H., Hoth, D. F., Gelman, E., Groopman, J. E., Resnick, L., Gallo, R. C., Myers, C. E., & Fauci, A. S. (1985) *Lancet ii*, 627–630.
Harada, S., Koyanagi, Y., & Yamamoto, N. (1985) *Science (Washington, D.C.) 229*, 563–566.
Lindell, T. J., Weinberg, F., Morris, P. W., Roeder, R. G., & Rutter, M. J. (1967) *Science (Washington, D.C.) 170*, 447–449.
Matsukage, A., Sivarajan, M., & Wilson, S. H. (1976) *Biochemistry 15*, 5305–5314.
Mitsuya, H., & Broder, S. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1911–1915.
Mitsuya, H., Popovic, M., Yarchoan, R., Matsushita, S., Gallo, R. C., & Broder, S. (1984) *Science (Washington, D.C.) 226*, 172–174.
Mitsuya, H., Weinhold, K. J., Furman, P. A., St. Clair, M. H., Nusinoff-Lehrman, S., Gallo, R. C., Bolognesi, D., Barry, D. W., & Broder, S. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7096–7100.
Nakajima, K., Ono, K., & Ito, Y. (1974) *Intervirology 3*, 324–341.
Nakane, H., Balzarini, J., De Clercq, E., & Ono, K. (1988) *Eur. J. Biochem. 177*, 91–96.
Okamura, S., Crane, F., Messner, H. A., & Mak, T. W. (1978) *J. Biol. Chem. 253*, 3765–3767.
Ono, K., Ohashi, A., Tanabe, K., Matsukage, A., Nishizawa, M., & Takahashi, T. (1979) *Nucleic Acids Res. 7*, 715–726.
Ono, K., Iwata, Y., & Nakane, H. (1983) *Biomed. Pharmacother. 37*, 27–35.
Ono, K., Nakane, H., Barrê-Sinoussi, F., & Chermann, J.-C. (1988a) *Eur. J. Biochem. 176*, 305–310.
Ono, K., Nakane, H., & Fukushima, M. (1988b) *Eur. J. Biochem. 172*, 349–353.
Ono, K., Nakane, H., Fukushima, M., Chermann, J.-C., & Barrê-Sinoussi, F. (1989a) *Biochem. Biophys. Res. Commun. 160*, 982–987.
Ono, K., Nakane, H., Fukushima, M., Chermann, J.-C., & Barrê-Sinoussi, F. (1989b) *Biomed. Pharmacother.* (in press).
Ono, K., Nakane, H., Meng, Z.-M., Ose, Y., Sakai, Y., & Mizuno, M. (1989c) *Chem. Pharm. Bull. 37*, 1810–1812.
Rozenbaum, W., Dormont, D., Spire, B., Vilmer, E., Gentilini, M., Griscelli, C., Montagnier, L., Barrê-Sinoussi, F., & Chermann, J.-C. (1985) *Lancet i*, 450–451.
Yamaguchi, M., Matsukage, A., & Takahashi, T. (1980) *J. Biol. Chem. 255*, 7002–7009.
Yarchoan, R., Klecker, R. W., Weinhold, K. J., Markham, P. D., Lyerly, H. K., Durack, D. T., Gelmann, E., Nusinoff-Lehrmann, S., Blun, R. M., Barry, D. W., Shearer, G. M., Fischl, M. A., Mitsuya, H., Gallo, R. C., Collins, J. M., Bolognesi, D. P., Myers, C. E., & Broder, S. (1986) *Lancet i*, 575–580.

Characterization of cDNAs of the Human Pregnancy-Specific β1-Glycoprotein Family, a New Subfamily of the Immunoglobulin Gene Superfamily†,‡

Qiao-Xi Zheng, Le Ann Tease, W. Lesley Shupert, and Wai-Yee Chan*
Departments of Pediatrics, Biochemistry, Anatomy, and Cell Biology, Georgetown University Medical Center, 3800 Reservoir Road, NW, Washington, D.C. 20007
Received August 11, 1989; Revised Manuscript Received November 2, 1989

ABSTRACT: Three highly homologous cDNAs encoding human pregnancy-specific β1-glycoprotein (SP1) were isolated from a human placental cDNA library. These cDNAs share >90% nucleotide homology in their coding sequences, and >79% of the encoded amino acids are homologous. Proteins encoded by these cDNAs are very similar to members of the carcinoembryonic antigen family and contain repeating domains, conserved disulfide bridges, and β-sheet structure typical of the immunoglobulin gene superfamily. However, the high degree of sequence homology and relatively lesser degree of glycosylation among the SP1 proteins suggest that they exist as a unique family instead of being members of the CEA family. Both soluble and potentially membrane-bound forms of SP1 proteins were present in the placenta. Northern blot analysis using specific probes confirmed the expression of multiple mRNA species in human term placenta.

Pregnancy-specific β1-glycoprotein (SP1) is an early pregnancy protein that can be detected in maternal serum as early as 16 days after conception (Grudzinskas et al., 1977). The maternal serum level of SP1 increases with progression of gestation and has been shown to be a good index for monitoring fetal growth and pregnancy complications (Bischof, 1984). SP1 is also elevated in serum of patients with tumors of trophoblastic origin as well as some nontrophoblastic tumors (Sorensen, 1982). The physiological function of SP1 is not known. Like many of the early pregnancy proteins, it has been suggested to be an immunosuppressive agent enhancing survival of the fetus (Cerni et al., 1977). Human placental SP1 as first reported (Bohn, 1972) is a single-subunit glycoprotein with a molecular mass of 90 kDa, 29% of which is carbohydrate. This protein, besides being shown to be synthesized by the syncytiotrophoblasts of the placenta (Horne et al., 1976), was recently shown to be also produced in extraplacental sites including intestine, testis, and uterus (Chan et al., 1988a–c). Further studies indicate that SP1 as it was initially defined might be composed of a group of highly homologous proteins (Chan et al., 1988a). These different species of SP1 proteins are the products of a family of genes. Several species of

† Supported in part by NIH Grant HD 21793 and by the Oklahoma Medical Research Foundation.
‡ The nucleotide sequences reported for hPS12 and hPS2 have been submitted to GenBank under Accession Numbers J02892 and J02893.
* Author to whom correspondence should be addressed at the Department of Pediatrics.

0006-2960/90/0429-2845$02.50/0

placental SP1 cDNAs have been cloned (Chan et al., 1988a; Rooney et al., 1988; Streydio et al., 1988; Watanabe & Chou, 1988a). Computer analysis of the SP1 cDNAs cloned revealed that they are very similar to carcinoembryonic antigen (CEA), a subfamily of the immunoglobulin (Ig) gene superfamily (Beauchemin et al., 1987; Oikawa et al., 1987; Zimmermann et al., 1987; Chan et al., 1988a; Rooney et al., 1988; Streydio et al., 1988; Watanabe & Chou, 1988b). All the cDNAs reported encode proteins that are highly hydrophilic. However, in vitro translation studies suggested that SP1 protein might be amphiphilic (Hau et al., 1983). The present report describes the structure of two hitherto unreported human placental SP1 cDNAs, one of which has a sequence encoding a hydrophobic C-terminus making it a very likely candidate for the membrane-bound form of SP1. Detailed analysis of the three placental SP1 cDNAs cloned in our laboratory and the other SP1 cDNAs reported was performed. Results support the hypothesis that the placental SP1 proteins are products of a multigene family and constitute a new subfamily of the Ig gene superfamily (Williams & Barclay, 1988).

Materials and Methods

Screening of cDNA Library. A human placental cDNA expression library in λ gt11, generously provided by Dr. Brian Knoll of M. D. Anderson Hospital and Tumor Institute, was screened with ^{125}I-labeled anti-human SP1 antibody (Calbiochem, San Diego, CA) as described previously (Chan et al., 1988c). Two million phages were screened. The positive plaques picked were rescreened until purified. In order to obtain longer clones, another human placental cDNA library in λ gt11, purchased from Clontech Laboratories, Palo Alto, CA, was rescreened with confirmed partial SP1 cDNAs.

Preparation of DNA for Subcloning and Sequencing. Recombinant phage DNA was prepared by plate lysis followed by banding on a cesium chloride step gradient as described previously (Chan & Qiu, 1988). The cDNA insert was released from the phage vector by partial digestion with *Eco*RI and separated in a 1% SeaPlaque agarose gel (FMC BioProducts, Rockland, ME) by electrophoresing in 1× TAE buffer (40 mM Tris, 20 mM sodium acetate, 2 mM EDTA, pH 8.0) (Weislander, 1979). The cDNA insert was excised and the gel melted by incubation at 65 °C for 10 min (Dumais & Nochumson, 1987). For hPS12, the cDNA insert was subcloned into M13mp18 and M13mp19 vectors for single-strand sequencing (Messing, 1983) and into pBluescript for double-strand sequencing (Stratagene, 1988). For hPS2, the total cDNA insert was recovered, and a number of clones containing overlapping cDNA fragments were generated by Bal-31 deletion subcloning.

Bal-31 Deletion Subcloning. The hPS2 cDNA insert was blunt-ended by filling in with dNTPs by the action of DNA polymerase I in melted agarose at 37 °C (Dumais & Nochumson, 1987). The blund-ended DNA was ligated into pUC13, and the recombinant plasmid DNA was prepared for Bal-31 deletion subcloning as previously described (Maniatis et al., 1982). The recombinant plasmid was linearized by digestion with either *Pst*I or *Sst*I. Linearized plasmid DNA was subjected to sequential digestion with Bal-31 nuclease (Yoshitake et al., 1985). Sequentially deleted DNA fragments were first blunt-ended and then digested with either *Sst*I or *Pst*I. Deleted cDNA fragments were separated on 1.3% SeaPlaque agarose gels (Weislander, 1979) and ligated into M13mp18 or M13mp19 vectors for single-strand sequencing (Messing, 1983).

DNA Sequence Determination. The nucleotide sequence of the cDNAs was determined by a modified dideoxy chain termination method using Klenow fragment (Boehringer Mannheim Biochemicals, Indianapolis, IN) at 50 °C or Sequenase (U.S. Biochemical Corp., Cleveland, OH) at 37 °C (Johnston-Dow et al., 1987). M13 universal sequencing primer and T3 and T7 sequencing primers as well as synthetic oligonucleotide primers were used to prime the sequencing reaction in single-strand sequencing (Chan et al., 1989) or double-strand sequencing with the pBluescript system (Stratagene, 1988). All sequences were determined three or more times as well as from different subclones and from both strands.

DNA Sequence Analysis. DNA sequences were analyzed by the Sequence Analysis Software Package of the Genetics Computer Group at the University of Wisconsin (Devereux et al., 1984). SP1 cDNAs reported in this work were compared with those reported previously, including hPS11 (Chan et al., 1988a) PSG16 and PSG93 (Watanabe & Chou, 1988a), PSBGC, PSBGD, and PSBGE (Streydio et al., 1988), and pSP1-i (Rooney et al., 1988).

Northern Blot Analysis. Total RNA was extracted from human term placenta in the presence of a 4 M guanidine monothiocyanate–phenol–chloroform mixture (Chomczynski & Sacchi, 1987). Fifty micrograms of total placental RNA was denatured with formaldehyde, separated in an agarose/formaldehyde gel, and analyzed by Northern blotting with labeled SP1 cDNA probes (Chan et al., 1988c; Fourney et al., 1988).

Results

Screening of Human Placental cDNA Library. Six positive clones were initially obtained by immunological screening of a human placental cDNA library (Chan et al., 1988c). These clones were confirmed to be partial cDNAs encoding human placental SP1 (Chan et al., 1988a). One of these cloned cDNAs was used as a probe to rescreen another human placental cDNA library. A total of 26 clones were obtained. The nucleotide sequence of one of these clones (hPS11) has been reported (Chan et al., 1988a). The nucleotide sequences of two of the longer clones, hPS12 and hPS2, were determined. They both represented partial cDNA with an incomplete 5′ coding sequence. To obtain more 5′ sequence, the most 5′ *Eco*RI–*Bam*HI fragments of hPS12 and hPS2 were used as probes to rescreen the same cDNA library. Two clones, hPS89 and hPS90, with more 5′ sequence than hPS12 were identified, while no clone having more 5′ sequence than hPS2 was found.

Sequence Analysis of hPS12 and hPS2. The composite sequence of hPS12 is shown in Figure 1. The cDNA had 1573 bp with a 5′ noncoding sequence of 45 bp, an open reading frame of 1272 bp encoding 424 amino acids with a calculated molecular mass of 47.5 kDa, a stop codon of TAA, and a 3′ noncoding sequence of 253 bp. Even though no upstream stop codon could be identified, the sequence around the presumed translation initiation codon ACCATGG agreed with the consensus sequence for initiation of translation in vertebrates suggested by Kozak (1987). No poly(A) tail was found in any of the hPS12 clones sequenced. Eight potential glycosylation sites, two of the form Asn-X-Ser and six of the form Asn-X-Thr, were present. The encoded protein contained an N-terminal domain, two n-subdomains and one c-subdomain each containing two conserved Cys residues, and a C-terminal domain, all characteristic of SP1 proteins reported previously (Chan et al., 1988a; Rooney et al., 1988; Streydio et al., 1988; Watanabe & Chou, 1988a).

Figure 2 shows the nucleotide sequence of hPS2. It had 1744 bp with an open reading frame of 1053 bp encoding 351 amino acids, a stop codon TGA, and a 3′ noncoding sequence

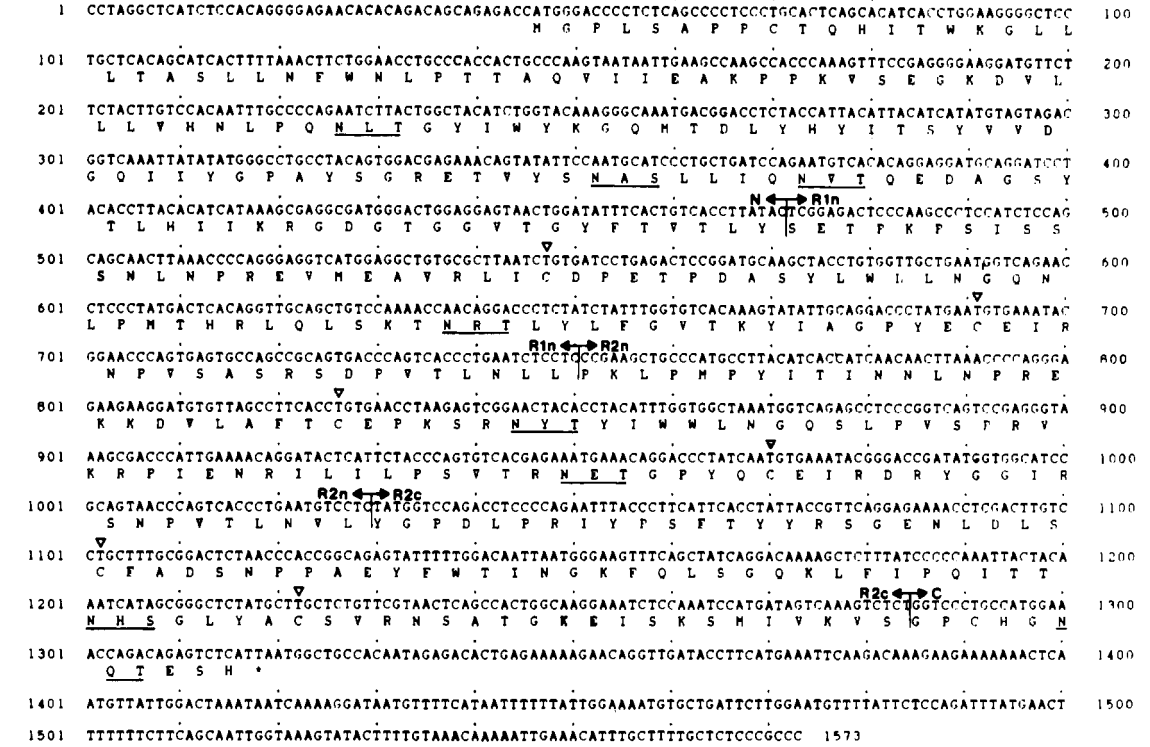


FIGURE 1: Nucleotide and predicted amino acid sequence of hPS12. Nucleotide numbers are indicated. Potential glycosylation sites are underlined. Open inverted triangles indicate conserved Cys residues. An asterisk indicates a stop codon. Boundaries of the different domains are indicated as follows: N, N-terminal domain; R1n, n-subdomain of repeat unit 1; R2n, n-subdomain of repeat unit 2; R2c, c-subdomain of repeat unit 2; C, C-terminal domain.

FIGURE 2: Nucleotide and predicted amino acid sequence of hPS2. Nucleotide numbers are indicated. Potential glycosylation sites are underlined. Open inverted triangles indicate conserved Cys residues. Solid squares indicate the polyadenylation site. An asterisk indicates a stop codon. Boundaries of the different domains are indicated as follows: N, N-terminal domain; Rn, n-subdomain of repeat unit; Rc, c-subdomain of repeat unit; C, C-terminal domain.

of 659 bp. A polyadenylation signal ATTAAA 14 bp upstream from a 29-bp poly(A) could be identified. The polypeptide encoded by hPS2 contained part of the N-terminal domain, an n-subdomain, a c-subdomain similar to that observed in hPS12, and an 81 amino acid C-terminal domain. In addition to the conserved Cys residues in the n- and c-subdomains, two Cys residues separated by 27 amino acids were present in the C-terminal domain. Five potential gly-

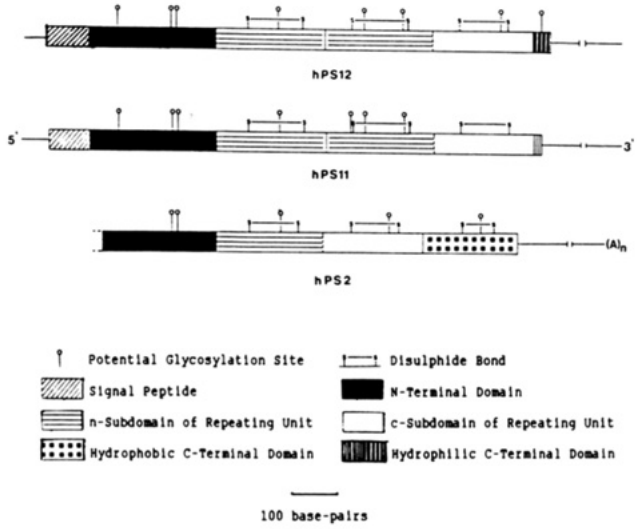


FIGURE 3: Comparison of domain structure of human placental pregnancy-specific β1-glycoprotein cDNAs hPS12, hPS11, and hPS2.

cosylation sites could also be identified.

The domain structures of hPS12, hPS2, and hPS11 were compared in Figure 3. hPS11 and hPS12 had the same domain structure. The most significant differences between hPS2 and the other two cDNAs cloned from the same placental library were the presence of only one complete repeating unit (i.e., one n-subdomain and one c-subdomain) and an 81 amino acid hydrophobic C-terminal domain. In spite of these differences, the Cys residues as well as the potential glycosylation sites were conserved in these three cDNAs.

Prediction of secondary structure with programs available in the Sequence Analysis Software Package of the Genetics Computer Group (Kyte & Doolittle, 1982) showed that both hPS12 and hPS2 were mainly in the form of β-sheets (results not shown). The PEPPLOT program (Goldman et al., 1986) and the PLOTSTRUCTURE program (Chou & Fasman, 1978) both indicated that hPS12 was largely hydrophilic except for the N-terminal 120 amino acids. The N-terminal 34–35 amino acids had the characteristics of a putative signal peptide as observed for hPS11 (Chan et al., 1988a). The N-terminal domain of hPS2, similar to that of hPS12, was also hydrophobic. However, in contrast to any of the SP1 sequences reported, the C-terminal domain of hPS2 was highly hydrophobic (results not shown).

Northern Blot Analysis of Human Placental RNA. Figure 4 shows results of Northern blot analysis of human placental RNA probed with a specific fragment of each cDNA, namely, the 3′ 620-bp *Eco*RI–*Eco*RI fragment (the entier 3′ noncoding sequence) of hPS11, the 3′ *Nco*I–*Eco*RI fragment (the C-terminal 24 bp of the coding sequence and the 3′ noncoding sequence) of hPS12, and the 280-bp *Bam*HI–*Eco*RI fragment (sequence encoding the C-terminal domain) of hPS2. Both the hPS11- and the hPS2-specific probes hybridized with a 2.25-kb mRNA band, while the hPS12-specific probe hybridized with a 1.65-kb mRNA band. In addition to the 2.25-kb mRNA band, the hPS2-specific probe also hybridized with the 1.65-kb mRNA band. Even though the ratio of the specific radioactivity of the three probes used to hybridize to the blot was roughly 7:4:3 (hPS2:hPS12:hPS11), the intensity of the hybridization signal was hPS11 > hPS12 > hPS2.

Comparison of Members of the SP1 Family. A comparison of the encoded amino acid sequences of all SP1 cDNAs whose nucleotide sequences had been reported and their consensus sequences is shown in Figure 5. An amino acid or nucleotide was considered to be consensus if it occurred at least three

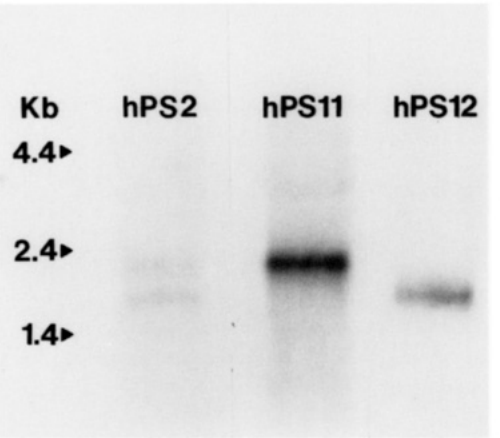


FIGURE 4: Human placental Northern blot probed with cDNA specific probes. Mobilities of size standards (in kb) are indicated. The specific probe used for each lane is shown.

times among the clones hPS12, hPS2, PSBGE, pSP1-i, and hPS11/PSG16/PSBGC-D. The percent homology with the consensus amino acid sequence ranged from 92.5 to 100 in the N-terminal domains and the n-subdomains except for the N-terminal domain of hPS2, which showed only 85.9% homology. Levels of homology among the c-subdomains were slightly lower than those of the other domains and ranged from 85.9% to 97.6%. Nucleotide sequence homology among the N-terminal domains and n- and c-subdomains was similar to that among the amino acid sequences and ranged from 90% to 99%. On the other hand, the amino acid and nucleotide sequences of the C-terminal domains of these cDNAs shared very little sequence homology, with two exceptions. First, the C-terminal domains of hPS11 and PSBGD were identical, and second, the 3′ noncoding sequence of PSG16 was identical with that of hPS11 except for the deletion of 86 bp near the 3′ end of the coding region (Chan et al., 1988a). The positions of the cysteine residue in the n- and c-subdomains of all cDNAs were conserved. The positions of the potential glycosylation sites in the N-terminal domain and the R1n-subdomain were also conserved.

Figure 5A shows that hPS12 differed from the other SP1 cDNAs in the deletion of three nucleotides encoding Ile-90 in the N-terminal domain. It was also apparent from Figure 5A that, albeit very similar, hPS12 and pSP1-i were different from each other and from the other SP1 cDNAs.

Both hPS2 and PSBGE had one less n-subdomain when compared to the other SP1 cDNAs. The percent homology with the R1n-subdomain consensus sequence was 93.5 and 92.5 and that of the R2n-subdomain was 52.7 and 48.4 for hPS2 and PSBGE, respectively. The c-subdomains of hPS2 and PSBGE were less homologous to the consensus sequence than those of the other SP1 cDNAs, being 87.1% and 85.9% respectively, while all other cDNAs showed >90% homology. The most significant difference between hPS2 and the other SP1s was the presence of the 81 amino acid hydrophobic C-terminus, while all other SP1s had only relatively short (14 amino acids or less) hydrophilic or amphiphilic C-termini.

Comparison of the 5′ noncoding sequence showed that there was over 91% similarity among all SP1 cDNAs. Comparison of the 3′ noncoding sequences was more varied. The 3′ noncoding sequence of hPS2 had very little homology with any of the SP1 cDNAs reported. On the other hand, the 3′ noncoding sequence of hPS12 was very homologous to those of PSBGE and pSP1-i. Comparison of these three sequences using the LINEUP program (Devereux et al., 1984) showed that aside from 27, 30, and 39 bp at the 5′ end of this region in

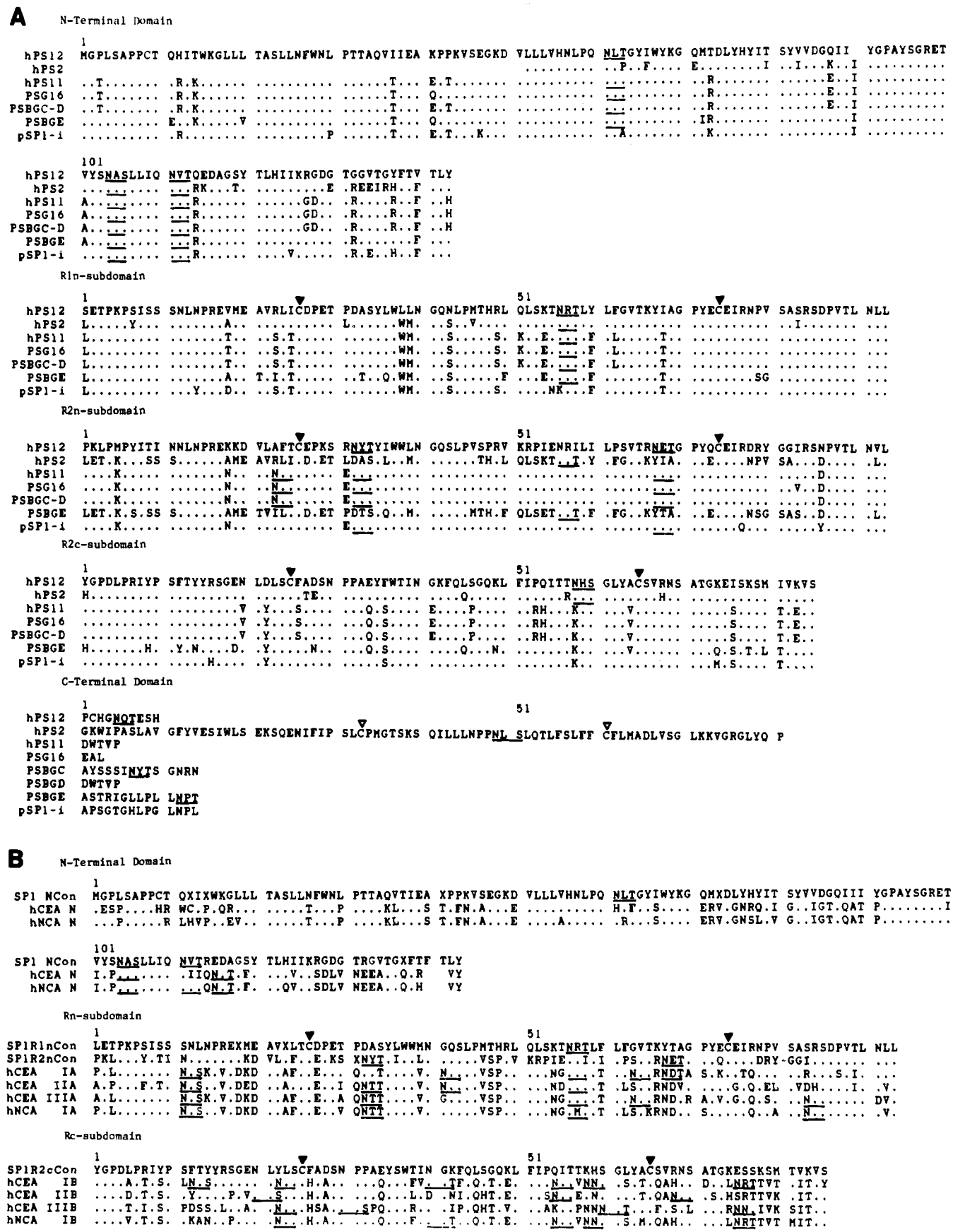


FIGURE 5: Comprison of (A) aligned amino acid sequences encoded by human placental SP1 cDNAs and (B) consensus amino acid sequences of SP1 cDNAs with those encoded by human CEA and NCA cDNAs. Identical amino acids are indicated by dots. Potential glycosylation sites are underlined. Solid inverted triangles indicate conserved Cys residues. Amino acids are numbered with reference to the beginning of each domain or subdomain. The Rn-subdomain of hPS2 and PSBGE was repeated under both R1n- and R2n-subdomains for the purpose of comparison. SP1 NCon, consensus sequence of the N-terminal domain of SP1 cDNAs; SP1R1nCon, consensus sequence of the R1n-subdomain of SP1 cDNAs; SP1R2nCon, consensus sequence of the R2n-subdomain of SP1 cDNAs; SP1R2cCon, consensus sequence of the R2c-subdomain of SP1 cDNAs. Domain notation of CEA and NCA is the same as described (Beauchemin et al., 1987; Tawaragi et al., 1988).

hPS12, PSBGE, and pSP1-i, respectively, there was 95% homology among the three cDNAs. The 3′ noncoding sequences of hPS11, PSG16, and PSBGD were almost identical except for the deletion of 70 bp after the stop codon in PSG16 and one mismatch in both PSG16 and PSBGD when compared with hPS11.

Comparison with Carcinoembryonic Antigen (CEA) and Nonspecific Cross-Reacting Antigen (NCA), Members of the Ig Gene Superfamily. Among the placental SP1 cDNAs, hPS2 and PSBGE had the domain structure closest to that of CEA and NCA. Both hPS2 and PSBGE, similar to NCA, had only one complete repeat unit, i.e., one n-subdomain and one c-subdomain (Neumaier et al., 1988; Tawaragi et al., 1988), while CEA had three complete repeating units (Beauchemin et al., 1987; Oikawa et al., 1987; Zimmermann et al., 1987). All other SP1 cDNAs had one complete repeating unit and an additional n-subdomain. Figure 5B shows the comparison of the consensus amino acid sequences of the different domains of the SP1 cDNAs with the corresponding domains in CEA and NCA. The percent homology of both CEA and NCA to SP1 was quite similar in all domains compared, ranging from 53.8 to 62.4 at the amino acid level. The amino acid sequences of the C-terminal domains were very variable among the SP1 cDNAs and were not compared with those of CEA or NCA. Nucleotide sequence comparison of the different domains of SP1, CEA, and NCA showed results similar to those of amino acid comparison.

Discussion

All the SP1 cDNAs reported demonstrate characteristics required for inclusion in the Ig gene superfamily, namely, sequence homology, characteristic domain structure, conserved disulfide bonds within domains, and β-sheet structure (Williams, 1987). The SP1 proteins therefore constitute a subfamily of the Ig gene superfamily.

There is increasing evidence suggesting the presence of multiple species of highly homologous SP1 proteins in human placenta. It has also been shown previously that the SP1 proteins are encoded by multiple genes (Chan & Qiu, 1988). Even though a recent report suggested the possibility that different SP1 proteins might be the differentially spliced products of one gene (Oikawa et al., 1988), results of these studies support the contention that SP1 protein in human placenta consists of products of more than one gene (Chan et al., 1988a; Rooney et al., 1988; Streydio et al., 1988). Two of the reported placental SP1 cDNAs, hPS11 and PSBGD, are identical. PSG16 and another partially sequenced cDNA, PSG93, differ from these two cDNAs at only four bases, which could correspond to polymorphisms (Chan et al., 1988a; Streydio et al., 1988; Watanabe & Chou, 1988a). On the basis of their almost identical protein coding sequences, hPS11, PSG16, PSBG93, PSBGC, and PSBGD are likely to be the products of the same gene with differentially spliced exons encoding the C-terminus and 3′ noncoding sequence. This conclusion is in line with that derived from studies of a cloned fragment of human SP1 gene (Oikawa et al., 1988) and CEA gene (Barnett et al., 1989). Three other cDNAs, hPS12, PSBGE, and pSP1-i, have 221 bp in their 3′ noncoding sequences that are highly homologous (95%). Considering that this is observed in the noncoding region and that the three clones were derived from two different libraries, it is conceivable that the few differences were individual polymorphisms. These three cDNAs are products of splicing different amino acid coding exons to one common exon that contains the 3′ noncoding sequence. The exon encoding the N-terminal domain of hPS12 is unique in that it has a 3-bp deletion when compared to the other cDNAs. PSBGE differs from hPS12 and pSP1-i by having one less n-subdomain, which could be the result of a difference in splicing. The gene encoding hPS12/PSBGE/pSP1-*i* could be different from that encoding hPS11/PSG16/PSG93/PSBGC-D. The other SP1 cDNA, hPS2, is unique. It has no significant homology with any of the SP1 cDNAs reported. Thus it might be the product of a third SP1 gene.

The presence of at least three different species of mRNA in human placenta was confirmed by the Northern blot analysis. The difference in hybridization signal intensity suggested that hPS2 mRNAs were present in lesser quantity than either hPS12 mRNA or hPS11 mRNA. It is interesting to notice that the hPS2-specific probe hybridized to both the 1.65-kb and 2.25-kb mRNA bands. These results suggest that each mRNA band contains more than one species of mRNA. Up to the present time only one placental SP1 cDNA with a sequence encoding a hydrophobic C-terminus has been found. The Northern blot results thus also suggest that there might be more than one species of membrane-bound SP1 present in human placenta. Even though the size of pSP1-i cDNA reported is 2.016 kb (Rooney et al., 1988), the hPS12-specific probe, which is supposed to be also capable of hybridizing with pPS1-i, hybridized only to a mRNA band of 1.65 kb. The cause of this discrepancy is not clear from this study. Nonetheless, the presence of multiple species of SP1 mRNA in human term placenta is demonstrated by Northern analysis. Multiple mRNAs have also been reported for the CEA gene family (Thompson & Zimmermann, 1988). Thus the SP1 genes, like the CEA family of genes, are probably localized in clusters (Thompson & Zimmermann, 1988). Naturally, there is also the possibility that all SP1 cDNAs so far reported were derived from one very large gene that gives rise to the above three groups of products by differential splicing.

Both hPS2 and PSBGE, unlike the other SP1 cDNAs, contain only one n-subdomain. These cDNAs could be formed by having one complete repeating unit containing an n-subdomain and a c-subdomain (i.e., the n-subdomain in hPS2 is an R2n-subdomain) or by splicing the n-subdomain of one repeating unit with the c-subdomain of another repeating unit (i.e, the n-subdomain in hPS2 is an R1n-subdomain). By comparison to the consensus sequences of R1n- and R2n-subdomains of the SP1 cDNAs, it is apparent from Figure 5 that the n-subdomain of hPS2 was probably derived from the R1 repeating unit instead of the R2 repeating unit. A similar conclusion was drawn for PSBGE (Streydio et al., 1988).

Figure 3 shows one distinctive feature of hPS2—the presence of an 81 amino acid hydrophobic C-terminus. This feature of hPS2 is very similar to that observed in CEA (Beauchemin et al., 1987; Oikawa et al., 1987; Zimmermann et al., 1987), TM-CEA (Barnett et al., 1989), and NCA (Tawaragi et al., 1988). Hydropathy plot analysis showed that this C-terminal domain of hPS2 is very hydrophobic and supports the inference that it represents the membrane-anchor region of the molecule. Thus it appears that there are two types of SP1 proteins in human placenta, namely, the cytoplasmic or soluble SP1 and the membrane-bound SP1. Analogous phenomena have been reported for the CEA family of proteins (Thompson & Zimmermann, 1988; Barnett et al., 1989).

Even though the SP1 proteins and CEAs appear to share many common properties, there are sufficient features unique to the SP1 proteins to qualify them as a separate subfamily of the Ig gene superfamily instead of being members of the CEA gene subfamily. Even though the SP1s are very similar

to CEA and NCA, the homology among the different SP1 family members is even higher, being consistently >90% at the nucleotide level and >85% at the amino acid level. The SP1 proteins are also less glycosylated. The number of potential glycosylation sites ranges from four in SP1-i to eight in PSBGC with the majority having six or seven. Most of these sites are also conserved. On the other hand, both CEA and NCA are more heavily glycosylated, with 27 potential sites in CEA (Beauchemin et al., 1987; Oikawa et al., 1987; Zimmermann et al., 1987) and 12 potential sites in NCA (Tawaragi et al., 1988; Neumaier et al., 1988). A number of these sites are conserved between CEA and NCA but not between CEAs and SP1s, as shown in Figure 5B.

The strong conservation of both nucleotides and amino acids among the internal repeats of SP1 and CEA genes suggests that both gene families evolved recently by the duplication of a primordial gene. It is interesting to notice that the percent homology at the amino acid level between the R1n-subdomain and R2n-subdomain of SP1 is 52.7, which is significantly lower than that between the R1n-subdomain of SP1 and the n-subdomains of CEA (62.4 with IA, 57.0 with IIA, and 59.1 with IIIA) and NCA (62.4). On the other hand, it is comparable with the percent similarity between R2n-subdomain and the n-subdomains of CEA and NCA. Comparison of nucleotide sequences gives the same result. This seems to suggest that duplication of the R1 repeat unit to the R2 repeat unit occurred before the divergence of SP1 and CEA and that the divergence of the two genes involved duplication of the R1 repeat unit only.

Similarity in structure to the Ig gene superfamily suggests potential immunological function of the SP1 proteins. Preliminary study using purified human placental SP1 proteins show that these proteins inhibit stimulated lymphocytes in a mixed lymphocyte assay but not those stimulated by phytohemagglutinin (L. Thompson and W. Y. Chan, unpublished). The specific species of SP1 protein(s) that has this immunosuppressive function, however, is not clear from these studies. Previous studies have indicated that SP1 might have a growth-promoting property (Chan et al., 1988b). This property of the SP1 proteins has been demonstrated by using murine megakaryocyte cultures (S. A. Burnstein and W. Y. Chan, unpublished). The presence of multiple species as well as soluble and membrane-bound members in the SP1 family might indicate different functions for different members. Cloning of the cDNA of the different species of SP1 will allow testing of the properties of each individual protein after expression of the specific cDNA with appropriate systems.

ACKNOWLEDGMENTS

We acknowledge the generosity of Dr. Brian Knoll of the Department of Pathology, M. D. Anderson Hospital and Tumor Institute, for use of the placental cDNA library.

REFERENCES

Barnett, T. R., Kretschmer, A., Austen, D. A., Goebel, S. J., Hart, J. T., Elting, J. J., & Kamarck, M. E. (1989) *J. Cell Biol. 108*, 267–276.

Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A., & Stanners, C. P. (1987) *Mol. Cell. Biol. 7*, 3221–3230.

Bischof, P. (1984) *Contrib. Gynecol. Obstet. 12*, 6–92.

Bohn, H. (1972) *Blut 24*, 292–302.

Cerni, C., Tatra, G., & Bohn, H. (1977) *Arch. Gynaekol. 223*, 1–7.

Chan, W. Y., & Qiu, W. R. (1988) *Am. J. Hum. Genet. 43*, 152–159.

Chan, W. Y., Borjigin, J., Zheng, Q. X., & Shupert, W. L. (1988a) *DNA 7*, 545–555.

Chan, W. Y., Tease, L. A., Bates, J. M., Jr., Borjigin, J., & Shupert, W. L. (1988b) *Hum. Reprod. 3*, 687–692.

Chan, W. Y., Tease, L. A., Borjigin, J., Chan, P. K., Rennert, O. M., Srinivasan, B., Shupert, W. L., & Cook, R. G. (1988c) *Hum. Reprod. 3*, 677–685.

Chan, W. Y., Lin, Q., Busch, H., Rennert, O. M., Tease, L. A., & Chan, P. K. (1989) *Biochemistry 28*, 1033–1039.

Chomczynski, P., & Sacchi, N. (1987) *Anal. Biochem. 162*, 156–159.

Chou, P. Y., & Fasman, G. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 145–147.

Devereux, J., Haeberli, P., & Marquess, P. (1984) *Nucleic Acids Res. 12*, 387–395.

Dumais, M. M., & Nochumson, S. (1987) *BioTechniques 5*, 62–67.

Fourney, R. M., Miyakoshi, J., Day, R. S., III, & Paterson, M. C. (1988) *Focus 10*, 5–7.

Goldman, A., Engelman, D. M., & Steiz, T. A. (1986) *Annu. Rev. Biophys. Biophys. Chem. 15*, 321–353.

Grudzinskas, J. G., Lenton, E. A., Gordon, Y. B., Kelso, I. M., Jeffrey, D., Sobowale, O., & Chard, T. (1977) *Br. J. Obstet. Gynaecol. 84*, 740–742.

Hau, J., Grudzinskas, J. G., Teisner, B., & Svendsen, P. (1983) *Placenta 4*, 291–294.

Horne, C. H. W., Towler, C. M., Pugh-Humphreys, R. G. P., Thompson, A. W., & Bohn, H. (1976) *Experientia 32*, 1197–1199.

Jameson, B. A., & Wolf, H. (1988) *Comput. Appl. Biosci. 4*, 181–186.

Johnston-Dow, L., Mardis, E., Heiner, C., & Roe, B. A. (1987) *BioTechniques 5*, 754–765.

Kozak, M. (1987) *Nucleic Acids Res. 15*, 8125–8148.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–131.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Messing, J. (1983) *Methods Enzymol. 101C*, 20–77.

Neumaier, M., Zimmermann, W., Shively, L., Hinoda, Y., Riggs, A. D., & Shively, J. E. (1988) *J. Biol. Chem. 263*, 3202–3207.

Oikawa, S., Imajo, S., Noguchi, T., Kosaki, G., & Nakazato, H. (1987) *Biochem. Biophys. Res. Commun. 144*, 634–642.

Oikawa, S., Inuzuka, C., Kosaki, G., & Nakazato, H. (1988) *Biochem. Biophys. Res. Commun. 156*, 68–77.

Rooney, B. C., Horne, C. H. W., & Hardman, N. (1988) *Gene 71*, 439–449.

Sorensen, S. (1982) *Clin. Chim. Acta 121*, 199–208.

Stratagene (1988) *pBluescript II Exo/Mung DNA sequencing system*, Stratagene, La Jolla, CA.

Streydio, C., Lacka, K., Swillens, S., & Vassart, G. (1988) *Biochem. Biophys. Res. Commun. 154*, 130–137.

Tawaragi, Y., Oikawa, S., Matsuoka, Y., Kosaki, G., & Nakazato, H. (1988) *Biochem. Biophys. Res. Commun. 150*, 89–96.

Thompson, J., & Zimmermann, W. (1988) *Tumor Biol. 9*, 63–83.

Watanabe, S., & Chou, J. Y. (1988a) *J. Biol. Chem. 263*, 2049–2054.

Watanabe, S., & Chou, J. Y. (1988b) *Biochem. Biophys. Res. Commun. 152*, 762–768.

Weislander, L. (1979) *Anal. Biochem. 98*, 305–309.
Williams, A. F. (1987) *Immunol. Today 8*, 298–303.
Williams, A. F., & Barclay, A. N. (1988) *Annu. Rev. Immunol. 6*, 381–405.
Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W., & Kurachi, K. (1985) *Biochemistry 24*, 3736–3750.
Zimmermann, W., Ortlieb, B., Friedrich, R., & von Kleist, S. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2960–2964.

An NMR Study of the Covalent and Noncovalent Interactions of CC-1065 and DNA

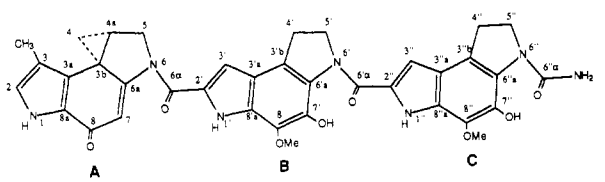
Terrence A. Scahill,* Randy M. Jensen, David H. Swenson,‡ Nicole T. Hatzenbuhler, Gary Petzold, Wendell Wierenga, and Nanda D. Brahme§

Research Laboratories, Pharmaceutical Research and Development Division, The Upjohn Company, Kalamazoo, Michigan 49001

Received August 8, 1989; Revised Manuscript Received November 16, 1989

ABSTRACT: The binding of the antitumor drug CC-1065 has been studied with nuclear magnetic resonance (NMR) spectroscopy. This study involves two parts, the elucidation of the covalent binding site of the drug to DNA and a detailed investigation of the noncovalent interactions of CC-1065 with a DNA fragment through analysis of 2D NOE (NOESY) experiments. A CC-1065–DNA adduct was prepared, and an adenine adduct was released upon heating. NMR (^{1}H and ^{13}C) analysis of the adduct shows that the drug binds to N3 of adenine by reaction of its cyclopropyl group. The reaction pathway and product formed were determined by analysis of the ^{13}C DEPT spectra. An octamer duplex, d(CGATTAGC·GCTAATCG), was synthesized and used in the interaction study of CC-1065 and the oligomer. The duplex and the drug–octamer complex were both analyzed by 2D spectroscopy (COSY,NOESY). The relative intensity of the NOEs observed between the drug (CC-1065) and the octamer duplex shows conclusively that the drug is located in the minor groove, covalently attached to N3 of adenine 6 and positioned from the 3′ → 5′ end in relation to strand A [d(CGATTA^{6}GC)]. A mechanism for drug binding and stabilization can be inferred from the NOE data and model-building studies.

CC-1065, a fermentation product of *Streptomyces zelensius*, was one of the most cytotoxic antitumor agents known when discovered (Hanka et al., 1978):



The isolation and characterization of this antibiotic has been reported (Martin et al., 1980, 1981), and the X-ray crystal study including an interaction study of CC-1065 with DNA was published shortly thereafter (Chidester et al., 1981). The antibiotic has a unique structure possessing a cyclopropyl ring with alkylating potential and a twist or pitch in the backbone of the molecule giving it a half-moon or banana shape with concave and convex sides. This structural feature permits a compatible fit in the minor groove of DNA. Experimental evidence for minor groove binding includes competitive binding studies with netropsin, site-exclusion studies, and reduction of the alkylation of minor groove sites after treatment of the DNA with methylation agents (Swenson et al., 1982). Molecular model building using CPK models and computer graphics have also been used to understand the drug–DNA interaction in a qualitative way.

(+)-CC-1065 was shown to be efficacious in the treatment of experimental tumors in mice (Reynolds et al., 1986; Martin et al., 1981). Unfortunately, (+)-CC-1065 also produced delayed death in mice at therapeutic doses (McGovern et al., 1984), and subsequently, its development was halted. However, this lead compound has led to the development of potent, efficacious analogues that do not cause delayed death and are promising as therapeutic agents (Warpehoski et al., 1986; Li et al., 1987).

Preliminary reports (Hurley et al., 1984; Scahill et al., 1986) detailed the structure of the adduct formed upon reaction with calf thymus (CT)[1] DNA, and two others demonstrating the sequence specificity of the drug with DNA have also appeared (Reynolds et al., 1986; Hurley et al., 1988). While CC-1065 interacts with DNA and is thought to exhibit its cytotoxic effects through disruption of DNA synthesis, the exact nature of the DNA binding has been elusive. Circular dichroism (CD) studies (Krueger et al., 1985, 1987) suggest strong binding to duplex DNA, especially sequences rich in adenine–thymine (A-T) base pairs. CC-1065 and DNA interact

* Author to whom correspondence should be addressed.
‡ Present address: Visiting Scientist at the National Center for Toxicological Research, Jefferson, AR 72079.
§ Present address: Bio-Rad Laboratories, 1000 Alfred Nobel Dr., Hercules, CA 94547.

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; COSY, homonuclear correlated spectroscopy; NOESY, two-dimensional NOE correlated spectroscopy; COLOC, correlated spectroscopy with long-range coupling; DEPT, distortionless enhancement with polarization transfer; CT, calf thymus; FID, free induction decay; CD, circular dichroism; HPLC, high-performance liquid chromatography.

0006-2960/90/0429-2852$02.50/0