The Cloning and Characterization of the Human Transcobalamin II Gene

By Annette Regec, Edward V. Quadros, Ovidiu Platica, and Sheldon P. Rothenberg

Transcobalamin II (TCII) is a plasma protein that binds vitamin B12 (cobalamin; Cbl) and facilitates the cellular uptake of the vitamin by receptor-mediated endocytosis. In genetic disorders that are characterized by congenital deficiency of TCII, intracellular Cbl deficiency occurs, resulting in an early onset of megaloblastic anemia that is sometimes accompanied by a neurologic disorder. To define the genetic basis for TCII deficiency, we have cloned and characterized the human gene that encodes this protein. The gene spans a minimum of 18 kb and contains nine exons and eight introns, with a polyadenylation signal sequence located 509 bp downstream from the termination codon and a transcription initiation site beginning 158 bp upstream from the ATG translation start site. The 5' flanking DNA does not have a TATA or CCAAT regulatory element, but a 34-nucleotide stretch beginning just upstream of the CAP site contains four tandemly organized 5'CCCC-3' tetramers. This sequence is a motif for a trans-acting transcription factor (ETF) that regulates expression of the epidermal growth factor receptor gene (EGFR), which also lacks TATA and CCAAT regulatory elements. A GC-rich sequence that binds the SP1 protein is located 356 nucleotides upstream from the first of the series of CCCC tetramers. Although this GC sequence is at an unusual location with respect to the CAP site, a 507-bp fragment containing this GC box drives the chloramphenicol acetyltransferase (CAT) reporter gene after transient transfection into NIH 3T3 cells. No CAT activity was observed when a 420-bp fragment lacking this GC box but containing the ETF-binding domains was similarly transfected into this cell line. One consensus and two atypical motifs for the c-myc ligand are located downstream and upstream, respectively, of the GC box, and this could explain the elevated plasma TCII observed in some patients with multiple myeloma, as the c-myc product is overexpressed in some myeloma cells. Restriction endonuclease digestion of genomic DNA from eight normal subjects with Taq I, Hinfl, Msp I, and Bgl I identified three patterns of restriction fragment length polymorphism (RFLP). A number of the exon/intron splice junctions of human TCII, TCI, and IF genes are located in homologous regions of these proteins, providing evidence that these genes have evolved by duplication of an ancestral gene. This characterization of the TCII gene and the RFLP should facilitate the identification of the mutation(s) responsible for the genetic abnormalities of TCII expression.

© 1995 by The American Society of Hematology.

MATERIALS AND METHODS

The enzymes used for cloning and restriction digestion of DNA were obtained from New England Biolabs (Beverly, MA) and Bethesda Research Laboratories (Bethesda, MD). Oligonucleotides, 17 to 24 nucleotides (nt) in length, were used as probes, as primers for DNA sequencing, and for amplification of DNA by the polymerase chain reaction (PCR). These primers were selected using the Primer Detective, version 1.0 software (Clontech Labs, Palo Alto, CA) and were synthesized in a Gene Assembler Plus synthesizer (Pharmacia Biotech Inc, Piscataway, NJ). Table 1 lists the location of the oligomers that were used for PCR and for Southern blotting of the genomic fragments.

Isolation of genomic clones. A human genomic DNA library prepared in λ Charon 4A phage was obtained from American Type Culture Collection (ATCC, Bethesda, MD). The library was plated at a density of approximately 50,000 plaque-forming units (pfu) and screened with the TCII cDNA labeled with [32P]deoxyadenosine triphosphate (dCTP) using a random primer labeling kit. A total of 2 × 106 pfu were screened, and 12 hybridizing clones were selected and plaque-purified. Three of these clones were identified as distinct from each other.

From The Division of Hematology/Oncology, State University of New York (SUNY)-Health Science Center, Brooklyn; and the Department of Veterans Affairs Medical Center, Brooklyn, NY.

Submitted July 26, 1994; accepted December 20, 1994.

Supported by Grants No. RO1 DK28561 from the National Institutes of Health and a Merit grant from the Medical Research Service of the Department of Veterans Affairs.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. section 1734 solely to indicate this fact.

© 1995 by The American Society of Hematology. 9006-4971/95/8510-00983.00/0


2711
This strategy identified most of the exodintron junctions. In addition, provided the sequence for approximately 1 kbp of the S' flanking to facilitate sequencing the 3' and 5' regions of the gene which, in Thermocycler (Nonvalk, CT) in a reaction volume of 100 µL containing 20 pmol of each sense and antisense primer, 50 pmol of each deoxyribonucleotide, 0.1 to 1 µg of parent plasmid or plasmid DNA as template, and 1 to 5 U of Taq DNA polymerase. Amplification was performed for 30 to 35 cycles of 3 minutes at 95°C, 3 minutes at 55°C, and 5 minutes at 72°C using primers complementary to the exon regions of the TCII gene. The identity of each PCR product was confirmed by Southern blot analysis using an oligonucleotide probe complementary to a region within the amplified fragment.

An additional strategy was used to confirm the size of the introns that were generated by PCR amplification and for one intron that could not be amplified by PCR. For this analysis, each clone was digested with a restriction enzyme selected to cut within a known intron sequence flanking the exon-intron junctions. The size of each intron was determined by comparing the sizes of the fragments generated by PCR using a sense primer complementary to this dC tail and an antisense orientation was combined with a control plasmid containing this GC box was obtained by digestion of GC2 (the ApaI-derived subclone) with Sma I and Sac I to release the insert. A 420-bp fragment, extending from 5' (−328) + 3' (−64) and that lacked this GC rich sequence, was prepared by PCR. Each fragment was subcloned into a plasmid containing the chloramphenicol acetyltransferase (CAT) reporter gene (pCAT Basic Plasmid: Promega, Madison, WI).

The pCAT vector containing the fragment in the sense or the antisense orientation was used for the pCAT control plasmid) was used to transfect the 3T3 cells as a method of Gorman et al using [14C]chloramphenicol, and the products of the reaction were analyzed by thin layer chromatography (TLC) on silica plates (Sigma, St Louis, MO). The volume of the assay mixture applied to the TLC plate was normalized to contain the same β-galactosidase activity to correct for variations in transcription efficiency. The acetylated products (mono- and diacetylated forms) and the nonacetylated substrate were identified by autoradiography, and the corresponding areas were scraped off the TLC plates and counted in a Packard Tri-Carb model 1900CA liquid scintillation analyzer (Meriden, CT).

Mobility-shift assays for DNA binding proteins. The binding of nuclear proteins to selected DNA motifs in the 5' flanking region of the gene was determined by the mobility-shift assay. A 507-bp fragment, extending from 5' (−834) → 3' (−326) and containing

Other by the restriction pattern obtained by digestion with ApaI and DraI and are identified as GC1, GC2, and GC3. The digested clones were separated in a 1% agarose gel and analyzed by Southern blotting11 using oligonucleotide probes complementary to the 3', middle, and 5' regions of the TCII cDNA to identify the coding and flanking regions of the gene.

The three genomic clones were digested with EcoRI and each insert, approximately 15 kb in size, was subcloned into pGEMZf(−) and amplified. These subcloned EcoRI fragments were sequenced using nine pairs of primers and 10 additional single primers encompassing the entire TCII cDNA, as previously described.12 This strategy identified most of the exon/intron junctions. In addition, to facilitate sequencing the 3' and 5' regions of the gene which, in these EcoRI subclones, spanned more than 2 kbp of flanking DNA, these subclones were further digested separately with Pst I and Sac I (GC1), Apa I and Dra I (GC2), and Apa I (GC3). The fragments identified by Southern blot analysis with oligonucleotide probes to these flanking regions were subcloned and sequenced. This strategy provided the sequence for approximately 1 kbp of the 5' flanking DNA (three subclones from GC2 and GC3) and approximately 0.6 kbp of the 3' flanking DNA (one subclone from GC1).

Mapping of the TCII gene. The pGEM subclones were sequenced by the method of Sanger et al14 using universal forward and reverse primers and the Sequenase-2 Kit (USB). The 5' region of the gene and intron 5 were completely sequenced using internal primers generated from the nucleotide sequence flanking the exon-intron junctions. The size of each intron was determined by amplifying the region between two adjacent exons using PCR. Amplification of the genomic fragments was performed in a Perkin-Elmer Thermocycler (Norwalk, CT) in a reaction volume of 100 µL containing 20 pmol of each sense and antisense primer, 50 pmol of each deoxyribonucleotide, 0.1 to 1 µg of parent plasmid or plasmid DNA as template, and 1 to 5 U of Taq DNA polymerase. Amplification was performed for 30 to 35 cycles of 3 minutes at 95°C, 3 minutes at 55°C, and 5 minutes at 72°C using primers complementary to the exon regions of the TCII gene. The identity of each PCR product was confirmed by Southern blot analysis using an oligonucleotide probe complementary to a region within the amplified fragment.

An additional strategy was used to confirm the size of the introns that were generated by PCR amplification and for one intron that could not be amplified by PCR. For this analysis, each clone was digested with a restriction enzyme selected to cut within a known sequence of an exon close to an intron and in a known downstream sequence, either in the same intron or the adjacent exon. A pair of oligonucleotide probes to internal sequences close to the restriction site were used for Southern blot analysis. If each set of probes hybridized to the same band on the Southern blot, it indicated that an entire intron must be located between these probes. The size of the intron was then estimated as the size of the fragment plus any known intron sequence flanking the restriction sites. If the two probes hybridized to two fragments, a minimum size was assigned to the intron that was computed as the sum of the size of the two fragments. We cannot exclude the possibility that additional restriction sites are present in the intron between these probes, and the estimated size of the intron, therefore, is a minimum value. Only the size of one intron (intron 8) was determined exclusively by this method.

Mapping the transcription start site. The 5' RACE assay5 (GIBCO-BRL, Gaithersburg, MD) was used to determine the transcription start site and the full length of the 5' region of the mRNA. For this analysis, mRNA was prepared from human umbilical vein endothelial (HUVE) cells using the BioMag Kit (Advanced Magnetix Inc, Cambridge, MA). First-strand synthesis was performed using 1 µg of this mRNA and an antisense oligonucleotide complementary to 5' (397) → 3' (375) of the TCII cDNA. A dC tail was attached to the 3' end of the first strand cDNA copy, which was then amplified by PCR using a sense primer complementary to this dC tail and an antisense primer complementary to the sequence 5' (361) → 3' (340) of the TCII cDNA. The PCR fragment was separated in a 1% agarose gel and subcloned into SK- (Stratagene, CA).

Six clones were isolated, and the insert size in each was determined by electrophoresis in a 1% agarose gel after digestion with Sac I plus Sac I. All six clones were sequenced11 to locate the transcription start site.

Characterization of the putative promoter elements. The nucleotide sequence of the 5' flanking region was analyzed using the Gene Works program (IntelliGenetics Inc, Mountain View, CA) to identify consensus promoter sequences, and a GC box promoter element (5'-GGGAGGCGGTCGTG-3') extending from 5' (−565) → 3' (−552) was identified. A 507-bp restriction fragment 5' (−834) → 3' (−328) containing this GC box was obtained by digestion of GC2 (the ApaI-derived subclone) with Sma I and Sac I to release the insert. A 420-bp fragment, extending from 5' (−328) → 3' (−64) and that lacked this GC rich sequence, was prepared by PCR. Each fragment was subcloned into a plasmid containing the chloramphenicol acetyltransferase (CAT) reporter gene (pCAT Basic Plasmid: Promega, Madison, WI).

The pCAT vector containing the fragment in the sense or the antisense orientation was used for the pCAT control plasmid) was used to transfect the 3T3 cells as a method of Gorman et al using [14C]chloramphenicol, and the products of the reaction were analyzed by thin layer chromatography (TLC) on silica plates (Sigma, St Louis, MO). The volume of the assay mixture applied to the TLC plate was normalized to contain the same β-galactosidase activity to correct for variations in transcription efficiency. The acetylated products (mono- and diacetylated forms) and the nonacetylated substrate were identified by autoradiography, and the corresponding areas were scraped off the TLC plates and counted in a Packard Tri-Carb model 1900CA liquid scintillation analyzer (Meriden, CT).
THE HUMAN TRANSCOBALAMIN II GENE

the putative GC box, was labeled with \( ^{32}P \)dCTP by filling in the 5' overhang with Klenow DNA polymerase, and then incubated with nuclear extract prepared from the 3T3 fibroblasts or with purified SP1 protein (Promega). To establish the specificity of the mobility-shift assays, the nuclear extract or the purified human SP1 protein was preincubated with a 10-fold molar excess of unlabeled 507-bp fragment. The mobility-shift assay was also performed with a \( ^{32}P \)-labeled synthetic 34-bp GC box fragment, and the specificity of the reaction was established by preincubation of the probe with a 10-fold molar excess of the unlabeled 34-bp fragment or the SP1-specific oligonucleotide sequence.

**Restriction fragment length polymorphism (RFLP) of the TCII gene.** Blood from normal donors was collected in citrate. After centrifugation, the plasma was aspirated, and the buffy coat overlaying the red cells was removed and used to prepare genomic DNA by a standard method. Samples of the DNA (10 µg) were digested to completion with the following restriction endonucleases: BamHI, Bgl I, Bgl II, Eco RI, Hae III, Hha I, HindIII, Hinfl, Kpn I, Msp I, Pst I, Sac I, Sma I, Tag I, and Xba I. After digestion, the DNA fragments were separated by electrophoresis in a 1% agarose gel and analyzed by Southern blotting using \( ^{32}P \)-labeled TCII cDNA as the probe.

**RESULTS**

The alignment of the three clones (GC1, GC2, and GC3) that encompass the TCII gene is shown in Fig 1A. GC1 contains the 3' flanking DNA and extends upstream to include most of the coding region. GC2 and GC3 contain part of the coding region and the 5' flanking DNA, with GC3 containing an additional 4 kb of upstream DNA. The gene spans approximately 18 kb and contains, in addition to the coding region, the 3' untranslated and flanking DNA, and the 5' flanking DNA that contains regulatory elements (vide infra). This size of the TCII gene is a minimum value, as the 5' and 3' extremities of GC1 and GC3 may contain additional regulatory components of the gene that have not been defined. The organization of the TCII gene is shown in Fig 1B. Within the defined 18-kbp span of the gene, there are nine exons that range in size from 117 bp to 581 bp. There are eight introns that range in size from 126 bp to approximated minimal values of 650 to 5,200 bp. Table 2 lists the sizes of introns, which were estimated by restriction endonuclease digestion of the genomic clones (GC1, GC2, and GC3). The restriction fragment(s) containing the indicated intron was identified by Southern analysis using probes to internal sequences in close proximity and between the restriction sites, as described in Materials and Methods.

Figure 2 shows the nucleotide sequence of the coding region, the region 850 bp upstream of the ATG start codon and the region 530 bp downstream from the stop codon. The size of introns 1, 2, 3, 4, 6, 7, and 8 are given as approximate values as they were not fully sequenced. Intron 5, however, was completed sequenced and is 126 bp in size. A polyade-

![Diagram](image-url)

**Table 2. Analysis of the Size of the Introns by Digestion of the Genomic Clones With Restriction Enzymes**

<table>
<thead>
<tr>
<th>Intron</th>
<th>Genomic Clone(s)</th>
<th>Restriction Enzyme(s)**</th>
<th>No. of Hybridizing Fragments†</th>
<th>Intron Size (bp)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3</td>
<td>Apa I + Cf × 10 I</td>
<td>2</td>
<td>3,620</td>
<td></td>
</tr>
<tr>
<td>2, 3</td>
<td>Fok I</td>
<td>1</td>
<td>1,745</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Hpa II</td>
<td>2</td>
<td>1,450</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Hpa II</td>
<td>2</td>
<td>1,895</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Stu I</td>
<td>2</td>
<td>3,700</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Ava II</td>
<td>2</td>
<td>5,200</td>
<td></td>
</tr>
</tbody>
</table>

* The restriction endonucleases were chosen to cut in the exons that were identified by sequencing the subcloned fragments of the genomic clones.
† The restriction fragment(s) containing the intron was identified by Southern analysis using oligonucleotide probes to specific nucleotide sequences in close proximity and between the restriction sites.
‡ The size of the intron was estimated as described in Materials and Methods.
The nucleotide sequence of the coding regions and the proximal 5' and 3' untranslated and flanking DNA of the TCI gene. Intron 5 has been fully sequenced. The size of the other introns have been determined by PCR amplification and restriction digestion as described in Materials and Methods. A GC-rich sequence in the 5' region is shown in the open box. The arrowhead indicates the transcription initiation site. The polyadenylation signal sequence in the 3' untranslated region is indicated by the shaded box. The double underlined sequences are motifs for the c-myc ligand. The single underlined sequences are motifs for the ets-encoded nuclear binding proteins. The dashed-underlined CCC tetramers and the dashed-overlined GGGGG tetramer are motifs for the ETF transcription factor.

Fig 2. The nucleotide sequence of the coding regions and the proximal 5' and 3' untranslated and flanking DNA of the TCI gene. Intron 5 has been fully sequenced. The size of the other introns have been determined by PCR amplification and restriction digestion as described in Materials and Methods. A GC-rich sequence in the 5' region is shown in the open box. The arrowhead indicates the transcription initiation site. The polyadenylation signal sequence in the 3' untranslated region is indicated by the shaded box. The double underlined sequences and the single underlined sequence are motifs for the c-myc ligand. The single underlined sequences are the motifs for the ets-encoded nuclear binding proteins. The dashed-underlined CCC tetramers and the dashed-overlined GGGGG tetramer are motifs for the ETF transcription factor.
Fig 3. The Southern blot of the fragment obtained by the 5'-RACE assay. Lane 1, the 500-bp fragment generated from HUVE cell mRNA; lane 2, control amplification using the TCII cDNA with the same anti-sense amplimer. This fragment is approximately 100 bp smaller than the 5'-RACE product, because the sense amplimer was located 5' (-34) to 3' (-11) downstream from the transcription start site.

A 420-bp fragment extending from 5' (-328) to 3' (+64) and lacking the GC box but containing the tandem CCCC tetramer repeats, all of exon 1, and 28 nt of intron 1, did not drive expression of the CAT gene (lanes 4 and 5).

Mobility-shift assay to identify the GC box. Figure 5 shows the autoradiograph of the mobility-shift assay. The radiolabeled 507-bp fragment (lane 1) containing the putative GC box shifted close to the top of the gel when incubated with purified human SPI protein (lane 2). A substantial component of this shift was blocked when the SPI protein was preincubated with the unlabeled 507-bp fragment (lane 3), with the 34-bp synthetic fragment containing the GC box (lane 4), or with the SPI-specific oligonucleotide (lane 5).

A mobility-shift of this radiolabeled fragment was also observed with the nuclear extract prepared from mouse NIH 3T3 cells (lane 6). Preincubating the nuclear extract with poly dI-dC followed by the addition of the labeled fragment did not alter this shift (lane 7), but preincubation of the nuclear extract with the unlabeled fragment blocked the shift of a fraction of the labeled probe (lane 8). Preincubating the nuclear extract with poly dI-dC and the unlabeled synthetic 34-bp fragment containing the putative GC box blocked the shift of a greater fraction of the labeled probe (lane 9), indicating that the nuclear extract contains a factor, which is likely to be SPI, that binds to this GC-rich motif in the 5' flanking DNA of the TCII gene.

Figure 6 shows the RFLP observed after digestion of human genomic DNA with four of the 15 different restriction endonucleases studied. The most distinctive RFLP was obtained by digestion with Taq I, HindIII, Msp I (Fig 6A), and Bgl I (Fig 6B), which resulted in either a gain or loss of fragment(s). Three distinctive RFLP types were observed and are summarized in Table 4. Whereas HindIII, Msp I, and Taq I identify RFLP type III, they do not distinguish RFLP
types I and II (each enzyme generates the same number and size of restriction fragments). Bgl I alone, however, identifies all three types.

The DNA prepared from fibroblasts cultured from three patients who lacked expression of a functional and immuno-reactive TCII was also analyzed for some distinctive differences in RFLP using 12 restriction endonucleases, including the four above, and a variance from any normal pattern was not identified (data not shown).

DISCUSSION

This report provides the characterization of the gene that encodes TCII, the Cbl-binding plasma protein that is essential for the cellular uptake of this vitamin. TCII is one of three structurally distinct Cbl-binding proteins, with the other two being intrinsic factor (IF) and transcobalamin I (TCI). Although these proteins are immunologically distinct, a comparison of the nucleotide sequence and the deduced amino acid sequence of the cDNA(s) encoding the rat IF22 and human TCI23 with human TCII cDNA12 shows greater than 50% nucleotide homology and 20% amino acid homology,12 with seven stretches of four or more amino acids common to the three proteins.

The human TCII gene lacks the CCAAT and TATA box promoter elements, and the 5'-RACE assay located a putative transcription start site 158 bp upstream from translation initiation. A TATA-less promoter region with or without multiple transcription sites is more typical of a housekeeping gene that encodes a protein that is constitutively expressed.24,25 This could explain the observation by Li et al26 of low-level expression of TCII mRNA in several different tissues. However, that study did not identify the cellular components of the tissue that express the message. This is an important consideration, because endothelial cells in culture and in situ synthesize TCII,27 and these cells may contribute a substantial fraction of RNA prepared from most tissues.

Although the 5' flanking region of the gene lacks the TATA and CCAAT sequences, there are a number of motifs that could regulate transcription of the TCII gene. The most discernable region lies between 5' (-175) -> 3' (-162) and contains three tandemly organized CCCC tetramers (dashed underline in Fig 2) separated by a G/A purine, another CCCC tetramer 17 nt upstream, and a GGGGGG hexamer (dashed overline in Fig 2) located 196 nt further upstream. These sequences, including the GC box, are motifs for the trans- active transcription factor (ETF) that regulates expression of the epidermal growth factor receptor (EGFR) gene.28 The ETF transcription factor is purported to regulate transcription in genes, like the TCII gene, that lack TATA and CCAAT regulatory elements.28 In our studies, a fragment of the 5' flanking DNA containing the GC box but lacking the downstream ETF motifs could drive expression of the CAT reporter gene after transient transfection into 3T3 cells, whereas a fragment containing the ETF motifs but lacking the GC box could not. This observation does not exclude a regulatory function of the ETF motifs for the TCII gene, because 3T3 cells may not express the ETF protein, whereas they do express the ubiquitous SPI transcription factor. The precise role of the ETF motifs in regulating expression of the TCII gene is now under investigation in our laboratory using cell lines that do and do not express the ETF protein. These studies may provide information to clarify why expression of TCII may be constitutive (ie, housekeeping) as well as inducible in response to rapid changes in plasma TCII.29

We are continuing the analysis of the 5' flanking DNA of the TCII gene for other transcription regulatory elements, because this region contains nucleotide sequences that may serve as binding motifs for transacting oncogene products. In the region between 5' (-525) -> 3' (-520), there is a consensus binding motif (CACGTG, single underline in Fig 2) for the c-myc protein,30 and there are two atypical motifs (double underline in Fig 2) in the region 5' (-811) to 3' (-843) that may also serve as DNA binding sites for this oncogene product. There are also three GGAA tetramers (5' (-402) -> 3' (-405), 5' (-309) -> 3' (-312), and 5' (-109) -> 3' (-112)) that have been identified as a motif to which products of the ets gene family bind.31
The identification of these motifs in the TCII gene may explain the elevation of TCII observed in a number of malignant disorders, especially multiple myeloma and lymphoproliferative diseases. In multiple myeloma, a translocation of chromosome 8q24, the locus of the c-myc protooncogene, to chromosome 14q32 has been observed, and an elevation of c-myc mRNA in the myeloma cells of some patients has been reported. An increase in the steady state concentration of any plasma protein may be initiated by several factors: i.e., a decrease in the plasma clearance rate, a decrease in the turnover of the mRNA by cytoplasmic stabilizing factor(s), or an increase in the rate of transcription. Thus, the elevated plasma level of TCII in some patients with multiple myeloma could be due to upregulation of transcription by the c-myc ligand, a nuclear protein that is believed to regulate transcription and gene expression. This activation of the TCII gene by a trans-activating transcription factor in cells that do not normally express TCII may be likened to the paraneoplastic syndromes in which neoplastic cells express hormones or other proteins not synthesized by the corresponding non-neoplastic cells. This could also explain the elevated plasma TCII observed in patients with breast cancer and histiocytic proliferation.

The similar Cbl binding properties of TCII, TCI, and IF indicate that the genes encoding these proteins are likely to have evolved from a common ancestral gene. Hewitt et al. have found 36% amino acid identity of human IF and human haptocorrin as deduced from their cDNAs, and our previous study also showed that human TCII contains several stretches of amino acid sequences that are homologous with human TCI and rat IF. A similar comparative analysis of the amino acid sequence homology of these three protein and porcine haptocorrin has identified four regions with 80% homology and two regions with 60% homology.
ston et al. have also identified very similar intron/exon junctions with a number of conserved positional splice sites for human TCII and human IF.

Figure 7 shows a comparative analysis of the placement of introns in the coding regions of human TCII, TCI, and IF genes that indicates substantial homology of the amino acids flanking these insertions and supports the notion that these genes are derived by duplication of an ancestral gene. Despite this homology, a number of important differences in the TCII and TCI genes indicate that a major event(s) occurred during this evolutional process that randomly translocated the TCII gene to chromosome 22 from chromosome 11, where the human TCI and IF genes are located. In addition, the TCII gene lacks both the TATA and CCAAT proximal promoter sequences that are contained in the human TCI and IF genes, indicating that the translocation of the TCII gene did not include the 5′ flanking promoter region. This could account, in part, for the differences in the tissue-specific expression of these proteins.

An interesting feature of the TCII gene is the relatively frequent occurrence of RFLP. In this small study of DNA polymorphism of the normal TCII gene, three different restriction patterns were generated by digestion of normal genomic DNA from seven individuals with the enzyme Bgl I. Similarly, the enzymes Hinf I, Msp I, and Taq I could each generate two different restriction patterns by digestion of normal genomic DNA from four individuals. The purpose for analyzing this RFLP of the TCII gene was to identify a unique restriction pattern in the genomic DNA from patients with congenital deficiency of functional and immunoreactive TCII that would indicate a major deletion of a part of the TCII gene. However, similar to the findings of Li et al., the RFLP of the TCII gene in three patients with this disorder did not differ from normal using 12 different restriction enzymes. (The fibroblast cell lines from these patients were provided by the late Dr Charles Hall as a collaborative project that is in progress to define the genetic basis for TCII deficiency.) This study is continuing with additional restriction enzymes as single- or even multiple-base mutations rather than a segmental deletion of the gene may be the basis of this disorder. This could be identified by the loss of a selective restriction site that would alter the RFLP. Such a finding would permit genetic linkage studies and kindred analysis for this abnormal TCII gene.

REFERENCES

9. Seligman PA, LaDonna L, Steiner LL, Allen RH, Lazerson J:

![Table 4. RFLP of the TCII Gene](image-url)

<table>
<thead>
<tr>
<th>RFLP Type</th>
<th>Restriction Endonuclease (no. of DNA samples)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bgl I (n = 7)</td>
</tr>
<tr>
<td><strong>I</strong></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6 (−3.3 kbp)</td>
</tr>
<tr>
<td></td>
<td>(−3.8 kbp)</td>
</tr>
<tr>
<td><strong>II</strong></td>
<td>7 (−3.3 kbp)</td>
</tr>
<tr>
<td></td>
<td>(−3.8 kbp)</td>
</tr>
<tr>
<td></td>
<td>(−2.4 kbp)</td>
</tr>
</tbody>
</table>

Data are numbers of autoradiographic bands. The negative sign in parentheses indicates the absence of a fragment of that size; the plus sign indicates the size of the additional fragment(s).

* The number of genomic DNA samples from normal subjects digested with each restriction endonuclease.

† This polymorphism was identified in subject 7 (shown in Fig 7A and B).

‡ This polymorphism was identified in subject 8 (shown in Fig 7A and B) and subject 2 (shown in Fig 7B).

§ This polymorphism was identified in subject 3 (shown in Fig 7A and B) and subjects 1, 4, 5, and 6 (shown in Fig 7B).


The cloning and characterization of the human transcobalamin II gene
A Regec, EV Quadros, O Platica and SP Rothenberg