Only three of the seven human chorionic gonadotropin beta subunit genes can be expressed in the placenta

Karen Talmadge[1*+], William R.Boorstein[2*], Nikos C.Vamvakopoulos[3*], Mary-Jane Gething[*] and John C.Fiddes[1*+]

*Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724, and +California Biotechnology, Inc., 2450 Bayshore Frontage Road, Mountain View, CA 94043, USA

ABSTRACT

     Human chorionic gonadotropin (hCG) is a placental hormone
essential for the maintenance of pregnancy. While the alpha
subunit of this hormone is encoded by a single gene, the beta
subunit is encoded by a complex family of seven very similar
genes or pseudogenes. Two approaches have been taken to
establish which of these genes are functional. First, we have
used two restriction enzyme site polymorphisms to correlate 15
independently isolated beta hCG cDNA clones with their
corresponding genes. Second, we have used transient expression
in COS cells to assay for correctly-initiated transcription from
six of the seven beta hCG gene promoters. From these data, we
conclude that, at most, only three of the seven beta hCG genes
are expressed in the placenta. Comparison of the sequences of a
functional and a non-functional beta hCG gene reveals no obvious
differences, such as promoter changes, that could account for
this differential expression.

INTRODUCTION

     Human chorionic gonadotropin (hCG) is a placental

glycoprotein hormone essential for the maintenance of pregnancy.

The protein is a dimer consisting of two dissimilar,

non-covalently associated subunits called alpha and beta (see

ref. 1 for a review). The alpha subunit is common to hCG and to

the three pituitary glycoprotein hormones, luteinizing hormone

(hLH), follicle ·stimulating hormone, and thyroid stimulating

hormone, while the beta subunit of each of the four glycoprotein

hormones is unique and confers biological specificity.

     In the human(2,3) and in the cow(4), it has been shown that

the common alpha subunit is encoded by a single gene, indicating

that the expression of the beta subunit genes controls the tissue

specificity of these hormones. We have also shown that the hLH

beta subunit is encoded by a single gene(5) while, in contrast,

the hCG beta subunit is encoded by a complex family of seven very
similar genes or pseudogenes(5,6). These beta subunit genes were
isolated in three distinct groups of cloned sequences, summarized
in Fig. 1. Although the restriction enzyme maps of these three
groups of genes do not overlap, there is evidence that the beta
hLH and beta hCG genes are all physically linked at one locus on
the chromosome(5,6). These genes have been individually cloned
and characterized by restriction enzyme mapping(5) and by DNA
sequencing(7).

It is intriguing that the hCG beta subunit is encoded by a
multigene family. The human placenta produces an excess of free
alpha subunit protein, and more alpha than beta subunit
mRNA(8,9), yet there are seven beta subunit genes and only a
single alpha subunit gene. Moreover, the beta hCG genes are very
young; homologous proteins with similar functions have been found
in baboons(10) and horses(11), but not in rats(12,13). In
contrast, most other multigene families, such as the globin and
chorion genes, are ancient. Both because of the medical
significance of hCG, and because of possible implications for the
evolution of pseudogenes, we would like to understand which of
the seven beta hCG genes are expressed.

Our detailed restriction enzyme analysis of the seven beta
hCG genes had shown that, although distinct, they were
exceedingly similar(5). Despite their near identity, we had a
preliminary indication that not all of them are expressed in the
placenta. The complete nucleotide sequence of beta hCG genes 5
and 6 (ref. 7) showed that although they are highly homologous,
they differ at codon 117. In beta hCG gene 5, an aspartic acid
is encoded at this position while, in beta hCG gene 6, it is an
alanine. Amino acid 117 has been identified as an aspartic acid
both in the published amino acid sequence from two
groups(14,15,16) and in our beta hCG cDNA sequence(17).
Expression of beta hCG gene 6 would thus give rise to a variant
beta hCG amino acid sequence as yet.undetected.

Two independent approaches have been taken to investigate
which of the seven beta hCG genes are expressed. First, we have
used two restriction enzyme site polymorphisms to correlate 15
independently isolated beta hCG cDNA clones with their

corresponding genes. Second, we have used transient expression
in COS cells(18,19) to assay for correctly-inititiated
transcription from six of the seven beta hCG gene promoters. The
data from these methods are in agreement and reveal that, at
most, only three of the seven beta hCG genes are expressed in the
placenta.

MATERIALS AND METHODS
Bacteria, DNA, and RNA

    Escherichia coli K-12 strain DH1(20) was used for
transformations. Plasmid pHX3C was the gift of D. Hanahan.
Plasmid DNA was isolated by the boiling method(21). Human first
trimester placental polyadenylated RNA was the gift of M. Evinger
and J. L. Roberts.
Isolation of beta hCG cDNA clones

    An oligo-dT primed cDNA library was generated in pBR322(22)
using human first trimester placental RNA and the sequential
linker cloning method(23). The beta hCG cDNA(17) was labeled
with [alpha-^{32}P]-dCTP by random priming(24) and used as a
hybridization probe(25) to isolate additional cDNA clones.
Restriction enzyme analysis of polymorphisms

    The construction of the beta hCG genomic subclones used in
the restriction enzyme analysis was described previously(5).
Plasmids bearing beta hCG genomic and cDNA inserts were digested
with PvuII and 3'-end labeled by incubation with DNA polymerase I
(Klenow fragment) and 0.5 uCi of each of the four [alpha-^{32}P]-
deoxyribonucleoside triphosphates at room temperature for 15 min.
The labeled DNA was extracted with phenol:chloroform (1:1) and
chloroform, and precipitated with ethanol. The DNA samples were
then digested with either AvaI or HinfI, fractionated on a 5%
polyacrylamide gel, and autoradiographed.
Construction of beta hCG gene plasmids for transient expression

    Bacteriophage lambda recombinants bearing human genomic
fragments containing beta hCG genes have been described
previously(5). DNA from recombinant phage lambda 4A-8 (encoding
all of beta hCG gene 2 and part of beta hCG genes 1 and 3),
recombinant phage lambda 4A-16 (encoding all of beta hCG gene 3
and part of beta hCG gene 2, as well as part of beta hLH),

recombinant phage lambda 4A-15 (encoding all of beta hCG genes 5 and 6), and recombinant phage lambda 28-11 (encoding all of beta hCG genes 7 and 8) was digested with KpnI. The KpnI-generated DNA fragments from each recombinant phage were shotgun-cloned into KpnI-digested pXH3C and transformed into E. coli strain DH1. Transformants were screened for the insertion of beta hCG gene sequences by the colony hybridization method(25) using a ^{32}P-labeled beta hCG cDNA probe as described above. Plasmids from positive colonies were digested with KpnI and the size of the KpnI insert was compared on 1.0% agarose gels to a KpnI digest of the appropriate recombinant phage. The orientation of each gene within the plasmid was mapped by HindIII digestion, which cuts once within each of the beta hCG genes and twice in the plasmid pXH3C.

DNA sequencing

M13 vectors mp8 and mp9(26) and the dideoxy sequencing method(27) were used.

Transfection into COS cells

COS-A2 cells, derived from COS-1(18), were a gift of Yakov Gluzman. The cells were cultured as described(18). For transfection experiments, the cells were split 1:5, seeded on 60 mm dishes, and grown overnight to 70% confluence. The same amount of recombinant plasmid, from 200 ng to 2 ug, was used in each transfection experiment, along with expression control plasmid (pXH3C without a beta hCG gene insert) and, in two experiments, transfection control DNAs pSVLHA8(28) or SVEHA20-A⁻(29). DNA was applied to the cells using 1 mg/ml DEAE-dextran(30) for 25 minutes at 37°. The cells were washed, treated with 100 uM chloroquine(31) in 5 ml of medium with serum for 4 hours, washed, fed with 5 ml of medium with serum, and harvested 60 hours after transfection. For transfection controls, cell-surface haemabsorption and haemagglutinin radioimmunoassays were performed as described(28).

RNA isolation

Transfected cells were washed and lysed by the method of Chirgwin et al.(32). The cell lysate was extracted two times with an equal volume of phenol:chloroform (1:1), once with chloroform, and stored at -20^oC as an ethanol precipitate.

SI nuclease analysis

RNA isolated from transfected cells was analyzed for correctly-initiated beta hCG gene transcription by the SI nuclease method of Berk and Sharp(33) as modified by Weaver and Weissmann(34). DNA from subclones containing the 5' end of beta hCG gene 5 and, in one experiment, beta hCG gene 6, were digested with HinfI and 5'-end labelled with [gamma-^{32}P]-ATP by polynucleotide kinase. The 290 base pair fragment spanning 235-535 nucleotides upstream from the beta hCG gene initiation codon(7) was isolated by polyacrylamide gel electrophoresis and used as the SI probe. SI-nuclease resistant fragments were electrophoresed on 40 cm 8% acrylamide, 7 M urea gels that were dried and autoradiographed on screen at -70°C.

RESULTS

Beta hCG 3 is expressed in the placenta

The organization of the seven beta hCG genes (1-3 and 5-8) and of the single beta LH gene is shown in Fig. 1 (refs. 5,6). We had previously subcloned these genes from the bacteriophage lambda recombinants in which they were isolated and compared them in detail by restriction enzyme analysis(5). We found that these genes were very similar, but not identical, with most of the differences located in the 5' untranslated regions and introns. However, we found one difference in the coding region of beta hCG 3 that distinguishes it from the other six beta hCG genes.

Beta hCG 3 has an AvaI site close to the 3' end of the coding region not present in the other six genes(5). We determined by DNA sequencing that this site is generated by a silent third position difference in the codon for amino acid 126 [CCG in beta hCG 5 and 6 (ref. 7); CCC in beta hCG 3 (data not shown)]. The strategy used to assay for the AvaI polymorphism is outlined in Fig. 2A. As shown in Fig. 2A, a 115 b.p. PvuII-AvaI fragment reveals the presence of an AvaI site at codon 126, whereas a 142 b.p. PvuII-AvaI fragment reveals the absence of an AvaI site at codon 126. Fig. 2B demonstrates, as we had shown earlier(5), that only beta hCG gene 3 has the AvaI site at codon 126. The beta hCG gene inserts subcloned from the recombinant phage contain different lengths of beta hCG gene sequences, as
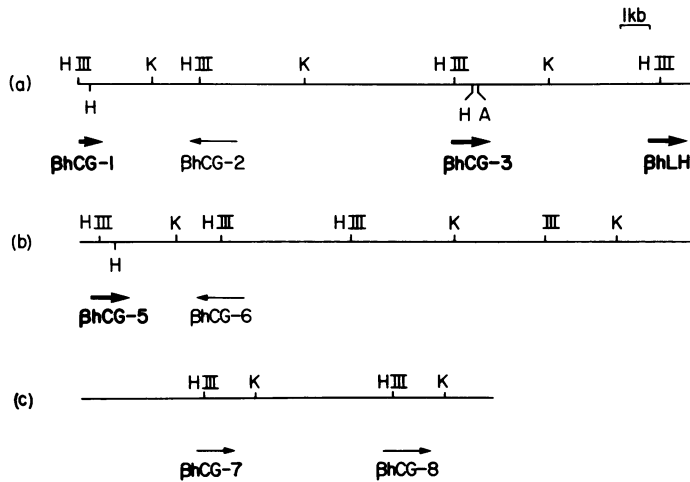
Figure 1   Organization of the three groups, a, b, and c, of genes which hybridize to the cDNA(17) for the hCG beta subunit. The figure is adapted from refs. 5 and 6.   The positions, and 5'-3' orientations, of the seven beta hCG genes and pseudogenes (1-3 and 5-8), and of the single beta hLH gene, are shown by arrows.   The expressed genes, beta hCG genes 1, 3, and 5 [demonstrated in this paper], and beta hLH [assumed to be expressing because it is a single gene(5)], are shown with bold arrows.   Restriction enzyme sites used to subclone each gene for transient expression assays, HindIII (HIII) and KpnI (K), are shown above the lines.   Polymorphic restriction enzyme sites, HinfI (H) and AvaI (A), are shown below the lines.

Fig. 1 shows, and hence have different PvuII-AvaI digestion patterns.   However, only one of the two diagnostic fragments appears in each digestion, and the results are unambiguous.

   We have used the presence of this AvaI site in beta hCG cDNA clones to establish that beta hCG gene 3 is actively expressed in the human placenta.   Using the beta hCG cDNA clone(17) as a hybridization probe, we isolated 14 more independent cDNA clones from a cDNA library constructed from human first-term placental polyadenylated RNA.   These clones were characterized by restriction enzyme mapping to determine that they contain the 3' end of the beta hCG coding region.   When subjected to the analysis of Fig. 2A, two of the 15 clones generated the 115 b.p. PvuII-AvaI fragment diagnostic of beta hCG gene 3, while the other 13 had the 142 b.p. fragment.   Digests of six representative cDNA clones are shown in Fig. 2C.   We

Figure 2   Characterization of the AvaI polymorphism at codon 126. (a) Strategy for analysing beta hCG sequences for the presence or absence of the AvaI site at codon 126. (b) Analysis of the beta hCG genes 1-3 and 5-8. The positions of the 115 and 142 b.p. PvuII-AvaI fragments are shown. Size markers are in b.p. (c) Analysis of six representative beta hCG cDNAS, 1-6. Five of the six clones have the 142 b.p. fragment (142); one has the 115 b.p. fragment (115).

confirmed the presence at codon 126 of the AvaI site in one of
the cDNA clones by DNA sequencing (data not shown). These data
demonstrate that beta hCG gene 3 is an actively expressed gene in
human first-trimester placenta, accounting for about 15% of the
total beta hCG gene expression.

Expression of beta hCG genes 2, 6, 7, and 8, encoding a variant
beta hCG sequence, was not detected in the placenta

        When we determined the complete nucleotide sequences of beta
hCG genes 5 and 6 (ref. 7), we found that beta hCG gene 6 encodes
alanine (GCC) at position 117, rather than the aspartic acid
(GAC) found in beta hCG gene 5. Aspartic acid is found in this
position in beta hCG gene 3, in the previously isolated beta hCG
cDNA clone(17), and in the amino acid sequences from two
different laboratories(14,15,16). Translation of transcripts
from beta hCG 6 would result in a protein with a different
sequence whose change from aspartic acid (an acidic amino acid)
to alanine (a neutral amino acid) might alter the protein
structure and function. Because of this, we wanted to determine
both how many of the genes could potentially encode the alanine
117 variant and whether any of these genes are expressed in the
placenta.

        The alanine 117 variant is associated with a restriction
site difference, so we analyzed the beta hCG genes and cDNA
clones with a strategy similar to the one we used for the AvaI
polymorphism. This sequence difference, GACTC in gene 5 and
GCCTC in gene 6, corresponds to the loss in gene 6 of a HinfI
site (GANTC), as shown in Fig. 3A. The presence of a HinfI site
at this position thus indicates the usual aspartic acid rather
than the alanine variant. [It is also possible to maintain the
HinfI site at this position and make a conservative change from
aspartic acid to glutamic acid, or to lose the HinfI site by many
changes other than aspartic acid to alanine. However, there are
only two nucleotide differences between the coding regions of
beta hCG genes 5 and 6 (ref. 7), and the extreme similarity of
the restriction maps of the other beta hCG genes to the maps of
these two genes(5) argues against the likelihood of many other
changes.]

        All seven beta hCG genes were therefore analysed by a
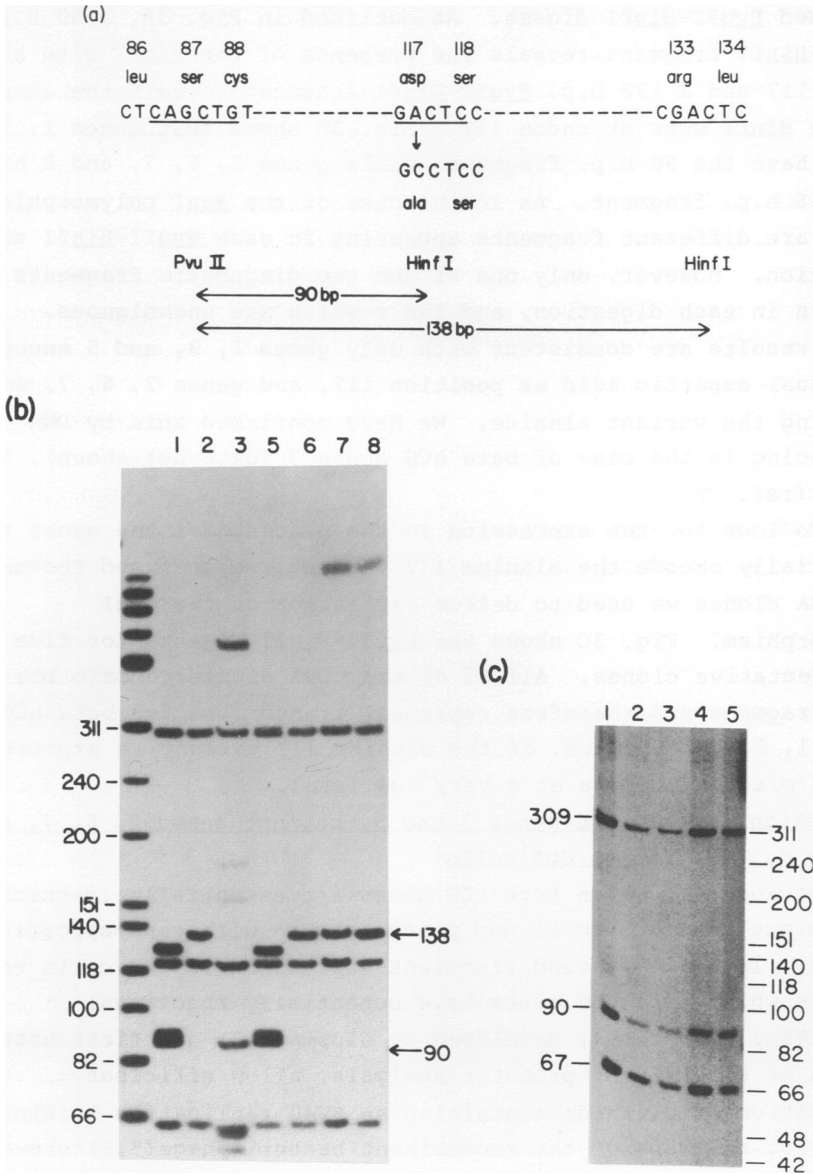
Figure 3 Characterization of the HinfI polymorphism at codon 117. (a) Strategy for analysing beta hCG sequences for the presence or absence of the HinfI site at codon 117. (b) Analysis of the beta hCG genes 1-3 and 5-8. The positions of the 90 and 138 b.p. PvuII-HinfI fragments are shown. Size markers are in b.p. (c) Analysis of five representative beta hCG cDNAS, 1-5. All five had 90 b.p. fragments (90).

combined PvuII-HinfI digest. As outlined in Fig. 3A, a 90 b.p.
PvuII-HinfI fragment reveals the presence of the HinfI site at
codon 117 and a 138 b.p. PvuII-HinfI fragment reveals the absence
of the HinfI site at codon 117. Fig. 3B shows that genes 1, 3,
and 5 have the 90 b.p. fragment, while genes 2, 6, 7, and 8 have
the 138 b.p. fragment. As in the case of the AvaI polymorphism,
there are different fragments appearing in each PvuII-HinfI gene
digestion. However, only one of the two diagnostic fragments
appears in each digestion, and the results are unambiguous.
These results are consistent with only genes 1, 3, and 5 encoding
the usual aspartic acid at position 117, and genes 2, 6, 7, and 8
encoding the variant alanine. We have confirmed this by DNA
sequencing in the case of beta hCG genes 3 (data not shown), 5,
and 6 (ref. 7).

To look for the expression in the placenta of the genes that
potentially encode the alanine 117 variant, we analysed the same
15 cDNA clones we used to detect expression of the AvaI
polymorphism. Fig. 3C shows the PvuII-HinfI digests for five
representative clones. All 15 of the cDNA clones contain the 90
b.p. fragment and therefore represent transcripts for beta hCG
genes 1, 3, or 5. Thus, if the alanine 117 variant is expressed
in the placenta, it is at a very low level.

Expression of beta hCG genes 3 and 5, but not genes 2, 6, 7, and
8, can be detected in COS cells

Because the seven beta hCG genes are essentially identical,
it was not possible to S1 map placental RNA with gene-specific
probes. Instead, we used transient expression in COS cells to
analyse which beta hCG genes have potentially functional
promoters. COS cells, developed by Gluzman(18) and first used by
Mellon et al.(19) for promoter analysis, allow efficient
replication of plasmids containing an SV40 replication origin.

From the maps of the recombinant bacteriophage(5,6) shown in
Fig. 1, we knew that beta hCG genes 2, 3, 6, and 8 can be
isolated intact on individual KpnI fragments of known sizes using
the KpnI sites between the genes. In the case of beta hCG genes
5 and 7, one of the two KpnI sites comes from within the
bacteriophage lambda vector charon 4A(35). (Beta hCG gene 1 was
not subcloned because the gene isolate lacks the 5',
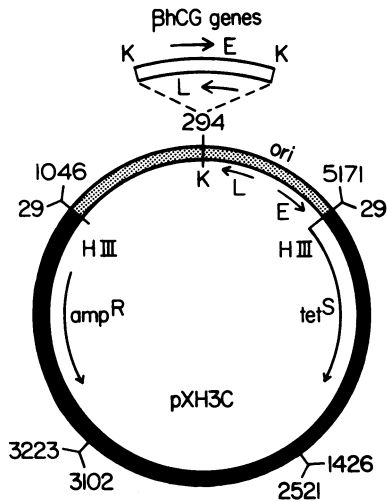
Figure 4 Structure of the vector pXH3C with the beta hCG genes inserted at the unique KpnI site. The shaded region is from pXf3 (ref. 20), a poisonless(36) pBR322 (ref. 22) derivative with deletions between nucleotides 1426-2521 and 3102-3223 (numbers from ref. 51). The SV40 HindIII C fragment (nucleotides 5171-1046, numbers from ref. 52)(stippled) is inserted at the unique HindIII (HIII) site of pXf3 (pBR322 nucleotide 29, ref. 51). The location of the SV40 replication origin (ori) and the direction of early (E) and late (L) transcription are shown with arrows. The beta hCG genes were isolated as KpnI (K) fragments (unshaded) and cloned into the SV40 KpnI site (SV40 nuclotide 294, ref. 52). The genes were isolated oriented in the same direction as the SV40 early (E) or late (L) transcription. amp^R = ampicillin resistance gene; tet^S = tetracyline resistance gene destroyed by the insertion of the SV40 HindIII C fragment.

promoter-containing, end.) Thus, we subcloned each beta hCG gene as a KpnI fragment into the shuttle vector, pXH3C. pXH3C is pXf3 (ref. 20), a poisonless(36) pBR322 (ref. 22) derivative, with the HindIII C fragment of SV40 inserted into the plasmid HindIII site (see Fig. 4). The HindIII C fragment contains the SV40 origin of replication, allowing the vector to be efficiently replicated in COS cells, and the unique KpnI site used for subcloning the beta hCG genes. To control for the potential influence of the enhancer and the early and late SV40 promoters, which are also on the SV40 HindIII C fragment, we attempted to isolate plasmids with beta hCG gene inserts in both orientations, calling the construction E (for early) if the beta hCG gene promoter was

oriented on the plasmid in the same direction as the SV40 early promoter and L (for late) if the beta hCG gene promoter was oriented on the plasmid in the same direction as the SV40 late promoter (see Fig. 4). We obtained beta hCG genes 2, 3, 5, 6, and 7 in both orientations, but despite multiple attempts could only isolate gene 8 in the late orientation.

We had previously determined by Northern blot analysis(37) that human first-trimester placental polyadenylated RNA has a single beta hCG message(2). The size of the protected fragment obtained by SI nuclease mapping showed that this message initiates at around 352-353 b.p. from the start of translation(6), although the steric effect of the cap structure may mean that the 5'-untranslated region is actually 3-5 bases shorter(34) than this estimate.

To assay the beta hCG genes for promoters functional in COS cells, we transfected pXH3C and beta hCG gene bearing derivatives of pXH3C into COS cells using the DEAE-dextran method(30), enhanced the transfection efficiency with chloroquine treatment(31), and harvested the RNA 60 hours after transfection. SI nuclease analysis was used to map the 5' ends of the beta hCG transcripts generated in COS cells and to compare these to the beta hCG message in placental RNA. As the SI mapping strategy in Fig. 5A shows, a 290 b.p. beta hCG HinfI fragment spanning the region of transcription initiation yields a 126 b.p. protected fragment as the major species. This result, with human first trimester placental polyadenylated RNA, is shown in Fig. 5B. When RNA from COS cells is probed using the same strategy, we found that only RNA isolated from cells transfected with plasmids bearing beta hCG genes 3 and 5 in either orientation contain correctly-initiated transcripts (Fig. 5C, lanes 3E, 3L, 5E, and 5L), and that RNA isolated from cells transfected with the plasmid alone or with plasmid bearing beta hCG genes 2, 6, 7, and 8, do not contain these transcripts (Fig. 5C, lanes C, 2E, 3E, 7E, 8L). We have tested the expression of beta hCG genes 2, 6, and 7 in the orientation not shown in Fig. 5C, and again found no transcripts (data not shown).

Because the level of correctly-initiated transcription in the beta hCG gene 3 transfections is quite low, we have shown a

long exposure in Fig. 5C. A lighter exposure of this same gel
shows that the correctly-initiated beta hCG gene 3 and 5
transcripts are 126 b.p. and, in the case of beta hCG gene 5, but
not 3, that this is the major transcript (data not shown). The
SI analyses of both beta hCG genes 3 and 5 cloned in either
orientation show identical higher molecular weight bands which
may represent transcripts initiating as far as 80 nucleotides
upstream from the correct initiation site (Fig. 5C, lanes 3E, 3L,
5E, 5L). Many of these transcripts are also present as minor
species in the placental polyadenylated RNA control (compare Fig.
5B, lane A+, with the overexposed Fig. 5C, lane A+). It is also
possible that these bands do not come from extra upstream
initiation sites, but are the result of readthrough
transcription, either from SV40 promoters or from the cloned beta
hCG promoters, followed by splicing to generate new 5' ends.
However, this is unlikely since, even after a long exposure,
these larger transcripts are totally absent from cells
transfected with beta hCG genes that do not give
correctly-initiated transcripts (Fig. 5C, lanes 2E, 6E, 7E, and
8L), as well as from cells transfected with the pXH3C vector
control (Fig. 5C, lane C). SI products in the genes 2, 6, 7, and
8 lanes of Fig. 5C are also present in the control lane, and are
thus not due to the beta hCG gene insert. The larger transcripts
of beta hCG genes 3 and 5 are therefore specific to expressing
beta hCG genes.

Because of the positions of the KpnI sites used to subclone
each of the beta hCG genes, there is a different amount of
upstream sequence in each construction (see Fig. 1).
Specifically, there are 3.2 (gene 2), 4.8 (gene 3), 0.3 (gene 5),
7.8 (gene 6), 4.0 (gene 7), and 4.7 (gene 8) kb of beta hCG gene
sequence upstream from the start of transcription in these
constructions. Since 0.3 kb, the shortest amount of upstream
sequence in any of the constructions for gene transfer, allows
the expression of beta hCG gene 5 in COS cells (Fig. 5C, lanes
5E, 5L), we conclude that the entire promoter region of every
gene is present. Moreover, since the beta hCG gene 3 promoter,
with 4.7. kb of upstream sequence, is functional in COS cells
(Fig. 5C, lanes 3E, 3L), while the beta hCG gene 2 promoter, with
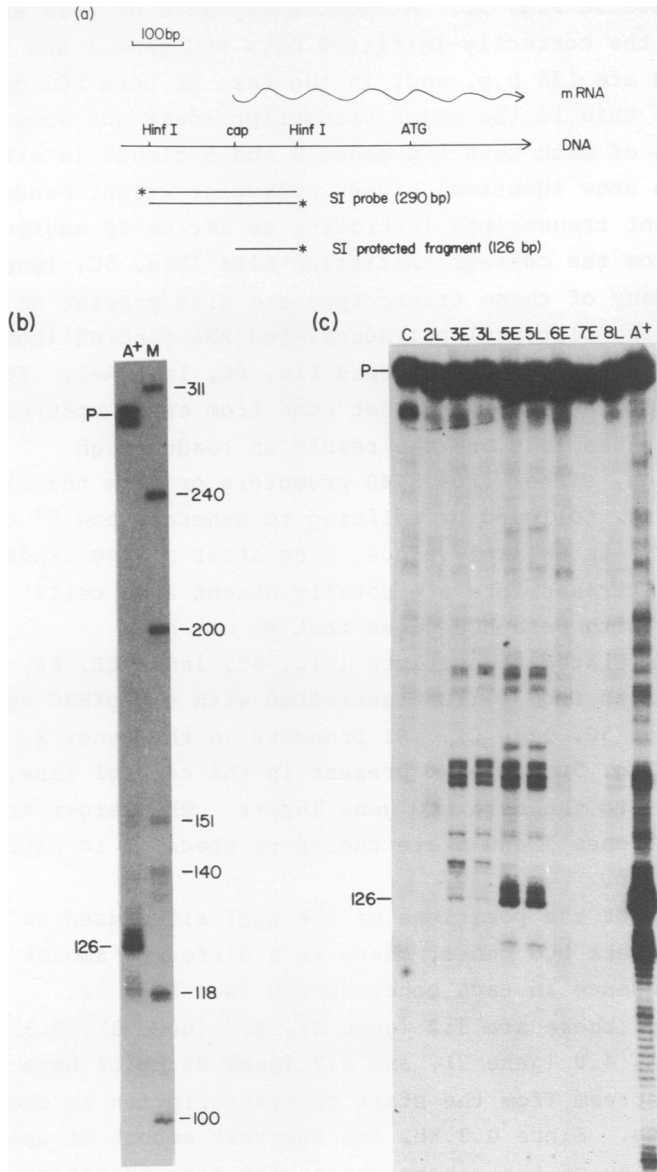
Figure 5 SI analysis of beta hCG gene transcripts. (a) Strategy of SI analysis. Asterisks indicate radioactive labelling. (b) SI analysis of placental RNA. Markers (M) are in b.p. The 126 b.p. SI-protected fragment (126) and 290 b.p. probe (P) are indicated in the SI experiment with polyadenylated placental RNA (A+). The dried gel was exposed on film for 12 hours. (c) SI analysis of RNA from COS cells transfected with pXHC3 (C) or

pXHC3 derivatives bearing beta hCG genes 2 in the late
orientation (2L), 3 in the early orientation (3E), 3 in the late
orientation (3L), 5 in the early orientation (5E), 5 in the late
orientation (5L), 6 in the early orientation (6E), 7 in the early
orientation (7E), and 8 in the late orientation (8L), compared to
an SI analysis of polyadenylated placental RNA (A+). The probe
(P) and 126 b.p. SI-protected fragment (126) are indicated. The
same dried gel shown in (b) was exposed on film for 10 days.

only 3.2 kb of upstream sequence, is not functional in COS cells
(Fig. 5C, lane 2E), we believe that these extra upstream
sequences do not contain a region that is artificially inhibitory
in this promoter assay system.

We controlled for transfection efficiency in two different
ways. In one experiment, we cotransferred pXH3C or its beta hCG
gene derivatives into COS cells with pSVLHA8 (ref. 28), a
construction bearing the influenza haemagglutinin gene fused to
the SV40 early promoter which produces haemagglutinin when
transfected into COS cells. We did duplicate cotransfections and
analysed the cells of one group for expression of haemagglutinin
by cell-surface haemabsorption, then harvested the RNA of the
other group and analysed it for correctly-initiated beta hCG gene
transcripts. In this case, the transfection efficiency varied by
less than a factor of two, and we did not adjust the amount of
RNA we used for the beta hCG SI analysis. In the second
experiment, we cotransferred pXH3C or its beta hCG gene
derivatives with SVEHA20-A⁻, a haemagglutinin gene deletion
mutation lacking its anchor sequence(29). COS cells secrete
haemagglutinin when transfected with this DNA(29). When we
analysed the culture medium by radioimmunoassay for levels of
haemagglutinin production, the haemagglutinin levels varied as
much as threefold, and we adjusted the amount of RNA isolated
from the same cells for the beta hCG SI analysis to account for
this. In both these experiments, we saw essentially the same
results as those presented in Fig. 5C (data not shown).

In most experiments, the SI probe was from beta hCG gene 5.
It was possible that we were not detecting transcripts from genes
2, 6, 7, and 8 as a result of the SI nuclease digesting the DNA
at nucleotide differences across the probed region. To test this
possibility, we isolated the SI probe from beta hCG gene 6 as
well as from beta hCG gene 5, and hybridized them separately both

```
                                    -250
5                        AATAATCAGTTAAATCACCTGAAGCACACGCATTTCCGGGGACCGCTCCGGGCA
6                        AATAATCAGTTAAATCACCTGAAGCACACGCATTTCCGGGGACCGCTCCGGGCA


                         -200                                               -150
5   TCCTGGCTTGAGGGTAGAGTGGGCGGAGGTTCCTAAGGGAGAGGTGGGGCTCGGGCTGAATCCCTCGTTGGGGGG
6   TCCTGGCTTGAGGGTAGGGTGGGCGGAGGTTCCTAAGGGAGAGGTGGGGCTCGGGCTGAATCCCTCGTTGGGGGG
                       *


                                    -100
5   CATCTGGGTCAAGTGGCTTCCCTGGCAGCACAGTCACGGGGAGGCCCTCTCTCATTGGGCAGAAGCTAAGTCCGA
6   CATCTGGGTCAAGTGGCTTCCCTGGCAGCACAGTCACGGGGAGACCCTCTCTCACTGGGCAGAAGCTAAGTCCGA
                                                   *          *


                         -50                                                -1
5   AGCCGCGCCCCTCCTGGGAGGTTGAACTGTGGTGCAGGAAAGCCTCAAGTAGAGGAGGGTTGAGGCTTCAATCC
6   AGCCGCGCCCCTCCTGTTAGGTTGGACTGTGGTGCAGGAATGGCTCAAGTAGAGGAGAGTTGAGGCTTCAGTCC
                  **       *                 * *                 *            *
```

Figure 6 Comparison of beta hCG genes 5 and 6 sequences 277 b.p. upstream from the initiation of beta hCG transcription. The base identified as -1 is nucleotide 190 in ref. 7. Differences between the two sequences are marked below with an asterisk. The potential TATA box at -25, enhancer at -49, and CCAAT box at -89, are underlined.

to human first-trimester placental polyadenylated RNA and to RNA isolated from COS cells transfected with the beta hCG gene plasmids. The results with both probes were identical (data not shown), and were the same as those shown in Fig. 5C.

    We analysed the detection limits of our COS cell assay system by various exposures of the gel shown in Fig. 5C. At eight hours, beta hCG-specific transcripts from beta hCG genes 3 and 5 were visible (data not shown), where the 30-fold longer exposure of Fig. 5C did not reveal any beta hCG-specific transcripts from beta hCG genes 2, 6, 7, and 8. Thus, if beta hCG genes 2, 6, 7, and 8 are expressed in COS cells, they must function at a level less than 3% of beta hCG genes 3 and 5. (Beta hCG gene 1 could not be tested in this system.)

DNA sequence of the beta hCG genes 5 and 6 promoters

    We have previously presented the complete DNA sequence of beta hCG genes 5 and 6, including 190 bases upstream from the region of transcription initiation(7). In Fig. 6, we show a comparison of these sequences, and we now include another 87 bases not previously published, determined by the strategy in ref. 7. As Fig. 6 reveals, these sequences are exceedingly similar, with only 10 nucleotides different out of 277.

Table 1

Summary of evidence that beta hCG genes 3 and 5, and possibly 1, but not 2. 6, 7, or 8, are expressed in the placenta.

Beta hCG	AvaI site at codon 117	HinfI site at codon 126	Functional promoter in COS cells	Splice sites intact
Gene 1	-	+	ND	ND
Gene 2	-	-	-	-*
Gene 3	+	+	+	ND
Gene 5	-	+	+	+
Gene 6	-	-	-	+
Gene 7	-	-	-	ND
Gene 8	-	-	-	ND
cDNAs	2/15	15/15	NA	NA

+ = present, – = absent, ND = not determined, NA = not applicable,
* = ref. 38.

DISCUSSION

The beta subunit of hCG is encoded by a complex family of seven very similar genes or pseudogenes (see Fig. 1). We have used two restriction enzyme site polymorphisms, and transient expression in COS cells, to analyse which of the seven beta hCG genes are functional. These two polymorphisms are the only restriction site differences in the coding region we have yet detected among these seven genes. The data from these two independent approaches are in agreement and demonstrate that, at most, only three of the seven beta hCG genes are expressed in the placenta. The COS cell experiments do not necessarily reflect which beta hCG genes are expressed in vivo, as these genes may have tissue specific control elements. However, we note that, for these genes, the COS cell data and the independent restriction site polymorphism analysis agree.

Table 1 summarizes our evidence that beta hCG genes 3 and 5, and possibly 1, but not genes 2, 6, 7, or 8, are expressed in the placenta. An AvaI site present in beta hCG gene 3, but not the other six beta hCG genes, was found in 2 of our 15 independently

isolated beta hCG cDNA clones, demonstrating conclusively that beta hCG gene 3 is expressed in the placenta. Consistent with this result, beta hCG gene 3 produces correctly-initiated transcripts when transfected into COS cells. A HinfI site, present in beta hCG genes 1, 3, and 5, but not 2, 6, 7, or 8, is found in the same 15 cDNA clones, indicating that only beta hCG genes 1, 3, and 5 can be expressed in the placenta. This result is consistent both with the data from the AvaI polymorphism, which demonstrates directly that beta hCG gene 3 is expressed in placenta, and with the COS cell experiments, where only beta hCG genes 3 and 5, but not 2, 6, 7, and 8, produce detectable transcripts. (Beta hCG gene 1 was not tested in COS cells because our gene isolate lacks the 5', promoter-containing, end.) Moreover, these data are supported by Policastro et al.(38), who obtained partial nucleotide sequences of three of the seven beta hCG genes. In one of these genes, which corresponds to beta hCG gene 2, the splice site donor sequence in the first intron is GA(38), rather than the consensus GT(39). This suggests that beta hCG gene 2 is a pseudogene as correct splicing of the mRNA would not be expected to take place.

Because we isolated only 15 independent cDNA clones, our restriction site polymorphism analysis would not detect a low level of gene expression. We can only safely conclude from these data that the combined transcripts from beta hCG genes 2, 6, 7, and 8 represent less than about 10% of the total beta hCG gene expression. However, the COS cell experiments, which also demonstrate the absence of beta hCG gene 2, 6, 7, and 8 expression, show that these genes are expressed at less than 3% of the levels of beta hCG genes 3 and 5. Thus, if beta hCG genes 2, 6, 7, and 8 are functional, they function at very low levels. The SI analysis would not have detected RNA from beta hCG genes 2, 6, 7, or 8, if their transcriptional start sites are very different from those of genes 3 and 5. However, if they are different, these genes must be expressed in tissues other than the placenta, because we detect only one message by Northern blot analysis of placental RNA(2). It is possible that these genes are expressed at low levels in other tissues not normally thought to be making hCG or in ectopic tumors that synthesize hCG(40).

We have sequenced(7) beta hCG genes 5 (functional) and 6 (non-functional), and find no splice site mutation in gene 6. We have also sequenced 277 bases upstream from the start of beta hCG gene transcription, as shown in Fig. 6. Two characteristic canonical sequences which precede the start of transcription have been identified in eukaryotic genes (see compilation in ref. 41). The first of these sequences, the TATA box, occurs at about 30 bases upstream from the start of transcription. In both beta hCG genes 5 and 6, we find a weak match to this sequence, TAGAGGA (at -25), which has 4 out of 7 bases of the canonical sequence, TATA(T/A)A(T/A). However, transcription from some viral promoters occurs without the presence of a TATA box (see ref. 42 and the refs. therein). The second of the these sequences, the CCAAT box, usually occurs at about 60-80 nucleotides from the start of transcription. We find another weak match for this sequence for both genes 5 and 6, AGCTAAGTC (at -88), with 4 of 9 bases matching the canonical sequence GG(C/T)CAATCT. Thus, both genes, one of which produces beta hCG gene-specific transcripts in COS cells and the other of which does not, have the same TATA and CCAAT boxes.

Two other regions required for efficient transcription have been identified in some genes. The first of these is the enhancer(43,44). We find a reasonable match to the enhancer consensus sequence, GXTGTGG(T/A)(T/A)(T/A)(45), at -49 (see Fig. 6). There are even better matches at three positions in the first intron, one position in the first exon, and two positions in the second intron(7). We do not know if these potential enhancer sequences have any physiological significance for beta hCG gene expression, but there are no differences between the two genes that could explain why beta hCG gene 5, but not 6, is expressed at detectable levels. However, one enhancer sequence has been identified at 340 nucleotides upstream from the Adenovirus-2 ElA gene transcription start site(45), and our sequence comparison does not extend that far upstream. A second region, identified in some genes and required for efficient transcription, is found upstream from the CCAAT box, around -100 or further (see ref. 46 and refs. therein). These regions do not resemble each other or the upstream sequences of the beta hCG

genes 5 and 6. However, as with these other genes, it is likely
that some of the sequences upstream from the beta hCG CCAAT box
are necessary for transcription.

There are 10 nucleotides different between the upstream
sequences of beta hCG genes 5 and 6, or 3.6%, across the 277
nucleotides sequenced (see Fig. 6). This is comparable to the
3.2% homology in the beta hCG gene 5 and 6 introns(7). However,
9 of the 10 nucleotide changes occur within the first 109
nucleotides, a difference of 8.3%, and the genes become more
similar beyond this 109 base region. Thus, the differences
between beta hCG genes 5 and 6 cluster in the promoter region.
Although genes have elements necessary for transcription in
common, the need for other upstream sequences, specific to gene
type, indicates that the alterations that abolish beta hCG gene 6
transcription could be contained among these 9 base changes. All
of the naturally occurring promoter mutations, identified in
humans with beta-thalassemias, lower but do not abolish
transcription. Three of these are single bases changes in the
beta-globin gene TATA box(47,48,49), and one is a single base
change in the -100 region(50). Thus, it is likely that multiple
changes have been necessary to so greatly reduce the level of
beta hCG gene transcription. Since the beta hCG gene family is
young, these data indicate that the evolution of pseudogenes
might in some cases begin with a small number of promoter region
changes that abolish or greatly lower transcription.

[1]Present address: California Biotechnology, Inc., 2450 Bayshore
Frontage Rd., Mountain View, CA 94043 USA
[2]Present address: Department of Molecular Biology, Madison, WI
53706 USA
[3]Present address: 28 Kononos, Pagrati, Athens, Greece

REFERENCES
1. Pierce, J.G., and Parsons, T.F. (1981) Ann. Rev. Biochem.
    50, 465-495.
2. Fiddes, J.C., and Goodman, H.M. (1981) J. Mol. Appl. Genet.
    1, 3-18.
3. Boothby, M., Ruddon, R.W., Anderson, C., McWilliams, D., and
    Boime, I. (1981) J. Biol. Chem. 256, 5121-5127.
4. Goodwin, R.G., Moncman, C.L., Rottman, F.M., and Nilson, J.H.
    (1983) Nucl. Acids Res. 11, 6873-6882.
5. Talmadge, K., Boorstein, W.R., and Fiddes, J.C. (1983) DNA
    2, 281-289.
6. Boorstein, W.R., Vamvakopoulos, N.C., and Fiddes, J.C. (1982)
    Nature 300, 419-422.
7. Talmadge, K., Vamvakopoulos, N.C., and Fiddes, J.C. (1984)
    Nature 307, 37-40.
8. Chatterjee, M., and Munro, H.N. (1977) Biochem. Biophys.
    Res. Commun. 77, 426-433.
9. Boothby, M., Kukowska, J., and Boime, I. (1983) J. Biol.
    Chem. 258, 9250-9253.
10. Canfield, R.E., Birken, S., Morse, J.H., and Morgan, F.J.
    (1976) In Peptide Hormones (ed. Parsons, J.A.), 299-315
    (University Park Press, Baltimore).
11. Ward, D.N., Moore, W.T. Jr., and Burleigh, B.D. (1982) J.
    Prot. Chem. 1, 263-280.
12. Wurzel, J.M., Curatola, L.M., Gurr, J.A., Goldschmidt, A.M.,
    and Kourides, I.A. (1983) Endocrinology 113, 1854-1852.
13. Tepper, M.A., and Roberts, J.L. (1984) Endocrinology, in
    press.
14. Morgan, F.J., Birken, S., and Canfield, R.E (1975) J. Biol.
    Chem. 250, 5247-5258.
15. Birken, S., and Canfield, R.D. (1977) J. Biol. Chem. 252,
    5386-5392.
16. Keutmann, H.T., and Williams, R.M. (1977) J. Biol. Chem.
    252, 5393-5397.
17. Fiddes, J.C., and Goodman, H.M. (1980) Nature 286, 684-687.
18. Gluzman, Y. (1981) Cell 23, 175-182.
19. Mellon, P., Parker, V., Gluzman, Y., and Maniatis, T. (1981)
    Cell 27, 279-288.
20. Hanahan, D. (1983) J. Mol. Biol. 166, 559-580.
21. Holmes, D.S., and Quigley, M. (1981) Anal. Biochem. 114,
    193-197.
22. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C.,
    Heyneker, H.L., and Boyer, H.W. (1977) Gene 2, 95-113.
23. Helfman, D.M., Feramisco, J.R., Fiddes, J.C., Thomas, G.P.,
    and Hughes, S.H. (1983) Proc. Natl. Acad. Sci. USA 80,
    31-35.
24. Taylor, J.M., Illmensee, R., and Summers, S. (1976) Biochem.
    Biophys. Acta 442, 324-330.
25. Grunstein M., and Hogness, D.S. (1975) Proc. Natl. Acad.
    Sci. USA 72, 3961-3965.
26. Messing, J., and Vieira, J. (1982) Gene 19, 269-276.
27. Sanger, F., Nicklen, S., and Coulsen, A.R. (1977) Proc.
    Natl. Acad. Sci USA 74, 5463-5467.
28. Gething, M.-J., and Sambrook, J. (1981) Nature 293, 620-625.
29. Gething, M.-J., and Sambrook, J. (1982) Nature 300, 598-603.
30. McCutchan, J.H., and Pagano, J.S. (1968) J. Natl. Cancer
    Inst. 41, 351-357.

31. Luthman, H., and Magnusson, G. (1983) Nucl. Acids Res. 11, 1295-1307.
32. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979) Biochem. 18, 5294-5299.
33. Berk, A.J., and Sharp, P. (1978) Proc. Natl. Acad. Sci. USA 75, 1274-1278.
34. Weaver, R.F., and Weissmann, C. (1979) Nucl. Acid Res. 7, 1175-1193.
35. Blattner, F., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Fuber, H.E., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L., and Smithies, O. (1977) Science 196, 161-169.
36. Lusky, M., and Botchan, M. (1981) Nature 293, 79-81.
37. Alwine, J.C., Kemp, D.J., and Stark, G.R. (1977) Proc. Natl. Acad. Sci. (USA) 74, 5350-5354.
38. Policastro, P., Ovitt, C.E., Hoshina, M., Fukuoka, H., Boothby, M.R., and Boime, I. (1983) J. Biol. Chem. 258, 11492-11499.
39. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978) Proc. Natl. Acad. Sci (USA) 75, 4853-4857.
40. In (1976) Nature 260, 480-481.
41. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
42. Catterall, J.F., O'Malley, B.W., Robertson, M.A., Staden R., Tanaka, Y., and Brownlee, G.G. (1978) Nature 275, 510-513.
43. Benoist, C. and Chambon, P. (1981) Nature 290, 304-310.
44. Fromm, M. and Berg, P. (1982) J. Mol. Appl. Genet. 1, 457-481.
45. Hen, R., Borelli, E., Sassone-Corsi, P. and Chambon, P. (1983) Nucl. Acids Res. 11, 8747-8760.
46. Dierks, P., van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J., and Weissmann, C. (1983) Cell 32, 695-706.
47. Poncz, M., Ballantine, M., Solowiejczyk, D., Barak, I., Schwartz, E. and Surrey, S. (1982) J. Biol. Chem. 257, 5994-5996.
48. Orkin, S. H., Sexton, J. P., Cheng, T.-C., Goff, S. C., Giardina, J. V., Lee, J. I., and Kazazian, H. H., Jr. (1983) Nucl. Acids Res. 11, 4727-4734.
49. Antonarkis, S. E., Irkin, S. H., Cheng, T.-C., Scott, A. F., Sexton, J. P., Trusko, S. P., Charache, S., and Kazazian, H. H., Jr. (1984) Proc. Natl. Acad. Sci. USA 81, 1154-1158.
50. Orkin, S.H., Kazazian, H.H., Jr., Antonarkis, S.E., Goff, S.C., Boehm, C.D., Sexton, J.P., Waber, P.G., and Giardina, P.J.V. (1982) Nature 296, 627-631.
51. Sutcliffe, J.G. (1979) Cold Spring Harbor Symp. 43, 77-90.
52. Tooze, J. (1981) In DNA Tumour Viruses, 2nd edn., J. Tooze, Ed., pp. 801-840, Cold Spring Harbor, NY.