Prior studies have established that the expression of the human apolipoprotein B (apoB) gene in the intestine is dependent on DNA sequences located a great distance from the structural gene. To identify the location of those sequences, we used recA-assisted restriction endonuclease (RARE) cleavage to truncate the 5'- or 3'-flanking sequences from a 145-kilobase (kb) bacterial artificial chromosome spanning the entire human apoB gene. Seven RARE cleavage-modified bacterial artificial chromosomes with different lengths of flanking sequences were used to generate transgenic mice. An analysis of those mice revealed that as little as 1.5 kb of 3' sequences or 5 kb of 5' sequences were sufficient to confer apoB expression in the liver. In contrast, apoB gene expression in the intestine required DNA sequences 54–62 kb 5' to the structural gene. Those sequences retained their ability to direct apoB expression in the intestine when they were moved closer to the gene. These studies demonstrate that the intestinal expression of the apoB gene is dependent on DNA sequences located an extraordinary distance from the structural gene and that the RARE cleavage/transgenic expression strategy is a powerful approach for analyzing distant gene-regulatory sequences.

The B apolipoproteins, apoB48\(^1\) and apoB100, play important roles in the formation of the triglyceride-rich lipoproteins (1, 2). ApoB48 is essential for the assembly of chylomicrons in the intestine, and apoB100 is essential for the formation of VLDL in the liver. When apoB gene mutations prevent the synthesis of apoB48 and apoB100, neither chylomicrons nor VLDL can be detected in the plasma, and the plasma concentrations of triglycerides and cholesterol are extremely low (3, 4).

Although the basic function of apoB (the assembly of triglyceride-rich lipoproteins) is the same in the liver and intestine, the genetic control of apoB gene expression in these two tissues is strikingly different. In transgenic mouse expression studies, an ~80-kb P1 bacteriophage (p158) spanning the entire human apoB gene (and containing 19 kb of 5' sequences and 17.5 kb of 3' sequences) yielded human apoB expression in the liver of transgenic mice, but expression was completely absent in the intestine (5, 6). Similarly, a P1 bacteriophage clone spanning the mouse apoB gene (and containing 33 kb of 5'-sequences and 11 kb of 3' sequences) did not direct transgene expression in the intestine of transgenic mice (7). More recently, we identified a 145-kb bacterial artificial chromosome (BAC) clone spanning the human apoB gene and used it to generate human apoB transgenic mice. That clone, which contained 70 kb of 5' sequences and 22 kb of 3' sequences, conferred completely appropriate levels of apoB gene expression in the absorptive enterocytes of the intestine (8).

In the current study, we have used transgenic mouse expression studies with BACs to further define the sequences that are important for the expression of the apoB gene in the intestine. Our approach involved the use of recA-assisted restriction endonuclease (RARE) cleavage to delete portions of the 5'- or 3'-flanking sequences from the 145-kb BAC. The RARE cleavage-modified BACs were then used to generate multiple lines of human apoB transgenic mice. This experimental approach allowed us to document that the intestinal expression of the apoB gene is controlled by DNA sequences located an extraordinary distance upstream from the human apoB gene.

**MATERIALS AND METHODS**

*Generating Modified BACs by RARE Cleavage—All of the BACs used in this study were derived from a 145-kb BAC, BAC(70,22), spanning the human apoB gene. BAC(70,22) contained 70 kb of 5'-flanking sequences and 22 kb of 3'-flanking sequences. To delete portions of the 5'- and 3'-flanking sequences from BAC(70,22), we deleted segments of DNA between two EcoRI sites or two HindIII sites. BAC(70,22) contains...*
more than 40 EcoRI and more than 40 HindIII sites. To cleave specifically at only two of the EcoRI sites or at only two of the HindIII sites, we used RARE cleavage (9–15). In the presence of the bacterial protein recE, oligonucleotides can be used to “cover up” restriction sites in a BAC clone and thereby protect them from methylation by a specific DNA methylase. In the absence of recE, the inactivation was complete on both restriction sites of interest was obtained by sequencing back across the restriction sites using P1–70 as template. For these sequencing reactions, we used the oligonucleotide primers 1 (5′-GGG GAA GGA GGG GAG GAT GGT GG-3′ (to define the sequence surrounding the EcoRI site located –5 kb 5′ to the gene)), 2 (5′-TCT AGT AAA AAT CTA TGA CCA ATG-3′ (to define the sequence surrounding the HindIII site located –5 kb 5′ to the gene)), 3 (5′-AGC CCC AGG TAG ACC CCA TTC TTG TC-3′ (to define the EcoRI site located –62 kb 5′ to the gene)). Oligonucleotide primers 5 (5′-TCT AGA ATT GTC AGG GAA TAA C-3′) and 6 (5′-GCT GGA AGG GAG ATG TGC TG-3′) (corresponding to BAC vector sequences) were used to sequence across the HindIII cloning site at the 3′ and 5′ ends of BAC(70,22), respectively. All sequencing was done with the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer) according to the manufacturer’s protocol, except for sequencing of BAC or P1 DNA, in which we used 2 or 4 μg of DNA in each sequencing reaction.

To clone both BAC(30,22) and BAC(54,22), we used a single 60-mer oligonucleotide (recE) to cleave the HindIII site in the BAC vector located 49 bp from the 5′ EcoRI site and an 18-mer oligonucleotide consisting of sequences immediately 5′ to the HindIII cloning site in BAC(70,22), while the last 30 bp consisted of sequences immediately 3′ to a HindIII site located –30 or –54 kb 5′ to the apoB gene. To generate the remainder of the BACs, we used pairs of different 60-mer oligonucleotides, each of which spanned specific HindIII or EcoRI sites. BAC(54,22) was also cloned with a two-oligonucleotide approach. Oligonucleotides were purchased from Life Technologies or Oligos Etc. (Wilsonville, OR). The oligonucleotides used to generate the BACs used in this study were as follows: BAC(70,1.5), 7 (5′-ATC CTT CAC CCA CCA CCA TTC TTG CAT AAC GCT TCA CAT TCT CTT GGT GAA GCA GCC CAC TGT-3′ (covering a HindIII site +1.5 kb to the apoB gene)) and 8 (5′-CTA TTA AGG TCA CAG TAT ATG ATG TCC AAG ATG TTG GCT TTC TTG CTA AGA GAT TAT TTG ATG-3′ (covering the HindIII site at the 5′ cloning site within BAC(70,22)); BAC(30,22), 9 (5′-TCT TCC AGA GTC GAG CAC GAG TCG CAG GAC TGC AAG CTA TAT ATG TAC AAT TTG AGG GAG ATG TCA GAG GAA ATC TGA CCA CCA CCA TTC TTG CAT AAC GCT TCA CAT TCT CTT GGT GAA GCA GCC CAC TGT-3′ (covering both the 5′ HindIII cloning site and a HindIII site located –30 kb 5′ to the apoB gene)); BAC(54,22) (using the single oligonucleotide approach), 10 (5′-TCT TCC AGA GTC GAG CAC GAG TCG CAG GAC TGC AAG CTA TAT ATG TAC AAT TTG AGG GAG ATG TCA GAG GAA ATC TGA CCA CCA CCA TTC TTG CAT AAC GCT TCA CAT TCT CTT GGT GAA GCA GCC CAC TGT-3′ (covering both the 5′ HindIII cloning site and HindIII site located –30 kb 5′ to the apoB gene)); BAC(54,22) (using the two-oligonucleotide approach), 11 (5′-GAT CAA TGA TGA AGG ATG TGA GAA AAG CAG CTT CCT CCA CCA ACC AAA AAC AAT TTA-3′ (covering a HindIII site –54 kb 5′ to the apoB gene)) and 12 (5′-TCT TCC AGA GTC GAG CAC GAG TCG CAG GTC AAG CTA TAC CTC TCC TTT GCT GTC GGT GCT-3′ (covering the HindIII site at the 5′ cloning site within BAC(70,22)); BAC(30,22), 9 (5′-TCT TCC AGA GTC GAG CAC GAG TCG CAG GAC TGC AAG CTA TAT ATG TAC AAT TTG AGG GAG ATG TCA GAG GAA ATC TGA CCA CCA CCA TTC TTG CAT AAC GCT TCA CAT TCT CTT GGT GAA GCA GCC CAC TGT-3′ (covering both the 5′ HindIII cloning site and a HindIII site located –30 kb 5′ to the apoB gene)) and 14 (5′-AGT GAA TTT TAA TAC GAC TCA TGT AGG GGC TAA GTC GAC TTC GGT GCA AGG CAG CAA CCA TCC ATG CTT GGT GAA GCA GCC CAC TGT-3′ (covering both the 5′ HindIII cloning site and a HindIII site located –30 kb 5′ to the apoB gene)); BAC(54,22) (using the single oligonucleotide approach), 10 (5′-TCT TCC AGA GTC GAG CAC GAG TCG CAG GAC TGC AAG CTA TAT ATG TAC AAT TTG AGG GAG ATG TCA GAG GAA ATC TGA CCA CCA CCA TTC TTG CAT AAC GCT TCA CAT TCT CTT GGT GAA GCA GCC CAC TGT-3′ (covering both the 5′ HindIII cloning site and HindIII site located –30 kb 5′ to the apoB gene)); BAC(54,22) (using the two-oligonucleotide approach), 11 (5′-GAT CAA TGA TGA AGG ATG TGA GAA AAG CAG CTT CCT CCA CCA ACC AAA AAC AAT TTA-3′ (covering a HindIII site –54 kb 5′ to the apoB gene)) and 12 (5′-TCT TCC AGA GTC GAG CAC GAG TCG CAG GTC AAG CTA TAC CTC TCC TTT GCT GTC GGT GCT-3′ (covering the HindIII site at the 5′ cloning site within BAC(70,22)); BAC(30,22), 9 (5′-TCT TCC AGA GTC GAG CAC GAG TCG CAG GAC TGC AAG CTA TAT ATG TAC AAT TTG AGG GAG ATG TCA GAG GAA ATC TGA CCA CCA CCA TTC TTG CAT AAC GCT TCA CAT TCT CTT GGT GAA GCA GCC CAC TGT-3′ (covering both the 5′ HindIII cloning site and HindIII site located –30 kb 5′ to the apoB gene)) and 14 (5′-AGT GAA TTT TAA TAC GAC TCA TGT AGG GGC TAA GTC GAC TTC GGT GCA AGG CAG CAA CCA TCC ATG CTT GGT GAA GCA GCC CAC TGT-3′ (covering both the 5′ HindIII cloning site and a HindIII site located –30 kb 5′ to the apoB gene)).
Transgenic founders were identified by a human apoB-specific radioimmunoassay (5) or enzyme-linked immunosorbent assay. To determine transgene copy numbers, 5 or 10 μg of genomic DNA from human leukocytes or tails of mice was digested with EcoRI or HindIII. The digested DNA was separated on 1% agarose gels and blotted onto a nylon membrane. Southern blot analysis was performed with a 32P-labeled 1857-bp BamHI–EcoRI fragment from exon 26 of the human apoB gene (apoB cDNA nucleotides 4650–5607) using Quickhyde (Stratagene). Transgene copy number was assessed by comparing the intensity of the human apoB signal in transgenic mouse genomic DNA with that in human genomic DNA with a phosphor imager (BAS1000, Fuji). In some experiments, transgene copy number was determined with slot-blot analysis of genomic DNA.

**PCR Analysis of the Intactness of BAC Transgenes**—To determine whether any of the transgenic animals had incorporated only a portion of the BAC transgene, we used PCR to enzymatically amplify BAC vector sequences located at the far 5′ or far 3′ ends of the transgene (2). In these experiments, a 272-bp fragment of the BAC vector at the 5′ end of the transgene and a 322-bp fragment at the 3′ end of the transgene were amplified using oligonucleotide primers that have been described previously (8). PCR products were resolved on 1% agarose gels and visualized by ethidium bromide staining.

**Analysis of ApoB Gene Expression by RNase Protection Assays**—Total cellular RNA was isolated from mouse liver and duodenum with the Totally RNA kit (Ambion, Austin, TX). To analyze apoB expression, RNase protection assays were performed with the RPA II ribonuclease protection assay kit (Ambion) according to the manufacturer’s instructions. Two human apoB antisense riboprobes were transcribed from linearized plasmids with T7 RNA polymerase (Boehringer Mannheim); one probe spanned the first 121 nucleotides of exon 1 of the apoB gene, and the other spanned 220 bp of exon 26 (DNA nucleotides 4457–5607) (8). A 316-bp mouse glyceraldehyde-3-phosphate dehydrogenase (GAPDH) probe was used as an internal control (Ambion, catalogue no. 7431). The riboprobes were mixed with 5–25 μg of total RNA from liver and intestine. Yeast RNA (25–40 μg) was used as a negative control. After the digestion of the RNA/riboprobe mixture, the samples were electrophoresed on 6% polyacrylamide gels containing either 7 or 8.3 M urea (Novex, San Diego, CA). Dried gels were exposed to autoradiographic film to visualize protected RNA fragments.

**RESULTS**

**Generation of RARE Cleavage–modified BACs**—We recently isolated a 145-kb BAC spanning the human apoB gene (designated BAC(70,22)) because it contains 70 kb of sequences and 22 kb of 3′ ends of the transgene (2). In these experiments, a 272-bp fragment of the BAC vector at the 5′ end of the transgene and a 322-bp fragment at the 3′ end of the transgene were amplified using oligonucleotide primers that have been described previously (8). PCR products were resolved on 1% agarose gels and visualized by ethidium bromide staining.

**Localization of DNA Sequences That Control apoB Gene Expression**—To determine whether the 3′-flanking sequences played a role in controlling apoB gene expression, we analyzed apoB gene expression in transgenic mice generated with BAC(70,1.5), BAC(70,1.5) differed from the parental BAC, BAC(70,22), in that it lacked the sequences 1.5–22 kb 3′ to the gene. In each of eight independent transgenic mouse lines (four F1s and four founders), the apoB gene was expressed at high levels in both the liver and intestine (Fig. 5A). The level of human apoB expression in the intestines of the BAC(70,15) mice appeared to be appropriate. As judged by phosphor imaging analysis, the relative human apoB expression levels in the liver and intestine paralleled those of the endogenous mouse apoB gene (data not shown). These data indicated that the sequences located between +1.5 and +22 kb are not important for either intestinal or liver expression of the apoB gene.

To define the 5′-flanking DNA sequences that control apoB gene expression, we initially generated mice from BACs that

---

2 In prior experiments, we generated transgenic mice using BAC(120,35) and BAC(70,22) (8). In those experiments, we used two different approaches to analyze whether the BAC transgenes had been incorporated intact into the chromosomal DNA. First, we demonstrated by PCR that BAC vector sequences at either end of the transgene were present in the genomic DNA. Second, we used pulsed field gel electrophoresis of NotI-digested genomic DNA (obtained from hepatocyte plugs), followed by Southern blotting, to directly identify the transgene and verify that it was intact within the genomic DNA. We have also used both approaches to show the “intactness” of an 86-kb P1 bacteriophage transgene and of YAC transgenes (7, 49). Because our Southern blotting studies have invariably supported the results obtained by PCR, we performed only the PCR analyses in this study. We recognize that these PCR studies cannot provide definitive evidence for transgene intactness (50). However, in view of our prior experience, the PCR studies suggest that the BAC transgenes were incorporated intact into the genomic DNA.
The amounts of 5' and 3' flanking sequences in the BACs are indicated by their lengths (e.g., BAC(70,22) contains 70 kb 5' and 22 kb 3' to the apoB gene). Each BAC contains two NotI sites that flank the insert, so a NotI digest cuts out the 7-kb BAC vector fragment. Each BAC also contains a NotI site in intron 1 of the apoB gene, 221 bp 3' to the transcription start site. Thus, digestion of the BACs results in generation of a 7-kb BAC vector fragment and two additional DNA fragments: one fragment representing the apoB 5' region and the other representing the 43 kb of the structural gene (located downstream to the intron 1 NotI site) and the 3' flanking sequences.

Fig. 2. A, map of the 5' region of the apoB gene showing the positions of EcoRI or HindIII sites used in RARE cleavage reactions. The positions of these sites were mapped previously by quantitative DNA fiber mapping (19, 45). The colored fiber shows the results of multicolor fluorescence in situ hybridization (FISH) of a single DNA molecule (P1-70) that spans from intron 1 of the apoB gene to 70 kb upstream from the gene. The P1–70 DNA fiber was immobilized on glass and visualized (in green) by FISH, using the entire P1–70 DNA as a probe (the probe was labeled with fluorescein). Two HindIII subclones of P1–70 were labeled with digoxigenin (blue) and biotin (red), respectively. By FISH, these subclones were positioned 36 and 20 kb 5' to the 3' end of the DNA fiber. In some cases, the orientation of the DNA fiber was also determined with a digoxigenin-labeled probe spanning a 2.1-kb portion of the P1 bacteriophage vector at the 3' end of the DNA fiber. B and C, ethidium bromide– stained pulsed field gels showing NotI digests of modified BACs generated by RARE cleavage reactions. Each of the BACs contains the entire apoB structural gene, but they differ in the lengths of 5' and 3' flanking sequences. The amounts of 5' and 3' flanking sequences in the BACs are indicated by their lengths (e.g., BAC(70,22) contains 70 kb 5' and 22 kb 3' to the apoB gene). Each BAC contains two NotI sites that flank the insert, so a NotI digest cuts out the 7-kb BAC vector fragment. Each BAC also contains a NotI site in intron 1 of the apoB gene, 221 bp 3' to the transcription start site. Thus, digestion of the BACs results in generation of a 7-kb BAC vector fragment and two additional DNA fragments: one fragment representing the apoB 5' region and the other representing the 43 kb of the structural gene (located downstream to the intron 1 NotI site) and the 3' flanking sequences.
segment of the upstream sequences had been deleted. This construct, BAC(Δ5–47,1.5), was generated by deleting sequences between 5 and 47 kb upstream of the apoB gene in BAC(62,1.5). BAC(Δ5–47,1.5) yielded robust apoB gene expression in the intestine (two founders examined) (Fig. 7). These studies revealed that the sequences between 25 and 47 kb do not play an essential role in intestinal apoB gene expression and that the distant intestinal regulatory sequences function normally when moved closer to the structural gene.

**DISCUSSION**

Elevated plasma levels of the apoB-containing lipoproteins play a central role in the pathogenesis of atherosclerotic disease (20, 21). The relationship of apoB to atherosclerotic disease has led to intensive study of the apoB gene, including studies of the regulatory sequences that control apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression. Early attempts to define the apoB gene's regulatory sequences utilized the expression of reporter gene constructs in cultured cells (22–24). Those studies strongly suggested that the proximal promoter sequences located immediately upstream from the apoB gene were sufficient for apoB gene expression.

There are precedents for gene regulatory elements, or locus control regions (LCRs), being located at a distance from the structural gene (30–34). For example, liver expression of four genes at the apoE/CI/CIV/CI gene cluster is controlled by the same hepatic control region (15, 35–38), and expression of the ε-, γ-, and β-globin genes in the β-globin locus is regulated by a 5′-LCR located 5–22 kb upstream from the ε-globin gene (34, 39–43). Many of the previously characterized LCRs have occurred in the setting of multigene families that arose by ancient gene duplication events, and it is not difficult to imagine how these duplication events could result in a regulatory element being placed at a significant distance from the genes it controls. In contrast, the apoB gene is not known to have any adjacent family members, which makes the distance of the apoB gene's LCR (more than 54 kb from the structural gene) unique and extraordinary. It is important to note that the DNA sequences that control apoB gene expression in the intestine, we generated apoB transgenic mice using seven different BACs containing different lengths of 5′- and 3′-flanking sequences. Analysis of the tissue-specific expression pattern of 2–8 independently generated lines from each BAC construct was unambiguous and suggested 1) that sequences located between 54 and 62 kb upstream from the apoB gene are essential for the intestinal expression of the apoB gene, 2) that these distant sequences are effective in conferring intestinal expression when they are moved closer to the structural gene by deleting intervening sequences, and 3) that the sequences located between 1.5 and 22 kb downstream from the structural gene are not important for apoB gene expression, either in the intestine or in the liver.

There are precedents for gene regulatory elements, or locus control regions (LCRs), being located at a distance from the structural gene (30–34). For example, liver expression of four genes at the apoE/CI/CIV/CI gene cluster is controlled by the same hepatic control region (15, 35–38), and expression of the ε-, γ-, and β-globin genes in the β-globin locus is regulated by a 5′-LCR located 5–22 kb upstream from the ε-globin gene (34, 39–43). Many of the previously characterized LCRs have occurred in the setting of multigene families that arose by ancient gene duplication events, and it is not difficult to imagine how these duplication events could result in a regulatory element being placed at a significant distance from the genes it controls. In contrast, the apoB gene is not known to have any adjacent family members, which makes the distance of the apoB gene's LCR (more than 54 kb from the structural gene) unique and extraordinary. It is important to note that the DNA sequences that control apoB gene expression in the intestine, we generated apoB transgenic mice using seven different BACs containing different lengths of 5′- and 3′-flanking sequences. Analysis of the tissue-specific expression pattern of 2–8 independently generated lines from each BAC construct was unambiguous and suggested 1) that sequences located between 54 and 62 kb upstream from the apoB gene are essential for the intestinal expression of the apoB gene, 2) that these distant sequences are effective in conferring intestinal expression when they are moved closer to the structural gene by deleting intervening sequences, and 3) that the sequences located between 1.5 and 22 kb downstream from the structural gene are not important for apoB gene expression, either in the intestine or in the liver.

There are precedents for gene regulatory elements, or locus control regions (LCRs), being located at a distance from the structural gene (30–34). For example, liver expression of four genes at the apoE/CI/CIV/CI gene cluster is controlled by the same hepatic control region (15, 35–38), and expression of the ε-, γ-, and β-globin genes in the β-globin locus is regulated by a 5′-LCR located 5–22 kb upstream from the ε-globin gene (34, 39–43). Many of the previously characterized LCRs have occurred in the setting of multigene families that arose by ancient gene duplication events, and it is not difficult to imagine how these duplication events could result in a regulatory element being placed at a significant distance from the genes it controls. In contrast, the apoB gene is not known to have any adjacent family members, which makes the distance of the apoB gene's LCR (more than 54 kb from the structural gene) unique and extraordinary. It is important to note that the DNA sequences that control apoB gene expression in the intestine, we generated apoB transgenic mice using seven different BACs containing different lengths of 5′- and 3′-flanking sequences. Analysis of the tissue-specific expression pattern of 2–8 independently generated lines from each BAC construct was unambiguous and suggested 1) that sequences located between 54 and 62 kb upstream from the apoB gene are essential for the intestinal expression of the apoB gene, 2) that these distant sequences are effective in conferring intestinal expression when they are moved closer to the structural gene by deleting intervening sequences, and 3) that the sequences located between 1.5 and 22 kb downstream from the structural gene are not important for apoB gene expression, either in the intestine or in the liver.
distant regulation of apoB gene expression in the intestine is not a peculiarity of the human apoB gene. A mouse apoB gene clone containing 33 kb of 5′ sequences did not confer intestinal expression of the apoB gene, indicating that this distant form of apoB gene regulation has existed for more than 60 million years of mammalian evolution (44).

In our study, we used a new approach to localize distant gene regulatory sequences of large genes: mapping the gene’s flanking sequences with quantitative DNA fiber mapping, using RARE cleavage to modify a large BAC clone, and then analyzing gene expression patterns in a series of BAC transgenic mice. There are several attractive features of this approach. First, high redundancy BAC libraries of human and mouse genomic DNA are commercially available. Once a clone is in hand, quantitative DNA fiber mapping (45) makes it possible to define, rapidly and accurately, useful EcoRI and HindIII sites for RARE cleavage reactions (19). Truncating a BAC using RARE cleavage, followed by ligating the ends of the BAC, can be performed efficiently. In addition, our studies demonstrated that the generation of transgenic mice with BAC DNA can be performed with very high efficiency, comparable with that obtained with short fragments of plasmid DNA. We believe that our approach establishes a very attractive method for delineating the regulation of large gene loci by distant cis-acting elements.

In the case of the β-globin gene locus, mutations in the LCRs have been shown to cause human disease. There are several examples of deletions in the 5′ LCRs that abolish β-globin synthesis and cause thalassemia, despite the fact that the protein-coding sequences of the β-globin gene remained intact (46). Therefore, we would not be surprised if mutations in theapoB gene's distant intestinal LCR cause human disease. A deletion of the apoB gene's distant LCR would be expected to eliminate chylomicron formation (leading to intestinal fat mal-

Fig. 5. A, RNase protection assays with a 121-bp 32P-labeled human apoB riboprobe showing human apoB gene expression in liver (5 μg of RNA) and intestine (25 μg of RNA) of five lines of BAC(70,1.5) mice. B, RNase protection assay with 220-bp human apoB and 316-bp mouse GAPDH 32P-labeled riboprobes showing human apoB and mouse GAPDH gene expression in liver (10 μg of RNA) and intestine (10 μg of RNA) of BAC(5,22) and BAC(30,22) mice. ApoB gene expression in the BAC(5,22) and in BAC(30,22) mice was analyzed in different experiments, but under similar conditions.

Fig. 6. A, RNase protection assay with a 121-bp 32P-labeled human apoB riboprobe showing liver (5 μg of RNA) and intestinal (25 μg of RNA) expression of the human apoB gene in three lines of BAC(62,1.5) mice. B, RNase protection assay with 32P-labeled 121-bp human apoB and 281-bp mouse GAPDH riboprobes showing human apoB and mouse GAPDH gene expression in the livers and intestines of three lines of BAC(54,22) mice. To analyze human apoB expression, we used 5 μg of liver RNA and 25 μg of intestinal RNA. To analyze mouse GAPDH expression, we used 2 μg of liver RNA and 2 μg of intestinal RNA.
Defining a Distant LCR with BAC Transgenic Mice

absorption) without affecting the ability to generate lipoproteins in the liver (28). In humans, the inability to assemble and secrete chylomicrons with a preserved capacity to secrete hepatic lipoproteins has been termed Anderson’s syndrome, or chylomicron retention disease (3). In one human family, the apoB gene locus was excluded as the culprit (47). However, this syndrome is heterogeneous, both clinically and pathologically (3), and it would not be surprising if at least a few of these cases were due to mutations in the apoB gene’s intestinal LCR.

While our current studies have added new insights into understanding the control of apoB gene expression, they have undeniably cast a spotlight on a series of new issues. For example, why the apoB gene’s intestinal LCR is located so far away from the structural gene is not known. Nor is it known whether the distant regulatory sequences exist only to govern the apoB gene or whether they might also control the expression of another, yet unidentified, related gene located within the upstream sequences. Finally, the precise identity of the regulatory sequences and how they act over such an enormous distance remain mysterious. Although finding answers to each of these issues will undoubtedly pose challenges, the current studies have opened doors for future investigations. First, the localization of the apoB gene’s intestinal element to a fairly short stretch of DNA means that it is now feasible to embark on DNase I-hypersensitive site mapping (using tissues from transgenic mice or cultured cell lines) to more precisely identify potential regulatory sequences. Second, these studies have provided the underpinnings for a different transgenic mouse approach. Our studies demonstrated that the apoB gene’s intestinal LCR can function when moved closer to the structural gene, and our mapping and subclone sequencing studies have made it possible for any laboratory to obtain the crucial regulatory sequences.

REFERENCES

1. Havel, R. J., and Kane, J. P. (1995) in The Metabolic and Molecular Bases of Inherited Disease (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds) 7th Ed, Vol. 2, pp. 1841–1853, McGraw-Hill Inc., New York
2. Young, S. G. (1990) Circulation 82, 1574–1594
3. Gregg, R. E., and Wetterau, J. R. (1994) Curr. Opin. Lipidol. 5, 81–86
4. Linton, M. F., Farese, R. V., Jr., and Young, S. G. (1993) J. Lipid Res. 34, 521–541
5. Linton, M. F., Farese, R. V., Jr., Chiesa, G., Grass, D. S., Chin, P., Hammer, R. E., Hobbs, H. H., and Young, S. G. (1993) J. Clin. Invest. 92, 3029–3037
6. Callow, M. J., Stoltzfus, L. J., Lawn, B. M., and Rubin, E. M. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 2130–2134
7. McCormick, S. P. A., Ng, J. K., Veniant, M., Boren, J. Pierotti, V., Flynn, L. M., Grass, D. S., Connolly, A., and Young, S. G. (1996) J. Biol. Chem. 271, 11963–11970
8. Nielsen, L. B., McCormick, S. P. A., Pierotti, V., Tam, C., Gunn, M. D., Shizuya, H., and Young, S. G. (1997) J. Biol. Chem. 272, 29752–29758
9. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) J. Mol. Graph. 6, 13–27
10. Ferrin, L. J., and Camerini-Otero, R. D. (1986) Science 234, 1494–1497
11. Ferrin, L. J., and Camerini-Otero, R. D. (1994) Nat. Genet. 6, 379–383
12. Ferrin, L. J. (1995) Genet. Eng. 17, 21–30
13. Bore´n, J., Lee, I., Callow, M. J., Rubin, E. M., and Innerarity, T. L. (1996) Genome Res. 6, 1123–1130
14. Callow, M. J., Ferrin, L. J., and Rubin, E. M. (1994) Nucleic Acids Res. 22, 4348–4349
15. Allan, C. M., Taylor, S., and Taylor, J. M. (1997) J. Biol. Chem. 272, 11919–11921
16. McCormick, S. P. A., Allan, C. M., Taylor, J. M., and Young, S. G. (1997) in Transgenic Animals: Generation and Use (Houdebine, L. M., eds) pp. 273–281, Harwood Academic Publishers, The Netherlands
17. Schiedl, A., Larin, Z., Montoulie, L., Thies, E., Kelsey, G., Lebrach, H., and Schutz, G. (1993) Nucleic Acids Res. 21, 4783–4787
18. Schiedl, A., Montoulie, L., Kelsey, G., and Schutz, G. (1993) Nature 362, 258–261
19. Duell, T., Nielsen, L. B., Jones, A., Young, S. G., and Weier, H-U. (1997) Cytogeten. Cell Genet. 79, 64–70
20. Steinberg, D. and Witztum, J. L. (1991) J. Am. Med. Assoc. 264, 3047–3052
21. Young, S. G., and Parthasarathy, S. (1994) West. J. Med. 160, 153–164
22. Paul, M., Wagner, J., Hoffmann, S., Urata, H., and Ganten, D. (1994) Anna. J. Biol. Chem. 270, 7932–7936
23. Paulweber, B., Brooks, A. R., Nagy, B. P., and Levy-Wilson, B. (1991) J. Biol. Chem. 266, 21956–21961
24. Paulweber, B., and Levy-Wilson, B. (1991) J. Biol. Chem. 266, 24161–24168
25. Paulweber, B., and Fortier, C. (1989) J. Biol. Chem. 264, 21196–21204
26. Levy-Wilson, B. (1995) Prog. Nucleic Acid Res. Mol. Biol. 50, 161–190
27. Purcell-Huyhn, D. A., Farese, R. V., Jr., Johnson, D. F., Flynn, L. M., Pierotti, V., Newland, D. L., Linnenbach, F., Sanan, D. A., and Young, S. G. (1995) J. Clin. Invest. 95, 2246–2257
28. Young, S. G., Chang, C. M., Pitao, R. E., Burri, B. J., Connolly, A., Flynn, L., Pappu, A. S., Wong, J. S., Hamilton, R. L., and Farese, R. V., Jr. (1995) J. Clin. Invest. 96, 2932–2946
29. Fan, J., McCormick, S. P. A., Krauss, R. M., Taylor, S., Quan, R., Taylor, J. M., and Young, S. G. (1995) Circulation 92, 154 (abstr).
30. Cotes, S., Sarner, N., Vidal, M., Cox, A., and Grufsveld, F. (1990) New Biol. 2, 801–811
31. Walsh, A., Azrulan, N., Wang, K., Marciglano, A. O’Connell, A., and Breslow, J. L. (1990) J. Lipid Res. 31, 617–622
32. Montoulie, L., Umland, T., and Schutz, G. (1996) EMBO J. 15, 6026–6034
33. Jones, B. K., Monks, B. R., Liebler, S. A., and Cooke, N. E. (1995) Mol. Cell. Biol. 15, 7010–7021
34. Lowrey, C. H., Bodine, D. M., and Nienhaus, A. W. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 1143–1147
35. Simonet, W. S., Bucay, N., Lauer, S. J., Pitta, R., Weisgraber, K. H., and Taylor, J. M. (1990) Trans. Assoc. Am. Phys. 103, 119–128
36. Simonet, W. S., Bucay, N., Lauer, S. J., and Taylor, J. M. (1993) J. Biol. Chem. 268, 8221–8229
37. Schachter, N. S., Zhu, Y., Walsh, A., Breslow, J. L., and Smith, J. D. (1993) J. Lipid Res. 34, 1609–1707
38. Taylor, J. M., Simonet, W. S., Lauer, S. J., Zhu, G., and Walker, D. (1993) Curr. Opin. Lipidol. 4, 84–89
39. Grosveld, F., van Assendelft, G. B., Greaves, D. R., and Kollia, G. (1987) Cell 51, 975–985
40. Hug, B. A., Wesselchsmeid, R. L., Fiering, S., Bender, M. A., Epper, E., Groudine, M., and Ley, T. J. (1996) Mol. Cell. Biol. 16, 2906–2912
41. Martin, D. I., K. Fiering, S., and Groudine, M. (1996) Curr. Opin. Genet. Dev. 6, 488–495
42. Caterina, J. J., Ryan, T. M., Pawlik, K. M., Palmer, R. L., Behringer, R. R., and Townes, T. M. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 1626–1630
43. Stamatosyanopoulos, J. A., Goodwin, A. J., Joyce, T., and Lowrey, C. H. (1995) EMBO J. 14, 106–116
44. Silver, L. M. (1995) *Mouse Genetics: Concepts and Applications*, Oxford University Press, New York
45. Weier, H.-U. G., Wang, M., Mullikin, J. C., Zhu, Y., Cheng, J.-F., Greulich, K. M., Bensimon, A., and Gray, J. W. (1995) *Hum. Mol. Genet.* 4, 1903–1910
46. Townes, T. M., and Behringer, R. R. (1990) *Trends Genet.* 6, 219–223
47. Strich, D., Goldstein, R., Phillips, A., Shemer, R., Goldberg, Y., Razin, A., and Freier, S. (1993) *J. Pediatr. Gastroenterol. Nutr.* 16, 257–264
48. Hammer, R. E., Maika, S. D., Richardson, J. A., Tang, J.-P., and Taurog, J. D. (1990) *Cell* 63, 1099–1112
49. McCormick, S. P. A., Peterson, K. R., Hammer, R. E., Clegg, C. H., and Young, S. G. (1996) *Trends Cardiovasc. Med.* 6, 16–24
50. Young, S. G., Lusis, A. J., and Hammer, R. E. (1998) in *Molecular Basis of Heart Disease* (Chien, K. R., Breslow, J. L., Leiden, J. M., Rosenberg, R. D., and Seidman, C., eds) W. B. Saunders, Philadelphia, in press
51. McCormick, S. P. A., Linton, M. F., and Young, S. G. (1994) *Genet. Anal. Tech. Appl.* 11, 158–164