Evolutionary Duplication of a Hepatic Control Region in the Human Apolipoprotein E Gene Locus

IDENTIFICATION OF A SECOND REGION THAT CONFER HIGH LEVEL AND LIVER-SPECIFIC EXPRESSION OF THE HUMAN APOLIPOPROTEIN E GENE IN TRANSGENIC MICE

Charles M. Allan†‡§, David Walker†, and John M. Taylor†‡§

From the G Gladstone Institute of Cardiovascular Disease, the Cardiovascular Research Institute, and the Department of Physiology, University of California, San Francisco, California, 94141-9100

We have identified a second hepatic control region (HCR-2) in the human apolipoprotein (apo) E gene locus that confers liver expression of the human apoE gene in transgenic mice. This HCR-2 sequence is located 27 kilobases downstream of the apoE gene and 10 kilobases downstream of the previously described liver-specific enhancer (HCR-1). Nucleotide sequence analysis of the HCR-2 region revealed a sequence that shares 85% identity to the functional 319-base pair domain of HCR-1. To test its activity, transgenic mice were prepared with a fusion construct containing a human apoE gene fragment, which is not normally expressed in the liver, ligated to a 632-base pair region containing the HCR-2 sequence. This construct resulted in high levels of liver-specific apoE transgene expression, indicating that HCR-2 can function as a hepatic enhancer and has an activity similar to that of HCR-1. Hence, these findings suggest that there are at least two hepatic control regions, HCR-1 and HCR-2, capable of controlling the liver expression of this human apolipoprotein gene locus.

The genes encoding human apolipoprotein (apo) E, apo-C-I, and apo-C-II are located within a 45-kb cluster on chromosome 19 (1, 2). We recently identified and characterized another human gene in this cluster, the apoC-IV gene, that shares structural characteristics with the other apolipoprotein genes of this locus (3). The human locus also contains an apoC-I pseudogene (4), located between the apoC-I and apoC-II genes, which appears to have arisen from a duplication event early in the primate lineage, approximately 39 million years ago (5, 6).

The apoE, apoC-I, and apoC-II genes code for apolipoproteins that are components of plasma lipoproteins. These apolipoproteins have evolved distinct functions in lipid metabolism. Apolipoprotein E mediates lipoprotein clearance from the plasma by acting as a ligand for the low density lipoprotein receptor (7, 8) and the low density lipoprotein receptor-related protein (9). A role for apoE in neuron growth and homeostasis, as well as the pathology of Alzheimer's disease, has been suggested by recent studies (for review, see Ref. 10). Apolipoprotein C-II is an essential cofactor for lipoprotein lipase and therefore has an important role in the hydrolysis of lipoprotein triglycerides (11). The precise function of apoC-I is uncertain, but it may inhibit the apoE-mediated cellular uptake of lipoproteins (12, 13). This would suggest a role in modulating lipoprotein catabolism. It may also function as one of several cofactors for the enzyme lecithin:cholesterol acyltransferase (14).

The primary site of synthesis for apolipoproteins E, C-I, and C-II is the liver (4, 15, 16). Previous studies indicated that the regulatory elements required for hepatic expression of the human apoE and apoC-I genes are contained in a specific region, known as the hepatic control region, denoted HCR (17). The expression of various genomic constructs containing the human apoE and/or apoC-I genes in transgenic mice demonstrated that a 774-bp region, located 15 and 5 kb downstream of the apoE and apoC-I genes, respectively, directs high level liver expression of both genes (17). More recent studies in our laboratory have determined that full activity of the HCR is provided by a 319-bp sequence, which contains at least three domains that are cooperatively involved in directing high level and liver-specific expression of the human apoE transgene. A 154-bp subfragment of this region has been reported to have liver enhancer capability (18).

We now report the identification of a second HCR sequence (denoted HCR-2) at this locus, located ~5.5 kb downstream of the apoC-I' pseudogene. HCR-2 shares 85% homology to the functional 319-bp HCR (henceforth referred to as HCR-1) sequence. In addition, we show that HCR-2, which appears to have arisen from the duplication event that formed the apoC-I' pseudogene, has retained the necessary sequences for directing high level and liver-specific expression of human apoE mRNA in transgenic mice.

MATERIALS AND METHODS

Isolation and Characterization of Human Genomic Sequences—A P1 plasmid clone (p1.198) containing the entire human apoE/C-I/C-IV/C-II gene locus was identified by screening a human P1 genomic library (Genome Systems, Inc., St. Louis, MO), and the 90-kb genomic insert was mapped as described previously (3). A 10.5-kb BamHI fragment of p1.198, located about 5.5 kb downstream of the apoC-I' pseudogene, was cloned into pBluescript SK+ (Stratagene) for nucleotide sequence analysis. A 632-bp genomic fragment containing the HCR-2 sequence was isolated from this 10.5-kb BamHI fragment and then ligated to purified HEG1, a 10-kb genomic sequence containing the human apoE gene (19). The resulting fusion construct, HCR-2-HEG1, was isolated byelectrophoresis in 1.0% SeaPlaque GTG agarose (FMC). Gel slices containing the construct were digested with β-agarase (FMC Bioproducts, Rockland, ME) and then purified using a Magic DNA Clean-up System kit (Promega).

* This research was funded in part by National Institutes of Health Grant HL-37063 (to J. M. T.) and a Fellowship from the American Heart Association, California Affiliate (to C. M. A.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Gladstone Institute of Cardiovascular Disease, P. O. Box 419100, San Francisco, CA 94141-9100. Tel.: 415-826-7500; Fax: 415-285-5632.

‡ The abbreviations used are: apo, apolipoprotein; kb, kilobase(s); bp, base pair(s); HCR, hepatic control region.

2 Q. Dang, D. Walker, S. Taylor, C. Allan, P. Chin, J. Fan, and J. M. Taylor, manuscript in preparation.
Hepatic Control Regions of the Human ApoE Gene Locus

Nucleotide Sequence Analysis—The nucleotide sequences of the DNA fragments were determined using a Taq Dye-Deoxy Terminator cycle sequencing kit (Applied Biosystems, Inc., Foster City, CA) and automated sequencing or by using a Sequenase sequencing kit (U. S. Biochemical Corp.). Selected subclones of the 10.5-kb BamHI fragment of p1.198 DNA were sequenced as described above or by aliquonucleotide walking, as described previously (20). The complete nucleotide sequences of both complementary strands of overlapping genomic clones were determined. The nucleotide sequence of the intergenic region between the human apoC-I gene and the human apoC-I pseudogene was determined using similar sequencing strategies and techniques. 3

Transgenic Mice—Transgenic animals were generated essentially as described previously (19). The HCR-2. HEG1 construct at 2–3 ng/ml in 10 ml Tris-HCl, pH 7.4, 0.1 ml EDTA buffer was microinjected into the fertilized embryos of superovulated FVB/N mice. Four independent transgenic founder lines were identified by Southern blot analysis of tail DNA, using an apoE cDNA probe as described previously (19). Copy numbers of the transgenes were determined by Southern dot blot analysis and quantification using a Fujix Bas1000 Bio-imaging Analyzer (Fuji Photo Co., Ltd., Japan). Human genomic DNA from cultured HepG2 cells was used as standards.

RNA Preparation and RNase Protection Analysis—Total RNA from mouse tissues and from cultured HepG2 cells was isolated using guanidine thiocyanate (21). Human liver total RNA (a gift from Dr. Clive Pullinger) was isolated as described previously (22). An antisense RNA probe for human apoE mRNA was transcribed (19) using [α-32P]UTP (800 mol/Ci, DuPont NEN) in the presence of T7 RNA polymerase (Stratagene). A mouse actin mRNA probe (pTRI-actin, Ambion Inc.) was used to confirm that equivalent levels of RNA were present in each tissue sample. RNase protection analysis was performed as described previously (23), using 5 μg of total cellular RNA per sample.

Identification of a CDNA Clone, pC4.1. Containing HCR-related Sequence—A partial cDNA clone (pC4.1) was obtained during the characterization of the 5′ end of the human apoC-IV gene. Analysis of human liver cDNA was performed using human liver 5′-rapid amplification of cDNA ends ready cDNA (Clontech Laboratories, Inc., CA) as described previously (3). The polymerase chain reactions for this analysis were performed using GeneAmp polymerase chain reaction Core Reagents (Perkin Elmer, Cetus, NJ) according to the manufacturer’s recommendations. Secondary polymerase chain reaction products were cloned into pBluescript SK + (Stratagene) for nucleotide sequence analysis. An antisense RNA probe to the 173-bp 5′ sequence of the cDNA insert of pC4.1 was generated using [α-32P]UTP (800 mol/Ci, DuPont NEN) in the presence of T3 RNA polymerase (Stratagene). The genomic location of this 5′ region was mapped by Southern blotting of p1.198 DNA using the RNA probe.

RESULTS AND DISCUSSION

Isolation and Characterization of Human Genomic Clones—The hepatic control region (HCR-1) that directs high level and liver-specific expression of the human apoE and apoC-I genes in transgenic mice is located between two closely related sequences, the apoC-I gene and apoC-I pseudogene (17). To determine if the duplication event that formed the human apoC-I pseudogene also may have resulted in a duplication of the HCR, we analyzed the nucleotide sequence of a genomic 10.5-kb BamHI fragment that was located approximately 5.5 kb downstream of the apoC-I pseudogene. This BamHI fragment, derived from a P1 plasmid (p1.198) containing the entire human apoE/C-I/C-IV/C-II gene locus, was recently shown to contain the human apoC-IV gene (3). Nucleotide sequence analysis revealed that the 5′ end of the 10.5-kb fragment contained a sequence that shares 86% homology to the HCR-1 sequence (Fig. 1). The striking homology between this newly identified region (HCR-2) and the HCR-1 sequence implies a common evolutionary origin and conserved function of both regions.

These results suggest that HCR-2, which is located approximately 10 kb downstream of HCR-1 (see Fig. 2), arose from the duplication event that formed the apoC-I pseudogene. In comparison to the 3′-flanking nucleotide sequence of both HCR-1 and HCR-2 showed that the sequence homology ended at two

3 D. Walker and J. Taylor, unpublished results.
The level of transgene expression in both tissues, including the liver. These transgenic lines incorporate 5 and 13 copies of the HCR-2.HEG1 construct, respectively. HEG1 without HCR-2 is expressed in only the kidney, as shown. The bands shown here correspond to the expected protected fragment that was examined.

apoE gene HEG1 fragment (Fig. 3). The resulting construct, HCR-2.HEG1, was used to generate transgenic mice. We examined two independent founder lines of transgenic mice for expression of the HCR-2.HEG1 apoE transgene in six different tissues, including the liver. These transgenic lines incorporated 5 and 13 copies of the HCR-2.HEG1 construct, respectively. By RNase protection analysis, each transgenic line showed the same pattern of human apoE transgene expression (Fig. 3), with both the liver and kidney expressing high levels of human apoE mRNA. The levels of transgene expression in both tissues were comparable with the expression levels that were previously reported using HEG1 constructs containing the 774-bp HCR-1 sequence (17). Therefore, like the HCR-1 sequence, the 632-bp HCR-2 sequence can direct liver-specific and high level expression of the human apoE gene.

The specific roles of two HCR domains in the apoE gene locus remain to be determined. It is possible that each gene in the locus is controlled primarily by an individual HCR. Previous findings demonstrated that HCR-1 is sufficient to direct high levels of apoE and apoC-I expression in the liver (17). Our results here indicate that HCR-2 also can function to control the expression of the apoE gene. However, the downstream location of HCR-2 suggests that this latter domain may have a more important function in directing apoC-IV and apoC-II gene expression. Alternatively, the activities of HCR-1 and HCR-2 might be combined to yield increased levels of liver expression for one or more genes.

It is noteworthy that the duplication event that gave rise to two copies of the apoC-I gene (which would include the HCR domain) was estimated to have occurred early in the primate lineage (5, 6). The corresponding mouse (25) and rat (26) apoE gene loci lack the equivalent of the human apoC-I pseudogene. Therefore, unlike the human locus, the mouse and rat gene loci may contain only one HCR-like domain located between the apoC-I and apoC-IV genes (27). Thus, the liver-specific expression of the apoE gene locus may be controlled by a single HCR domain in nonprimate species.

Characterization of a cDNA Clone, pC4.1, Containing HCR-related Sequence—During characterization of the 5' end of the human apoC-IV gene by the 5'-rapid amplification of cDNA ends technique, we identified 11 cDNA clones that contained sequences corresponding to the expected first and second exons of the apoC-IV gene (3). Subsequently, an additional cDNA clone was identified that contained the second exon sequence of the apoC-IV gene and 173 bp of upstream sequence that showed no homology to the expected first exon of the apoC-IV gene (refer to Fig. 1). Comparison of this 173-bp sequence with sequences in the GenBank and EMBL data bases revealed no homology to any previously identified human gene. However, Southern blot analysis of p1.198 DNA, which contains the entire human apoE/C-I/C-IV/C-II gene locus, showed that an antisense probe to the 173-bp cDNA sequence hybridized to the genomic 10.5-kb BamHI fragment of p1.198 that contains the apoC-IV gene and extends ~6 kb upstream (data not shown). Sequence analysis of this 10.5-kb fragment revealed that this 173-bp sequence was identical to the 3' portion of the HCR-2 sequence (Fig. 1).

The presence of this 173-bp sequence at the 5' end of a human liver-derived cDNA clone (pC4.1) suggests that a genomic region positioned upstream of this sequence may initiate transcription in the liver. However, the adjacent 5' sequence of HCR-2 lacks typical promoter motifs, such as TATA boxes or Sp1-binding sites (see Fig. 1), suggesting that a conventional gene promoter is not present in this region. While it
is possible that this 173-bp sequence represents a distant exon of the apoC-IV gene. Previous studies of both the human and mouse apoC-IV genes provide compelling evidence that this gene consists of only three exons in both species (3, 27). RNase protection analysis of human liver total RNA using an antisense probe to the 173-bp sequence did not detect any liver RNA transcript (data not shown), indicating that this sequence is not highly expressed in the liver and that the partial cDNA insert of pC4.1 probably results from a rare transcript. Thus, the pC4.1 sequence may represent an alternative splicing product of this rare primary transcript (see Fig. 4). Similar transcripts have been reported for the homologous monkey and Balb/c mouse apoC-I/IV and apoC-II genes, in which alternative splicing resulted in extended apoC-II transcripts that also contained exons of the apoC-I/IV gene (3, 27). Since they amount to only minor amounts of the transcribed products of both genes, the functional significance of these hybrid transcripts is not clear (3, 28).

The similarity of this 173-bp sequence to an exon is consistent with the presence of AA at its 3′ end and GTAAGT at the 5′ end of the adjacent intron-like region (see Fig. 1). This potential exon-intron boundary sequence (AA/GTAAGT) is similar to the consensus (A/C)G/GT(A/G)AGT for a splice site donor sequence (29). Interestingly, a consensus sequence for a splice acceptor site is not found adjacent to the 5′ end of the 173-bp exon, which indicates the absence of an upstream exon. It is noteworthy that the consensus splice site donor sequence is not found in the corresponding HCR-1 sequence (Fig. 1), which indicates the absence of an upstream exon. It is noteworthy that the consensus splice site donor sequence is not found in the corresponding HCR-1 sequence (Fig. 1), which suggests that HCR-1 is unlikely to produce comparable transcripts (i.e. containing both HCR-1 and apoC-I′ sequences).

Thus, it appears that part of HCR-2 can direct rare transcription of a sequence that resembles an exon. It is interesting that the corresponding segment of HCR-1 (i.e. the region upstream of the sequence homologous to the exon-like 173-bp region of HCR-2) is the domain that is required for directing high level and liver-specific expression of the human apoE gene.2 In addition, it has been shown that a 154-bp segment of this 5′ portion of HCR-1 can confer transgene expression in the liver of transgenic mice (18). Thus, the key regulatory region of HCR-1, and possibly HCR-2, may be functionally distinct from the 3′ portion that was detected as a transcribed sequence.

In summary, we have identified a new regulatory sequence, HCR-2, in the human apoE/C-I/C-IV/C-II gene locus that is closely related to HCR-1. HCR-2, like HCR-1, can function to direct high level and liver-specific expression of the human apoE gene in transgenic mice. Previous studies reported that no liver enhancers were located within 30 kb of the 5′-flanking region of the apoE gene (17), and our preliminary analysis of apoC-II gene constructs in transgenic mice have indicated that no liver enhancer lies within 3 kb of its 3′-flanking sequence.4

Thus, there are two HCR domains in an ~80 kb genomic segment that contains four genes that are expressed in the liver. The potential roles of the two liver-specific control regions in this gene locus remain to be determined. Our present findings suggest that the tissue-specific expression of the human apoE gene and other genes at this locus may be more complicated than previously believed.

Acknowledgments—We thank John Carroll for assistance with graphics, Gary Howard for editorial assistance, Stacy Taylor for excellent technical assistance, and Susannah White for help with manuscript preparation.

REFERENCES

1. Myklebost, O., and Rogne, S. (1988) Hum. Genet. 78, 244–247
2. Smit, M., van der Kooi-Meijs, E., Frants, R. R., Havekes, L., and Klasen, E. C. (1988) Hum. Genet. 78, 90–93
3. Allan, C. M., Walker, D., Segrest, J. P., and Taylor, J. M. (1995) Genomics 28, 291–300
4. Lauer, S. J., Walker, D., Elshourbagy, N. A., Reardon, C. A., Levy-Wilson, B., and Taylor, J. M. (1988) J. Biol. Chem. 263, 7277–7286
5. Luo, C.-C., Li, W.-H., and Chan, L. (1989) J. Lipid Res. 30, 1735–1746
6. Raisonier, A. (1991) J. Mol. Evol. 33, 211–219
7. Mahley, R. W., and Innerarity, T. L. (1983) Biochim. Biophys. Acta 737, 197–222
8. Brown, M. S., and Goldstein, J. L. (1986) Science 232, 34–47
9. Beisiegel, U., Weber, W., Hirke, G., Herz, J., and Stanley, K. K. (1989) Nature 341, 162–164
10. Weisgraber, K. H., Pitas, R. E., and Mahley, R. W. (1995) in Atherosclerosis X (Woodford, F. P., Davignon, J., and Sniderman, A., eds) pp. 670–674, Elsevier Science, Amsterdam
11. Breckenridge, W. C., Little, J. A., Steiner, G., Chow, A., and Poapst, M. (1978) N. Engl. J. Med. 298, 1265–1273
12. Weisgraber, K. H., Mahley, R. W., Kowal, R. C., Herz, J., Goldstein, J. L., and Brown, M. S. (1990) J. Biol. Chem. 265, 22453–22459
13. Sehayek, E., and Eisenberg, S. (1991) J. Biol. Chem. 266, 18259–18267
14. Steyer, E., and Kostner, G. M. (1988) Biochim. Biophys. Acta 958, 484–491
15. Elshourbagy, N. A., Liu, W. S., Mahley, R. W., and Taylor, J. M. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 203–207
16. Wu, A.-L., and Windmueler, H. G. (1979) J. Biol. Chem. 254, 7316–7322
17. Simonet, W. S., Bucay, N., Lauer, S. J., and Taylor, J. M. (1993) J. Biol. Chem. 268, 8221–8229
18. Shachter, N. S., Zhu, Y., Walsh, A., Breslow, J. L., and Smith, J. D. (1993) J. Lipid Res. 34, 1699–1707
19. Simonet, W. S., Bucay, N., Lauer, S. J., Wirak, D. O., Stevens, M. E., Weisgraber, K. H., Pitas, R. E., and Taylor, J. M. (1990) J. Biol. Chem. 265, 10809–10812
20. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, p. 13.15, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
21. MacDonald, R. J., Swift, G. H., Przybyla, A. E., and Chirgwin, J. M. (1987) Methods Enzymol. 152, 219–227
22. Auffray, C., and Rougeon, F. (1980) Arch. Biochem. Biophys. 203, 203–207
23. Gilman, M. (1993) in Methods Enzymol. 219, 219–227
24. Fazio, S., Lee, Y.-L., Ji, Z.-S., and Rall, S. C., Jr. (1993) J. Clin. Invest. 92, 1497–1503
25. Hoffer, M. J. V., Hofker, M. H., van Eck, M. M., Havekes, L. M., and Frants, R. R. (1993) Genomics 15, 62–67
26. Shen, P., and Howlett, G. J. (1992) Arch. Biochem. Biophys. 297, 345–353
27. van Eck, M. M., Hoffer, M. J. V., Havekes, L. M., Frants, R. R., and Hofker, M. H. (1994) Genomics 21, 110–115
28. Hoffer, M. J. V., van Eck, M. M., Havekes, L. M., Hofker, M. H., and Frants, R. R. (1993) Genomics 17, 45–51
29. Mount, S. M. (1982) Nucleic Acids Res. 10, 459–472

4 C. Allan, D. Walker, and J. Taylor, unpublished results.