J. D., & Patchett, A. A. (1982) *Biochem. Biophys. Res. Commun. 109*, 1303–1309.
Nishino, N., & Powers, J. C. (1978) *Biochemistry 17*, 2846–2850.
Orlowski, M., & Wilk, S. (1981) *Biochemistry 20*, 4942–4950.
Orlowski, M., Michaud, C., & Chu, T. G. (1983) *Eur. J. Biochem. 135*, 81–88.
Pachett, A. A., Harris, E., Tristram, E. W., Wyvratt, M. J., Wu, M. T., Taub, D., Peterson, E. R., Ikeler, T. J., ten Broeke, J., Payne, L. G., Ondeyka, D. L., Thorsett, E. D., Greenlee, W. J., Lohr, N. S., Hoffsommer, R. D., Joshua, H., Ruyle, W. V., Rothrock, J. W., Aster, S. D., Maycock, A. L., Robinson, F. M., Hirschmann, R., Sweet, C. S., Ulm, E. H., Gross, D. M., Vassil, T. C., & Stone, C. A. (1980) *Nature (London) 288*, 280–283.
Pfleiderer, G. (1970) *Methods Enzymol. 19*, 514–521.
Roques, B. P., Fournie-Zaluski, M. C., Soroca, E., Lecomte, J. M., Malfroy, B., Llorens, C., & Schwartz, J. C. (1980) *Nature (London) 288*, 286–288.
Schechter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157–162.
Sullivan, S., Akil, H., & Barchas, J. D. (1978) *Commun. Psychopharmacol. 2*, 525–531.
Wolfenden, R. (1972) *Acc. Chem. Res. 5*, 10–18.

Primary Structure of Human Preangiotensinogen Deduced from the Cloned cDNA Sequence†

Ryoichiro Kageyama, Hiroaki Ohkubo, and Shigetada Nakanishi*

ABSTRACT: Cloned cDNA sequences for human preangiotensinogen have been isolated from a human liver cDNA library by hybridization with a restriction fragment derived from a previously cloned cDNA for rat preangiotensinogen. Analyses by nucleotide sequence determination, S1 nuclease mapping, and RNA blot hybridization indicate that human preangiotensinogen is encoded by two mRNAs that differ only in the length of the 3′-untranslated region. The deduced amino acid sequence shows that the mature angiotensinogen consists of 452 amino acid residues with the angiotensin sequence at its amino-terminal portion. Two potential initiation sites have been discussed. These are the methionine codon located at the position exactly corresponding to the initiation site of rat preangiotensinogen mRNA and an additional methionine codon positioned nearest the 5′ end of the mRNA. The amino acid sequences starting at either of the initiation sites and preceding the angiotensin sequence constitute a large number of hydrophobic amino acid residues, thus representing the signal peptide characteristic of the secretory proteins. Human and rat preangiotensinogens show that 63.6% of the amino acid positions of the two proteins are identical. However, the amino-terminal portions directly distal to angiotensin I diverge markedly between the two proteins and differ in their possible glycosylation sites. These structural differences may contribute to the known species specificity exhibited by renin.

The renin–angiotensin system plays an important role in the regulation of blood pressure and hydromineral balance [see reviews from Reid et al. (1978) and Skeggs et al. (1980)]. Activation of this system is initiated by the release of angiotensin I from its precursor angiotensinogen by processing with the enzyme renin (EC 3.4.99.19). Angiotensin-converting enzyme (dipeptidyl carboxypeptidase, peptidyldipeptide hydrolase, EC 3.4.15.1) then cleaves a dipeptide from angiotensin I to form the octapeptide angiotensin II. Angiotensin II is the principal biologically active peptide that causes arteriolar vasoconstriction and stimulates aldosterone secretion. Several lines of evidence indicate that angiotensinogen is as important as renin in determining the rate of formation of angiotensin and therefore the activity of the renin–angiotensin system (Reid et al., 1978).

A number of inhibitors of angiotensin-converting enzyme have recently been developed, and their wide applicability for treating hypertensive patients suggests that the renin–angiotensin system may involve the pathogenesis of various forms of human hypertension including essential hypertension [see review from Ondetti & Cushman (1982)]. Human angiotensinogen has been purified to homogeneity in several laboratories [see review from Tewksbury (1983)], and its amino-terminal amino acid sequence up to the 25 amino acid residues has been reported (Tewksbury et al., 1981). However, the primary structure of a large carboxyl-terminal portion of human angiotensinogen remains to be determined. Furthermore, little is known about the regulatory mechanism responsible for the biosynthesis of human angiotensinogen, although it has been reported that the concentration of human plasma angiotensinogen varies under various physiological and pathological conditions (Reid et al., 1978).

We have previously reported the whole primary structure of rat preangiotensinogen by determining the nucleotide sequence of cloned DNA complementary to its mRNA (Ohkubo et al., 1983). In the present study, the construction and sequence analysis of cDNAs for human preangiotensinogen have been undertaken as an initial approach to investigate the regulatory mechanism involved in the biosynthesis of human angiotensinogen and to study the possible involvement of the renin–angiotensin system in the pathogenesis of human hypertension. We here report the entire amino acid sequence of human preangiotensinogen deduced from the nucleotide

† From the Institute for Immunology, Kyoto University Faculty of Medicine, Kyoto 606, Japan. *Received January 12, 1984.* This investigation was supported in part by research grants from the Ministry of Education, Science and Culture of Japan and the Institute of Physical and Chemical Research.

0006-2960/84/0423-3603$01.50/0 © 1984 American Chemical Society

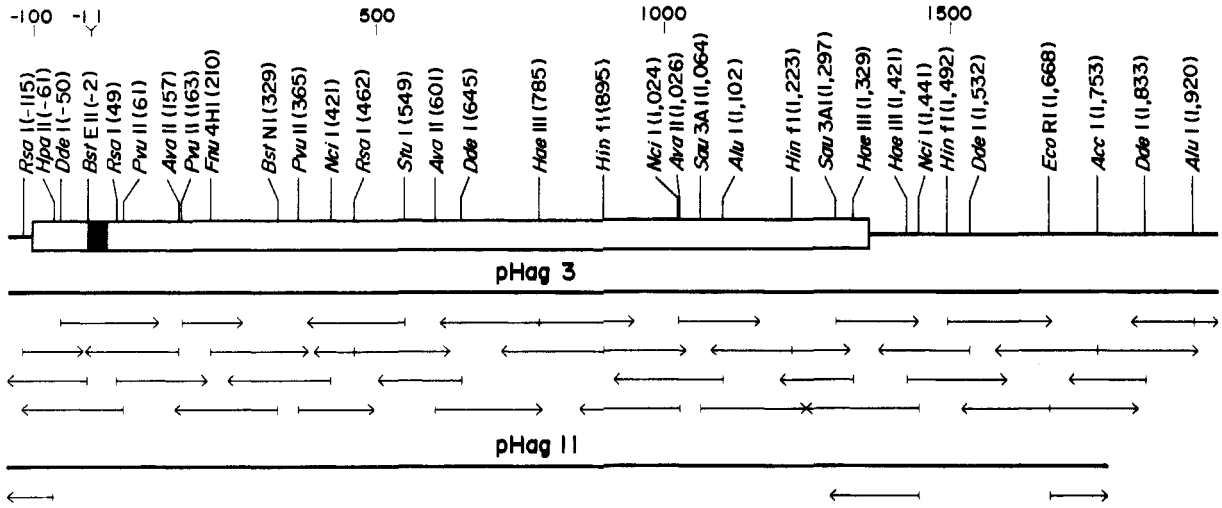

FIGURE 1: Strategy for determining the sequences of the cDNA inserts in clones pHag 3 and pHag 11. The map displays only the relevant restriction endonuclease sites, which are identified by numbers indicating the 5′-terminal nucleotide generated by cleavage (for the nucleotide numbers, see Figure 2). The sequence corresponding to the coding region is indicated by the outlined box; the solid box indicates the coding region for angiotensin I. 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 under each clone used; the sites of 5′-end labeling are indicated by short vertical lines at the ends of arrows.

sequence and its comparison with the sequence of rat preangiotensinogen.

Experimental Procedures

Materials. Reagents were obtained as follows: [α-^{32}P]-dCTP (3000 Ci/mmol) and [γ-^{32}P]ATP (5000 Ci/mmol) from Amersham; oligo(dT)–cellulose (type 7) and *Escherichia coli* ribonuclease H from P-L Biochemicals; *E. coli* DNA polymerase I and *E. coli* ligase from New England Nuclear; S1 nuclease from Bethesda Research Labs; the Klenow fragment of *E. coli* DNA polymerase I from Boehringer-Mannheim; aminobenzyloxymethyl paper from Schleicher & Schuell; terminal deoxynucleotidyl transferase, bacterial alkaline phosphatase, and T4 polynucleotide kinase from Takara Shuzo Co.; and restriction endonucleases from Takara Shuzo Co., Bethesda Research Labs, and New England Biolabs. The mixture of oligodeoxyribonucleotides used for identifying the angiotensin-coding sequence was the same as described for the isolation of the cDNA clone for rat preangiotensinogen (Ohkubo et al., 1983); the oligodeoxyribonucleotides were synthesized by the modified triester methods (Hirose et al., 1978).

Cloning Procedures. Isolation of cloned cDNAs for human preangiotensinogen was performed according to the procedure described previously (Nakanishi et al., 1978; Ohkubo et al., 1983). In brief, total RNA was extracted from a human liver as described (Chirgwin et al., 1979), and poly(A)-containing RNA was isolated by subjecting the total RNA extracted to oligo(dT)–cellulose chromatography (Aviv & Leder, 1972). A human liver cDNA library was constructed by the method of Okayama & Berg (1982) from 17 μg of poly(A)-containing RNA and 5.6 μg of the vector-primer DNA. *E. coli* HB101 was transformed and selected for ampicillin resistance (Morrison, 1979). About 70000 transformants obtained were screened by hybridization (Hanahan & Meselson, 1980) at 55 °C with the 1097-base-pair *Acc*I fragment derived from the cloned cDNA for rat preangiotensinogen [residues 81–1177; see Figure 3 of Ohkubo et al. (1983)], labeled by nick translation with [α-^{32}P]dCTP. 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 (Ohkubo et al., 1983). The 3′-end labeling was carried out as described by Challberg & Englund (1980), except that the Klenow fragment of DNA polymerase I was used instead of T4 DNA polymerase. Nick translation was conducted with use of [α-^{32}P]dCTP (Rigby et al., 1977). DNA sequence analysis was carried out by the procedure of Maxam & Gilbert (1980). DNA blot hybridization analysis was performed according to the procedure of Southern (1975); when the mixture of synthetic oligodeoxyribonucleotides was used as a probe, hybridization and filter washing were carried out at 36 °C (Wallace et al., 1979). RNA blot hybridization analysis was carried out by the procedure of Alwine et al. (1977); poly(A)-containing RNA was denatured with 1 M glyoxal/50% dimethyl sulfoxide (McMaster & Carmichael, 1977), electrophoresed on a 1.5% agarose gel, and transferred to diazobenzyloxymethyl paper. For S1 nuclease mapping analysis, an appropriate probe was denatured, hybridized to total liver poly(A)-containing RNA in 80% formamide at 40 or 50 °C for 3 h, and digested with S1 nuclease (Berk & Sharp, 1978). The products of S1 digestion and the size markers were electrophoresed on 7 M urea/4% polyacrylamide gel.

Results and Discussion

Isolation of cDNA Clones. cDNA clones specific for human preangiotensinogen were isolated from a human liver cDNA library by hybridization with a cDNA fragment containing most of the protein-coding sequence for rat preangiotensinogen. The cDNA library was constructed with the plasmid DNA vector of Okayama & Berg (1982) by using poly(A)-containing RNA extracted from human liver. Ten hybridization-positive clones were isolated from approximately 70000 transformants. Upon restriction enzyme analysis, these 10 clones were found to contain cDNA inserts that shared common restriction fragments. Angiotensin I constitutes the sequence of the first 10 amino acids for all angiotensinogens studied to date, including human angiotensinogen (Tewksbury, 1983). Therefore, cDNA inserts that contained a nearly full length of cDNA sequence were identified from the 10 clones by the blot hybridization method (Southern, 1975) with the use of a mixture of chemically synthesized oligodeoxyribonucleotides, 5′-TGRAANGGRTG$^{R}_{T}$AT-3′ (R = G or A and N = A, G, C, or T), the sequences of which are comple-

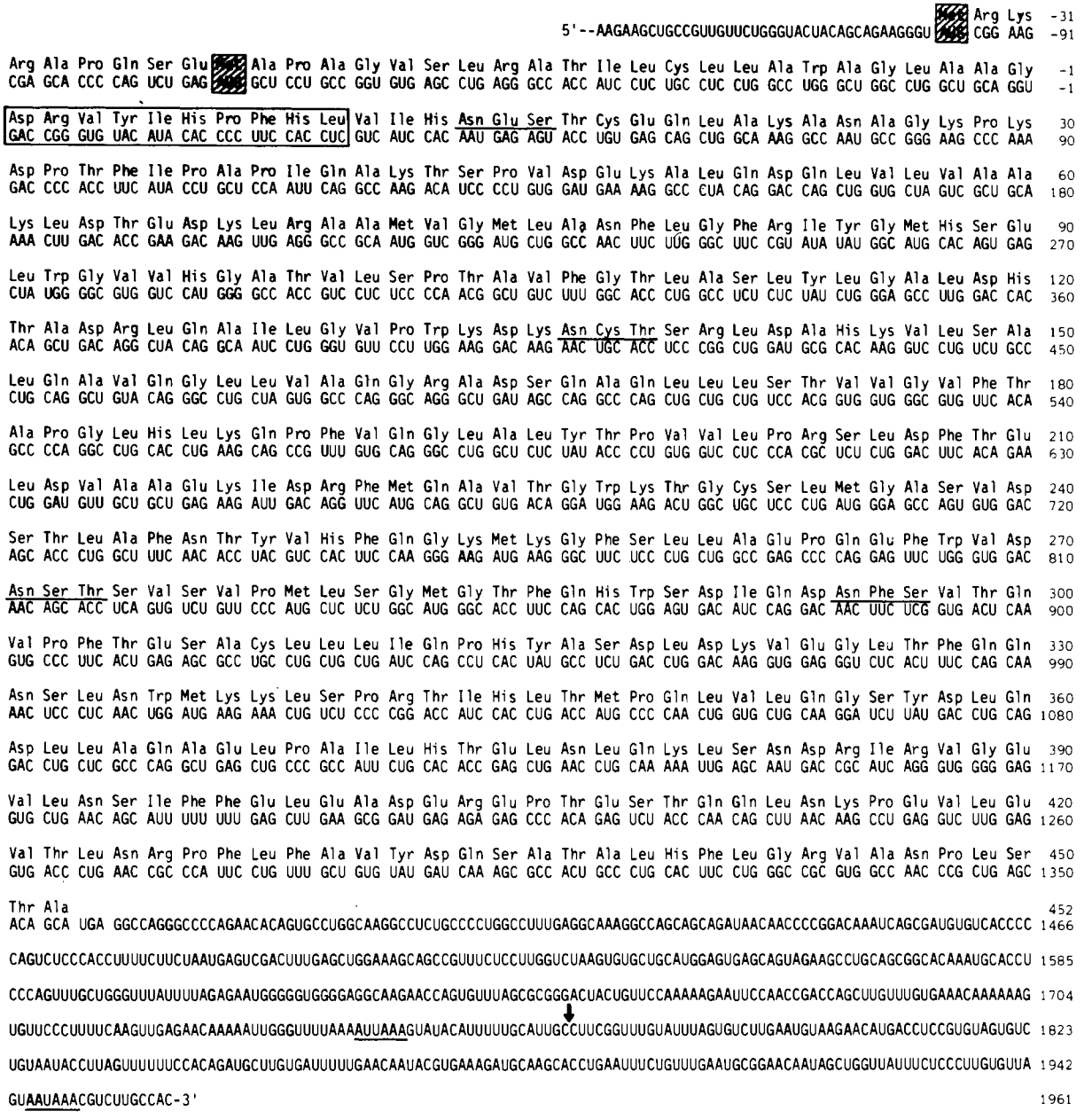

FIGURE 2: Primary structure of human preangiotensinogen mRNA. The nucleotide sequence of the mRNA was deduced from that of the cDNA insert of clone pHag 3. Nucleotide residues are numbered in the 5′ to 3′ direction, beginning with the first residue coding for the angiotensin sequence. The nucleotides on the 5′ side of residue 1 are indicated by negative numbers. The 3′ sequence deduced from the cDNA insert of clone pHag 11 is included up to the position indicated by the arrow. At the 5′-proximal region of this mRNA, one nucleotide (residue −138) is lacking, probably due to the incomplete synthesis of cDNA. The two possible polyadenylation signals AAUAAA and AUUAAA are underlined. The predicted amino acid sequence of preangiotensinogen is shown above the nucleotide sequence. The amino acid residues are numbered beginning with the first residue of angiotensin. The angiotensin I sequence is boxed with a solid line. The two possible initiation sites are indicated by the hatched boxes. The possible glycosylation sites conforming to the Asn-X-$\frac{Ser}{Thr}$ sequence are underlined at the amino acid sequence.

mentary to the codons specifying the carboxyl-terminal sequence Ile-His-Pro-Phe-His of angiotensin I; restriction fragments derived from the 10 clones were electrophoresed on agarose gels, transferred to nitrocellulose filters, and hybridized to the ^{32}P-labeled oligodeoxyribonucleotide mixture. It was found that two clones, pHag 3 and pHag 11, contained a hybridization-positive restriction fragment (data not shown) and shared a common restriction map, except for the 3′-terminal region of the cDNA sequences (Figure 1). The whole nucleotide sequence of the cDNA insert in clone pHag 3 as well as the partial sequence of the insert in clone pHag 11 was determined according to the strategy indicated in Figure 1.

Assignment of Amino Acid Sequence of Human Preangiotensinogen. The primary structure of human preangiotensinogen mRNA was deduced from the 2099-nucleotide sequence determined for the cDNA of clone pHag 3. The results are summarized in Figure 2. Angiotensin I is encoded at nucleotide residues 1–30 and is followed by a large open-reading frame that terminates at amino acid residue 452 with the termination codon UGA. The human mature angiotensinogen is thus composed of 452 amino acid residues, with a calculated M_r of 49 770. This value agrees well with the M_r of 52 800 for the carbohydrate-free angiotensinogen, which was calculated from the glycosylated form of the purified protein (M_r 61 400) (Tewksbury, 1983). The deduced sequence following the angiotensin I moiety also coincides with the sequence of the 15 amino acid residues reported for human angiotensinogen, except that the residues at positions 18 and 25 are Cys and Asn instead of Ser and Asp, respectively (Tewksbury et al., 1981). The residue at position 14 was

blanked in the protein sequence determination for human angiotensinogen but was suggested to be Asn that was linked to a carbohydrate moiety (Tewksbury et al., 1981). In fact, this residue was found to be Asn. There are three other potential glycosylation sites conforming to the canonical Asn-X-$^{Ser}_{Thr}$ sequence (Marshall, 1974) at positions 137–139, 271–273, and 295–297. The amino acid composition deduced from the nucleotide sequence agrees with the composition reported from amino acid analysis of purified human angiotensinogen (Tewksbury, 1983). All of these results support the authenticity of the amino acid sequence deduced from the cloned cDNA sequence.

Two possible initiation sites can be predicted from the nucleotide sequence. One is the methionine codon at position −24, which is located at the position exactly corresponding to the initiation site of rat preangiotensinogen mRNA (Ohkubo et al., 1983). The other is an additional methionine codon at position −33, which is located nearest the 5′ end of the mRNA. The presence of the two AUG triplets at the 5′-proximal region of the mRNA was also confirmed by sequence determination of clone pHag 11 (Figure 1). The eukaryotic translation generally begins at the first AUG codon at the 5′ end of the mRNA (Kozak, 1981). However, there are several eukaryotic mRNAs in which translation appears to be initiated at an internal AUG triplet (Dhruva et al., 1980; Kozak, 1981; Hendy et al., 1981) or which may possess two functional initiation codons (Jay et al., 1981; Bos et al., 1981). Furthermore, Lomedico & McAndrew (1982) have shown that the sequence surrounding AUG, and not simply its position relative to the 5′ end of mRNA, is important in determining initiation efficiency, on the basis of the introduction of additional in-frame AUG codons in the 5′-untranslated region of the rat insulin II gene by in vitro mutagenesis. The second but not the first AUG codon relative to the 5′ end of the human preangiotensinogen mRNA is contained in the sequence $^{A}_{G}$XXAUGG, which is most frequently found in the functional initiation site (Kozak, 1981). Thus, the possible use of the second AUG codon as an initiation site canot be excluded. Which one or whether both of the AUG triplets can function as initiation codons, however, remains to be determined.

The amino acid sequences starting with either of the two initiation methionines and proceeding to the angiotensin sequence include a large number of hydrophobic amino acids characteristic of the secretory proteins (Blobel & Dobberstein, 1975). Thus, either or both of these sequences of 24 and 33 amino acid residues would represent the signal peptide involved in the process of the protein secretion.

Presence of Two Human Preangiotensinogen mRNAs. The nucleotide sequences deduced from the two clones pHag 3 and pHag 11 differ in the length of the extreme 3′ terminus of the 3′-untranslated region. Because the termination codon UGA is located at the same position between the nucleotide sequences determined for pHag 3 and pHag 11 (Figure 1), the 3′-untranslated region of the mRNA is composed of either 605 or 413 nucleotides [excluding the poly(A) tract]. The presence of two such preangiotensinogen mRNAs in the human liver was confirmed by the S1 nuclease mapping analysis as well as the RNA blotting method (Figures 3 and 4). When the cDNA probe corresponding to the 3′ sequence of the mRNA was hybridized to total liver poly(A)-containing RNA, S1 nuclease digestion of the DNA–RNA hybrid formed yielded two bands upon urea/polyacrylamide gel electrophoresis (Figure 3a, probe C). The size difference of the two bands, estimated from their mobilities on the electrophoretic gel, was

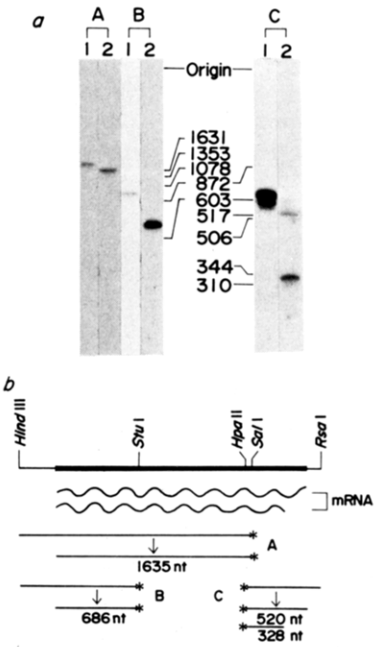

FIGURE 3: Identification of two preangiotensinogen mRNAs by S1 nuclease mapping. Autoradiographs of S1 nuclease mapping analysis are shown in (a). Lane 1 and lane 2 show the cDNA probe without S1 digestion and the S1 digestion products of the RNA–DNA hybrids, respectively. The size markers used were the 5′-end-labeled *Hin*fI cleavage products of pBR322 and the 5′-end-labeled *Hae*III cleavage products of ϕX174. Probes A, B, and C were isolated from clone pHag 3. Probe A was the *Sal*I–*Hin*dIII fragment containing the 1635-nucleotide cDNA, the 22-nucleotide oligo(dC) tail, and the 272-nucleotide vector DNA. Probe B was the *Stu*I–*Hin*dIII fragment containing the 686-nucleotide cDNA and its flanking DNA described in probe A. Probe C was the *Hpa*II–*Rsa*I fragment containing the 520-nucleotide cDNA and the approximately 60–140-nucleotide oligo(dT) tail and the 2-nucleotide vector DNA; the several bands of probe C observed in lane 1 were due to the heterogeneity of the length of the poly(dT) tail, which probably occurred during the amplification of clone pHag 3. In (b), the probes used in this experiment and the predicted S1 nuclease resistant DNA fragments are schematically illustrated. The upper line displays the restriction map relevant to this experiment; the thick line indicates the cDNA sequence. The asterisks denote the sites of 5′-end labeling (probes A and B) and 3′-end labeling (probe C). Nucleotide numbers (nt) indicate the predicted sizes of ^{32}P-labeled cDNA fragments that should be protected from S1 nuclease digestion. Note that the duplex formation between the poly(A) tail of the mRNAs and the oligo(dT) tail of the cDNA was digested by S1 nuclease under the experimental condition used.

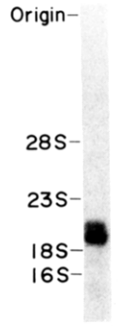

FIGURE 4: Identification of two preangiotensinogen mRNAs by blot hybridization analysis. A sample of human liver poly(A)-containing RNA (4.8 μg) was analyzed. The hybridization probe used as the 1262-base-pair *Bst*EII fragment (residues −2 to 1260) derived from clone pHag 3. The size markers used were human and *E. coli* rRNA.

consistent with the difference in the lengths of the 3′-untranslated region deduced from the nucleotide sequences determined. In addition, RNA blotting analysis showed the presence of two bands, the mobilities of which corresponded to sizes of approximately 2200 and 2000 nucleotides (Figure

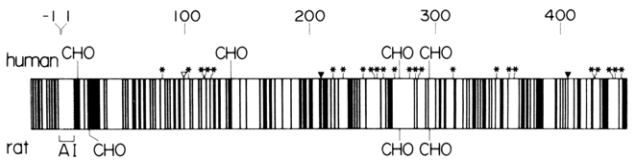

FIGURE 5: Comparison of the amino acid sequences of human and rat preangiotensinogens. Vertical lines indicate positions in which the two sequences have the different amino acids. Asterisks denote positions in which all five sequences of human and rat angiotensinogens, human α_1-antitrypsin (Carrell et al., 1982), human antithrombin III (Chandra et al., 1983), and chicken ovalbumin (McReynolds et al., 1978) have the same amino acid; for this comparison, the human preangiotensinogen sequence was added to the sequence alignments proposed by Doolittle (1983) for the latter four proteins. The angiotensin I sequence (AI) is indicated by the dotted box. The white and the black triangles indicate the positions in which the human sequence possesses or lacks the amino acid as compared with the rat sequence, respectively. Amino acid numbers of human preangiotensinogen are given above the structure. Note that the amino-terminal portion of nine amino acid residues (residues −33 to −25) starting with the 5′-proximal methionine codon is deleted from the human preangiotensinogen sequence, because the corresponding sequence is lacking in rat preangiotensinogen. CHO indicates a position of the asparagine residue that may serve as a possible glycosylation site.

4). In contrast, when the cDNA sequences corresponding to the 5′ sequence of the mRNA were used as hybridization probes, a single band appeared upon S1 nuclease mapping analysis (Figure 3a, probes A and B). Therefore, it is concluded that the two mRNAs share a common sequence at the 5′ side and differ only in the length of the 3′-untranslated regions.

The sequences AAUAAA and AUUAAA were found to be located at 12 and 20 nucleotides upstream from the polyadenylation sites of the larger and the smaller mRNAs, respectively. It has been suggested that the sequence AAUAAA and its variant AUUAAA, both of which are present at a similar position of other eukaryotic mRNAs, serve as a signal for polyadenylation after transcription (Proudfoot & Brownlee, 1976; Nevins, 1983). Because preliminary DNA blotting analysis of total human liver DNA supports the notion that human preangiotensinogen is encoded by a single gene, we postulate that the two preangiotensinogen mRNAs arise from the single preangiotensinogen gene by the utilization of the two polyadenylation signals. The generation of multiple mRNAs from a single gene has been reported for several mRNAs such as mouse dihydrofolate reductase mRNA (Setzer et al., 1980) and mouse α-amylase mRNA (Tosi et al., 1981). It is also noteworthy that, in both S1 mapping and RNA blotting analysis, the hybridization signal of the smaller mRNA is roughly 3 times more intense than that of the larger mRNA. This suggests that the sequence of AUUAAA is used more effectively than that of AAUAAA as a polyadenylation signal for the generation of the preangiotensinogen mRNAs.

Comparison of Amino Acid Sequences of Human and Rat Preangiotensinogens. In our previous study (Ohkubo et al., 1983), the primary structure of rat preangiotensinogen was elucidated by determining the nucleotide sequence of the cloned cDNA. The amino acid sequences of human and rat preangiotensinogens were compared, and the comparison is schematically illustrated in Figure 5. In this figure, the positions of the amino acids that differ between the two proteins are indicated by vertical lines. To align the two proteins, a line is introduced into the human protein at the position between 208 and 209, and between 405 and 406, because these positions have an additional amino acid residue in the rat protein. In contrast, the amino acid residue at position 99 is lacking in the rat sequence when compared with the human counterpart. It has been found that overall, 63.6% of the amino acid positions in the two proteins are identical. Furthermore, the homology between the two proteins increases to 80.3% if the amino acids are compared on the basis of their belonging to one of four groups having a characteristic side chain: hydrophobic (Phe, Leu, Ile, Met, Val, Pro, Ala, Tyr, Trp), polar (Ser, Thr, Gln, Asn, Cys, Gly), acidic (Glu, Asp), or basic (His, Lys, Arg). Among these four groups, amino acids of the hydrophobic group (87.2%) are more conserved than those of the polar (78.4%), acidic (69.6%), or basic group (65.3%). The proline residue and glycine residue that frequently appear in the β-turn are also well conserved between the two mature angiotensinogens; 20 of the 23 proline residues (87.0%) and 19 of the 25 glycine residues (76.0%) in the human sequence are found at the corresponding positions in the aligned rat sequence. Reflecting this homology, the hydropathy profiles of the two proteins (Kyte & Doolittle, 1982) show the very similar patterns as indicated in Figure 6.

Renin cleaves a Leu–Val bond in human angiotensinogen whereas this enzyme processes a Leu–Leu bond in the other mammalian angiotensinogens studied, including rat angiotensinogen (Skeggs et al., 1957; Tewksbury et al., 1979). The comparison of the amino acid sequences of the two angiotensinogens further indicates that there is little homology in the amino-terminal portions directly distal to angiotensin I (Figure 5). Only 5 out of 21 amino acids (23.8%) are homologous in the region between amino acid residues 11 and 31. Furthermore, the glycosylation site conforming to the Asn-X-Ser sequence is located at positions 14–16 in human angiotensinogen, whereas it is found at positions 23–25 in rat angiotensinogen. Human angiotensinogen is cleaved only by human or primate renin whereas most other mammalian angiotensinogens are cleaved by most mammalian renins, including human renin (Braun-Menendez et al., 1946). Thus, it is possible that the marked substitutions seen at the amino-terminal portion of the angiotensinogens as well as the different attachment sites for the asparagine-linked carbohy-

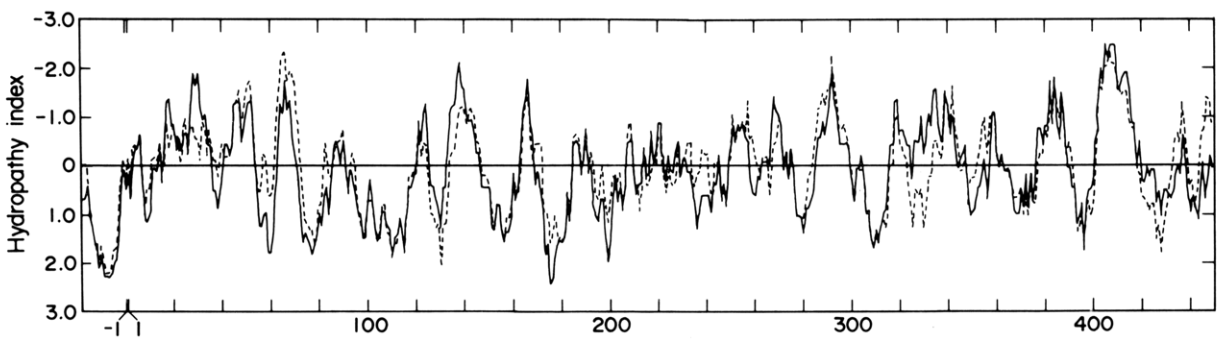

FIGURE 6: Comparison of hydropathy profiles of human and rat preangiotensinogens. The averaged hydropathy index of a nonapeptide is plotted against the amino acid number of human preangiotensinogen. The solid and dotted lines represent the human and rat sequences, respectively.

drate unit may contribute to the species specificity of renin.

Recently, Doolittle (1983) reported a computer analysis that revealed a significant resemblance of rat angiotensinogen to α_1-antitrypsin, which itself is known to be related to antithrombin III and ovalbumin (Hunt & Dayhoff, 1980). He found that there are no sequences of the α_1-antitrypsin family corresponding to the amino-terminal one-fifth of rat angiotensinogen but that the carboxyl-terminal four-fifths of rat angiotensinogen is significantly related to the sequences of the α_1-antitrypsin family. An alignment of the two preangiotensinogens with the three proteins of the α_1-antitrypsin family indicates that all five sequences have the same amino acids at 24 out of the 31 positions in which rat angiotensinogen was found to be identical to the α_1-antitrypsin family (Doolittle, 1983). The 24 positions having the common amino acids are mainly localized at the regions where the sequences of human and rat angiotensinogens are well conserved with each other (Figure 5). However, two relatively large regions have been identified near the middle (residues 138–191) and at the carboxyl-terminal portion (residues 387–423), where the sequences of human and rat angiotensinogens are highly homologous but do not match the consensus sequence of the α_1-antitrypsin family. Further structural comparisons of these two regions have revealed that, at more than a third of the positions, the former sequence has residues in common with at least one of the three proteins of the α_1-antitrypsin family, whereas the latter sequence rarely matches either of these three proteins. Thus, these comparisons suggest that angiotensinogen may resemble the α_1-antitrypsin family in the tertiary structure formed by the sequence composed of the amino acids common to angiotensinogen and the other three proteins but may differ from the α_1-antitrypsin family in the amino-terminal and carboxyl-terminal structures. Interestingly, the carboxyl-terminal portion described here corresponds to the amino acid sequences containing the active sites of α_1-antitrypsin and antithrombin III (Travis & Salvesen, 1983).

We have recently cloned the specific cDNA sequence for human renal renin (Imai et al., 1983). Because little is known about the molecular basis for the regulation of the human renin–angiotensin system, the cloned cDNA for human angiotensinogen together with that for human renin may provide the tools necessary for a more detailed study of the regulation of blood pressure in humans.

Acknowledgments

We sincerely thank Dr. Takashi Miyata for computer analysis.

Registry No. Preangiotensinogen, 89800-05-5; angiotensinogen, 11002-13-4; angiotensin, 1407-47-2; DNA (human angiotensin I messenger RNA complimentary), 90669-95-7; preangiotensinogen (human reduced), 90669-94-6; angiotensinogen (human reduced), 90669-93-5; angiotensin I (human), 484-42-4; 5-L-isoleucineangiotensin II, 4474-91-3.

References

Alwine, J. C., Kemp, D. J., & Stark, G. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5350–5354.

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Berk, A. J., & Sharp, P. A. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 1274–1278.

Blobel, G., & Dobberstein, B. (1975) *J. Cell Biol. 67*, 852–862.

Bos, J. L., Polder, L. J., Bernards, R., Schrier, P. I., van den Elsen, P. J., van der Eb, A. J., & van Ormondt, H. (1981) *Cell (Cambridge, Mass.) 27*, 121–131.

Braun-Menendez, E., Fasciolo, J. C., Leloir, L. F., Munoz, J. M., & Taquini, A. C. (1946) *Renal Hypertension*, p 134, Thomas, Springfield, IL.

Carrell, R. W., Jeppsson, J. O., Laurell, C. B., Brennan, S. O., Owen, M. C., Vaughan, L., & Boswell, D. R. (1982) *Nature (London) 298*, 329–334.

Challberg, M. D., & Englund, P. T. (1980) *Methods Enzymol. 65*, 39–43.

Chandra, T., Stackhouse, R., Kidd, V. J., & Woo, S. L. C. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1845–1848.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Dhruva, B. R., Shenk, T., & Subramanian, K. N. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 4514–4518.

Doolittle, R. F. (1983) *Science (Washington, D.C.) 222*, 417–419.

Hanahan, D., & Meselson, M. (1980) *Gene 10*, 63–67.

Hendy, G. N., Kronenberg, H. M., Potts, J. T., & Rich, A. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 7365–7369.

Hirose, T., Crea, R., & Itakura, K. (1978) *Tetrahedron Lett. 28*, 2449–2452.

Hunt, L. T., & Dayhoff, M. O. (1980) *Biochem. Biophys. Res. Commun. 95*, 864–871.

Imai, T., Miyazaki, H., Hirose, S., Hori, H., Hayashi, T., Kageyama, R., Ohkubo, H., Nakanishi, S., & Murakami, K. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 7405–7409.

Jay, G., Nomura, S., Anderson, C. W., & Khoury, G. (1981) *Nature (London) 291*, 346–349.

Kozak, M. (1981) *Nucleic Acids Res. 9*, 5233–5252.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Lomedico, P. T., & McAndrew, S. J. (1982) *Nature (London) 299*, 221–226.

Marshall, R. D. (1974) *Biochem. Soc. Symp. 40*, 17–26.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

McMaster, G. K., & Carmichael, G. G. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4835–4838.

McReynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Givol, D., Fields, S., Robertson, M., & Brownlee, G. G. (1978) *Nature (London) 273*, 723–728.

Morrison, D. A. (1979) *Methods Enzymol. 68*, 326–331.

Nakanishi, S., Inoue, A., Kita, T., Numa, S., Chang, A. C. Y., Cohen, S. N., Nunberg, J., & Schimke, R. T. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 6021–6025.

Nevins, J. R. (1983) *Annu. Rev. Biochem. 52*, 441–466.

Ohkubo, H., Kageyama, R., Ujihara, M., Hirose, T., Inayama, S., & Nakanishi, S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2196–2200.

Okayama, H., & Berg, P. (1982) *Mol. Cell. Biol. 2*, 161–170.

Ondetti, M. A., & Cushman, D. W. (1982) *Annu. Rev. Biochem. 51*, 283–308.

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.

Reid, I. A., Morris, B. J., & Ganong, W. F. (1978) *Annu. Rev. Physiol. 40*, 377–410.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., & Berg, P. (1977) *J. Mol. Biol. 113*, 237–251.

Setzer, D. R., McGrogan, M., Nunberg, J. H., & Schimke, R. T. (1980) *Cell (Cambridge, Mass.) 22*, 361–370.

Skeggs, L. T., Kahn, J. R., Lentz, K., & Shumway, N. P. (1957) *J. Exp. Med. 106*, 439–453.

Skeggs, L. T., Dorer, F. E., Levine, M., Lentz, K. E., & Kahn, J. R. (1980) in *The Renin-Angiotensin System* (Johnson, J. A., & Anderson, R. R., Eds.) pp 1–27, Plenum, New York.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.

Tewksbury, D. A. (1983) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 42*, 2724–2728.
Tewksbury, D. A., Dart, R. A., & Travis, J. (1979) *Circulation 59/60* (Suppl. II), 132 (Abstr. 513).
Tewksbury, D. A., Dart, R. A., & Travis, J. (1981) *Biochem. Biophys. Res. Commun. 99*, 1311–1315.
Tosi, M., Young, R. A., Hagenbüchle, O., & Schibler, U. (1981) *Nucleic Acids Res. 9*, 2313–2323.
Travis, J., & Salvesen, G. S. (1983) *Annu. Rev. Biochem. 52*, 655–709.
Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., & Itakura, K. (1979) *Nucleic Acids Res. 6*, 3543–3557.

Condensation of Bacteriophage ϕW14 DNA of Varying Charge Densities by Trivalent Counterions†

Jülyet Aksiyote Benbasat*

ABSTRACT: Bacteriophage ϕW14 DNA carries the hypermodified, positively charged (2+) base α-putrescinylthymine (putThy) and consequently exhibits a decreased average linear charge density compared to the conventional B-form DNA helix. Noting that the unusual physical characteristics may contribute to the collapse properties of this DNA and facilitate the exceptionally high density of packaging of its genome in ϕW14, I used total intensity light scattering to determine in vitro the critical concentrations of spermidine (Spd, 3+) required to induce the cooperative, monomolecular collapse of wild-type and mutant ϕW14 DNA samples and quasi-elastic light scattering to compare the dynamic characteristics of the compacted particles. The DNA samples carried various percentages of the modified base with average charge spacings ranging from 1.3 to 2.2 Å in comparison to T4 phage DNA (1.7 Å). The results are analyzed and discussed both from a general theoretical point of view according to the counterion condensation theory of Manning [Manning, G. S. (1978) *Q. Rev. Biophys. 11*, 179–246] and from the more specialized aspect of DNA packaging in ϕW14. In accord with theory, DNAs of lower charge density require a considerably higher critical counterion concentration (up to 118 μM Spd), whereas the outside diameter of the toroidal condensates, which they form, varies only marginally. Specific ion effects were probed by substituting hexaamminecobalt(III) (Hc, 3+) for Spd. Hc appears to be more efficient than Spd: it induces the collapse of all DNA samples at only one-sixth the critical concentration of Spd, and its condensates are 30% smaller (1072–1142 Å vs. 744–800 Å) except for wild-type ϕW14 DNA, which forms Hc-collapsed particles indistinguishable from Spd-induced condensates. Collapse occurs, again with the exception of wild-type ϕW14 DNA, when ~89% of the charges on each DNA are neutralized by territorially bound Spd. I conclude that the driving force for condensation clearly is a function of the charge density of the DNA and that the charge distribution may be an important factor in determining the degree of neutralization at which collapse becomes possible. The sample with the lowest charge density, wild-type ϕW14 DNA, does not follow the trends set by the other members of the series. The possibility is discussed that lowering the charge density by covalent modification beyond a threshold may result in the compression of the DNA double helix, thus allowing more information to be carried on a genome, the packagable length of which is determined by the encapsidation mechanism.

The polyamine-induced cooperative transition (Post & Zimm, 1979) of the DNA[1] helix from an extended conformation to a highly condensed structure in vitro (Chattoraj et al., 1978; Gosule & Schellman, 1976, 1978; Wilson & Bloomfield, 1979; Widom & Baldwin, 1980; Allison et al., 1981) is of focal interest in biology, since, in vivo, nucleic acids often occur tightly packaged and require polyamines to achieve and stabilize the compact form (Cohen, 1978, and references cited therein). When the energy requirements for such a conformational change were estimated by using the encapsidation of DNA by bacteriophage T4 as a model system (Riemer & Bloomfield, 1978), it became obvious that electrostatic repulsions vastly dominate other forces, which oppose the collapse of DNA, and that the process can, therefore, be facilitated by neutralization of the high charge density within the condensed DNA domain. According to the purely electrostatic counterion condensation theory of polyelectrolytes (Manning, 1978), neutralization of the polyelectrolyte depends on the condensation of counterions along its backbone in a process driven strictly by the charge density of the polymer and the valence of the ions. The extent of neutralization realized under specified counterion concentrations can be easily estimated from the theory, which therefore provides the means to determine the critical charge conditions beyond which the collapse of DNA becomes spontaneous (see Theory and Data Analysis). These conditions conceivably depend not only on the charge of the counterion and the extended polyelectrolyte but also on the geometry and mechanism of collapse, the balance between DNA–solvent and DNA–DNA contacts (Post

† From the Department of Biochemistry, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada. *Received November 15, 1983.* This work was aided by a grant from the Banting Research Foundation (65-0570) and a grant (MA-7606) from the Medical Research Council of Canada. Various aspects of this work were supported by National Institutes of Health Grant GM 17855 to V. A. Bloomfield and by Medical Research Council Grant MA-6340 (P. P. Dennis, principal investigator).

* Address correspondence to the Department of Microbiology, University of British Columbia, Vancouver, B.C., Canada V6T 1W5.

[1] Abbreviations: DNA, deoxyribonucleic acid; Spd, spermidine; Hc, hexaamminecobalt(III) $[Co^{3+}(NH_3)_6]$; putThy, α-putrescinylthymine; hmPPUra, 5-(hydroxymethyl)uracil 5α-*O*-pyrophosphate; hmCyt, 5-(hydroxymethyl)cytosine; w.t., wild type; am37, conditional lethal mutant of bacteriophage ϕW14 defective in gene 37; am42, conditional lethal mutant of bacteriophage ϕW14 defective in gene 42; QLS, quasi-elastic light scattering.

0006-2960/84/0423-3609$01.50/0 © 1984 American Chemical Society