The EMBO Journal vol.6 no.4 pp.971–981, 1987

Tissue-specific expression of the human growth hormone gene is conferred in part by the binding of a specific *trans*-acting factor

Christophe Lefevre, Masayoshi Imagawa, Sharon Dana, Joan Grindlay, Mordechai Bodner and Michael Karin

Division of Pharmacology, M-036 H, School of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

Communicated by R.Cortese

The molecular basis for the pituitary-specific expression of the human growth hormone (hGH) gene was investigated, by gene transfer and protein footprinting experiments. Plasmid constructs in which CAT or Neo transcription units are fused to a 0.5 kb fragment of the hGH 5′ sequences were efficiently expressed in GC and GH_3 cells, derived from a pituitary tumor, but not in cell lines of other origins, indicating the presence of a tissue-specific promoter. DNaseI footprinting experiments have identified at least three factors that specifically bind to the hGH 5′ region. While two of these factors were also detected in extracts of non-expressing cells, the third factor, GHF-1, was detected only in extracts of GH expressing pituitary tumor cells. Mutagenesis experiments suggest that binding of GHF-1 and some of the other more ubiquitous factors is required for optimal hGH promoter activity *in vivo*. Tissue specificity of the hGH promoter therefore seems to be determined by the binding of at least one tissue-specific *trans*-acting factor, acting in concert with several other more ubiquitous, yet specific, DNA binding proteins.

Key words: gene-expression/growth-hormone/promoters/transcription factor

Introduction

Many of the genes expressed by multicellular organisms are transcribed in one cell type but not in others. Recent studies have shown that tissue-specific expression of differentiated cell products is controlled by the two general classes of *cis*-acting genetic elements: promoters (McKnight and Kingsbury, 1982) and enhancers (Khoury and Gruss, 1983) some of which exhibit cell-type specificity. While tissue-specific promoters (Walker *et al.*, 1983; Charnay *et al.*, 1984; Wright *et al.*, 1984; Grosschedl and Baltimore, 1985; Mason *et al.*, 1985; Ott *et al.*, 1984; Ciliberto *et al.*, 1985) and enhancers (Gillis *et al.*, 1983; Banerji *et al.*, 1985; Queen and Baltimore, 1983; Edlund *et al.*, 1985) have been described, the biochemical basis for their cell-type specificity is not clearly understood. To further characterize the mechanisms involved in the establishment of tissue-specific gene expression, we chose to study the human growth hormone gene family.

This gene family contains genes coding for three related hormones, growth hormone (GH), chorionic somatomammotropin (CS), and prolactin (Prl), which have evolved through duplication of a common ancestor, giving rise to Prl and GH. More recent duplication events unique to primates gave rise to the normal hGH gene, hGH-N, the major hCS gene, hCS-A, and three other related genes, hCS-B, hCS-L, and hGH-V (Barsh *et al.*, 1983). Due to their recent divergence, hGH and hCS exhibit a very high degree of sequence conservation not only in their coding region but also in their 5′ flanking regions, introns, and the immediate 3′ flanking regions (see Miller and Eberherdt, 1983; Moore *et al.*, 1982; Seeburg, 1982 for reviews). Despite their considerable degree of sequence homology, these genes are expressed in a distinct tissue-specific manner. hGH and hPrl are produced in the anterior pituitary, whereas the hCS genes are expressed in the placental syncytiotrophoblast. The difference in the levels of hGH-N and hCS-A mRNAs in these cell types is at least 10 000× (P.Seeburg, personal communication), which is quite striking, considering the sequence similarity of the two genes (Seeburg, 1982; Selby *et al.*, 1984). On the other hand, the hPrl and hGH-N genes are much less similar in their nucleotide sequences (Truong *et al.*, 1984), but are expressed in related cell types derived from a common precursor. For these reasons the human growth hormone gene family is an interesting system not only for investigating the molecular mechanisms which control tissue-specific gene expression but also for studying the co-evolution of *cis* and *trans*-acting genetic elements.

As a first step towards this end, we located *cis*-acting genetic elements responsible for tissue-specific expression of the hGH gene, and characterized their recognition by cellular *trans*-acting regulatory proteins. Here we demonstrate that tissue-specific expression of hGH is conferred in part by the combinatorial recognition of its promoter by several *trans*-acting factors, one of which seems to be present only in cells that are permissive for expression of this gene.

Results

The 5′ flanking region of the hGH-N gene contains a tissue-specific promoter element

Some of the *cis*-acting involved in the tissue-specific expression of the β globin (Charnay *et al.*, 1984; Wright *et al.*, 1984), immunoglobulin (Grosschedl and Baltimore, 1985; Mason *et al.*, 1985), insulin (Edlund *et al.*, 1985), chymotrypsin (Walker *et al.*, 1983), albumin (Ott *et al.*, 1984) and α_1-antitrypsin (Ciliberto *et al.*, 1985) genes are located within a few hundred bp 5′ to their transcriptional start sites. Therefore, we examined whether a similar region of the hGH-N gene contains a tissue-specific promoter element. A 0.5 kb fragment containing the hGH-N 5′ region (DeNoto *et al.*, 1981) was fused to the bacterial gene coding for chloramphenicol acetyltransferase (CAT) whose expression can be easily monitored in transfected cells (Gorman *et al.*, 1982) to yield the vector phGH−CAT. Transfection experiments showed that phGH−CAT was not expressed to a significant extent in a number of cell lines of fibroblastic and epithelial origins, including Rat1, Ltk-, HeLa, JEG-4, or HepG2, which do not express their endogenous GH genes. In contrast, efficient expression was obtained in GC and GH_3 cells, both of which were derived from a pituitary GH-expressing tumor (Tashjian *et al.*, 1968). On the other hand, pSV2CAT and pXCAT3-, in which the CAT gene is controlled by the SV40 early promoter and the enhancer elements of either SV40 (Gorman *et al.*, 1982) or the human metallothionein II_A ($hMT\text{-}II_A$) gene (Haslinger and

Table I. Summary of transient-expression experiments

Plasmid	Cell line						
	HeLa-B	JEG-4	HepG2	Rat1	LTK-	GC	GH_3[a]
pSV2CAT (CAT u.)	1500	8500	6940	2500	3500	245 (± 60)	25
pSV2CAT (%)	100	100	100	100	100	100 (± 25)	–
pXCAT3- (%)	53	–	43	–	–	160 (± 30)	38
phGH−CAT (%)	<1	<1	7	2	3	180 (± 45)	660
pAlb−CAT (%)	–	–	3	–	<1	<1	–
pUC−CAT- (%)	<1	<1	<1	<1	<1	<1	<1

Ten μg's of each plasmid were transfected into the different cell lines. 48 h later the levels of CAT activity were determined. The results shown are averages of two separate experiments (each done in duplicate) with the exception of the results of the GC transfection experiments, which reflect the averages (± standard deviation) of six different experiments. The levels of CAT activity were converted to percent of pSV2CAT expression.
[a]Due to the low level of pSV2CAT activity in GH_3 cells, the conversion described above was not made, and the data reflects the averages of two separate experiments expressed in pmoles of [^{14}C]chloramphenicol converted/mg protein/h (CAT units).

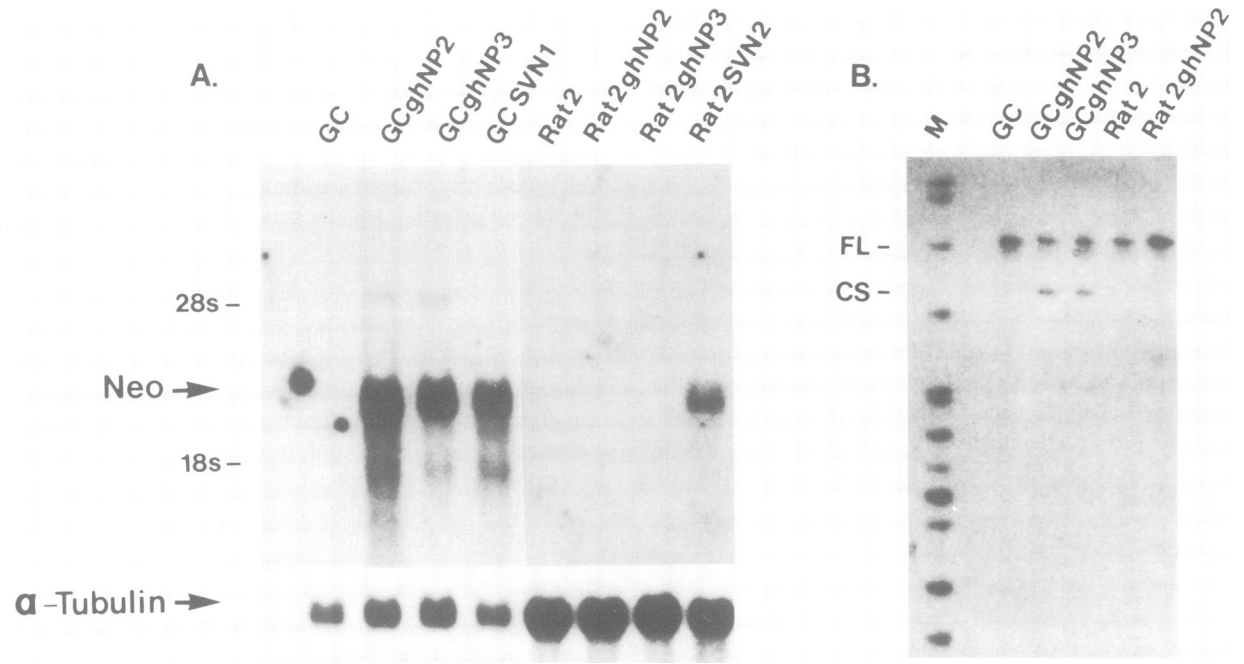


Fig. 1. Tissue-specific expression of the hGH−Neo gene. **(A)** Northern blot analysis of 5 μg polyA$^+$ RNA extracted from untransfected GC or Rat2 cells or cells transfected with hGH−Neo (GCghNP2, GCghNP3, Rat2ghNP2, and Rat2ghNP3) or pSV2Neo (GCSVN1 and Rat2SVN2). The hybridization probes were complimentary either to the Neo or α-tubulin mRNAs. The positions of migration of the ribosomal RNAs are indicated. **(B)** S1 nuclease analysis of hGH−Neo transcription, 5 μg of polyA$^+$ RNA extracted from the various pooled cultures, as indicated above each lane, were hybridized to an end-labelled single stranded Nsi (−84) to *Bgl*II (+320) probe labelled at position +320 relative to the expected start site of phGHNeo transcription. FL - full-length probe. CS - correct start. Markers are *Hpa*II fragments of pBR322. RNA extraction, blot hybridization and S1 nuclease mapping were done as previously described (Heguy *et al.*, 1986).

Karin, 1985) were efficiently expressed in these cell lines with the exception of GH_3. The vectors pAlb−CAT, in which the rat albumin promoter is fused to the CAT gene (Ott *et al.*, 1984) and pUC−CAT-, which contains a promotorless CAT gene, were not expressed in any of these cell lines (Table I). Due to substantial differences in transfection efficiencies among these lines, we normalized the expression of the various vectors relative to that of pSV2CAT. As shown in Table I, the 5′ flanking region of hGH-N leads to approximately 100× over-expression of CAT activity in GC compared to HeLa cells, while the hMT-II_A enhancer is equally active in both lines. In addition, we have stably introduced the hGH−CAT gene, by co-transfection with pSV2Neo (Southern and Berg, 1982) into GC and Rat2 cells. In pools of stably transfected GC cells hGH−CAT was expressed more than 100× more efficiently than in pools of stably transfected Rat2 cells, even though its copy number was somewhat higher in the latter (S.Dana, unpublished results).

To extend the results obtained with the hGH−CAT vector, we constructed the vector phGH−Neo, a fusion of the hGH promoter to the bacterial gene coding for resistance to the aminoglycoside, G418. While G418-resistant colonies were obtained at a frequency of approximately 5×10^{-6} after transfection of

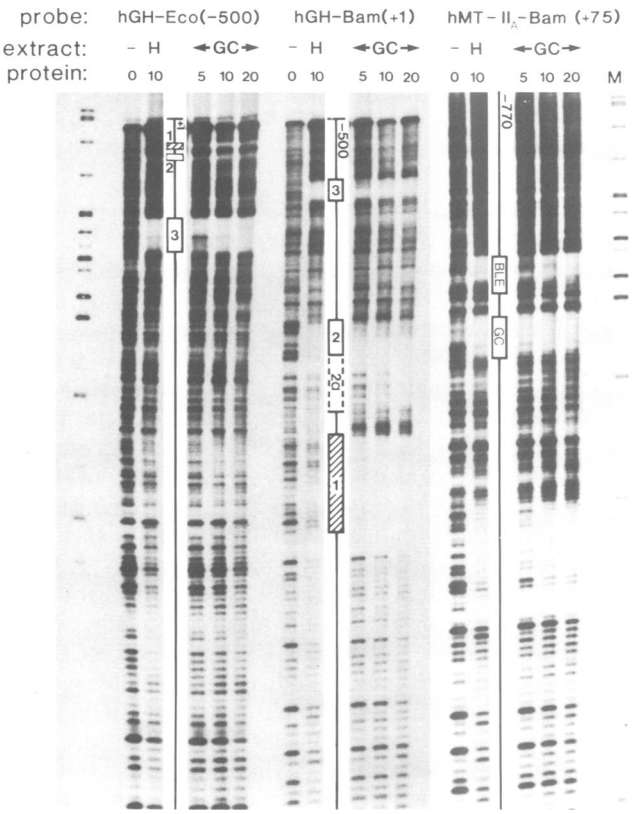


Fig. 2. GC cells contain a specific hGH binding factor not present in HeLa cells. The protection pattern of the 0.4 M heparin–agarose fraction of the GC WCE (GC) was compared to that of an equivalent fraction of a HeLa WCE (H). 5, 10 or 20 μg of the WCEs (as indicated above each lane) were incubated with ~2 ng of hGH–*Eco*RI probe (end-labelled at −500) or hGH–*Bam*HI (end-labelled at +3) and subjected to DNaseI footprinting. As a control, the protection pattern of the two WCEs with an hMT-II$_A$–*Bam*HI probe (end-labelled at +75), were compared, as the expression of this gene is not tissue-specific. Control reactions (0) were done incubated with 10 μg of BSA. The boxes on the side panel denote the various binding domains which were observed. The tissue-specific site 1 is indicated by the cross-hatched box.

GC cells with phGH–Neo, no such colonies could be derived by direct selection of transfected Rat2 cells. On the other hand, with pSV2Neo (Southern and Berg, 1982), the transformation frequency of Rat2 cells (nearly 10^{-4}) was 100× higher than that of GC cells ($<10^{-6}$). Therefore phGH–Neo was introduced into Rat2 cells by cotransfer with the HSV-TK gene. Pools of tk^+ Rat2 cells and $G418^r$ GC cells were grown up and analyzed for the presence and expression of the hGH–Neo gene. While all analyzed pools contained the fusion gene (data not shown), only the transfected GC cells expressed correctly initiated Neo mRNA (Figure 1A,B). In agreement with the essentially undetectable level of Neo mRNA in the Rat2 pools, the majority of these cells failed to grow even in a relatively low concentration of G418 (200 μg/ml). These results, taken together with the results of the transient and stable expression experiments with the CAT vectors support the notion that the first 500 bp of the hGH 5′ flanking region contain a pituitary-specific promoter element.

The hGH 5′ flanking region is recognized by factors present in GC cells

The distinct tissue-specific expression of phGH–CAT and phGH–Neo suggested that the hGH promoter is recognized by positive *trans*-acting factors present in GC cells. As a direct test of this hypothesis, we have prepared whole-cell extracts (WCE) from GC cells and subjected them to partial purification on a heparin-agarose column (Davison *et al.*, 1979). A fraction which was step-eluted between 0.2 M to 0.4 M KCl (0.4 M fraction) was examined for presence of specific DNA binding proteins using the DNaseI footprinting assay (Galas and Schmitz, 1978). Previously, we found that this fraction contains all of the factors that bind to the hMT-II$_A$ gene (M.Imagawa, unpublished results). As shown in Figure 2 the GC WCE contained factors which specifically bind to at least three sites within the hGH 5′ control region. A strong hypersensitive site can be observed between sites 1 and 2. An identical pattern of protection of the hGH 5′ region was also obtained with unfractionated GC extract (see Figure 7).

To test whether the tissue specificity of the hGH promoter is due to differential distribution of *trans*-acting factors between expressing and non-expressing cells, we compared the protection pattern observed with the GC WCE to that conferred by an extract from HeLa cells, in which the hGH promoter is not active. The HeLa extract contained factors which protected sites 2 and 3, but not site 1 (Figure 2). In addition, the footprint on site 2 conferred by the HeLa extract was not as extensive as that conferred by the GC extract. In subsequent experiments (see Figures 3A and 4) it became apparent that the GC extract gave protection of an additional site not observed with HeLa extracts. The additional sequence protected by the GC WCE was named site 2a. Furthermore, the strong hypersensitive site present between sites 1 and 2 was specific to the GC extract. As a control we examined the binding of the two WCE's to the control region of a gene whose promoter is active in both cell types: hMT-II$_A$ (see Table I). In contrast to the marked differences between the two extracts observed with the hGH probes, the protection patterns of the hMT-II$_A$ 5′ probe were essentially identical (Figure 2, hMT-II$_A$-Bam). As observed earlier (Lee *et al.*, 1987), both extracts contain factors that bind to the regions important for the basal expression of the hMT-II$_A$ gene: the GC box and the basal level enhancer element (BLE) (Karin *et al.*, 1987). As an additional control we have footprinted the HSV-TK gene and found that both extracts gave results identical to those of Jones *et al.* (1985), suggesting that HeLa and GC cells have similar levels of SpI, AP1 and CTF (see Figure 6).

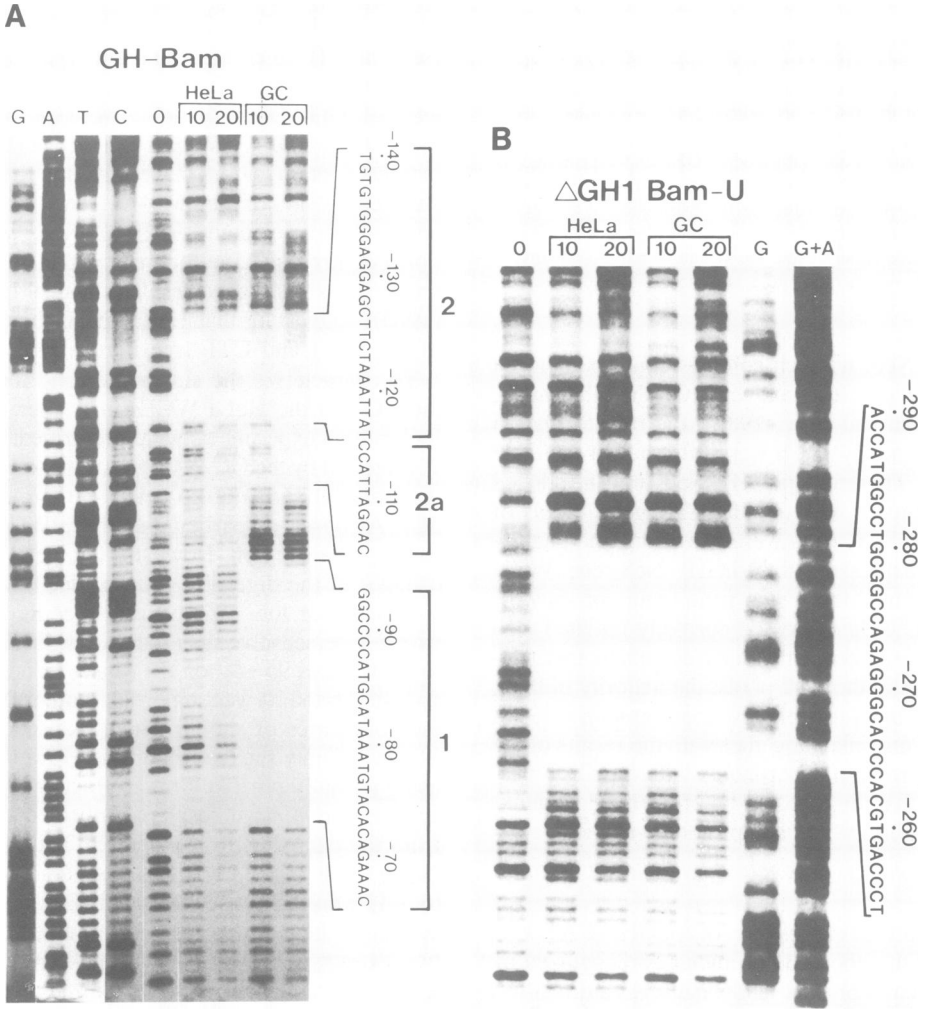
To characterize the similarities and the differences between the GC and the HeLa WCEs in further detail, high-resolution footprinting experiments were performed. The boundaries of site 1, protected by the GC WCE, are −66 to −93 on the sense strand (Figure 3A) and −66 to −97 on the antisense strand (Figure 3C). As observed earlier, the HeLa WCE did not lead to protection of this region. However, the HeLa WCE produced subtle changes in the digestion pattern, due either to the presence of a nuclease or low affinity, non-specific binding proteins. The strong hypersensitive site produced by the GC WCE was localized to nucleotides 95–97 on the sense strand, and 100–102 on the opposite strand. It was somewhat more difficult to map accurately the boundaries of site 2a which is also GC specific because it is a low-affinity site (see Figure 4). However, further experiments using partially purified protein fractions (M.Bodner and M.Imagawa, unpublished results) confirmed the conclusions drawn from the data shown in Figure 3. Thus, the boundaries of site 2a are −106 to −115 on the sense strand (Figure 3A) and −107 to −120 on the antisense strand (Figure 3C). While the boundaries of site 2 on the sense strand seem to be similar for the two extracts, i.e., from −116 to −140 (Figure 3A), the patterns of protection are somewhat different on the antisense strand.

The GC WCE provides strong protection between −121 to −134, but protection by the HeLa WCE extends all the way to −142 (Figure 3C) a region which is partially protected by the GC WCE. These differences could be due to a lower level of the factor which binds to site 2 in GC cells. The boundaries of site 3 were identical for both WCEs, which protected the region between −254 and −290 on the sense strand (Figure 3B) and approximately from −256 to −294 on the antisense strand (Figure 3D, data not shown for HeLa extracts). The distinct binding patterns of HeLa and GC WCEs were obtained with at least three different preparations. In addition, WCE prepared from a human hepatoma cell line and nuclear extracts of rat liver and a human osteosarcoma gave essentially the same results as the HeLa WCE, while a GH_3 nuclear extract was similar to the GC WCE (data not shown).

To determine whether one or two different factors are responsible for the GC specific recognition of sites 1 and 2a competition experiments were performed using an oligonucleotide whose sequence matches that of site 1. As shown in Figure 4, this oligonucleotide inhibited the protection of both of these sites but had no effect on protection of sites 2 and 3 (not shown). Furthermore, it competed more efficiently with site 2a than with site 1. These results, suggest that a single pituitary cell-specific factor binds to sites 1 and 2a on the hGH gene, both of which share the common core sequence 5′-TAAAT-3′. Further support for this conclusion is provided by fractionation experiments. Partial purification of a factor that binds to site 1, by gel-filtration, heparin agarose and DNA-sequence affinity chromatography leads to co-purification of site 2a binding activity as well (M.Bodner and M.Imagawa, unpublished results). We named this factor GHF-1 for growth hormone factor 1. The sequences of the various binding sites and their arrangement are shown in Figure 5.

GHF-1 is not a CAAT box binding factor

Interestingly, site 1 contains the sequence 5′-CATAAAT-3′, which resembles the highly-conserved sequence 5′-CCAAT-3′, known as the CAAT box, found in the −80 region of a large number of eukaryotic mRNA coding genes (Benoist *et al.*, 1980; Efstratiadis *et al.*, 1980). Because the 5′-CATAAAT-3′ sequence occupies a similar position (−84 to −79) on the hGH and hCS genes, it was proposed to be equivalent to the CAAT box (Miller and Eberhardt, 1983). Recently, two factors were identified that bind to the CAAT boxes of the HSV-TK and MSV-LTR promoters. These factors were isolated either from HeLa cells and named CAT transcription factor (CTF) by Jones *et al.* (1985), or from rat liver and named CAT binding protein (CBP) by Graves *et al.* (1986). It is not clear at this point whether the two factors are identical or not. For the sake of simplicity we will refer to them as a single factor, CTF/CBP. Since CTF/CBP was found in both HeLa cells and in rat liver, and its binding is important for expression of the HSV-TK gene (Jones *et al.*, 1985; Graves *et al.*, 1986), whose promoter is not tissue-specific, it seemed rather unlikely that GHF-1 is the same factor. However,



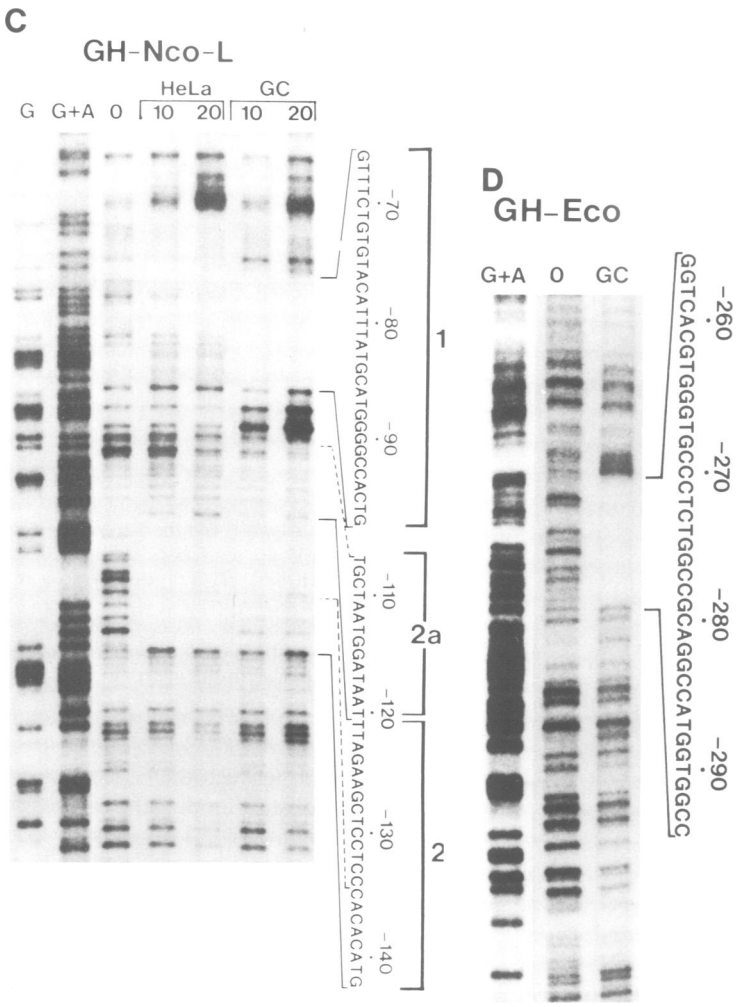


Fig. 3. High-resolution footprint analysis of the hGH 5′ probes incubated with GC and HeLa extracts. DNaseI footprinting was performed with either 10 or 20 μg of the 0.4 M heparin−agarose fractions of the two extracts as described in Materials and methods. **(A)** To determine the boundaries of regions 1 and 2 on the sense strand, the hGH 5′ region was labelled at the *Bam*HI site by filling in, and the DNaseI cleavage products were analyzed side by side with the chemical sequencing products of the same probe. 0- indicates control reaction incubated with 10 μg of BSA prior to DNaseI digestion. **(B)** To analyze the exact boundaries of region 3, a probe was prepared from an internal deletion mutant ph5′Δ1 in which, due to an internal deletion, the *Bam*HI site is closer to region 3. Binding to the sense strand was examined, by labelling the *Bam*HI−*Eco*RI probe, at the *Bam*HI site by filling in. **(C)** To map the boundaries of regions 1 and 2 on the antisense strand, the hGH 5′ region was labelled at the *Nco*I site at −285 and digested with *Bam*HI. The *Nco*I−*Bam*HI (+3) fragment was used as a probe. **(D)** To map the boundaries of region 3 on the antisense strand, the *Eco*RI site at −500 was labelled by filling in and the *Eco*RI−*Bam*HI (+3) fragment was isolated to use as a probe. Only 10 μg of GC extract were used in this experiment.

to rule out the possibility that the CTF/CBP activity was preferentially lost during the preparation of the HeLa WCE and to provide further evidence that GHF-1 is a distinct factor, we compared the ability of the two WCEs to protect specific regions of the HSV-TK gene. As shown in Figure 6A, the footprints obtained with the two extracts on the HSV-TK probe were essentially identical. Strong protection is observed over the distal GC box, while the CAAT box is moderately protected. Very little protection is observed over the proximal GC box. Similar differences in binding efficiencies between the different sites were observed by Jones *et al.* (1985). These results indicate that the WCEs prepared from the two different cell lines contain similar levels of SP1 (which binds to the GC box) and CTF/CBP.

As a further proof that GHF-1 is not CTF/CBP, we have performed competition-footprinting experiments between the labelled hGH 5′ probe and restriction fragments derived from the normal hGH 5′ region (hGH 5′; see below), a linker-scanning mutant which destroys site 1 (hGHΔNsi1), and HSV-TK deletion mutant which contains the CAAT box (TK-109), and an HSV-TK deletion mutant which lacks the CAAT box (TK-46). Significant competition for GHF-1 binding was observed only when the wild-type hGH 5′ region was used as a competitor, and only marginal competition was observed with the other DNA fragments (Figure 6B). Therefore it is unlikely that the GC specific factor GHF-1 is CTF/CBP. Further support to the non-identity of the two factors can be derived from the mutational analysis of the HSV-TK CAAT box, performed by Graves *et al.* (1986), who found that most of the point mutations introduced into that sequence led to a dramatic decrease in factor binding. The hGH CAAT box homolog differs from the canonical CAAT box at three positions, a difference which probaly abolishes CTF/CBP binding altogether.

Binding of GHF-1 and some of the other factors is essential for hGH promoter activity in vivo

To test whether the binding of GHF-1 and the other factors is required for optimal hGH promoter activity *in vivo*, we constructed several mutant derivatives of the hGH 5′-flanking region. The first mutant hGHΔNsi1 is a linker-scanning mutant, in which an 8 bp *Bgl*II linker (5′-CAGATCTG-3′) replaces positions −87

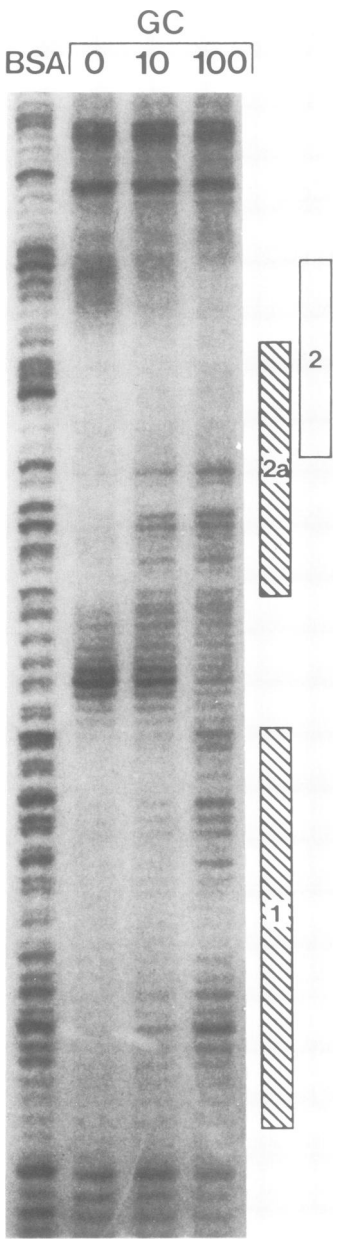


Fig. 4. Sites 1 and 2a are recognized by a common factor. To determine whether the two tissue-specific binding sites, 1 and 2a, are recognized by a common factor, competition experiments were performed. Binding reactions contained 1 ng of hGH 5′ probe and either 0, 10 or 100 fold excess of a synthetic site 1 formed by annealing of the oligonucleotides 5′-GATCCC-ATGCATAAATGTACACAG-3′ and 5′-AATTCTGTGTACATTTATGCA-TGG-3′. The different binding sites are demarcated by the numbered boxes on the side panel.

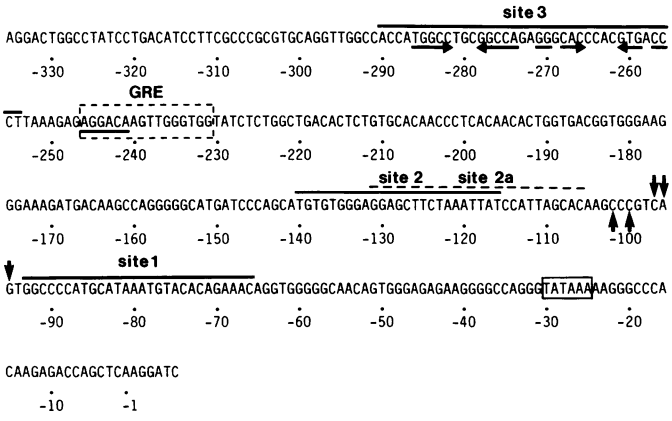


Fig. 5. Structure of the hGH promoter region. The sequence of the sense strand is shown together with the locations of the different binding sites, the putative GRE and the TATAA box. The broken line under site 2 demarcates a sequence which is conserved among the hGH, hCS, rGH, bGH, hPrl and rPrl genes (Truong *et al.*, 1984). The vertical arrows indicate the locations of the strong DNase I hypersensitive site. Arrows above the line indicate cleavage positions on the sense strand and arrows below the line indicate cleavage positions on the antisense strand. The horizontal arrows under site 3 denote two dyad symmetries present at that region.

to −82 which are part of site 1. As shown in Figure 7A, this mutation abolishes recognition of site 1 by GHF-1 while binding to the other sites was not affected. To mutate site 2a mutant hGH-LS(−121/−126) was generated, in which the sequence 5′-TTCTAA-3′ (positions −126 to −121) is replaced by the sequence 5′-AGATCT-3′. As shown in Figure 7B, this mutation abolishes binding to site 2a and reduces binding to site 2. However, neither of these mutations seems to have a significant effect on GHF-1 binding to the remaining site. On the other hand, it appears that inactivation of either site 1 or 2a inhibits the formation of the strong DNaseI hypersensitive site normally present between them (Figures 7A and B). The 2a mutation also abolishes the hypersensitive site present 5′ to this binding site (Figure 7B). These results also indicate that in addition to the common core 5′-TAAAT-3′ other residues are also important for recognition by GHF-1.

In an attempt to mutate site 3, a *Bgl*II linker was inserted at position −284 (hGH−NK) and between positions −290 and −283 (hGH−ΔNco1). These mutants, which affected sequences at the 5′ border of site 3, had only a marginal effect on the protection pattern at the edge of the site, but no significant effect elsewhere, indicating that site 3 was still intact (data not shown). Therefore a larger internal deletion was generated, hGH−ΔNco2, which replaced positions −386 to −206 with a *Bgl*II linker, to delete site 3 altogether.

These mutants and two additional ones, a 5′ deletion, hGHΔ5′-289, and an internal deletion, hGHΔ(−82/−128), which removes sites 1 and 2a, were fused to the CAT gene and their expression was determined after transfection into GC cells. Both transiently and stably transfected cells were examined. All of the mutants that inactivated or deleted binding sites 1, 2a or 3 exhibited CAT levels which were significantly lower than those of the WT construct (Table II). The expression level of two of the three mutants, which did not affect factor binding, was similar to that of the WT gene, while the 5′ deletion to position −289 caused a small decrease. Similar results were obtained by transient and stable transfection experiments. As shown in Figure 8, the decrease in CAT expression is due to lower levels of correctly initiated hGH−CAT mRNA. In summary, these results strongly suggest that the binding of *trans*-acting factors to sites 1 and 2a observed *in vitro* is also important for expression of the hGH promoter *in vivo*. The role of site 3 and its binding factor, however, are not completely clear, because the internal deletion that eliminates that site also includes sequences between −250 to −210. These sequences could function as recognition sites for other factors such as the glucocorticoid and thyroid hormone receptors, which are also involved in growth hormone expression.

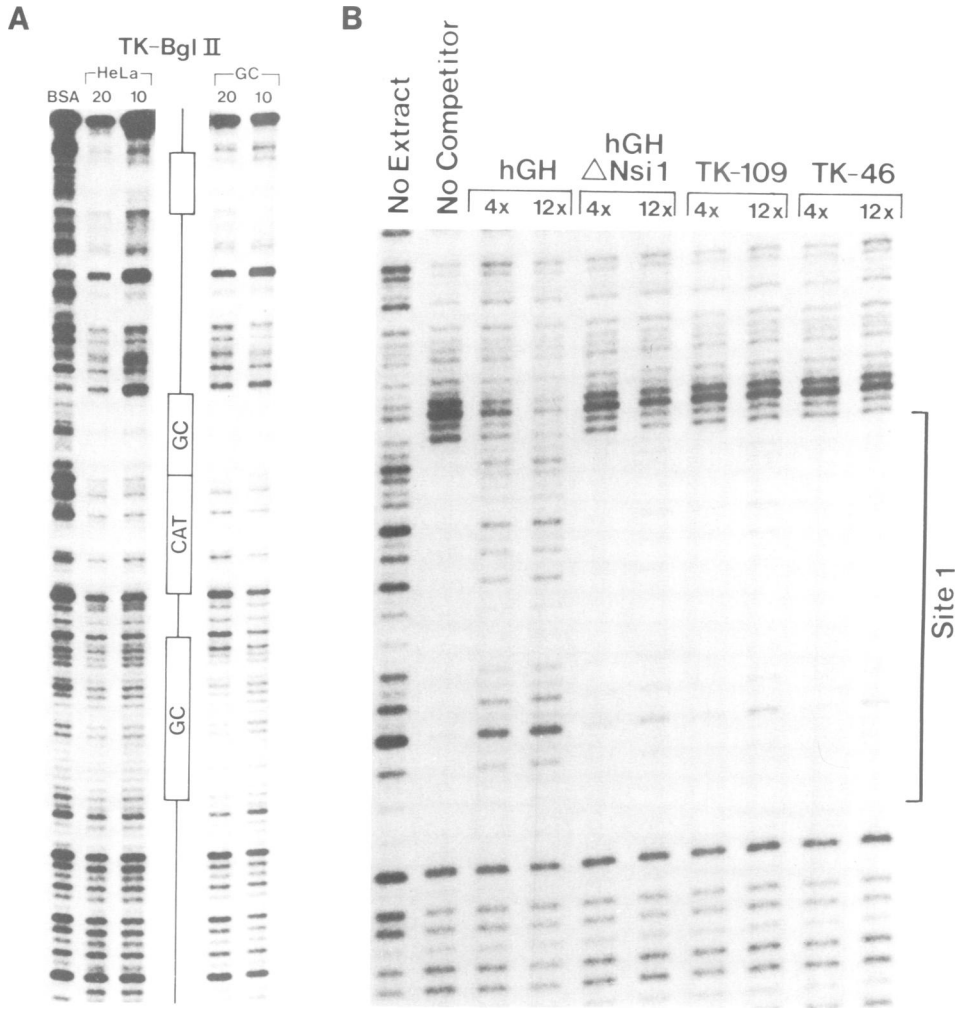


Fig. 6. GHF-1 is different from CAAT box binding factors. **(A)** Footprinting of HeLa and GC extracts on the HSV-TK gene. HSV-TKΔ5′-109 subcloned in pUC8, was labelled at the *Bgl*II site and digested with *Pvu*II to release a 350 bp fragment that was used as a probe. 2 ng of the TK probe were incubated with either 10 μg of BSA, 10 or 20 μg of the heparin−agarose 0.4 M fractions of the HeLa or GC WCE's, and subjected to DNaseI footprinting. The different binding sites present within the HSV-TK promoter are symbolized by the boxes in the middle panel. The upper box is a strongly-protected region within the plasmid sequence flanking the TK gene. **(B)** Competition footprinting. 2 ng of the hGH−*Bam*HI probe were incubated with either 10 μg of BSA (0) or with 10 μg of the GC extract 0.4 M fraction in the absence or presence of 4X or 12X molar excess of the following competitors: a 500 bp *Bam*HI−*Eco*RI fragment derived from the wild-type hGH gene, a 500 bp *Bam*HI−*Eco*RI fragment derived from phGH5′ΔNsi1, in which binding site 1 is destroyed by insertion of *Bgl*II linker, a 350 bp *Bgl*II−*Pvu*II fragment containing the HSV−TKΔ5′-109 promoter, and a 300 bp *Bgl*II−*Pvu*II fragment containing the HSV-TKΔ5′-46 promoter.

Discussion

The role of GHF-1 in GH gene expression

As has been shown for several other mammalian genes, (Grosschedl and Baltimore, 1985; Mason *et al.*, 1985; Edlund *et al.*, 1985; Ott *et al.*, 1984; Ciliberto *et al.*, 1985; Walker *et al.*, 1983), the 5′ flanking region of the hGH gene contains elements responsible at least in part for its tissue-specific expression. A simple working hypothesis that can account for these results assumes that tissue-specific promoter elements are recognized by *trans*-acting factors that are either present or active only in cell types permissive for these promoters. The results presented in this paper lend a general support to this hypothesis. The hGH 5′ flanking region, which functions as a pituitary-specific promoter, is recognized by at least three distinct factors, one of which, GHF-1, can be detected only in GH-expressing pituitary tumor cell lines. Two other factors that bind to the hGH 5′ region seem to be present in all cell types examined thus far. This interpretation of the data is based on the assumption that identical footprints reflect the binding of identical factors; however, the exact identity of the factors present in the different cell lines will be clear only after their purification to homogeneity. The present results also do not indicate whether GHF-1 is completely absent from cells in which the hGH promoter is not active, is present in much reduced amounts, or is present in a form which does not bind to DNA. However, it is clear that only cell lines derived from the anterior pituitary contain detectable amounts of GHF-1, and this differential distribution can account for the cell type specificity of the hGH promoter.

GHF-1 binds to two sites upstream of the TATA box of the hGH gene, both of which serve as functional upstream promoter elements *in vivo*, required for optimal expression, even though binding of GHF-1 to the distal site is of much lower affinity than the proximal site. In contrast, in the HSV-TK promoter, the low affinity Sp1 binding site is the proximal site (Jones *et al.*, 1985). In their normal arrangement both sites are required for optimal

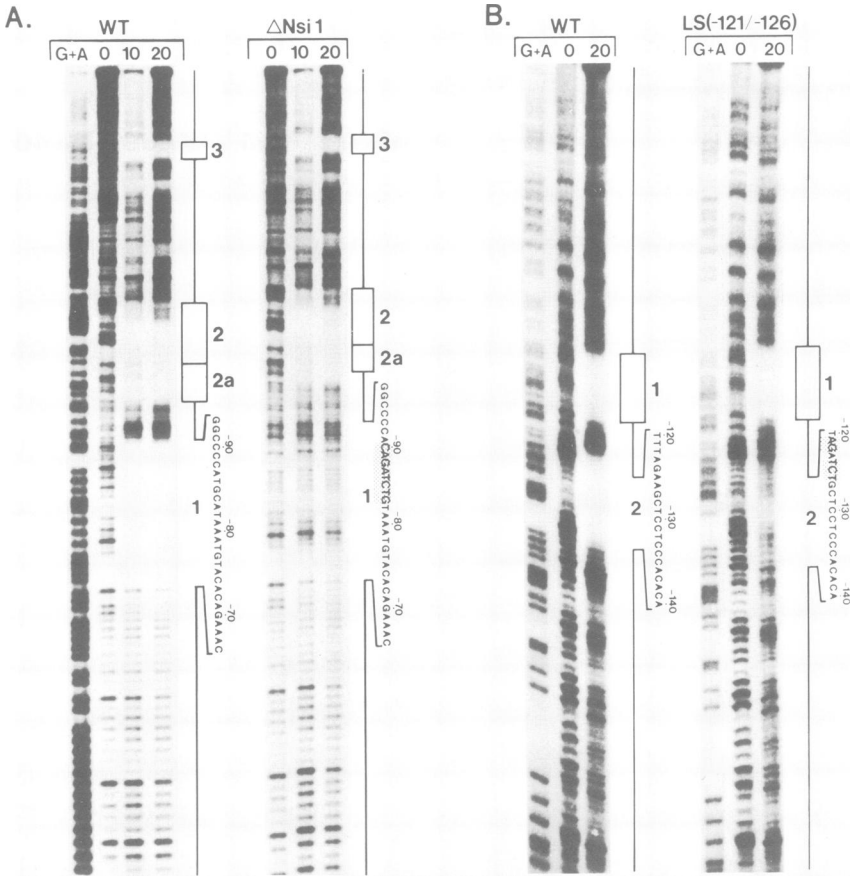


Fig. 7. Mutations of sites 1 and 2a abolish GHF-1 binding. The binding of factors present in GC WCE to the different mutants and the WT plasmids was determined by DNaseI footprinting, as described above. **(A)** Binding of factors to the site 1 mutant phGHΔNsi1. This plasmid and the WT plasmid, phGH5′, were labelled to approximately the same specific activity at their *Bam*HI sites (+3) and the corresponding *Bam*HI (+3) to *Eco*RI (−500) fragments were used on the footprinting experiments. The sequences shown are of the sense strands. **(B)** Binding of factors to the site 29 mutant phGH-LS(−121/−126). This and the WT plasmid phGH5′ were labelled at their *Nco*I sites (−285) and the *Nco*I (−285) to *Bam*HI (+3) fragments were used as footprinting probes. The sequence shown is of the antisense strand. The sequences of the WT and mutant sites are shown, with the mutated bases highlighted. In these experiments we used an unfractionated WCE.

Table II. Expression of hGH promoter mutants

Plasmid	Type of mutation	Sequence affected	Binding site affected	Expression level (% of WT)	
				transient	stable
hGH-CAT	W.T.	–	–	100 ± 16	100
hGHΔ5′-289−CAT	5′ deletion	−289 to −500	–	53 ± 7	77
NK−CAT	linker insertion	−284	–	145 ± 15	–
ΔNco1−CAT	linker scanner	−283 to −293	–	111 ± 20	–
ΔNco2−CAT	internal deletion	−206 to −386	3	15 ± 3	35
hGH−LS(−121/−126)−CAT	linker scanner	−121 to −126	2a	9 ± 2	12
ΔNsi1−CAT	linker scanner	−82 to −86	1	0	7
hGHΔ (−82/−128)−CAT	internal deletion	−82 to −128	1 + 2a	0	11

The results of the transient transfections represent the average relative levels of expression (W.T. = 100%) ± standard deviation, determined by four independent transfection experiments, using at least two different preparations of each plasmid. The results of the stable transfections are averages of two CAT assays done on pools of stable transfectants (co-transfection with phGHNeo) in which the copy number of the hGH−CAT fusion genes did not vary considerably. The binding of factors to these mutants was determined by DNaseI footprinting (see Figure 7).

expression. However, after the two sites are inverted, and the high affinity site is proximal, the distal low affinity site is no longer required (Jones *et al.*, 1985; McKnight *et al.*, 1984). The different behavior of the hGH promoter suggests that an interaction between two GHF-1 molecules bound to DNA could be required for optimal expression. This interaction may also be the cause for the strong hypersensitive site present between the two GHF-1 sites.

Interestingly the hGH-V gene which has a much weaker promoter compared to hGH-N (N.Eberhardt, personal communication) contains several nucleotide sequence substitutions within the proximal GHF-1 site (Seeburg, 1982). These sequence changes which resemble our *in vitro* generated mutations could therefore be responsible for the low activity of that promoter.

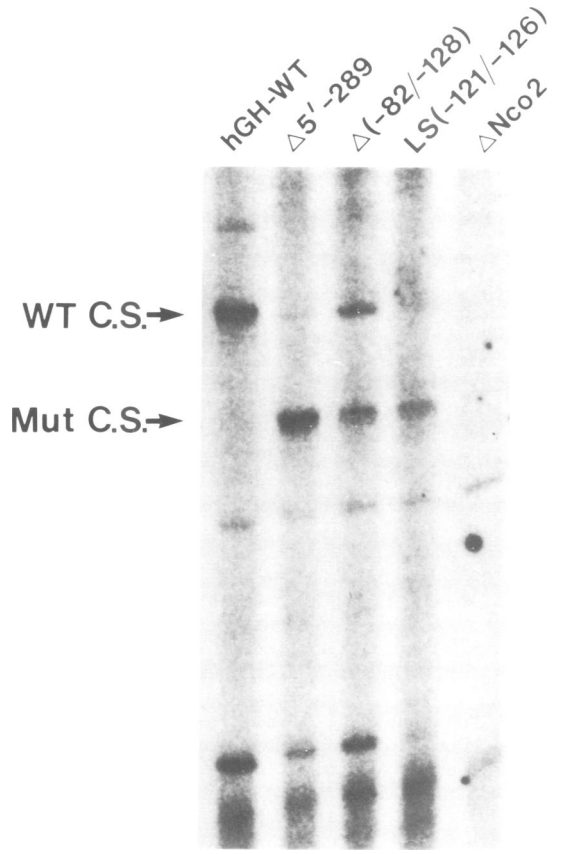


Fig. 8. Expression of hGH-CAT mutant genes in transfected GC cells. 5 μg polyA$^+$ RNA extracted from pools of GC cells, stably ransfected with the GH-CAT fusion genes indicated above (except for ΔNco2 for which only 1 μg poly A$^+$ was used), were subjected to primer extension analysis using a synthetic CAT specific primer (Haslinger and Karin, 1985). GH-WT CS- indicates the correctly initiated transcript from the WT hGH−CAT gene fusions. Mut GH CS- indicates the correctly initiated transcripts from the various mutants. Due to differences in their 5′ leader sequences the WT fusion generates a transcript which is 10 bases longer than these generated by the various mutants.

This observation and the good correlation between the reduced binding of GHF-1 to mutants of sites 1 and 2a and their decreased expression *in vivo* strongly suggest that GHF-1 is required for expression of the hGH-N gene in the anterior pituitary. Both sites 1 and 2a are almost perfectly conserved in the rat GH gene (Miller and Eberhardt, 1983), further supporting the importance of GHF-1. On the other hand, neither of these sites is well conserved in the hPrl (Truong *et al.*, 1984) and rPrl (Miller and Eberhardt, 1983) genes, suggesting that these genes which are expressed in a different cell type are activated by another factor.

The binding site for GHF-1 bear some resemblance to the more common promoter element known as the CAAT box (Benoist *et al.*, 1980; Efstratiadis *et al.*, 1980) and were originally suggested to be the hGH CAAT box equivalent (Miller and Eberhardt, 1983). However direct footprinting of the HSV-TK gene, which contains a bona fide CAAT box (Jones *et al.*, 1985; Graves *et al.*, 1986) did not reveal any differences between HeLa and GC extracts. In addition, *in vitro* competition experiments failed to demonstrate competition between HSV-TK and hGH for GHF-1 binding. These results rule out the possibility that GHF-1 is one of the CAAT-box binding factors characterized previously (Jones *et al.*, 1985; Graves *et al.*, 1986).

The hGH 5′ control region is also recognized by at least two other factors none of which seems to be cell type specific. One of these factors binds to site 2 immediately upstream to the distal GHF-1 site. It is not certain whether binding of this factor is actually required for expression of the hGH gene. Site 3 is present within a region (−289 to −206) which is required for optimal hGH promoter activity, yet its contribution to hGH promoter activity is not absolutely clear because the deleted region contained additional sequences which are not part of site 3. Site 3 is rather large in comparison to the other binding sites and could be recognized, however, by more than a single factor. Preliminary results suggest that a DNA fragment which contains site 3 functions as a transcriptional enhancer (S.Dana, unpublished results). In this respect, the hGH control region resembles that of the rat insulin 1 gene, but the overall organization of *cis*-elements is somewhat different. The insulin gene contains two elements which exhibit cell-type specificity: and enhancer and an upstream promoter element (Edlund *et al.*, 1985). So far, tissue-specific recognition has been demonstrated only for one of the components of the enhancer (Ohlsson and Edlund, 1986). Tissue-specific interaction with *trans*-acting factors was also demonstrated for the immunoglobulin enhancer (Ephrussi *et al.*, 1985; Sen and Baltimore, 1986); and promoter (Landolfi *et al.*, 1986; Staudt *et al.*, 1986) elements. Two tissue specific factors which bind to the immediate 5′ flanking region of the chicken β globin gene were found by Emerson *et al.* (1985) in erythrocyte extracts. However, the relevance of all these factors for expression of these genes is not yet clear, as the importance of their binding sites has not yet been directly tested by construction of the appropriate clustered or individual point mutations.

The recognition of promoter regions of higher eukaryotic protein coding genes by a large number of *trans*-acting factors is not unique to tissue-specific genes. For example, the hMT-II$_A$ promoter requires binding of at least three different trans-acting proteins for maintaining its maximal basal activity (Lee *et al.*, 1987; Karin *et al.*, 1987). Since these factors seem to be ubiquitously distributed in most cell types, the hMT-II$_A$ promoter is not tissue-specific. Likewise the SV40 early promoter is also recognized by multiple factors (Dynan and Tjian, 1983; Wildeman *et al.*, 1986; Lee *et al.*, 1987), however in this case some of the factors appear to be cell-type specific and recognize different but overlapping sequence motifs within the enhancer (Davidson *et al.*, 1986). Yet, the end result is that the SV40 early promoter is active in most cell types. From recent results it seems that such trans-acting factors have to bind simultaneously to a given promoter to cause its maximal activation (Jones *et al.*, 1985; Lee *et al.*, 1987). Therefore a single factor present only in a limited number of cell types could be sufficient for conferring cell type specificity. The combinatorial recognition of promoters by multiple, distinct *trans*-acting factors is particularly suitable for complex, multicellular organisms, as it allows the switching on and off of different classes of structural genes during cellular differentiation, while minimizing the number of regulatory genes.

Finally, while in this work we have concentrated on one element contributing to the tissue-specific expression of the hGH gene, namely the 5′ promoter, we have not ruled out the possible involvement of other DNA sequence elements in establishing the strict cell type specificity of that gene. Elements found downstream to the start of transcription were found to contribute to the differential expression of globin (Charnay *et al.*, 1984; Wright *et al.*, 1985) and immunoglobulin (Grosschedl and Baltimore, 1985; Gillis *et al.*, 1983; Banerji *et al.*, 1983; Queen and Baltimore, 1983) genes. In fact, at least one *cis* element, a glucocorticoid responsive element (GRE) is present in the first intron of the hGH gene (Moore *et al.*, 1985; Slater *et al.*, 1985).

Materials and methods

Cells

For transfection assays GC cells (Tashjian *et al.*, 1968) were grown on 100 mm tissue culture dishes in DMEM supplemented with 12.5% horse serum and 5% fetal calf serum (FCS), penicillin and streptomycin. HeLa cells were grown in DMEM with 2.5% calf serum and 2.5% FCS. All other cell lines were grown in DMEM with 10% FCS.

For preparation of extracts, GC and HeLa cells were grown in spinner flasks in Joklik modified Eagle's medium supplemented as indicated above. The cells were maintained and harvested while still at the exponential growth phase (i.e., $< 10^6$ cells/ml). Some of the HeLa extracts were prepared from cells provided by Don Giard at the MIT Cell Culture Center.

Transfection and determination of gene expression

Transfections were performed by the calcium-phosphate–DNA co-precipitation technique (Graham and Van der Eb, 1973). Cells were incubated with the precipitate for 5 h. The precipitate was removed and the cells were incubated for 2–3 min in Hepes buffered saline containing 15% glycerol at 37°C (glycerol shock), after which the cells were incubated for another 48 h in normal growth medium. For stable transformation the glycerol shock was omitted and cells were given an expression time of 48 h, before applying selective pressure. Cells were harvested, and processed for determination of CAT activity and primer extension analysis as described earlier (Haslinger and Karin, 1985). RNA extraction, blot hybridization and S1 nuclease protection were as previously described (Heguy *et al.*, 1986).

Preparation and fractionation of whole cell extracts

Whole cell extracts were prepared essentially as described by Manley *et al.* (1980). Lysates were dialyzed against TM buffer consisting of 50 mM Tris–HCl (pH 7.9), 1 mM EDTA, 1 mM DTT, 12.5 mM $MgCl_2$, 20% glycerol and 0.1 M KCl. Heparin-agarose was prepared as described by Davison *et al.* (1979). 40 ml of extract at approximately 4–8 mg protein/ml were loaded on a 10 ml heparin–agarose column equilibrated in TM containing 0.1 M KCl. Washing and elution were as described by Dynan and Tjian (1983) Protein concentrations were determined by the Bradford (1976) assay using BSA as a standard. Aliquots were quickly frozen in liquid nitrogen and stored at −80°C.

DNase footprinting assay

DNaseI footprinting reactions were performed as described by Jones *et al.* (1985) with slight modifications. End-labelled DNA probes (1-2 ng), prepared by either 'filling in' with the Klenow fragment of DNA polymerase I or by kinasing using T4 polynucleotide kinase, were incubated in a total volume of 50 μl containing 10% v/v glycerol, 2% polyvinylalcohol, 25 mM Tris–HCl (pH 7.9), 6.25 mM $MgCl_2$, 50 mM KCl, 1 mM EDTA, 0.5 mM DTT and 1 μg poly(dI–dC) (Sigma) with up to 20 μg of extract. Incubation was for 15 min on ice, followed by 2 min at room temperature. An equal volume of 5 mM $CaCl_2$, 1.5 mM EDTA was added, followed by 1–5 μl freshly diluted DNase solution (Worthington) (5 μg/ml) for 1 min at room temperature. Reactions were stopped, extracted and analyzed on 8% polyacrylamide 42% urea gels, as described by Jones *et al.* (1985).

Plasmid constructions

All plasmids were constructed using standard recombinant DNA procedures. pUC-CAT- was constructed by subcloning a *Bgl*II–*Bam*HI fragment containing the CAT gene from pA10CAT-3M (Laimins *et al.*, 1982) into the *Bam*HI site of pUC13. phGH-N (DeNoto *et al.*, 1981) was digested with *Bam*HI (+3), filled in and digested again with *Hin*dIII (site is in pBR322, 23 bp upstream to the *Eco*RI site at position −500 of this gene). The fragments containing the 5′ flanking regions of these genes were inserted between the *Hin*dIII and the filled in *Sal*I sites of pUC-CAT- to yield phGH–CAT. To generate the promoter mutants the *Eco*RI–*Bam*HI 5′ flanking fragment of hGH was cloned between the *Eco*RI and *Bam*HI sites of pUC13 to yield phGH-5′. This plasmid was digested with *Nsi*I, blunt-ended with T4 polymerase, and recircularized in the presence of *Bgl*II linkers to yield phGHΔNsi1. phGH-5′ was also digested with *Nsi*I, treated with S1 nuclease, and religated in the presence of *Bgl*II linkers to yield pΔhGH1, which has a 200 bp deletion around the original *Nsi*I site. phGHΔNco-1 and -2 were generated by digestion of phGH5′ with *Nco*I, treatment with S1 nuclease, and religation in the presence of *Bgl*II linkers. phGH-LS (−121/−126) was generated by substitution of bases −287 to −84 (*Nco*I–*Nsi*I) of phGH-5′ by a hybrid fragment consisting of bases −287 to −130 (*Nco*I–*Bgl*II) from phGHΔ(−82/−128) covalently linked to a synthetic oligonucleotide homologous to positions −121 to −84 of hGH with a *Bgl*II 'sticky end' at position −121. phGHΔ5′-289 was generated by a deletion of the presence of a *Bgl*II linker. The structures of all mutants and constructs were confirmed by restriction mapping and nucleotide sequencing.

Acknowledgements

We would like to thank Dr Norman Eberhardt for providing us with hGH and hCS clones, Ms Fran Denoto and Dr Ron Evans for GC and GH_3 cells, and Ms Sandy Dutky and Jennifer Meek for preparation of the manuscript. The research was supported by funds from the NIH and Chicago Community Trust/Searle Scholars Program to M.K., who is a Searle Scholar. C.L. was supported by a fellowship from the Ministere des Relations Exterieur (France), and M.B. by an EMBO long-term fellowship.

References

Banerji,J., Olsen,L. and Schaffner,W. (1983) *Cell*, **33**, 729–740.
Barsh,G.S., Seeburg,P.H. and Gelinas,R.E. (1983) *Nucleic Acids Res.*, **11**, 3939–3958.
Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.*, **8**, 127–142.
Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248–254.
Charnay,P., Treisman,R., Mellon,P., Chau,M., Axel,R. and Maniatis,T. (1984) *Cell*, **38**, 251–263.
Ciliberto,G., Dente,L. and Cortese,R. (1985) *Cell*, **41**, 531–540.
Davison,B.L., Leighton,T. and Rabinowitz,J.C. (1979) *J. Biol. Chem.*, **254**, 9220–9226.
Davison,I., Fromental,C., Augereau,P., Wildeman,A., Zenke,M. and Chambon,P. (1986) *Nature*, **323**, 544–548.
DeNoto,F.M., Moore,D.D. and Goodman,H.M. (1981) *Nucleic Acids Res.*, **9**, 3719–3730.
Dynan,W.S. and Tjian,R. (1983) *Cell*, **32**, 669–680.
Edlund,T., Walker,M.D., Barr,P.J. and Rutter,W.J. (1985) *Science*, **230**, 912–918.
Efstratiadis,A., Posakony,J.W., Maniatis,T., Lawn,R.M., O'Connell,C., Spritz,R.A., DeRiel,J.K., Forget,B.G., Weisseman,S.M., Slightom,J.L., Blechl,A.E., Smithies,V., Baralle,F.E., Shoulders,C.C. and Proudfoot,N.J. (1980) *Cell*, **21**, 653–668.
Emerson,B.M., Lewis,C.D. and Felsenfeld,G. (1985) *Cell*, **41**, 21–30.
Ephrussi,A.,Church,G.M., Tonegawa,S. and Gilbert,W. (1985) *Science*, **230**, 912–918.
Galas,D. and Schmitz,A. (1978) *Nucleic Acids Res.*, **5**, 3157–3170.
Gillis,S.D., Morrison,S.L., Oi,V.T. and Tonegawa,S. (1983) *Cell*, **33**, 717–728.
Gorman,C., Moffat,L.F. and Howard,B.H. (1982) *Mol. Cell. Biol.*, **2**, 1044–1051.
Graham,R. and Van der Eb,A. (1973) *Virology*, **52**, 456–467.
Graves,B.J., Johnson,P.F. and McKnight,S.L. (1986) *Cell*, **44**, 565–576.
Grosschedl,R. and Baltimore,D. (1985) *Cell*, **41**, 885–897.
Haslinger,A. and Karin,M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8572–8576.
Heguy,A., West,A., Richards,R.I. and Karin,M. (1986) *Mol. Cell. Biol.*, **6**, 2149–2157.
Jones,K.A., Yamamoto,K.R. and Tjian,R. (1985) *Cell*, **42**, 559–572.
Karin,M., Haslinger,A., Heguy,A., Deitlin,T. and Cooke,T. (1987) *Mol. Cell. Biol.*, **7**, 606–613.
Khoury,G. and Gruss,P. (1983) *Cell*, **33**, 313–314.
Laimins,L., Khoury,G., Gorman,C., Howard,B.H. and Gruss,P. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6453–6457.
Landolfi,N.F., Capra,J.D. and Tucker,P.W. (1986) *Nature*, **323**, 548–551.
Lee,W., Haslinger,A., Karin,M. and Tjian,R. (1987) *Nature*, **325**, 368–372.
Manley,J.L., Fire,A., Cano,A., Sharp,P.A. and Gefter,M.L. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3855–3859.
Mason,J.O., Williams,G.T. and Neuberger,M.S. (1985) *Cell*, **41**, 479–487.
McKnight,S.L. and Kingsbury,R. (1982) *Science*, **212**, 316–323.
McKnight,S.L., Kingsbury,R.C., Spence,A. and Smith,M. (1984) *Cell*, **57**, 253–262.
Miller,W.L. and Eberhardt,N.L. (1983) *Endoc. Rev.*, **4**, 97–130.
Moore,D.D., Conkling,M.A. and Goodman,H.M. (1982) *Cell*, **29**, 285–286.
Moore,D.D., Marks,A.R., Buckley,D.I., Kapler,G., Payvar,F. and Goodman,H.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 699–702.
Ohlsson,H. and Edlund,T. (1986) *Cell*, **45**, 35–44.
Ott,M.O., Sperling,L., Herbomel,P., Yaniv,M. and Weiss,M.C. (1984) *EMBO J.*, **3**, 2505–2510.
Queen,C. and Baltimore,D. (1983) *Cell*, **33**, 741–748.
Seeburg,P. (1982) *DNA*, **1**, 239–249.
Selby,M.J., Barta,A., Baxter,J.D., Bell,G.I. and Eberhardt,N.L. (1984) *J. Biol. Chem.*, **259**, 13131–13138.
Sen,R. and Baltimore,D. (1986) *Cell*, **46**, 705–716.
Slater,E.P., Rabenau,O., Karin,M., Baxter,J.D. and Beato,M. (1985) *Mol. Cell. Biol.*, **5**, 2984–2992.

Southern,P.J. and Berg,P. (1982) *J. Mol. Appl. Genet.*, **1**, 327–341.
Staudt,L.M., Singh,H., Sen,R., Wirth,T., Sharp,P.A. and Baltimore,D. (1986) *Nature*, **323**, 640–643.
Tashjian,A.H., Yasamura,Y., Levine,L., Sato,G.H. and Parker,M.L. (1968) *Endocrinology*, **82**, 342–352.
Truong,A.T., Derez,C., Belayew,A., Renard,A., Pictet,R., Bell,G.I. and Martial,J.A. (1984) *EMBO J.*, **3**, 429–436.
Walker,M.D., Edlund,T., Boulet,A.M. and Rutter,W.J. (1983) *Nature*, **300**, 557–561.
Wildeman,A.G., Zenke,M., Schatz,C., Wintzerith,M., Grundstrom,T., Matthes,H., Takahashi,K. and Chambon,P. (1986) *Mol. Cell. Biol.*, **6**, 2098–2105.
Wright,S., Rosenthal,A., Flavell,R. and Grosveld,F. (1984) *Cell*, **38**, 265–273.

Received on December 23, 1986; revised on February 3, 1987