Vol. 183, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 28, 1992 Pages 8-13

MOLECULAR CLONING AND SEQUENCING OF THE GENE ENCODING HUMAN ANGIOTENSIN II TYPE 1 RECEPTOR*

Hiroaki Furuta, Deng-Fu Guo, and Tadashi Inagami

Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, TN 37232

Received January 11, 1992

Summary: The gene of human angiotensin II type 1 (AT_1) receptor was isolated from a lymphocyte genomic library. The coding region of the human AT_1 receptor gene was contained in a single exon coding segment of the gene indicating an intronless structure of the coding region. The amino acid sequence of human AT_1 receptor deduced from its base sequence has 359 amino acids and showed a high degree of sequence identity to bovine and rat AT_1 receptor sequences. Amino acid substitutions specific to the human AT_1 receptor were mostly confined to the carboxy terminal half of the molecule. The seven-transmembrane domains are well-conserved in those sequences.

Angiotensin II plays an important role in the homeostasis of blood pressure, fluid volume, electrolyte balance and cardiovascular hypertrophy through its action on specific receptors in its peripheral and central target tissues (1). Reflecting the versatility of the biological effects mediated by angiotensin II, its receptors have been detected in a wide variety of tissues: the vascular wall, adrenal, liver, kidney, lung, brain, and pituitary (2). Recently, Sasaki et al.(3) and Murphy et al.(4) reported the expression cloning and sequence analyses of bovine adrenal and rat aortic angiotensin II type 1 (AT_1) receptors, respectively. However, the structure of human AT_1 receptor and its gene remained unresolved. In this paper, we describe cloning of the human AT_1 receptor gene and its amino acid sequence deduced from the nucleotide sequence. Interestingly, the entire coding region was contained in a single exon.

*Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. Z11162.

0006-291X/92 $1.50

Materials and Methods

Materials. Radiolabeled nucleotides [α-^{32}P] dCTP (110TBq/mmol) were obtained from New England Nuclear (Boston,MA). Oligonucleotide primers were synthesized in a MilliGen/Biosearch Cyclone Plus DNA synthesizer. Dideoxy sequencing kits were from United States Biochemicals (Cleveland, OH).

Screening of genomic library. A human lymphocyte genomic library in lambda dash was obtained from Stratagene (La Jolla, CA). A total of 1.2×10^6 lambda phage plaques were screened on nylon membranes (Hybond-N, Amersham). Hybridization was performed at 42°C in a 6×SSC solution containing 2.5×Denhardt's, 30% formamide, 1% SDS with a labeled probe from rat SacI-KpnI DNA fragment for the AT_1 receptor (4,5). The membranes were washed twice with 2×SSC at room temperature for 5 min, then soaked twice at 55°C in 2×SSC containing 1% SDS and then exposed to radiographic film at -70°C with an intensifying screen. Positive clone was subcloned into Bluescript KS(+) plasmids.

Nucleotide sequencing. The nucleotide sequences of DNAs were determined by the dideoxy chain-termination method (6). Commercially available primers corresponding to plasmid sequences (T3 and T7) and synthetic oligonucleotide primers were used. Analysis of DNA data was performed by the DNA strider 1.0 program.

Results and Discussion

Out of approximately 1.2×10^6 recombinant phage five positive clones were selected by plaque hybridization of a genomic library. One clone contained a 6.1 kb *Bam* H1-fragment which were found to contain the complete reading frame of the human AT_1 receptor plus approximately 3.6 and 1.4 kb of 5'-and 3'-flanking regions, respectively (Fig. 1). The nucleotide sequence of the human AT_1 receptor gene and the corresponding amino acid sequence are shown in Fig. 2. The open reading frame consisting of 1,080 bp contained an initiation codon ATG and a termination codon TGA, and encoded 359 amino acids which accounted for a molecular mass of 41,066 daltons. The ATG sequence designated as the translational initiation codon agreed with Kozak's consensus sequence (7): C in nucleotide position -4, A in -3, and A in +4. Beyond the stop codon, two

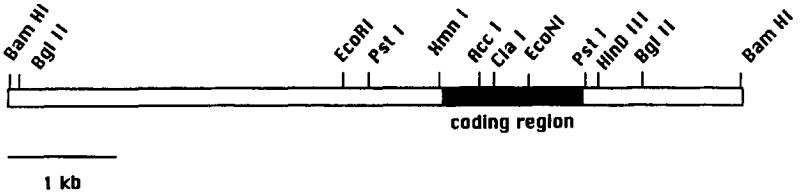

Figure 1. Restriction map of human AT_1 receptor gene. A partial restriction map of the 6.1-kb *Bam* H1 fragment containing human AT_1 receptor gene is shown. The *black bar* represents the human AT_1 receptor open reading frame.

putative polyadenylation signals (AATAAA) were located in the 3'-noncoding region, at +1947 and +2154, respectively. Six ATTTA nucleotide sequences (Fig. 2) are seen in the 3'-franking regions which are considered as a motief common to transiently expressed labile mRNAs (8). Those seqences might cause the instability of the mRNA of human AT_1 receptor for posttrascriptional regulation. An exon-intron junction site was identified at the nucleotide position between -47 and -48 upstream of the initiation codon, and a poly(A) tail was located starting at +1968 in comparison with the base sequence of an AT_1 cDNA cloned from human liver (unpublished result from this laboratory). These results indicate that the human AT_1 receptor gene has an exon in which a complete uninterupted open reading frame for the entire receptor is contained. The amino acid sequence of the human AT_1 receptor was 95.3 % identical to the bovine adrenal AT_1 receptor sequence (3) (see Fig. 3A) and 94.7 % to the rat (renal or vascular smooth muscle cell) AT_1 receptor sequence (4,5) all having uniformly 359 amino acid residues. At the base sequence level the human and bovine sequences showed 91.0% identity.

Hydropathy analysis (Fig. 3B) and comparison with bovine (3) and rat (4,5) AT_1 receptor sequences permitted assignment of seven putative transmembrane domains I-VII. One potential N-glycosylation site (Asn-X-Ser/Thr) is located in the N-terminal extracellular domain and two in the second extracellular loop. Each of the four extracellular domain contained one cysteine residue which would form two disulfide bonds essential for ligand binding. Positions of all of these amino acid residues are well conserved in rat and bovine AT_1 receptors. The amino terminal extracellular domain, the third intracellular loop and carboxy terminal cytoplasmic segment contain distinctly high hydrophilic sequences indicating their functions in the extra membrane space.

```
                TTAATGATAAATGAATTGGTCCTGCTTACCTCAGGAAAAACTTTCAAGTCTTTCTGAAAAACTAATTT  -396
AATTCAGTAGTATTTTCTAAGATTTAGGTTATGTTTTTAATCAATTTGGAAACCAAGATTTACTTATAGAAAAAAAGGA  -317
AAAGGACCTAGATAGGTTTATTCACATAGAATCCCAATTTCACTTCTCTGGATGATACCATTTTCTACAAAAGCAATTA  -238
TGTTCTAAAATTTAAGTGTGCTTTCTTAGGCTTTATCAGTTCACAGTGTTTCCTTAAGAAATATGATCCAGTATTTTTT  -159
CCTAAGACTAAAGTTGAGTTACTACGTTTATGACTGAGAAATGAATGTTTGTTAGTTTGTTTGTTTACAATAAGAATTT   -80
TTTCTTTACCATTTTATTTTTATTTTCCCCAGGTGTATTTGATATAGTGTTTGCAACAAATTCGACCCAGGTGATCAAA    -1

ATG ATT CTC AAC TCT TCT ACT GAA GAT GGT ATT AAA AGA ATC CAA GAT GAT TGT CCC AAA    60
Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp Asp Cys Pro Lys   20
GCT GGA AGG CAT AAT TAC ATA TTT GTC ATG ATT CCT ACT TTA TAC AGT ATC ATC TTT GTG   120
Ala Gly Arg His Asn Tyr Ile Phe Val Met Ile Pro Thr Leu Tyr Ser Ile Ile Phe Val   40
GTG GGA ATA TTT GGA AAC AGC TTG GTG GTG ATA GTC ATT TAC TTT TAT ATG AAG CTG AAG   180
Val Gly Ile Phe Gly Asn Ser Leu Val Val Ile Val Ile Tyr Phe Tyr Met Lys Leu Lys   60
ACT GTG GCC AGT GTT TTT CTT TTG AAT TTA GCA CTG GCT GAC TTA TGC TTT TTA CTG ACT   240
Thr Val Ala Ser Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr   80
TTG CCA CTA TGG GCT GTC TAC ACA GCT ATG GAA TAC CGC TGG CCC TTT GGC AAT TAC CTA   300
Leu Pro Leu Trp Ala Val Tyr Thr Ala Met Glu Tyr Arg Trp Pro Phe Gly Asn Tyr Leu  100
TGT AAG ATT GCT TCA GCC AGC GTC AGT TTC AAC CTG TAC GCT AGT GTG TTT CTA CTC ACG   360
Cys Lys Ile Ala Ser Ala Ser Val Ser Phe Asn Leu Tyr Ala Ser Val Phe Leu Leu Thr  120
TGT CTC AGC ATT GAT CGA TAC CTG GCT ATT GTT CAC CCA ATG AAG TCC CGC CTT CGA CGC   420
Cys Leu Ser Ile Asp Arg Tyr Leu Ala Ile Val His Pro Met Lys Ser Arg Leu Arg Arg  140
ACA ATG CTT GTA GCC AAA GTC ACC TGC ATC ATC ATT TGG CTG CTG GCA GGC TTG GCC AGT   480
Thr Met Leu Val Ala Lys Val Thr Cys Ile Ile Ile Trp Leu Leu Ala Gly Leu Ala Ser  160
TTG CCA GCT ATA ATC CAT CGA AAT GTA TTT TTC ATT GAG AAC ACC AAT ATT ACA GTT TGT   540
Leu Pro Ala Ile Ile His Arg Asn Val Phe Phe Ile Glu Asn Thr Asn Ile Thr Val Cys  180
GCT TTC CAT TAT GAG TCC CAA AAT TCA ACC CTT CCG ATA GGG CTG GGC CTG ACC AAA AAT   600
Ala Phe His Tyr Glu Ser Gln Asn Ser Thr Leu Pro Ile Gly Leu Gly Leu Thr Lys Asn  200
ATA CTG GGT TTC CTG TTT CCT TTT CTG ATC ATT CTT ACA AGT TAT ACT CTT ATT TGG AAG   660
Ile Leu Gly Phe Leu Phe Pro Phe Leu Ile Ile Leu Thr Ser Tyr Thr Leu Ile Trp Lys  220
GCC CTA AAG AAG GCT TAT GAA ATT CAG AAG AAC AAA CCA AGA AAT GAT GAT ATT TTT AAG   720
Ala Leu Lys Lys Ala Tyr Glu Ile Gln Lys Asn Lys Pro Arg Asn Asp Asp Ile Phe Lys  240
ATA ATT ATG GCA ATT GTG CTT TTC TTT TTC TTT TCC TGG ATT CCC CAC CAA ATA TTC ACT   780
Ile Ile Met Ala Ile Val Leu Phe Phe Phe Phe Ser Trp Ile Pro His Gln Ile Phe Thr  260
TTT CTG GAT GTA TTG ATT CAA CTA GGC ATC ATA CGT GAC TGT AGA ATT GCA GAT ATT GTG   840
Phe Leu Asp Val Leu Ile Gln Leu Gly Ile Ile Arg Asp Cys Arg Ile Ala Asp Ile Val  280
GAC ACG GCC ATG CCT ATC ACC ATT TGT ATA GCT TAT TTT AAC AAT TGC CTG AAT CCT CTT   900
Asp Thr Ala Met Pro Ile Thr Ile Cys Ile Ala Tyr Phe Asn Asn Cys Leu Asn Pro Leu  300
TTT TAT GGC TTT CTG GGG AAA AAA TTT AAA AGA TAT TTT CTC CAG CTT CTA AAA TAT ATT   960
Phe Tyr Gly Phe Leu Gly Lys Lys Phe Lys Arg Tyr Phe Leu Gln Leu Leu Lys Tyr Ile  320
CCC CCA AAA GCC AAA TCC CAC TCA AAC CTT TCA ACA AAA ATG AGC ACG CTT TCC TAC CGC  1020
Pro Pro Lys Ala Lys Ser His Ser Asn Leu Ser Thr Lys Met Ser Thr Leu Ser Tyr Arg  340
CCC TCA GAT AAT GTA AGC TCA TCC ACC AAG AAG CCT GCA CCA TGT TTT GAG GTT GAG TGA  1080
Pro Ser Asp Asn Val Ser Ser Ser Thr Lys Lys Pro Ala Pro Cys Phe Glu Val Glu END  359

CATGTTCGAAACCTGTCCATAAAGTAATTTTGTGAAAGAAGGAGCAAGAGAACATTCCTCTGCAGCACTTCACTACCAA  1159
ATGAGCATTAGCTACTTTTCAGAATTGAAGGAGAAAATGCATTATGTGGACTGAACCGACTTTTCTAAAGCTCTGAACA  1238
AAAGCTTTTCTTTCCTTTTGCAACAAGACAAAGCAAAGCCACATTTTGCATTAGACAGATGACGGCTGCTCGAAGAACA  1317
ATGTCAGAAACTCGATGAATGTGTTGATTTGAGAAATTTTACTGACAGAAATGCAATCTCCCTAGCCTGCTTTTGTCCT  1396
GTTATTTTTTATTTCCACATAAAGGTATTTAGAATATATTAAATCGTTAGAGGAGCAACAGGAGATGAGAGTTCCAGAT  1475
TGTTCTGTCCAGTTTCCAAAGGGCAGTAAAGTTTTCGTGCCGGTTTTCAGCTATTAGCAACTGCTGCTACACTTGCACC  1554
TGGTACTGCACATTTTGTACAAAGATATGCTAAGCAGTAGTCGTCAAGTTGCAGATCTTTTTGTGAAATTCAACCTGTG  1633
TCTTATAGGTTTACACTGCCAAAACAATGCCCGTAAGATGGCTTATTTGTATAATGGTGTTACTAAAGTCACATATAAA  1712
AGTTAAACTACTTGTAAAGGTGCTGCACTGGTCCCAAGTAGTAGTGTCTTCCTAGTATATTAGTTTGATTTAATATCTG  1791
AGAAGTGTATATAGTTTGTGGTAAAAAGATTATATATCATAAAGTATGCCTTCTGTTTAAAAAAAGTATATATCTACAC  1870
ATATATATATATGTATATCTATATCTCTAAACTGCTGTTAATTGATTAAAATCTGGCAAAGTTATATTTACTTTAAAAT  1949
AAAATAATTTTATTGCAATGTATTTATCTTCATTACTTAAAATAGATGCTAATTTATTTTAAAATAAGACTACCTTGAA  2028
TGAGTATGAATATATTTTTATTTAAATTTTGATACAACTGATAGTTTAATACTATTGGTTATAGATTTTTTATCCTGAC  2107
ATTGAAAAGTTAAAGAAAAAACATTTTGTTCTACTGCATGTCATGGAATAAACACATCGTTT
```

Figure 2. Nucleotide sequence and deduced amino acid sequence of a human AT_1 receptor gene. A stop codon is indicated as *END*. Two potential polyadenylation signals (AATAAA) are *underlined*. The nucleotide motiefs corresponding to AUUUA, a putative mRNA destabilization signal (8) are enclosed within *boxes*. Both an exon-intron junction site and poly(A) tail start site in a cDNA encoding human AT_1 receptor are indicated with *asterisks*.

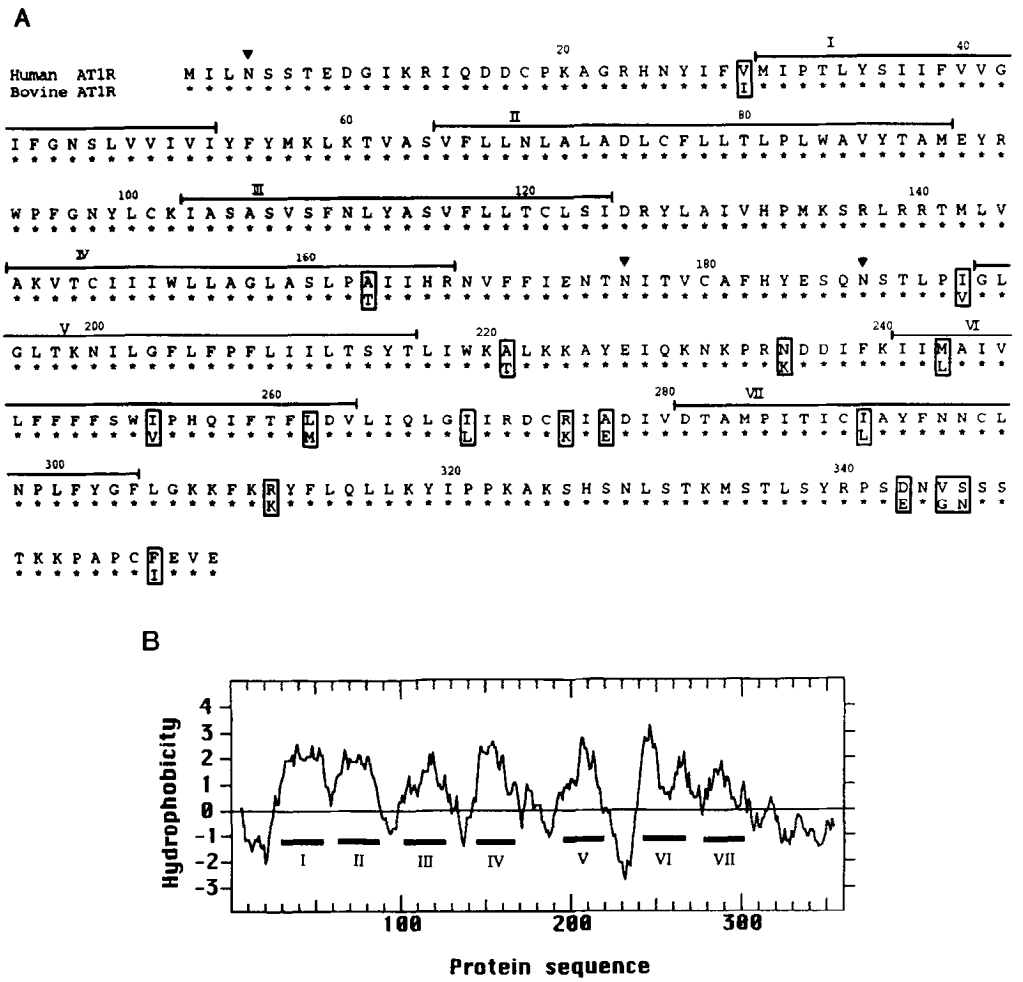

Figure 3. The alignment of the predicted protein sequence of the AT_1 receptor (A) and hydropathy analysis (B). *A:* Comparison of the primary structure of the human and bovine AT_1 receptor. Identical amino acids are shown by *asterisks* in the bovine sequence and non-identical residues are enclosed in *boxes*. The seven putative transmembrane domains are indicated above the amino acid sequences. The potential N-glycosylation sites are marked by *triangles*. *B:* Kyte-Doolittle (9) hydropathy analysis of the amino acid sequence of the human AT_1 receptor. The putative seven transmembrane domains are indicated with *black bars*.

Besides the similarity with bovine and rat AT_1 receptor structure, the present results uncovered that several specific substitutions are localized predominantly in the carboxy terminal half of the sequence.

Elucidation of the structural details of human AT_1 receptor will be very useful for the design of selective receptor antagonists for therapeutic purposes and identification of genetic disorders of the receptor in humans.

Acknowledgments

We thank Drs Marion L. Dodson and Makoto Yoshida for their aid on the computer analysis of sequences, Tim Atkinson for synthesizing oligonucleotides for sequencing, and Tina Stack for secretarial assistance. This work was supported by Research Grants HL14192 and HL 35323 from the National Institutes of Health of U.S.A.

References

1. Peach, M. J. (1977) *Physiol. Rev.*, 57, 313-370
2. Millan M. A., Carvallo P., Izumi S., Zemel S., Catt K. J., and Aguilera, G. (1989) *Science* 244, 1340-1342
3. Sasaki, K., Yamano, Y., Bardhan, S., Iwai, N., Murray, J. J., Hasegawa, M., Matsuda, Y. and Inagami, T. (1991) *Nature*, 351, 230-233
4. Murphy, T. J., Alexander, R. W., Griendling, K. K., Runge, M. S. and Bernstein, K. E. (1991) *Nature*, 351, 233-236
5. Iwai, N., Yamano, Y., Chaki, S., Konishi, F., Bardhan, S., Tibbetts,C., Sasaki, K., Hasegawa, M., Matsuda, Y. and Inagami, T. (1991) *Biochem. Biophys. Res. Commun.*, 177, 299-304
6. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 5463-5467
7. Kozak, M. (1984) *Nucleic Acids Res.*, 12, 857-872
8. Shaw, G., and Kamen, R. (1986) *Cell*, 46, 659-667
9. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* 157, 105-132