Volume 11 Number 12 1983 Nucleic Acids Research

The human growth hormone gene family: structure and evolution of the chromosomal locus

Greg S.Barsh, Peter H.Seeburg[1] and Richard E.Gelinas*

Program in Tumor Biology, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104, and [1]Department of Molecular Biology, Genentech, Inc., South San Francisco, CA 94080, USA

Received 8 April 1983; Revised and Accepted 27 May 1983

ABSTRACT

The structure of the human growth hormone gene cluster has been determined over a 78 kilobase region of DNA by the study of two overlapping cosmids. There are two growth hormone genes interspersed with three chorionic somatomammotropin genes, all in the same transcriptional orientation. One of the growth hormone genes lies in an active chromatin conformation in the pituitary and at least one of the chorionic somatomammotropin genes lies in an active chromatin conformation in the placenta. The two groups of genes are highly homologous throughout their 5' flanking and coding sequences, but diverge in their 3' flanking regions which raises the paradox of how genes so similar in structural and flanking sequences can be so differentially regulated. Analysis of the sequences of the genes and identification of at least three different classes of duplication units interspersed throughout the five gene cluster suggests that the cluster evolved quite recently and that the mechanism of gene duplication involved homologous but unequal exchange between middle repetitive elements of the Alu family.

INTRODUCTION

Human growth hormone (HGH, also known as somatotropin) is part of a protein family which also includes the human chorionic somatomammotropin (HCS, also known as placental lactogen) and prolactin (1). HGH and HCS are more closely related to each other than to prolactin; amino acid sequence comparisons manifest 85% homology for HGH/HCS but only 35% for HGH/prolactin (1-3). The genes for HGH and HCS are clustered together at band q22-q24 on chromosome 17 (4-6), but the prolactin gene is located on chromosome 6 (7). In spite of such close proximity, expression of HGH and HCS genes seems strictly tissue-specific; HGH production is restricted to the pituitary and HCS is made only in the placenta.

HGH and HCS genes were among the first human genes to be isolated and a great deal is known about their nucleotide sequences and transcriptional processing (8-15). Comparison of gene sequences with cloned cDNAs revealed that two non-allelic genes are responsible for HCS production (15), but only one gene expresses HGH. Very little is known about the mechanisms

responsible for developmental regulation of these genes, but by analogy with other multigene families (16-17,48), these mechanisms may involve heritable changes in chromatin structure and thus depend on the linkage arrangement and orientation of HGH and HCS genes along the chromosome.

Two approaches have been taken to establish how the HGH and HCS genes might be arranged: analysis of recombinant bateriophage (9,11,13,18-20) and hybridization of HGH or HCS cDNA probes to digests of genomic DNA (21-23). Although the sizes of certain genomic restriction fragments which contain HGH and HCS coding sequences are well established, there has been no consensus as to either the true number or the linkage arrangement of these genes (hereafter referred to as the HGH gene family).

In order to understand the molecular anatomy of the HGH gene family and the differential and tissue-specific expression of its members, we have isolated HGH and HCS genes from a cosmid library of human DNA. We now report the structure of two overlapping cosmid clones which, for the first time, demonstrates the exact number of genes and the physical linkage of the entire gene family. Results from preliminary experiments on HGH chromatin structure and on the location of internal duplication units suggest models for the expression and molecular evolution of the HGH locus.

MATERIALS AND METHODS

Preparation of the cosmid library

A cosmid library was prepared in *E. coli* strain HB101 from high molecular weight placental DNA (26) partially digested with *Mbo*I to an average size of 30-50 kb. The 35-45 kb fraction was isolated from a 10%-40% sucrose gradient and treated with calf intestinal alkaline phosphatase to prevent self ligation of target fragments. Cosmid vector arms were prepared from the plasmid pHC79 (27) by adapting a strategy designed to prevent the formation of vector concatemers (28). Briefly, pHC79 was digested with *Eco*RI or *Sal*I, treated with calf intestional alkaline phosphatase, then digested with *Bam*HI. *Eco*RI - *Bam*HI arms or *Bam*HI - *Sal*I arms were then purified from a 10%-40% sucrose gradient and ligated to target DNA at a 2 to 1 molar ratio. The conditions for ligation, packaging, transduction, plating, hybridization, secondary screening, and preparation of cosmid DNA have all been described (29). Typical packaging efficiencies were $2\text{-}5\times10^5$ colonies/ug of target DNA and 1×10^3 colonies/ug of vector arms alone. Approximately 500,000 colonies were screened with two probes labelled to a specific activity of 2×10^8 cpm/ug.

Physical Mapping of the Cosmids

Digestion of each cosmid with all possible pairwise combinations of ClaI, HindIII, SalI, and XhoI led to a unique solution for the location of most of these relatively infrequent sites. We then used three different strategies to deduce the more complex BamHI and EcoRI restriction maps: 1) double and triple limit digests were performed on each cosmid with various combinations of BamHI, EcoRI, ClaI, HindIII, SalI, and XhoI; 2) every EcoRI and HindIII restriction fragment was isolated and redigested with BamHI and EcoRI, respectively; and 3) selected ClaI and HindIII restriction fragments were isolated, asymmetrically end-labeled, and then partially digested with BamHI and EcoRI. Every restriction fragment linkage was confirmed with at least two independent strategies.

Blot Hybridizations

DNA fragments to be labelled as probes were purified from vector sequences and were nick-translated (32) with ^{32}p-dCTP and dTTP to a specific activity of at least $2x10^8$ cpm/ug. Nitrocellulose filters were prehybridized at 42° for 4 hr in 50% formamide, 5xSSC, 50 mM sodium phosphate pH 6.5, 5xDenhardt's, 200 ug/ml sonicated denatured herring sperm DNA, and 100 ug/ml polyadenlyic acid. Hybridizations were performed at 42° for 12-16 hr in fresh prehybridization solution supplemented with 10% dextran sulfate and 10^6 cpm/ml of denatured hybridization probe. Filters were washed at 67° with at least four changes of 2xSSC, 0.1% SDS for 24 hr, then 0.5xSSC, 0.1% SDS for 30 min.

DNAaseI Digestion of Nuclei

Single cell suspensions from human placenta or pituitary were dissected away from vascular and connective tissue and then homogenized with a B-Dounce in 1xSSC, 10mM TrisHCl, pH 7.5. Cells were lysed with RSB (0.01 M TrisHCl (pH7.4), 0.01 M NaCl, 3 mM $MgCl_2$)+0.5% NP-40 and the nuclear pellet was washed in RSB with no detergent (35). Ten aliquots of the nuclear suspension were dispensed at a DNA concentration of 1 mg/ml and DNAseI was added to samples 2 through 10 to final concentrations of 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1.0, 2.0, and 5.0 ug/ml respectively. After incubation at 37° for 5 min an equal volume of stop solution (10mM TrisHCl pH 7.5, 300 mM NaCl, 5mM EDTA, 0.5% SDS, and 500 ug/ml Proteinase K) was added and the DNA was purified by phenol-chloroform extraction and ethanol precipitation.

RESULTS

Molecular Cloning of the HGH Gene Family

To isolate the HGH gene family on large segments of DNA, we constructed a cosmid library from human placental DNA and screened it with two radiolabelled probes: a 500 basepair EcoRI-BamHI fragment from the 5' flanking region of HGH-N which cross hybridizes with the corresponding region in the HGH-V and HCS genes (13); and a 120 basepair PstI-PvuII fragment from the 5' end of an HCS cDNA clone (13). Four colonies hybridized with these probes and the cosmids isolated from two of these, cGH4 and cGH21, were analyzed in detail.

Measurement of EcoRI and BamHI restriction fragments from the two cosmids revealed that they carried inserts of 43 and 39 kb. Figure 1 presents restriction maps which were generated with the enzymes EcoRI, BamHI, ClaI, HindIII, KpnI, SalI, SstII, and XhoI. The mapping strategies which we used are discussed in more detail in Materials and Methods. The alignment of the two cosmids with respect to one another was facilitated by the presence of a 1.1 kb BamHI fragment in both cosmids which contains the 5' coding sequences for the HGH-V gene.

Linkage Arrangement of HGH and HCS Genes in Cloned and Genomic DNA

From the restriction map of the 78 kb of cloned DNA, we inferred the exact linkage arrangement and transcriptional orientation of five HGH and HCS genes. All five genes have the same transcriptional orientation and are arranged 5' : HGH-N : HCS-L : HCS-A : HGH-V : HCS-B : 3' with the HGH-N gene separated by about 45 kb from the HCS-B gene.

Our rationale for using the above nomenclature is as follows. The HGH-N and HGH-V genes correspond to the well-described normal and variant growth hormone genes, respectively (13,22). They are both located on 2.6 kb EcoRI fragments but can be readily distinguished by the presence of a BamHI site within the fourth intron of the variant gene (13). The HCS-A and HCS-B genes are both located on 2.9 kb EcoRI fragments, but the HCS-A gene contains two PvuII sites whereas the HCS-B gene contains only one (data not shown). Thus, the reported HCS gene sequence from a genomic 2.9 kb EcoRI fragment (13) corresponds to that of the HCS-B gene, but the DNA sequence of a cloned HCS cDNA (8,13) corresponds to the HCS-A gene. We have used the suffixes -A and -B since both genes have been shown to be expressed *in vivo* giving rise to distinct mRNAs that code for identical polypeptides (15). The gene located on the 9.5 kb EcoRI fragment is more closely related to the HCS-A and HCS-B genes than the HGH-N and HGH-V genes by DNA sequence analysis

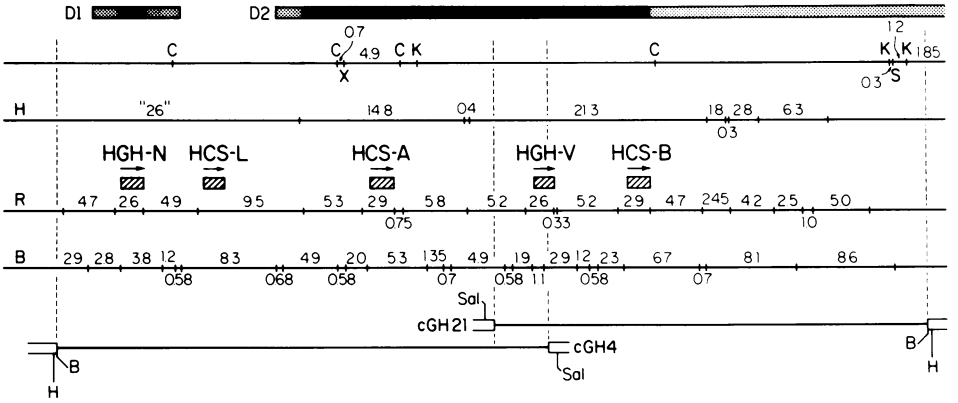

Figure 1. Restriction Map of the Cloned DNA Representing the HGH Locus.

Complete restriction maps of cGH4 and cGH21 are shown for BamHI, EcoRI, HindIII, KpnI, SstII, and XhoI. Only ClaI sites which were apparent in the cloned DNA (grown in a dam⁻ host) are indicated. Vector and insert sequences are indicated by open bars and solid lines, respectively, in the lower portion of the figure. The sizes of restriction fragments or location of restriction sites not present in the cloned DNA which are inferred by comparison with genomic blots are indicated by quotation marks. The approximate location of breakpoints for two deletion mutants D1 and D2 of HGH (22) and HCS (24) are indicated in the upper portion of the figure and were determined as follows. The individual homozygous for a 7.5 kb deletion which encompasses HGH-N coding sequences (22) has an intact 8.3 kb BamHI fragment which contains the HCS-L gene. This places the right endpoint of the HGH deletion between the 3' end of the HGH-N gene (approximately 800 bp from the end of the 2.6 kb EcoRI fragment) and the 5' BamHI site of the 8.3 kb BamHI fragment. The left endpoint of this deletion (D1) is indicated 7.5 kb distant. The individual with a deletion of the HGH-V gene and the HCS genes normally found on 2.9 kb EcoRI fragments also has an intact 8.3 kb BamHI fragment (24) which contains the HCS-L gene, but the 9.5 EcoRI fragment which normally contains the HCS-L gene is missing in this patient and is apparently replaced with a 10.0 kb EcoRI fragment. This places the left endpoint of the HCS deletion between the 3' end of the 8.3 kb BamHI fragment and the 5' end of the 5.3 kb EcoRI fragment. The right endpoint of this deletion is unknown, but must lie to the right of the 3' end of the HCS-B gene and places a minimum estimate of 30 kb on the size of the HCS deletion (D2).

(H.A.Barrera-Saldana and P.H.S., unpublished data). It cross-hybridizes to a probe from the 3' flanking region of HCS-B which is HCS-specific (ref.13, and Fig. 2). Therefore, we have named this gene HCS-L ("HCS-like").

To determine whether the restriction map derived from the cloned DNA accurately represents the genomic arrangement of the HGH gene family, we compared the map with the results of blot hybridization experiments utilizing the genomic DNA used to construct the cosmid library. For example, since the immediate 5' flanking sequences of the HGH-N, HCS-A, HCS-B, and HGH-V genes

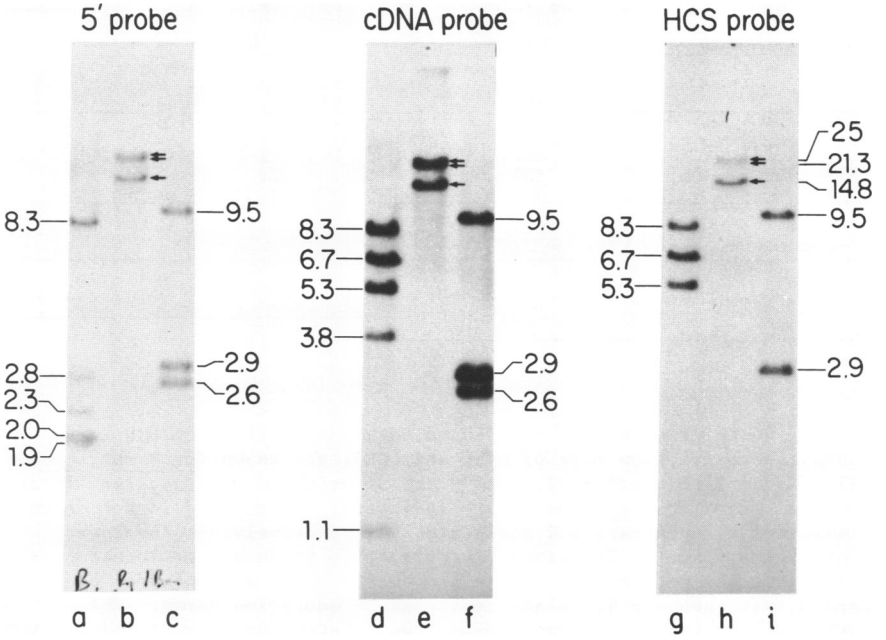

Figure 2. Linkage arrangement of the HGH and HCS genes in genomic DNA.
Ten ug of the placental DNA used to construct the cosmid library was digested with BamHI (lanes a,d,g), HindIII (lanes b,e,h), or EcoRI (lanes c,f,i), and transferred to nitrocellulose after electrophoresis on a 0.7% agarose gel (21). 5' probe: a 500 base pair EcoRI-BamHI fragment from the 5' flanking region of HGH-V (13). cDNA probe: a 120 base pair PstI-XbaI fragment from the 5' end of an HCS-A cDNA clone. HCS probe: a 700 base pair BglII-EcoRI fragment from the 3' flanking region of HCS-B.
We cannot predict the size of the genomic HindIII fragment which contains the HGH-N and HCS-L genes from the restriction map of cGH4 because there are no HindIII sites between the vector/insert junction and the beginning of the HGH-N gene (see Fig.1). However, the HindIII site located 22 kb in from the 5' end of the cloned DNA, is presumably part of the 25 kb HindIII fragment detected in genomic DNA as shown. Molecular weight estimates of hybridizing bands were inferred by comparison with HindIII fragments of phage lambda DNA.

are approximately 95% homologous and are contained on 500-bp long EcoRI-BamHI fragments (ref. 13 and P.H.S., unpublished data), the data in Figure 1 would predict that a labelled 500 bp EcoRI-BamHI fragment from this region of the HGH-V gene should hybridize to genomic BamHI fragments of 2.8, 2.0, 1.9, and 2.3 kb, containing immediate and further 5' flanking areas of the HGH-N, HCS-A, HGH-V, and HCS-B genes, respectively. This prediction is borne out as shown in Figure 2, a; in addition, the 8.3 kb BamHI fragment containing HCS-L gene sequences hybridizes to the same probe. This latter observation

suggests that 5' flanking sequence homology is maintained in HCS-L, but that the BamHI site corresponding to the site of transcriptional initiation in the other genes is absent (13). The 5' flanking probe also hybridizes to 2.6, 2.9, and 9.5 kb EcoRI fragments with the intensities expected for two HGH genes (N and V), two HCS genes (A and B), and one HCS-L gene as shown in Figure 2, c.

Similarly consistent results were obtained with a probe derived from a 120 bp fragment from the 5' end of a cloned HCS-A cDNA (13), which hybridized to genomic BamHI fragments of 3.8, 8.3, 5.3, 1.1, and 6.7 kb as shown in Figure 2, d. These fragments carry the HGH-N, HCS-L, HCS-A, HGH-V, and HCS-B genes, respectively. The reduced intensity of hybridization to the 3.8 and 1.1 kb BamHI, and 2.6 kb EcoRI fragments visible in Figure 2, d and f, is due to the sequence mismatch of about 10% between the HCS-A probe and the HGH-N or -V genes in this region (13). Figure 2, lanes g through i show the hybridization pattern of an HCS-specific 3' flanking probe which, as predicted, hybridized to 8.3, 5.3, and 6.7 kb BamHI fragments (Fig. 2, g), and 9.5 and 2.9 kb EcoRI fragments (Fig. 2, i).

As predicted by the restriction map of the cloned DNA, the hybridization pattern of genomic DNA digested with HindIII shows that 14.8, 21.3, and 25 kb fragments of about equal intensity were detected with all three probes mentioned above which is shown in lanes b,e, and h of Figure 2. Thus, the 21.3 kb HindIII fragment contains the HGH-V and HCS-B genes, and the 14.8 kb HindIII fragment contains the HCS-A gene. Every hybridizing genomic fragment generated by BamHI, EcoRI, and HindIII is accounted for in the restriction map derived from the cosmid DNAs. Furthermore, allowing for slight variations in estimating fragment sizes from genomic blots, results consistent with the map shown in Figure 1 have been obtained from over 50 different individuals examined in several different laboratories in which genomic DNA was digested with BamHl, HindIII, or EcoRl and probed with HGH coding sequences (13,22-24).

Other groups have suggested that there are at least seven members in the HGH family based on the results of molecular cloning experiments with lambda vectors (18-20), but their conclusions are not well-supported by genomic blotting data. The additional genes postulated by these groups (18,19) might be attributed to allelic polymorphisms. Perhaps more likely, the high density of repetitive elements within the HGH locus could have led to rearrangements during the cloning process, as has been well-documented with similar phage vectors in other systems (25,34).

One striking feature of the HGH gene cluster is its very high density of middle repetitive DNA. A probe made from nick translated total human DNA hybridized to every EcoRl and BamHl fragment of cGH4 and cGH21, (with the exception of the 0.33 kb EcoRl fragent) which indicates that a large number of middle repetitive DNA elements exist in the HGH gene family (hybridization not shown). By direct sequence analysis (13 and Barrera-Saldama and P.H.S., unpublished), and in one study of recombinant lambda clones containing HGH or related genes (18), both partial and complete elements of the Alu family were found 3' to every HGH and HCS gene,respectively, and at multiple locations between the genes. Based on these results, we estimate there are at least twenty Alu elements within the HGH cluster, which is about four times more than are found in the beta-globin cluster (46).

Chromatin Structure of the HGH Locus

Because HGH and HCS are expressed in different tissues, we asked if the chromatin structure of the locus might vary in a tissue specific pattern. It is generally believed that production of HGH-N is restricted to somatotrophic cells in the anterior pituitary, and that production of HCS-A and -B is restricted to trophoblastic cells in the mature placenta. Expression is thought to be regulated at the level of transcription (8,14). At present there is no evidence to suggest that HGH-V or HCS-L are expressed in either tissue.

We determined the pattern of DNAse I sensitivity in HGH chromatin from pituitary and placental cells by digesting isolated nuclei with a range of DNAse I concentrations. After purifying the DNA and digesting with BamHI, a 500 bp EcoRI-BamHI probe from the 5' flanking region of HCS-B was used in blot hybridizations to detect the unique BamHI fragments from the five HGH and HCS genes, as shown in Figure 3. Initial observations suggested that HGH chromatin was only slightly more sensitive to DNAse I digestion than bulk or inactive chromatin (data not shown). Closer inspection revealed distinct differences between the DNAse I sensitivity of pituitary and placental HGH chromatin.

In particular, Figure 3 shows that several DNAse I hypersensitive sites are present in HGH chromatin from placenta which are not apparent in HGH chromatin from pituitary. The appearance of discrete, low-intensity "sub-bands" at concentrations of DNAse I which are approximately 100-fold below that required to digest active chromatin (as shown in Figure 3, lane 4) indicate the presence of at least three and possibly four hypersensitive sites within the 8.3 BamHI fragment which contains HCS-L. The sizes of these

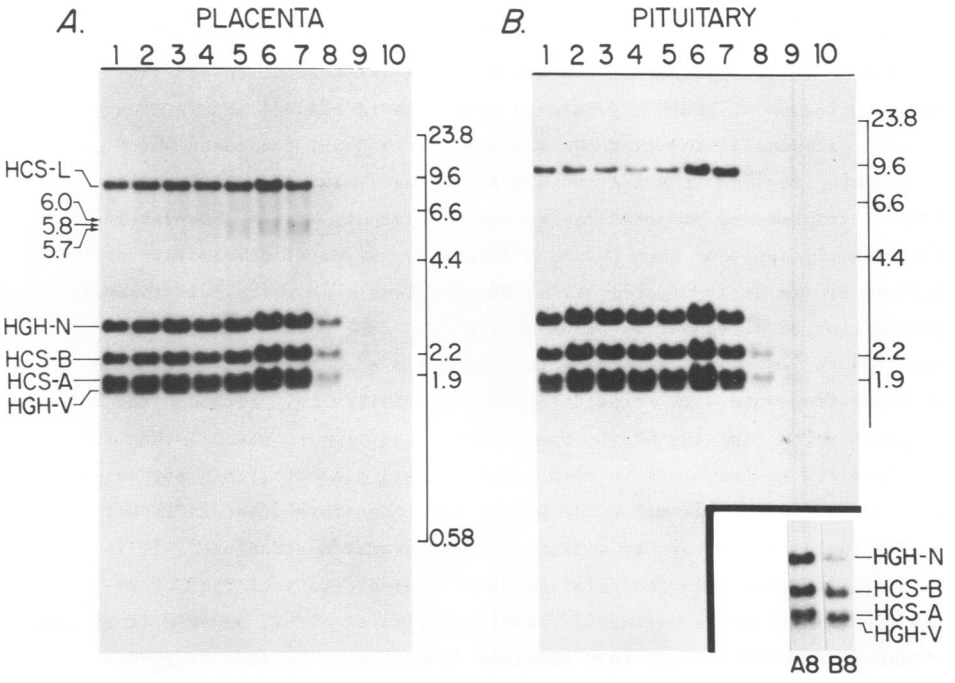

Figure 3. DNAse sensitivity of the HGH-like gene cluster in placenta and pituitary.

Nuclei were prepared from human placenta or pituitary and digested with DNAse I at concentrations which varied from 0.01 to 5.0 ug/ml as described in Materials and Methods. After digestion, the DNA was purified and ten ug from each sample was digested with BamHI, fractionated on 0.7% agarose gels and transferred to nitrocellulose membranes (21). The filters were processed for hybridization and the 500 bp EcoRI-BamHI fragment from the 5' flanking region of the HCS-B gene was used as probe. A, placental nuclei with the identity of each band (Fig.1) indicated in the left margin. Arrows indicate the hypersensitive sites detected 3' to the HCS-L gene in lanes 5 through 9 in placental nuclei. B, pituitary nuclei treated as in A. Insert in panel B: lane 8 from a darker exposure of panels A and B remounted side by side to emphasize the differential DNAseI sensitivity of the HCS-B gene in placental nuclei and the HGH-N gene in pituitary nuclei.

sub-bands (6.0, 5.8, and 5.7 kb) indicate that hypersensitive sites map 2.1,2.2, and 2.4 kb 3' to the polyadenlyation consensus sequence of HCS-L. Hypersensitive sites close (within a few kb) to the 5' end of any of the five genes, or close to the 3' end of any of the other genes might not have been demonstrable with the particular restriction enzyme and probe combination used here. Although the characterization of the hypersensitive sites between HCS-L and HCS-A is continuing, their presence in placenta but not in

pituitary is direct evidence in support of tissue-specific differences in HGH chromatin structure.

Figure 3 also shows that the 2.8 kb BamHI fragment which contains the 5' flanking region of HGH-N may be more sensitive to DNAse I than HCS -A or -B in pituitary nuclei and that the 2.3 and 2.0 kb BamHI fragments which contain 5' flanking regions of HCS-A and HCS-B, respectively, are more sensitive to DNAse I in placenta compared to the HGH-N fragment. These observations are most clearly apparent when the band intensity ratios of all 3 fragments (2.8, 2.3 and 2.0 kb) are compared at the DNAse I levels at which bulk chromatin just begins to disappear as shown in Figure 3 lane 8 and inset. From these experiments, we cannot determine the chromatin structure of the 8.3 and 1.9 kb BamHI fragments with respect to DNAse I sensitivity. Although the 8.3 kb fragment which contains HCS-L appears more sensitive to DNAse I than the 2.8, 2.3, and 2.0 kb fragments in both pituitary and placenta, this may be caused by the relatively high molecular weight (and therefore greater "target size") of this band rather than by differences in chromatin structure. It is difficult to determine the relative DNAse I sensitivity of the 1.9 kb fragment, which contains the 5' flanking region of HGH-V, because it is less intense than, and incompletely resolved from, the other four fragments. Nonetheless, preferential digestion of the 2.8 kb fragment in pituitary and preferential digestion of the 2.0 kb fragment in placenta is consistent with a simple model for HGH chromatin structure. In pituitary, an active chromatin conformation may be restricted to the left end of the cluster (as defined by transcriptional orientation, but in placenta, an active chromatin conformation may be restricted to the right end of the cluster. A determination of the relative DNAse I sensitivity of HCS-L, HGH-V, and intergenic regions using other restriction enzyme and probe combinations, especially oligonucleotides, which will discriminate between the five genes will help to confirm and refine this model by determining the borders of DNAse I sensitive regions more precisely.

Duplication Units in the HGH Cluster

An analysis of the DNA sequences of the HGH-N, HGH-V, and HCS-B genes (13) and the HCS-A and HCS-L genes (H.A. Barrera-Saldana, and P.H.S., unpublished data) combined with genomic blot hybridizations was used to study the duplicaiton units in the HGH-family.

DNA sequence comparisons show that the five genes are highly conserved from at least 500 nucleotides 5' to the start of transcription to the polyadenlyation consensus sequence. Within this region of about 2200 base

pairs, there are no systematic differences between genes of the HGH and HCS group. In the 3' flanking region however, high sequence conservation is maintained between HCS genes for at least 700 nucleotides, and between HGH genes for at least 400 nucleotides, but there is no detectable homology between the two groups of genes. The HGH gene group diverges from the HCS gene group within an inverted element of the middle repetitive Alu family which begins about 100 nucleotides downstream of the polyadenylation site (13). This Alu sequence continues for about 270 nucleotides in HGH-N and HGH-V, but is suddenly truncated after about 25 nucleotides in HCS-A, -B, and -L.

To extend the homology relationships beyond the sequenced regions, we examined the genomic hybridization patterns of probes that flanked the 2.6 kb HGH-V EcoRI fragment. The 700 bp BglII-EcoRI fragment 5' to HGH-V hybridizes to 1.9, 2.0, 2.3, 2.8, and 8.3 kb BamHI fragments; 4.7, 4.9, 5.2, and 5.3 kb EcoRI fragments; and 14.8, 21.3, and 25 kb HindIII fragments which are visible in the blot-hybridizations presented in lanes a through c of Figure 4. This demonstrates that the high degree of sequence conservation in the 5' flanking region of all five genes extends through some if not all of the 700 bp covered by this probe.

Data in Figure 2 shows that a probe derived from the 3' flanking region of HCS-B cross-hybridizes to 3' flanking regions of HCS-L and HCS-A but not to the corresponding regions in HGH-N and HGH-V which contain Alu sequences (see also Fig. 2 in ref.13). The 330 bp EcoRI fragment located 3' to the EcoRI fragment carrying the HGH-V gene (see Fig. 1) hybridizes to 2.9 and 3.8 kb BamHI fragments, a 4.9 kb EcoRI fragment, and 21.3 and 25 kb HindIII fragments which is shown in lanes d through f of Figure 4. Thus the high degree of homology between the 3' flanking sequences of HGH-N and HGH-V, not shared by the HCS genes, extends at least part way through the 330 bp EcoRI fragment 3' to HGH-V.

Taken together, these results show that the HGH cluster contains at least three types of duplication units (Fig. 4, bottom). One type contains the coding regions for five genes and their 5' flanking sequences, a second type contains 3' flanking sequences of the two HGH genes, and a third type contains 3' flanking sequences of the three HCS genes. This pattern suggests that the unit we recognize in the present-day locus which contains 5' flanking and coding sequences of all five genes is the descendant of the ancient unit which gave rise by duplication to progenitors for HGH and HCS. Furthermore, because HGH-N is at the 5' end and HCS-B at the 3' end of the

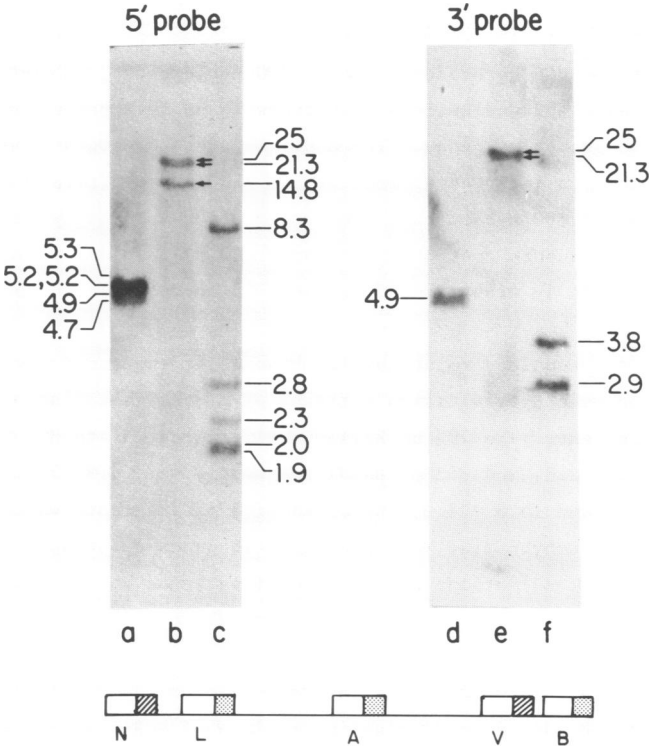

Figure 4. Duplication units in the HGH cluster.

A, 10 ug of the placental DNA used to construct the cosmid library was digested with BamHI (lanes a and d), HindIII (lanes b and e), or EcoRI (lanes c and f), and transferred to nitrocellulose after electrophoresis on a 0.7% agarose gel. Hybridizations were as described in Methods and the legend to Fig. 2. 5' probe: a 700 base pair BglII-EcoRI fragment immediately 5' to the 2.6 kb EcoRI fragment which contains the HGH-V gene. 3' probe: a 330 base pair EcoRI fragment immediately 3' to the 2.6 kb EcoRI fragment which contains the HGH-V gene. B, Lower portion: diagrammatic representation of the duplication units. Open boxes contain 5' flanking and coding regions from all five genes; hatched boxes and dotted boxes contain 3' flanking regions of HGH genes (-N and -V) and HCS genes (-A, -B, and -L), respectively. Sequences distantly related to the hatched boxes may also be present 5' to the HGH-N, HCS-A, and HGH-V genes (see Fig. 5). The left-hand border of the open boxes has not been precisely mapped, but is at least several hundred nucleotides 5' of the 2.6, 2.9, or 9.5 kb EcoRI fragments which contain the genes. The right-hand borders of the hatched and dotted boxes have also not been precisely mapped, although they are at least 700 nucleotides 3' to the HGH and HCS termination codons. The boundary between open boxes and dotted or hatched boxes corresponds to the 3' endpoint of the 25 nucleotide truncated Alu sequence found 3' of HCS-A, -B, and -L.

present day five gene locus, the structure of the proposed ancestral two gene locus was probably 5' : HGH : HCS : 3'.

Molecular Evolution of the HGH Cluster and the Structure of Deletion Mutants

Although we have not yet mapped the precise 5' boundary of the initial duplication unit, the 3' boundary is located within the nearly full-length 270 nucleotide Alu sequence 3' of HGH-N and HGH-V (see above). Because the corresponding Alu sequence 3' of HCS-A, -B, and -L is truncated after approximately 25 nucleotides, and because HCS-B lies at the 3' end of the present day five gene cluster, the primordial GH-like gene probably contained a truncated Alu sequence on its 3' side (line A of Figure 5). Such reasoning suggests that the ancestral two gene locus was produced from a single GH-like gene by homologous recombination between truncated and complete Alu sequences (Figure 5 lines A and B). Thus, the nearly full-length Alu sequence 3' to HGH-N and HGH-V may represent the recombinant product of the initial HGH/HCS duplication. Although it is not clear how a truncated Alu sequence 3' to the primordial GH-like gene might have been generated initially, we predict a complete Alu sequence will coincide with the 5' end of the ancestral HGH/HCS duplication unit.

Formation of the present five gene locus from the ancestral two gene locus probably involved homologous but unequal exchange between the initial duplication units. However, the three HCS genes share 3' flanking sequences which lie outside of the initial HGH/HCS duplication unit, which suggests that expansion of the HGH cluster from two genes to five genes may also have involved recombination of unrelated sequences. In considering several molecular pathways that could account for this expansion, we found that the structure of the present day locus is most easily explained by a simple combination of both mechanisms which are presented on lines C through E of Figure 5. We propose that a four gene locus, 5' : HGH : HCS : HGH : HCS : 3', was generated from the two gene cluster, 5' : HGH : HCS : 3', by recombination between two points outside the boundaries of the initial HGH/HCS duplication. Homologous but unequal exchange of the four gene cluster then produced the present day five gene locus.

Although the intermediate structures we have diagrammed are hypothetical, this pathway predicts a characteristic pattern of homology between intergenic regions which agrees almost perfectly with the pattern of intergenic BamHI and EcoRI restriction fragments as determined from the cloned genomic DNA. Furthermore, this pathway predicts that the divergence of HCS-B from HCS-A

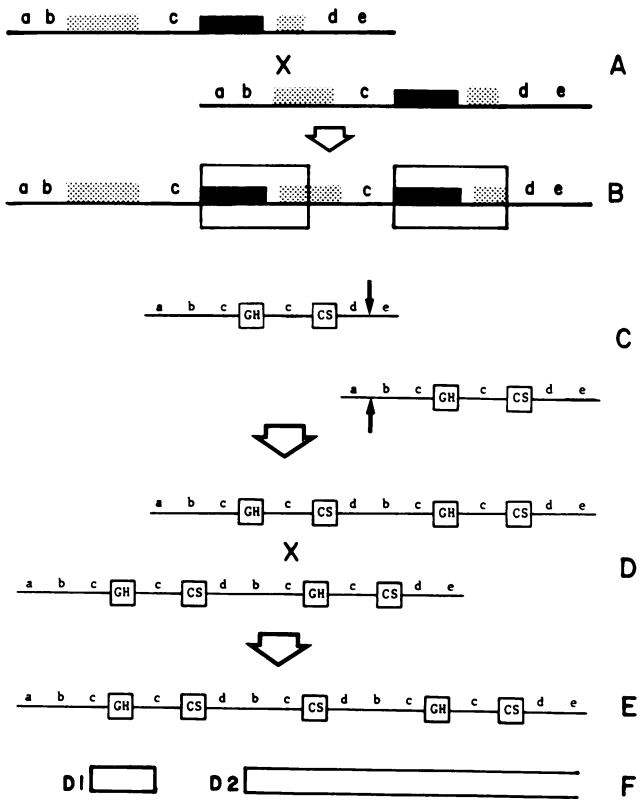

Figure 5. Molecular evolution of the HGH gene cluster.

A, a primordial GH-like gene (solid boxes) with full length and truncated Alu sequences (dotted boxes) on 5' and 3' sides, respectively. Lower case letters represent the DNA sequences of flanking regions. B, homologous but unequal crossing over between the two Alu sequences explains the structure of an ancestral two gene locus. We believe the left hand gene gave rise to HGH-N and -V, which both have full-length Alu sequences in their 3' flanking regions; whereas the right hand gene gave rise to HCS-A, -B, and -L, which all have truncated Alu sequences in their 3' flanking regions. Open boxes contain portions of the initial duplication unit helpful in depicting the subsequent steps in molecular evolution of the cluster. The 5' border of all of the open boxes is arbitrarily placed at the 5' ends of coding sequence; the 3' border of all of the open boxes corresponds to the 3' end of the initial duplication unit. C and D, intermediate structures in the expansion of the locus from two to five genes. Because the only characteristic sequence differences that we can detect between HGH and HCS genes in the present day cluster are in 3' flanking regions, we have labelled individual genes in intermediate structures as "GH" or "CS" based on the identity of their 3' flanking sequences, c or d, relative to the ancestral two gene locus. C,formation of a four gene locus from a two gene locus by breakage and reunion between intergenic regions a:b and d:e produces the recombinant region d:b (solid arrows mark breakpoints). This event cannot simply be explained by unequal crossing over between tandemly repeated homologies

because the region d:e lies outside of the initial duplication unit. D, formation of a five gene locus from a four gene locus by homologous but unequal crossing over at sequence element c, which lies within the initial duplication unit. E, the predicted pattern of intergenic homologies in the present day five gene locus. This pattern corresponds very well with the one determined experimentally (Fig. 1,4). F, the extent of HGH and HCS deletion mutants described in the legend to Fig.1.

and HCS-L predates the divergence of HCS-A from HCS-L, which is substantiated by DNA sequence analysis (G.S.B., P.H.S., & R.E.G., manuscript in preparation). It is possible that one, two, and four gene loci with characteristic Alu flanking sequences as diagrammed in Figure 5 may be represented in the genomes of lower primates.

The mechanism of evolution of the HGH gene cluster could be reflected in the generation of the deletion mutants which have been described (22,24). Lane f of Figure 5 shows the breakpoints of the deletions based on the available data from genomic blot hybridizations which was presented in Figure 1. In both cases terminal portions of the cluster are missing. In one case, a 7.5 kb fragment which contains the HGH-N gene has been deleted (22); in another, a fragment at least 30 kb in length which contains the HCS-A, HGH-V,and HCS-B genes is missing (24). These deletions could not have been generated by homologous but unequal crossing over between tandemly repeated HGH or HCS genes, because this type of recombination could have removed only internal fragments. These deletions were probably generated by recombination between repetitive elements unrelated to HGH, similar to the mechanisms involved in the initial HGH/HCS duplication. DNA sequence analysis of the intergenic regions in the normal cluster and a determination of the precise breakpoints of the deletion mutants will yield more insight into the particular sequences involved in the crossing over.

DISCUSSION

We have now isolated the entire HGH gene family. The structure of the HGH chromosomal locus as determined from analysis of two DNA segments carried in cosmid vectors agrees very well with extensive mapping data generated from genomic blots (22,23), although we believe the 16, 23, and 26 kb *Hin*dIII fragments described in earlier experiments correspond more precisely to fragments we observe of 14.8, 21.3, and 25 kb. To some extent, our results agree with restriction maps of recombinant lambda phages isolated by other groups (18,19) but there are several notable discrepancies. In particular, we have found no evidence for an HGH/HCS hybrid gene (20), an HGH or HCS gene

missing the first exon (20), an HCS gene with a transcriptional polarity opposite to that of other members of the family (18), or an HGH or HCS gene contained within a 3.4 kb HindIII fragment (18). The linkage map we propose is supported by genomic blot experiments with DNA from many individuals, but very rare polymorphisms in restriction fragment size and gene number might still explain some of the above discrepancies. It seems most likely, however, that the high density of repetitive sequences in the HGH locus (18) may have predisposed certain recombinant phage to rearrange in E. coli perhaps by pathways formally analogous to the major evolutionary steps of this locus in humans (18,25).

Like several other gene families studied to date the growth hormone gene family is clustered rather that dispersed. However, the relatively recent expansion of this cluster offers some unique insight into mechanisms of molecular evolution. All five genes in the present day HGH cluster manifest a high degree of homology throughout 5' flanking and coding regions. For any homologous pair of introns, the magnitude of sequence divergence is less than 10% (G.B., R.G., & P.H.S., manuscript in preparation). By analogy to the globin and preproinsulin gene family, in which the rate of non-selective divergence is at least 8% every 10 million years (39,40), the ancestral HGH gene probably diverged from the ancestral HCS gene very recently, less than 15 million years ago. Thus, the HGH gene family has expanded from one to five genes in approximately one twentieth of the time for a similar expansion to occur in the human beta-like globin gene family (40). This rapid expansion in gene number may have been driven by a combination of selective forces and the structural characteristics of the DNA surrounding the progenitor to the HGH-N gene.

We believe that the high density of Alu elements in the HGH gene cluster is causually related to the mechanism of gene duplication. The pattern of duplication units in the present day cluster suggests that unequal exchange between homologous (but "non-allelic") Alu family members played a prominent role in the evolution of the cluster. It has been suggested that Alu sequences in particular, and middle repetitive elements in general, may serve as "hot spots" for recombination, by a mechanism which remains speculative (37). The few reported examples of recombination possibly mediated by Alu sequences involve breakage within the consensus sequence, followed by a reunion with non-homologous DNA (37,38). Our results do not speak to the relative frequency of different recombinational mechanisms, but do suggest that in some cases, recombination of ostensibly unrelated sequences may occur

by a mechanism very similar to meiotic exchange, mediated by pre-existing repetitive elements instead of allelic sequences on homologous chromosomes.

What are the selective forces responsible for the expansion of the HGH gene cluster? Although the biology of growth hormone (HGH-N) is well understood and the selective advantage of a hormone whose action results in greater size and strength is easily rationalized, the function(s) of the other four genes in the cluster are less clear. The HCS-A and HCS-B genes differ very slightly in DNA sequence and they give rise to identical mature polypeptide chains (13,15). Chorionic somatomammotropin is thought to function as a metabolic hormone of pregnancy which may allow fetal growth in times of maternal starvation as well as having mammotrophic and lactogenic effects (25,30-31,41-43). However, the existence of a normal individual who has a homozygous deletion of HCS-A and -B demonstrates these genes are not required for human gestation (24). However, CS may have a subtle and unrecognized beneficial influence on maternal physiology and thus could have been selected for during the evolution of lower primates (24). There is no evidence for HCS production by any tissue other than placenta, but it is conceivable that HCS-A and -B could be synthesized at different times during placental development. It is likely that the HCS-L gene does not give rise to a polypeptide analogous to the HCS-A or -B product because of a G to A transition at the start of the second exon which abolishes a consensus splice donor sequence (H.A. Barrera-Saldana and P.H.S., unpublished data). A similar mutation has been proposed as rendering the human beta globin gene inactive in a form of beta^{o}-thalassemia(33).

Transgenomic expression experiments show that HGH-V can be expressed (44). But, because nine of the fifteen substitutions predicted to occur in HGH-V produce a marked change in side chain hydrophobicity or charge, any HGH-V produced *in vivo* might well have end organ effects different from HGH-N (13). Patients who are homozygous for HGH-N deficiency but retain the HGH-V gene are severely dwarfed (24); conversely, patients homozygous for HGH-V deficiency with intact HGH-N genes appear normal (24). Thus, there is no apparent physiological role for HGH-V at present, although HGH-V might have had a function during the evolution of lower primates. It is even possible that deletion of 4 of the 5 genes in the cluster (HCS-L, HCS-A, HCS-V and HCS-B) would have no detrimental effects for an individual.

Regardless of whether or not HCS-L and HGH-V are expressed *in vivo*, our preliminary analysis of chromatin structure helps to understand the factors which control expression of the HGH-N, HCS-A, and HCS-B genes. We assume

that there is some structural difference between the gene expressed in pituitary (HGH-N) and those expressed in placenta (HCS-A and -B). This difference could be based on both position, as reflected in the chromatin structure of the HGH locus, and sequence-specific regulatory elements around the individual genes. There are tissue-specific differences in the chromatin structure of HGH and HCS genes, consistent with a two domain model for position-dependent regulation of the HGH gene family. A "left-hand" domain containing the HGH-N gene seems to be activated in pituitary, and a "right-hand" domain containing the HCS-A and -B genes is activated in placenta. Although this model is probably an oversimplification, it suggests that the HGH-V gene, which lies between the HCS-A and -B genes within the putative "right-hand" domain, might be expressed in placenta. It will also be important to see if HGH-N and HCS-L are expressable in other tissues of ectodermal origin. With respect to position-independent factors, there are no systematic sequence differences within 500 nucleotides of 5' flanking and coding sequences of the HGH-N, HCS-A, and HCS-B genes. Thus, if tissue-specific regulatory elements are involved in the expression of these genes, they might lie further than 500 base pairs 5' from the start of transcription or 250 base pairs 3' to the termination codons, where the HGH and HCS genes begin to diverge. Sequences which mediate the sensitivity of HGH-N to glucocorticoids after cotransformation into murine fibroblasts are located within the conserved 5' region (47). As systems for transgenomic expression become more refined (44-45), it will be interesting to see what effect the flanking sequences of the HGH and HCS genes have on the expression of unrelated genes.

ACKNOWLEDGEMENTS

The authors wish to thank Richard Harvey, Mark Groudine, George Sale, Hal Weintraub, Hugo A. Barrera-Saldana, Mary Peretz, Tina Roush, Mike Walls, and Ann Kritzberger who helped with various aspects of this work. These studies were supported by grants from the American Cancer Society (NP-342), American Heart Association (80-728), and NIH (AM31232) to R.E.G., and support from Genentech, Inc. to P.H.S. Greg Barsh is a predoctoral fellow in the Medical Scientist Training Program of the School of Medicine, University of Washington (USPHS, NIH GM 07266), and was supported in part by a grant from the March of Dimes, Birth Defects Foundation (6-298).

*To whom correspondence should be addressed

REFERENCES

1. Niall,H.D.,Hogan,M.,Sauer,K.,Rosenblum,I.,and Greenwood,F. Proc.Natn.Acad.Sci.U.S.A.68,866-869(1971).
2. Bewley,T.A.,Dixon,J.S.&Li,C.H. Int.J.Peptide Prot.Res.4,281-287(1972).
3. Shome,B.&Parlow,A.F. J.Clin.Endocrinol.Metab.45,1112-1115(1977).
4. Owerbach,D.,Rutter,W.,Martial,J.,and Baxter,J. Science209,289-292(1980).
5. George,D.L.,Phillips,J.,Francke,U.,and Seeburg,P. Hum.Genet.57,138-141 (1981).
6. Harper, M.E., Barrera-Saldana, H.A. and Saunders, G.F. Am. J. Hum.Genetics34,227-234(1982).
7. Cooke,N.E.,Coit,D.,Shine,J.,Baxter,J.,and Martial, J.Biol.Chem.256, 4007-4016(1981).
8. Shine,J.,Seeburg,P.,Martial,J.,Baxter,J.,and Goodman,H. Nature270, 494-499(1977).
9. Fiddes,J.,Seeburg,P.,Martial,J.,Baxter,J.,and Goodman,H. Proc.Natn.Acad.Sci.U.S.A.76,4294-4298(1979).
10. Roskam,W.G.&Rougeon,F. NucleicAcidsRes.7,305-320(1979).
11. Seeburg, P. In Polypeptides Hormones. R.F. Beers and E.G. Bassett. Eds. (Raven Press, New York),. 19-31(1979)
12. DeNoto,F.M.,Moore,D.D.&Goodman,H. NucleicAcidsRes.9,3719-3730(1981).
13. Seeburg,P. DNA1,239-249(1982).
14. Seeburg,P.,Shine,J.,Martial,J.,Ullrich,A.,Baxter,J.,and Goodman,H. Cell12,157-165(1977).
15. Barrera-Saldana, H.A.,Robberson,D.L.,and Saunders,G. J.Biol.Chem. 257,12399-12404(1982).
16. Weintraub,H.,Larsen,A.&Groudine,M. Cell24,333-344(1981).
17. Elgin,S.C. Cell27,413-415(1981).
18. Kidd,V.J.&Saunders,G.F. J.Biol.Chem.157,10673-10680(1982).
19. Moore,D.D. Recent Progress in Hormone Research38,197-222(1982).
20. Moore,D.D.,Conkling,M.A.&Goodman,H.M. Cell29,285-286(1982).
21. Southern,E.M. J.Molec.Biol.98,503-517(1975).
22. Phillips,J.A.,Hjelle,B.,Seeburg,P.and Zachmann,M. Proc.Natn.Acad.Sci.U.S.A.78,6372-6375(1981).
23. Phillips,J.A.,Parks,J.,Hjelle,B.,Herd,J.,Plotnick,L.,Migeon,C.and Seeburg,P. J.Clin.Invest.70,489-495.(1982).
24. Wurzel,J.M.,Parks,J.S.,Herd,J.E.,and Neilsen,P.V. DNA1,251-257(1982).
25. Lauer,J.,Shen,C.&Maniatis,T.M. Cell20,119-130(1980).
26. Blin,N. & Stafford,D.W. Nucleic Acids Res.3,2303-2308(1976).
27. Hohn,B. & Collins,J.C. Gene11,291-298(1980).
28. Ish-Horowicz,D. & Burke,J. Nucleic Acids Res.9,2989-2998(1981).
29. Grosveld,F.G.,Dahl,H.,deBoer,E.,and Flavell,R. Gene 13,227-237(1981).
30. Saxena,B.N. Vitam.and Hormones 29,95-151(1971).
31. Friesen,H.G.,Suwa,S.and Pare,P. Recent Prog. Horm. Res.25,161-205(1969).
32. Rigby,P.J.,Dieckmann,M.,Rhodes,C.,and Berg,P. J.Molec.Biol.113,237-251 (1977).
33. Baird, M.,Driscoll,C.,Schreiner,H.,Sciarratta,G.,Sansome,G.,Niazi,G., Ramirez,F.,and Bank,A. Proc.Natn.Acad.Sci. U.S.A.78,4218-4221(1981).
34. Liu,C.-P.,Tucker,P.,Mushinski,F.,and Blattner,F. Science209,1348-1353 (1980).
35. Stalder,J.,Groudine,M.,Dodgson,J.,Engel,J.,and Weintraub,H. Cell19,973-980(1981).
36. Shen,S.,Slightom,J.L.&Smithies,O. Cell26,191-203(1981).
37. Jagadeeswaran,P.,Pan,J.,Spritz,R.,Duncan,C.,Biro,P.,Tuan,D.,Forget,B.and Weissman,S. Nature 296,469-470(1982).
38. Dhruva,B.R.,Shenk,T.&Subramanian,K.N. Proc.Natn.Acad.Sci.U.S.A. 77,4514-4518(1980).

39. Perler,F.,Efstratiadis,A.,Lomedico,P.,Gilbert,W.,Kolodner,R.,and Dodgson,J. Cell20,555-566(1980).
40. Efstratiadis,A.,Posukony,J.,Maniatis,T.,Lawn,R.,O'Connell,C.,Spritz,R., Deriel,J.,Forget,B.,Weissman,S.,Slightom,J.,Blechl,A.,Smithies,O., Barralle,F.,Shoulders,C.,and Proudfoot,N. Cell21,653-668(1980).
41. Bolander,F.F.&Fellows,R.E. J.Biol.Chem.251,2703-2708(1976).
42. Talamantes,F.,Ogren,L.,Markoff,E.,Woodard,S.andMadrid,J. Federation Proc.39,2582-2587 (1980).
43. Grumbach,MM.,Kaplan,S.,Sciarra,J.,and Burr,I. AnnalsNewYorkAcad.Sci. 148,501-531(1968).
44. Pavlakis,G.,Hizuka,N.,Gorden,P.,Seeburg,P.,and Hamer,D. Proc.Natn.Acad.Sci. U.S.A.78,7398-7402(1981).
45. Palmiter,R.D.,Brinster,R.,Hammer,R.,Trunbauer,M.,Rosenfeld,M., Brinberg,N.and Evans,R. Nature300,611-615(1982).
46. Fritsch,E.F.,Lawn,R.M.,&Maniatis,T. Cell19,959-972(1980).
47. Robins,D.M.,Paek,I.,Seeburg,P.H.,andAxel,R. Cell29,623-631(1982).
48. Wu,C.&Gilbert,W. Proc.Natn.Acad.Sci. U.S.A.78,1577-1580(1981).