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 kbp 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 tandemly organized 5'-CCCC-3' tetramers. This sequence is responsible for these congenital abnormalities of TCII, and there was no evidence for the presence of TATA boxes, CCAAT boxes, or GC boxes in the 5' flanking region. To assist in the identification of the genetic mutations that are responsible for these congenital abnormalities of TCII, we have cloned and characterized the gene encoding this protein.

The cellular uptake of vitamin B12 (cobalamin; Cbl) is mediated by a plasma protein, transcobalamin II (TCII). The TCII-Cbl complex binds to a specific receptor on the plasma membrane and is then internalized by a process of endocytosis. In genetic disorders that result in congenital TCII deficiency, intracellular Cbl deficiency rapidly occurs, and this impairs the methylation of homocysteine in the de novo synthesis of methionine and the rearrangement of methylmalonyl CoA to succinyl CoA, the two essential metabolic pathways that require Cbl cofactors. The clinical sequelae of TCII deficiency are megaloblastic anemia, which usually occurs in early infancy, and at times an accompanying neurologic disorder. Hakami et al. were the first to identify TCII deficiency as the cause of megaloblastic anemia in two infants at 3 and 5 weeks of age. Other reports of congenital TCII deficiency followed, and in one case reported by Carmel and Ravindranath, the presentation was atypical as the serum Cbl concentration was low (it is usually normal in congenital TCII deficiency) and there was early neurologic dysfunction.

Most patients with congenital TCII deficiency lack both a functional and immunoreactive TCII. However, Haurani et al. described the occurrence of Cbl deficiency due to an abnormal TCII protein that bound Cbl normally but did not facilitate the cellular uptake of the vitamin. Selbagian et al. have reported studies of a patient with megaloblastic anemia due to an abnormal TCII that did not bind Cbl but was immunoreactive and could bind to the TCII receptor. Another clinical variant of TCII deficiency has been described in a patient with pancytopenia and a hypoproliferative bone marrow who lacked a functional form of TCII. To assist in the identification of the genetic mutations that are responsible for these congenital abnormalities of TCII, we have cloned and characterized the gene encoding this protein.

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]-deoxycytidine triphosphate (dCTP) using a random primer labeling kit [United States Biochemicals (USB), Cleveland, OH]. A total of 2 × 105 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. RO1DK28561 from the National Institutes of Health and a Merit grant from the Medical Research Service of the Department of Veterans Affairs.
Address reprint requests to Sheldon P. Rothenberg, MD, SUNY-Health Science Center, Box 20, 450 Clarkson Ave, Brooklyn, NY 11203.

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.
9006-4971/95/8510-0093$3.00/0

From www.bloodjournal.org by guest on October 27, 2017. For personal use only.
This strategy identified most of the exodintron junctions. In addition, provided the sequence for approximately 1 kbp of the S' flanking DraI to facilitate sequencing the 3' and 5' regions of the gene which, in these subclones were further digested separately with 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 pg of parent phage 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 cDNA. The identity of each PCR product could not be amplified by ER. For this analysis, each clone was digested with EcoRI and each fragment was subcloned into the coding strand of the gene (see Fig 2).

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 blotting using oligonucleotide probes complementary to the 3', middle, and 5' regions of the TCIII cDNA to identify the coding and flanking regions of the gene.

The three genomic clones were digested with EcoRI and each insert, approximately 15 kbp in size, was subcloned into pGEM9Zf(−) and amplified. These subcloned EcoRI fragments were sequenced using nine pairs of primers and pGEM9Zf(-) and amplified. These subcloned EcoRI fragments were sequenced using nine pairs of primers and the Sequenase-2 Kit (USB). The S' region was sequenced using nine pairs of primers and reverse primers and the Sequenase-2 Kit (USB). The S' region of the gene and intron S were completely sequenced using internal primers encompassing the entire TCII cDNA, as previously described.I2 The PCR fragment was separated in a 1% agarose gel and subcloned into SCR-Script (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 sequencedI3 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 (−328)[GGGAGGCGGTGCTG−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 Smal 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 combined with a control plasmid containing the β-galactosidase gene and cotransfected into NIH 3T3 fibroblasts (ATCC) by calcium phosphate precipitation of the DNA using a standard protocol.I4 The cells were harvested by trypsination at 24 and 40 hours after transfection and lysed by freezing and thawing, and the soluble product was recovered after centrifugation at 15,000g for 10 minutes. A plasmid containing the SV40 promoter (pCAT control plasmid) was used to transfect the 3T3 cells as a positive promoter control. Activity of β-galactosidase was measured by the method of Gorman et al.I5 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 transfection 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.I6 A 507-bp fragment, extending from 5′ (−834) → 3′ (−328) and containing

<table>
<thead>
<tr>
<th>Table 1. Primers Used for the Analysis of the Size of the Introns by PCR Amplification of the Genomic Clones</th>
<th>Primers*</th>
<th>Sense (5′ → 3′)</th>
<th>Antisense (5′ → 3′)</th>
<th>Product Size (bp)</th>
<th>Intron Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intron</td>
<td>Sense</td>
<td>Antisense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>23 → 39</td>
<td>105 → 88</td>
<td>3,750</td>
<td>3,670</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>196 → 216</td>
<td>361 → 340</td>
<td>2,000</td>
<td>1,840</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>383 → 406</td>
<td>554 → 537</td>
<td>1,800</td>
<td>1,630</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>450 → 468</td>
<td>753 → 743</td>
<td>950</td>
<td>650</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>850 → 890</td>
<td>1052 → 1030</td>
<td>1,900</td>
<td>1,610</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1061 → 1083</td>
<td>1154 → 1153</td>
<td>1,900</td>
<td>1,810</td>
<td></td>
</tr>
</tbody>
</table>

* The nucleotide numbers indicate the position of the primer on the coding strand of the gene (see Fig 2).
† See Materials and Methods for computing the intron size from the size of the PCR product.
the putative GC box, was labeled with [32P]dCTP by filling in the 5' overhang with Klenow DNA polymerase, and then incubated with the 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 32P-labeled synthetic 34-bp GC box fragment, and the specificity of the reaction was established by preincubation of the probe with a 50-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, Hinfl, HindIII, 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 the 32P-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 kbp of upstream DNA. The gene spans approximately 18 kbp 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 completely sequenced and is 126 bp in size. A polyade-
nlation signal sequence (boxed and shaded) is located 510 bp downstream from the stop codon, and a 14-nt, GC-rich sequence (boxed) is located 552 bp upstream from the ATG site. Similar substitutions of a pyrimidine for a purine in this cation and/or restriction mapping the other introns downstream from the stop codon, and a 14-nt, GC-rich purine as the second upstream base flanking the donor splice AGTGCC~;CCTCTGAC~CA~ATGGCCTGTGTGGACCTGC

DVLKKAHELGGFT
CCAAGCATCTTCCCTW~GAAG~~TTTCTGGCCRM~TCTGGCCAGCCT~~~CCCTGC~TCTCCCATG~CCACCCCA~TC~ATG~~~C~G~A 1575

the introd exon junctions. An unusual feature of the donor assays using mRNA prepared from

GGGGCTATGGCCCTGACCCCCTCTCCAC~TCTTffiffi~AGCTCTGffi~GTTGTG~~GTAGCTGGGGffiAC~ffiCA~CTGCTC 1775

TGCTGCC~CTCCTGTGCPAGCAATGCCCCCTGGGATCRCCCC~CACAA~CCTTCG~~~GGCCCTAT~CATGGCCCACC~GGAGCAGAG~~~ 1475

AGCTGGTffiCCCCTGAGCTCCCCCTCATCCCffi~GCCTCGCACA~CCTAGGCTTCTACCCTCCCTCCTCATGCCTCCCTGGAACAGG~~CGCCTGACCC

TfAGCCCCCTAC~~~CGTCA~~ffiCCGGffi~ffiffi~CTGCCA~TTCTCCGffiACCCC~AC~CACTGTTGCARGgtg~~t
GATGTCCTCARGAIGOCCCA~ffiTT~A~Agtaaggt

TTCATCTGAWTTCCCAGACl'GTCTGGCACCACGAgtagcccaactttttgtgga

TTCAACCCTGGTCCCACACRRCGGATCRCCA~CCATCffi-~CGffiffiG~TC~~CCCffiACCCCCGA~CCAC~G~~~T 730 FNPGRRQRITMAIRTVREEILKAQTPEGHFGNVY

GCffiGCTGGCACClUCGARG~TCTAC~ACffiCCTCARGC~~ACc~AGTG~TCCTffiG~attgcc~~actctctttt

427

SGPYLTSVUGKRAGEREFWQLLRDPNTPLLQG

TCCTC~~~RCCCIUC~~~ATC~~~~TGTCACG~AGGTGCTT~~~TCTC~CCGC~TAC~~~AC~~~TCCACCTCTCTG~~CGGGTCC~GTGGAA

KARVALLASLQDGAFQNALMISQLLPVLNHKTYI

LALCLHQ,KRVHDSVVDKLLYAVEPFHQGHHSVD

DLIPPDCLAPR  VHLEPAAETI

Fig 2. The nucleotide sequence of the coding regions and the proximal 5' and 3' untranslated and flanking DNA of the TO1 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-my ligand. The single overlined sequences are the motifs for the ets-encoded nuclear binding proteins. The dashed-undelineed CCC tetramers and the dashed-overlined GGGG tetramer are motifs for the ETF transcription factor.

Table 3

Table 3. Analysis of the Intron-Exon Splice Juncitons

<table>
<thead>
<tr>
<th>Intron No.</th>
<th>5' Donor Splice Site</th>
<th>Intron Size* (bp)</th>
<th>3' Acceptor Splice Site</th>
<th>Intron Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TG gtagt</td>
<td>3,645</td>
<td>tctag AA</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>GG gtagt</td>
<td>1,790</td>
<td>tctag GT</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>TG gtagc</td>
<td>1,630</td>
<td>tctag GG</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>GG gtagc</td>
<td>690</td>
<td>tctag AC</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>AG gtagg</td>
<td>128</td>
<td>cogag TT</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>GA gtagc</td>
<td>1,530</td>
<td>cogag GT</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>CA gtagg</td>
<td>1,860</td>
<td>cogag CA</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>AG gtagt</td>
<td>5,200</td>
<td>cogag GT</td>
<td>1</td>
</tr>
</tbody>
</table>

*Except for intron 5, which was completely sequenced, the size of the other introns is an approximate value estimated by PCR amplification and/or restriction mapping of the genomic clones.
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 SP1 protein (lane 2). A substantial component of this shift was blocked when the SP1 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 SP1-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 SP1, 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, Hinf I, 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 Hinf I, Msp I, and Taq I identify RFLP type III, they do not distinguish RFLP

A 420-bp fragment extending from 5' (-328) → 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).
types I and II (each enzyme generates the same number and size of restriction fragments). BglI alone, however, identifies all three types.

The DNA prepared from fibroblasts cultured from three patients who lacked expression of a functional and immunoreactive 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 IF and human TCI shows greater than 50% nucleotide homology and 20% amino acid homology 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 underline in Fig 2) located 196 nt further upstream. These sequences, including the GC box, are motifs for the trans-acting 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 transcription factor. 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′ (-843) → 3′ (-811) that may also serve as DNA binding sites for this oncogene product. There are also three GGAA tetramers (5′ (-465) → 3′ (-402), 5′ (-312) → 3′ (-309), and 5′ (-112) → 3′ (-109)) that have been identified as a motif to which products of the ets gene family bind.31

**Fig 5.** The mobility-shift assay of a 32P-labeled, 507-bp 5′ flanking DNA fragment containing the GC-rich sequence. Lane 1, labeled fragment alone; lane 2, labeled fragment plus purified SP1 protein; lane 3, purified SP1 protein preincubated with the unlabeled 507-bp fragment followed by the labeled fragment; lane 4, purified SP1 protein preincubated with the 34-bp synthetic fragment containing the putative GC box followed by the labeled fragment; lane 5, purified SP1 protein preincubated with the SP1-specific oligonucleotide followed by the labeled fragment; lane 6, labeled fragment plus 3.5 pg NIH 3T3 nuclear extract proteins; lane 7, nuclear extract preincubated with poly dl-dC and unlabeled 507-bp fragment followed by the labeled fragment; lane 8, nuclear extract preincubated with poly dl-dC followed by labeled fragment; lane 9, nuclear extract preincubated with poly dl-dC and unlabeled 34-bp synthetic fragment containing the putative GC box followed by the labeled fragment.
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.\textsuperscript{35} In multiple myeloma, a translocation of chromosome 8q24, the locus of the c-myc protooncogene, to chromosome 14q32 has been observed,\textsuperscript{35} and an elevation of c-myc mRNA in the myeloma cells of some patients has been reported.\textsuperscript{35} 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.\textsuperscript{35} This activation of the TCII gene by a trans-active 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.\textsuperscript{35} This could also explain the elevated plasma TCII observed in patients with breast cancer\textsuperscript{36} and histiocytic proliferation.\textsuperscript{36}

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\textsuperscript{37} have found 36% amino acid identity of human IF and human haptocorrin as deduced from their cDNAs, and our previous study\textsuperscript{12} also showed that human TCII contains several stretches of amino acid sequences that are homologous with human TCI\textsuperscript{13} and rat IF.\textsuperscript{22} A similar comparative analysis of the amino acid sequence homology of these three protein and porcine haptocorrin\textsuperscript{39} has identified four regions with 80% homology and two regions with 60% homology. John-
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 indicate that a major event(s) occurred during this evolution 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 HinfI,Msp I, and Tag 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.

### Table 4. RFLP of the TCII Gene

<table>
<thead>
<tr>
<th>RFLP Type</th>
<th>Bgl I (n = 7)</th>
<th>Hinf I (n = 4)</th>
<th>Msp I (n = 4)</th>
<th>Tag I (n = 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>If</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Il</td>
<td>6 (−3.3 kbp)</td>
<td>4</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>IlIII</td>
<td>10 (−0.8 kbp)</td>
<td>8 (−1.2 kbp)</td>
<td>10 (−2.4 kbp)</td>
<td></td>
</tr>
<tr>
<td>IlII</td>
<td>5 (−0.7 kbp)</td>
<td>10 (−0.8 kbp)</td>
<td>8 (−1.2 kbp)</td>
<td></td>
</tr>
<tr>
<td>IlIIII</td>
<td>7 (−3.3 kbp)</td>
<td>5 (−0.7 kbp)</td>
<td>10 (−0.8 kbp)</td>
<td>8 (−1.2 kbp)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(−3.8 kbp)</td>
<td>(−0.4 kbp)</td>
<td></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 in parentheses 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).

### REFERENCES


9. Seligman PA, LaDonna L, Steiner LL, Allen RH, Lazerson J:
The cloning and characterization of the human transcobalamin II gene

A Regec, EV Quadros, O Platica and SP Rothenberg