, including 23 lysines, compared to only 17 acidic residues. There are 15 cysteines in the protein. The positioning of several of these cysteines along with several histidines occurs in a repeated pattern. This pattern, repeated four times, describes a 25 amino acid domain with the overall consensus structure of Cys-$(X)_2$-Cys-$(X)_{17}$-His/Cys-$(X)_2$-Cys. These domains are best aligned pairwise since the first and third domains end with His-$(X)_2$-Cys while the second and fourth domains end with Cys-$(X)_2$-Cys (Fig. 2B). The amino acid identity between the first and third domains (12/25) and the second and fourth (14/25) domains are significantly higher than that between the adjacent first and second (7/25) or third and fourth domains (5/25). Each repeated unit conforms quite well to a zinc-finger motif (5). A diagram of this protein containing four putative zinc fingers is shown in figure 2C. If the amino acid comparison among the fingers is limited to the 17 residue loops, the pattern of finger similarity is even more striking. Only a single amino acid is conserved between the adjacent fingers while there is a 7 and 8 out of 17 match between the first and third and between the second and fourth fingers, respectively. The first and second fingers, as well as the third and fourth fingers are separated by an unusually short linker region of two amino acids. The pattern of internal homology suggests that the four finger domains arose by duplication of a preexisting two domain unit.

Each of the finger doublets are immediately followed by a glycine-rich domain containing a high proportion of aromatic and basic residues (Fig. 2C). The two glycine-rich regions are highly similar with a common sequence of ØGPKGØGØGQGAG where Ø is an amino acid residue with an aromatic R-group. The amino terminus of the first domain is overlapped by a sequence of six basic amino acids containing an internal proline, KKYGPK. This sequence resembles the nuclear localization signal the *S. cerevisiae* MATα2 protein but is somewhat less basic than the signal of the SV40 T-antigen (33,34). There may however be a significant difference in these sequences in that two uncharged polar residues in the hCRP sequence occupy positions corresponding to hydrophobic residues in the two examples cited.

A computer-based search of the National Biomedical Research Foundation and Swiss Protein databases identified a single highly significant alignment between hCRP and mouse and rat cysteine-rich intestinal peptides (r/mCRIP) (35). These two rodent proteins are identical to each other in structure despite divergence in the respective mRNA sequences. Further searching of Genbank and EMBL databases at the nucleotide sequence level failed to detect other genes with significant levels of structural similarity to hCRP. The primary sequence of the r/mCRIP is aligned with the two repeated domains of the hCRP protein in figure 2B. The finger doublet and glycine-rich region are present in a single copy in r/mCRIP, consistent with the hypothesis that a pre-existing single two finger domain unit was duplicated to form hCRP. Over the 68 amino acid residues encompassing the two zinc-finger and glycine-rich region in m/rCRIP there is a 28 residue match to hCRP which allowed us to derive a consensus sequence (Fig. 2B). Based on the sequence similarity to r/mCRIP, we have named the 5F/A4 protein human cysteine-rich peptide (hCRP).

Relationship of the hCRP cDNA to hPrl

The hCRP cDNA was identified and isolated on the basis of reproducible hybridization at low stringency to the hPrl cDNA probe. The hCRP cDNA and encoded protein were therefore analyzed for similarity to hPrl. Direct analysis of hCRP cDNA by hybridization to hPrl cDNA and by computer analysis for matching regions suggests that clone identification resulted from hybridization over short regions of similarity; 70% nucleotide identity (43/67 with a 4 base gap) was found between exon 3 of hPrl (codons 23–44) and the region bridging the first and second fingers of hCRP (codons 16–37). There is no significant amino acid homology in this region. The hybridization studies also suggest that the poly A tails of the cDNA and the hPrl cDNA probe may have contributed to the detection despite the presence of oligo(dT) in the hybridization mix. In further distinction from hPrl, hCRP lacks a signal peptide and has none of the conserved residues characteristic of the GH-Prl family of genes (36). These data suggest that hCRP and hPrl contain no regions of significant structural similarity.

In vitro synthesis of the hCRP

To confirm the predicted translational start site and reading frame of the hCRP mRNA, the 5F and A4 cDNAs were ligated to form

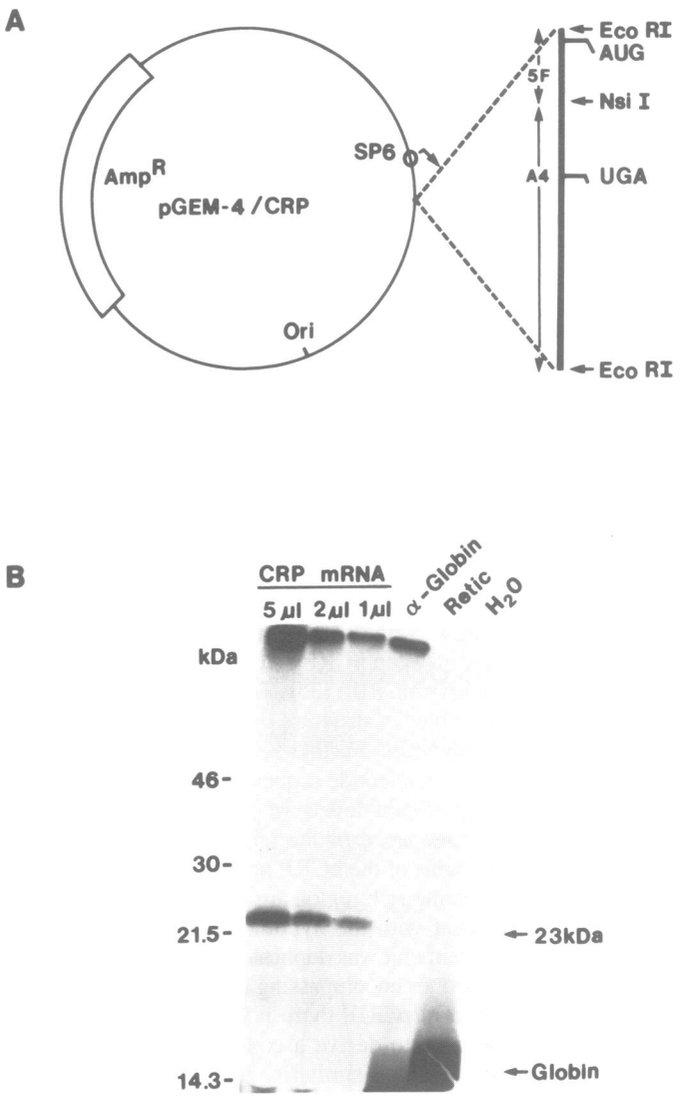

Figure 3. Synthesis and *in vitro* translation of hCRP mRNA. **A.** Insertion of a full-length hCRP cDNA into transcription vector pGEM4. A full coding length hCRP cDNA was constructed and inserted into the EcoRI site of the polylinker region of the pGEM4 vector in the indicated orientation (AUG initiator codon to UGA terminator codon). The landmarks of the β-lactamase gene (Amp^R), the plasmid origin of replication (Ori) and the SP6 polymerase promotor (SP6) are all noted. The position of the Nsi I site used to generate a full-length template, and the regions contributed by the A4 and 5F clones are indicated. **B.** Translation of hCRP mRNA SP6 polymerase catalyzed capped mRNA transcripts of the hCRP cDNA were translated in a rabbit reticulocyte lysate at three different concentrations (5 ul, 2 ul, 1 ul) in parallel with three controls: synthetic human α-globin mRNA (α-globin), human reticulocyte mRNA (Retic), and water (H_2O). Translations were carried out in the presence of [^{3}H]-leucine, then directly analyzed on a 15% SDS-polyacrylamide gel. The positions of molecular weight markers (kDa) are noted on the left of the resultant autoradiograph, and the positions of the hCRP and globin translation products are noted to the right of the gel.

a near full-length hCRP cDNA (see Methods for details). This plasmid was used as template to transcribe an hCRP mRNA (Fig. 3A). *In vitro* translation of this mRNA resulted in a single protein band with an estimated molecular weight of 23.4 kDa (Fig. 3B). The 14% difference between the calculated (20.5 kDa) and measured molecular weights is ascribed to inaccuracies in molecular weight determinations by SDS-polyacrylamide gel electrophoresis (37).

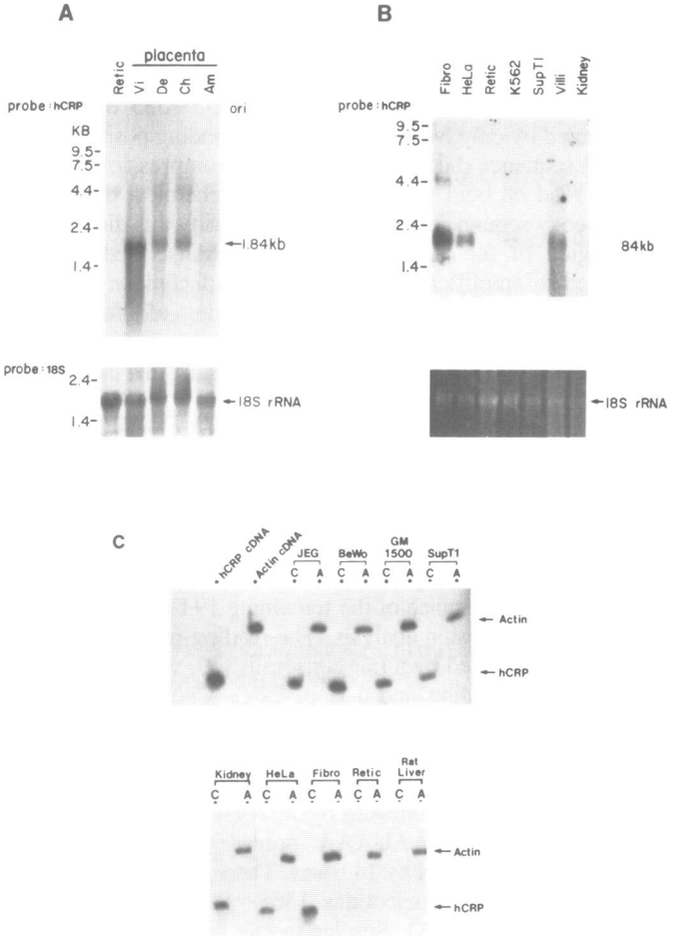

Figure 4. The tissue distribution of hCRP mRNA. **A.** Distribution of hCRP mRNA in term placental tissues. Top panel. Equal quantities of total cellular RNA isolated from the four dissected layers of a term placenta, amnion (Am), chorion (Ch), decidua (De), and villi (Vi), as well as from normal human reticulocytes (Retic) were analyzed by Northern blotting using a [^{32}P]-labeled hCRP cDNA probe (5F). The position of RNA size markers are noted on the left of the autoradiograph and the size and position of the hCRP mRNA band is noted on the right. Bottom panel. Hybridization of a companion Northern blot with a [^{32}P]-labeled 18S ribosomal genomic DNA clone. The position of the 18S rRNA signal is labeled. **B.** Detection of hCRP mRNA in a variety of human tissues and cell lines. Top panel. Equal quantities of total cellular RNA from a human skin fibroblasts (Fibro), HeLa cells (HeLa), human reticulocytes (Retic), the human K562 cell line (K562), the human T-cell derived line SupT1 (SupT1), the villous layer of a human term placenta (Villi), and a human kidney (Kidney) were analyzed as in A. The 1.84 kb hCRP mRNA is indicated by an arrow. Bottom panel. Ethidium bromide strain of the agarose gel prior to Northern transfer demonstrating the 18S rRNA band. **C.** Reverse transcriptase/PCR analysis of RNAs from a variety of sources. A cDNA copy of each total RNA sample was generated using either a specific 3′ primer for actin or hCRP and reverse transcriptase. Each cDNA was amplified using the set of primers specific for either actin or hCRP (see Methods). In each case the sense primer was ^{32}P-end labeled prior to the reaction. Reaction products were resolved on a denaturing acrylamide gel and directly visualized by autoradiography. hCRP and actin cDNAs were amplified to demonstrate the expected position of the specific amplification products (noted by arrows to the right of each panel). The lanes containing hCRP reverse transcriptase/PCR are labeled 'C' and those containing actin amplification products are labeled 'A'.

Distribution of hCRP mRNA

The tissue distribution of hCRP was determined by Northern analysis and by the more sensitive reverse transcription/PCR assay. Northern analysis demonstrated the expected 1.8 kb hCRP

mRNA band in all four tissue layers of the placenta: amnion, chorion, villi, and decidua. This result suggests a lack of strict tissue specificity for hCRP expression. An additional minor band of 4.4 kb was seen in some lanes. The identity of this band is not clear but its presence in the reticulocyte sample which is devoid of the 1.8 kb hCRP band suggests it represents minor crosshybridization to 28S rRNA. The concentrations of the 1.8 kb hCRP mRNA differed among the four placental tissues with the highest concentration in the villi (Fig. 4A). The lack of apparent tissue specificity was strengthened by a survey of nonplacental tissues. hCRP mRNA could be detected in a variety of nucleated cells by Northern analysis but as in the placental tissue survey the concentrations differed markedly. The highest levels were observed in skin fibroblasts, with intermediate levels in HeLa cells and placental villi, and very low levels in myeloid, lymphoid and kidney cells (clearly seen on the original autoradiograph). No signal was detected in reticulocytes. hCRP mRNA was also detected in brain, liver, primary B-lymphocytes, and two human hepatoma cell lines, Hep 3B and Hep G2 (data not shown).

To confirm the presence of hCRP mRNA in several of the samples with low levels by Northern analysis, the more sensitive reverse transcription/PCR assay was used (Fig. 4C). hCRP amplified fragments were detected in all samples with the specific exception of the reticulocytes. Rat liver mRNA, used as a negative control in this assay, demonstrates the species specificity of the hCRP amplification primers. In contrast, an actin cDNA band was detected in the amplification of all samples including reticulocyte and rat liver (Fig. 4C). The actin primers are located in regions of actin mRNA homologous between human and rat.

Evolutionary conservation of gene sequence

To determine the degree to which the structure of the CRP gene has been conserved during evolution we analyzed DNA from a spectrum of eukaryotic species selected from disperate segments of the phylogenetic tree. DNA was isolated from fungi (*Saccharomyces cerevisiae*), flatworm (*Schistosoma mansoni*), arthropod (*Drosophila melanogaster*), and chordates including bird (chicken) and mammals (mouse and human). Under high stringency hybridization conditions, EcoRI digestion of each of the DNAs gave a simple pattern of bands consistent with the presence of a single copy gene (Fig. 5). In addition, in a survey of primates including rhesus, orangutan, and gorilla the banding patterns were identical to that seen in humans (data not shown). These data are consistent with the presence of a highly conserved single gene locus.

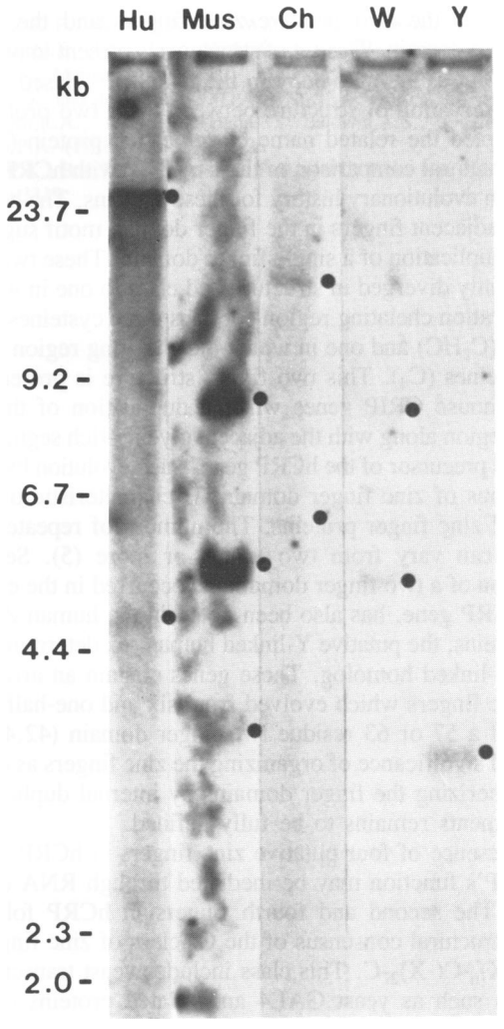

Figure 5. Detection of genomic sequences in distantly related species which cross-hybridize with the hCRP cDNA. Equal quantities of high molecular weight total cellular DNA from man (Hu), mouse (Mus), chicken (Ch), *Schistosoma mansoni* (W) and *Saccharomyces cerevesiae* (Y) were digested with EcoRI then resolved on an 0.8% agarose gel and analyzed by Southern blotting and hybridization at high stringency (see Methods) using a [^{32}P]-labeled hCRP (5F) probe. The origin of each of the samples is noted at the top of each lane, a dot is positioned to the right of each hybridizing band and the position of DNA molecular weight markers in kilobases (kb) is noted to the left of the autoradiograph.

DISCUSSION

The mRNA transcript described in this report was identified by screening a placental cDNA library at low stringency with an hPrl cDNA probe. The goal was to identify transcripts structurally similar to hPrl mRNA. An hPrl mRNA has been detected in the decidual layer of placenta (29) and a cDNA clone nearly identical to pituitary hPrl has been characterized (38). The assumption that other hPrl-related mRNAs might exist was based upon the detection of Prl-related mRNAs in bovine and rodent species (39,40,41). In addition, we had detected hPrl-related mRNAs in human non-decidual placental RNA samples by low stringency Northern analysis (unpublished data). Despite the fact that the cDNA described in this report hybridized reproducibly to the hPrl cDNA probe at low stringency, its structure and that of its predicted protein product lack any significant evolutionary or functional relationship to hPrl. We conclude that the isolated cDNA is not related to hPrl and was probably a fortuitous cross-hybridization based upon regions of limited sequence homology.

We have searched the protein databases with the predicted primary structure of hCRP for sequence similarity to previously reported proteins. We identified a single highly significant match with CRIP characterized in both rat and mouse (35). The r/mCRIP gene encodes an 8.55 kDa protein. rCRIP mRNA is present in a wide variety of tissues including lung, spleen, adrenal, and testis, but is absent from liver, kidney, and brain as assessed by Northern analysis. It is most actively expressed in the small intestine and colon in the time interval between suckling and weaning. Alignment of hCRP with r/mCRIP demonstrates a high degree of structural similarity (Fig. 2B). This match is most striking in the 68 amino acid region which

encompasses the two putative zinc fingers and the adjacent glycine-rich domain. This set of structures is present in two copies in hCRP but as a single copy in the r/mCRIP. Based upon the clear conservation of structure between these two proteins, we have adopted the related name cysteine-rich protein (hCRP).

The structural comparison of the r/mCRIP with hCRP support a common evolutionary history for these proteins. The homology between adjacent fingers in the finger doublet motif suggests an original duplication of a single finger domain. These two fingers subsequently diverged in structure and class to one in which the putative cation chelating region contains three cysteines and one histidine (C_2HC) and one in which the chelating region contains four cysteines (C_4). This two finger structure is present in the rat and mouse CRIP genes while a duplication of the finger doublet region along with the adjacent glycine-rich segment gave rise to the precursor of the hCRP gene. Such evolution by internal duplications of zinc finger domains is characteristic of certain classes of zinc finger proteins. The number of repeated finger domains can vary from two to ten or more (5). Segmental duplication of a two finger domain, as occurred in the evolution of the hCRP gene, has also been noted in the human ZFY and ZFX proteins, the putative Y-linked human sex determinant gene and its X-linked homolog. These genes contain an array of 13 C_2H_2 zinc fingers which evolved from six and one-half tandem repeats of a 57 or 63 residue two-finger domain (42,43). The functional significance of organizing the zinc fingers as doublets or multimerizing the finger domains by internal duplication of gene segments remains to be fully defined.

The presence of four putative zinc fingers in hCRP suggests that hCRP's function may be mediated through RNA or DNA binding. The second and fourth fingers in hCRP follow the general structural consensus of the C_4 class of zinc fingers, C-$(X)_2$-C-$(X)_n$-C$(-X)_2$-C. This class includes yeast transcriptional activators such as yeast GAL4 and related proteins (44), the mammalian steroid receptors (45,5,6), and the mouse and chicken erythroid specific transcription factor homologs, GF-1 and Eryf1 (46,47). The specific consensus sequence and spacing of cysteine residues in each of these families of proteins are distinct. The Cys-$(X)_2$-Cys-$(X)_{17}$-Cys-$(X)_2$-Cys configuration which appears in GF-1 and Eryf1 is found in exact copy in the second and fourth putative hCRP fingers. The actual involvement of Zn coordination in this group of proteins has been difficult to document (47). On the basis of structrual conservation of the C_4 motif finger we propose that the second and fourth fingers of hCRP may function by sequence specific interaction with DNA or RNA.

The first and third putative finger domains in hCRP are of the C_2HC class. These fingers are of particular interest as they have specific structural similarities to the finger consensus Cys-$(X)_2$-Cys-$(X)_4$-His-$(X)_4$-Cys found in a number of nucleic acid binding proteins (NBPs) in mammalian, avian, drosophila, and plant lineages (48). Specific examples of these NBPs include the retroviral GAG genes that encode RNA binding proteins necessary for packaging the retroviral genome, the cellular nuclear binding protein that encodes the mammalian sterol regulatory element (CNBP) (49), and the bacteriophage T4 gene 32 protein (50). The first two have C_2HC fingers while the last has a CHC_2 finger. In all three cases the proteins appear to function as single-stranded nucleic acid binding proteins being associated with either DNA or RNA. Both GAG and CNBP share two additional structural characteristics with hCRP and r/mCRIP. First, they all share a conserved glycine in their C_2HC finger loops located four residues before the zinc-coordinated histidine. Second, they all contain a glycine-rich region containing a high concentration of basic and aromatic residues immediately after the finger domain. It should be noted however that the sizes of the finger loops in the NBP's four residues are significantly smaller than the 17 residue finger loops found in the r/mCRIP and hCRP. In further similarity with hCRP, all of these proteins appear to be highly conserved in evolution, have a wide tissue distribution, and contain long 3'-nontranslated regions. Recently a chromatin-associated enzyme, poly (ADP-ribose) polymerase, has been cloned and found to contain two C_2HC fingers with loops of 28 and 30 residues. These fingers are associated with zinc ions which have been shown to be essential for the protein's ability to bind to DNA (51). These structural comparisons suggest that the first and third C_2HC finger domains of the hCRP may function by interaction with nucleic acids.

The information presently available on the hCRP gene suggests that it may encode a protein with a fundamental function(s). At present we have no information on what that function might be but the data suggest certain possibilities. The specific absence of hCRP mRNA in reticulocytes (post-mitotic cells which have extruded their nuclei) and the presence of a nuclear targeting consensus signal in the hCRP protein suggests that its function may relate to nuclear activity or cellular replication. The lack of detectable hCRP mRNA in the reticulocytes may also reflect the presence of the AUUUA motifs in its 3' translated region resulting in a shortened mRNA half- life (32). The possibility that hCRP is a nuclear protein is supported by the presence of four putative zinc-finger motifs of two distinct classes, both of which suggest function by nucleic acid binding. The ubiqitous tissue distribution of hCRP along with the remarkable conservation of CRP-related sequence during evolution further suggest that its evolution has been constrained by an activity that may be fundamental to cell function. In fact, a number of proteins important to the regulation of transcription have been found to be structurally conserved from higher eucaryotics through yeast (52,53). These function(s) may in fact be served by a number of structurally related proteins, since comparison of Southern blot analyses of total genomic DNA probed with hCRP and m/rCRIP cDNAs suggests that these two genes are members of a dispersed multigene family. Structural analysis of these closely related genes in a number of disparate species may indicate conserved regions of these proteins of particular interest for further study.

ACKNOWLEDGEMENTS

This investigation was partially supported by National Institutes of Health grant RO1-HD25147 (NEC and SAL) and National Foundation March of Dimes Basic Research Grant 1-1015 (NEC). The BIONET national computer resource was supported by National Institutes of Health Grant RR01865-05. The authors are grateful to M. Lazar and W. M-F. Lee for critical comments and Susan Kelchner for expert secretarial assistance. S.A. Liebhaber is an Associate Investigator in the Howard Hughes Medical Institute.

REFERENCES

1. Struhl,K. (1989) TIBS **14**, 137–141.
2. Klug,A. and Rhodes,D. (1987) TIBS **12**, 464–469.
3. Miller,J., McLachlan,A.D. and Klug,A. (1985) EMBO J. **4**, 1609–1614.
4. Brown,R.S., Sander,C. and Argos,P. (1985) FEBS Lett. **186**, 271–274.

5. Evans,R.M. and Hollenberg,S.M. (1988) Cell **52**, 1–3.
6. Berg,J.M. (1989) Cell **57**, 1065–1068.
7. Frankel,A.D. and Pabo,C.O. (1988) Cell **53**, 675.
8. Hanas,J.S., Hazuda,D.J., Bogenhagen,D.F., Wu,F.Y. and Wu,C.W. (1983) J. Biol. Chem. **258**, 14120–14125.
9. Cooke,N.E., Coit,D., Shine,J., Baxter,J.C., and Martial,J.A. (1981) J. Biol. Chem. **256**, 4007–4016.
10. Wilson,G.N., Hollar,B.A., Watterson,J.R., and Schmickel,R.D. (1978) Proc. Natl. Acad. Sci. USA **75**, 5367–5371.
11. Maxam,A.M. and Gilbert,W. (1977) Proc. Natl. Acad. Sci. USA **74**, 560–564.
12. Messing,J. (1983) in Methods in Enzymology 101 (part C); Recombinant DNA (Wu,R., Grossman,L., Moldave,K, eds) Vol. 101, pp. 20–78, Academic Press, New York.
13. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
14. Henthorn,P.S., Knoll,B.J., Raducha,M., Rothblum,K.N., Slaughter,C., Weiss,M., Lafferty,M.A., Fisher,T. and Harris,H. (1986) Proc. Natl. Acad. Sci. USA **83**, 5597–5601.
15. Goossens,M. and Kan,Y.W. (1981) Methods Enzymol. **76**, 805–817.
16. Benton,W.D. and Davis,R.W. (1977) Science **196**, 180–182.
17. Maniatis,T.E., Fritsch,E.F. and Sambrook,J. (1982) Cold Sring Harbor Laboratory, Cold Sring Harbor, New York.
18. Aviv,H. and Leder,P. (1972) Proc. Natl. Acad. Sci. USA **69**, 1408–1412.
19. Liebhaber,S.A. and Kan,Y.W. (1981) J. Clin. Invest. **68**, 439–446.
20. Knowles,B.B., Howe,C. and Aden,D.P. (1980) Science **209**, 497–499.
21. Chirgwin,J.M., Przbyla,A.E., MacDonald,R.J. and Rutter,W.F. (1979) Biochem. **18**, 5294–5299.
22. Strohman,R.C., Moss,P.S., Micon-Eastwood,H., Spector,D., Przybyla,A. and Patterson,B. (1977) Cell **10**, 265–273.
23. Lehrach,H., Diamond,D., Wozney,J.M. and Boedtder,H. (1977) Biochem. **16**, 4743–4751.
24. Saiki,R.K., Bugawan,T.L., Horn,G.T., Mullis,D.B., and Erlich,H.A. (1986) Nature **324**, 163.
25. Ng,S.Y., Gunning,P., Eddy,R., Ponk,P., Leavitt,J., Shows,T., and Kedes,L. (1985) Mol. Cell Biol. **5**, 2720–2732.
26. Liebhaber,S.A., Urbanek,M., Ray,J., Tuan,R.S. and Cooke,N.E. (1989) J. Clin. Invest. **83**, 1985–1991.
27. Pelham,H.R.B. and Jackson,R.J., Eur. J. Biochem. (1976) **67**, 247–256.
28. Liebhaber,S.A., Cash,F.E. and Shakin,S.H. (1984) J. Biol. Chem. **259**, 15597–16502.
29. Clements,J., Whitfield,P., Cooke,N., Healy,B., Matheson,B., Shine,J. and Funder,J., Endocrinol. (1983) **112**, 1133–1134.
30. Kozak,M. (1986) Cell **44**, 483–492.
31. Proudfoot,N.J. and Brownlee,G.G. (1976) Nature **263**, 211–214.
32. Shaw,G. and Kamen,R. (1986) Cell **46**, 659–667.
33. Kalderon,D., Roberts,B.L., Richardson,W.D. and Smith,A.E. (1984) Cell **39**, 499–509.
34. Hall,M.N., Hereford,L. and Hershowitz,I. (1984) Cell **36**, 1057–1065.
35. Birkenmeier,E.H. and Gordon,J.I. (1986) Proc. Natl. Acad. Sci. USA **83**, 2516–2520.
36. Nicoll,C.S., Mayer,G.L. and Russell,S.M. (1986) Endocrine Reviews **7**, 169–203.
37. Weber,K., and Osborn,M. (1969) J. Biol. Chem. **244**, 4406–4412.
38. Takahashi,H., Nabeshima,Y., Nabeshima,Y., Ogata,K., and Takeuchi,S., J. Biochem. (1984) **95**, 1491–1499.
39. Robertson,M., Gillespie,B., and Friesen,H.G. (1982) Endocrinol. **111**, 1862–1866.
40. Linzer,D.I.H., Lee,S-J., Ogren,L., Talamantes,F., and Nathans,D. (1985) Proc. Natl. Acad. Sci. USA **82**, 4356–4359.
41. Schuler,L.A., and Hurley,W.L. (1987) Proc. Natl. Acad. Sci. USA **84**, 5650–5654.
42. Page,D.C., Mosher,R., Simpson,E.M., Fisher,E.M.C., Mardon,G., Pollack,J., McGillivray,B., de la Chapelle,A. and Brown,L.G. (1987) Cell **51**, 1091–1104.
43. Schneider-Gadicke,A., Beer-Romero,P., Brown,L.G., Nussbaum,R. and Page,D.C. (1989) Cell **57**, 1247–1258.
44. Messenguy,F., Dubois,E. and Deschamps,F. (1986) Eur. J. Biochem. **157**, 77–81.
45. Evans,R.M. (1988) Science **240**, 889–895.
46. Tsai,S-F., Martin,D.I.K., Zon,L.I., D'Andrea,A.D., Wong,G.G. and Orkin,S.H. (1989) Nature **339**, 446–451.
47. Evans,T. and Felsenfeld,G. (1989) Cell **58**, 877–885.
48. Covey,S.N. (1986) Nucleic Acids Res. **2**, 623–633.
49. Rajavashisth,T.B., Taylor,A.K. Andalibi,A., Svenson,K.L., and Lusis,J. (1989) Science **245**, 640–643.
50. Giedroc,D.P., Keating,K.M., Williams,K.R., Konigsberg,W.H. and Coleman,J.E. (1986) Proc. Natl. Acad. Sci. **83**, 8452–8456.
51. Mazen,A., Menissier-de Murcia,J., Molinete,M., Simonin,F., Gradwhol,G., Poirer,G., and de Murcia,G. (1989) Nucl. Acids Res. **17**, 4689–4698.
52. Chodosh,L.A., Olesen,J., Hahn,S., Baldwin,A.S., Guarente,L. and Sharp,P.A. (1988) Cell **53**, 25–35.
53. Vogt,P.K., Bos,T.J. and Doolittle,R.F. (1987) Proc. Natl. Acad. Sci. **84**, 3316–3319.