New York.
Diliberto, E. J., & Axelrod, J. (1976) *J. Neurochem. 26*, 1159-1165.
Duerre, J. A., & Gaitonde, M. K. (1971) *J. Neurochem. 18*, 1921-1929.
Duerre, J. A., Traynor, M. D., & Leatherbury, K. C. (1985) in *Progress in Nonhistone Chromosomal Proteins* (Beckhor, I., Ed.) Vol. 1, pp 103-116, CRC Press, Boca Raton, FL.
Gagnon, C. (1979) *Biochem. Biophys. Res. Commun. 88*, 847-853.
Gagnon, C. (1982) in *Biochemistry of S-Adenosylmethionine and Related Compounds* (Usdin, E., Borchardt, R., & Creveling, C. R., Eds.) pp 55-64, Macmillan, London.
Gagnon, C., Viveros, J., Diliberto, E. J., & Axelrod, J. (1978) *J. Biol. Chem. 253*, 3778-3781.
Gerner, E. W., & Humphrey, R. M. (1973) *Biochim. Biophys. Acta 331*, 117-127.
Green, G., Sikka, S. C., & Kalra, V. J. (1983) *J. Biol. Chem. 258*, 12958-12966.
GuGuen-Guillouzo, C., Tichonicky, L., & Kruk, J. (1979) *Eur. J. Biochem. 95*, 235-238.
Kim, S., & Paik, W. K. (1970) *J. Biol. Chem. 245*, 1806-1813.
Kim, S., & Li, C. H. (1979) *Int. J. Pept. Protein Res. 13*, 282-286.
Kim, S., Nochumson, S., Chin, W., & Paik, W. K. (1978) *Biochemistry 17*, 415-422.
Kloog, Y., Axelrod, J., & Spector, I. (1983) *J. Neurochem. 40*, 522-529.
Kostraba, N. C., & Wang, T. Y. (1973) *Exp. Cell Res. 80*, 291-296.
Krayse, M. O., Kleinsmith, L. J., & Stein, G. S. (1975) *Life Sci. 16*, 1047-1058.
Laemmli, U. K. (1970) *Nature (London) 227*, 680-685.
Layne, E. (1957) *Methods Enzymol. 3*, 447.
Liss, M., & Edelstein, L. M. (1967) *Biochem. Biophys. Res. Commun. 26*, 497-504.
McFadden, P., & Clarke, S. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2460-2464.
O'Connor, C., & Clarke, S. (1983) *J. Biol. Chem. 258*, 8485-8492.
O'Dea, R. F., Viveros, O. H., & Axelrod, J. (1978) *Nature (London) 272*, 462-464.
O'Dea, R. F., Viveros, O. H., & Diliberto, E. J. (1981) *Biochem. Pharmacol. 30*, 1163-1168.
Paik, W. K., & Kim, S. (1980) in *Protein Methylation* (Meister, A., Ed.) Vol. 1, pp 202-231, Wiley, New York.
Quick, D. P., Orchard, P. J., & Duerre, J. A. (1981) *Biochemistry 20*, 4724-4729.
Schlenk, F., & Depalma, R. F. (1957) *J. Biol. Chem. 229*, 1051-1057.
Stock, J. B., & Koshland, D. E. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 3659-3663.
Swanson, R. J., & Applebury, M. L. (1983) *J. Biol. Chem. 258*, 10599-10605.
van Waarde, A. (1982) *FEBS Lett. 149*, 266-270.
van Waarde, A., & van Haastert, P. J. M. (1984) *J. Bacteriol. 157*, 368-3749.
Venkatasubramanian, K., Hirata, F., Gagnon, C., Corcoran, B. A., O'Dea, R. F., Axelrod, J., & Schiffman, E. (1980) *Mol. Immunol. 17*, 201-207.
Webb, J. L. (1963) in *Enzyme and Metabolic Inhibitors*, Vol. 1, p 84, Academic Press, New York.
Zukerman, S. H., O'Dea, R. F., Olson, J. M., & Douglas, S. D. (1982) *Mol. Immunol. 19*, 281-286.

Characterization of the Bovine Prothrombin Gene†

David M. Irwin,‡ Kevin G. Ahern,§ George D. Pearson,§ and Ross T. A. MacGillivray*,‡
Department of Biochemistry, The University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5, and Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331-6503
Received April 12, 1985

ABSTRACT: The bovine prothrombin gene was characterized by Southern blot analysis of bovine genomic DNA using bovine prothrombin cDNA fragments as hybridization probes. These analyses suggested that the bovine genome contains a single prothrombin gene that is at least 10 kilobase pairs (kbp) in size. To characterize the gene more thoroughly, two bovine genomic phage libraries were screened by using prothrombin cDNAs as hybridization probes. Heteroduplex analysis of the cloned genomic DNA and cDNA showed that the prothrombin gene is 14.9 kbp in size and contains at least 14 exons interrupted by 13 introns. The exons vary in size from 28 to 317 base pairs (bp), while the introns vary in size from <100 to 6940 bp. Regions of self-complementarity were observed within some of the introns, suggesting the presence of inverted repeat sequences. The bovine prothrombin gene shows similarities in structure to both the human prothrombin gene and the human factor IX gene.

The serine proteases are a family of structurally related proteins that are involved in a variety of essential physiological functions [for reviews, see Neurath & Walsh (1976) and Neurath (1984)]. These proteases are involved in both basic physiological processes, such as digestion and fertilization, and highly specialized physiological processes, such as the vertebrate immune response and blood coagulation systems. The serine proteases share regions of amino acid sequence identity particularly around the active site and the substrate binding sites (Hartley, 1970; Greer, 1981). This sequence identity is also found in prokaryotic serine proteases such as trypsin and proteases A and B from *Streptomyces griseus* (Olafson et al.,

† This work was supported in part by grants from the Medical Research Council of Canada and the British Columbia Health Care Research Foundation (to R.T.A.M.), by a Faculty Research Award (FRA-236) from the American Cancer Society, by a grant from the National Cancer Institute (CA17699), and by a grant from the Murdock Charitable Trust (to G.D.P.).
‡ The University of British Columbia.
§ Oregon State University.

0006-2960/85/0424-6854$01.50/0 © 1985 American Chemical Society

1975; Johnson & Smillie, 1974; Jurasek et al., 1974) and α-lytic protease from *Myxobacter sorangium* 495 (Olson et al., 1970). X-ray crystallographic studies have shown that mammalian chymotrypsin and *S. griseus* protease B have very similar three-dimensional structures (Delbaere et al., 1975), suggesting that the catalytic mechanism of proteolytic cleavage is similar in each of these proteins. It has been proposed that the present-day serine protease genes have evolved from a common ancestral gene as a result of a series of gene duplication events [see Neurath (1984)]. Following the duplications, the genes have diverged to give rise to the large number of genes coding for the serine proteases found today. For example, the clotting of blood involves at least seven serine proteases participating in two different pathways (Jackson & Nemerson, 1980). The proteases are found in plasma as inactive zymogens that are activated by limited proteolysis. The clotting proteases share amino acid sequence identity with the other members of the serine protease family but differ by having limited substrate specificities and by containing additional functional domains [see Jackson & Nemerson (1980)].

Prothrombin is a plasma glycoprotein of M_r 65 000 (Magnusson et al., 1975) that has been detected in many species including lamprey (Doolittle et al., 1962). The complete amino acid sequences of bovine (Magnusson et al., 1975) and human (Butkowski et al., 1977; Walz et al., 1977) prothrombin have been determined. The carboxy-terminal region of prothrombin contains the catalytic domain that shares amino acid sequence homology with trypsin [see Jackson & Nemerson (1980)]. Trypsinogen and prothrombin are activated by limited proteolysis of homologous peptide bonds in the respective sequences

Lys-↓Ile-Val-Gly-Gly and Arg-↓Ile-Val-Glu-Gly

These bond cleavages result in the release of activation peptides and the formation of the proteases trypsin and thrombin, respectively. In contrast to trypsinogen, which has an activation peptide of only six residues, activation of prothrombin by factor Xa leads to the release of an activation peptide of 274 residues (Owen et al., 1974; Magnusson et al., 1975). The amino-terminal region of this activation peptide contains 12 residues of γ-carboxyglutamic acid (Gla)[1] and is homologous to the amino-terminal regions of factor IX, factor X, factor VII, protein C, and protein S [see Jackson & Nemerson (1980)]. The Gla residues arise as a result of the posttranslational modification of glutamate residues by a vitamin K dependent carboxylase [see Stenflo & Suttie (1977) for a review]. The Gla residues bind calcium and help to anchor these clotting factors to the site of injury via calcium bridges. The activation peptide of prothrombin also contains two homologous regions called kringles (Magnusson et al., 1975). Kringles have also been found in plasminogen (Sottrup-Jensen et al., 1978), plasminogen activator (Pennica et al., 1983), urokinase (Verde et al., 1984), and factor XII (McMullen & Fujikawa, 1984; Cool et al., 1985), but their function is unclear. Thus, prothrombin contains the core catalytic region found in all serine proteases but also contains regions found in a variety of other proteins.

It has been suggested that complex genes (such as the prothrombin gene) may have evolved by recombination of separate exons that code for domains of proteins (Gilbert, 1979; Lunberg & Gilbert, 1985). This type of an evolutionary process may be reflected by the organization of the introns within a gene. In this paper, we describe the cloning and characterization of the bovine prothrombin gene.

Experimental Procedures

Materials

Restriction enzymes were purchased from Bethesda Research Laboratories (BRL) (Gaitherburg, MD) and New England Biolabs (Beverly, MA) and were used according to the manufacturers' recommendations. BRL also supplied T4 DNA ligase. *Escherichia coli* DNA polymerase I, DNA polymerase I (Klenow fragment), and deoxyribonuclease I were purchased from Boehringer Mannheim. ^{32}P-labeled nucleotides were purchased from New England Nuclear and Amersham. P-L Biochemicals supplied deoxy- and dideoxynucleotides and the M13 sequencing primer (pentadecanucleotide). Nitrocellulose was obtained from Schleicher & Schuell and from Millipore. Formamide (BRL) was deionized by treatment with AG 501-X8 mixed-bed ion-exchange resin (Bio-Rad Laboratories).

Methods

Isolation of Bovine Genomic DNA. Bovine genomic DNA was isolated from liver by the method of Blin & Stafford (1976).

Southern Blot Analysis. Bovine liver genomic DNA (10 μg) was digested with various restriction endonucleases (20 units) at 37 °C for 16 h, and the resulting DNA fragments were separated by electrophoresis on either 0.8% or 1.0% agarose gels. After denaturation, the DNA was transferred to nitrocellulose as described by Southern (1975). Blots were analyzed by using restriction endonuclease fragments previously labeled to specific activities of (0.5–1) × 10^8 cpm/μg by nick translation (Maniatis et al., 1975) as hybridization probes. Hybridization and washing conditions were as described by Kan & Dozy (1978). Blots were exposed to Kodak XRP-1 X-ray film at −70 °C with intensifying screens for 1–6 days.

Northern Blot Analysis. Total bovine liver RNA (10 μg) was treated with glyoxal, electrophoresed on a 1% agarose gel, and transferred to nitrocellulose as described by Thomas (1980). Hybridization and washing conditions were as described by Thomas (1980) except that dextran sulfate was omitted and the hybridization time was increased to 48 h. After being washed, the blot was exposed to X-ray film with intensifying screens for 7 days at −70 °C.

Construction of a Bovine Liver DNA Library in λ1059. Bovine liver genomic DNA was partially digested with *Sau*3AI, and fragments (15–25 kbp in size) were recovered by agarose gel electrophoresis. These DNA fragments (5 μg) were ligated with λ1059 DNA (10 μg) (Karn et al., 1980) that had been digested with *Bam*HI. The ligated DNA was packaged into phage λ particles in vitro (Sternberg et al., 1977), and aliquots of the packaged mix were plated on *E. coli* strains Q358 and Q359 (Karn et al., 1980). The library contained 1.3 × 10^6 independent recombinants when plated on Q359. The phage library was amplified as described by Maniatis et al. (1982).

Screening of Bovine Genomic Phage Libraries. In addition to the λ1059 phage library, a bovine phage library constructed in Charon 28 was generously provided by Dr. Fritz Rottman (Department of Molecular Biology and Microbiology, Case Western Reserve University, Cleveland, OH). Both phage libraries were screened by plaque hybridization (Benton & Davis, 1977; Woo, 1980) using cloned bovine prothrombin cDNAs (MacGillivray & Davie, 1984) as hybrdization probes.

[1] Abbreviations: bp, base pair(s); kbp, kilobase pair(s); Gla, γ-carboxyglutamic acid; EDTA, ethylenediaminetetraacetic acid; ir, inverted repeat.

Table I: Subclones of Cloned Phage DNA Used in Heteroduplex Analysis

subclone	flanking restriction sites	location of cloned fragment[a]
pHE2	*Hin*dIII, *Eco*RI	11.7–16.6
pBGII21	*Bam*HI, *Sau*3AI	15.4–20.4
pBGII3	*Bam*HI, *Bam*HI	15.4–21.9

[a] Location is given in kilobase pairs relative to the restriction map in Figure 1.

Positive phage were plaque-purified, and DNA was extracted (Maniatis et al., 1982).

Restriction Endonuclease Mapping. DNA from positive phage was digested with restriction enzymes, and the resulting DNA fragments were analyzed by agarose gel electrophoresis. The *Bam*HI, *Eco*RI, and *Hin*dIII fragments of each λ clone were subcloned into pUC13 (Vieira & Messing, 1982) to facilitate restriction mapping.

Heteroduplex Analysis. Heteroduplexes were formed between *Eco*RI- or *Pst*I-cut prothrombin cDNA plasmids (pBII102 or pBII111) and DNA either from λ clones containing bovine genomic sequences (λBII1, λBII2, or λBII3) or from appropriately cleaved subclones of bovine genomic sequences (pHE2, pBGII3, or pBGII21; see Table I). DNA molecules (100 ng of each species) were denatured together in 10 μL of 80% formamide by heating at 70 °C for 10 min. Hybridization occurred at 37 °C for 1 h in reaction mixtures (20 μL) containing 50% formamide and 200 mM NaCl. DNA spreading conditions were essentially as described (Chow & Broker, 1981). The entire duplex mixture was spread as hyperphase (40 μL) containing 50% formamide/100 mM NaCl/5 mM EDTA, 100 ng of DNA added as a length standard, and cytochrome *c* at a concentration of 40 μg/mL. The hypophase (20 mL) was 20% formamide. The DNA–protein film was adsorbed to a Parlodion-coated grid, stained with uranyl acetate, and rotary-shadowed with platinum–palladium. Grids were examined with a Zeiss EM-10A electron microscope operating at 60 kV. Molecular lengths were measured on a video display using a Videoplan II image analysis system. Single-stranded DNA measurements were converted to double-stranded lengths by using the factor 1.16 to correct for compression during spreading.

Sequence Analysis. *Sau*3AI fragments of pBGII21, a subclone of λBII2 (see Table I), were cloned into the *Bam*HI site of M13mp7 (Messing, 1983). Plaques containing prothrombin coding sequence were identified by plaque hybridization (Benton & Davis, 1978) using nick-translated cDNA probes. Positive M13 clones were sequenced by the chain termination method (Sanger et al., 1977). DNA sequence data were stored and analyzed by using the computer programs of Delaney (1982).

Results

Southern Blot Analysis of the Bovine Prothrombin Gene. As an initial step toward the characterization of the bovine prothrombin gene, bovine liver DNA was digested with several restriction endonucleases, and the resulting fragments were separated by agarose gel electrophoresis. After denaturation, the DNA fragments were transferred to nitrocellulose and analyzed with ^{32}P-labeled hybridization probes derived from cloned bovine prothrombin cDNAs. Several bovine prothrombin cDNA clones have been described (MacGillivray & Davie, 1984) including pBII111 (that contains DNA coding for 5 bp of 5′-untranslated sequence and DNA coding for residues −43 to 579 of prothrombin) and pBII102 [that contains DNA coding for residues 69–582, a stop codon, 119 bp of 3′-untranslated sequence, and a poly(A) tail]. When the Southern blots of bovine genomic DNA were analyzed with the cDNA inserts of both pBII111 and pBII102 as hybridization probes, several fragments were detected with each of the restriction enzymes used (data not shown). The intensities of the bands were similar to those found when pBII111 DNA was included in the blot at a concentration equivalent to a single copy gene (data not shown). When the 5′ or 3′ ends of the cDNA were used as hybridization probes, single restriction fragments were detected with many of the enzymes used (data not shown), suggesting that the bovine genome contains a single gene coding for prothrombin. From these blots, it was estimated that the prothrombin gene was at least 10 kbp in length.

Northern Blot Analysis of Bovine Prothrombin mRNA. To determine the size of prothrombin mRNA, bovine liver RNA was denatured with glyoxal and electrophoresed on an agarose gel. After transfer to nitrocellulose, the RNA was hybridized to the ^{32}P-labeled cDNA insert of pBII111. Autoradiography of the blot revealed a single band that was 2150 ± 100 nucleotides in size (data not shown). This is similar to the size of human prothrombin mRNA (Degen et al., 1983). The cloned bovine cDNAs contain 1998 bp of coding sequence plus 3′-untranslated sequence. Thus, the 5′-untranslated sequence plus poly(A) tail represents 150 ± 100 nucleotides in prothrombin mRNA. As eukaryotic mRNAs usually contain poly(A) tails of 180–200 nucleotides (Perry, 1976), we conclude that our cDNA clones may be missing about 50 bp of 5′-untranslated sequence.

Cloning of the Bovine Prothrombin Gene. To study the bovine prothrombin gene more thoroughly, a bovine genomic phage library was constructed by using bovine liver DNA cloned into the *Bam*HI site of λ1059. One million phage from this library were screened by using the cDNA insert of pBII102 as a hybridization probe. Two independent positives were isolated, λBII1 and λBII2. Restriction endonuclease mapping and Southern blot analysis showed that these phage contained overlapping DNA and represented 25 kbp of contiguous bovine genomic DNA (Figure 1). Southern blot analysis showed that these phage contained most of the prothrombin gene but lacked the 3′ region. The λ1059 library was subsequently rescreened by using the 3′ *Bam*HI–*Pst*I fragment of pBII102 as a hybridization probe, but these screens only resulted in the reisolation of λBII1.

To isolate the 3′ end of the prothrombin gene, 10^6 phage of a second bovine liver genomic library (from Dr. Fritz Rottman) were screened by using the *Bam*HI–*Pst*I fragment of pBII102 as a probe. Three different clones, λBII3, λBII4, and λBII5, were identified and plaque-purified. Restriction enzyme mapping showed that these phage clones overlapped λBII1 and λBII2 at positions that were 3′ to the mapped prothrombin gene (Figure 1). λBII3 and λBII4 contained restriction fragments that were consistent with those detected in the genomic Southern blots with the 3′ probe. λBII5 did not contain these 3′-most exons (see Figure 1) but was isolated because it contained exon 12, a part of which is contained in the *Bam*HI–*Pst*I fragment used as a probe. A total of 42.4 kbp of continuous genomic DNA was represented by the five phage (λBII1–λBII5). This region contained all restriction enzyme fragments detected in the genomic Southern blot analysis. The prothrombin gene maps to a 15 kbp region in the middle of this cloned DNA (see next section).

Heteroduplex Analysis of the Cloned Bovine Prothrombin Gene. The positions of introns within the prothrombin gene

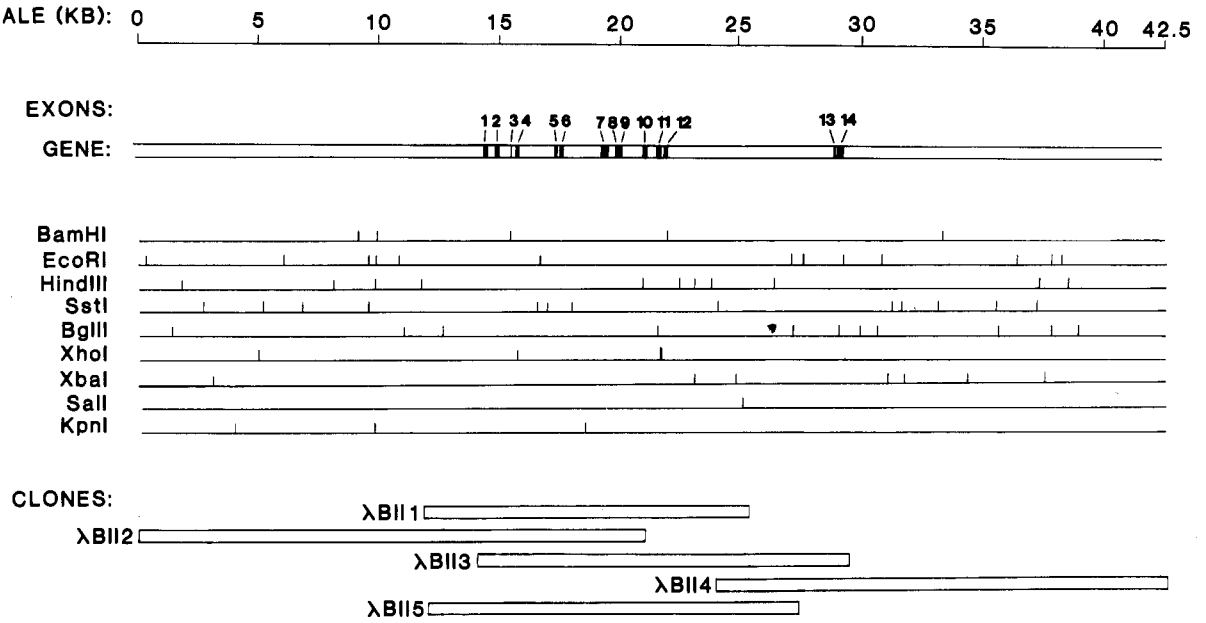

FIGURE 1: Restriction map of the region of the bovine genome containing the prothrombin gene. The restriction map was determined by analysis of the five recombinant phage λBII1–5 and subclones derived from them. The location of the prothrombin gene within this region is indicated (data derived from Table II). Exons are represented by black boxes and have been numbered from the 5′ end of the gene. The scale at the top represents nucleotides in kilobase pairs.

Table II: Lengths of Exons and Introns in the Prothrombin Gene

exon	length[a]	intron	length[a]
1	98 ± 14 (43)	A	261 ± 46 (48)
2	168 ± 18 (46)	B	601 ± 62 (49)
3	28 ± 8 (52)	C	170 ± 39 (68)
4	53 ± 13 (68)	D	1574 ± 73 (50)
5	103 ± 13 (56)	E	112 ± 19 (50)
6	139 ± 15 (54)	F	1381 ± 99 (54)
7	317 ± 26 (60)	G	235 ± 23 (58)
8	137 ± 15 (41)	H	<100
9	117 ± 16 (41)	I	1055 ± 94 (51)
10	170 ± 19 (56)	J	397 ± 46 (48)
11	159 ± 19 (56)	K	216 ± 29 (56)
12	160 ± 17 (62)	L	6940 ± 255 (41)
13	65 ± 10 (90)	M	<100
14	227 ± 17 (39)		

[a] Expressed as mean ± standard deviation in base pairs where the number of measurements is in parentheses.

Table III: Location of Restriction Sites within the Prothrombin Gene

restriction site	coordinate[a]	location in gene[b]
*Bam*HI	15.4	<100 bp 5′ from exon 3 in intron B
*Sst*I	16.4[c]	673 ± 69 (38) bp 3′ from exon 4 in intron D
*Sau*3AI	20.4	798 ± 69 (14) bp 3′ from exon 8 in intron I
*Bam*HI	21.9	124 ± 17 (48) bp 3′ from intron K in exon 12

[a] See Table I. [b] Expressed as mean ± standard deviation where the number of measurements is in parentheses. [c] *Sst*I-cut pHE2.

Table IV: Lengths and Locations of Inverted Repeat Sequences within Introns

feature[a]	length[b]
stem	119 ± 26 (9)
ir	387 ± 27 (45)
a	586 ± 57 (40)
b	129 ± 23 (45)
c	4456 ± 186 (46)
d	2117 ± 109 (30)
loop[c]	5692 ± 234 (37)

[a] See Figure 4. [b] Same as Table II. [c] Separation between ir sequences in "snapback" DNA; see text.

were determined by heteroduplex analysis using the electron microscope. Heteroduplexes were formed between prothrombin cDNAs and either recombinant phage λ DNAs or subclones derived from the phage DNA (Table I). Typical heteroduplexes are shown in Figure 2. As summarized in Table II and Figure 1, the prothrombin gene is encoded by at least 14 exons extending over 15 kbp of DNA. Introns H and M were less than 100 bp in length, but detectable. The exons varied in size from 28 to 317 bp, and the introns ranged in size from <100 to 6940 bp. The total length of exons measured by electron microscopy was 1941 bp, accounting within experimental error for all of the cDNA sequence (1998 bp). The 5′ to 3′ orientation of the gene was deduced from the observation that heteroduplexes between genomic prothrombin DNA and pBII102, which lacks 5′ cDNA sequences, did not contain intron loops A through D (not shown). Several restriction endonuclease cleavage sites within the gene were also mapped by electron microscopy (Table III). This information allowed the alignment of the gene with the restriction map of the cloned bovine genomic sequences. On the basis of these measurements, the 5′ end of the gene was positioned at coordinate 14.3 and the 3′ end at coordinate 29.2, a distance of 14.9 kbp. Regions of self-complementarity were observed within the λ genomic clones, an indication of inverted repeat sequences; Figure 2C, D shows a stem at the base of intron F, and Figure 2E, identifies inverted repeat (IR in figure) sequences shared between introns I and L. Table IV summarizes the measurements of these features. Under the hybridization conditions used to form the heteroduplexes, the stem and ir sequences were not observed in every case, suggesting that these repeat sequences do not represent perfectly matched hybrids. Although the stem in intron F is roughly the size of an Alu-like element already described in the bovine genome (Watanabe et al., 1982), further characterization will require nucleotide sequence analysis. Using data in Table IV, it was possible to position the ir sequences within introns I and L. Direct measurement of the loop between ir sequences in "snapback" genomic DNA established a separation of 5692 bp. Analysis of the lengths of the introns and exons in this

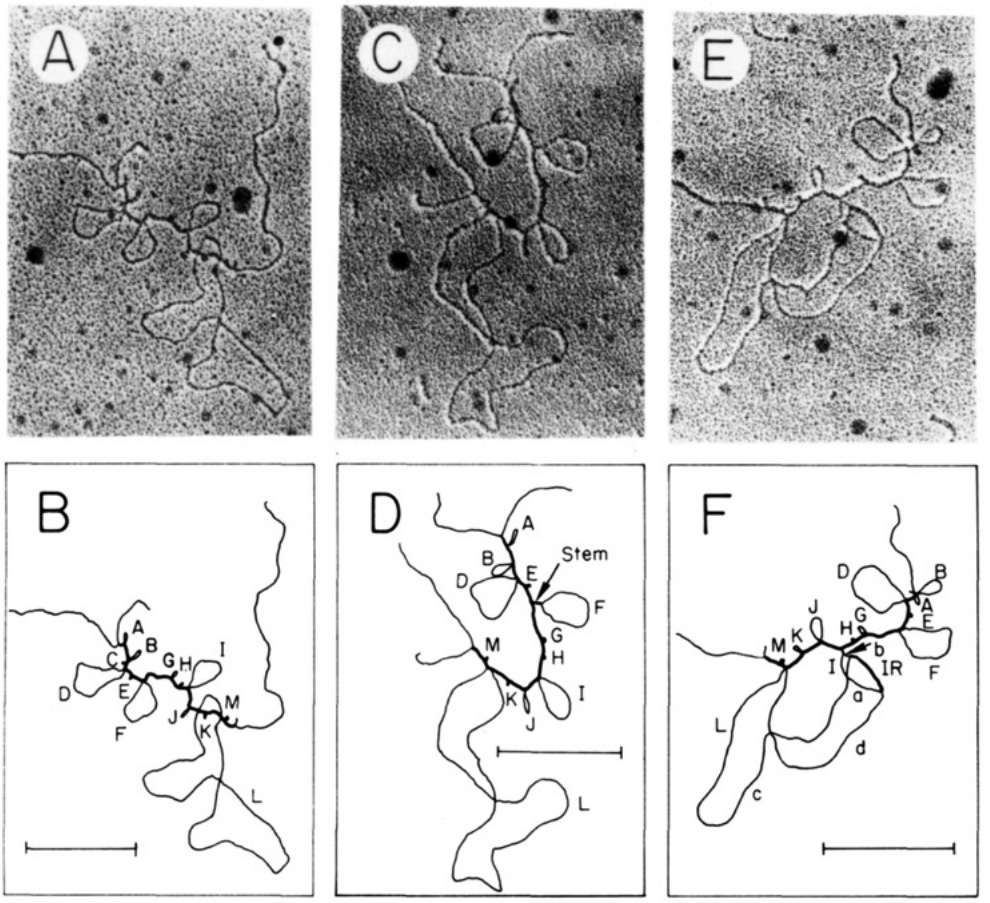

FIGURE 2: Analysis by electron microscopy of heteroduplexes formed between cloned bovine genomic DNA (λBII3) and cloned prothrombin cDNA (pBII111). Three representative heteroduplexes are shown together with interpretive drawings below each photograph. The thin line is single-stranded DNA, and the thick line is double-stranded DNA. The bar in each panel represents 1 kbp. Introns are lettered A–M starting from the 5′ end of the gene where intron A is flanked by exons 1 and 2 (see Figure 1). Stem refers to inverted repeat sequences in intron F. IR indicates inverted repeat sequences shared between introns I and L where a–d locate the positions of IR within each intron (see Table IV).

```
GAT CCC TTT CAC AGG CAA GCC GAA TGC GGC TCA GAG AGG TTA AGT GAC TTG ACA TCG TAC
                          20                          40                          60

AGG TCA AAG GTC AGG GGG CAG AGC TTC CCA TCC TGT CTG GCT CCA CTG TGG TCC AAC GCT
                          80                          100                         120

                                                  145                         154
                                              (G)ly Gln Asp Arg Val Thr Val Glu Val Ile
CTC TGC CCC CTC TCT CTC CTC ACC CAC CAG ▲GC CAG GAC CGA GTC ACA GTG GAG GTG ATC
                       140                        160                         179
```

FIGURE 3: DNA sequence analysis of a region of λBII2. The DNA sequence of the coding strand is shown together with the predicted amino acid sequence of the exon (coding for amino acid residues 145–154 of bovine prothrombin). The position of the putative splice site is shown by the arrow. The first nucleotide of the codon for Gly-145 is probably located in the preceding exon.

region of the gene showed that the size of this loop corresponds best to the arrangement 5′-b-10-J-11-K-12-c-3′ (a distance of 5684 bp), where 10, 11, and 12 refer to exons, J and K refer to introns, a and b represent DNA flanking the ir sequence in intron I, and c and d represent DNA flanking the ir sequence in intron L (see Figure 2F). Therefore, the order in intron I is 5′-a-ir-b-3′, and in intron L, it is 5′-c-ir-d-3′. Another inverted repeat shared between intron L and the genomic sequences distal to the 3′ end of the gene was identified but not measured.

Partial DNA Sequence Analysis of the Bovine Prothrombin Gene. To prove that we had isolated the bovine prothrombin gene and not a highly homologous gene or pseudogene, pBGII21 (a subclone of λBII2, see Table I) was digested with *Sau*3AI, and the resulting DNA fragments were cloned into M13mp7. Exon-containing recombinants were detected by plaque hybridization. One *Sau*3AI fragment was isolated in both orientations and had the DNA sequence shown in Figure 3. Nucleotides 151–179 were identical with nucleotides 568–596 of the prothrombin cDNA sequence (MacGillivray & Davie, 1984) and coded for amino acid residues 145–154 of plasma prothrombin (Magnusson et al., 1975). The DNA sequence prior to nucleotide 151 (Figure 3) diverges from the cDNA sequence. A consensus splice acceptor sequence (Cech, 1983) occurs at the divergence point, suggesting that there is an intron/exon junction at amino acid residue 145 in bovine prothrombin. An intron has been reported in the corresponding position of the human prothrombin gene (Degen et al., 1983).

Discussion

We have isolated and characterized five overlapping clones that together encode 42.4 kbp of contiguous bovine genomic

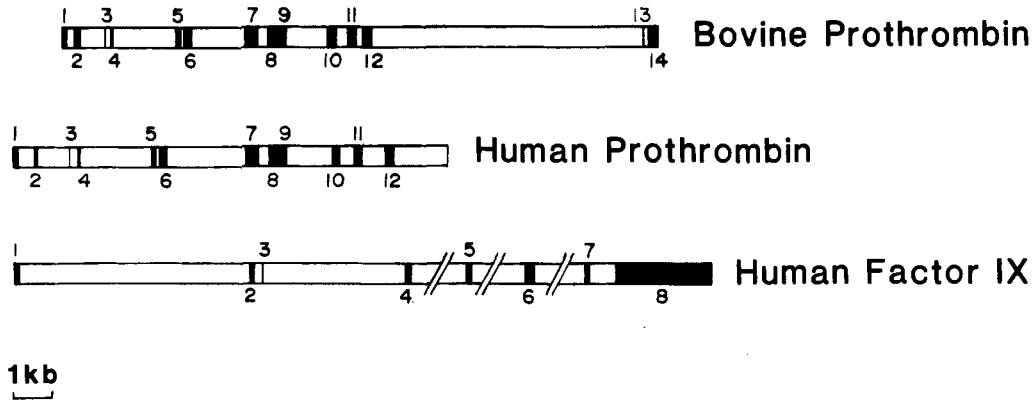

FIGURE 4: Comparison of the structures of the genes coding for bovine prothrombin, human prothrombin (Degen et al., 1983; Davie et al., 1983), and human factor IX (Anson et al., 1984). Exons are shown as black boxes and are numbered from the 5′ ends of the genes. Introns are shown as open boxes. The 3′ end of the human prothrombin gene has not been characterized. Introns D–F of the human factor IX gene have been abbreviated and are 7.5, 2.6, and 10 kbp in size, respectively. The bar at the bottom represents 1 kbp.

A

```
                10        20        30        40        50        60        70        80        90
Bovine: GATCCCTTTCACAGGCAAGCCGAATGCGGCTCAGAGAGGTTAAGTGACTTGACATCGTACAGGTCAAAGGTCAGGGGGCAGAGCTTCCCA
          •   .   *  **  **  •     *     •         * *     ***       * *    * • * ****       *  *
Human:  TTTAAAGGCAAACGGTCAGAAGCCCAGAGAGGTTAAGTAACCTGAGGTCACACAGGCAGAAAGCAGCAAGACCGGGGTTCACACCCCTGT
     1805 1810      1820      1830      1840      1850      1860      1870      1880      1890

                100       110       120       130       140       150       160       170       179
Bovine: TCCTGTCTGGCTCCACTGTGGTCCAACGCTCTCTGCCCCCTCTCTCTCCTCACCCACCAGGCCAGGACCGAGTCACAGTGGAGGTGATC
          • * * *       *   **          **  **  • ********************* * ****** ** • • ***
Human:  CTGTTCCGGTCCATGTGTGGTCTCACTCACTCTGCTGCCTCCTTGCCCCTCACCCACCAGGCCAGGATCAAGTCACTGTAGCGATGACT
          1900      1910      1920      1930      1940      1950      1960      1970      1980
```

B

```
Bovine: TTCACAGGCAAGCCGAATGCGGCTCAGAGAGGTTAAGTGACTTGACATCGTACAGGTCAAAGGTCAGGGGGCAGAGCTTCCCAT-CCTGT
        ** • ****** * *   •  ** ************** ** ***  **  *****   ** *      * * • • *** **  *****
Human:  TTTAAAGGCAAACGGTCAGAAGCCCAGAGAGGTTAAGTAACCTGAGGTCACACAGGCAGAAAGCAGCAAGACCGGGGTTCACACCCCTGT
              10        20        30        40        50        60        70        80        90

Bovine: CT----GGCTCCAC-TGTGGTCCAACGCTCTCTGCCCCCTCTCT-CTCCTCACCCACCAGGCCAGGACCGAGTCACAGTGGAGGTGAT
        **    • ****  *******  ** * ******  ****  * • ********************** • ******  ** • * ***
Human:  CTGTTCCGGTCCATGTGTGGTCTCACTCACTCTGCTGCCTCCTTGCCCCTCACCCACCAGGCCAGGATCAAGTCACTGTAGCGATGAC
              100       110       120       130       140       150       160       170     179
```

FIGURE 5: Comparison of the DNA sequences of part of the bovine and human prothrombin genes. The bovine sequence is taken from Figure 3. The human sequence represents nucleotides 1805–1983 of λ10 (Degen et al., 1983). Identical nucleotides in corresponding positions are denoted by asterisks between the sequences. (A) Comparison of the two sequences. (B) Comparison of the two sequences after the addition of seven gaps (denoted by hyphens) in the bovine sequence to maximize the homology. Arrows indicate putative splice sites.

DNA. Under the conditions used for hybridization, no other phage were identified, suggesting that there is a single gene for prothrombin in the bovine genome. Southern blot analysis of bovine genomic DNA using the complete prothrombin cDNA as a hybridization probe resulted in the detection of two to five fragments depending on the restriction enzyme used to cleave the genomic DNA. Hybridization probes from the 5′ or 3′ ends of the cDNA detected only one to two fragments, again suggesting that there is a single gene for prothrombin. All of the genomic DNA fragments detected in the genomic blot were accommodated in the restriction enzyme map obtained from the cloned phage DNAs, suggesting that no gross rearrangements of the gene had occurred during the cloning procedures.

Heteroduplex analysis of the cloned genomic DNA and cloned cDNA shows that the prothrombin gene is approximately 15 kbp in size and contains 14 exons separated by 13 introns. Because our cDNA clones contain only 5 bp of 5′-untranslated sequence, we cannot exclude the presence of further intron(s) in the 5′-untranslated region of the gene. Thus, the size of the gene and the number of exons are minimal estimates. It is also possible that other small introns were not detected by the heteroduplex analysis because of the difficulty in detecting introns smaller than 100 bp. Since there are few introns less than 80 bp in size (Wieringa et al., 1984), however, it is probable that there are no other introns in the prothrombin gene. From the Northern blot analysis, the 5′-untranslated region of prothrombin mRNA is probably about 50 nucleotides in length. As we have cloned 14 kbp of DNA upstream of the 5′-most exon of the gene and as the largest intron detected in the rest of the gene is 6.9 kbp, it is probable that λBII2 contains the 5′ end of the gene plus any 5′-regulatory sequences. The cDNA clones pBII111 and pBII102 together contain the complete coding region and the complete 3′-untranslated region. The genomic Southern blot using a 3′ probe from the cDNA shows that the 3′ end of the gene is located in a 2 kbp *Eco*RI fragment (data not shown). In that case, we have isolated 13 kbp of 3′-flanking sequence, including possible 3′-regulatory sequences. However, the identities of any 5′- or 3′-regulatory sequences are unknown at present.

Although the prothrombin gene is 15 kbp in size, the cytoplasmic mRNA is only 2150 ± 100 nucleotides in size, including a poly(A) tail. Thus, 86% of the prothrombin gene is comprised of introns. The average size of exons in the prothrombin gene is 143 bp which is similar to the average length of exons found in other eukaryotic genes (Blake, 1983).

For those regions of the bovine and human prothrombin genes that have been characterized, the numbers and sizes of the exons are very similar (Figure 4). This similarity is supported by the localization of an intron in the same position (corresponding to amino acid residue 145) in the genes of both species. This intron separates exons 6 and 7 in the human gene [see Figure 4 and Degen et al. (1983)]. Exons 13 and 14 at the 3′ end of the bovine gene are separated from the rest of the exons by a large intron that is 6.9 kbp in size. The presence of a large intron in the corresponding position of the human gene could explain why no phage containing the 3′ end of the human gene have been reported to date. There are no obvious changes in the sizes of any of the exons in the two genes.

Introns E, G, and H appear to be similar in size in both the human and bovine genes, whereas the other introns appear to be significantly larger in the human gene (Figure 4). Repetitive sequences of the *Alu*I type have been identified within the larger introns of the human gene (Degen et al., 1983). Repetitive elements of an unknown type were also found within at least some of the larger introns of the bovine gene (Figure 2). A repetitive element similar to the human *Alu* element has been isolated from the bovine genome and is 120 nucleotides in length (Watanabe et al., 1982). This is much shorter than the size of the human *Alu* repeat (300 bp; Jelinek & Schmid, 1982) and may explain in part the difference in length of the introns between the two species. An alternative explanation may be either the deletion of DNA from the bovine introns or the insertion of DNA into the human introns. Evidence in support of this is shown in Figure 5 where several gaps must be inserted into the bovine intron to maximize the homology with the corresponding human intron (see following paragraph).

Partial DNA sequence analysis of λBII2 confirms that we have cloned the bovine prothrombin gene. Comparison with the cDNA shows that an intron interrupts the cDNA at the codon for amino acid residue 145 (Figure 3). A comparison of this region with the corresponding region of the human gene (Degen et al., 1983) is shown in Figure 5A. The exon regions (nucleotides 151–179 of the bovine sequence) show 72% sequence identity (21/29 nucleotides in corresponding positions are identical). The homology extends into the intron and is similar to the consensus sequence derived from the sequences of other introns (Cech, 1983). Further upstream into the intron, the human and bovine sequences diverge to only 27% identity (37/137 nucleotides are identical), but this identity can be increased to 66% by the addition of seven gaps into the bovine sequence (Figure 5B). Thus, for the short region of the prothrombin genes shown in Figure 5, the intron and exon sequences have each mutated to approximately the same extent.

Heteroduplex analysis does not allow us the accuracy to compare the exact locations of the other intron/exon junctions within the gene. However, exons 1–3 of the human prothrombin and factor IX genes are in analogous positions (Davie et al., 1983; Anson et al., 1984; see Figure 4). These exons code for the homologous leader peptides and Gla regions. The amino acid sequences of prothrombin and factor IX diverge after the Gla region, and this divergence is reflected in the sizes of the exons from exon 4 onward (Figure 4). Prothrombin and factor IX also share amino acid sequence homology in the catalytic regions (at the carboxy-terminal ends of the polypeptide chains). It is not obvious from the gross structure of the prothrombin genes whether the catalytic regions have introns in homologous positions to those in the factor IX gene. However, the positions of some introns in the serine protease genes have been conserved [see Campbell et al. (1984), Anson et al. (1984), and Swift et al. (1984)]. We are currently determining the exact locations of the introns within the bovine prothrombin gene. This should allow a more detailed analysis of the evolution of prothrombin within the serine protease gene family.

ACKNOWLEDGMENTS

We thank Dr. F. Rottman and D. Ayers for providing a sample of their bovine genomic phage library, Drs. Mark Zoller and Jim Stone for helpful suggestions, and our colleagues in the laboratory for critically reading the manuscript.

Registry No. Prothrombin, 9001-26-7.

REFERENCES

Anson, D. S., Choo, K. H., Rees, D. J. S., Giannelli, F., Gould, K., Huddleston, J. A., & Brownlee, G. G. (1984) *EMBO J. 3*, 1053–1060.

Benton, W. D., & Davis, R. W. (1977) *Science (Washington, D.C.) 196*, 180–182.

Blake, C. (1983) *Nature (London) 306*, 535–537.

Blin, N., & Stafford, D. W. (1976) *Nucleic Acids Res. 3*, 2303–2308.

Butkowski, R. J., Elion, J., Downing, M. R., & Mann, K. G. (1977) *J. Biol. Chem. 252*, 4942–4957.

Campbell, R. D., Bentley, D. R., & Morley, B. J. (1984) *Philos. Trans. R. Soc. London, B 306*, 367–378.

Cech, T. R. (1983) *Cell (Cambridge, Mass.) 34*, 713–716.

Chow, L. T., & Broker, T. R. (1981) in *Electron Microscopy in Biology* (Griffith, J. D., Ed.) Vol. 1, pp 139–188, Wiley, New York.

Cool, D. E., Edgell, C.-J. S., Louie, G. V., Zoller, M. J., Brayer, G. D., & MacGillivray, R. T. A. (1985) *J. Biol. Chem.* (in press).

Davie, E. W., Degen, S. J. F., Yoshitake, S., & Kurachi, K. (1983) *Dev. Biochem. 25*, 45–52.

Degen, S. J. F., MacGillivray, R. T. A., & Davie, E. W. (1983) *Biochemistry 22*, 2087–2097.

Delaney, A. D. (1982) *Nucleic Acids Res. 10*, 61–67.

Delbaere, L. T. J., Hutcheon, W. L. B., James, M. N. G., & Thiessen, W. E. (1975) *Nature (London) 257*, 758–763.

Doolittle, R. F., Oncley, J. L., & Surgenor, D. M. (1962) *J. Biol. Chem. 237*, 3123–3127.

Gilbert, W. (1979) in *Eukaryotic Gene Regulation* (Axel, R., Maniatis, T., & Fox, C. F., Eds.) pp 1–12, Academic Press, New York.

Greer, J. (1981) *J. Mol. Biol. 153*, 1027–1042.

Hartley, B. S. (1970) *Philos. Trans. R. Soc. London, B 257*, 77–87.

Jackson, C. M., & Nemerson, Y. (1980) *Annu. Rev. Biochem. 49*, 765–811.

Jelinek, W. R., & Schmid, C. W. (1982) *Annu. Rev. Biochem. 51*, 813–844.

Johnson, P., & Smillie, L. B. (1974) *FEBS Lett. 47*, 1–6.

Jurasek, L., Carpenter, M. R., Smillie, L. B., Gertler, A., Levy, S., & Ericsson, L. H. (1974) *Biochem. Biophys. Res. Commun. 61*, 1095–1100.

Kan, Y. W., & Dozy, A. M. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5631–5635.

Karn, J., Brenner, S., Barnett, L., & Cesareni, G. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5172–5176.
Lunberg, N., & Gilbert, W. (1985) *Cell (Cambridge, Mass.) 40*, 81–91.
MacGillivray, R. T. A., & Davie, E. W. (1984) *Biochemistry 23*, 1626–1634.
Magnusson, S., Peterson, T. E., Sottrup-Jensen, L., & Claeys, H. (1975) in *Proteases and Biological Control* (Reich, E., Rifkin, D. B., & Shaw, E., Eds.) pp 123–149, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
Maniatis, T., Jeffrey, A., & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1184–1188.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
McMullen, B., & Fujikawa, K. (1984) *Circulation 70*, II–351.
Messing, J. (1983) *Methods Enzymol. 101*, 20–78.
Neurath, H. (1984) *Science (Washington, D.C.) 224*, 350–357.
Neurath, H., & Walsh, K. A. (1976) in *Proteolysis and Physiological Regulation* (Ribbons, D. W., & Brew, K., Eds.) pp 29–40, Academic Press, New York.
Olafson, R. W., Jurasek, L., Carpenter, M. R., & Smillie, L. B. (1975) *Biochemistry 14*, 1168–1177.
Olson, M. O. J., Nagabhushan, N., Dzwiniel, M., Smillie, L. B., & Whitaker, D. R. (1970) *Nature (London) 228*, 438–442.
Owen, W. G., Esmon, C. T., & Jackson, C. M. (1974) *J. Biol. Chem. 249*, 594–605.
Pennica, D., Holmes, W. F., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennet, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V., & Collen, D. (1983) *Nature (London) 301*, 214–221.
Perry, R. P. (1976) *Annu. Rev. Biochem. 45*, 605–629.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Sottrup-Jensen, L., Claeys, H., Zajdel, M., Petersen, T. E., & Magnusson, S. (1978) in *Progress in Chemical Fibrinolysis and Thrombolysis* (Davidson, J. F., Rowan, R. M., Samana, M. M., & Desnoyers, P. C., Eds.) Vol. 3, pp 191–209, Raven Press, New York.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
Stenflo, J., & Suttie, J. W. (1977) *Annu. Rev. Biochem. 46*, 157–172.
Sternberg, N., Tiemeier, D., & Enquist, L. (1977) *Gene 1*, 255–280.
Swift, G. H., Craik, C. S., Stary, S. J., Quinto, C., Lahaie, R. G., Rutter, W. J., & McDonald, R. J. (1984) *J. Biol. Chem. 259*, 14271–14278.
Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5201–5205.
Verde, P., Stoppelli, M. P., Galeffi, P., Di Nocera, P., & Blasi, F. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4727–4731.
Vieira, J., & Messing, J. (1982) *Gene 19*, 259–268.
Walz, D. A., Hewett-Emmett, D., & Seegers, W. H. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 1969–1972.
Watanabe, Y., Tsukada, T., Notake, M., Nakanishi, S., & Numa, S. (1982) *Nucleic Acids Res. 10*, 1459–1469.
Wieringa, N., Hofer, E., & Weissmann, C. (1984) *Cell (Cambridge, Mass.) 37*, 915–925.
Woo, S. L. C. (1980) *Methods Enzymol. 68*, 389–395.

Stimulation of Deoxyribonucleic Acid Excision Repair in Human Fibroblasts Pretreated with Sodium Butyrate†

Steven L. Dresler
Department of Pathology, Washington University School of Medicine, St. Louis, Missouri 63110
Received April 3, 1985

ABSTRACT: The effect of pretreatment with sodium butyrate on DNA excision repair was studied in intact and permeable confluent (i.e., growth-inhibited) diploid human fibroblasts. Exposure to 20 mM sodium butyrate for 48 h increased subsequent ultraviolet (UV)-induced [*methyl*-^{3}H]thymidine incorporation by intact AG1518 fibroblasts by 1.8-fold and by intact IMR-90 fibroblasts by 1.2–1.3-fold. UV-induced incorporation of deoxy[5-^{3}H]cytidine, deoxy[6-^{3}H]cytidine, and deoxy[6-^{3}H]uridine, however, showed lesser degrees of either stimulation or inhibition in butyrate-pretreated cells. This result suggested that measurements of butyrate's effect on DNA repair synthesis in intact cells are confounded by simultaneous changes in nucleotide metabolism. The effect of butyrate on excision repair was also studied in permeable human fibroblasts in which excision repair is dependent on exogenous nucleotides. Butyrate pretreatment stimulated UV-induced repair synthesis by 1.3–1.7-fold in permeable AG1518 cells and by 1.5–2-fold in permeable IMR-90 cells. This stimulation of repair synthesis was not due to changes in repair patch size or composition or in the efficiency of DNA damage production but rather resulted from a butyrate-induced increase in the rate of damage-specific incision of DNA. The increased rate of incision in butyrate-pretreated cells could be due either to increased levels of enzymes mediating steps in excision repair at or before incision or to alterations in chromatin structure making damage sites in DNA more accessible to repair enzymes.

Chromatin structure has a significant influence on mammalian DNA excision repair, as it does on other cellular processes which involve nuclear DNA [see, for instance, DePamphilis & Wassarman (1980) and Mathis et al. (1980)]. Transient changes in chromatin structure at the nucleosome level, which render newly synthesized repair patches highly susceptible to digestion by exogenous nucleases, are a consistent feature of excision repair of DNA damage in mammalian cells (Smerdon & Lieberman, 1978, 1980; Tlsty & Lieberman,

† Supported by Grant IN-36 from the American Cancer Society, by Grant CA 20513 from the National Institutes of Health, and by Brown and Williamson Tobacco Corp., Phillip Morris Inc., and the U.S. Tobacco Co.

0006-2960/85/0424-6861$01.50/0 © 1985 American Chemical Society