Volume 9 Number 15 1981 Nucleic Acids Research

Human growth hormone DNA sequence and mRNA structure: possible alternative splicing

Frances M.DeNoto*, David D.Moore† and Howard M.Goodman*

*Howard Hughes Medical Institute, and †Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA

Received 20 April 1981

ABSTRACT

We have determined the complete sequence of the human growth hormone (hGH) gene and the position of the mature 5' end of the hGH mRNA within the sequence. Comparison of this sequence with that of a cloned hGH cDNA shows that the gene is interrupted by four intervening sequences. S1 mapping shows that one of these intervening sequences has two different 3' splice sites. These alternate splicing pathways generate hGH peptides of different sizes which are found in normal pituitaries. Comparison of sequences near the 5' end of the hGH mRNA with a similar region of the α subunit of the human glycoprotein hormones reveals an unexpected region of homology between these otherwise unrelated peptide hormones.

INTRODUCTION

The synthesis and release of human growth hormone (hGH) by the anterior pituitary has a number of effects, including both short term metabolic alterations and long term promotion of growth (1). Some of the biological activities of hGH overlap with those of two other hormones which are structurally related to hGH. hGH is 92% homolgous at the cDNA level with human chorionic somatomammotropin (hCS, placental lactogen), which is synthesized in the placenta (2, 3, 4). These two are more distantly related to human prolactin (hPL), which is synthesized in the anterior pituitary. This gene family provides an excellent system for the study of both evolution of related genes and tissue specific gene expression. In order to more fully understand the structural relationships among these genes and, possibly, the signals which regulate their expression, we have cloned several genes related to hGH. We present here the results of nucleotide sequence analysis of the gene which encodes hGH.

MATERIALS AND METHODS

The 2.6 kb EcoRI fragment containing the hGH gene was isolated from a

size fractionated library of human placental DNA cloned in λWES, as decribed (5), and subcloned into pBR322 for further analysis. DNA sequence determinations were by the method of Maxam and Gilbert (35). Restriction fragments were end labelled using T4 polynucleotide kinase (P-L Biochemicals) and γ^{32}P-ATP after dephosphorylation with calf intestine alkaline phosphatase (Sigma). Assymmetrically labelled fragments were generated by redigestion with an appropriate second restriction enzyme and prepared by soaking out of gel slices after electrophoresis on thin polyacrylamide gels.

S1 mapping was carried out according to the protocol of Weaver and Weissman (15), Poly A^+ RNA was prepared from individual human pituitaries obtained at autopsy, and was the gift of Robert Hallewell. Nuclease S1 was obtained from Sigma.

RESULTS AND DISCUSSION

We have previously described the cloning of a 2.6 kb EcoRI fragment from human placental DNA which hybridizes to an hGH cDNA probe (5). Partial DNA sequence results (R. Hallewell, personal communication) show that this fragment, designated here 2.6-1 does not encode hGH, although it might encode a peptide very closely related to hGH (>95% homology). It is not known whether this potential peptide corresponds to any of the minor variant hGH peptide variants found in the pituitary (6). Several additional genomic clones were isolated using analogous approaches. One with a restriction map similar to 2.6-1, designated 2.6-3, was studied further by sequence analysis. The analysis of the other hGH related genes will be described elsewhere.

The strategy for sequence determination of 2.6-3 is indicated in figure 1. The hGH hybridizing 2.6 kb EcoRI fragment was subcloned into the plasmid pBR322 from the λ phage in which it was isolated, and the sequence of the region encoding hGH was determined by the Maxam and Gilbert technique (34). The restriction endonuclease sites used are shown in figure 1. Most of the sequence was determined on both strands and from more than one restriction site, and all sites used for end labelling were sequenced across, except the BglII downstream of IVSD.

The sequence obtained is presented in Figure 2. Comparison with the two independently derived cloned hGH cDNA sequences (4, 7) shows that 2.6-3 contains the entire hGH gene. However, there are two differences between the gene and the cDNA's in noncoding regions (see fig. 2). The gene and

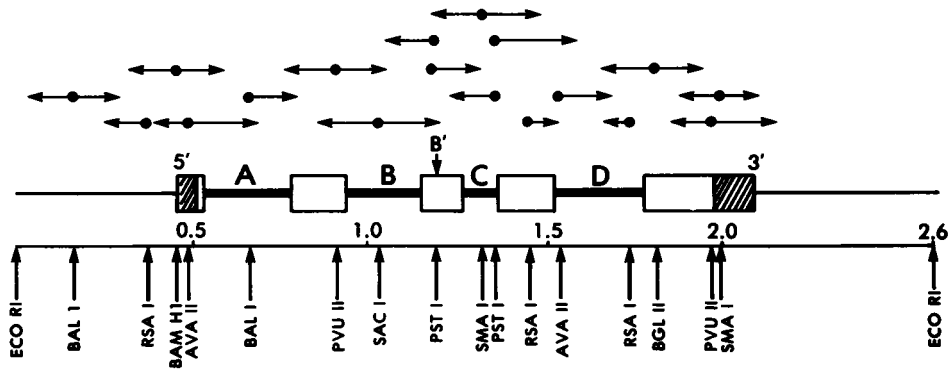


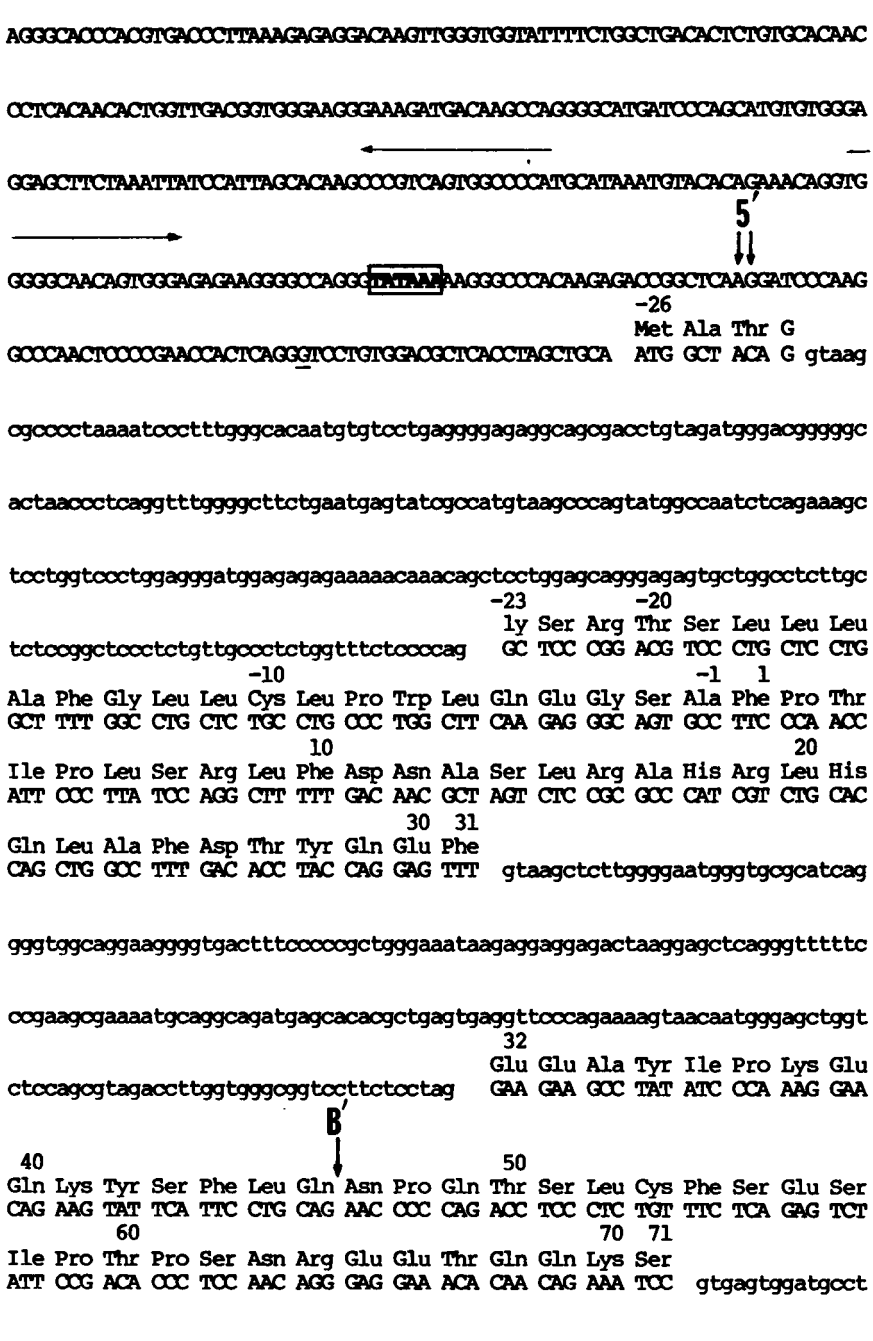
Fig. 1. Map of the hGH gene. Restriction sites labelled for DNA sequencing are indicated. The scale of the fragment is shown in kilobases (kb). Horizontal arrows show direction and approximate extent of sequence determination. Boxes indicate mRNA coding regions; 5' and 3' untranslated segments are stippled, and intervening sequences are thick lines. B' indicates the position of an alternative 3' end for IVS B.

one cDNA differ at one position in the 5' untranslated region. This difference is likely to be due to a reverse transcriptase artifact in the cDNA cloning, since such artifacts are relatively frequent at the end of reverse transcriptase copies of mRNA's and the 2.6-3 sequence is identical to the other cDNA sequence. Another difference between the gene and both cDNA's in the 3' untranslated region presumably reflects allelic variation. (The 2.6-3 fragment is not from the same individual as either of the cDNA clones.)

INTERVENING SEQUENCES

Comparison of the sequence of 2.6-3 with the sequences of the hGH cDNA's shows that the hGH gene is interrupted by four intervening sequences (IVS) of 256 (A), 209 (B), 93 (C), and 253 (D) bases, rather than three as previously estimated from restriction analysis (5). In agreement with the general rules deduced for such sequences, all the hGH intervening sequences could begin with GT and end with AG (8). As diagrammed in figure 3, the 5' and 3' boundaries of all four are also partially complementary to a portion of the the U1 RNA sequence, AGAGGGACGGUCCAUUCA, which has been proposed to be a guide for the proper removal of intervening sequences from nuclear mRNA precursors (9, 10).

The rat growth hormone (rGH) gene also has four intervening sequences


AGGGCACCCACGTGACCCTTAAAGAGAGGACAAGTTGGGTGGTATTTTCTGGCTGACACTCTGTGCACAAC
CCTCACAACACTGGTTGACGGTGGGAAGGGAAAGATGACAAGCCAGGGGCATGATCCCAGCATGTGTGGGA
GGAGCTTCTAAATTATCCATTAGCACAAGCCCGTCAGTGGCCCCATGCATAAATGTACACAGAAACAGGTG
5'
GGGGCAACAGTGGGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCGGCTCAAGGATCCCAAG
-26
Met Ala Thr G
GCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACGCTCACCTAGCTGCA ATG GCT ACA G gtaag
cgccccctaaaatcccctttgggcacaatgtgtcctgaggggagaggcagcgacctgtagatgggacgggggc
actaaccctcaggtttggggcttctgaatgagtatcgccatgtaagcccagtatggccaatctcagaaagc
tcctggtccctggagggatggagagagaaaaacaaacagctcctggagcagggagagtgctggcctcttgc
-23 -20
ly Ser Arg Thr Ser Leu Leu Leu
tctccggctccctctgttgccctctggtttctccccag GC TCC CGG ACG TCC CTG CTC CTG
-10 -1 1
Ala Phe Gly Leu Leu Cys Leu Pro Trp Leu Gln Glu Gly Ser Ala Phe Pro Thr
GCT TTT GGC CTG CTC TGC CTG CCC TGG CTT CAA GAG GGC AGT GCC TTC CCA ACC
10 20
Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Ser Leu Arg Ala His Arg Leu His
ATT CCC TTA TCC AGG CTT TTT GAC AAC GCT AGT CTC CGC GCC CAT CGT CTG CAC
30 31
Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
CAG CTG GCC TTT GAC ACC TAC CAG GAG TTT gtaagctcttggggaatgggtgcgcatcag
gggtggcaggaaggggtgactttcccccgctgggaaataagaggaggagactaaggagctcagggttttc
ccgaagcgaaaatgcaggcagatgagcacacgctgagtgaggttcccagaaaagtaacaatgggagctggt
32
Glu Glu Ala Tyr Ile Pro Lys Glu
ctccagcgtagaccttggtgggcggtccttctcctag GAA GAA GCC TAT ATC CCA AAG GAA
B'
40 50
Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser
CAG AAG TAT TCA TTC CTG CAG AAC CCC CAG ACC TCC CTC TGT TTC TCA GAG TCT
60 70 71
Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser
ATT CCG ACA CCC TCC AAC AGG GAG GAA ACA CAA CAG AAA TCC gtgagtggatgcct
tgaccccaggcggggatgggggagacctgtagtcagagccccgggcagcacaggccaatgcccgtccttc

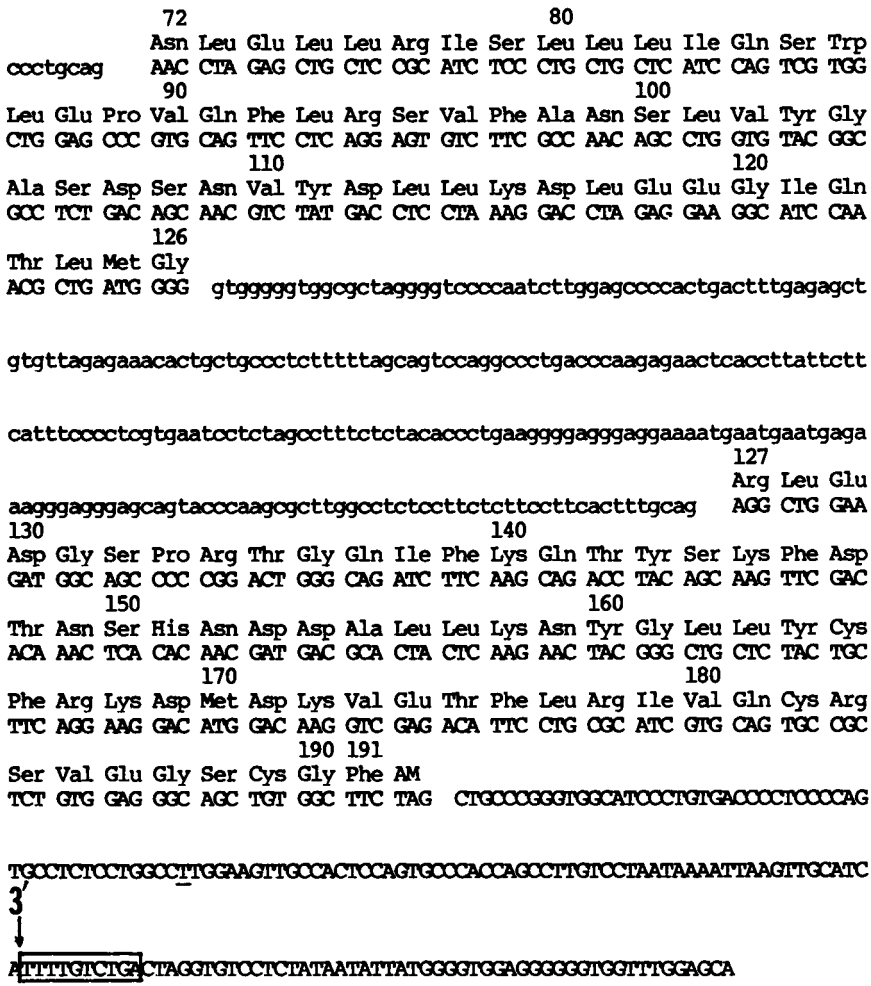


Fig. 2. DNA sequence of the hGH gene. Predicted hGH amino acid sequence is shown. B' indicates a potential alternative 3' splice site for intervening sequence B. Horizontal arrows indicate regions of partial dyad symmetry near the mRNA 5' end. The positions of the 5' end of the mRNA as determined by S1 mapping and the poly A addition site are indicated by vertical arrows. The "TATA box" sequence near the 5' end and a common sequence near the 3' end are boxed. Two differences from the reported cDNA sequences are underlined.

located in the same positions as those in the hGH gene when two three base pair spaces are put in the rat sequence to maximize homology (11). This is strong confirmation for the supposition that the two genes evolved from a

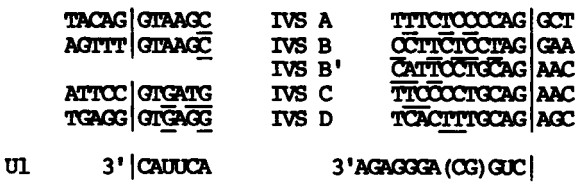


Fig. 3. Comparison of the hGH intervening sequence junctions with the complement of the sequence of a portion of the rat U1 RNA. Vertical lines indicate the presumed splice sites. IVS B' is a potential alternative 3' splice site for IVS B. Differences from the complement of the U1 RNA sequences are underlined.

common precursor.

Several examples of alternate splicing patterns for the removal of intervening sequences from a single nuclear mRNA precursor are known. For instance, changes in the splicing of the mouse immunoglobulin heavy chain precursor result in the synthesis of IgM molecules with different biological activities (12,13). The DNA sequence of the hGH gene was therefore searched for other potential splice sites. We constructed arbitrary canonical splice sites by combining the consensus sequences (9, 10, 14) derived for splice sites with the sequence homologous to the appropriate segment of U1 RNA. 5' sites were defined as sequences with three or fewer mismatches from AGGTAAGT, not including mismatches with the highly conserved, underlined GT. 3' sites were defined as sequences with six or fewer mismatches from TNTTCTCCCTNCAGG, not including mismatches with the conserved, underlined AG and with no AG dinucleotide less than fourteen bases upstream from the conserved AG (14). All of the correct splice sites as well as one additional 5' site and thirteen additional 3' sites were found by this search.

S1 mapping was used to locate regions of non-homology between the cloned hGH cDNA and total pituitary mRNA which would result from alternative splicing of a nuclear precursor. A cDNA probe fragment was prepared which was end-labelled at the RsaI site between IVS C and IVS D and which extended to the 5' untranslated region, covering slightly more than 50% of the mRNA. An excess of this probe fragment was hybridized to poly A^+ pituitary RNA under the conditions of Weaver and Weissman (15), and the resultant mRNA-DNA hybrids were digested with the single strand specific nuclease S1. As shown in Figure 4 (lane B), two different mRNA's were detected. As expected, the major fragment was the same size as the undigested probe fragment (lane A), and therefore results from mRNA identi-

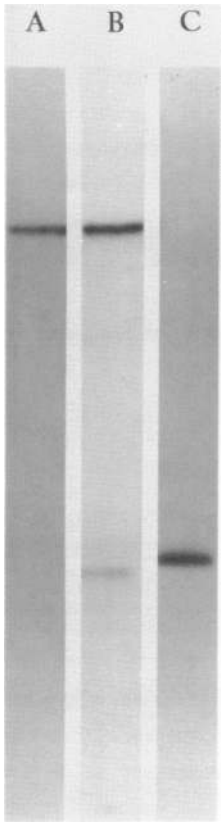


Fig. 4. S1 mapping of hGH mRNA. Total pituitary poly A^+ RNA was annealed with RsaI end labelled probe fragment from the cloned cDNA, and the mRNA-DNA duplexes were digested with S1. Products were separated by electrophoresis on an 8% polyacrylamide sequencing gel and visualized by autoradiography. A) RsaI probe fragment; no RNA or S1 digestion. B) RsaI probe fragment plus pituitary RNA and S1. C) Marker RsaI-PstI fragment.

cal to the cDNA. The smaller protected fragment present at 10-20% of the level of the full length fragment must result from a divergent mRNA. This fragment is slightly smaller than a marker fragment prepared by digestion of the probe fragment with PstI (lane C). This locates the point of divergence from the cDNA precisely at a potential 3' splice site which happens to contain the PstI recognition site CTGCAG (see Fig 3) 45 nucleotides downstream from the 3' end of IVS B. (Such a protected fragment would be shorter than the marker fragment because the PstI cleavage site is five nucleotides upstream from the expected splice site.) An internally deleted hGH peptide with fifteen missing amino acids comprises 10% of the GH in normal pituitaries (16). Since the deleted amino acids correspond precisely to the sequence from the 3' end of IVS B to this alternative splice site, we conclude that it is likely that in normal pituitaries IVS B has two alternative 3' sites. However, it is also possible that the deleted

peptide is the product of another hGH gene virtually identical to the sequenced 2.6-3 gene.

Some evidence has been obtained which suggests that the smaller hGH peptide has somewhat different biological activities from the major hGH (16), but the physiological significance of this is not clear. This difference in activities may reflect division of hGH into functional domains defined by the positions of the intervening sequences, as is the case in the mouse immunoglobulin heavy chain genes (17), and perhaps the β-type globins (18).

One functional domain in secreted proteins such as hGH is the presequence or signal peptide, which is required during the process of secretion. IVS A is located within the hGH presequence, partially separating this region from the remainder of the gene. It is striking that all of the genes encoding secreted proteins sequenced to date, including hGH, human (19) and rat (20,21) insulin; the α-subunit of the human glycoprotein hormones (22), mouse immunoglobulins (23), and chicken ovalbumin (24, 25) and conalbumin (26), have an intervening sequence within or just preceding the presequence. (Ovalbumin, which does not have a presequence, has an intervening sequence 18 nucleotides from the translational start). Of the several non-viral, non-secreted genes studied so far, only mouse dihydrofolate reductase (27) and yeast actin (28, 29) also have a similarly placed intervening sequence. The hypothesis that intervening sequences serve in evolution to allow combinations of functional domains (31) suggests that such intervening sequences may separate a functional domain contained within the 5' end of the mRNA from the body of the coding region. Such an mRNA domain could serve to direct the message for secretary proteins to the appropriate membrane bound ribosomes, for example. It is perhaps interesting in this respect that there is an area of striking homology (19 of 26 bases identical) in the 5' untranslated regions of the messages for hGH and the otherwise unrelated α subunit of the human glycoprotein hormones. This homology is shown in Figure 5. Alternatively, these common 5' intervening sequences could serve to separate the structural gene from the promoter. This might allow greater flexibility in regulation of expression and perhaps more rapid evolution of these two functional domains.

hGH mRNA 5' END

The position of the 5' end of the mature hGH mRNA within the sequence was determined directly using the S1 nuclease mapping technique. Poly A^+

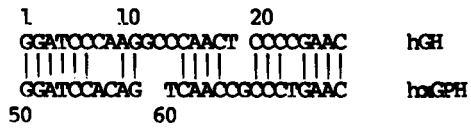


Fig. 5. Comparison of the 5' untranslated regions of human peptide hormone mRNA's. Partial sequences from the 5' untranslated regions of the mRNA's are shown. Numbers indicate estimated number of nucleotides from the 5' end of the mRNA. The hGH sequence is from this paper and the sequence for the α-subunit of the human glycoprotein hormone is from Fiddes and Goodman (22).

pituitary mRNA was hybridized to a 5' end-labelled probe extending from an AvaII site in the 5' untranslated region to an EcoRI site approximately 400 base pairs upstream. The length of the DNA fragments protected by the RNA from nuclease digestion was determined precisely by comparison with a partial DNA sequence ladder generated from the end labelled probe fragment (Fig. 6). The length of the protected fragments was dependent on the temperature of the S1 digestion, increasing in length by one nucleotide as the temperature dropped from 30°C to 17°C. Presumably this is due to steric hindrance by the mRNA cap. The protected fragments were also heterogeneous in size at all temperatures varying by approximately four bases. It is not known whether this is also a result of interference by the mRNA cap or whether it reflects some heterogeneity of 5' ends. The position determined for the 5' end(s) is indicated by vertical arrows in Figure 2. The position of the 5' end determined by S1 mapping agrees precisely with that estimated by reverse transcriptase extension of a specific end labelled probe (7).

5' FLANKING REGION

A sequence related to TATAAAA is found approximately 25 base pairs upstream from the start of transcription in nearly all genes transcribed by RNA polymerase II (31). That sequence is found upstream of the estimated start of the hGH mRNA (boxed in Figure 2). The related sequence CATAAAT is also found 54 nucleotides upstream. It is not known whether under certain conditions there may also be a secondary transcription start near this sequence but the S1 mapping shows that the level of such initiations must be less than 10% of the primary starts in adult pituitaries. In either case, translation should start at the first AUG of the transcript, which is the start expected from the amino acid sequence of hGH.

It has been pointed out that some eukaryotic promoters have a sequence

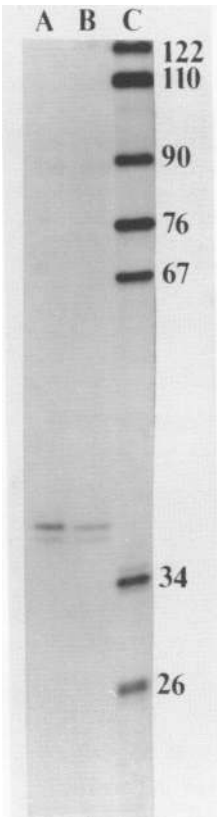


Fig. 6. S1 mapping of the 5' end of the hGH mRNA. Total polyA+ RNA was annealed with an AvaII to EcoRI subfragment from the 2.6-3 fragment and digested with S1 at 24°. The protected fragments were separated by electrophoresis on a 10% polyacrylamide sequencing gel and visualized by autoradiography (lanes A and B. HinfI fragments of ØX174 served as size markers (lane C).

similar to GG(CT)CAATCT approximately 40 base pairs upstream from the TATAAAA sequence (32). There is not a strong homology with that model sequence at the analogous position in the hGH promoter, although the sequence at that position, GCCCCATGC, matches the consensus at five of nine positions.

Several eukaryotic promoters, such as those of human insulin and mouse α-globin, have regions of dyad symmetry upstream of the transcriptional start (19, 32). Such structures in procaryotic promoters are often binding sites for regulatory proteins. A large imperfect palindrome, found upstream of the hGH transcriptional start is indicated by horizontal arrows in Figure 2.

3' END

Comparison of the sequence of the hGH gene with the cloned hGH cDNA shows that the hGH mRNA ends at one of the two positions indicated by vert-

ical arrows in Figure 2, depending on whether or not the A in the gene sequence is a part of the post transcriptionally added poly A tail. As previously noted for the hGH cDNA clone, the highly conserved mRNA sequence AAUAAA (33) is found upstream of that position. Another conserved sequence similar to TTTTCACTGC is present in some but not all genes at or just downstream of the 3' end of the transcipt (32). The related sequence TTTTGTCTGA is found in hGH just downstream of the 3' end. Although the functional role, if any, of these two conserved sequences is not known, their location at the end of the gene suggests involvement in poly A addition and/or transcription termination.

SUMMARY

We have determined the DNA sequence of the human growth hormone gene. Comparison of the gene sequence to the published cDNA sequence shows that the gene is interrupted by four intervening sequences. Searching the gene sequence for additional splice sites using a combination of published consensus sequences and homology to rat U1 RNA reveals a number of potential alternative splicing sites. Splicing at one alternative 3' site located 45 base pairs downstream from the normal 3' site of intervening sequence B is likely to generate the mRNA for one smaller hGH peptide found at low levels in normal pituitaries. The position of the 5' end of the mature mRNA was located using S1 mapping. The signal sequence TATAAA, thought to be involved in initiation of transcription (29), is found approximately 25 base pairs upstream from the 5' end of the mature mRNA. Conserved sequences are also found in the gene near the sequence where poly A is added to the mRNA. Surprisingly, there is a region of homology in the 5' untranslated regions of the mRNA's of hGH and the otherwise unrelated α subunit of the human glycoprotein hormones.

ACKNOWLEDGEMENTS

Funding for this research was supplied by NIH grants CA14026 and AM19997. HMG is an Investigator of the Howard Hughes Medical Institute.

REFERENCES

1. Daughaday, W.H., Herrington, A.C. and Phillips, L.S. (1975) Ann. Rev. Physiol. 35, 211-244
2. Niall, H.D., Hogan, M.L. Sauer, R., Rosenblum, I.Y. and Greenwood, F.C. (1971) Proc.Natl.Acad.Sci. USA 68, 866
3. Shine, J., Seeburg, P.H., Martial, J.A., Baxter, J.D. and Goodman, H.M. (1977) Nature 270, 494-499
4. Martial, J.A., Hallewell, R.A., Baxter, J.D., and Goodman, H.M.

(1978) Science 205, 602
5. Fiddes, J.C., Seeburg, P.H., DeNoto, F.M., Hallewell, R.A., Baxter, J.D., and Goodman, H.M. (1978) Proc. Natl. Acad. Sci. USA 76, 4294
6. Lewis, V.J., Singh, R.N.P., Bonewald, L., Lewis, L., and Vanderlaan, W. (1979) Endrocrinology 104, 1256-1262
7. Roskam, W.G. and Rougeon, R. (1979) Nucl. Acids. Res. 7, 305-320
8. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4853-4857
9. Lerner, M., Boyle, J., Mount, S., Wolin, S. and Steitz, J (1980) Nature 283, 220-224
10. Rogers, J., and Wall, R. (1980) Proc. Natl. Sci. USA 77, 1877-1879
11. Page, G.S., Smith, S.L., and Goodman, H.M. Nucl. Acids. Res. (in press)
12. Early, P., Rogers, J., Davis, M., Calami, K., Bond, M., Wall, R., and Hood, L. (1980) Cell 20, 313-319
13. Alt., F., Bothwell, A., Knapp, M., Siden, E., Mather, E., Koshland, M., and Baltimore, D. (1980) Cell 20, 293-301
14. Seif, I., Khoury, G., and Shar, R. (1979) Nucleic Acids Res. 6, 3387-3398
15. Weaver, R. and Weissman, C. (1979) Nucleic Acids Res. 7, 1175-1193
16. Lewis, V.J., Bonewald, I.F., and Lewis, L.J. (1980) Biochem. Biophys. Res. Comm. 92, 511-518
17. Sakano, H., Rogers, J.H., Huppi, K., Brach, C., Traunecker, A., Maki, R., Wall, R., and Tonegawa, S. (1978) Nature 277, 629-633
18. Craik, C., Buchman, S., and Beychok, S. (1980) Proc. Natl. Acad. Sci. USA 77, 1384-1388
19. Bell, G.I., Pictet, R.L., Rutter, W.J., Cordell, B., Tischer, E., and Goodman, H.M. (1980) Nature 284, 26-32
20. Cordell, B., Bell, G., Tischer, E., DeNoto, F., Ullrich, A., Pictet, R., Rutter, W.J., and Goodman, H.M. (1979) Cell 18, 533-543
21. Lomedico, P., Rosenthal, N., Efstradiatis, A., Gilbert, W., Kolodner, R., and Tizard, R. (1979) Cell 18, 545-558
22. Fiddes, J.C., and Goodman, H.M. J. Molec. App. Gen. (in press)
23. Tonegawa, S., Maxam, A., Tizard, R., Bernard, O., and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA 75, 1485-1489
24. Dugaiczyk, A., Woo, S., Colbert, D., Lai, E., Mace, M., and O'Malley, B. (1979) Proc. Natl. Acad. Sci. USA 76, 2253-2257
25. Cochet, M., Gannon, F., Hen, R., Maroteau, L., Perrin, F., and Chambon, P. (1979) Nature 282, 567-574
26. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perin, F. and Chambon, P. (1979) Nature 282, 567-574
27. Nunberg, J., Kaufman, R., Chang, A., Cohen, S., and Schimke, R. (1980) Cell 19, 355-364
28. Gallwitz, D., and Sures, I. (1980) Proc.Natl. Acad. Sci. USA 77, 2546-2550
29. Ng, R., and Abelson, J. (1980) Proc.Natl. Acad. Sci. USA 77, 3912-3916
30. Gilbert, W. (1978) Nature 271, 501
31. Goldberg, M. (1979) Thesis, Stanford University
32. Benoist, C., O'Hare, K., Chambon, P. (1980) Nucl. Acids Res. 8, 127-142
33. Nishioka, Y., and Leder, P. (1979) Cell 18, 875-882
34. Proudfoot, N.J., and Brownlee, G.G. (1976) Nature 263, 211-214
35. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology Grossman, L., and Moldave, K., Eds., Vol. 65, pp. 499-560 Academic Press, New York.