Vol. 208, No. 1, 1995 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
March 8, 1995 Pages 339-344

ISOLATION AND STRUCTURAL ANALYSIS OF THE 5' FLANKING REGION OF THE GENE ENCODING THE HUMAN GLUCAGON RECEPTOR

Joseph Buggy, Jacqueline Hull and Heeja Yoo-Warren[1]

Signal Transduction Section, Institute for Metabolic Disorders,
Miles Inc.,
West Haven, CT 06516

Received January 23, 1995

The gene encoding the human glucagon receptor, including several kb of upstream sequence, was isolated from a bacteriophage λ FIX II library constructed from human placental DNA. We report here the novel sequence of the 5' flanking region of the gene, the identification of a previously unreported intron of 5 kb, and the identification of the transcription start point of the glucagon receptor-specific transcript, which estimates the length of the first exon to be 300 bp.

Glucagon is a 29 amino acid hormone that helps regulate blood glucose levels through its major target organ, the liver (1). The alpha cells of the pancreas secrete glucagon which then binds to receptors present on the surface of liver cells (2). The glucagon-glucagon receptor complex activates adenylate cyclase by coupling with heterotrimeric GTP binding proteins (G-proteins), which results in increased intracellular levels of cAMP. This activation ultimately

[1]Fax (203) 937-2686; e-mail: hyw@mrc.com.

Abbreviations: bp, base pair(s); cAMP, cyclic adenosine 3', 5'-monophosphate; cDNA, complementary DNA; G-protein, guanine nucleotide-binding protein; GPCR, G-protein coupled receptor; GR, glucagon receptor; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); PCR, polymerase chain reaction; RT reverse-transcription; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl/0.015 M Na_3·citrate, pH, 7.6; UTR, untranslated region(s).

0006-291X/95 $5.00

leads to increased glucose synthesis (gluconeogenesis) and glycogen breakdown (glycogenolysis) (3).

Recently, cDNA clones for both human and rat GR were isolated (4,5). Sequence comparisons with other known receptors place the GR in a well characterized family of G-Protein coupled receptors (GPCR) that have seven transmembrane domains (6). Among GPCRs, the GR shares the highest degree of sequence similarity with a subfamily which includes the receptors for secretin (7), calcitonin (8), parathyroid hormone (9) and the glucagon-like peptide (10).

In addition to understanding the basic features of the GR, it is important to develop an understanding of the structure and regulation of the encoding gene. This paper describes the cloning and structural characterization of the human *GR* gene, the identification of a large intron present in the 5' UTR, the mapping of the 5' end of the *GR* specific mRNA, and novel sequence analysis of the 5' region of the gene.

MATERIALS AND METHODS

A cDNA clone encoding the human GR was isolated from a human liver cDNA library using the previously described rat *GR* cDNA (5) as a probe. Nucleotide sequencing of the human *GR* cDNA has demonstrated that this clone is approximately 1.7 kb in length , consisting of ~ 70 bases of the 5' untranslated region, the full coding sequence, and ~230 bases of the 3' untranslated segment. Nucleotide sequence of the cDNA clone was identical to the corresponding portion of the published human *GR* (Genbank accession #L20316, 11). Sequence analysis was performed using chain termination sequencing reactions (Promega T7 sequencing kit) on a double stranded genomic clone template.

Primer Extension Analysis:

An antisense oligonucleotide primer specific for the 5' untranslated region of the human *GR* (5'-TCGGCTTGGCTCCCCGCCGTCCTCTG -3') was 5' end labeled by mixing 90 ng of oligonucleotide with 100µCi of [g-^{32}P]ATP and 5 U of T4 polynucleotide kinase in a 50µl reaction mixture (kinase buffer contains 50mM Tris-HCL [pH 9.0], 10mM $MgCl_2$, 5mM dithiothreitol, and 5% glycerol) for 1 h at 37°C. The labeled primer (100KCPM) was added to 2 µg of human liver poly A+ RNA (Clontech) in a 30µl hybridization reaction containing 80% formamide, 0.4 M NaCl, 400 mM piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES), pH 6.4, and 1mM EDTA. The mixture was placed in a waterbath at 85° C for 10 minutes, then slowly lowered to 34° C over a period of 4-5 h and held at 34° C overnight. The mixture was precipitated by adding 170 µl of 0.3 M sodium acetate and 500 µl ethanol, chilled at -80° C for 20 minutes, and then pelleted by microcentrifugation for 10 minutes. The pellet was dissolved in a 20 µl "reverse transcription mix": 50mM Tris-HCL, pH

8.5, 8 mM $MgCl_2$, 30 mM KCl, 1 mM dithiothreitol, 0.5 mM dNTPs, and 40 units AMV reverse transcriptase (Boehringer Mannheim Biochemica). The reaction was incubated at 37° C for 30 minutes, then 42° C for 30 minutes, then mixed with 15 µl of formamide loading dye to stop the reaction. The reaction was then heated to 90° C for 10 minutes and run on a 6% Long-Ranger polyacrylamide gel (AT Biochem Corp.) alongside a sequencing ladder, which was created by use of chain termination sequencing reactions (Promega T7 sequencing kit) on a single stranded genomic clone template using the same oligonucleotide primer as for primer extension (above).

RESULTS AND DISCUSSION

Structure of the human GR gene

A genomic clone encoding the full length human *GR* was isolated from a human placenta genomic library using a restriction fragment generated from the human *GR* cDNA (1.7 kb) as a probe. A clone containing a 13.5 kb insert, designated λ 58111, was chosen for further analysis because it contained the greatest amount of DNA 5' to the coding sequence.

To estimate the number and size of each intron, a series of reverse transcriptase-polymerase chain reaction (RT-PCR) experiments was carried out using primers from the cDNA sequence and the genomic clone λ 58111 as a template (12). Our results indicate that the *GR* gene contains 12 introns downstream of the ATG start codon (Fig. 1), in agreement with previous reports (11). However a large, previously unreported intron of approximately 5 kb was also found to lie within the 5' untranslated region of the gene (later determined to be the first of thirteen total introns). Nucleotide sequencing of the genomic clone and sequence alignment with the cDNA defined each splice site. The sizes of each exon range between 42 bases (exon 13) and >500 bases (exon 14).

Nucleotide sequence and organization of the 5' region of the GR gene

The 5' end of the *GR* specific transcript was mapped by primer extension analysis using a poly A+ RNA preparation extracted from human liver (14). A 27 base synthetic oligonucleotide whose sequence was derived from the *GR* cDNA was used as a primer in a reaction which yielded one major product as shown in Fig. 2. The reaction yielded no product with tRNA only (data not shown). When the primer extension product was run on an acrylamide gel alongside a standard Sanger sequencing ladder generated from the cloned genomic DNA template, it locates the start site for

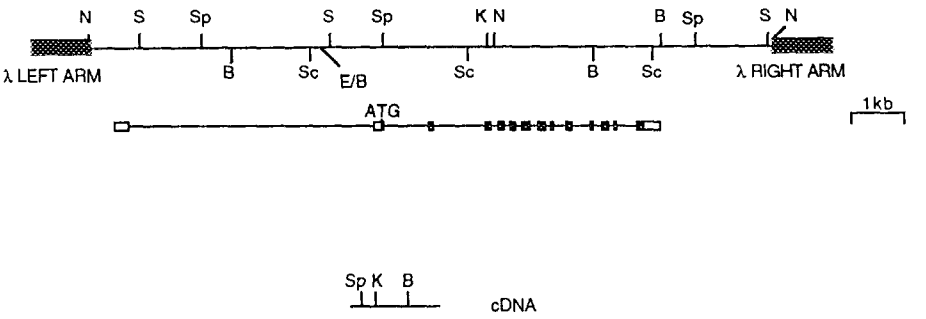

Figure 1. Diagram of the structure of the human *GR* gene. Top line, a detailed restriction map of the 13.5-kb locus containg the entire *GR* gene. B, Bam H1; E, Eco R1; K, Kpn 1; N, Not 1; S, Sal 1; Sc, Sac 1; Sp, Sph 1. Middle line, intron-exon organization of the *GR* gene. Filled and open boxes represent the coding and noncoding exons, respectively, and solid lines introns. Bottom line, Partial restriction map of cDNA for the *GR*.

transcription to the region boxed in Fig. 2, a region that lies approximately 475 bases upstream of the AUG start codon of the coding sequence within the mature mRNA.

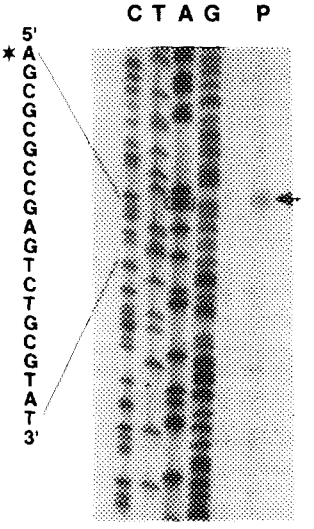

Figure 2. Mapping of the 5' end of human *GR* specific mRNA. A strand complementary to the *GR* specific transcript was synthesized by use of a ^{32}P-labeled antisense primer. A DNA sequencing ladder was generated from the same primer using an M13 recombinant single stranded genomic *GR* clone as template. The position of the resulting single major product, indicating the start site of transcription, is indicated by an arrow. The corresponding nucleotide sequence is shown, with the start site of transcription indicated by an asterisk (*).

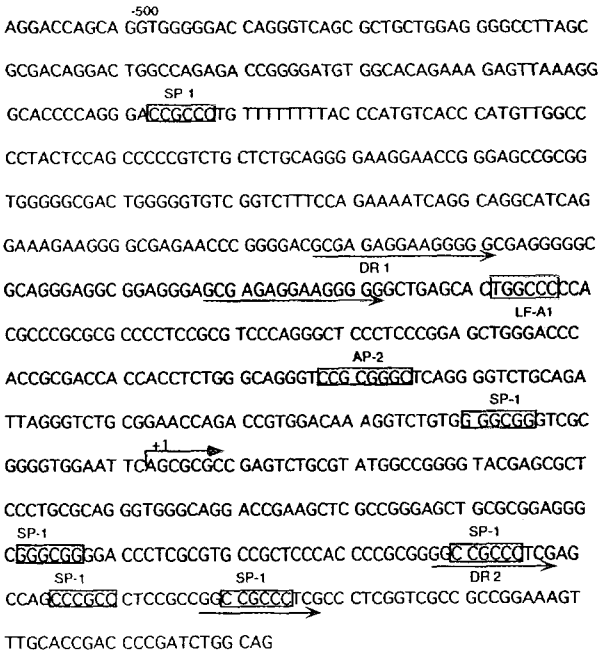

Figure 3. Arrangement of the putative promoter region of the human *GR* gene. The start site of transcription is indicated (arrow and +1). Highlighted sequences are: DR, direct repeat; LF-A1, liver specific factor; as well as the known consensus sequences for the Sp 1 and AP 2 transcription factor binding sites.

The nucleotide sequence of the 5' region is reported in Fig. 3. Underlined in Fig. 3 are multiple sequences corresponding to consensus Sp 1 binding sites (15), one AP 2 site (16), as well as a motif that is known to bind the liver specific transcription factor LF-A1 (17). Also indicated are two perfect direct repeats. Interestingly, we neither observe a consensus "TATA" box, nor a consensus "CAAT" box in this region. The *GR* gene promoter may thus represent another example of a gene which, like the promoter region of the human insulin receptor gene (18), lacks such conventional *cis*-acting elements. In contrast to the human insulin receptor gene, which has multiple heterogeneous start sites located over a 400 bp region, only a single start site is found for the *GR*. It is not known how transcription initiation is controlled accurately in TATA-less genes, although it is noted that they often contain multiple Sp 1 sites and the "initiator" sequences which bind factors such as YY 1 and E2F (19). We are currently investigating whether

the 5' flanking region shown in Fig. 3 contains all of the necessary sequences for appropriate transcription of the human *GR* gene.

ACKNOWLEDGMENTS

We would like to thank James N. Livingston and Ann Gore Willse for helpful discussions and comments on this project and Peiling Jiang for her assistance in sequencing the gene.

REFERENCES

1. Unger, R.H. and Orci, L. (1990) Ellenberg and Rifkin's Diabetes Melitus, Theory and Practice. Elsevier, New York.
2. Unger, R.H. and Foster, D. (1992) Williams' Textbook of Endocrinology Ed. 8, Saunders, Philadelphia.
3. Smith, R.J. (1989) Endocrinology (L.J . DeGroot, Ed.), Vol. 2, WB Saunders Co., Philadelphia.
4. MacNeil, D.J., Occi, J.L., Hey, P.J., Strader, C.D. and Graziano, M.P. (1994) Biochem Biophys Res Commun 198, 328-334.
5. Jelinek, L.J., Lok, S., Rosenberg, G.B., Smith, R.A., Grant, F.J., Biggs, S., Bensch, P.A., Kuijper, J.L., Sheppard, P.O., Sprecher, C.A., O'Hara, P.J., Foster, D., Walker, K.M., Chen, L.H.J., McKernan, P.A. and Kindsvogel, W. (1993) Science 259, 1614-1616.
6. Segre, G.V. and Goldring, S.R. (1993) TEM 4, 309-314.
7. Ishihara, T., Nakamura, S., Kaziro, Y., Takahashi, T., Takakhashi, K. and Nagata, S. (1991) EMBO J. 10, 1635-1641.
8. Lin, H.Y., Harris, T.L., Flannery, M.S., Aruffo, A., Kaji, E.H., Gorn, A., Kolakowski, L.F., Lodish, H.F. and Goldring, S.R. (1991) Science 254, 1022-1023.
9. Juppner, H., Abou-Samra, A-B., Freeman, M., Kong, S.F., Schipani, E., Kolakowski, L.F. Jr., Hock , J., Potts, J.T. Jr., Kronenberg, H.M. and Segre, G.V. (1991) Science 254, 1024-1026.
10. Thorens, B. (1992) Proc Natl Acad Sci USA, 89, 8641-8645.
11. Lok, S., Kuijper, J.L., Jelinek, L.J., Kramer, J.M., Whitmore, T.E., Sprecher, C.A., Mathewes, S., Grant, F.J., Biggs, S.H., Rosenberg, G.B., Sheppard, P.O., O'Hara, P.J., Foster, D.C. and Kinsvogel, W.K. (1994) Gene 140, 203-209.
12. Matzuk, M. and Bradley, A. (1992) Biochem Biophy Res Comm 185, 404-413.
13. Shapiro, M.B. and Senepathy, P. (1987) Nucleic Acids Res 15, 7155-7174.
14. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning, A laboratory manual (second edition), Cold Spring Harbor Laboratory Press, New York.
15. Briggs, R., Kadonaga, J.T., Bell, S.P. and Tjian,R (1986) Science 234, 47- 52.
16. Mitchell, P.J., Wang, C. and Tjian, R. (1987) Cell 50, 847- 861.
17. Hardon, E.M., Frain, M., Paonessa, G. and Cortese, R. (1988) EMBO J 7, 1711-1719.
18. Cameron, K.E., Resnik, J. and Webster, N.J.G. (1992) J Biol Chem 267, 17375-17383.
19. Azizkhan, J.C., Jensen, D.E., Pierce, A.J. and Wade, M. (1993) Critical review in eukaryotic gene expression 3, 229-254.