Kurachi, K., Davie, E. W., Strydom, D. J., Riordan, J. F., & Vallee, B. L. (1985) *Biochemistry* (third of three papers in this issue).
Kurosky, A., Barnett, D. R., Lee, T. H., Touchstone, B., Hay, R. E., Arnott, M. S., Bowman, B. H., & Fitch, W. M. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 3388-3392.
Lazure, C., Leduc, R., Seidah, N. G., Thibault, G., Genest, J., & Chretien, M. (1984) *Nature (London) 307*, 555-558.
McLachlan, A. D. (1971) *J. Mol. Biol. 61*, 409-424.
Meek, J. L. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1632-1636.
Podell, D. N., & Abraham, G. N. (1978) *Biochem. Biophys. Res. Commun. 81*, 176-185.
Ruegg, U. T., & Rudinger, J. (1977) *Methods Enzymol. 47*, 111-116.
Scheffer, A. J., & Beintema, J. J. (1974) *Eur. J. Biochem. 46*, 221-233.
Schwarz, R. M., & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 3, pp 353-358, National Biomedical Research Foundation, Washington, DC.
Smyth, D. G., Stein, W. H., & Moore, S. (1963) *J. Biol. Chem. 238*, 227-234.
Sperling, R., Burstein, Y., & Steinberg, I. Z. (1969) *Biochemistry 8*, 3810-3820.
Strydom, D. J., & Bazzone, T. J. (1985) *J. Protein Chem.* (in press).
Tarr, G. E., Beecher, J. F., Bell, M., & McKean, D. J. (1978) *Anal. Biochem. 84*, 622-627.
Titani, K., Sasagawa, T., Resing, K., & Walsh, K. A. (1982) *Anal. Biochem. 123*, 408-412.
Vallee, B. L., Riordan, J. F., Lobb, R. R., Higachi, N., Fett, J. W., Crossley, G., Bühler, R., Budzik, G., Breddam, K., Bethune, J. L., & Alderman, E. M. (1985) *Experientia 41*, 1-15.
Zalut, C., Henzel, W. J., & Harris, H. W., Jr. (1980) *J. Biochem. Biophys. Methods 3*, 11-30.

Sequence of the cDNA and Gene for Angiogenin, a Human Angiogenesis Factor†

Kotoku Kurachi,‡,§ Earl W. Davie,‡ Daniel J. Strydom,§ James F. Riordan,§ and Bert L. Vallee*,§

Center for Biochemical and Biophysical Sciences and Medicine, Harvard Medical School, Boston, Massachusetts 02115, and Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received June 14, 1985

ABSTRACT: Human cDNAs coding for angiogenin, a human tumor derived angiogenesis factor, were isolated from a cDNA library prepared from human liver poly(A) mRNA employing a synthetic oligonucleotide as a hybridization probe. The largest cDNA insert (697 base pairs) contained a short 5′-noncoding sequence followed by a sequence coding for a signal peptide of 24 (or 22) amino acids, 369 nucleotides coding for the mature protein of 123 amino acids, a stop codon, a 3′-noncoding sequence of 175 nucleotides, and a poly(A) tail. The gene coding for human angiogenin was then isolated from a genomic λ Charon 4A bacteriophage library employing the cDNA as a probe. The nucleotide sequence of the gene and the adjacent 5′- and 3′-flanking regions (4688 base pairs) was then determined. The coding and 3′-noncoding regions of the gene for human angiogenin were found to be free of introns, and the DNA sequence for the gene agreed well with that of the cDNA. The gene contained a potential TATA box in the 5′ end in addition to two Alu repetitive sequences immediately flanking the 5′ and 3′ ends of the gene. The third Alu sequence was also found about 500 nucleotides downstream from the Alu sequence at the 3′ end of the gene. The amino acid sequence of human angiogenin as predicted from the gene sequence was in complete agreement with that determined by amino acid sequence analysis. It is about 35% homologous with human pancreatic ribonuclease, and the amino acid residues that are essential for the activity of ribonuclease are also conserved in angiogenin. This provocative finding is thought to have important physiological implications.

Angiogenesis, the process leading to the development of a vascular network in normal as well as in malignant tissues, occurs under various physiological and pathological conditions including wound healing, embryonic development, rheumatoid diseases, diabetic retinopathies, and progressive tumor growth (Folkman & Cotran, 1976; Gullino, 1981; Schor & Schor, 1983; Vallee et al., 1985). The proteins that mediate angiogenesis have been difficult to characterize since they occur in trace amounts. Recently, the first human tumor derived angiogenesis factor, designated as angiogenin, has been isolated and characterized (Fett et al., 1985). It is a protein of molecular weight ~14000 and induces neovascularization on the chick chorioallantoic membrane at femtomolar levels and in the rabbit cornea at picomolar concentrations. It is purified from the serum-free conditioned medium of a human colon adenocarcinoma cell line (HT-29) (Alderman et al., 1985). The amino acid sequence of angiogenin isolated from this cell line is described in an accompanying paper (Strydom et al., 1985).

In the present report, the cloning and sequence of cDNAs and the gene for human angiogenin are described. The amino acid sequence predicted from the gene agrees completely with that obtained by protein sequencing studies. There is a high

† This work was supported in part by funds from the Monsanto Co. under agreements with Harvard University.
‡ University of Washington.
§ Harvard Medical School.

0006-2960/85/0424-5494$01.50/0

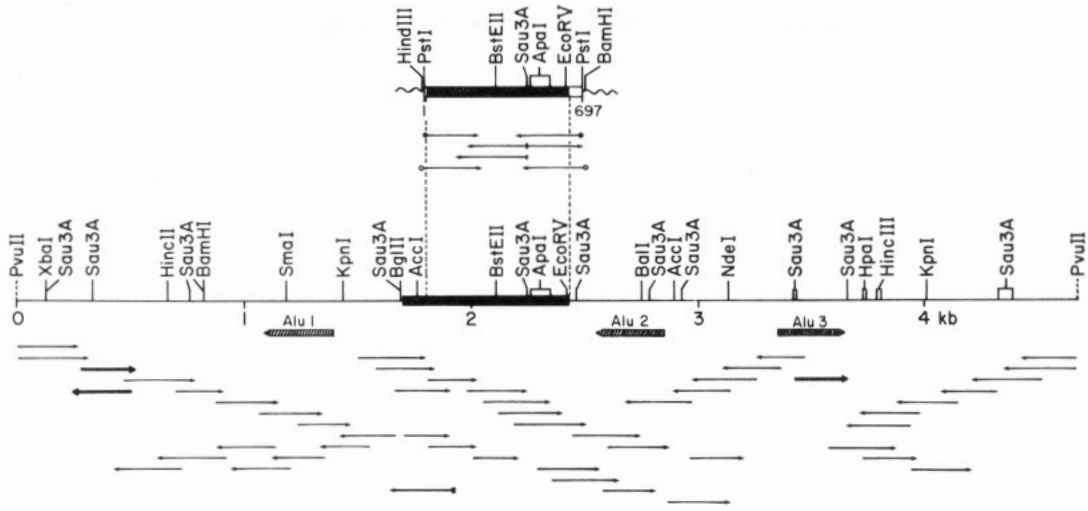

FIGURE 1: Restriction map and sequencing for the cDNA and the gene for human angiogenin. The top portion refers to the cDNA and the lower portion to the genomic DNA. Solid bars indicate the regions encoded by the cDNA and the gene. The wavy lines at the 5′ and 3′ ends of the cDNA refer to the pUC13 vector. The direction and length of the sequencing by the chemical base cleavage method are shown by thin arrows. Arrows with solid circles, open circles, and thin vertical bars indicate the DNA fragments that were labeled at the *Pst*I site, *Hin*dIII or *Bam*HI site, and *Sau*3A site, respectively. The locations and directions of the three Alu sequences flanking the gene are shown by large hatched arrows. The length and direction of the sequencing by the dideoxy chain-termination method are shown by the thin arrows immediately below the Alu sequences. Each analysis was started at a site which was generated by sequential deletion with endonuclease *Bal*31. The thicker arrows indicate that DNA sequencing was carried out with specific synthetic oligonucleotides as primers. The arrow with two vertical bars indicates that DNA sequencing was performed on a *Kpn*I fragment employing the synthetic probe as a primer. The positions of the two *Pvu*II sites shown with dashed lines are only approximate.

degree (35%) of sequence identity to the pancreatic ribonuclease family of proteins.

Materials and Methods

A human cDNA library was prepared from human liver poly(A) mRNA employing plasmid pUC13 as a cloning vector (Maniatis et al., 1982). This plasmid was previously tailed with G's at its *Pst*I site (Michelson & Orkin, 1982). A mixture of 26 synthetic oligonucleotides [CCCTGAGGCTTAGC-(A/G)TC(A/G)TA(A/G)TG(C/T)TG] was purchased from P-L Biochemicals and employed as a hybridization probe. The nucleotide mixture was complementary to nucleotide sequences that code for Gln-His-Tyr-Asp-Ala-Lys-Pro-Gln-Gly. This sequence is present in the amino-terminal region of human angiogenin isolated from the colon adenocarcinoma cell line HT-29 (Strydom et al., 1985). The nucleotide mixture was radiolabeled with T4 kinase and [^{32}P]ATP to a specific activity of approximately 3×10^8 cpm/μg and employed for the screening of 350 000 transformants from the liver library by the method of Wallace et al. (1981). Recombinant plasmids that hybridized strongly with the probe were isolated and purified by cesium chloride gradient centrifugation. The DNA inserts in each of the positive clones were digested with various restriction enzymes and analyzed by polyacrylamide gel electrophoresis. Their sequence was determined by the chemical degradation method of Maxam & Gilbert (1980). Each sequence was determined 2 or more times, and greater than 85% of the sequence was determined on both strands.

A human genomic library (Maniatis et al., 1978) consisting of about 3×10^6 λ Charon 4A bacteriophage was screened with a cDNA coding for human angiogenin (pHAG1) previously radiolabeled by nick translation (Rigby et al., 1977). One strongly hybridizing phage clone identified by the method of Benton & Davis (1977) was plaque purified, and the phage DNA was isolated by the plate lysis method (Maniatis et al., 1982). The genomic insert was analyzed by digestion with various restriction enzymes. A DNA fragment generated by digestion of the insert with *Pvu*II was about 5 kilobases in size and strongly hybridized to the cDNA probe. This fragment was subcloned into plasmid pBR322 and subjected to DNA sequencing by the dideoxy method (Messing et al., 1981; Norrander et al., 1983) employing [^{35}S]dATPαS as described in the Amersham cloning and sequencing manual. A DNA fragment generated by digestion of the phage genomic insert with *Kpn*I was about 3 kilobases in size and also strongly hybridized to the probe. It was then subcloned with the M13mp18 phage vector and also subjected to DNA sequencing employing the synthetic oligonucleotide probe as a primer. Systematic deletions of the genomic DNA with endonuclease *Bal*31 were carried out as described by Poncz et al. (1982), Guo & Wu (1983), and Yoshitake et al. (1985). About 95% of the genomic DNA sequence was determined 2 or more times, and greater than 50% of the genomic sequence was determined on both strands.

Restriction endonucleases, T4 DNA ligase, T4 kinase, alkaline phosphatase, endonuclease *Bal*31, and the Klenow fragment (*Escherichia coli*) were purchased from Bethesda Research Laboratories or New England Biolabs. Reverse transcriptase (avian myeloma virus) was obtained from Seikagaku U.S.A., Inc. Dideoxynucleotide triphosphates, deoxynucleotide triphosphates, pBR322, and pUC13 were purchased from P-L Biochemicals. Universal primers (heptadecamer) for dideoxy sequencing were purchased from New England Biolabs, and [α-^{32}P]dATP, [γ-^{32}P]ATP, and [^{35}S]-dATPαS were obtained from Amersham.

Results and Discussion

Seven positive clones for angiogenin were identified when a human liver cDNA library of 350 000 clones was screened with a radiolabeled synthetic probe corresponding to the amino acid sequence Gln-His-Tyr-Asp-Ala-Lys-Pro-Gln-Gly. The plasmid containing the largest cDNA insert (pHAG1) was then sequenced by the method of Maxam & Gilbert (1980) according to the strategy shown in the top of Figure 1. This cDNA insert contained 697 nucleotides and included 12 G's at the 5′ end, a short noncoding sequence, a leader sequence coding for a signal peptide of 24 (or 22) amino acids, 369 nucleotides coding for the mature protein of 123 amino acids,

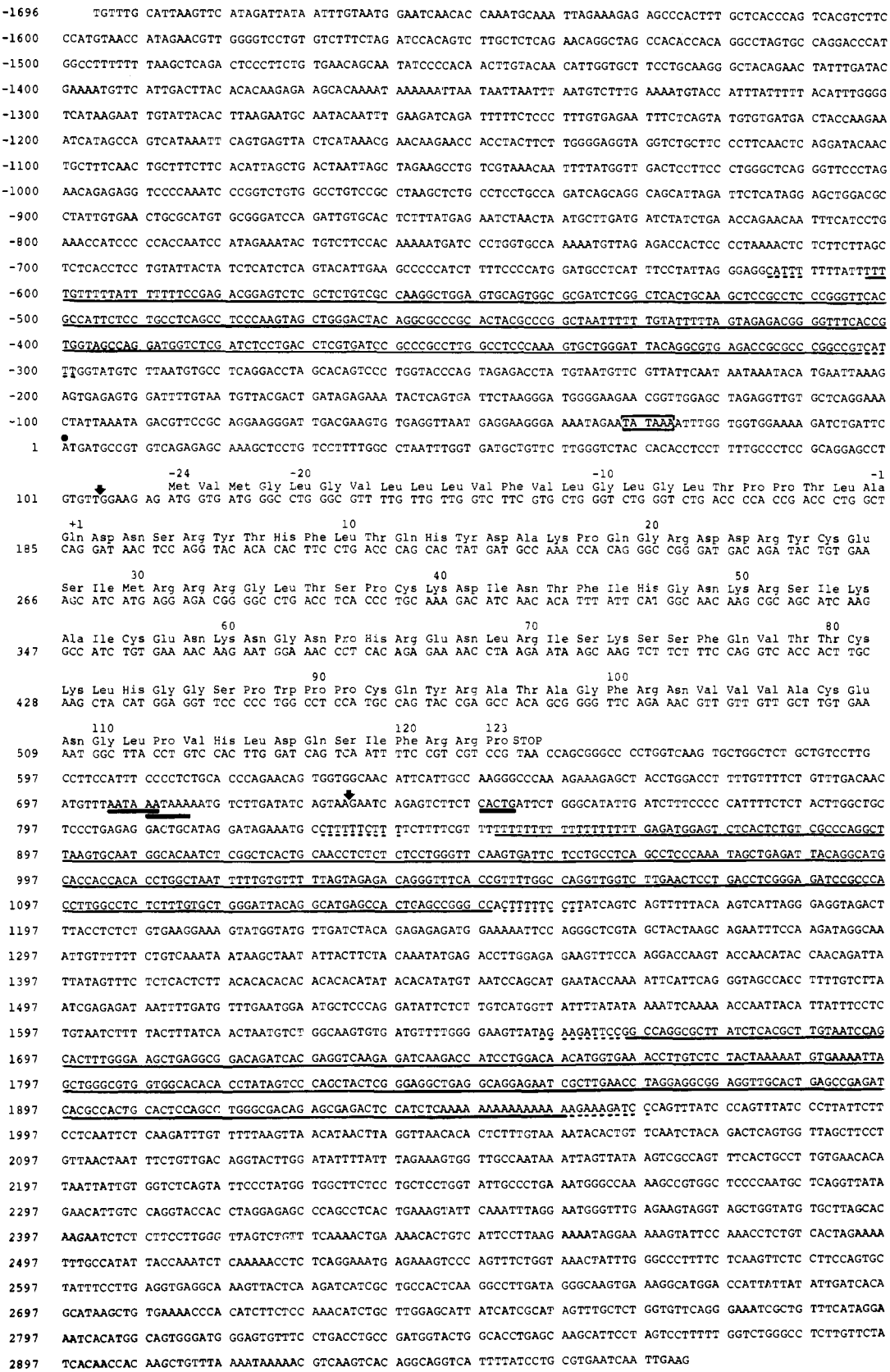

FIGURE 2: Nucleotide sequence for the cDNA for human angiogenin as well as its gene and flanking regions. The sequence for the cDNA is located between the two vertical arrows and corresponds to nucleotides 106–731 where the polyadenylation occurs. Amino acids present in the leader sequence are identified by negative numbers. The proposed transcription initiation site in the gene (nucleotide 1) is identified by a solid circle, while the proposed TATA box (nucleotide −32) is boxed. The two AATAAA sequences and the CACTG sequence are underlined with solid bars. The three Alu repeat sequences in the 5′- and 3′-flanking regions of the gene are underlined, while the direct repeat sequences flanking each Alu repeat are identified with dashed underlines.

a stop codon, 175 nucleotides of 3′-noncoding sequence, a poly(A) tail of 36 nucleotides, and 23 C's on the 3′ end. The sequence for this cDNA is shown in Figure 2 and corresponds to nucleotides 106–731 in the genomic DNA. It differs from the genomic sequence (see below) at nucleotide 252 (G in the cDNA and A in the genomic DNA), corresponding to an Asp

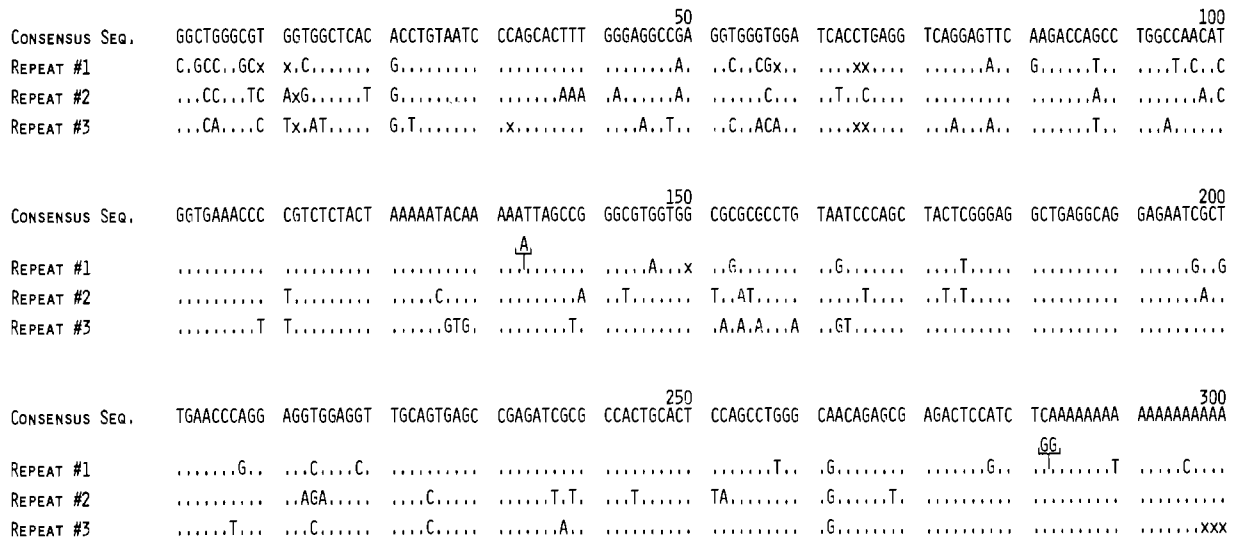

FIGURE 3: Sequence of the three Alu sequences that flank the gene for angiogenin. Alu repeats 1 and 2 are the inverted repeats from the complementary strand shown in Figure 2. Dots represent identity with the consensus sequence (Schmid & Jelinek, 1982), while x represents a deletion.

at residue 23 rather than a Gly. The amino acid sequence of the protein isolated from the HT-29 cell line also contains Asp at residue 23 as did two other cDNAs. These results suggest that the Gly predicted in position 23 by the cDNA insert in pHAG1 was due to a cloning artifact or polymorphism in the gene.

The amino-terminal residue in the mature angiogenin is pyroglutamic acid (Strydom et al., 1985). This indicates that angiogenin is synthesized with a leader sequence of 24 (or 22) amino acids which is removed during processing by signal peptidase. Like other leader sequences in secreted proteins (Blobel et al., 1979), the sequence is rich in hydrophobic amino acids including Val, Leu, and Phe. At the present time, it is not known whether angiogenin biosynthesis starts at Met-24 or Met-22.

The amino acid composition of the mature protein was calculated as follows: Asp_6, Asn_9, Thr_7, Ser_9, Glu_4, Gln_6, Pro_8, Gly_8, Ala_5, Val_5, Met_1, Ile_7, Leu_6, Tyr_4, Phe_5, Lys_7, His_6, Arg_{13}, Trp_1, and $^1/_2$-Cys_6. The molecular weight for the protein was calculated to be 14 193. The protein is free of potential carbohydrate chains bound to Asn in the sequence of Asn-X-Ser or Asn-X-Thr.

A human genomic library consisting of about 3×10^6 recombinant λ phage was then screened with the cDNA from plasmid pHAG1 as a probe. A strongly hybridizing clone (λHAG1) was identified and plaque purified, and the DNA insert was subjected to Southern blotting after digestion with various restriction enzymes. A DNA fragment that was generated by digestion with *Kpn*I was found to contain the gene for angiogenin, and this DNA was subcloned into M13mp18. It was then sequenced with the synthetic probe as a primer, and this confirmed the presence of the gene for angiogenin in this phage. The gene for angiogenin was also found in a DNA fragment of about 5 kilobases that was generated by digestion of λHAG1 with *Pvu*II. This DNA fragment was subcloned into pBR322 and subjected to DNA sequencing by the dideoxy chain-termination method employing the strategy shown in the bottom of Figure 1. The complete sequence of the gene for human angiogenin (Figure 2) indicated that the gene contains about 800 nucleotides and is free of intervening sequences in the coding and 3′-noncoding regions of the gene. The possibility of an intron(s) in the 5′-flanking region cannot be excluded, however, since the largest cDNA did not extend into this region.

The gene for angiogenin contains three Alu repetitive sequences (Schmid & Jelinek, 1982) in its flanking regions (Figure 1 and 2). The first Alu repeat was located in the immediate 5′-flanking region of the gene, while the second was present in the immediate 3′-flanking region. These two Alu repeats were in the same inverted orientation. The third Alu repeat was located about 500 nucleotides downstream from the second Alu sequence in the 3′-flanking region of the gene and was in the typical orientation with the poly(A) on the 3′ end of the 300-nucleotide sequence. Furthermore, each Alu repeat was flanked by a pair of short direct repeat sequences (Figure 2). The nucleotide sequences for the three Alu repeats were about 87% homologous to the consensus Alu sequence of Schmid & Jelinek (1982) (Figure 3).

A tentative TATA box and a transcription initiation site were identified at nucleotides −32 and +1, but no potential CAAT box was found in the immediate vicinity. A sequence of TCAAT was identified, however, at nucleotide −225 which is about 190 base pairs upstream from the proposed TATA box. Two sequences of AATAAA which are involved in the polyadenylation or processing of the messenger RNA at the 3′ end (Proudfoot & Brownlee, 1976) were identified at nucleotides 703 and 707. Polyadenylation of the mRNA occurs at nucleotide 731 which is 20 nucleotides downstream from the end of the second AATAAA sequence. The consensus sequence of CACTG, which also may be involved in polyadenylation or cleavage of the mRNA at the 3′ end (Berget, 1984), was present starting at nucleotide 747. A stretch of 32 nucleotides with alternating purine and pyrimidine was found starting with nucleotide 1416. This sequence provides a potential region for a left-handed helix structure or Z DNA in the gene (Rich et al., 1984).

It is of interest that the gene for human angiogenin is devoid of introns. This is not entirely unique among mammalian genes, however, since several other genes including those coding for proteins in mitochondria (Anderson et al., 1981), histones (Kedes, 1979), and α- and β-interferon (Nagata et al., 1980) are also free of introns.

Southern blotting of human leukocyte DNA digested with *Eco*RI, *Pst*I, or *Pvu*II indicated the presence of only one DNA fragment for each enzymatic digest that hybridized with the cDNA probe (pHAG1). These data are consistent with the conclusion that there is only one copy of the gene for angiogenin in human DNA. Mammalian genes without introns,

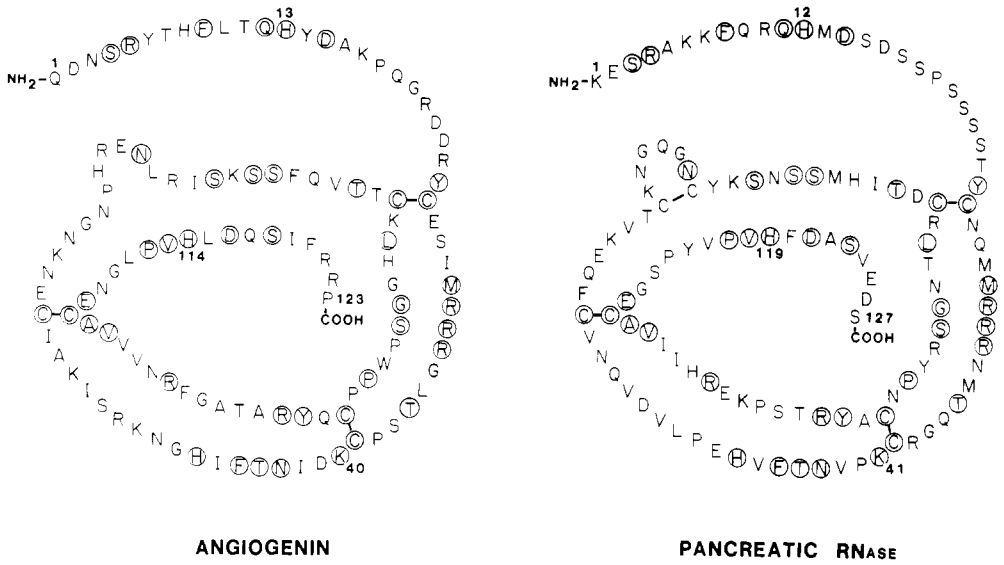

FIGURE 4: Comparison of amino acid sequences for human angiogenin and human pancreatic ribonuclease [from Bientema et al. (1984)]. The positions of the disulfide bonds in angiogenin have been established (Strydom et al., 1985) and are analogous to those in ribonuclease. Amino acids that are homologous between the two proteins are circled. Chemical evidence (Strydom et al., 1985) indicates that the N-terminal residue of angiogenin is pyroglutamic acid (or glutamine).

however, are often present in more than one copy in order to protect the gene from inactivation by mutations (Naora & Deakon, 1982).

A computer search of the flanking regions of the gene for angiogenin as well as in the complementary strand showed no open reading frames. An extensive search for sequences homologous to angiogenin using the NIH Genbank, however, revealed a number of short sequences that were homologous with other DNAs (18–23 identical nucleotides out of 25). Among these, the sequence starting at nucleotide −1635 showed 23 out of 25 nucleotides that were identical with the mRNA for rabbit muscle actin starting at residue 86. The significance of this finding is not clear.

The amino acid sequence of human angiogenin is about 35% homologous with human ribonuclease. The structures of the two proteins are compared in Figure 4 in which residues in common are circled. Amino acid residues Gln-11, His-12, Lys-41, Thr-45, His-119, and Asp-121 that are involved in the active site of ribonuclease (Wlodawer et al., 1982; Blackburn et al., 1982; Stern et al., 1984; Bientema et al., 1984) are also conserved in angiogenin (Gln-12, His-13, Lys-40, Thr-44, His-114, and Asp-116). The location of the disulfide bonds, determined by direct protein sequence analysis (Strydom et al., 1985), further emphasizes the homology to ribonuclease. This structural relationship has been discussed (Strydom et al., 1985), and a more detailed analysis of the homology between the two proteins will be presented elsewhere. To date, angiogenin has not been found to exhibit ribonuclease activity toward a number of potential RNA substrates.

At the present time, very little is known about the mechanism of action of angiogenin in the physiological processes leading to angiogenesis. The identification and chemical characterization of molecular entities involved in this phenomenon have been the subject of intense investigation for many years [see Vallee et al. (1985) and references cited therein]. The isolation and structural analysis of angiogenin are therefore a most significant accomplishment in this long-term study. Many of the past obstacles were due to the miniscule quantities of material available even from very large-scale cell culture facilities. The cloning of the gene for this protein now makes it possible to prepare adequate quantities of it in mammalian, yeast, and various microbial expression systems which will permit the study of its biochemical and physiological properties in greater detail. In addition, means are now available to explore the profound clinical and pharmacological implications of these findings so as to develop substantial and decisive inroads to the more general aspects both of organogenesis and of its antagonization.

ACKNOWLEDGMENTS

We thank Dr. Fred Hagen and Charl Gray for assistance in the preparation of the cDNA library, Dr. Tom Maniatis for kindly making available the genomic library constructed in λ Charon 4A phage, and John Ward for excellent technical assistance. The many helpful discussions with Drs. Edward M. Alderman, J. Lemuel Bethune, James W. Fett, Roy R. Lobb, and Susanna Ryback are also gratefully acknowledged.

REFERENCES

Alderman, E. M., Lobb, R. R., & Fett, J. W. (1985) *Proc. Natl. Acad. Sci. U.S.A.* (in press).

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijin, M. H. L., Couldon, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., & Young, I. G. (1981) *Nature (London) 290*, 451–465.

Benton, W. D., & Davis, R. W. (1977) *Science (Washington, D.C.) 196*, 181–182.

Berget, S. M. (1984) *Nature (London) 309*, 179–182.

Bientema, J. J., Wietzes, P., Weickmann, J. L., & Glitz, D. J. (1984) *Anal. Biochem. 136*, 48–64.

Blackburn, P., & Moore, S. (1982) *Enzymes, 3rd Ed. 15*, 317–433.

Blobel, G., Walter, P., Chang, C. N., Goldman, B. M., Erickson, A. H., & Lingappa, R. (1979) *Symp. Soc. Exp. Biol. 33*, 9–36.

Fett, J. W., Strydom, D. J., Lobb, R. R., Alderman, E. M., Bethune, J. L., Riordan, J. F., & Vallee, B. L. (1985) *Biochemistry* (first of three papers in this issue).

Folkman, J., & Cotran, R. S. (1976) *Int. Rev. Exp. Pathol. 16*, 207–248.

Gullino, P. M. (1981) *Handb. Exp. Pharmacol. 57*, 427–449.

Guo, L. H., & Wu, R. (1983) *Methods Enzymol. 100*, 60–96.

Kedes, L. H. (1979) *Annu. Rev. Biochem. 49*, 837–870.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., & Efstratiadis, A. (1978) *Cell (Cambridge, Mass.) 15*, 687–702.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
Messing, J., Crea, R., & Seeburg, P. H. (1981) *Nucleic Acids Res. 9*, 309–321.
Michelson, A. M., & Orkin, S. H. (1982) *J. Biol. Chem. 257*, 14773–14782.
Nagata, S., Mantei, N., & Weissman, C. (1980) *Nature (London) 287*, 401–408.
Naora, H., & Deakon, N. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6196–6200.
Norrander, J., Kempe, T., & Messing, J. (1983) *Gene 26*, 101–106.
Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E., & Surrey, S. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4298–4302.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.
Rich, A., Nordheim, A., & Wang, A. H.-J. (1984) *Annu. Rev. Biochem. 53*, 791–846.
Rigby, P. W. J., Dieckman, M., Rhodes, C., & Berg, P. (1977) *J. Mol. Biol. 113*, 237–251.
Schmid, C. W., & Jelinek, W. R. (1982) *Science (Washington, D.C.) 216*, 1065–1070.
Schor, A. M., & Schor, S. L. (1983) *J. Pathol. 141*, 385–413.
Stern, M., & Doscher, M. S. (1984) *FEBS Lett. 171*, 253–256.
Strydom, D. J., Fett, J. W., Lobb, R. R., Alderman, E. M., Bethune, J. L., Riordan, J. F., & Vallee, B. L. (1985) *Biochemistry* (second of three papers in this issue).
Vallee, B. L., Riordan, J. F., Lobb, R. R., Higachi, N., Fett, J. W., Crossley, G., Bühler, R., Budzik, G., Breddam, K., Bethune, J. L., & Alderman, E. M. (1985) *Experientia 41*, 1–15.
Wallace, R. B., Johnson, M. J., Hirose, T., Miyske, T., Kawashima, E. H., & Itakura, K. (1981) *Nucleic Acids Res. 9*, 879–894.
Wlodawer, A., Bott, R., & Sjölin, L. (1982) *J. Biol. Chem. 257*, 1325–1332.
Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W., Kurachi, K. (1985) *Biochemistry 24*, 3736–3750.

Methylated Pyrimidines Stabilize an Alternating Conformation of Poly(dA-dU)·Poly(dA-dU)†

Hai-young Wu and Michael J. Behe*
Department of Chemistry, City University of New York/Queens College, Flushing, New York 11367
Received February 19, 1985

ABSTRACT: We have investigated the effect of increasing percentages of methylated pyrimidines on the structure of poly(dA-dU)·poly(dA-dU). This was done by synthesizing analogous polynucleotides that contained deoxythymidine residues as well as deoxyuridine residues and observing their ^{31}P NMR spectra in increasing amounts of CsF. The results show that methylated pyrimidines play a large role in the stabilization of the "alternating B" conformation of DNA.

The effect of DNA methylation on the expression of genes in eukaryotic cells has been reviewed several times in the past few years (Bird, 1984; Felsenfeld & McGhee, 1982). Extensive methylation often is associated with gene inactivation, while undermethylation is sometimes required for a gene to be expressed. The methylated sites in chromosomal DNA are most often at the C5 position of cytosine in a -CG- dinucleotide sequence. There has been speculation that the effect of methylation is to alter DNA conformation since methylation of cytosines in a polymer of that sequence, poly(dG-m^{5}dC)·poly(dG-m^{5}dC), was shown (Behe & Felsenfeld, 1981) to greatly facilitate the B to Z transition. However, no B–Z transition was seen when a plasmid containing the chicken adult β-globin gene was methylated at its -CCGG- sites by *Hpa*II methylase (Nickol & Felsenfeld, 1983).

The presence of a methylated pyrimidine can cause other conformational variations besides the B–Z transition, however. It has been observed (Patel et al., 1982; Chen et al., 1983) that the methylated synthetic polymer poly(dG-m^{5}dC)·poly(dG-m^{5}dC) is in an "alternating B" conformation in solutions of moderate NaCl concentration but the analogous unmethylated polymer is not. The alternating B conformation was first proposed by Klug et al. (1979) for poly(dA-dT)·poly(dA-dT), another polymer containing methylated pyrimidines, and the appearance of a closely spaced doublet (0.2–0.3 ppm separation) in its ^{31}P NMR spectrum greatly supported the alternating model (Shindo et al., 1979; Chen & Cohen, 1983; Patel et al., 1981). Klug et al. (1979) speculated that the alternating structure was stabilized by a stacking interaction involving the methyl group of the pyrimidine base.

In order to elucidate the dependence of the alternating B conformation on methylated pyrimidines, we have examined the ^{31}P NMR spectra of several synthetic polynucleotides. The polymers studied include poly(dA-dU)·poly(dA-dU), poly(dA-dT)·poly(dA-dT), and analogous polynucleotides con-

† This work was supported by National Institutes of Health Grants GM31898 and CA00945 and by Professional Staff Congress of CUNY Grant 6-64106.
* Address correspondence to this author at the Department of Chemistry, Lehigh University, Bethlehem, PA 18015.

0006-2960/85/0424-5499$01.50/0