*Proc. Natl. Acad. Sci. USA*
Vol. 78, No. 8, pp. 4867–4871, August 1981
Biochemistry

# Primary structure and evolution of rat growth hormone gene

(gene organization/introns/repetitive sequences/gene evolution)

ANDREA BARTA*, ROBERT I. RICHARDS†, JOHN D. BAXTER*, AND JOHN SHINE†

†Genetics Department, Research School of Biological Sciences, Australian National University, Canberra, A.C.T., Australia; and *The Howard Hughes Medical Institute Laboratories; Department of Medicine, Metabolic Research Unit, and Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

*Communicated by Walter Gilbert, May 6, 1981*

**ABSTRACT** **The rat growth hormone gene was isolated on a cloned 11.4-kilobase *Eco*RI-generated DNA fragment from a bacteriophage "library" of chromosomal DNA. The structural gene sequence, ≈2.1 kilobases long, was identified by hybridization to the corresponding cloned rat growth hormone cDNA and shown to contain four intervening sequences. The complete primary structure of the gene and the 5′- and 3′-flanking regions was determined. The mosaic structure of exons and introns can be related to the different biological activities of growth hormone and to the evolution from ancestral sequences of a gene that was the precursor to the growth hormone and the related prolactin and placental lactogen (chorionic somatomammotropin) genes. The largest intron was found to contain a dispersed repetitive DNA sequence flanked by perfect 18-base pair direct repeats. The mobility of sequences of this kind could play a role in the observed variation of intron sizes and in rearrangements of mammalian genes.**

Growth hormone (GH) is a member of a set of structurally related peptide hormones that includes prolactin (Prl) and placental lactogen (PL; chorionic somatomammotropin) (1, 2). Analysis of the primary structures of the proteins and of cDNAs to each of their mRNAs suggests that these hormone genes evolved from a common precursor (3–6). There are also regions of internal homology within each protein suggesting that the precursor of this gene family was formed by duplications of a smaller ancestral gene (3). Human (h) GH has been found to be a mixture of slightly different proteins (7). Although some of the heterogeneities may be explained by posttranslational modifications, others may reflect variant genes or differences in processing of the same mRNA (8). Thus, knowledge of the primary GH gene structure may provide information about the evolution of this family of genes and about the mechanisms by which several hormone forms can be produced. The GH gene is expressed in the pituitary and in cultured rat pituitary tumor cells (2). This expression is regulated by hormones and other factors (2). Thus, the gene is also of interest to study hormone action. Knowledge of its primary structure should facilitate such investigations. This paper presents the complete nucleotide sequence of the cloned rat (r) GH gene isolated from a chromosomal DNA library.

## MATERIALS AND METHODS

**Hybridization Probes and Screening of the Rat DNA Library.** The rat chromosomal DNA library, which has been described by Sargent *et al.* (9) and was provided by that group, was composed of a partial *Eco*RI digest of rat liver DNA cloned into the bacteriophage λ Charon 4A. Screening of the library was carried out as described (10). The rGH cDNA fragment isolated from the recombinant plasmid pRGH-1 (5) was used as hybridization probe after radioactive labeling with random DNA primers (11). Nitrocellulose filters were pretreated and hybridized as described (12). Positive plaques were purified by two further cycles of plating at much lower densities. λ Charon 4A phage were grown and purifed as described (13).

**DNA Sequence Analysis.** DNA sequence analysis was carried out by either the chemical degradation method (14) or the chain termination method (15). In the latter case, fragments of the cloned DNA were subcloned into the bacteriophage M13 mp7 (16) and analyzed by using a synthetic primer complementary to the codons for amino acids 13–17 of the β-galactosidase gene (kindly provided by R. Crea).

**Biosafety Conditions.** All experiments involving recombinant plasmids and phages were carried out under containment conditions recommended by ASCORD (Australia) or National Institutes of Health guidelines for recombinant DNA research.

## RESULTS AND DISCUSSION

**Isolation of Cloned rGH Sequences.** A cloned rGH cDNA probe was used to screen the rat chromosomal DNA library. Of the $8 \times 10^5$ recombinant plaques screened, five clones hybridized with the $^{32}$P-labeled cloned cDNA. Restriction endonuclease analysis of DNA isolated from these clones showed that they fall into two groups. The first group contains both an 11.4-kilobase (kb) and a 1.5-kb *Eco*RI fragment. The 11.4-kb fragment hybridizes strongly with the rGH cDNA and contains the GH gene. The GH gene on an 11.4-kb *Eco*RI fragment was also found in DNA from rat liver and cultured pituitary ($GH_3$) cells. This fragment appears to be identical to that characterized by heteroduplex and restriction endonuclease mapping by Chien and Thompson (17). The second type of clone contains four *Eco*RI fragments (1.7, 2.1, 3.6, and 4 kb). Only the 4-kb fragment hybridizes (albeit only weakly) with the rGH cDNA probe, and it probably does not contain the rGH gene as no additional *Eco*RI fragment of this size could be detected on autoradiography of the genomic blots. Digestion of this type of clone with several other restriction endonucleases also gave rise to a completely different pattern of fragments than that found in the rGH gene (Fig. 1). The nature of these sequences is unknown; they could conceivably encode the closely related PL gene or a GH "pseudogene."

One clone (λrGH-13) carrying the rGH gene was analyzed further. Digestion of the 11.4-kb *Eco*RI fragment of λrGH-13 with *Hin*dIII gives rise to a 5.8-kb fragment that hybridizes to the cloned rGH cDNA. This fragment (prGHeh 5.8) was subcloned into the *Eco*RI/*Hin*dIII site of pBR322 for further char-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: GH, growth hormone; kb, kilobase(s); rGH and hGH, rat and human growth hormones, respectively; PL, placental lactogen; Prl, prolactin; bp, base pair(s).

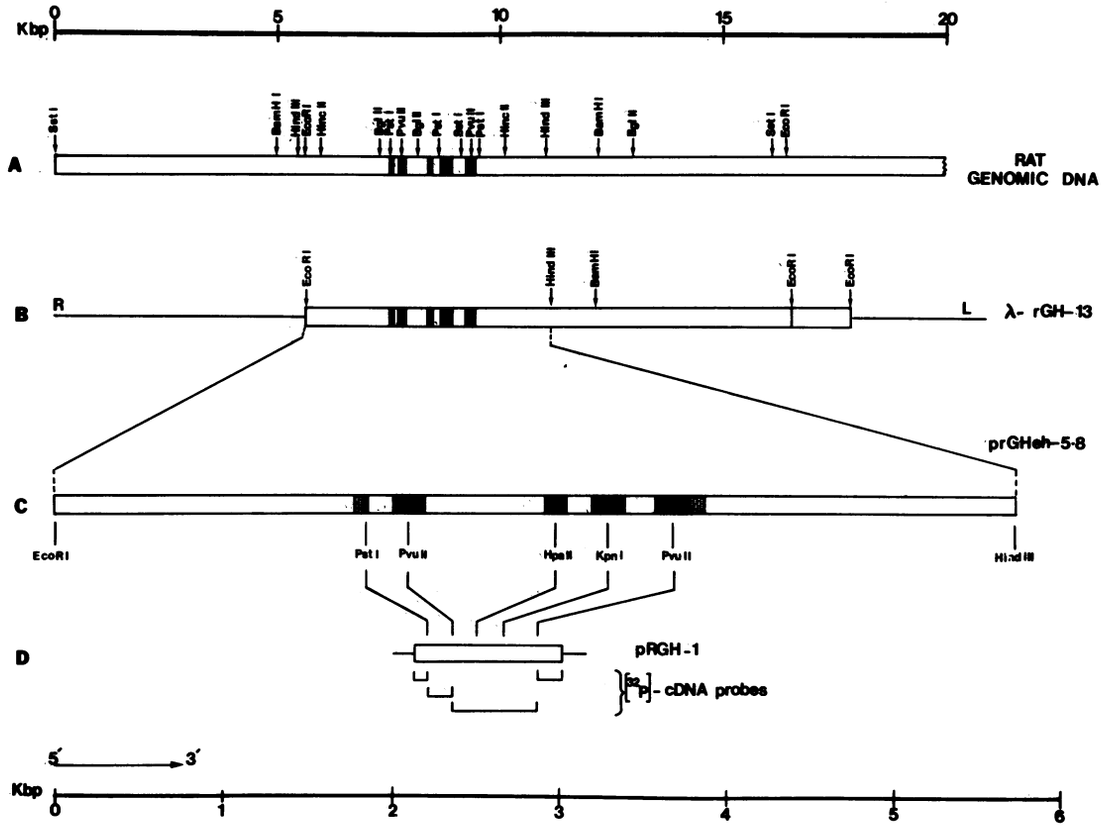


FIG. 1. Map of the rGH gene. (*A*) Restriction endonuclease map of GH gene derived from rat liver DNA. (*B*) Schematic of the λ-rGH-13 clone containing two *Eco*RI fragments; the 11.4-kb fragment comprised the rGH gene. R and L, right and left arms, respectively, of λ Charon 4A. (*C*) Enlargement of the 5.8-kb *Eco*RI/*Hin*dIII fragment, which contained the structural GH sequences and was subcloned in pBR322 (prGHeh 5.8). (*D*) Fragments of pRGH-1 used for establishing the restriction map. ■, coding regions; ▩, 5'- and 3'-untranslated regions. For clarity, only a few restriction endonuclease sites are shown in *B* and *C*. kbp, kilobase pairs.

acterization. The structure of this fragment and rGH cDNA is compared with the restriction endonuclease sites in λrGH-13 in Fig. 1. This comparison shows at least four intervening sequences interrupting the rGH coding region. The arrangement of introns and exons in this gene was determined by using specific fragments of the cloned cDNA to hybridize with restriction digests of prGHeh 5.8 that had been fractionated by agarose gel electrophoresis and transferred to nitrocellulose filters (18). Similar digestions and hybridizations were carried out on DNA isolated from rat liver and $GH_3$ cells; by these criteria, the structural organization of the cloned rGH gene is identical to that found in normal liver and is indistinguishable from that in cultured pituitary tumor cells in which the gene is expressed (Fig. 1).

**Primary Structure of GH Gene.** The complete nucleotide sequence of the gene (2088 bp) and adjacent regions (400 bp at the 5' and 70 bp at the 3' end of the gene) in prGHeh 5.8 was determined (Fig. 2). All restriction sites previously determined in the gene and the surrounding regions from mapping data were confirmed by this analysis. The first intron of the gene ($I_A$) interrupts the coding sequence at codon −23 (where it splits the GAC that encodes aspartate). The other introns, $I_B$, $I_C$, and $I_D$ occur between codons 31/32, 70/71, and 124/125, respectively.

The GH mRNA sequence contained in the five exons agrees completely with that determined from the cDNA sequence except that the codon for the first amino acid of the mature protein was found to be that for phenylalanine (UUC), as originally predicted from protein-sequence data (4), instead of that for leucine (UUA) found in rGH cDNA and that an extra ribosylthymine residue occurs in the cDNA clone between position 2075 and 2076 in the 3'-untranslated region.

**Putative Promoter Region of rGH Gene.** A feature common to eukaryotic genes is the Goldberg–Hogness box found 27–33 nucleotides to the 5' side of the initiation point of mRNA transcription (19). In the rGH gene such a sequence (5' T-A-T-A-A-A) occurs in the anticipated position (−30) with respect to the putative mRNA cap site. We assigned nucleotide 1 as the putative cap site as, in preliminary studies, reverse transcription of GH mRNA (from cultured rat pituitary $GH_3$ cells) primed with a rGH cDNA fragment terminated at the nucleotide at position 1 (data not shown).

If GH mRNA transcription initiates as indicated in Fig. 2, then the largest RNA precursor should be ≈2100 nucleotides. Although mRNA precursors of this size have been identified, much larger precursor forms have also been reported (20) and identified by G. Cathala (personal communication). The nature of such large mRNA precursors is currently unkown.

The 5' end of the gene is of interest in terms of possible structures that may be important in regulating its expression. Apart from the Goldberg–Hogness box, the region around nucleotide −80 has been found to be conserved in other mammalian gene families (21, 22). Although the structure of other GH (or related) genes has not been published, a partial 5' sequence of one of the hGH genes made available to us by P. H. Seeburg (personal communication) shows that the sequence between nucleotides −88 and −74—C-C-A-T-G-A-A-T-A-A-A-T-G-T-A—is completely conserved in the human gene except for a cytidine substitution at position −83. Therefore, the question must be raised of whether this structure has importance for GH gene

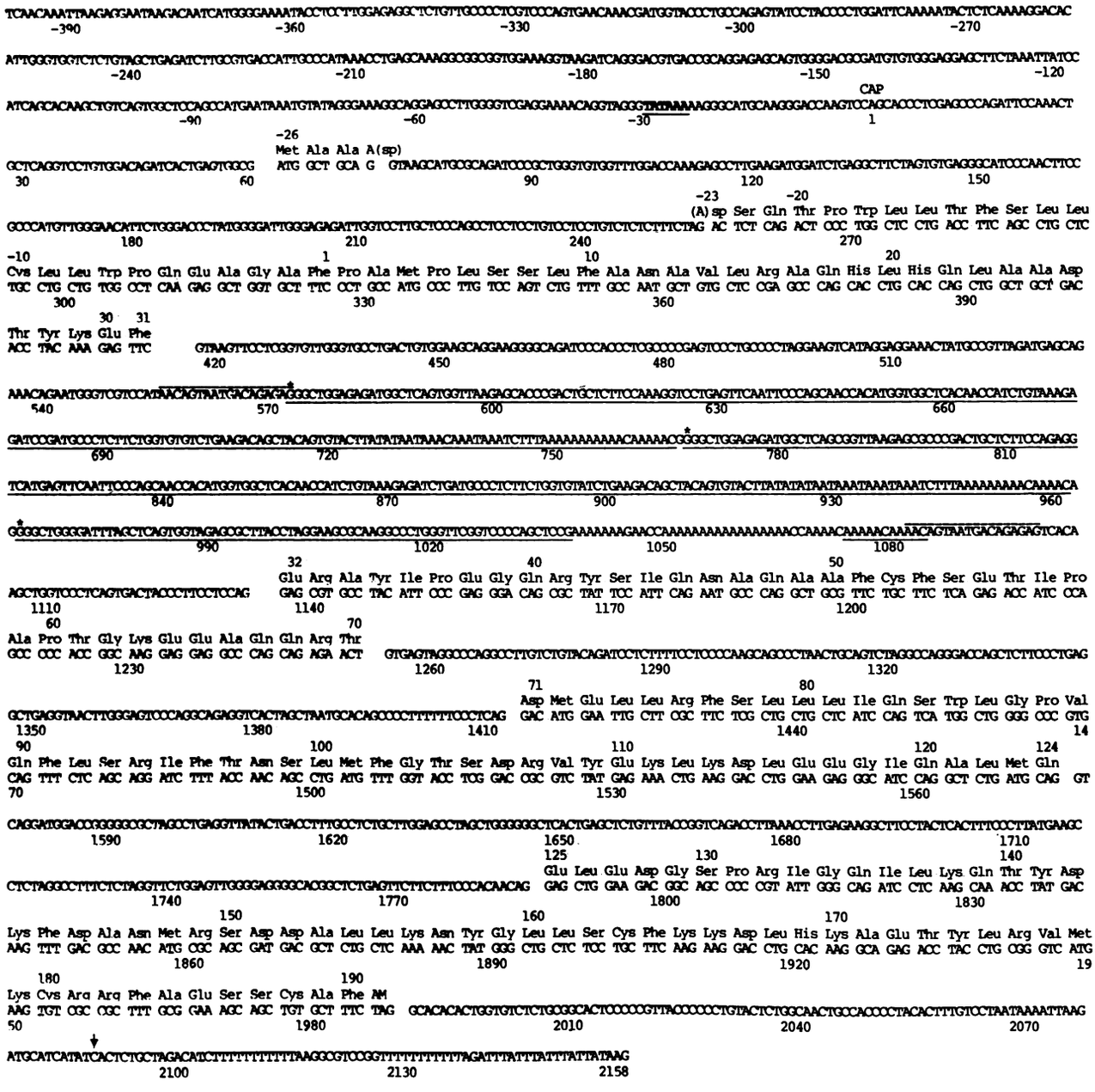

FIG. 2. Sequence of the rGH gene and adjacent regions. Translation into the amino acid sequence of rat pre-GH is shown; amino acid 1 indicates the first amino acid of the mature protein and amino acids −26 to −1 represent the signal peptide sequence. The Goldberg–Hogness sequence T-A-T-A-A-A starting at position −30 is underlined. Nucleotide +1 represents the putative 5′ end of the mRNA (capping site) and is designated "CAP." The poly(A) addition site is indicated by an arrow. In intron B (nucleotides 417–1133), the three long repeated sequences are underlined; asterisks indicate the start of each repeat. The 18-bp direct repeats flanking these long repeated structures are overlined.

expression. By contrast, comparison of the GH gene structure with that of the ovalbumin gene (also regulated by steroids) did not show obvious similarities.

**Termination Region of rGH Gene.** As determined previously from cloned rGH cDNA, the 3′-untranslated region of the mRNA is 100 nucleotides long and contains the characteristic A-A-U-A-A-A sequence (23). This portion of the mRNA is contiguous in the gene and is immediately followed by a region rich in ribosylthymine residues with two stretches of poly(T)$_{11}$. Similar structures have been observed in the 3′-flanking region of some, but not all, eukaryotic genes (24).

**Evolution of GH Gene Family—Structure–Function Relationships.** It is of interest to compare the GH gene with those of the related PL and Prl genes. Although complete sequence data for these other genes have not been reported, some comparisons can be made from heteroduplex mapping data (17), partial sequence data of the rPrl gene (25), and the unpublished observations of N. Cooke (rPrl) and P. H. Seeburg (hGH). Like the rGH gene, the hGH and hPrl genes both contain four introns, although the introns of the Prl gene are larger than those of the GH genes. Further, the exon-intron junctions in all three genes occur at identical locations when the coding regions are aligned to maximize amino acid homology (6). The similar organization of these genes is thus consistent with the hypothesis that they have a common evolutionary origin. It is of interest to compare regions of internal homology previously observed in GH, Prl, and PL (3) to the location of the exons (Fig. 3). As three of these homologous regions are separated by introns, it

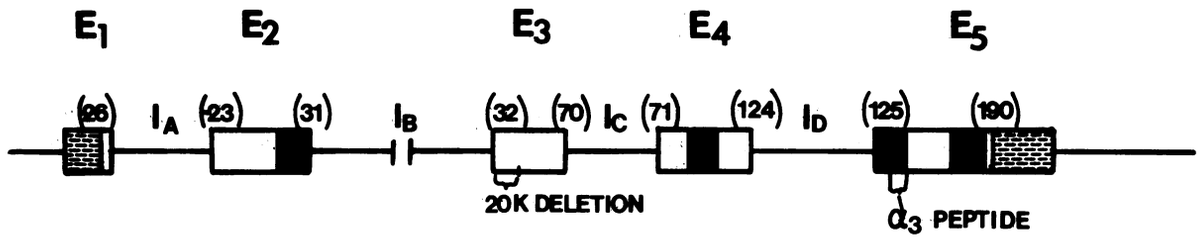


FIG. 3. Schematic of the structural organization of the rGH gene. Exons ($E_1$–$E_5$) are shown as boxes and introns and flanking regions are shown as bars. 20K refers to a deletion form of hGH that omits amino acids 32–46. $\alpha_3$ peptide, site (positions 136–145) at which hGH is proteolytically cleaved to a form that shows enhanced growth-promoting activity; ▩, untranslated gene sequences; ■, homologous sequences; numbers in parentheses represent amino acid numbers.

is conceivable that the gene did evolve by duplications after which an intron separating the third and fourth regions of homology in exon 5 was removed (Fig. 4). The absence of a homologous region in exon 3 suggests that it was added after duplication of the ancestral sequence, perhaps with the addition of a new biological function.

The physiological actions of GH can be divided into growth-promoting and "anti-insulin-like" (diabetogenic). Several proteolytically modified forms of hGH have been examined (7). A peptide (amino acids 44–77) containing mainly amino acids coded by exon 3 (amino acids 32–70) has anti-insulin-like but not growth-promoting activity (26). Thus, this activity may be contained completely within exon 3. This idea is further supported by the demonstration that a variant form of hGH (20K variant) lacking amino acids 32–46 (coded for by exon 3) does not possess certain anti-insulin-like but does have growth-promoting activity (Fig. 3; ref. 7).

Proteolytic cleavage of GH can remove amino acids 135–146 (Fig. 3) to yield a two-chain structure connected by a disulfide bridge. This $\alpha_3$ peptide shows enhanced growth-promoting activity although destruction of the cysteine-53 (exon 3) — cysteine 165 (exon 5) bond abolishes this activity (7). Thus, although sequences contained in exon 3 appear to be responsible for diabetogenic activity, some amino acids encoded in this exon are involved, along with some other portion of the molecule, in maintenance of the structure needed for growth-promoting activity. Exon 3 might therefore have been used initially to generate the tertiary strucure required for the activity of the product of the gene that was the precursor of GH and related genes.

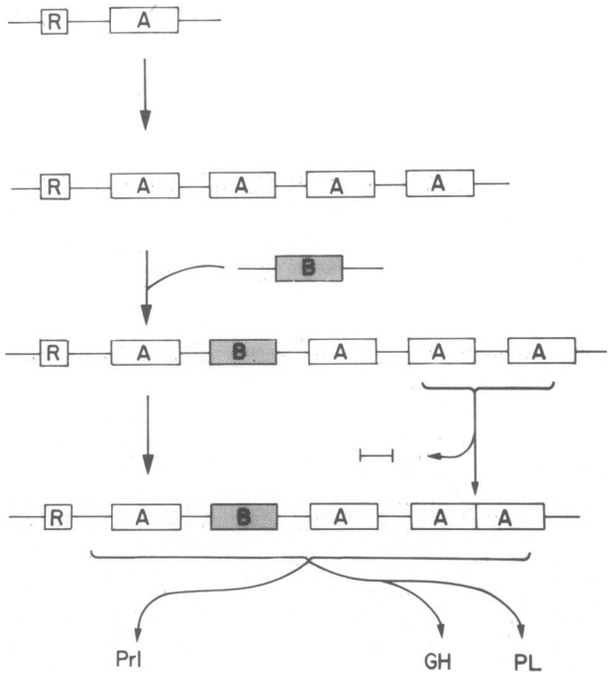


FIG. 4. Model for the evolution of GR, Prl, and PL genes. R, regulatory sequence; A and B, coding sequences.

The possible sequence of events in evolution of the GH gene family shown in Fig. 4 suggests that the most primitive gene may have included regulatory sequences (R) required for expression, as well as coding sequences (A) that later were duplicated. Exon R carries the promoter region as well as the sequences for the 5′-untranslated region of the mRNA. The presence of an intron separating a "regulatory exon" at the 5′ end of the gene from the functional domains of the coding portion is also seen in other mammalian genes (22, 24, 27, 28). Additional coding sequences (B) were added later. The intron separating the last two homologous regions of coding sequences may have been removed. Genes that have this overall structure appear to be the precursor for the entire GH gene family. Based on the cDNA (4–6) and protein structural analysis (3), it appears that the first duplication of this precursor led to one gene that became the Prl gene and a second one that served as the precursor to both GH and PL.

**Structure of Introns in the rGH Gene.** The border sequences of all four introns in the rGH gene are in good agreement with the consensus sequence for such regions (29). As with intron-exon junctions of other eukaryotic genes, there are ambiguities in the precise splicing points, but each intron begins with G-T at the 5′ terminus and ends with A-G at the 3′ terminus. Noteworthy is that the human 20K variant protein (lacking amino acids 32–46; Fig. 3) does not arise by proteolytic cleavage. Instead, from a previous partial sequence of the hGH gene (8), it appears that a second RNA splicing point in exon 3 can account for removal of the codons on the 5′ portion of the exon and generation of the 20K form. Such a potential alternative RNA splicing point ending in A-G and containing other nucleotides consistent with a consensus sequence (29) for RNA splicing is also found in the rat gene, although whether or not RNA splicing at this alternative point occurs in the rat is not known.

Three of the rGH gene introns are relatively small and similar in size ($I_A$, 194 bp; $I_C$, 160 bp; $I_D$, 213 bp) while the second is much larger ($I_B$, 718 bp). Intron B also has a lower G-C content (45.5%) than the other introns (56%). This results from three long A-T-rich regions in $I_B$.

$I_B$ also exhibits some interesting structural features. It contains a 195-bp tandem repeat; these repeats have 96% homology (eight nucleotide substitutions and three insertions/deletions). This structure is immediately followed by a 73-bp fragment homologous (66%) to the first part of the 195-bp repeat (Fig. 2). Each repeat ends with a long poly(A) stretch. The 508-bp segment of DNA containing the three long repeats is bounded by perfect 18-bp repeats (Fig. 2). The sequence of this repeat, 5′ A-A-C-A-G-T-A-A-T-G-A-C-A-G-A-G-A-G, itself

contains a palindromic symmetry with the center between the 7th and 8th residues.

The overall structure of this 508-bp DNA segment resembles that of dispersed repeated sequences in mammalian DNA (30–32). Such sequences are present in $\approx 3 \times 10^5$ copies in the haploid mammalian genome (30–32). The biological significance of these sequences is not known, although they can be transcribed, and it has been proposed that they play a role in processing of pre-mRNA (30). For instance, the Alu family of dispersed repetitive sequences in humans contains (*i*) internal repeats with various levels of homology, (*ii*) 3′-teminal adenosine-rich regions, and (*iii*) direct terminal repeats. The terminal repeats differ in each case analyzed so far, suggesting that they were generated during insertion (30, 32) and represent duplicated chromosomal DNA sequences where insertion occurred. Thus, the data raise the possibility that the 508-bp structure of $I_B$ occurs elsewhere in the rat genome. Preliminary results using $I_B$ as a hybridization probe indicate that this sequence is a member of the dispersed repetitive sequences and is furthermore found in at least three other positions in the next 7 kb on the 3′ side of the rGH gene. It is of interest in this respect that intron B of the hGH gene is only 200 bp long and does not contain large repeated sequences (P. H. Seeburg, personal communication).

Although the 508-bp sequence in $I_B$ is a member of a family of dispersed repetitive sequences, it is somewhat unusual in that the length of such sequences in the eukaryotic genome are usually ≈300 bp (30). One explanation may lie in the structure of the $I_B$ insertion itself. As the 195-bp tandem repeat is 96% homologous, its duplication (via unequal crossing over) must have occurred recently, possibly after the insertion into the intron. If this is the case, then the inserted sequence was only 313 bp long, which is similar to the size of interspersed repeated sequences. However, we cannot exclude the possibility that this duplication event occurred elsewhere in the DNA and that the whole 508-bp fragment was inserted into the intron.

The finding of repetitive DNA in an intron suggests mechanisms whereby such sequences can affect intron size. This could be due not only to the insertion itself but also to the potential for the sequences, once inserted, to be duplicated or deleted. Short stretches of simple repeated sequences such as A-T-A-T-A-T, adenosine-rich regions and the three long repeats themselves found in the gene are factors that can increase the frequency of unequal crossing over and deletions (33, 34). Analysis of the same repetitive sequences lying 3′ to the rGH gene should provide information as to whether the terminal direct repeats are an integral part of this sequence or are created on insertion.

It has been proposed that dispersed repetitive sequences are "hot spots" for genetic recombination and may be responsible for rearrangements of genes and exons (35). The idea that these sequences could be involved in the movement of genes is consistent with the findings in this and other (22, 32) studies of such sequences in the proximity of particular genes. However, the proposal that these repetitive sequences are involved in the rearrangements of particular exons requires that they occur between exons; the current finding of such sequences in an intron of the rGH gene is consistent with such an idea.

We thank Nancy Cooke, Peter H. Seeburg, and Guy Cathala for providing us with unpublished results and for advice; T. Sargent and J. Bonner for providing to us the rat chromosomal DNA library; R. Crea for the synthetic primer for dideoxy sequence analysis; N. Eberhardt, G. Bell, F. Savouret, and M. Wegnez for helpful discussions; and S. Bock and H. Martinez for help with computer programs in analyzing DNA sequences. A.B. is a recipient of a Max Kade fellowship. J.D.B. is an Investigator of the Howard Hughes Medical Institute. This work was supported by National Institutes of Health Grants AM18878-05 and AM19997-04.

1. Frantz, A. G. (1976) in *Peptide Hormone*, ed. Parson, J. A. (University Park Press, Baltimore, MD). p. 199.
2. Baxter, J. D., Seeburg, P. H., Shine, J., Martial, J. A., Ivarie, R. D., Johnson, L. K., Fiddes, J. C. & Goodman, H. M. (1979) *Cold Spring Harbor Conferences on Cell Proliferation* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Vol. 6 317–337.
3. Niall, H. D., Hogan, M. L., Sayer, R., Rosenblum, I. Y. & Greenwood, F. C. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 866–869.
4. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. & Goodman, H. M. (1977) *Nature (London)* **270**, 486–494.
5. Shine, J., Seeburg, P. H., Martial, J. A., Baxter, J. D. & Goodman, H. M. (1977) *Nature (London)* **270**, 494–499.
6. Cooke, N. E., Coit, D., Weiner, R. I., Baxter, J. D. & Martial, J. A. (1980) *J. Biol. Chem.* **255**, 6502–6510.
7. Lewis, U. J., Singh, R. N. P., Tutwiler, G. F., Sigel, M. B., Vander Laan, E. F. & Vander Laan, W. P. (1980) *Rec. Prog. Horm. Res.* **36**, 477–508.
8. Wallis, M. (1980) *Nature (London)* **284**, 512.
9. Sargent, T. D., Wu, J., Sela-Trepat, J. M., Wallace, R. B., Reyes, A. A. & Bonner, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3256–3260.
10. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
11. Taylor, J. M., Illmensee, R. & Summers, J. (1976) *Biochim. Biophys. Acta.* **442**, 324–330.
12. Wahl, G. M., Stern, M. & Stark, G. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3683–3687.
13. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15**, 687–701.
14. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9**, 309–321.
17. Chien, Y. & Thompson, E. B. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4583–4587.
18. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
19. Corden, J., Wasylk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980) *Science* **209**, 1406–1414.
20. Maurer, R. A., Gubbins, E. J., Erwin, C. R. & Donelson, J. E. (1980) *J. Biol. Chem.* **255**, 2243–2246.
21. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980) *Nucleic Acids Res.* **8**, 127–142.
22. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., Deriel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21**, 653–668.
23. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
24. Jones, C. W. & Kafatos, F. C. (1980) *Cell* **22**, 855–867.
25. Gubbins, E. J., Maurer, R. A., Lagrimini, M., Erwin, C. R., Donelson, J. E. (1980) *J. Biol. Chem.* **255**, 8655–8662.
26. Lohstroh, A. J. & Krahl, M. E. (1978) *Diabetes* **27**, 597.
27. Bell, G. I., Pictet, R. L., Rutter, W. J., Cordell, B., Tischer, E. & Goodman, H. M. (1980) *Nature (London)* **284**, 26–32.
28. Stein, J. P., Catterall, J. F., Kristo, P., Means, A. R., O'Malley, B. W. (1980) *Nature (London)* **283**, 220–224.
29. Sharp, P. A. (1981) *Cell* **23**, 643–646.
30. Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L. & Schmid, C. W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1398–1402.
31. Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T. & Schmid, C. W. (1980) *Nature (London)* **284**, 372–374.
32. Bell, G. I., Pictet, R. & Rutter, W. J. (1980) *Nucleic Acids Res.* **8**, 4091–4109.
33. Farabaugh, P. J. (1978) *Nature (London)* **274**, 756–769.
34. Slightom, J. L., Blechl, A. E. & Smithies, O. (1980) *Cell* **21**, 627–638.
35. Gilbert, W. (1978) *Nature (London)* **271**, 501.