Cloning and sequence analysis of cDNA for rat angiotensinogen

(angiotensin/recombinant DNA/DNA sequence/blood pressure)

Hiroaki Ohkubo*, Ryoichiro Kageyama*, Mayumi Ujihara*, Tadaaki Hirose†, Seiichi Inayama‡, and Shigetada Nakani*°

*Institute for Immunology, Kyoto University Faculty of Medicine, Kyoto 606, Japan; and †Pharmaceutical Institute, Keio University School of Medicine, Tokyo 160, Japan

Communicated by Joseph L. Goldstein, January 20, 1983

ABSTRACT A mixture of tetradecamer oligodeoxyribonucleotides complementary to the codons specifying the carboxy-terminal sequence, Ile-His-Pro-Phe-His, of angiotensin was chemically synthesized as two pools and used for the isolation of a cDNA clone specific for angiotensinogen from a cDNA bank of rat liver mRNA sequences. The two pools (oligo 1 and oligo 2), each containing 24 oligodeoxyribonucleotides, were first used as primers to initiate reverse transcription of rat liver mRNA. One of the pools (oligo 1) was found to prime a specific 32P-labeled cDNA of approximately 160 nucleotides that contained the anticoding sequence corresponding exactly to the amino acid sequence of rat angiotensin. This cDNA, in turn, was used to rescreen cDNA clones that were isolated by initially selecting the rat liver cDNA bank by hybridization with the oligo 1 mixture. One clone thus obtained, designated pRag16, was subjected to nucleotide sequence analysis and verified to contain a nearly full-length cDNA sequence coding for rat angiotensinogen precursor. The deduced amino acid sequence indicates that the precursor molecule consists of angiotensinogen of 453 amino acid residues and a putative signal peptide of 24 amino acid residues. The predicted molecular weight and amino acid composition of angiotensinogen agree well with those determined by using the purified protein. An angiotensin moiety is located at the amino-terminal part of angiotensinogen, preceded directly by the signal peptide and followed by a large carboxyl-terminal sequence that contains two internally homologous sequences and three potential glycosylation sites.

Angiotensinogen (renin substrate), a glycoprotein mainly synthesized by the liver and secreted into the circulating blood, is cleaved by the enzyme renin (EC 3.4.99.19), thus releasing the decapptide angiotensin I. The latter is then cleaved by angiotensin converting enzyme (dipeptidyl carboxypeptidase, peptidylpeptide hydrolase, EC 3.4.15.1), forming the octapeptide angiotensin II. Angiotensin II is the principal biologically active peptide that causes arterial vasoconstriction and stimulates aldosterone secretion from adrenal cortex (reviewed in refs. 1 and 2). Angiotensin formation has also been shown to occur in the central nervous system, where angiotensin appears to be involved in causing thirst and in control of the secretion of vasopressin and corticotropin (ACTH) (3). Through these actions, the renin-angiotensin system plays an important role in the control of blood pressure and hydromineral balance.

Evidence is accumulating that the level of angiotensinogen in the circulation is as important as that of renin in determining the rate of formation of angiotensin and therefore the activity of the renin-angiotensin system (1). Angiotensinogen production is regulated by several factors such as glucocorticoids, estrogens, and angiotensin II (1, 4). However, the mechanisms responsible for the regulation of angiotensinogen production have not been elucidated.

Recently, angiotensinogen has been purified to homogeneity from several species and the estimated molecular weights of the protein molecules of rat, human, and hog angiotensinogens have been reported to be about 50,000 (4, 5). Direct sequence determination of amino-terminal portions of rat and human angiotensinogens has shown that angiotensin I decapptide is located at the amino-terminal part of the precursor molecule (6, 7). However, neither the primary structure of a large remaining portion of the precursor molecule nor its possible biological function have been elucidated.

To study the regulatory mechanisms involved in the production of angiotensinogen and to reveal its entire primary structure, we have constructed an angiotensinogen cDNA clone from rat liver mRNAs. Here, we present the nucleotide sequence of the cloned cDNA which allows us to predict the entire coding sequence for rat angiotensinogen together with the structure of the 3′ untranslated region and most of the 5′ untranslated region.

MATERIALS AND METHODS

Primer Extension Reaction. As glucocorticoids have been reported to stimulate the synthesis of rat liver angiotensinogen (4, 8), total liver RNA was prepared as described (9) from adult male Wistar rats that were administered dexamethasone (10 μg/ml of drinking water) for 4 days. Poly(A)-containing RNA was obtained by twice subjecting the total RNA extracted to oligo(dT)-cellulose chromatography (10). The primer extension reaction was carried out with 50 ng of a mixture of 24 synthetic oligodeoxyribonucleotides labeled with 32P at the 5′ end (3 × 106 cpm/pmol), 150 μg of poly(A)-containing RNA as template, and 525 units of avian myeloblastosis virus reverse transcriptase in a total volume of 150 μl as described (11). The mixture of oligodeoxyribonucleotides was synthesized by the modified triester method (12). The cDNA synthesized was fractionated in a 7% polyacrylamide/7 M urea gel as described (13).

Cloning Procedure. A cDNA library was constructed by the method of Okayama and Berg (14) using 20 μg of poly(A)-containing RNA and 5 μg of the vector/primer DNA. Escherichia coli HB101 or χ1776, transformed and selected for ampicillin resistance as described (15), was first screened by hybridization at 36°C (17) with a mixture of 24 synthetic oligodeoxyribonucleotides described in Results. The hybridization-positive clones obtained were then screened by hybridization at 50°C (16) with a radioactive angiotensin-specific cDNA probe. All of the cloning procedures were conducted in accordance with the guidelines for research involving recombinant DNA molecules issued by the Ministry of Education, Science, and Culture of Japan.

Analytical Procedures. Procedures for restriction endonuclease digestion and 5′-end labeling of DNA were as described
RESULTS AND DISCUSSION

Isolation and Characterization of a cDNA Clone. Our approach for the isolation of a cloned DNA sequence specific for angiotensinogen involved chemical synthesis of two mixtures of 24 oligodeoxyribonucleotides each, 5' T-G-A-A-G-G-T-T-G-A-T 3' (oligo 1) and 5' T-G-A-A-G-G-T-T-G-A-T 3' (oligo 2), as shown in Fig. 1a. These tetradecamers represent all possible complimentary sequences corresponding to the pentapeptide sequence Ile-His-Pro-Phe-His of rat angiotensin I (excluding the third nucleotide residue of the ninth histidine codon for angiotensin I). The oligodeoxyribonucleotide mixtures were used as primers to initiate reverse transcription of rat liver mRNA. The oligo 1 and oligo 2 primers were labeled with 32P at the 5' end and single-stranded cDNAs were extended from these primers with reverse transcriptase. The cDNAs synthesized were analyzed by urea/polyacrylamide gel electrophoresis (Fig. 1b). Similar electrophoretic patterns were observed between the cDNAs extended from the oligo 1 primer and those from the oligo 2 primer, except that a band with a mobility corresponding to a size of approximately 160 nucleotides was missing in the oligo 2-primed cDNAs. To identify the primed cDNA that contained an angiotensin-specific nucleotide sequence, two major transcripts with mobilities corresponding to sizes of approximately 160 and 250 nucleotides were isolated and their sequences were determined by the procedure of Maxam and Gilbert (13). A sequence of 12 nucleotide residues immediately upstream from the primer sequence in the cDNA of approximately 160 nucleotides but not in that of approximately 250 nucleotides coincided precisely with the antisensing sequence corresponding to the amino-terminal amino acid sequence Asp-Arg-Val-Tyr of angiotensin I (data not shown). Thus, the cDNA of approximately 160 nucleotides represents a specific reverse transcription of angiotensinogen mRNA.

Because the radioactivity of the angiotensin-specific cDNA thus obtained did not seem to be enough for direct use as a hybridization probe to screen the rat liver cDNA bank, cDNA clones that carry DNA sequences complementary to one of the oligo 1 tetradecamer sequences were first selected by hybridization with the 32P-labeled oligo 1 mixture. In this way, 17 hybridization-positive clones were isolated from about 70,000 transformants derived from the rat liver cDNA bank. When these 17 hybridization-positive clones were rescreened with the angiotensin-specific cDNA probe, one clone, pRag16, was found to hybridize positively.

Clone pRag16 carried a cDNA insert of about 1,900 nucleotide residues, including the poly(dA)poly(dT) and poly(dC)poly(dG) tails. Restriction endonuclease analysis (Fig. 2) of the cloned DNA indicated that all the restriction sites observed in the nucleotide sequence determined for the primer-extended approximately 160-nucleotide cDNA were located in the 5' terminal portion of the cloned cDNA with direct one-to-one correspondence. In agreement with this result, nucleotide sequence analysis of the cloned cDNA in both directions from one of these restriction sites, Acc I (nucleotide residue 81), showed that the nucleotide sequence surrounding the Acc I site corresponded exactly to the amino-terminal sequence of 17 amino acid residues, including angiotensin I, reported for rat angiotensinogen (6).

Nucleotide Sequence of mRNA Coding for Angiotensinogen Precursor and Assignment of Protein Sequence. Clone pRag16 was subjected to nucleotide sequence analysis according to the strategy indicated in Fig. 2. The nucleotide sequence was determined on both strands of the cDNA for all but 53 and 16 residues corresponding to the 5' and 3' ends of the mRNA, respectively; for these regions, sequence determination on both strands was technically difficult, but the sequence data were reliable. The primary structure of the mRNA coding for the angiotensinogen precursor (preangiotensinogen) was deduced from the 1,762-nucleotide sequence determined (Fig. 3). Angiotensin I is encoded by nucleotide residues 73–102. The 5' terminal sequence of the mRNA deduced from the inserted cDNA se-

Biochemistry: Ohkubo et al.

(18). DNA sequence analysis was carried out by the procedure of Maxam and Gilbert (13). RNA blot hybridization analysis was carried out according to the procedure of Alwine et al. (19); poly(A)-containing RNA isolated from livers of dexamethasone-treated rats was denatured with 1 M glyoxal/50% dimethyl sulfoxide (20), electrophoresed on a 1.5% agarose gel, and transferred to diazobenzyloxymethyl-paper. The hybridization probe was labeled by nick-translation (21).

FIG. 1. (a) Synthetic oligodeoxyribonucleotides used for priming reverse transcription of rat liver mRNA and for probing cloned cDNA for rat angiotensinogen. All possible coding sequences and the corresponding tetradecamers synthesized as two pools (designated oligo 1 and oligo 2) are given for the carboxyl-terminal pentapeptide sequence of rat angiotensin I (fifth to ninth residue of angiotensinogen I). (b) Urea/polyacrylamide gel electrophoresis of cDNAs primed by the mixture of synthetic oligodeoxyribonucleotides. cDNA was synthesized with rat liver poly(A)-containing RNA using two kinds of oligodeoxyribonucleotide mixtures as primers and analyzed on a 7% polyacrylamide/7 M urea gel. An autoradiogram is shown: lane A, oligo 1-primed cDNAs; lane B, oligo 2-primed cDNAs. The sizes of the primed cDNA species given in nucleotides on the right side of the autoradiogram were estimated by comparison with the 5'-end labeled Hae III cleavage products of dX174 DNA, the sizes of which are given in nucleotides on the left side of the autoradiogram.
sequence extended to 159 nucleotides upstream from the nucleotide residue corresponding to the 5' end of the oligodeoxyribonucleotides used for primer extension. This size is closely similar to that of the angiotensin-specific cDNA extended from the oligo 1 primer (approximately 160 nucleotides long) (Fig. 1b). This indicates that the cDNA insert in clone pRag16 contains almost all of the sequence corresponding to the 3'-terminal sequence of the mRNA. On the basis of this finding, the translational initiation site was assigned to the methionine codon AUG at position 1–3, which is the first AUG triplet within the mRNA sequence deduced. This assignment was further supported by the presence of an in-frame nonsense codon UAG (residues −60 to −58) located upstream from the putative initiation methionine. The first 24 amino acid residues starting with the initiation methionine and preceding to the amino acid sequence of angiotensin include a large number of hydrophobic amino acids (16 nonpolar residues) and terminate in a residue with a small neutral side chain, glycine. This characteristic structure would represent the signal peptide similar to that of other secretory proteins (22).

The reading frame shown in Fig. 3 is the only one without multiple termination codons and consists of 1,431 nucleotide residues encoding 477 amino acid residues, which are followed by the termination codon, UGA. Removal of the putative signal peptide by cleavage between the glycine at position 24 and the first amino acid (aspartate-25) of angiotensin would result in a mature angiotensinogen of 453 amino acid residues, with a calculated molecular weight of 49,548. This value agrees well with the molecular weight of 50,000 for the carbohydrate-free angiotensinogen, which was calculated from the glycosylated form of the purified protein (Mr = 56,400) (5). Relative to this, the amino acid sequence deduced from Fig. 3 was found to possess three potential glycosylation sites conforming to the canonical Asn-X-Ser (or threonine) sequence (23) at amino acid positions 47–49, 295–297, and 319–321. The amino acid composition deduced from the nucleotide sequence agrees with the composition reported from amino acid analysis of purified rat angiotensinogen (5) (Table 1). All of these results support the authenticity of the amino acid sequence deduced from the cloned cDNA sequence.

The 3' untranslated region of the mRNA is 270 nucleotides long [excluding the poly(A) tract]. It has been suggested that the hexanucleotide A-A-U-A-A-A found 15–25 nucleotides upstream from the polyadenylation site in most eukaryotic mRNAs serves as a signal for polyadenylation (24). However, instead of this common sequence, 10 consecutive adenosine residues (nos. 1,671–1,680) were found at the corresponding region of rat preangiotensinogen mRNA, although the sequence A-A-U-A-A-A (residues 1,572–1,577) appeared 125 nucleotides upstream from the polyadenylation site. To correlate the length of rat preangiotensinogen mRNA with the size of the cDNA insert, rat liver poly(A)-containing RNA was analyzed by the blot hybridization technique. One band was observed corresponding to a length for the mRNA of approximately 1,850 nucleotides (Fig. 4). Assuming a length for the poly(A) tail of approximately 90 nucleotides, the size of the mRNA estimated is consistent with that of the inserted cDNA sequence in clone pRag16. Thus, the consecutive adenosine residues may be involved in polyadenylation after transcription, although the nucleotide sequence deduced from a single cDNA clone must be viewed cautiously because of the possibility of errors occurring during in vitro cDNA synthesis or DNA cloning.

It should be noted that preangiotensinogen exhibits a molecular organization similar to that of the arginine vasopressin/neurophysin II precursor, in that the biologically active peptide (angiotensin or vasopressin) in both precursor molecules is preceded directly by a signal peptide and followed by a long carboxy-terminal sequence (25). Moreover, as with the neurophysin II sequence, the carboxy-terminal region of angiotensinogen contains two internally homologous sequences located between residues 161–171 and residues 340–350; 6 out of 11 amino acid residues, including a cysteine residue, are present at equivalent positions. Neurophysin II has a carrier function for arginine vasopressin after cleavage from its precursor protein (26). Thus, it would be interesting to investigate whether

---

**Fig. 2.** Strategy for determining the sequence of the cDNA insert in clone pRag16. The map displays only the relevant restriction endonuclease sites, which are identified by numbers indicating the 5'-terminal nucleotide generated by cleavage (for the nucleotide numbers, see Fig. 3). The sequence corresponding to the coding region is indicated by the open box; the closed box indicates the coding region for angiotensin I. The sequence used as a hybridization probe for screening a cDNA bank and for priming the reverse transcription is indicated by the line directly beneath the closed box; the wavy line represents the primer-extended cDNA. The poly(dA)-poly(dT) and poly(dG)-poly(dC) tails are not included in the map. The direction and extent of sequence determinations are shown by horizontal arrows; the sites of 5'-end labeling are indicated by short vertical lines on the arrows.

**Fig. 3 (on next page).** Primary structure of rat preangiotensinogen mRNA. The nucleotide sequence of the mRNA was deduced from that of the cDNA insert in clone pRag16. Nucleotide residues are numbered in the 5' to 3' direction, beginning with the first residue of the AUG triplet coding for the initiation methionine. The nucleotides on the 5' side of residue 1 are indicated by negative numbers. The predicted amino acid sequence of preangiotensinogen is shown above the nucleotide sequence. The amino acid residues are numbered beginning with the initiation methionine. The angiotensin I sequence is boxed with a solid line. The internally homologous amino acid sequences described in the text are underlined.
Biochemistry: Ohkubo et al.

**FIG. 3.** (Legend appears at the bottom of the preceding page.)
Table 1. Comparison of amino acid compositions of rat angiotensinogen

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Predicted from nucleotide sequence</th>
<th>Reported*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspartic acid</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>Asparagine</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>Threonine</td>
<td>31</td>
<td>36</td>
</tr>
<tr>
<td>Serine</td>
<td>37</td>
<td>36</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>25</td>
<td>60</td>
</tr>
<tr>
<td>Glutamine</td>
<td>29</td>
<td>33</td>
</tr>
<tr>
<td>Proline</td>
<td>32</td>
<td>33</td>
</tr>
<tr>
<td>Glycine</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>Alanine</td>
<td>31</td>
<td>33</td>
</tr>
<tr>
<td>Valine</td>
<td>39</td>
<td>35</td>
</tr>
<tr>
<td>Methionine</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Leucine</td>
<td>63</td>
<td>65</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>Lysine</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Histidine</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Arginine</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Cysteine</td>
<td>3 ND</td>
<td>3 ND</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>4 ND</td>
<td>4 ND</td>
</tr>
<tr>
<td>Total</td>
<td>453</td>
<td>466</td>
</tr>
</tbody>
</table>

Predicted numbers of amino acid were calculated by assuming that angiotensinogen is formed by removal of the amino-terminal signal peptide of 24 amino acid residues. ND, not determined.

* Data taken from Hilgenfeldt et al. (5).

The long carboxyl-terminal region of angiotensinogen has some biological role after the release of angiotensin I.

We have recently cloned specific cDNA sequences for both low molecular weight and high molecular weight kininogens from bovine liver mRNA sequences (18). Both kininogens are precursor proteins of bradykinin, which is released by limited proteolysis of kininogens by kallikreins and causes hypotension (references cited in ref. 18). Thus, angiotensin and bradykinin exert opposite effects on the regulation of blood pressure. Because neither the molecular basis responsible for angiotensinogen biosynthesis nor that for kininogen biosynthesis are well understood, the cloned cDNA for angiotensinogen together with those for the kininogens may provide the tools necessary for detailed study of the regulation of blood pressure.

This investigation was supported in part by research grants from the Ministry of Education, Science, and Culture of Japan; the Institute of Physical and Chemical Research; and the Mitsubishi Foundation.