Structure of genes for human growth hormone and chorionic somatomammotropin

*(in vitro packaging/filter hybridization/intervening sequences/DNA sequences)*


1The Howard Hughes Medical Institute Laboratory, Departments of *Biochemistry and Biophysics and †Medicine, and the  †Metabolic Research Unit, University of California, San Francisco, California 94143

Communicated by Donald F. Steiner, June 1, 1979

**ABSTRACT** A 2.6-kilobase (kb) EcoRI restriction endonuclease fragment containing human growth hormone (hGH; somatomammotropin) gene sequences and a 2.5-kb EcoRI fragment containing human chorionic somatomammotropin (hCS; choriomammotropin) gene sequences have been identified by hybridization to cloned cDNA. Human DNA was cleaved with EcoRI and fractionated by preparative agarose gel electrophoresis; DNA in the size range 2–3 kb was ligated to λgt WES-λDNA and viable recombinant bacteriophage were recovered by *in vitro* packaging. After infection of *Escherichia coli* and screening of phage plaques, single isolates of hGH and hCS gene sequences were obtained. Restriction endonuclease mapping showed that the hGH gene contains three intervening sequences interrupting the coding sequence. Partial DNA sequence analysis of the hGH gene, obtained by the chain termination method, confirmed the location of the intervening sequences and the identity of the fragment.

Human growth hormone (hGH) and human chorionic somatomammotropin (hCS) are two closely related polypeptide hormones which have more than 80% of their 191 amino acids in common (1, 2). They have different biological activities and are synthesized in different tissues: hGH in the pituitary and hCS in the placenta. The genes coding for these two hormones provide a good system for studying the organization of structurally related sequences and their tissue-specific expression.

A 550-base-pair *Hae* III fragment cDNA clone coding for amino acids 24–191 of the hCS sequence has been described (3). Recently, the analogous 550-base-pair *Hae* III fragment coding for hGH has also been cloned and its nucleotide sequence has been determined (unpublished data). The nucleotide sequences of these two fragments are 93% homologous.

Further analysis of this system requires the isolation of genomic DNA fragments that contain the entire coding sequences and the regulating elements that may be involved in the differential expression of the genes. This report describes the construction of genomic DNA clones for sequences of both the hGH and hCS genes by utilizing bacteriophage λgt WES-λ as the vector (4) and the bacteriophage λ *in vitro* packaging system of Blattner *et al.* (5).

**MATERIALS AND METHODS**

Human Placental DNA. High molecular weight DNA was extracted from human placenta obtained by caesarian section. The frozen tissue was dispersed in a blender, treated with proteinase K, extracted with phenol/chloroform, and then incubated sequentially with RNase A and proteinase K (6). After extraction with phenol/chloroform, the DNA was precipitated with ethanol.

Preparative Agarose Gel Electrophoresis. EcoRI-digested human placental DNA (17.5 mg) was fractionated on a continuous elution, horizontal, 0.8% agarose gel, essentially as described by Polsky *et al.* (6) except that Seakem HGT (P) agarose was used because this appeared to contain fewer contaminants that inhibit subsequent enzymatic steps. The DNA was collected in approximately 80 fractions, and aliquots of these were assayed by analytical agarose gel electrophoresis followed by filter hybridization to a radioactive hGH cDNA probe.

Filter Hybridization. The filter hybridization technique of Southern (7) was used with the following modifications. Filters [Schleicher & Schuell, 0.45 μm (BA85) nitrocellulose; approximately 15 × 15 cm] were preannealed, in sealed plastic bags, at 42°C in 10 ml 50% formamide/50 mM Hepes, pH 7.0/5-fold concentrated Denhardt’s solution (8, 9)/0.45 M NaCl/0.045 M sodium citrate, pH 7 (3-fold concentrated NaCl/Cit) containing 18 μg of sheared and denatured *Micrococcus lysodeiketicus* DNA and 40 μg of yeast tRNA per ml. Heat-denatured hybridization probe (approximately 0.5–1 × 10⁶ cpm) was added in 0.5 ml of the same solution and the filters were hybridized at 42°C for 1–2 days. The filters were washed at 50°C for 1–2 hr with two or three changes (about 200–500 ml per filter) of 0.1-strength NaCl/Cit and 0.1% sodium dodecyl sulfate (10) and then briefly with 0.1-strength NaCl/Cit at room temperature and were exposed at −70°C to X-Omat R films and Du Pont Cronex Lightning Plus intensifying screens.

Hybridization Probe. Approximately 200 ng of polyacrylamide gel-purified 550-base-pair hGH cDNA fragment was incubated (room temperature for 60–90 min) in 20 μl of a reaction mixture containing 10 mM Tris·HCl at pH 7.5, 10 mM MgCl₂, 1 mM dithiothreitol, 2.5 mg of calf thymus DNA oligonucleotides [prepared by DNase I digestion (11)] per ml, dATP, dGTP, and dTTP at 10 μM each, 0.5 μM [α-³²P]dCTP [2000–3000 Ci/mmol (1 Ci = 3.7 × 10¹⁰ becquerels); NEN or Amersham], and 1 unit of DNA polymerase I (Klenow modification, Boehringer Mannheim). The cDNA fragment had previously been heat denatured in the presence of the calf thymus oligonucleotides. Routinely, specific activities in the order of 2 × 10⁶ cpm/μg were obtained. The unincorporated

Abbreviations: hGH, human growth hormone (somatomammotropin); hCS, human chorionic somatomammotropin (choriomammotropin); NaCl/Cit, 0.15 M NaCl/0.015 M sodium citrate, pH 7.0; kb, kilobase.  
† Present address: Genentech Inc., San Francisco, CA 94080.
[\alpha^{-32}p]dCTP was then removed by gel filtration on Sephadex G-50.

**Bacteriophage \lambda Packaging.** The bacteriophage \lambda in vitro packaging system was used exactly as described by Blattner et al. (5). The EK2 vector \lambdagt WES-\lambdaB (4) was used after either separation of the B fragment from the vector arms on a 5–20% sucrose gradient (0.1 M NaCl/10 mM Tris-HCl, pH 7.5/1 mM EDTA in an SW 40 rotor at 35,000 rpm for 4 hr) or digestion of the DNA with Sst I, for which there are two recognition sites in the B fragment but none in the vector arms.

In each packaging experiment, 3.5 \mu g of vector DNA was ligated to 400 ng of donor DNA in a total volume of 5.5 \mu l. Packaging and subsequent cloning were performed in a P3 physical containment facility in accordance with the National Institutes of Health Guidelines for Recombinant DNA Research. All packaging components were tested for packaging of endogenous DNA as well as for am^+ phage in control \lambdagt WES-AB DNA packaging experiments. In all cases the ratios were <10^{-6}. An identified \hGH clone in \lambdagt WES was subcloned into the EcoRI site of phosphatase-treated pBR322 by using χ 1776 as host and colony hybridization (12) with \alpha^{32}P-labeled \hGH cDNA as probe to detect recombinants.

**Plaque Screening.** Filters were made as described by Benton and Davis (13) except that the filters were pre-wetted with water and then 6-fold concentrated NaCl/Cit before application to the petri dish for 4 min; denaturation was for 4 min in 0.2 M NaOH/1.5 M NaCl; neutralization was for 4 min in 0.5 M Tris, pH 7.5/3 M NaCl, followed by a brief wash in 2-fold concentrated NaCl/Cit. The filters were hybridized to the \hGH cDNA probe as described above.

**DNA Sequence Analysis.** The chain termination method (14) was used with single-stranded template DNA generated by Escherichia coli exonuclease III (BioLabs) digestion of linear plasmid DNA (15). Thin sequencing gels (40 × 20 × 0.35 cm) were used (16).

**RESULTS**

**Identification of hGH and hCS Coding Sequences in Human DNA.** Restriction endonuclease fragments of high molecular weight human placental DNA containing sequences that hybridized to cloned \hGH cDNA were identified by filter hybridization (Fig. 1). The hybridization probe used was the cloned 550-base-pair \Hae III fragment of \hGH cDNA that extends from amino acid 24 of \hGH for 48 bases into the 3′ untranslated region of the mRNA (subsequently called \hGH 550) (unpublished data).

Two hybridizing EcoRI fragments, approximately 2.6 and 2.8 kb long, were identified. The 2.6-kb fragment hybridized more strongly than the 2.8-kb fragment, suggesting that, because the probe is \hGH 550, the 2.6-kb band contains \hGH sequences and the 2.8-kb band contains hCS sequences. This identification is supported by restriction endonuclease analysis of the 2.6- and 2.8-kb fragments (Fig. 2). The equivalent 550-base-pair \Hae III fragments in \hGH and hCS cDNAs are distinguished from each other by several restriction enzyme sites. hGH cDNA contained single sites for Bgl II, Pst I, and Sma I whereas hCS cDNA had none of these but did contain...
The hGH recombinant contained a 2.6-kb fragment that hybridized to the chGH 550 probe and had a restriction enzyme pattern for the enzymes *EcoRI*, *PstI*, *BglII*, *PvuII*, and *XbaI* (Fig. 3A) that was identical to that observed in total human DNA (Fig. 2). Similarly, the digestion pattern of the 2.8-kb hCS fragment (Fig. 3B) corresponded to that of total human DNA. This restriction enzyme data and the specific hybridization observed confirms the identities of the recombinants.

In the case of the hGH isolate, digestion of plaque-purified recombinant phage DNA with *EcoRI* revealed that three *EcoRI* fragments of human DNA had been cloned in the same phage, presumably as the result of multiple ligation of the human *EcoRI* DNA fragments. Two of these fragments were 2.6 kb and one was less than 2 kb. The 2.6-kb fragment that hybridized to the chGH 550 probe was transferred to the *EcoRI* site of the plasmid vector pBR322 by using *c* 1776 as the host strain. This recombinant plasmid has been used for most of the subsequent restriction analysis and DNA sequencing of the hGH gene and is designated pBR322-hgGH(2.6).

The hGH Gene Contains at Least Two Intervening Sequences. A restriction enzyme cleavage map of the hGH gene has been determined for the enzymes *EcoRI*, *BamHI*, *PstI*, *PvuII*, *SmaI*, and *BglII* (Fig. 4). Certain features of this map lead to the conclusion that the hGH coding sequence is interrupted by at least two intervening sequences.

The locations of the intervening sequences have been confirmed by DNA sequence analysis. The nucleotide sequence of the junction between intervening sequence B and the coding region at approximately nucleotide 1200 was obtained by the chain terminator method (14) using as primer a restriction fragment corresponding to amino acid residues 45–56 of the hGH protein. This was obtained by *PstI* and *HindIII* digestion of the chGH 550 recombinant. Template DNA was prepared from the pBR322-hgGH(2.6) recombinant by first digesting it with *HindIII* to produce a linear molecule and then converting it to single-stranded DNA with *E. coli* exonuclease III (15). An autoradiograph of a sequencing gel is shown in Fig. 5 and the sequence of the junction between the coding and intervening sequence is shown in Fig. 6.

In addition to locating the boundary between the coding sequence and intervening sequence B, this confirms the identity of the 2.6-kb fragment as hGH and not hCS because the sequence encompasses a region where the two hormones differ in amino acid sequence.

The sequence of the junction between the coding sequence and intervening sequence C at approximately nucleotide 1300

---

**Fig. 3.** Autoradiographs of filter hybridizations using the chGH 550 probe with the 2.6-kb hGH fragment (A) and the 2.8-kb hCS fragment (B). The digests were fractionated on two separate 1% agarose gels. Lanes 1–5 were obtained with the hGH-λgt WES recombinant digested with: lane 1, *EcoRI*; 2, *EcoRI* and *XbaI*; 3, *EcoRI* and *BglII*; 4, *EcoRI* and *PstI*; 5, *EcoRI* and *PvuII*. Lanes 6–10 were obtained with the hCS-λgt WES recombinant digested with: lane 6, *EcoRI*; 7, *EcoRI* and *PstI*; 8, *EcoRI* and *PvuII*; 9, *EcoRI* and *BglII*, 10, *EcoRI* and *XbaI*.

---

**Fig. 4.** Restriction enzyme map of the 2.6-kb *EcoRI* fragment containing the hGH gene sequences. Those sites written above the line are present in the cDNA; those below the line are unique to the gene fragment. The region containing sequences homologous to the 550-base-pair cDNA fragment used to probe the gene is indicated. This extends from amino acid 24 to 48 bases into 3' untranslated region. The 3' end of the coding region is marked. The poly(A) addition site is located 108 bases from the 3' end of the coding region. Coding sequences are indicated by shading. The intervening sequences are labeled A, B, and C. The precise boundary at the 3' end of region A is not clear (see text).
An 800-base-pair “full-length” cDNA clone of hGH (chGH 800) has been obtained recently (20). This contains, in addition to the entire coding sequence, 29 nucleotides of the 5′ untranslated region and all 109 nucleotides of the 3′ untranslated region preceding the poly(A) sequence. When this 800-base-pair fragment is used as a hybridization probe with pBR322-gbGH(2.6), the PstI/Peu II fragment between positions 600 and 1000 hybridizes strongly. This fragment does not hybridize to the chGH 550 probe, thus indicating that a large portion of the 5′ end of the hGH gene is contained in this region (Fig. 4). A very low level of hybridization of the chGH 800 probe to the 5′-terminal EcoRI/Pst I fragment has also been detected, indicating the possibility of a third intervening sequence (A) in the hGH gene. This will require further DNA sequence analysis for confirmation.

DNA sequence analysis has also established that the 2.6-kb hGH fragment contains the poly(A) addition site (Fig. 4).

**DISCUSSION**

EcoRI restriction endonuclease fragments 2.6 and 2.8 kb long containing coding sequences for the hGH gene and the hCS gene, respectively, have been identified by hybridization to a 550-base-pair cloned Hae III fragment of hGH cDNA. In addition to these two fragments, a third fragment of about 9.5 kb, that hybridizes weakly to the chGH 550 probe, is also detected routinely (see Figs. 1 and 2). The origin of this is not clear because the other known sequence related to hGH and hCS is the pituitary hormone prolactin which only has 16% amino acid sequence homology with hGH and 13% homology with hCS (21). The hybridization conditions used would not detect prolactin genes (assuming that the nucleotide sequences had similar levels of homology).

Both the 2.6-kb hGH and the 2.8-kb hCS fragments have been isolated by cloning in the bacteriophage Agt WES-LB vector with the in vitro packaging approach. A restriction enzyme cleavage map (Fig. 4) derived for the hGH gene indicates that, in common with other eukaryotic genes (22–26), the hGH gene contains intervening sequences that interrupt the coding sequences.

One of the two identified intervening sequences (sequence B) is about 220 base pairs long; the other (sequence C) is about 300 base pairs. A third intervening sequence (sequence A) very close to the 5′ end of the gene may also exist. The possibility that other very small intervening sequences are present—i.e., that some of the intervening sequences are interspersed within short coding regions—has not yet been excluded.

Preliminary DNA sequence analysis of the pBR322-gbGH(2.6) recombinant has confirmed the existence of the intervening regions shown in Fig. 4. The sequence at the 5′ end of intervening sequence B (at approximately position 1200) is shown in Figs. 5 and 6. As in other systems, the 3′ end is marked by an A-G doublet (27).

A detailed restriction enzyme map has not yet been determined for the hCS gene but the preliminary evidence suggests that the pattern of intervening sequences may be similar to that of the hGH gene. Total human placental DNA, when digested with both EcoRI and Pvu II (Fig. 2), contains a single 980-
base-pair fragment that hybridizes to the chGH 550 probe. This probe also detects hCS sequences. Because this fragment contains intervening and coding sequences in the hGH gene, the appearance of a single band implies that the hCS gene may also contain a similar structure in this region.

Several nucleotide sequence differences have been detected between the ghGH(2.6) gene and the cloned cDNA fragment. One of these is shown in Fig. 6: a CCA proline codon at amino acid position 37 in the cDNA is changed to a CTG leucine codon in the gene. The cDNA proline codon corresponds to the amino acid sequence established for the hGH protein (28). This codon change is in a region where tryptic mapping data of the hGH protein indicates that amino acid sequence variants occur (29). The ghGH(2.6) fragment may therefore correspond to an allelic variant of growth hormone. Alternatively, the differences may be the result of the different sources of material used: the hGH cDNA came from a pool of pituitary tumors and the hGH gene came from a single placenta.

The amino acid data in Fig. 6 exclude the possibility that a growth hormone variant of molecular weight 20,000 (29), rather than the normal 22,000, has been cloned. This variant, which exists as 5–10% of the total growth hormone, lacks the tryptic peptide located between amino acids 38 and 41. The DNA sequence analysis presented in Fig. 6 shows that the coding region corresponding to this peptide is present in ghGH(2.6). Nucleotide sequence differences between the cDNA and ghGH(2.6) have also been detected in the 3′ untranslated region (unpublished data).

An independently isolated 2.6-kb fragment has been recently cloned (unpublished data). This fragment gives rise to an identical hybridization pattern as the first isolate when the enzymes Pvu II, Pst I, Bgl II, and Sma I are used. However, a difference between the two is noticed with the enzyme Hae III. The possibility therefore exists that this second isolate corresponds to the expressed growth hormone gene. DNA sequence analysis of this isolate will be required in order to determine whether this is the case.

Note Added in Proof. The existence of the third intervening sequence (A) has been confirmed by DNA sequence analysis. The splice point is located in the second codon of the pre-sequence of hGH.

J.C.F. is the recipient of a postdoctoral fellowship from the Anna Fuller Fund. R.A.H. is the recipient of a postdoctoral fellowship from the British Science Research Council. This work was supported by Grants CA14026 to H.M.G. and AM19997 to H.M.G. and J.D.B.; H.M.G. and J.D.B. are Investigators of the Howard Hughes Medical Institute.