

- Grunberger, D., Nelson, J. H., Cantor, C. R., & Weinstein, I. B. (1970) *Proc. Natl. Acad. Sci. U.S.A.* 66, 488-494.
- Grunberger, D., Santella, R. M., Hanau, L. H., & Erlanger, B. F. (1985) *Carcinog.—Compr. Surv.* 10, 465-480.
- Hanau, L. H., Santella, R. M., Grunberger, D., & Erlanger, B. F. (1984) *J. Biol. Chem.* 259, 173-178.
- Herr, W. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82, 8009-8013.
- Hingerty, B., & Brody, S. (1982) *Biochemistry* 21, 3243-3252.
- Johnson, B. H., & Rich, A. (1985) *Cell (Cambridge, Mass.)* 42, 713-724.
- Kuo, K. C., McCune, R. A., Gehrke, C. W., Midgett, R., & Ehrlich, M. (1980) *Nucleic Acids Res.* 8, 4763-4776.
- Lacour, F., Nahon-Merlin, E., & Michelson, M. (1973) *Curr. Top. Microbiol. Immunol.* 62, 1-39.
- Lafer, E. M. (1982) Ph.D. Dissertation, Tufts University.
- Lafer, E. M., Sousa, R., Rosen, B., Hsu, A., & Rich, A. (1985) *Biochemistry* 24, 5070-5076.
- Lee, J. S., Woodsworth, M. L., & Latimer, L. J. P. (1984) *Biochemistry* 23, 3277-3281.
- Leslie, A. G. W., Arnott, S., Chandrasekaran, R., & Ratliff, R. L. (1980) *J. Mol. Biol.* 143, 49-72.
- Malfoy, B., Rousseau, N., & Leng, M. (1982) *Biochemistry* 21, 5463-5467.
- Moller, A., Nordheim, A., Kozlowski, S. A., Patel, D. J., & Rich, A. (1984) *Biochemistry* 23, 54-62.
- Munns, T. W., Liszewski, M. K., & Hahn, B. H. (1984) *Biochemistry* 23, 2958-2964.
- Nahon, E., Michelson, A. M., & Lacour F. (1967) *Biochim. Biophys. Acta* 149, 127-139.
- Pohl, F. M., & Jovin, T. M. (1972) *J. Mol. Biol.* 67, 375-396.
- Pohl, F. M., Thomae, R., & DiCapua, E. (1982) *Nature (London)* 300, 545-546.
- Rich, A., Nordheim, A., & Wang, A. H.-J. (1984) *Annu. Rev. Biochem.* 53, 791-846.
- Sage, E., & Leng, M. (1980) *Proc. Natl. Acad. Sci. U.S.A.* 77, 4597-4601.
- Santella, R. M., Grunberger, D., Weinstein, I. B., & Rich, A. (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78, 1451-1455.
- Stollar, B. D. (1975) *CRC Crit. Rev. Biochem.* 3, 45-69.
- Stollar, B. D. (1980) *Methods Enzymol.* 70, 70-85.
- Stollar, B. D. (1986) *CRC Crit. Rev. Biochem.* 20, 1-36.
- Stollar, B. D., Huang, B.-S., & Blumenstein, M. (1986a) *Biomol. Stereodyn.* 3, 69-84.
- Stollar, B. D., Zon, G., & Pastor, R. W. (1986b) *Proc. Natl. Acad. Sci. U.S.A.* 83, 4469-4473.
- Strauss, A. J. L., Seegal, B. C., Hsu, K. C., Burkholder, P. M., Nastuk, W. L., & Osseman, K. E. (1960) *Proc. Soc. Exp. Biol. Med.* 105, 184-191.
- Sundquist, W. I., Lippard, S. J., & Stollar, B. D. (1986) *Biochemistry* 25, 1520-1524.
- Wallace, S. S., Erlanger, B. F., & Beiser, S. M. (1971) *Biochemistry* 10, 679-683.
- Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., & Rich, A. (1979) *Nature (London)* 282, 680-686.
- Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H., & Rich, A. (1981) *Science (Washington, D.C.)* 211, 171-176.
- Zarling, D. A., Arndt-Jovin, D. J., Robert-Nicoud, M., McIntosh, L. P., Thomae, R., & Jovin, T. M. (1984) *J. Mol. Biol.* 176, 369-415.
- Zouali, M., & Stollar, B. D. (1986) *J. Immunol. Methods* 90, 105-110.

## Primary Structure of the Mouse Laminin B2 Chain and Comparison with Laminin B1<sup>†,‡</sup>

Marian E. Durkin,<sup>§</sup> Barbara B. Bartos,<sup>§</sup> Shu-Huang Liu,<sup>§</sup> Stephen L. Phillips,<sup>||</sup> and Albert E. Chung<sup>\*,§</sup>

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, and Department of Microbiology, Biochemistry, and Molecular Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261

Received December 1, 1987; Revised Manuscript Received February 25, 1988

**ABSTRACT:** One of the major components of basement membranes is the glycoprotein laminin, made up of three disulfide-bonded subunits, the A, B1, and B2 chains. We have isolated and sequenced overlapping mouse laminin B2 chain cDNA clones covering 7562 base pairs. The deduced amino acid sequence predicts that the mature B2 chain consists of 1572 residues, has an unglycosylated molecular weight of 173 541, and possesses 14 potential N-linked glycosylation sites. Analysis of the predicted secondary structure shows the presence of six domains, two rich in  $\alpha$ -helical structure, two composed of homologous cysteine-rich repeat units, and two globular regions. The organization of the molecule is very similar to that of the mouse laminin B1 chain, and significant sequence homology between the B1 and B2 chains was found in their two cysteine-rich domains and in their amino-terminal globular domains.

**L**aminin is a major and widely distributed component of basement membranes (Chung et al., 1979; Timpl et al., 1979).

<sup>†</sup> This work was supported by Grants GM 25690 and CA 21246 from the National Institutes of Health, U.S. Public Health Service.

<sup>‡</sup> The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02930.

<sup>\*</sup> Author to whom correspondence should be addressed.

<sup>§</sup> Department of Biological Sciences.

<sup>||</sup> Department of Microbiology, Biochemistry, and Molecular Biology.

The molecule consists of three nonidentical glycoprotein subunits linked by disulfide bonds, the A ( $M_r$  400 000), B1 ( $M_r$  230 000), and B2 ( $M_r$  220 000) chains (Cooper et al., 1981; Howe & Dietzschold, 1983). Laminin is involved in a variety of biological processes including cell adhesion, migration, morphogenesis, metastasis, differentiation, and neurite outgrowth [reviewed by Kleinman et al. (1985)]. Binding of laminin to other basement membrane constituents including type IV collagen, heparan sulfate proteoglycan, and entactin

to form supramolecular complexes has been demonstrated (Kleinman et al., 1986). Several cell-surface laminin binding molecules have been identified that may mediate the interaction of cells with the extracellular matrix, such as the  $M_r$  68 000 laminin receptors of muscle and tumor cells [reviewed by von der Mark and Kuhl (1985)], the chicken fibronectin receptor (Horwitz et al., 1985), sulfated glycolipids (Roberts et al., 1985), and cranin, an  $M_r$  120 000 integral membrane protein (Smalheiser & Schwartz, 1987).

Electron microscopy of rotary-shadowed laminin reveals that it has an asymmetric cruciform shape, consisting of three short arms, each with two globular domains, and a long arm with one large terminal globular domain (Engel et al., 1981). The current model for the structure of laminin proposes that the short arms consist of the amino termini of the A, B1, and B2 chains, the rodlike portion of the long arm is composed of a triple coiled-coil  $\alpha$ -helix formed by the three chains, and the large globule at the end of the long arm is formed by the carboxyl terminus of the A chain (Engel & Furthmayr, 1987). Considerable information on the structure of laminin has been obtained from cDNA cloning. Sequencing of cDNAs corresponding to the carboxyl ends of the B1 and B2 chains first revealed that they had the potential to form a coiled-coil  $\alpha$ -helix (Barlow et al., 1984), and this was later confirmed by biophysical and chemical methods (Paulsson et al., 1985). Recently, the complete cDNA sequences of the mouse and human laminin B1 chains have been determined, and they indicate that the polypeptide is organized into seven distinct structural domains (Sasaki et al., 1987; Pikkariainen et al., 1987). In this paper, we report the isolation of cDNA clones encoding the mouse laminin B2 chain. The deduced amino acid sequence shows that the B2 and B1 chains share many structural features.

## MATERIALS AND METHODS

Restriction endonucleases and other enzymes were purchased from New England Biolabs, Boehringer Mannheim, and Bethesda Research Laboratories. *Eco*RI-digested, alkaline phosphatase treated  $\lambda$ gt11 DNA was obtained from Promega Biotec. Gigapack packaging extracts, Bluescript KS plasmid DNA, and sequencing primers were purchased from Stratagene. [ $\alpha$ - $^{32}$ P]dCTP for nick-translations, [ $\gamma$ - $^{32}$ P]ATP for end-labeling, and [ $\alpha$ - $^{35}$ S]dATP for sequencing were purchased from Amersham and New England Nuclear. Oligonucleotides used in library screening, cDNA synthesis, and sequencing were synthesized on an Applied Biosystems DNA synthesizer at the Pittsburgh DNA Synthesis Facility.

Poly(A<sup>+</sup>) RNA was isolated by guanidine thiocyanate extraction and oligo(dT)-cellulose chromatography as previously described (Durkin et al., 1986) from basement membrane secreting aggregates of mouse M1536-B3 cells (Chung et al., 1979) grown in bacterial culture dishes for 6 days. An oligo(dT)-primed double-stranded cDNA library was synthesized by using M1536-B3 poly(A<sup>+</sup>) RNA according to the method of Gubler and Hoffman (1983). For insertion into the vector  $\lambda$ gt11 (Young & Davis, 1983), the cDNA was methylated at the *Eco*RI sites, and the ends were filled in and joined to *Eco*RI linkers (Huynh et al., 1985). After digestion with *Eco*RI and removal of linker monomers by two cycles of ethanol precipitation, 0.3  $\mu$ g of cDNA was ligated to 1  $\mu$ g of *Eco*RI-cut  $\lambda$ gt11 DNA, and the ligation reaction was packaged with Gigapack extracts. A specifically primed cDNA library was constructed in the same way, except that 6.3  $\mu$ g of a synthetic 20-mer, 5'-dTTCCTTCTCTA-GAAGGT-3', was annealed to 10  $\mu$ g of M1536-B3 poly(A<sup>+</sup>)

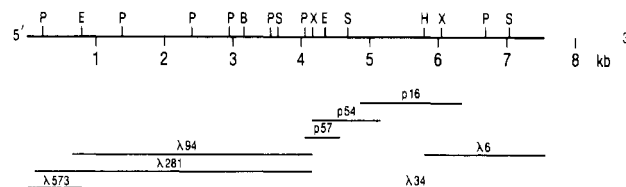


FIGURE 1: Laminin B2 cDNA clones. A map of eight overlapping laminin B2 cDNA clones spanning 7562 bp is shown.  $\lambda$ 573 was obtained from the specifically primed library; the internal *Eco*RI site was apparently cleaved during construction of the library.  $\lambda$ 34, also from the specifically primed library, was evidently synthesized via nonspecific annealing of the primer. Other cDNA clones not shown were isolated and sequenced. Restriction endonuclease sites for *Bam*HI (B), *Eco*RI (E), *Hind*III (H), *Pst*I (P), *Sac*I (S), and *Xba*I (X) are indicated.

RNA for 1 h at 25 °C in 0.4 M KCl prior to first-strand synthesis.

The  $\lambda$ gt11 libraries were screened by plaque hybridization (Maniatis et al., 1982) to cDNA inserts isolated by electroelution and labeled by nick-translation (Rigby et al., 1977). Recombinant phage were adsorbed to *Escherichia coli* Y1090 (r<sup>-</sup>) and incubated 8–15 h at 42 °C, and the plaques were transferred to nitrocellulose filters. A differentiated F9 cDNA library in pBR322 (Durkin et al., 1986) was screened by colony hybridization (Maniatis et al., 1982) to a synthetic 20-mer, 5'-dCGGACCATAGCTGAAGCCAA-3', labeled with [ $\gamma$ - $^{32}$ P]ATP and T4 polynucleotide kinase. Hybridization was performed according to conditions described by Wallace et al. (1979).

Large-scale preparation of phage DNA was performed by the procedure of Davis et al. (1980). The cDNA inserts were isolated by *Eco*RI digestion and agarose gel electrophoresis and subcloned into the Bluescript KS M13 (–) plasmid. Purification of plasmid DNA and restriction endonuclease mapping of the cDNA inserts were performed by using standard protocols (Maniatis et al., 1982).

Sequencing of cDNA inserts was accomplished by the dideoxy chain termination method (Sanger et al., 1977), using both the conventional single-stranded procedure in M13 and the double-stranded modification (Chen & Seeburg, 1985) in the Bluescript KS plasmid. The large *Eco*RI fragments of  $\lambda$ 94,  $\lambda$ 34, and  $\lambda$ 6 were sequenced by constructing nested deletions according to the exonuclease III/S1 nuclease technique (Henikoff, 1984; Lawler & Hynes, 1986). Gaps in the sequence were filled in by using specific synthetic oligonucleotide primers. Analysis of protein sequences was performed with MacGene Plus software.

## RESULTS

**Isolation of Laminin B2 Chain cDNA Clones.** A 1.5-kilobase (kb)<sup>1</sup> laminin B2 cDNA clone, p16, was isolated from a library in pBR322 constructed using poly(A<sup>+</sup>) RNA from mouse F9 embryonal carcinoma cells induced to differentiate to parietal endoderm (Durkin et al., 1986). Comparison of the sequence of p16 to that of a B2 clone isolated by Barlow et al. (1984) revealed that it consisted of the final 120 bp of the open reading frame, the TAG stop codon, and 1.4 kb of the 3' noncoding region. The library was rescreened using a  $^{32}$ P-labeled synthetic oligonucleotide probe (nucleotides 4366–4385 in Figure 2) derived from the published B2 sequence (Barlow et al., 1984); this yielded two clones, p57 and p54 (Figure 1). Screening of an oligo(dT)-primed M1536-B3 library in  $\lambda$ gt11 with nick-translated p16 insert resulted in the isolation of  $\lambda$ 6, which encoded the remainder of the 3' un-

<sup>1</sup> Abbreviations: kb, kilobase(s); bp, base pair(s); SDS, sodium dodecyl sulfate.

the 5' end of p57 (nucleotides 4148–4167 in Figure 2).  $\lambda$ 94 was isolated by screening the library with nick-translated p57 insert. The specifically primed library was probed several more

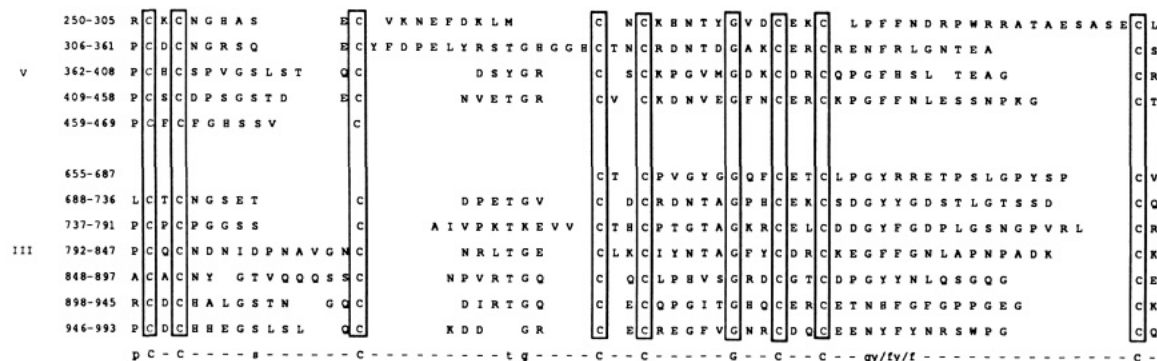


FIGURE 3: Homologous cysteine-rich repeat units of laminin B2. The cysteine-rich repeat units in domains V and III of laminin B2 were aligned with the aid of the MacGene homology program. The numbering of the residues is taken from Figure 2. The eight identical cysteine residues and one identical glycine residue in each repeat are boxed. Below is a consensus sequence that includes residues occurring in the same position greater than 50% of the time.

times, each time using a restriction fragment from the 5' end of a clone isolated in the previous round, until  $\lambda$ 573 was isolated. The cDNA clones span 7562 bp, close to the size of 8 kb estimated for the laminin B2 message by Northern blot analysis (Barlow et al., 1984; Durkin et al., 1986; Kleinman et al., 1987).

**Nucleotide and Amino Acid Sequences.** The complete laminin B2 cDNA sequence and the deduced amino acid sequence are shown in Figure 2. A 4815 bp open reading frame extends from a presumptive initiator methionine codon at nucleotide 193 (Kozak, 1987) to a TAG stop codon at nucleotide 5008, followed by a 2547 bp 3' untranslated region and 8 bp of the poly(A) tail. The open reading frame encodes a 1605 amino acid polypeptide of  $M_r$  176 836. Inspection of the amino acid sequence suggests that the mature laminin B2 chain begins at the alanine located 34 amino acids from the translation start site, and this has been designated +1. Residues -33 to -1 have the features of a signal sequence, including a positively charged amino terminus followed by a stretch of hydrophobic residues from -17 to -6 (von Heijne, 1985), and the sequence VRA immediately upstream of amino acid +1 conforms to the -3, -1 rule for the signal peptidase cleavage site (von Heijne, 1984). The mature B2 polypeptide consists of 1572 amino acids with an unglycosylated molecular weight of 173 541 and possesses 14 potential N-linked glycosylation sites (N-X-S/T) and 98 cysteine residues, most of which are located in 2 cysteine-rich domains.

**Structural Features of the B2 Chain.** Inspection of the amino acid distribution and prediction of the protein secondary structure according to the method of Garnier et al. (1978) reveals that the B2 molecule is organized into distinct structural domains highly analogous to those reported for the mouse and human laminin B1 chains (Sasaki et al., 1987; Pikkariainen et al., 1987). Starting at the carboxyl terminus, domain I (residues 1140-1572) of the B2 chain is rich in  $\alpha$ -helical structure and contains the heptad repeats characteristic of coiled-coil  $\alpha$ -helices noted previously (Barlow et al., 1984). Domain II (residues 994-1139) contains stretches of  $\alpha$ -helical structure interspersed with  $\beta$ -sheets and  $\beta$ -turns. Domains I and II contain 9 of the 14 potential N-linked glycosylation sites. The 2 cysteine-rich regions, domain III (residues 655-993) and domain V (residues 250-469), have 6.5 and 4.5 copies, respectively, of the homologous repeat unit first observed in the mouse B1 chain (Sasaki et al., 1987). The cysteines are organized into a repeat pattern of 47-56 amino acids containing 8 cysteines per repeat (Figure 3). Eleven other positions in the repeat are highly conserved. There is an invariant number of amino acids between the fifth and seventh cysteines, with 5 of the 11 intervening amino acids

Table I: Comparison of the Domains of Laminin B2 and B1<sup>a</sup>

domain	feature	B2 residues	B1 residues
I	$\alpha$ -helix	433	356
$\alpha$	cysteine rich		33
II	$\alpha$ -helix	146	220
III	homologous repeats	339	406
IV	globular	185	231
V	homologous repeats	220	271
VI	globular	249	248

<sup>a</sup> The protein secondary structure of residues 1-1572 in Figure 2 was predicted by the method of Garnier et al. (1978). Data for the mouse laminin B1 chain were taken from Sasaki et al. (1987).

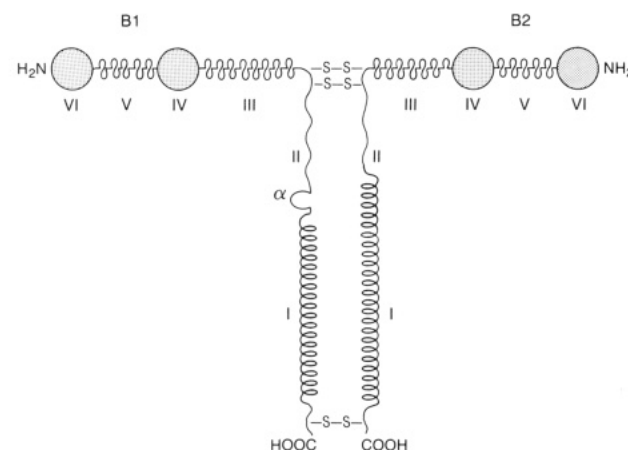


FIGURE 4: Structural model of the laminin B1-B2 dimer. Domains I and II of the laminin B1 and B2 chains form a paired coiled-coil  $\alpha$ -helix in the rodlike segment of the long arm of the cross, and domains III-VI form two of the short arms. Cysteine residues that may participate in interchain disulfide bond formation are indicated (-S-). The A chain is presumed to form the third short arm and the globular domain at the end of the long arm and also to participate in the  $\alpha$ -helical region of the long arm. A summary of the experimental data that contributed to the formulation of the model can be found in Engel and Furthmayr (1987).

being highly conserved. Domain IV (residues 470-654), which separates the two cysteine-rich domains, has no regular secondary structural features and probably adopts a globular conformation. The amino-terminal region, domain VI (residues 1-249), is probably globular like domain IV, since it has no extended secondary structures and consists of  $\beta$ -turns, random coils, and  $\alpha$ -helices. There are eight cysteine residues in this segment that could be involved in intramolecular disulfide bonding. Each of the globular domains has two potential N-linked glycosylation sites.

**Model of the Laminin B1-B2 Dimer.** In Table I, the corresponding domains of the mouse laminin B1 and B2 chains

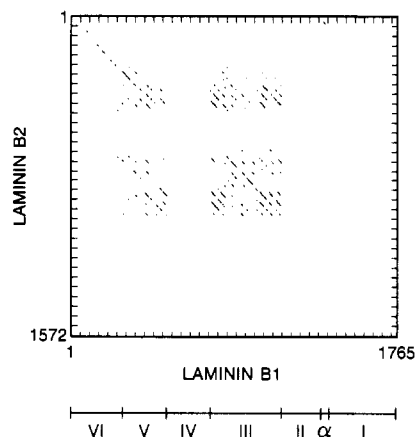


FIGURE 5: Comparison of the mouse laminin B1 and B2 chain amino acid sequences by dot matrix. Residues 1–1572 of laminin B2 from Figure 2 and residues 1–1765 of laminin B1 from Sasaki et al. (1987) were compared by dot matrix using the Homology program. The range was 12 residues, and the homology cutoff was 68%. Below the B1 axis is a diagram showing the corresponding domains of the B1 polypeptide.

are compared, and in Figure 4 is a model of the proposed structure formed by the two. Domains I and II of the B1 and B2 chains run parallel to each other, forming a paired coiled-coil  $\alpha$ -helix in the rodlike portion of the long arm of the cross. The B2 chain lacks the 33 amino acid cysteine-rich domain of the B1 chain (Sasaki et al., 1987) that separates the  $\alpha$ -helical domains. The cysteine residues at 1764 in the B1 chain and at 1563 in the B2 chain probably form the covalent linkage between the two chains. Additional cysteines at 995 and 998 in B2 and at 1158 and 1161 in B1 could also take part in interchain disulfide bond formation. Domain III in B2 is smaller than its counterpart in B1, having 6.5 versus 8 copies of the cysteine-rich repeat unit. The two cysteine repeat domains (III and V) of the B1 and B2 chains presumably form the rodlike segments of the short arms, and domains IV and VI form the globular domains of the short arms.

**Sequence Homologies of the Laminin B Chains.** The sequences of the mouse laminin B1 and B2 chains were compared by dot matrix analysis (Figure 5), and significant homology between the two is observed in domains III, V, and VI. Each of the cysteine-rich regions has a high degree of homology to the other cysteine-rich regions. Residues 967–975 in domain III of B2 have the sequence CREGFVGNR, and this is similar to a sequence (CDPGYIGSR, residues 925–933) in domain III of B1 that shows homology to epidermal growth factor and was found to be involved in cell attachment (Graf et al., 1987). The sequence similarities of the amino-terminal globular domains (VI) of the two chains are shown in detail in Figure 6. Both chains have eight cysteines in domain VI, and the positions of seven of these are conserved. There is little homology between B1 and B2 in the  $\alpha$ -helical domains (I and II) and in the other globular domain (IV).

**Features of the 3' Noncoding Region.** The laminin B2 cDNA has a 2.5 kb 3' untranslated region containing four AATAAA polyadenylation signals at positions 5203, 5464, 7468, and 7527 (Figure 2). Cleavage and polyadenylation apparently occur only after the final AATAAA, as Northern blot analysis has not detected more than one B2 RNA species (Barlow et al., 1984; Durkin et al., 1986; Kleinman et al., 1987). An unusual feature of the noncoding region of the B2 cDNA is the presence of a retroposon-like sequence between nucleotides 5984 and 6215, on the opposite strand. The sequence is bordered by a 7 bp direct repeat at both ends (GAAATGC) and has an A-rich region at the 3' end, char-

LAMININ	RESIDUES	SEQUENCE	SIMILARITY
B1	79–99	VTFAPNRLKIWWQSENGVEN	
B2	70–90	LTDYNNQADTTTHQSQMTLAG	5/21
B1	139–153	WGVYRYFYDCESSF	
B2	137–151	WIPXQYSGSCENTY	5/15
B1	166–184	DIICDSRYSDIEPSTEGEV	
B2	167–185	QALCTDEFSDISELTGGNY	8/19
B1	237–259	YYYAVYDMVVRGNCFCYGHASEC	
B2	238–260	YYYAISDFAYGGRCKCNHASEC	15/23

FIGURE 6: Comparison of the amino termini of the laminin B1 and B2 chains. Residues 1–260 of laminin B2 and 1–259 of laminin B1 were compared using the Homology program, and regions of similarity between the two are shown. Amino acids that are identical in both chains are underlined.

6170–6182 AGGCTTGGAGTGG  
A box consensus TGGC<sup>U</sup>NN--AGTGG

6129–6138 GTTGGCAGCC  
B box consensus GTTCGANNCC

FIGURE 7: RNA polymerase III promoter in the 3' noncoding region of the laminin B2 cDNA. The sequences on the opposite strand of nucleotides 6170–6182 and 6129–6138 in Figure 2 are compared to the consensus sequences for the A and B boxes of the RNA polymerase III split promoter (Galli et al., 1981). Nucleotides that match are underlined.

acteristic of retroposons (Weiner et al., 1986). Within the retroposon are sequences very similar to the consensus sequences for the A and B boxes of the RNA polymerase III split promoter (Galli et al., 1981) (Figure 7). The sequence shows no similarity to those of the two most abundant families of murine short interspersed repeats, B1 (Alu) and B2 (Krayev et al., 1982).

## DISCUSSION

Determination of the primary structure of the laminin B2 chain through cDNA cloning provides further support for the current model of laminin structure. Although the B1 and B2 chains are the products of separate genes, they have homologous multidomain structures. The similar organization of the chains and the sequence homology between the two in domains III, V, and VI suggest that they arose by duplication of an ancestral gene. The amino-terminal two-thirds of the B1 and B2 chain presumably form two of the morphologically similar short arms of the cross, and their carboxyl ends form a paired  $\alpha$ -helix in the elongated segment of the long arm. The model predicts that the laminin A chain will have regions similar to domains I through VI as well as a carboxyl-terminal extension to form the large globule at the end of the long arm (Engel & Furthmayr, 1987). This will be verified when the sequence of the A chain is known.

Evidence from biosynthetic studies indicates that the first step in the assembly of the laminin complex is the formation of a disulfide-linked B1–B2 dimer in the endoplasmic reticulum, followed by the addition of the A chain and transport to the Golgi apparatus for further processing (Morita et al., 1985; Peters et al., 1985). Covalent linkage of the B chains occurs through the cysteines close to their carboxyl termini (Paulsson et al., 1985) and possibly through additional cysteines located near the intersection of the arms. The B1 chain has the cysteine-rich  $\alpha$  domain not present in B2, and this may also be involved in disulfide bond formation, possibly with the A chain.

The amino terminus of the mature B2 chain is not known for certain, since attempts to sequence the polypeptide directly revealed that the amino terminus was blocked (unpublished observations). The alanine at position 1 was chosen as the amino terminus of the mature chain because residues 1–249

have many similarities with residues 1–248 of the B1 chain, and residues –33 to –1 are typical of a signal peptide. This predicts the unglycosylated B2 chain has a molecular weight of 173 541, close to the molecular weight of 180 000 estimated by SDS–polyacrylamide gel electrophoresis of the unglycosylated laminin subunits from tunicamycin-treated cells (Howe & Dietzschold, 1983).

After this work was complete, the sequence of the mouse laminin B2 cDNA was determined by Sasaki and Yamada (1987). Their sequence agrees with ours, with the following exceptions. Single-nucleotide substitutions change a D to an E at residue 231, a C to an S at residue 304, a P to an L at residue 414, and an S to a T at residue 629, and a two-nucleotide substitution changes an S to an R at residue 415. Their sequence also contains two additional amino acids: a P inserted between residues 852 and 853, and an A between residues 1121 and 1122. Several nucleotide differences were also observed in the 5' and 3' untranslated regions. The differences may represent sequence polymorphisms, errors in copying by reverse transcriptase, or dideoxy sequencing ambiguities.

The laminin B2 cDNA has 3' untranslated region of 2.5 kb, considerably longer than that of the B1 cDNA, 0.15 kb (Sasaki et al., 1987). The functional significance of this, in terms of differential mRNA stability or translatability, is not known. The noncoding region of the B2 cDNA contains a retroposon on the opposite strand; whether this is independently transcribed by RNA polymerase III has not been determined.

Several of the functions of laminin have been mapped to specific regions of the molecule. One cell binding site has been localized to a proteolytic fragment consisting of the central rodlike portions of the three short arms (Terranova et al., 1983; Timpl et al., 1983). This corresponds to domain III in the B chains, and a synthetic peptide with the sequence CDPGY-IGSR derived from domain III of the B1 chain has been found to promote cell attachment and migration (Graf et al., 1987). There is a similar sequence in domain III of B2, but whether it is also biologically active remains to be determined. Like B1, the B2 chain has no RGD sequence, the cell recognition site of extracellular matrix proteins such as fibronectin (Ruoslahti & Pierschbacher, 1986). Another cell binding site has been found in a fragment derived from the long arm of the cross (Aumailley et al., 1987), and this fragment also contains the neurite outgrowth-promoting activity (Edgar et al., 1984). The entactin (nidogen) binding site has been localized to the center of the cross (Paulsson et al., 1987), and the heparin binding activity has been mapped to the globular domain at the end of the long arm (Ott et al., 1982). The four terminal globules of the long and short arms have been implicated in laminin self-assembly (Yurchenco et al., 1985) and type IV collagen binding (Charonis et al., 1985).

The fact that the laminin B chains are similar but not identical raises the possibility that they may have somewhat different functions. The ratio of the mRNAs for the laminin A, B1, and B2 subunits has been found to vary greatly in different tissues (Kleinman et al., 1987). There may be tissue-specific differences in basement membrane structure and function due to variations in the subunit composition of laminin. Elucidation of the primary structures of the laminin chains through cDNA cloning will make it possible to analyze the role that each of the three chains plays in the many biological functions of laminin.

#### ACKNOWLEDGMENTS

We thank Ruth Ann Schulte for typing the manuscript, William Brent and Hana Novak for preparation of the figures,

Roger Hendrix for computer assistance, Jane Vergnes for help with the sequencing, Patricia Wells for oligonucleotide synthesis, and Joseph Kulkosky for analysis of the 3' noncoding sequence.

#### REFERENCES

- Aumailley, M., Nurcombe, V., Edgar, D., Paulsson, M., & Timpl, R. (1987) *J. Biol. Chem.* 262, 11532–11538.
- Barlow, D. P., Green, N. M., Kurkinen, M., & Hogan, B. L. M. (1984) *EMBO J.* 3, 2355–2362.
- Charonis, A. S., Tsilibary, E. C., Yurchenco, P. D., & Furthmayr, H. (1985) *J. Cell Biol.* 100, 1848–1853.
- Chen, E. Y., & Seeburg, P. H. (1985) *DNA* 4, 165–170.
- Chung, A. E., Jaffe, R., Freeman, I. L., Vergnes, J.-P., Braginski, J. E., & Carlin, B. (1979) *Cell (Cambridge, Mass.)* 16, 277–287.
- Cooper, A. R., Kurkinen, M., Taylor, A., & Hogan, B. L. M. (1981) *Eur. J. Biochem.* 119, 189–197.
- Davis, R. W., Botstein, D., & Roth, J. R. (1980) *Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Durkin, M. E., Phillips, S. L., & Chung, A. E. (1986) *Differentiation (Berlin)* 32, 260–266.
- Edgar, D., Timpl, R., & Thoenen, H. (1984) *EMBO J.* 3, 1463–1468.
- Engel, J., & Furthmayr, H. (1987) *Methods Enzymol.* 145, 3–78.
- Engel, J., Odermatt, E., Engel, A., Madri, J., Furthmayr, H., Rohde, H., & Timpl, R. (1981) *J. Mol. Biol.* 150, 97–120.
- Galli, G., Hofstetter, H., & Birnstiel, M. L. (1981) *Nature (London)* 294, 626–631.
- Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol.* 120, 97–120.
- Graf, J., Iwamoto, Y., Sasaki, M., Martin, G. R., Kleinman, H. K., Robey, F. A., & Yamada, Y. (1987) *Cell (Cambridge, Mass.)* 48, 989–996.
- Gubler, U., & Hoffman, B. J. (1983) *Gene* 25, 263–269.
- Henikoff, S. (1984) *Gene* 28, 351–359.
- Horwitz, A., Duggan, K., Greggs, R., Decker, C., & Buck, C. (1985) *J. Cell Biol.* 101, 2134–2144.
- Howe, C. C., & Dietzschold, B. (1983) *Dev. Biol.* 98, 385–391.
- Huynh, T. V., Young, R. A., & Davis, R. W. (1985) in *DNA Cloning, A Practical Approach* (Glover, D., Ed.) vol. I, pp 49–78, IRL, Oxford.
- Kleinman, H. K., Cannon, F. B., Laurie, G. W., Hassell, J. R., Aumailley, M., Terranova, V. P., Martin, G. R., & DuBois-Dalcq, M. (1985) *J. Cell. Biochem.* 27, 317–325.
- Kleinman, H. K., McGarvey, M. L., Hassell, J. R., Star, V. L., Cannon, F. B., Laurie, G. W., & Martin, G. R. (1986) *Biochemistry* 25, 312–318.
- Kleinman, H. K., Ebihara, I., Killen, P. D., Sasaki, M., Cannon, F. B., Yamada, Y., & Martin, G. R. (1987) *Dev. Biol.* 122, 373–378.
- Kozak, M. (1987) *Nucleic Acids Res.* 15, 8125–8148.
- Krayev, A. S., Markuseva, T. V., Kramerov, D. A., Ryskov, A. P., Skryabin, K. G., Bayev, A. A., & Georgiev, G. P. (1982) *Nucleic Acids Res.* 10, 7461–7475.
- Lawler, J., & Hynes, R. O. (1986) *J. Cell Biol.* 103, 1635–1648.
- Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Morita, A., Sugimoto, E., & Kitagawa, Y. (1985) *Biochem. J.* 229, 259–264.
- Ott, U., Odermatt, E., Engel, J., Furthmayr, H., & Timpl, R. (1982) *Eur. J. Biochem.* 123, 63–72.

- Paulsson, M., Deutzmann, R., Timpl, R., Dalzoppo, D., Odermatt, E., & Engel, J. (1985) *EMBO J.* 4, 309-316.
- Paulsson, M., Aumailley, M., Deutzmann, R., Timpl, R., Beck, K., & Engel, J. (1987) *Eur. J. Biochem.* 166, 11-19.
- Peters, B. P., Hartle, R. J., Krzesicki, R. F., Kroll, T. G., Perini, F., Balun, J. E., Goldstein, I. J., & Ruddon, R. W. (1985) *J. Biol. Chem.* 260, 14732-14742.
- Pikkarainen, T., Eddy, R., Fukushima, Y., Byers, M., Shows, T., Pihlajaniemi, T., Saraste, M., & Tryggvason, K. (1987) *J. Biol. Chem.* 262, 10454-10462.
- Rigby, P. W. J., Dieckmann, M., Rhodes, C., & Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
- Roberts, D. D., Rao, C. N., Magnani, J. L., Spitalnik, S. L., Liotta, L. A., & Ginsburg, V. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82, 1306-1310.
- Ruoslahti, E., & Pierschbacher, M. (1986) *Cell (Cambridge, Mass.)* 44, 517-518.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463-5467.
- Sasaki, M., & Yamada, Y. (1987) *J. Biol. Chem.* 262, 17111-17117.
- Sasaki, M., Kato, S., Kohno, K., Martin, G. R., & Yamada, Y. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 935-939.
- Smalheiser, N. R., & Schwartz, N. B. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 6457-6461.
- Terranova, V. P., Rao, C. N., Kalebic, T., Margulies, I. M., & Liotta, L. A. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 444-448.
- Timpl, R., Rohde, H., Robey, P. G., Rennard, S. I., Foidart, J.-M., & Martin, G. R. (1979) *J. Biol. Chem.* 254, 9933-9937.
- Timpl, R., Johansson, S., van Delden, V., Oberbaumer, I., & Hook, M. (1983) *J. Biol. Chem.* 258, 8922-8927.
- Von der Mark, K., & Kuhl, U. (1985) *Biochim. Biophys. Acta* 823, 147-160.
- Von Heijne, G. (1984) *J. Mol. Biol.* 173, 243-251.
- Von Heijne, G. (1985) *J. Mol. Biol.* 184, 99-105.
- Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., & Itakura, K. (1979) *Nucleic Acids Res.* 6, 3543-3557.
- Weiner, A. M., Deininger, P. L., & Efstratiadis, A. (1986) *Annu. Rev. Biochem.* 55, 631-661.
- Young, R. A., & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 1194-1198.
- Yurchenco, P. D., Tsilibary, E. C., Charonis, A. S., & Furthmayr, H. (1985) *J. Biol. Chem.* 260, 7636-7644.

## 5-Fluorocytosine in DNA Is a Mechanism-Based Inhibitor of *HhaI* Methylase<sup>†</sup>

David G. Osterman,<sup>†</sup> Gia D. DePillis,<sup>‡</sup> John C. Wu,<sup>‡</sup> Akira Matsuda,<sup>§</sup> and Daniel V. Santi<sup>\*,†,§</sup>

Department of Biochemistry and Biophysics and Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143, and Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo 060, Japan

Received February 12, 1988; Revised Manuscript Received March 22, 1988

**ABSTRACT:** 5-Fluorodeoxycytidine (FdCyd) was incorporated into a synthetic DNA polymer containing the GCGC recognition sequence of *HhaI* methylase to give a polymer with about 80% FdCyd. In the absence of AdoMet, poly(FdC-dG) bound competitively with respect to poly(dG-dC) ( $K_i = 3$  nM). In the presence of AdoMet, the analogue caused a time-dependent, first-order ( $k = 0.05$  min<sup>-1</sup>) inactivation of the enzyme. There is an ordered mechanism of binding in which enzyme first binds to poly(FdC-dG), then binds to AdoMet, and subsequently forms stable, inactive complexes. The complexes did not dissociate over the course of 3 days and were stable to heat (95 °C) in the presence of 1% SDS. Gel filtration of a complex formed with *HhaI* methylase, poly(FdC-dG), and [methyl-<sup>3</sup>H]AdoMet gave a peak of radioactivity eluting near the void volume. Digestion of the DNA in the complex resulted in a reduction of the molecular weight to the size of the methylase, and the radioactivity in this peak was shown to be associated with protein. These data indicate that the complexes contain covalently bound *HhaI* methylase, poly(FdC-dG), and methyl groups and that 5-fluorodeoxycytidine is a mechanism-based inactivator of the methylase. By analogy with other pyrimidine-modifying enzymes and recent studies on the mechanism of *HhaI* methylase (Wu & Santi, 1987), these results suggest that an enzyme nucleophile attacks FdCyd residues at C-6, activating the 5-position for one-carbon transfer. Subsequent transfer of the methyl group of AdoMet to the activated FdCyd forms a stable complex in which the enzyme is covalently bound to the 6-position of FdCyd in the polymer and a methyl group is attached to C-5. The effect of 5-fluorodeoxycytidine on the inhibition of DNA-cytosine methyltransferases is thus due to irreversible, covalent inactivation.

**D**NA-cytosine methyltransferases (DCMT)<sup>1</sup> (EC 2.1.1.37) catalyze the transfer of the activated methyl group of *S*-adenosylmethionine (AdoMet) to the C-5 position of cytosine.

In bacteria, modification by methylation prevents the digestion of host DNA by the host restriction endonucleases. Foreign DNA lacking this pattern of methylation is degraded (Arber, 1974). In mammals, this postreplicative modification is an important element in the control of gene expression. Changes

<sup>†</sup> This investigation was supported by USPHS Grant CA 14394 from the National Cancer Institute and a grant from the University of California Cancer Research Coordinating Committee.

\* To whom correspondence should be addressed.

<sup>‡</sup> University of California.

<sup>§</sup> Hokkaido University.

<sup>||</sup> D.V.S. is a Burroughs Wellcome Scholar in Molecular Parasitology.

<sup>1</sup> Abbreviations: AdoMet, *S*-adenosyl-L-methionine; azaC, 5-azacytidine; BSA, bovine serum albumin; DCMT, DNA-cytosine methyltransferase; DNase I, deoxyribonuclease I; DTT, dithiothreitol; FdCyd, 5-fluorodeoxycytidine; FdUrd, 5-fluorodeoxyuridine; FdUMP, 5-fluorodeoxyuridylic acid; poly(FdC-dG), poly(5-fluoro-dC-dG).