# Human Microsomal Triglyceride Transfer Protein Large Subunit Gene Structure

Daru Sharp, Beverly Ricci, Bernadette Kienzle, Marie C. M. Lin, and John R. Wetterau*

*Department of Metabolic Diseases, Bristol-Myers Squibb, Princeton, New Jersey 08543-4000*

*Received February 21, 1994; Revised Manuscript Received May 5, 1994®*

ABSTRACT: Microsomal triglyceride transfer protein (MTP) is a heterodimer consisting of the multifunctional enzyme protein disulfide isomerase and a unique, large 97-kDa subunit. MTP is required for the assembly and secretion of very low density lipoproteins and chylomicrons by the liver and intestine, respectively. *In vitro*, MTP catalyzes the transport of triglyceride, cholesteryl ester, and phosphatidylcholine between phospholipid surfaces. We have characterized the gene encoding the large subunit of human MTP. It contains 18 exons and spans approximately 55–60 kb. Fluorescent *in situ* hybridization localized this gene to band 4q24 of chromosome 4. A $(CA)_n$ repeat polymorphic marker, which may be useful for investigating a link between the MTP gene and genetic defects in lipid metabolism, was identified in intron 10. Sequence analysis of the 5′ flanking region of the gene revealed potential sites which may bind transcriptional factors and control MTP expression.

Lipid transfer proteins promote the transport of insoluble lipid molecules between membranes. The microsomal triglyceride transfer protein (MTP)[1] appears unique among lipid transfer proteins in that it is an intracellular protein which promotes the transport of neutral lipid (triglyceride and cholesteryl ester) between membranes (Wetterau & Zilversmit, 1984), and it resides within the lumen of the microsomal fraction of hepatocytes and enterocytes (Wetterau & Zilversmit, 1986). Purified MTP from bovine liver contains two subunits with apparent molecular weights of 58 000 and 88 000 (Wetterau & Zilversmit, 1985). The 58 000 molecular weight component has been identified as the multifunctional enzyme protein disulfide isomerase (PDI) (Wetterau et al., 1990). Complementary DNAs encoding the large subunit of bovine and human MTP have been cloned and characterized (Sharp et al., 1993).

We have reported that functional MTP was absent in four unrelated human subjects with abetalipoproteinemia (Wetterau et al., 1992), a rare genetic disease characterized as a defect in the assembly and secretion of plasma lipoproteins which contain apolipoprotein B (Kane & Havel, 1989). In the two subjects characterized to date, a homozygous frame shift and a homozygous nonsense mutation in the gene encoding the large subunit of MTP were identified (Sharp et al., 1993). These mutations fully explain the absence of MTP protein and demonstrate that MTP is required for the secretion of chylomicrons and very low density lipoproteins from the intestine and liver, respectively. Shoulders et al. (1993) recently reported a homozygous point mutation in the gene encoding MTP that disrupts normal splicing in another abetalipoproteinemic subject. To further compare MTP to other proteins which play important roles in plasma lipoprotein metabolism, elucidate possible mechanisms for MTP regulation, and facilitate the identification of MTP gene defects, we have characterized the structure of the gene encoding the large subunit of human MTP.

* To whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* July 1, 1994.

[1] Abbreviations: MTP, microsomal triglyceride transfer protein; PDI, protein disulfide isomerase; LCAT, lecithin:cholesterol acyltransferase; CETP, cholesteryl ester transfer protein; apoB, apolipoprotein B; PCR, polymerase chain reaction; bp, base pair; SDS, sodium dodecyl sulfate; SSC, saline–sodium citrate buffer.

## MATERIALS AND METHODS

*Isolation and Characterization of Genomic Clones.* Two million recombinant phage plaques from a human genomic λ DASH (Stratagene) library were screened for genomic sequences homologous to MTP with either a 2.4-kb *Eco*RI fragment from a bovine cDNA clone for the large subunit of MTP, corresponding to bases 180–2600 of the coding region, or a 1.5-kb *Eco*RI to *Apa*I fragment encoding bases −64 to 1497 of the human cDNA. These fragments were purified by agarose gel electrophoresis and $^{32}P$ labeled using a Multiprime labeling kit (Amersham). Hybridization was performed at 60 °C in 6 × SSC (1 × SSC is 150 mM NaCl–15 mM sodium citrate, pH 7.0), 20 mM sodium phosphate, 2 × Denhardt's, and 0.1 mg/mL denatured salmon sperm DNA. Washes were at 60 °C in 1 × SSC and 0.1% SDS. Thirty-six positive clones were plaque purified and further characterized by restriction mapping and hybridization to oligonucleotide probes corresponding to various regions of the cDNA sequence. These two screens yielded overlapping genomic DNA inserts from exon 2 through exon 18. To isolate exon 1, a second human genomic library, in the Lambda Fix II vector (Stratagene), was probed with a 0.5-kb *Eco*RI to *Nco*I restriction fragment of the human cDNA encoding bases −64 to 385. Hybridization of this probe to one million phage plaques yielded seven positives from which exon 1 and the 5′ flanking region of the MTP gene were characterized.

An Applied Biosystems, Inc. (ABI, Foster City, CA), 373 Automated DNA Sequencer was used to directly sequence λ DNA templates using primers designed from human cDNA sequence. Dye-labeled dideoxynucleotide sequencing was performed using the Taq Dye Deoxy Terminator Cycle Sequencing Kit (ABI No. 40113). In a typical sequencing reaction, 1.25 μg of template was used with 4 pmol of primer. Reaction cycles were performed using a Perkin-Elmer Cetus thermal cycler as described in the Taq Dye Deoxy Terminator Cycle Sequencing Protocol. Intron–exon boundaries were identified by comparing cDNA and genomic sequence. Sequencing primers were then designed against intron sequences 5′ and 3′ to each exon. These were used in sequencing reactions to confirm the sequence of the boundaries.

*Analysis of the 5′ Region of the MTP cDNA.* Anchored PCR followed by direct sequencing of the amplified products

0006-2960/94/0433-9057$04.50/0 

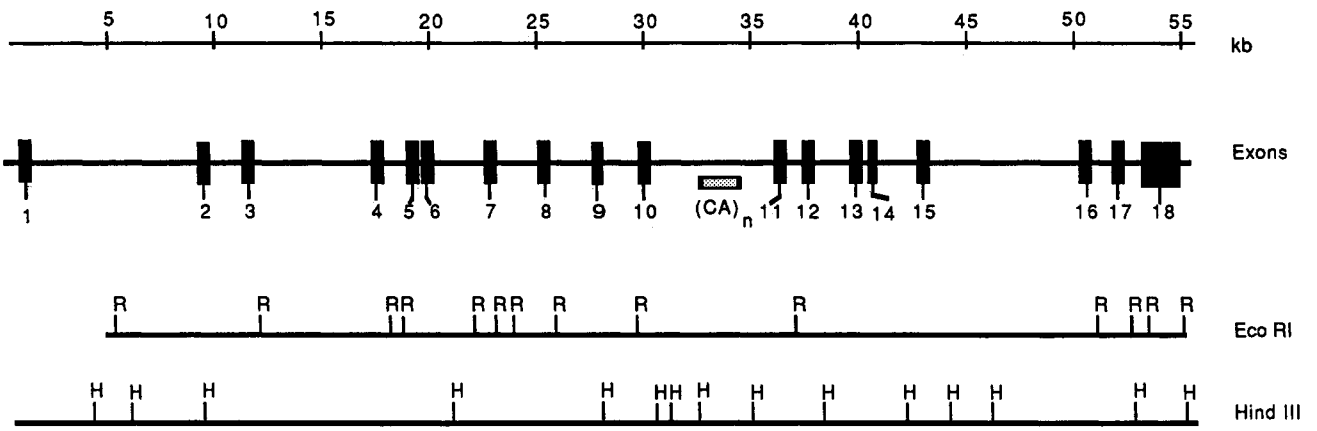


FIGURE 1: Organization of the human gene for the large subunit of MTP. The top line indicates genomic distance in kilobases. The 18 exons of the MTP gene are mapped on the second line. Refer to Tables 1 and 2 for exon and intron sizes. The distance between exons 1 and 2 is greater than 5 and less that 8 kb. The third and fourth lines are maps of the *Eco*RI (R) and *Hin*dIII (H) restriction sites, respectively. 5′ of the first *Eco*RI site indicated, there are five additional *Eco*RI sites which produce fragments of 2.2, 1.8, 1.5, 1.1, 0.59, and 0.58 kb. Exon 1 is encoded on the 1.5-kb *Eco*RI fragment. The approximate location of the CA repeat is indicated in intron 10.

was used to identify the 5′ end of the human message. Briefly, mRNA was isolated from HepG2 cells, a human hepatoblastoma derived cell line which expresses MTP. The RNA was reverse transcribed into cDNA with a primer complementary to the human cDNA sequence 64–113 base pairs downstream of the translation start site (Sharp et al., 1993). The sequences corresponding to the 5′ end of the cDNA were amplified utilizing a 5′AmpliFINDER RACE kit (Clonetech). An oligonucleotide was ligated to the 3′ end of the first strand of cDNA. PCR amplification was performed using one primer complementary to the anchored oligonucleotide and a second primer complementary to the human MTP cDNA sequence −15 to +18 relative to the translational start site. The single PCR product was isolated by agarose gel electrophoresis and sequenced directly as described for sequencing genomic DNA.

*Restriction Fragment Length Polymorphism Analysis.* Genomic blots of DNA from four individuals which had been digested with *Eco*RI, *Msp*I, *Taq*I, *Bgl*II, *Hin*dIII, *Pst*I, *Bam*HI, *Pvu*II, and *Rsa*I were obtained from Bios Laboratories. The blots were probed simultaneously with a bovine cDNA fragment encoding exons 2 through part of exon 18 and a 0.5-kb human cDNA fragment encoding another portion of exon 18. Both probes were $^{32}P$ random prime labeled as described above. Hybridization and washes were as described for the library screens.

*Microsatellite Analysis.* A $(CA)_9$ oligonucleotide was $^{32}P$ end labeled and used as a probe in Southern blot analysis of λ phage clones containing inserts of the human MTP gene. A region in intron 10 which strongly hybridized to this probe was subcloned into pUC18 and sequenced. A CA dinucleotide repeat was identified. Forward (5′-CATCTGGGTTCTG-GTCTTAT-3′) and reverse (5′-ATTTCCCCATAGTCT-CACCT-3′) primers designed from this sequence were used to amplify a 350-bp fragment containing this satellite from the genomic DNAs of 18 individuals. PCR conditions were as follows: the samples were denatured at 97 °C for 2 min followed by 35 cycles of denaturing for 30 s at 94 °C, annealing for 30 s at 50 °C, and extension at 72 °C for 1 min. The process was completed with a final extension at 72 °C for 7 min, and the samples were stored at 4 °C until further analysis. The amplification products were agarose gel purified, blunt ended, and subcloned into pUC18. Individual recombinant clones were sequenced to determine the structure of the 36 different alleles. A homozygous genotype was assigned when the PCR product gave a unique band after gel electrophoresis and 6–9 individual clones revealed the same CA repeat sequence.

*Chromosomal Localization.* Chromosomal localization was performed by Bios Laboratories. Briefly, a λ phage clone containing a 20-kb fragment of the MTP gene encoding from mid intron 10 to mid intron 15 was labeled by nick translation with digoxigenin dUTP. The labeled probe was hybridized to normal human metaphase chromosomes in a solution containing 50% formamide, 10% dextran sulfate, 2 × SSC, and sheared human DNA. The hybridization signal was detected with antidigoxigenin fluorescein isothiocyanate. Chromosomes were counterstained with propidium iodide. Double hybridization with the λ clone and a chromosome-specific centromere probe was used to localize the MTP gene.

## RESULTS AND DISCUSSION

*Organization of the Gene for the Large Subunit of MTP.* A continuous genomic map of the MTP gene, constructed from overlapping phage clones, is shown in Figure 1. The gene spans 55–60 kb. There are 18 exons corresponding to known human cDNA sequence (see Table 1 for summary). Exons 1 and 18 also include 5′ and 3′ noncoding sequences, respectively. The average length of the coding exons is 153 bp, excluding exons 1 and 18 which contain noncoding sequences. Intron sizes range from 91 bp to greater than 7 kb. Nucleic acid and protein databank searches with exon sequences revealed no known functional domains or homologies. All intron/exon junctions followed the GT·AG convention (Breathnach et al., 1978). Table 2 lists the 34 junctions and intron sequences adjacent to each junction.

*Sequence of the 5′ Flanking Region of the Gene.* The 5′ noncoding sequence of human MTP cDNA reported earlier (Sharp et al., 1993) contains 64 bases upstream from the translation start codon. Anchored PCR was used to determine the 5′ end of the cDNA. Direct sequencing of two independent PCR products revealed 86 bp of noncoding sequence upstream of the initiation codon. All 86 bases of the 5′ nontranslated sequence are encoded by exon 1. Exon 1 also has 61 base pairs which encode the signal peptide and two amino acids of the mature protein. The initiation codon is preceded by an ATG, 25 bases upstream and out of frame of the translational start signal; however, the surrounding nucleotide sequences of this ATG do not create a favorable initiation environment. The second ATG lies within a much more favorable context for initiation of translation (Kozak, 1987).

Table 1: Exon Size in Base Pairs and Amino Acids[a]

| exon | nucleotide no. | size (bp) | amino acid no. | size (aa) |
|---|---|---|---|---|
| 1 | −86 to 61 | 149 | 1–21 | 20+ |
| 2 | 62–249 | 188 | 21–83 | 62++ |
| 3 | 250–393 | 144 | 84–131 | 48 |
| 4 | 394–501 | 108 | 132–167 | 36 |
| 5 | 502–618 | 117 | 168–206 | 39 |
| 6 | 619–758 | 140 | 207–253 | 46++ |
| 7 | 759–909 | 151 | 253–303 | 50+ |
| 8 | 910–1067 | 159 | 304–356 | 53++ |
| 9 | 1068–1236 | 167 | 356–412 | 56+ |
| 10 | 1237–1344 | 108 | 413–448 | 36 |
| 11 | 1345–1557 | 213 | 449–519 | 71 |
| 12 | 1558–1769 | 212 | 520–590 | 70++ |
| 13 | 1770–1867 | 98 | 590–623 | 32++ |
| 14 | 1868–1989 | 122 | 623–663 | 40++ |
| 15 | 1990–2217 | 228 | 664–739 | 76 |
| 16 | 2218–2342 | 125 | 740–781 | 41++ |
| 17 | 2343–2513 | 170 | 781–838 | 56+++ |
| 18 | 2514–(>3200)[b] | >686 | 838–894 | 56+ |

[a] Nucleotides are numbered with the A of the translational start codon as +1. Amino acids are included as part of an exon if any part of the codon is encoded by the corresponding nucleotides. Exon sizes are indicated as the number of complete codons encoded (whole number). Additional bases contributing to codons are indicated by (+). [b] The size of exon 18 is based upon the length of the mRNA as determined by northern blot analysis (Sharp et al., 1993).

The sequences 5′ of the transcriptional start site do not contain a consensus sequence for a TATA box; however, there is an AT-rich sequence, AAAGATAAA, about 30 bases upstream of the transcriptional start site which may be recognized by the TATA binding protein, TFIID. The sequence at the transcription start site, GTCACTCCCT, conforms to the weak consensus, YYCAYYYYY, for a transcriptional start site initiator element for TATA-containing promoters (Corden et al., 1980). Assays for MTP activity (Wetterau & Zilversmit, 1986) and northern blot analysis (Sharp et al., 1993; Shoulders et al., 1993) have indicated that MTP is expressed in the liver and intestine. Low levels of MTP mRNA have also been detected by northern blot analysis in ovary, testis, and kidney (Shoulders et al., 1993). Computer-assisted analysis (University of Wisconsin Genetics Computer Group sequence analysis software) of 743 bases upstream of the transcription start site revealed sequences homologous to several known cis-acting elements that may regulate the tissue-specific expression of MTP (Figure 2).

```
-743        CCCCTCTTAATCTCTTCCTAGAAATGAGATTCAGAAAGGACAG
-700  GACTGCATCCAGCCTGTTTGGGAACTCAGACAAATGTGTGTTGTCACAGA
-650  CACAAATAGAGGTCTACTATGAAATAATTGGCTTGCTAGTGTGCTAATGA
-600  CAGACAATGCTGATTTGCTCCAACCTCATACAGTTTCACACATAAGGACA
-550  ATCATCTATGTTTCATGAAAGTTCTATCTACTTTAACATTATTTTGAAGT
-500  GATTGGTGGTGGTATGAATTAACAGTTTAAATTTAAATCCTAAAATTCAG
-450  TGTGAATTTTTTATAATAGCATAAAAATTCAAAGATGTCCATACAAGAAA
-400  AATTAAAATTTGGTTAGGTTTAGCAGAGTTTGAGAATCCTTACTACCCTC
-350  CCACATAGTATTGTAATGTGAATATAGGCAGTTACTATTACAGGCATAAT
-300  GATGATTATGTATTAAGCAGAAAGAAGTATCACCACCAGTTTTTTTCTTT
-250  GAATGCCCCTCAGTACTTCTGCATTTATAGGATGGTAGACTGGTTTGGTT
-200  TAGCTCTCAAAAGTGAAAACATTTAAAGTTTCCTCATTGGGTGAAAAAAA
-150  TTAAAAAGAGTGAGAGACTGAAAACTGCAGCCCACCTACGTTTAATCATT
-100  AATAGTGAGCCCTTCAGTGAACTTAGGTCCTGATTTTGGAGTTTGGAGTC
 -50  TGACCTTTCCCCAAAGATAAACATGATTGTTGCAGGTTCTGAAGAGGGTC
  +1  ACTCCCTCACTGGCTGCCATTGAAAGAGTCCACTTCTCAGTGACTCCTAG
  51  CTGGGCACTGGATGCAGTTGAGGATTGCTGGTCAAT ATG ATT CTT
                                           M   I   L
```

FIGURE 2: Nucleotide sequence of the 5′ flanking sequences for the MTP large subunit gene. Nucleotides are numbered with +1 corresponding to the first transcribed base. Negative numbers indicate 5′ flanking sequences. The first three codons of the translated protein are indicated with the encoded amino acid (single letter designation) below. The TATA-like sequence and consensus initiator sequence are underlined twice between bases −38 and −28 and bases −2 and +6, respectively. The first upstream and out-of-frame ATG is also underlined. Possible sites for interaction with known transcription factors of interest are HNF-5 (bases −649 to −643 and −186 to −180), C/EBP (−491 to −484, −143, to −136, and −55 to −48), HNF-1 (−174 to −162, −153 to −141, −112 to −100, and −108 to −96), SP-1 (−123 to −118), AP-1 (−109 to −103), and CTF/NF-1 (−624 to −618). Sites with exact matches to published consensus sequences are underlined. All other sequences contain a single base pair mismatch.

Possible upstream promoter elements within the first 200 base pairs of the transcription start site include sequences recognized by the liver-specific C/EBP (Costa et al., 1988), HNF1 (Kaling et al., 1991), and HNF-5 (Grange et al., 1991) transcription factors and the general SP-1 (Kadonaga et al., 1987) and AP-1 (Lee et al., 1987) transcription factors (Figure 2). The SP-1, C/EBP, HNF1, and HNF-5 sites in this region have a single base-pair mismatch when compared with proposed

Table 2: Intron/Exon Junction Sequences and Intron Sizes[a]

| 5′ INTRON | EXON | 3′ INTRON | INTRON SIZE |
|---|---|---|---|
| | 1 | ttaaag GTAAGTTTGTGTTGCCTTTTGCTAAACTTTAATTTCCATC | > 5 kb |
| TTACAATGAAAACTGGATATGTGTCATTATCTTTATGCAG gtcaca | 2 | ataacg GTGGGCATTTTCTACCAGATAAATGCAAAGATTAGATATC | 1.8 kb |
| TAGGAAGTTTTTTTTAACAGCTTTCTTTCTGTTACTCCAG atgaag | 3 | ggaaag GTAAAGGGGCGTTTAGATTCCACAACTTTTTCTCCAACTT | 6 kb |
| TTTTTGCTTCATTTGTGTTCTGTTCCCCTCTCCCCACCAG gtcaaa | 4 | aatgag GTACTTACCAATATTAATAAGGATTCAGCATCTCAATAAA | 1.5 kb |
| CAAGGAATCCCAAGCATTATGCCCTTGCCTTTCTTTTTAG gtagat | 5 | aatcag GTATGATAGATGTCACTTTCTTTGAGGCATTAAAATAATT | 0.3 kb |
| GATGATTACTTGTTATAAAGATGGCTATTTATTTATTTAG gtcttg | 6 | atcgaa GTAAGATAATGCTAAAATTTTTATTTTCTTTGCTATTCTT | 2.6 kb |
| GAATGATTATAATATAGCATTTCCCTTTGGTATTATGCAG gcagaa | 7 | ccttct GTAAGTGCAGACAAATATGGGAATAATCATGACATCAGAC | 2 kb |
| GTTATGAGTGGGGTATGAGCCTGCAGTGTATGTTTTGCAG ctctcg | 8 | agtatt GTAAGTTCCCCAACCTTTGTGTGGGGTTGTCTGTCAGAAA | 2 kb |
| CCTAAAACATTGATATCCATGATTATGCCTTTTTTTATAG acctca | 9 | ctcatt GTAAGTCAAATAGAAAATAAAGACCCTCAACTCCTATAAA | 3.5 kb |
| AATGTCTTGGTAACCTATTTTATCCCTGTTTGGGTAATAG agtaag | 10 | ctcaaa GTAAGTGCAAATCCAATCTCATGTATTACATCATTCTACA | 6 kb |
| TGTATGGTCATGCATATATCTAAGGTATATGATTTTTCAG gcagta | 11 | gatgag GTAAAATCTCCAAGAATATTTGCAACATTTACAGAAGAAA | 2.3 kb |
| AAATCTAGATGTGCACTAAGTTTGAACATCTTATGAACAG gtgaag | 12 | tgcaag GTATAATACATTGCACATGTCTCTCTGTGTATTCAAGCTT | 2 kb |
| ACAATGAATGTGCAGCTTTTTTTTTCCTCATATGTTGCAG caaaat | 13 | tagaac GTATGTACACCAAAAAGAGGTTCTCCTTCCATACCCCACA | 91 bp |
| CTGGAACTGCTATTAAATTACAGTTATTGTGTGTCATCAG gtagtc | 14 | agccag GTAACTCACTTCTCATGGATTTTGCTTAATAAAGTATGCA | 2 kb |
| AGGGAGCTTGCGTCATGAACATTATATTGATTTTATCCAG gtggtt | 15 | tctcag GTAATTCTTYCAGTCTGTGAGTATTTATTGAGTCCCTAAA | 7 kb |
| TTATTATTTTTATAACTATTATTATGCTTTTTTCTTCTAG gaactt | 16 | aaatag GTAAGTGTTTATGCATTATACATTTATGAATTACATATAA | 1 kb |
| TAATATGAACAAGTTTTTTCTTTTTTTCTCAAATGTTTAG ggtgac | 17 | attcag GTAAGATGCAGCGTACAGGTCATGTTCCAGGACCATCCCC | 1.5 kb |
| CTGTGTGTGTAATTCTGTTAATGTTGCTGTTGTTGTACAG gcaatt | 18 | | |

[a] Upper case letters represent the 40 nucleotides of the introns 5′ and 3′ to each exon. Lower case letters indicate the first or last six bases of each exon. All intron sizes are approximate unless indicated in base pairs.

Table 3: Allelic Variation of the CA Dinucleotide Repeat in Intron 10[a]

| $(CA)_n$, $n$ = | occurrence | frequency | $(CA)_n$, $n$ = | occurrence | frequency |
|---|---|---|---|---|---|
| 8 | 1 | 2.8 | 13 | 2 | 5.6 |
| 9 | 11 | 30.6 | 14 | 0 | |
| 10 | 2 | 5.6 | 15 | 7 | 19.4 |
| 11 | 9 | 25.0 | 16 | 0 | |
| 12 | 3 | 8.3 | 17 | 1 | 2.8 |

[a] A 350-bp region of intron 10 of the MTP gene encoding a CA dinucleotide microsatellite was PCR amplified and subcloned into pUC18 from 18 unrelated individuals. Eight variations to the structure $(CA)_4AA(CA)_3GA(CA)_4TA(CA)_nTACA$ were identified as indicated. Occurrence equals the overall number of alleles out of 36 total with a specific structure. Frequency represents the percent of total for each variation.

consensus sequences for transcription factor binding. Additional cis elements further upstream which are recognized by HNF-5, C/EBP, and CTF/NF-1 (Jones et al., 1987) were also found. The functional significance of these sites remains to be determined experimentally.

*Polymorphic Markers.* Restriction fragment length polymorphisms or polymorphic tandem repeats of simple sequences (Weber, 1990) linked to the MTP large subunit gene may be used to investigate an association between the gene for MTP and abnormalities in plasma lipid or lipoprotein metabolism. Preliminary restriction endonuclease analysis of genomic DNAs from four unrelated individuals with nine different restriction enzymes did not reveal any restriction fragment length polymorphisms. CA dinucleotide repeats that exhibit length polymorphism are distributed throughout the human genome. These repeats can be easily characterized using PCR, thus making them ideal genetic markers. To identify a CA dinucleotide repeat, a probe for a CA repeat was hybridized to MTP large subunit genomic clones. An imperfect CA repeat with the structure $(CA)_4AA(CA)_3GA(CA)_4TA(CA)_nTACA$ was identified in intron 10. Analysis of 36 alleles from 18 unrelated individuals revealed eight variants, where $n$ ranged from 8 to 17. Table 3 displays occurrence and frequencies for the different alleles sequenced. Sixteen of 18 individuals were heterozygous for this motif.

*Chromosomal Localization.* Fluorescent *in situ* hybridization analysis was used to localize the gene encoding the large subunit of MTP to chromosome 4 (see Figure 3). Twenty independent hybridizations with chromosome 4 localized the gene to a position 39% of the distance from the centromere to the telomere of chromosome arm 4q. This area corresponds to band 4q24. The only other known gene involved in lipid metabolism to be localized in the vicinity of MTP on chromosome 4 is the gene for the intestinal fatty acid binding protein which has been reported at band 4q28–q31 (Sweetser et al., 1987). The gene for PDI, the small subunit of the transfer protein complex, is located on chromosome 17 (Pajunen et al., 1988).

*Comparison of the Gene Structure of MTP to That of Other Proteins Involved in Plasma Lipoprotein Metabolism.* The size of the gene for the microsomal triglyceride transfer protein large subunit is formidable compared to previously characterized genes for apolipoproteins and enzymes involved in plasma lipoprotein metabolism. The exchangeable apolipoproteins AI, AII, CI, CII, CIII, and E share a common ancestral origin and have similar gene structures of four exons and three introns [reviewed by Li et al. (1988) except for the gene structure for apo CI which appears in Lauer et al. (1988)]. Apolipoprotein AIV has a similar gene structure except it

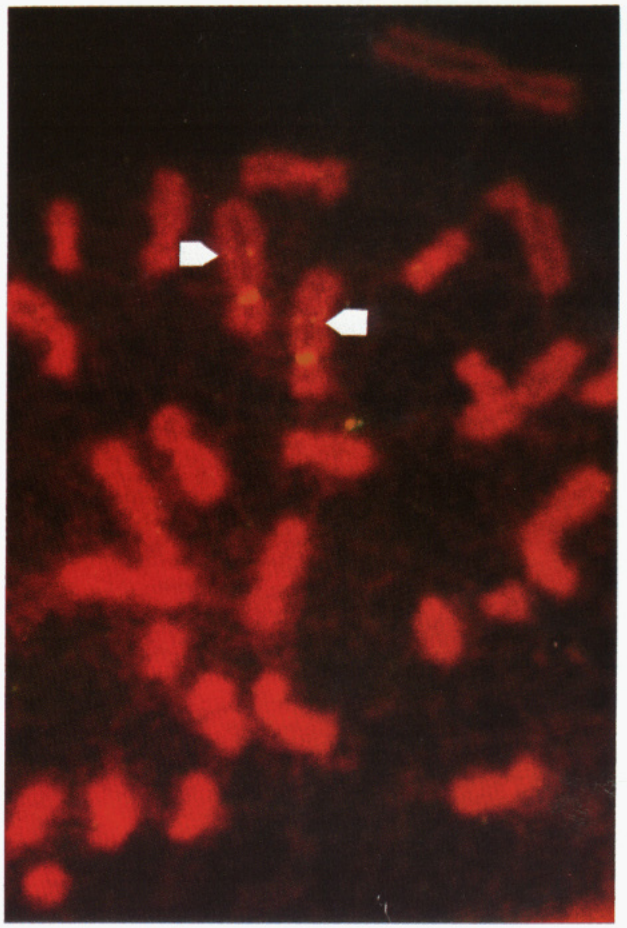

FIGURE 3: Chromosomal localization of the gene for the large subunit of MTP. Shown is an example of the fluorescent *in situ* hybridization performed on human metaphase chromosomes using a λ phage clone containing 20 kb of the MTP gene as probe. The arrows point to the area of hybridization of the MTP gene on chromosome 4.

only contains three exons and two introns. Human lipoprotein lipase (Kirchgessner et al., 1989), lecithin:cholesterol acyltransferase (LCAT) (McLean et al., 1986), cholesteryl ester transfer protein (CETP) (Agellon et al., 1990), low density lipoprotein receptor (Südhoff et al., 1985), and apoB (Blackhart et al., 1986) have 10, 6, 16, 18, and 29 exons, respectively, and except for the LCAT gene, which is about 4 kb, have genes which range in size from 25 to 45 kb. The MTP lipid transfer complex is unusual among proteins involved in lipoprotein metabolism in that the active protein is the product of two genes encoding PDI and the MTP large subunit which are located on different chromosomes.

The evolutionary origin of the transfer protein complex is of interest. By analogy to prolyl 4-hydroxylase, it is possible to speculate that the transfer protein originally consisted of the MTP large subunit which later evolved into its current heterodimeric form. Prolyl 4-hydroxylase is a tetrameric enzyme which contains two $\alpha$ subunits and two $\beta$ subunits which are PDI (Koivu et al., 1987). Immunological cross-reactivity studies of monomeric prolyl 4-hydroxylase from *Volvox carteri*, a multicellular green alga, and the tetrameric vertebrate enzyme suggest that the prolyl 4-hydroxylase tetramer evolved from an ancestral form of the enzyme which contained only the $\alpha$ peptide (Kaska et al., 1988).

On the basis of functional activity, the cholesteryl ester transfer protein is a likely candidate to have a relationship to the large subunit of MTP. Both proteins transport triglyceride, cholesteryl ester, and phospholipid between phospholipid

surfaces. However, sequence comparison between the MTP large subunit and CETP (Drayna et al., 1987), by a dot matrix comparison or exonic structure and sequence, provided no evidence of homology. In addition, the physical properties of the two proteins do not even appear similar. CETP is reported to be highly hydrophobic, with an overall index of hydropathy greater than +0.10. In contrast, the MTP large subunit has a negative index of hydropathy. Although CETP and MTP comprise a functional class of proteins, there appears to be little similarity between them. It is possible that the large subunit of MTP has evolved from a primordial protein not involved in lipid transport or lipoprotein metabolism. Identification of homologous genes from distantly related organisms may better our understanding of the evolution of the large subunit of MTP and its association with PDI to form the functional transfer complex.

## ACKNOWLEDGMENT

We thank Drs. Narayanan Hariharan, Todd Kirchgessner, and Richard Gregg for their critical reading of the manuscript.

## REFERENCES

Agellon, L. B., Quinet, E. M., Gillette, T. G., Drayna, D. T., Brown, M. L., & Tall, A. R. (1990) *Biochemistry 29*, 1372–1376.

Blackhart, B. D., Ludwig, E. M., Pierotti, V. R., Caiati, L., Onasch, M. A., Wallis, S. C., Powell, L., Pease, R., Knott, T. J., Chu, M.-L., Mahley, R. W., Scott, J., McCarthy, B. J., & Levy-Wilson, B. (1986) *J. Biol. Chem. 261*, 15364–15367.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., & Chambon, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4853–4857.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., & Chambon, P. (1980) *Science 209*, 1406–1414.

Costa, R. H., Grayson, D. R., Xanthopoulos, K. G., & Darnell, J. E., Jr. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3840–3844.

Drayna, D., Jarnagin, A. S., McLean, J., Henzel, W., Kohr, W., Fielding, C., & Lawn, R. (1987) *Nature 327*, 632–634.

Grange, T., Roux, J., Rigaud, G., & Pictet, R. (1991) *Nucleic Acids Res. 19*, 131–139.

Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J., & Tjian, R. (1987) *Cell 48*, 79–89.

Kadonaga, J. T., Carner, K. R., Masiarz, F. R., & Tjian, R. (1987) *Cell 51*, 1079–1090.

Kaling, M., Kugler, W., Ross, K., Zoidl, C., & Ryffel, G. U. (1991) *Mol. Cell Biol. 11*, 93–101.

Kane, J. P., & Havel, R. J. (1989) in *The Metabolic Basis of Inherited Disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S., & Valle, D., Eds.) pp 1139–1164, McGraw-Hill, Inc., New York.

Kaska, D. D., Myllylä, R., Günzler, V., Gibor, A., & Kivirikko, K. I. (1988) *Biochem. J. 256*, 257–263.

Kirchgessner, T. G., Chuat, J.-C., Heinzmann, C., Etienne, J., Guilhot, S., Svenson, K., Ameis, D., Pilon, C., D'Auriol, L., Andalibi, A., Schotz, M. C., Galibert, F., & Lusis, A. J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9647–9651.

Koivu, J., Myllylä, R., Helaakoski, T., Pihlajaniemi, T., Tasanen, K., & Kivirikko, K. I. (1987) *J. Biol. Chem. 262*, 6447–6449.

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

Lauer, S. J., Walker, D., Elshourbagy, N. A., Reardon, C. A., Levy-Wilson, B., & Taylor, J. M. (1988) *J. Biol. Chem. 263*, 7277–7286.

Lee, W., Mitchell, P., & Tjian, R. (1987) *Cell 49*, 741–752.

Li, W.-H., Tanimura, M., Luo, C.-C., Datta, S., & Chan, L. (1988) *J. Lipid Res. 29*, 245–271.

McLean, J., Wion, K., Drayna, D., Fielding, C., & Lawn, R. (1986) *Nucleic Acids Res. 23*, 9397–9406.

Pajunen, L., Myllyla, R., Helaakoski, T., Pihlajaniemi, T., Tasanen, K., Hoyhtya, M., Tryggvason, K., Solomon, E., & Kivirikko, K. I. (1988) *Cytogenet. Cell Genet. 47*, 37–41.

Sharp, D., Blinderman, L., Combs, K. A., Kienzle, B., Ricci, B., Wager-Smith, K., Gil, C. M., Turck, C. W., Bouma, M.-E., Rader, D. J., Aggerbeck, L. P., Gregg, R. E., Gordon, D. A., & Wetterau, J. R. (1993) *Nature 365*, 65–69.

Shoulders, C. C., Brett, D. J., Bayliss, J. D., Narcisi, T. M. E., Jarmuz, A., Grantham, T. T., Leoni, P. R. D., Bhattacharya, S., Pease, R. J., Cullen, P. M., Levi, S., Byfield, P. G. H., Purkiss, P., & Scott, J. (1993) *Hum. Mol. Genet. 2*, 2109–2116.

Südhof, T. C., Goldstein, J. L., Brown, M. S., & Russell, D. W. (1985) *Science 228*, 815–822.

Sweetser, D. A., Birkenmeier, E. H., Klisak, I. J., Zollman, S., Sparkes, R. S., Mohandas, T., Lusis, A. J., & Gordon, J. I. (1987) *J. Biol. Chem. 262*, 16060–16071.

Weber, J. L. (1990) *Genome Anal. 1*, 159–181.

Wetterau, J. R., & Zilversmit, D. B. (1984) *J. Biol. Chem. 259*, 10863–10866.

Wetterau, J. R., & Zilversmit, D. B. (1985) *Chem. Phys. Lipids 38*, 205–222.

Wetterau, J. R., & Zilversmit, D. B. (1986) *Biochim. Biophys. Acta 875*, 610–617.

Wetterau, J. R., Combs, K. A., Spinner, S. N., & Joiner, B. J. (1990) *J. Biol. Chem. 265*, 9800–9807.

Wetterau, J. R., Aggerbeck, L. P., Bouma, M.-E., Eisenberg, C., Munck, A., Hermier, M., Schmitz, J., Gay, G., Rader, D. J., & Gregg, R. E. (1992) *Science 258*, 999–1001.