

# Genomic analysis of human and mouse *TCL1* loci reveals a complex of tightly clustered genes

Cora Hallas\*, Yuri Pekarsky\*, Takahiro Itoyama, James Varnum, Roberta Bichi, Jay L. Rothstein, and Carlo M. Croce\*

Kimmel Cancer Institute, Thomas Jefferson University, 233 South 10th Street, Philadelphia, PA 19107

Communicated by Sidney Weinhouse, Jefferson Medical College, Philadelphia, PA, October 6, 1999 (received for review September 3, 1999)

*TCL1* and *TCL1b* genes on human chromosome 14q23.1 are activated in T cell leukemias by translocations and inversions at 14q32.1, juxtaposing them to regulatory elements of T cell receptor genes. In this report we present the cloning, mapping, and expression analysis of the human and murine *TCL1/Tcl1* locus. In addition to *TCL1* and *TCL1b*, the human locus contains two additional genes, *TCL1*-neighboring genes (*TNG*) 1 and 2, encoding proteins of 141 and 110 aa, respectively. Both genes show no homology to any known genes, but their expression profiles are very similar to those of *TCL1* and *TCL1b*. *TNG1* and *TNG2* also are activated in T cell leukemias with rearrangements at 14q32.1. To aid in the development of a mouse model we also have characterized the murine *Tcl1* locus and found five genes homologous to human *TCL1b*. *Tcl1b1*–*Tcl1b5* proteins range from 117 to 123 aa and are 65–80% similar, but they show only a 30–40% similarity to human *TCL1b*. All five mouse *Tcl1b* and murine *Tcl1* mRNAs are abundant in mouse oocytes and two-cell embryos but rare in various adult tissues and lymphoid cell lines. These data suggest a similar or complementary function of these proteins in early embryogenesis.

The *TCL1* gene at chromosome 14q32.1 is commonly activated in T cell neoplasms by chromosome rearrangements such as inversions *inv*(14)(q11;q32) and translocations *t*(14;14)(q11;q32) or *t*(7;14)(q35;q32). The gene normally is expressed in pre-B cells and immature IgM-expressing B cells and in early T cell progenitors (CD4<sup>+</sup>, CD8<sup>+</sup>, CD3<sup>+</sup>), but not in plasma cells or mature T cells (1). Introduction of a *TCL1* transgene under the control of the T cell-specific Ick promoter into mice causes T cell proliferative disorder and, at the age of 15 months, T cell leukemia (2). Another member of the *TCL1* gene family is the *MTCPI* gene on chromosome Xq28. *MTCPI* also is activated in rare cases of T cell leukemia by a *t*(X;14)(q28;q11) translocation (3). In rare cases of mature T cell leukemias with chromosomal abnormalities at 14q32.1, activation of the *TCL1* gene was not observed (4, 5). Recently, we have isolated a second putative oncogene in this region, the *TCL1b* gene, which is located approximately 16 kb centromeric to *TCL1* and shares 60% amino acid sequence similarity with *TCL1*. The expression profiles of both genes are very similar. *TCL1* and *TCL1b* are expressed at very low levels in normal bone marrow and peripheral blood lymphocytes (1, 6), but at higher levels in T cell lines containing rearrangements of the 14q32.1 region (1, 6). Because genes in close proximity to *TCL1* and *TCL1b* also may be activated in leukemias with rearrangements at 14q32.1, we have investigated the chromosomal region bracketed by two previously published breakpoint cluster regions observed in T cell neoplasias (1, 7) at 14q32.1 for the presence of additional genes.

Furthermore, we also examined the murine *Tcl1* locus to investigate the function of *TCL1* and *TCL1b*. In the mouse, the syntenic region of human chromosome 14q32 is the region of the murine chromosome 12 proximal to the Ig locus. The murine *Tcl1* protein shows a 50% homology to the human *Tcl1* (8) and is expressed in fetal hematopoietic organs and in immature T and B cells as well as in adult spleen and thymus (8). To identify other members of the murine *Tcl1* family we also have investigated the murine *Tcl1* locus for the presence of homologous genes.

## Methods

**Cell Lines.** Human leukemia cell lines MOLT 3, MOLT 4, CEM, and SupT11 (T cell leukemias), 697 (pre-B cell leukemia), CA-46, Raji, and Daudi (Burkitt's lymphomas) were obtained from American Type Culture Collection (Manassas, VA). Mouse lymphatic cell lines NFS-70 C-10 (pro-B cells), NFS-5 C-1 and WEHI-279 (pre-B cells), MOPC-31C and MPC-11 (plasma cells), and S49.1 and BW 5147 (thymocytes) also were purchased from American Type Culture Collection. All cell lines were grown in RPMI medium 1640 with 10% FBS.

**Rapid Amplification of cDNA Ends (RACE) and Reverse Transcription-PCR (RT-PCR) Analysis.** These experiments were carried out as described (9). Mouse and human tissue cDNAs for RT-PCR and RACE experiments were purchased from CLONTECH. Mouse egg and two-cell embryo cDNA libraries for embryonic expression studies in mouse have been described (10). The DNAs from these libraries were diluted to the same concentration of cDNA as in mouse tissue samples. RNA extractions and RTs from human and mouse cell lines and mouse embryonic stem cells were performed by using Trizol reagent (GIBCO/BRL). Two micrograms of total RNA was transcribed into cDNA in a total volume of 20  $\mu$ l by using the SuperScript RT kit (GIBCO/BRL) according to the manufacturer's instructions. One microliter of this reaction was used for PCR. RT-PCR for *TCL1* neighboring gene 1 (*TNG1*) was carried out with primers 1A (TGCATC-CCTCCAGCCAAGGAT) and 4A (TGGCCTGCAGAG-GCTCTCAAG) for 25–35 cycles. For *TNG2*, primers 3B (GT-GCCTGTCTCATTCGCCTCTG) and 8B (AGTGGGCACAT-GTTACAGCATTC) were used for the first round of 25 cycles and primers 4B (GCATCCAGGACTGTGCCAGCA) and 9B (TTCTGTTAGCCTTGCTGTCCGT) were used to amplify 0.1  $\mu$ l of the first reaction in a nested PCR of 20 cycles. PCR conditions were 94°C denaturation for 30 sec, 54–62°C annealing for 30 sec, and 72°C extension for 30 sec. *TCL1*, *TCL1b*, and, as control, *G3PDH* were amplified as described (6). RACE analysis was carried out in a nested reaction with 30 cycles in the first round and 25 in the second. The primers were, for *TNG1*, 1A and 2A (TTGAACCCAGGTCTCGTCTGAC, nested) for 3' RACE and 3A (AACGTAGGATGTGCACAGCA) and 4A (nested) for 5' RACE and, for *TNG2*, 3B and 4B (nested) for 3' RACE and 8B and 9B (nested) for 5' RACE together with primers AP1 and AP2 supplied by CLONTECH fitting to the adapters of the cDNA. The murine *Tcl1b* genes were amplified

Abbreviations: RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-PCR; EST, expressed sequence tag; BAC, bacterial artificial chromosome; T-PLL, T cell polymorphic leukemia.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. AF195488 to AF195493 (murine *Tcl1b1* to *Tcl1b5* and AF195820 and AF195821 (human *TNG1* and *TNG2*)].

\*C.H. and Y.P. contributed equally to this work.

<sup>†</sup>To whom reprint requests should be addressed at: Kimmel Cancer Institute, 233 South 10th Street, BLSB 1032, Philadelphia, PA 19107. E-mail: croce@calvin.jci.tju.edu.

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. §1734 solely to indicate this fact.

by using the respective R (reverse: 1R, GAGAACGGTCAG-GACCCAAACC; 2R, CAGGCTATCAAGACCTTTACTC; 3/5R, TCAACCTCGCATATTACTATGTC; 4R, CAAAG-GCACAAAGTGAGCAAGAG) and F (forward: 1F, AATGT-GGAAACTTCTCACTCAT; 2F, ACTGGAAACTTGTTC-CATTAC; 3/5F, CACTTGACATATGACCACAAT; 4F, CCTGGTCTGCACAAGAGATGA) primers for 28 cycles. Subsequently, the respective R and FN (forward nested: 1FN, CTGTCCACTTGTGGAAGTTAAT; 2FN, CACTTGTG-GCAGATGACCAGATA; 3/5FN, CCAGGAGCTACTC-CCCAGCAG; 4FN, GTGGCAGATGACCACACTCTT) primers were used in a seminested PCR for 25 cycles to amplify 1  $\mu$ l of the first reaction. PCR conditions were the same as described for human tissues. Because of the similarity of mouse *Tcl1b* genes, it was difficult to find specific primers for each of them. Subsequently, *Tcl1b3* and *Tcl1b5* were amplified with the same primers and sequenced to verify the expressed gene. However, in the case of embryonic tissue, unique forward primers were used to analyze the expression of *Tcl1b3* and *Tcl1b5* separately. The expression of both alternative first exons of *Tcl1b3* was verified by using the primers 3F hom (homologous exon 1: CATTACTATGGCTGATTCAGTTC) and 3F alt (alternative exon 1: GGAATGAGACTCTCAGGGCAC) instead of 3/5F. RT-PCR for *Tcl1* was carried out similarly with primers Tc11R (CCTGGGCAAGGCAGACAGGAGC) and Tc11F (TGCT-TCTTGCTCTTATCGGATG), followed by a nested PCR by using primers Tc11RN (TTCATCGTTGACTCCGAGTC) and Tc11FN (AATTCCAGGTGATCTTGCGCC). The quality of the cDNA was verified by 25 cycles of  $\beta$ -actin RT-PCR using primers actR (GTACCACCAGACAGCACTGTG) and actF (GACCCAGATCATGTTGAGACC). RACE analysis from mouse tissues was performed as described above for human tissues. The specific primers were allR (AAGCCATCTATA-AGGTCAGG) for the first step and the respective R primers for the nested step of 5' RACE and the respective F (first) and FN (nested) primers for 3' RACE.

**Sequencing.** Products from RACE and RT-PCR experiments were cut and extracted from agarose gels by using a QIAquick gel extraction kit (Qiagen, Chatsworth, CA) according to the manufacturer's instructions. Subsequently, they were sequenced by using an automated sequencer model 377 (Perkin-Elmer). Human bacterial artificial chromosome (BAC) (277A8) was partially digested with *Sau3A* and *TSP509I* and cloned into a pUC18 vector by using standard methods. One hundred random clones were isolated and sequenced from both ends by using a 377 automated sequencer. The DNA sequences were compared with the expressed sequence tag (EST) database. The mouse BAC 452-I24 was sequenced and analyzed as described (11). EST clones were purchased from Research Genetics (Huntsville, AL) and sequenced.

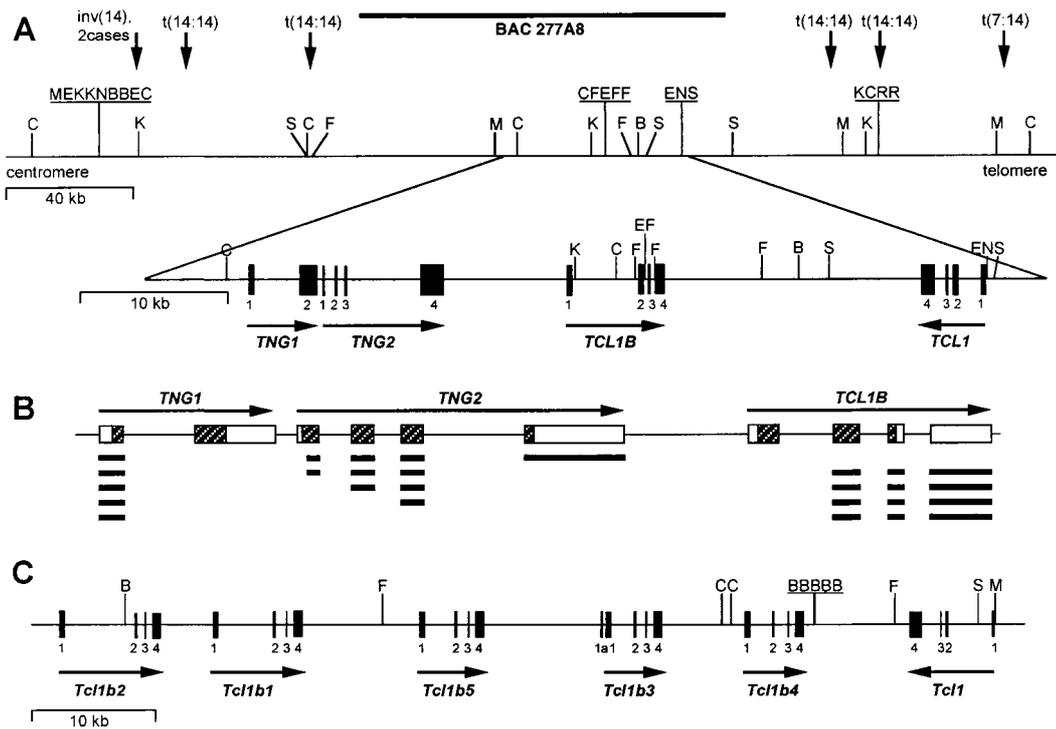
**Northern Blot and Pulse-Field Gel Electrophoresis (PFGE).** Total RNA for Northern blot experiments was isolated as described above. Poly(A)<sup>+</sup> RNA isolation, Northern blotting, and hybridization was performed as described (12). *TNG1* and *TNG2* probes were generated by RT-PCR. PFGE analysis was performed as described (6) by using BAC (277A8) DNA and *TNG1*, *TNG2*, *TCL1*, and *TCL1b* probes.

**Protein Structure.** A computer model was created for the human and murine Tc1b proteins based on their similarity to Tc11. The atomic coordinates for human TCL1 are derived from the crystal structure (13). The initial sequence alignment was generated by maximizing the correlation between the sequences. Modeling and analysis were done by using INSIGHTII (Biosym Technologies, San Diego).

## Results

**The Human *TCL1* Locus.** The *TCL1* and *TCL1b* genes both are located on chromosome 14q32.1 within a  $\approx$ 160-kb region between two previously published breakpoint cluster regions observed in T cell neoplasms (1, 7). Both genes are activated by translocations and inversions involving 14q32.1 (6). To investigate whether other unknown genes within this region also are activated by the same rearrangements, we analyzed a previously isolated BAC of 110 kb (277A8; ref. 6) covering the majority of this region. This BAC was partially digested with the restriction enzymes *Sau3A* and *TSP509I* and cloned into a pUC18 vector. One hundred clones (the equivalent of the length of the BAC) were picked randomly and sequenced from both sides. These sequences were compared with the EST database, and two different sets of ESTs homologous to the BAC sequences were found. We isolated two full-length cDNAs by using 3' and 5' RACE and RT-PCR of cDNA from human testis, peripheral blood lymphocytes, and the Burkitt's lymphoma cell line Raji by using primers made from the different ESTs. The 1.5-kb cDNA of *TNG1* contains an ORF coding for a protein of 141 aa with the start codon ATG at position 161. The 2-kb cDNA of *TCL1* neighboring gene 2 (*TNG2*) encodes a shorter protein of 110 aa with the start codon at position 36. Both genes do not show homology to any known genes found in the database. Relative positions of the genes and their distances from each other were determined by Southern hybridization and pulse-field Southern analysis. *TNG2* is located 8 kb centromeric of *TCL1b* and *TNG1* is only 118 bp centromeric of *TNG2*. *TNG1*, *TNG2*, and *TCL1b* have the same transcriptional orientation, opposite to *TCL1* (Fig. 1A). The *TNG1* gene is 4.5 kb in size and contains only two exons of 215 and 1,239 bp. The *TNG2* gene has a size of 8.6 kb containing four exons of 134, 136, 157, and 1,651 bp, all of which are coding. RT-PCR and RACE experiments revealed several alternatively spliced RNAs linking various exons of *TNG1* and *TNG2* to exon 2 of *TCL1b* (Fig. 1B). Only one of these RNAs, linking the exon 1 of *TNG1* in-frame to the second exon of *TCL1b*, contains a new ORF encoding a *TCL1b* protein with an alternative N-terminal end.

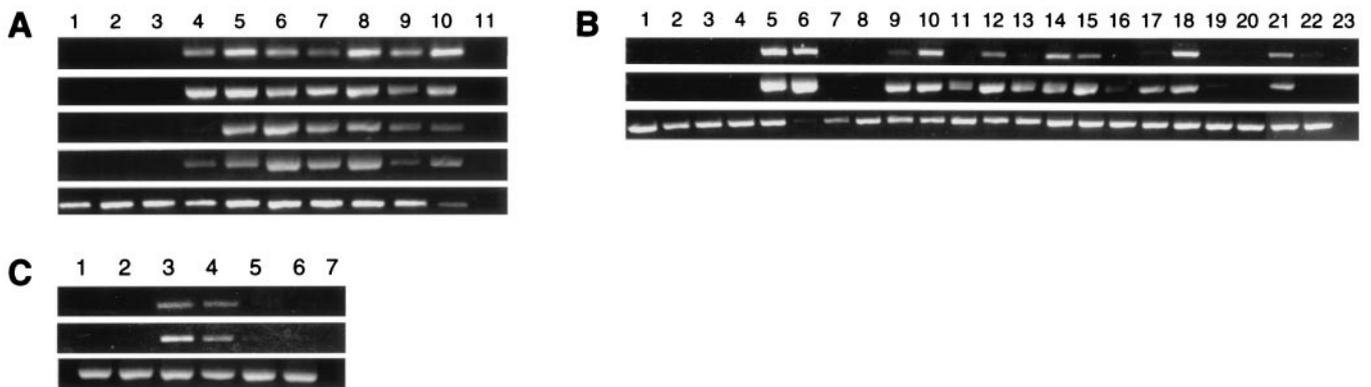
**The Murine *Tcl1* Locus.** To identify the murine *Tcl1b* gene we searched the murine EST database for sequences homologous to human *TCL1b*. We found three sets of ESTs that were very similar, but not identical, to each other and showed homology to human *TCL1b*. Additionally, we screened a BAC library and obtained three clones containing murine *Tcl1*. PCR analysis of these BAC clones confirmed the presence of all three EST sequences. By a combination of RACE and RT-PCR experiments, database analysis of EST sequences, and sequencing of selected EST clones, we isolated full-length cDNAs corresponding to these sequences. Because of the shared similarity among the cDNAs it was not possible to obtain unique probes for each. Thus, we could not determine the genomic structure of the region by conventional methods such as Southern hybridization and pulse-field gel analysis. Subsequently, we sequenced the BAC (452-I24) and determined the position and the exon-intron boundaries of the three cDNAs. Further analysis of the region revealed that it also contains two other sequence-related genes. RT-PCR experiments with specific primers for these two genes confirmed that they are transcribed (see below). Altogether, we isolated five full-length cDNAs located on murine chromosome 12 centromeric to the *Igh* locus homologous to human *TCL1b*. Murine *Tcl1b1*-*Tcl1b5* genes had a length of  $\approx$ 1 kb encoding for proteins ranging in size from 117 to 123 aa. They share 70–90% nucleic acid homology, 55–75% amino acid identity, and 65–80% amino acid similarity. The murine *Tcl1b* family members show  $\approx$ 25% identity and  $\approx$ 35% similarity to murine *Tcl1* and are 25–30% identical and 30–40% similar to human *TCL1b*. The five



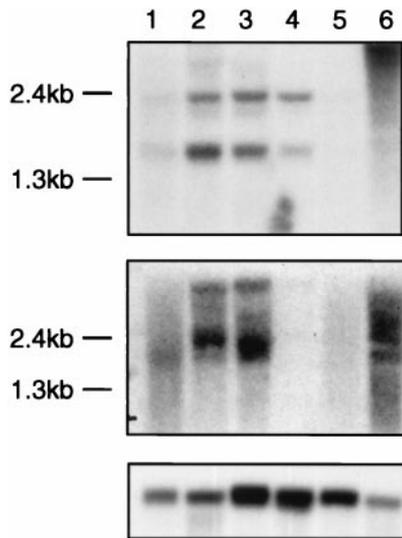
**Fig. 1.** Genomic organization of human and mouse *TCL1* loci. (A) Human *TCL1* locus. Vertical arrows refer to cloned 14q32.1 breakpoints (1, 7). Restriction sites are given for *Bss*III (B), *Clal* (C), *Eag*I (E), *Sfi*I (F), *Ksp*I (K), *Mlu*I (M), *Not*I (N), and *Sal*I (S). Solid boxes represent exons of the four genes. (B) Striped boxes indicate translated parts of exons, and open boxes indicate untranslated regions. Bold lines under the exons show various splicing products of *TNG1*, *TNG2*, and *TCL1B* genes. (C) Murine *Tcl1* locus. Restriction sites and exons are indicated as in A.

genes are aligned on murine chromosome 12 (Fig. 1C) in the order *Tcl1b2*, *Tcl1b1*, *Tcl1b5*, *Tcl1b3*, and *Tcl1b4* with distances of 4.5 kb, 9.7 kb, 9.9 kb, and 6.8 kb, respectively, from each other and 9.8 kb between *Tcl1b4* and *Tcl1*. The total sizes of the genes are: *Tcl1b1*, 6.9 kb; *Tcl1b2*, 8.2 kb; *Tcl1b3* and *Tcl1b4*, 4.6 kb; and *Tcl1b5*, 4.8 kb. The direction of transcription of *Tcl1b1*–*Tcl1b5* is opposite to that of *Tcl1*. Each of the murine *Tcl1b* genes contains four exons of approximately 200, 170, 70, and 590 bp in size. The only exceptions are the exons 3 of *Tcl1b2* and *Tcl1b4*, in which a different splicing site leads to a transcript 29 bp shorter. In addition, sequences of RT-PCR and RACE products

and ESTs derived from GenBank showed alternatively spliced cDNAs for *Tcl1b1* and *Tcl1b3*. *Tcl1b1* may have a deletion of 73 bp consisting of nearly the complete exon 3 and the first 6 bp of exon 4. Because this deletion includes the stop codon, the deduced protein sequence is slightly longer. For *Tcl1b3*, an alternative exon 1 was found leading to a shorter protein with an alternative N-terminal end without homology to other *Tcl1b* proteins. Although the homology of murine *Tcl1b* proteins to human *Tcl1b* is lower than typically observed between mouse and human homologues (70–100%), the position of the genes on the map, their direction of transcription, and their exon–intron



**Fig. 2.** RT-PCR analysis of *TNG1* and *TNG2* genes. (A) Leukemia cell lines. Lanes: 1–3, T-ALL cell lines MOLT3, MOLT4, and CEM; 4, pre-B-ALL cell line 697; 5, T-ALL cell line Supt11; 6–8, Burkitt’s lymphoma cell lines CA-46, Raji, and Daudi; 9 and 10, bone marrow and peripheral blood lymphocytes (PBL). Top gel, *TCL1* primers; second, *TCL1b* primers; third, *TNG1* primers; fourth, *TNG2* primers; bottom, control *G3PDH* primers (B) Normal human tissues. Lanes 1–23: heart, liver, brain, muscle, placenta, kidney, lung, pancreas, spleen, lymph node, thymus, tonsil, PBL, fetal liver, fetal brain, fetal lung, fetal kidney, fetal heart, fetal skeletal muscle, fetal spleen, fetal thymus, negative control. (C) Lanes: 1–4, T cell PLL samples 3047, 3046, 3050, and 3048; 5 and 6, bone marrow and PBL. (B and C) (Top) *TNG1* primers. (Middle) *TNG2* primers. (Bottom) Control *G3PDH* primers.



**Fig. 3.** Northern analysis of *TNG1* and *TNG2* genes. Lanes: 1–3, Burkitt's lymphomas Raji, Daudi, and CA-46; 4, T-ALL SupT11; 5 and 6, bone marrow and placenta. (Top) *TNG1* probe. (Middle) *TNG2* probe. (Bottom) Actin probe. Each lane contains 3  $\mu$ g of poly(A)<sup>+</sup> RNA.

structure are similar to the human *TCL1b* locus and indicate that these genes are authentic homologues to the human *TCL1b* gene.

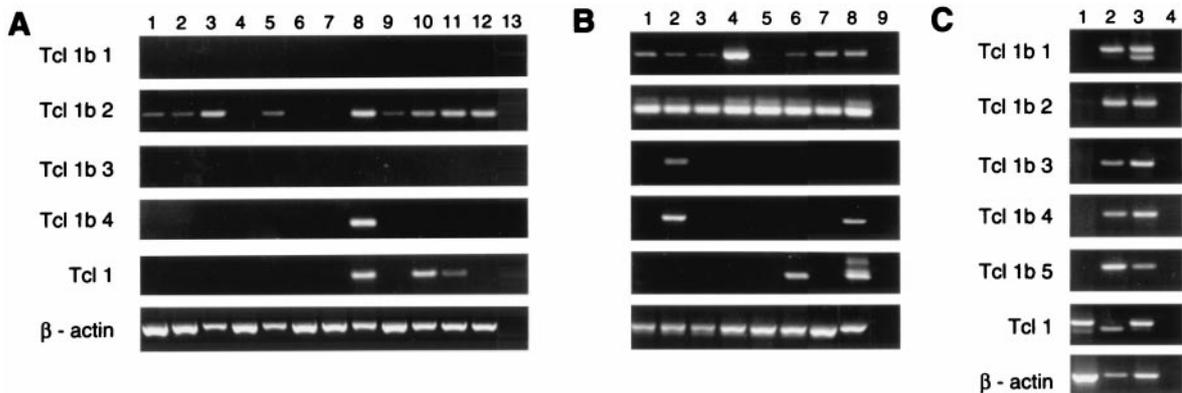
**Expression of Human *TNG1* and *TNG2*.** Because *TNG1* and *TNG2* are located at the same locus as *TCL1* and *TCL1b*, it seemed possible that they would exhibit similar expression patterns. To investigate this, we performed a series of Northern blot and RT-PCR experiments. *TNG1* and *TNG2* both are transcribed in a wide variety of normal tissues (Fig. 2B). The results demonstrate a low level of expression in most tissues examined including placenta, kidney, fetal kidney, fetal lung, and fetal heart and all lymphoid tissues including fetal liver and fetal spleen. The only exception is thymus, which showed transcripts of only *TNG2*, whereas fetal thymus expressed only *TNG1* (Fig. 2B). *TCL1b* was expressed in the same tissues as *TNG1* except thymus, fetal lung, and fetal heart (2). Northern blot analysis of normal adult and embryonic tissues was negative for *TNG1* and *TNG2* (not shown), probably because of the low level of expression.

Because of the similarity of transcription patterns of *TNG1*

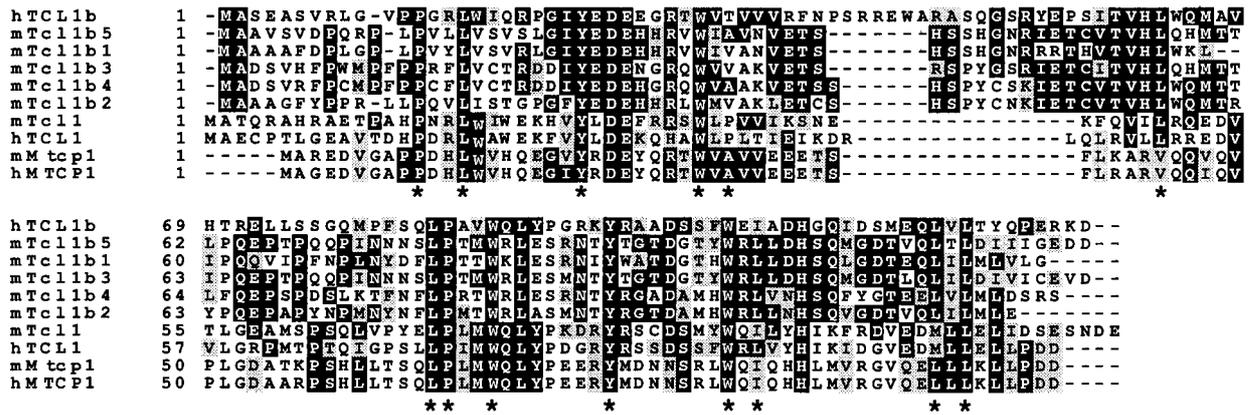
and *TNG2* to those of *TCL1* and especially *TCL1b* and the physical linkage of these genes, we investigated whether the *TNG* genes also could be activated by rearrangements at 14q32.1. Fig. 2A demonstrates that all four genes show an identical expression pattern in lymphoid tumor cell lines. They all are expressed in early B tumor cell lines (697, Raji, Daudi, and CA-46), but not in postthymic T cell lines without 14q32.1 rearrangements. Nevertheless, *TCL1*, *TCL1b*, *TNG1*, and *TNG2* all are transcribed in the T-ALL cell line SupT11 carrying a t(14;14)(q11;q32) translocation. Northern blot experiments confirmed these transcription patterns (Fig. 3). The 1.5-kb transcript of *TNG1* was found in Burkitt's lymphoma cell lines Daudi and CA-46 and, to a lesser extent, also in the Raji cell line and in the T cell acute lymphocytic leukemia cell line SupT11 (T-ALL) that carries a 14q32.1 translocation. The second band, a  $\approx$ 2.3-kb transcript, is probably a product of alternative splicing or incompletely processed hnRNA. However, activation of *TNG2* in the SupT11 cell line was not confirmed by Northern blotting, probably because of a lower expression level. The 2-kb *TNG2* transcript was detected in all three Burkitt's lymphoma cell lines, but not in the pre-B cell line 697 or in any of the T cell lines investigated. The diffuse signal around the bands is caused by the various alternative splicing products known to involve these genes.

To study further the possible activation of *TNG1* and *TNG2* by rearrangements at 14q32.1, we investigated the expression of *TNG1* and *TNG2* in four T cell prolymphocytic leukemias (T-PLL) overexpressing *TCL1*. Fig. 2C shows the activation of both genes in two of four cases. The transcripts of *TNG1* and *TNG2* were detected after 27 cycles of PCR in these two cases, even though at these conditions bone marrow and peripheral blood lymphocytes were negative. Interestingly, we previously found activation of *TCL1b* in one of two cases not expressing the *TNG* genes (6). These results indicate that juxtaposition of the *TCL1* locus at 14q32.1 to the  $\alpha/\delta$  locus of the T cell receptor may activate *TNG1* and *TNG2* as well as *TCL1* and *TCL1b*.

**Expression of Murine *Tcl1b* Genes.** To investigate the expression pattern of the murine *Tcl1b* genes, we carried out a series of RT-PCR experiments for each of the five genes. After a single round of PCR, no mRNA expression was found in a series of normal tissue, embryonic cDNA libraries, and lymphoid cell lines for any of the five genes (data not shown). However, nested PCR analysis revealed a low level of expression of *Tcl1b2* in all lymphoid cell lines (Fig. 4B) and nearly all normal tissues. Further, *Tcl1b2* expression increased during embryonic devel-



**Fig. 4.** RT-PCR analysis of murine *Tcl1b* genes. (A and B) Nested PCR, except  $\beta$ -actin. The gels are in the same order. (A) Normal mouse tissues. Lanes 1–13: heart, brain, spleen, lung, liver, skeletal muscle, kidney, testis, 7-day embryo, 11-day embryo, 15-day embryo, 17-day embryo, negative control. (B) Lymphoid cell lines. Lanes: 1–5, B cell lines NFS-5, NFS-70, WEHI-279, MOPC-31C, and MPC-11; 6 and 7, T cell lines S49.1 and BW5147; 8 and 9, embryonic stem cells and negative control. (C) Single round of PCR. Lanes 1–4: ES cells, mouse oocytes, two-cell embryos, and negative control.



**Fig. 5.** Sequence comparison of human and murine Tc1, Tc1b, and Mtcp1 proteins. Identities are shown in black boxes, and similarities are indicated by shaded boxes. \*, Conserved residues of the inner hydrophobic core.

opment (7- to 17-day-old embryos, Fig. 4A). A low level of expression also was found for *Tcl1b1* in nearly all lymphoid cell lines, but not in any other tissues. *Tcl1b4* only showed a low level of expression in testis and in the pro-B cell line NFS 5, which also expressed *Tcl1b3*, as confirmed by sequencing (Fig. 4). *Tcl1b5* expression was not detected in any tissue or cell line examined. In comparison with the *Tcl1b* genes, *Tcl1* was expressed at low levels in testis, 11- and 15-day-old embryos, and in the thymocyte cell line S49-1. Interestingly, murine *Tcl1* was not detected in any of the early B cell lines, although early B cells show expression of *TCL1* in humans (1, 2). Because the original three sets of ESTs all derived from a two-cell embryonic cDNA library, where they make up  $\approx 0.5\%$  of the total ESTs, we investigated cDNA from mouse embryonic stem cells, oocytes, and two-cell embryos for the expression of *Tcl1b* genes. After a single round of RT-PCR, expression of all five *Tcl1b* genes and *Tcl1* was found in mouse oocytes and two-cell embryos at a level comparable to that of  $\beta$ -actin expression (Fig. 4C). In two-cell embryos, both splicing variants of *Tcl1b1* were amplified. Interestingly, in the mouse oocyte cDNA library, only a shorter transcript of *Tcl1* was detected, missing a part of exon 2. Only *Tcl1* showed expression in embryonic stem cells after a single round of PCR, but nested PCR also revealed a low level of expression of *Tcl1b1*, *Tcl1b2*, and *Tcl1b4*. The high expression of all five *Tcl1b* genes and *Tcl1* in mouse oocytes and two-cell embryos suggests that an important function of these genes probably occurs in the early embryogenesis of the mouse.

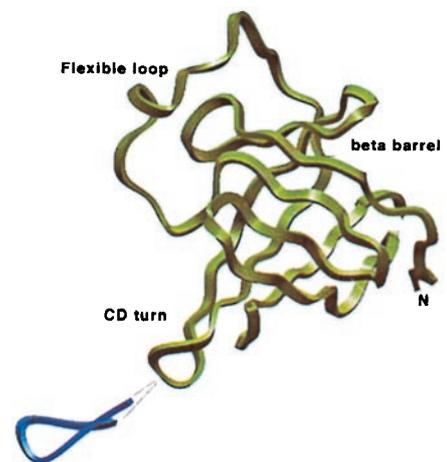
**Protein Structure of *TCL1* Family.** Tc1 and Mtcp1 proteins both consist of an eight-stranded antiparallel  $\beta$ -barrel with a hydrophobic core and are predicted to bind small hydrophobic ligands (14). Amino acid sequence alignment of these proteins with human and mouse Tc1b (Fig. 5) shows that, despite only an overall 30–40% homology, all 14 aa forming the hydrophobic core are conserved except Pro36. Ten of these 14 aa are identical in all 10 members of the Tc1 family, whereas three residues show conservative substitutions in some of the proteins (Leu-49  $\rightarrow$  Val, Leu-92  $\rightarrow$  Ile, Met-104  $\rightarrow$  Leu). This finding suggests that those residues may have an important function in all Tc1 family members. Human Tc1b shows a 14-residue insertion (Arg-44—Glu-58) relative to human Tc1 (Fig. 5). Mouse Tc1b has a smaller, 10- to 11-residue insertion in the same position. A molecular model was built for human and murine Tc1b based on the 35% similarity in amino acid sequence to Tc1. In this model, the Tc1b insertion aligns with a noncanonical, five-residue turn (Lys-42—Gln-46) observed in the crystal structure of human Tc1 (13). The additional residues in human and mouse Tc1b

may form a surface-accessible  $\beta$ -sheet extension or a flexible loop with conserved charged amino acids (Fig. 6).

### Discussion

In this report we presented the cloning, mapping, and expression analysis of the human and murine *TCL1/Tcl1* locus. Human *TCL1* and *TCL1b* genes are located between two clusters of chromosomal breakpoints and are activated by translocations and inversions at 14q32.1, juxtaposing them to regulatory elements of T cell receptor genes (6). Between these two sets of breakpoints we found and characterized two additional genes, *TNG1* and *TNG2*. Both show no homology to any known genes but similar expression patterns to that of *TCL1* and *TCL1b*. Both *TNG* genes also are activated in the T cell leukemia cell line SupT11 carrying a t(14;14) translocation and in two of four T-PLL samples. These results suggest that, like *TCL1* and *TCL1b*, these two genes also may be activated by rearrangements at 14q32.1 involved in T cell malignancies. T cell leukemias, therefore, in some cases may be induced by the activation of a single gene or in others by the cumulative activation of two or more of these four genes, although the oncogenic potential of *TNG1* and *TNG2* remains to be determined.

To assist in the structural and functional analysis of the human *TCL1* gene activation, we investigated the murine *Tcl1* locus



**Fig. 6.** Location of the insertion in human and murine Tc1b proteins. A side view of human *TCL1* (14) is shown in green. The proposed *TCL1b* insert into the C-D loop is shown in blue.

looking for homologues to human *TCL1b* and *TNG* genes. To our surprise, we found five genes homologous to human *TCL1b*. The high shared similarity between the five murine *Tcl1b* genes not only in the exons but also in intronic sequences suggests that they are most likely a result of duplications. All five genes are transcribed into mRNA, but it remains to be determined whether they all code for active proteins or whether some of them might be pseudogenes. Their genomic structure, though, is untypical for pseudogenes because it includes introns. The five *Tcl1b* genes show different expression patterns, suggesting different regulatory elements for each of them, but because the expression of all of the genes in all adult tissues and cell lines is very low, the significance of this finding is not clear. Interestingly, the expression of murine *Tcl1* and *Tcl1b* genes in lymphoid tissues and cell lines is much lower than the expression of their human homologues. The most striking feature of murine *Tcl1* and *Tcl1b* genes is their very high expression level (up to 0.5% of all mRNA) in mouse oocytes and two-cell embryos. This finding is consistent with the presence of human *TCL1b* in a syncytiotrophoblast-subtracted cDNA library (GenBank accession no. AF137027), suggesting a function of murine and human *TCL1b* genes in the early embryogenesis.

The identification of five more murine members of the *Tcl1* family provides a better understanding of the structural differences and similarities between the *Tcl1* family of proteins. A comparison of the protein sequences of all members of the family including murine and human *MTCP1* shows that, although

overall homologies between the genes are low, the hydrophobic core region as described by Fu *et al.* (14) is preserved. This may indicate a similar function for all of these proteins as transporters of small molecules such as retinoids, nucleosides, or fatty acids as suggested previously for *Tcl1* and *Mtcp1* (13, 14). However, compared with *Tcl1* and *Mtcp1*, mouse and human *Tcl1b* proteins show an insertion that may form a surface-accessible flexible loop or  $\beta$ -sheet extension. The conserved charged residues in the insert loop may play a significant role in mediating interactions with other proteins or ligands and also may influence the quaternary structure of mouse *Tcl1b* (13).

Altogether, murine and human *TCL1* loci show significant differences. There are five murine *Tcl1b* genes compared with one human *TCL1b*. The homology of human and mouse *TCL1b* is low, and the expression levels of murine *Tcl1* and *Tcl1b* in lymphoid tissues and cell lines are much lower than the expression levels of their human equivalents. Moreover, murine homologues of *TNG1* and *TNG2* were not found. This finding suggests that there also may be significant differences in the function of human and mouse *TCL1* loci. Further investigation should lead to a better understanding of the role of *TCL1* and *TCL1b* in normal development and T cell leukemia.

We thank Jean Letofsky for excellent technical assistance and Giandomenico Russo and Eric Martin for helpful discussions. This work was supported by Grant PO1 CA 76259 from the National Cancer Institute.

1. Virgilio, L., Narducci, M. G., Isobe, M., Billips, L. G., Cooper, M. D., Croce, C. M. & Russo, G. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 12530–12534.
2. Virgilio, L., Lazzeri, C., Bichi, R., Nibu, K., Narducci, M. G., Russo, G., Rothstein, J. L. & Croce, C. M. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 3885–3889.
3. Soulier, J., Madani, A., Cacheux, V., Rosenzweig, M., Sigaux, F. & Stern, M. H. (1994) *Oncogene* **9**, 3565–3570.
4. Sakashita, K., Kobayashi, H., Satake, N., Maseki, N., Sakurai, M., Izumo, T., Isobe, M. & Kaneko, Y. (1998) *Leukemia* **12**, 970–971.
5. Takizawa, J., Suzuki, R., Kuroda, H., Utsunomiya, A., Kagami, Y., Joh, T., Aizawa, Y., Ueda, R. & Seto, M. (1998) *Jpn. J. Cancer Res.* **89**, 712–718.
6. Pekarsky, Y., Hallas, C., Isobe, M., Russo, G. & Croce, C. M. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 2949–2951.
7. Virgilio, L., Isobe, M., Narducci, M. G., Carotenuto, P., Camerini, B., Kurosawa, N., Abbas-ar Rushdi, Croce, C. M. & Russo, G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 9275–9279.
8. Narducci, M. G., Virgilio, L., Engiles, J. B., Buchberg, A. M., Billips, L., Facchiano, A., Croce, C. M., Russo, G. & Rothstein, J. L. (1997) *Oncogene* **15**, 919–926.
9. Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y., Tillib, S., Draganescu, A., Wermuth, P., Rothman, J. H., Huebner, K., *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 8744–8749.
10. Rothstein, J. L., Johnson, D., DeLoia, J. A., Skowronski, J., Solter, D. & Knowles, B. (1992) *Genes Dev.* **6**, 1190–1201.
11. Inoue, H., Ishii, H., Alder, H., Snyder, E., Druck, T., Huebner, K. & Croce, C. M. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 14584–14589.
12. Hallas, C., Albitar, M., Letofsky, J., Keating, M. J., Huebner, K. & Croce, C. M. (1999) *Clin. Cancer Res.* **5**, 2409–2414.
13. Hoh, F., Yang, Y., Guignard, L., Padilla, A., Stern, M., Lhoste, J. & van Tilbeurgh, H. (1998) *Structure (London)* **6**, 147–155.
14. Fu, Z. Q., Du Bois, G. C., Song, S. P., Kulikovskaya, I., Virgilio, L., Rothstein, J. L., Croce, C. M., Weber, I. T. & Harrison, R. W. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 3413–3418.