Biochemical and Biophysical Research Communications **290,** 1054–1065 (2002)
doi:10.1006/bbrc.2001.6301, available online at http://www.idealibrary.com on IDEAL®

Analysis of the Human TrkB Gene Genomic Organization Reveals Novel TrkB Isoforms, Unusual Gene Length, and Splicing Mechanism

Peter Stoilov,* Eero Castren,† and Stefan Stamm*,[1]
**Friedrich-Alexander-University Erlangen, Institute for Biochemistry, Fahrstrasse 17, 91054 Erlangen, Germany; and †A.I. Virtanen Institute, University of Kuopio, P.O. Box 1627, 70211 Kuopio, Finland*

Received December 17, 2001

We determined the gene structure of the human TrkB gene. The gene is unusually large and spans at least 590 kbp. It contains 24 exons. Using alternative promoters, splicing, and polyadenylation sites, the gene can create at least 100 isoforms, that can encode 10 proteins. RT-PCR and Northern blot analysis reveals that only three major protein isoforms are generated by the gene: the full length receptor, an isoform lacking the tyrosine kinase domain, and a novel isoform lacking the tyrosine kinase domain but containing a Shc binding site. This novel isoform, TrkB-T-Shc is generated by the use of a new alternative exon 19. It is expressed only in brain. TrkB-T-Shc protein is located in the plasma membrane. Coimmunoprecipitation experiments show that TrkB-T-Shc is not phosphorylated by the full length receptor, indicating that it could be a negative regulator of TrkB signaling in the brain.

TrkB is a member of the protein tyrosine kinase family (1, 2) and together with trkA and trkC it forms the trk subfamily of RTKs. TrkB is the receptor for the neurotrophins brain derived neurotrophic factor (BDNF) and neurotrophin 4/5, whereas other neurotrophin family members preferentially bind to other trk receptors, nerve growth factor to trkA and neurotrophin-3 to trkC (3–6).

The analysis of its cDNA has revealed that TrkB contains several protein domains common to other receptor tyrosine kinases (7–10). The extracellular domain of the receptor contains two cysteine rich regions separated by a leucine rich domain. Between those domains and the plasma membrane are two IG like regions that are followed by the transmembrane domain. Intracellular, the receptor contains the juxtamembrane domain that includes a Shc binding site, a tyrosine kinase domain and a tail region containing a PLC-gamma binding site.

Northern blot analysis of TrkB messages has indicated that a number of mRNAs are generated from the TrkB locus (7) which is located on human chromosome 9 (11). Molecular cloning revealed the existence of several isoforms generated by alternative splicing. In the intracellular region, two isoforms are generated by this mechanism: a full length receptor containing the tyrosine kinase domain and a truncated form lacking it (TrkB-T1) (7, 9, 12). In rodents, another truncated form (TrkB-T2) has been described, with an apparently identical splice site as that of TrkB-T1, but with a unique intracellular tail region (7). The TrkB-T1 form can bind to neurotrophins and is internalized upon ligand binding (13) but due to the lack of a kinase domain cannot signal to the cytoplasm. Its exact function remains controversial. However, accumulating evidence suggest that the truncated TrkB isoforms act as dominant negative form (14, 15), possibly suppressing neurotrophin action.

In addition to the alternative splicing of the intracellular regions, alternative splice forms lacking one or all of the leucine rich motifs in the extracellular region have been described (16). The isoforms lacking the leucine rich motifs can no longer bind neurotrophins (16). In chicken, additional splice variants were described, one in the extracellular juxtamembrane region with leads to a reduced affinity to NT 4/5 (17, 18) and others with unknown function in the cytoplasmic juxtamembrane region (18). The splicing variant with reduced affinity to NT4/5 found in chicken was also found to be expressed in cultured human retinal pigmented epithelium cells (19). Together, these data show that alternative splicing is used to generate molecules with different function from the TrkB locus. This is typical

Abbreviations used: TrkB, tyrosine receptor kinase B; BDNF, brain derived neurotrophic factor; TK, tyrosine kinase; PAC, phage artificial chromosome; RT-PCR, reverse transcriptase polymerase chain reaction; EST, expressed sequence tag; RTK, receptor tyrosine kinase; IG, immunoglobulin.

[1] Corresponding author. Fax: +49 9131-85-22484. E-mail: stefan@stamms-lab.net.

0006-291X/02 $35.00

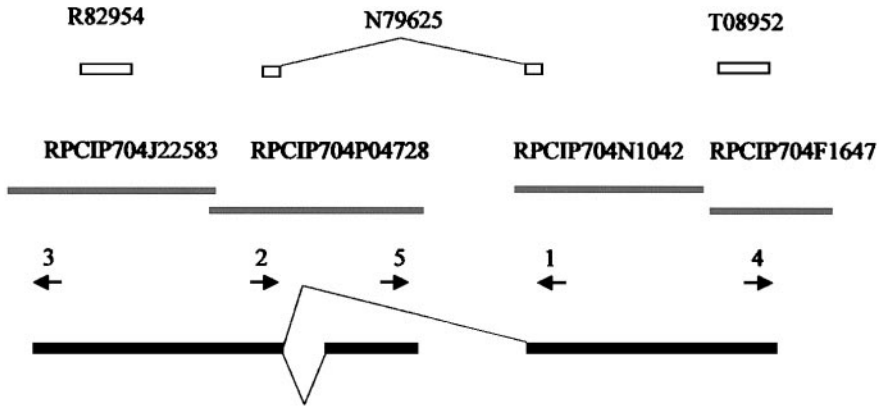

FIG. 1. Screening strategy. The three ESTs used to screen the PAC library are shown as empty boxes on the top. Numbers above the ESTs indicate their GeneBank accession numbers. EST N79625 hybridizes to two nonoverlaping genomic clones because it contains two exons that are separated by a large intervening sequence in the genomic DNA. Positive clones are indicated by gray lines and were verified by rescreening with primers TrkBseq1-5 (arrows). Numbers on top of the clones indicate their names from the German resource center in Berlin. A schematic gene structure derived from the available cDNA data is schematically shown below using black lines that correspond to the major mRNA isoforms encoding the full length TrkB and the TrkB-T1 receptors.

for higher eukaryotic gene regulation where about 60% of all genes are regulated by this mechanism (20–22). Since alternative splicing pathways can be regulated by extracellular signals (23), e.g. neuronal activity (24) the formation of splice variants is most likely important for neurotrophin signaling through TrkB.

As a first step to investigate the molecular mechanisms generating TrkB isoform diversity, we determined the gene structure of the human TrkB gene. We found that this gene is unusually large, contains a noncanonical intron and an extended repeat. The gene can produce 100 mRNA isoforms and 10 different proteins. However three forms, the full length receptor and two truncated receptors, one of which has a Shc binding site, are the major ones.

MATERIALS AND METHODS

Screening and sequencing. The human BAC library screen using ESTs matching the human TrkB cDNA sequence was kindly performed by the German Human Gene Center, Heidelberg. Additional library screens was performed using PCR products generated using primers TrkB-E18-F and TrkB-E19-R as a probe. Double stranded sequencing of the BAC clones was performed on an ABI Prism 370 sequencer using TrkB specific oligonucleotides. The sequencing protocol is from http://www.chori.org/bacpac/cyclesere.htm. Additional sequences were obtained by transposone insertion sequencing using the GPS-1 kit (NEB).

Cloning of the human full length and TrkB-T-Shc receptors. The full length human receptor and human TrkB-T-Shc isoform were amplified from SH-SY5Y RNA using Pfx DNA polymerase and cloned in pEGFP-N1 by standard procedures.

RT-PCR. Splicing isoforms were detected by RT-PCR using the primers listed at http://www.stamms-lab.net/pub_data/TrkB_splicing_primers.txt. Total RNA from human cortex was obtained from the National Brain Bank, Netherlands. Total RNA from SHSY-5Y cells was isolated using an RNeasy kit (Qiagen). Two micrograms RNA were used for first strand synthesis. One tenth of the RT reaction was later used as template for PCR. The PCR products were gel purified, cloned in a pCR 4 TOPO vector (Invitrogen) and sequenced.

Northern blot analysis. The primer pairs used for generating DNA probes can be found under http://www.stamms-lab.net/pub_data/TrkB_splicing_primers.txt. Double stranded probes were labeled either by random priming (Megaprime, Amersham) or when the probes were shorter than 200 bp by PCR. For the PCR labeling, 10 to 100 pg DNA was amplified for 15 cycles in the presence of 10 pmol of each primer, 10 μM dATP, dGTP, dTTP and 50 μCi γ-P32-dCTP (3000 Ci/mmol). dNTPs were added to final concentration of 150 μM and 5 additional cycles were performed. Single stranded DNA probes were generated by cycle extension using gel purified PCR products as a template. One nanogram of DNA was amplified for 35 cycles on presence of 10 pmol primer, 10 μM dATP, dGTP, dTTP and 50 μCi γ-P32-dCTP (3000 Ci/mmol). All probes were purified from the labeling mix using a dNTP removal kit (Qiagen). A human multiple tissue Northern blot (Clontech) was probed according to the manufacturer's instructions.

Computer analysis. Analysis was done using the GCG package (25). Contig assembly was performed using the GCG or the Staden Package. Calculation of splice site scores was as described (26). This calculation can be performed at http://www2.imcb.osaka-u.ac.jp/splice/score.html.

The human genome was downloaded from the NCBI ftp site. We utilized scripts written in Perl to extract the introns from the genomic DNA. The intron locations were determined using the information in the mRNA field in the GeneBank file headers. The perl scripts are available for download from http://stamms-lab.net/pub_data/.

Immunoprecipitation. Immunoprecipitations were performed as described (27).

RESULTS

Isolation of the Human TrkB Gene

To isolate clones corresponding to the TrkB gene, we used three ESTs from the 5′, 3′ and middle parts of the TrkB cDNA to probe a human genomic PAC library. 28 independent clones were obtained and verified by hybridization with TrkB specific oligonucleotides (Fig. 1).

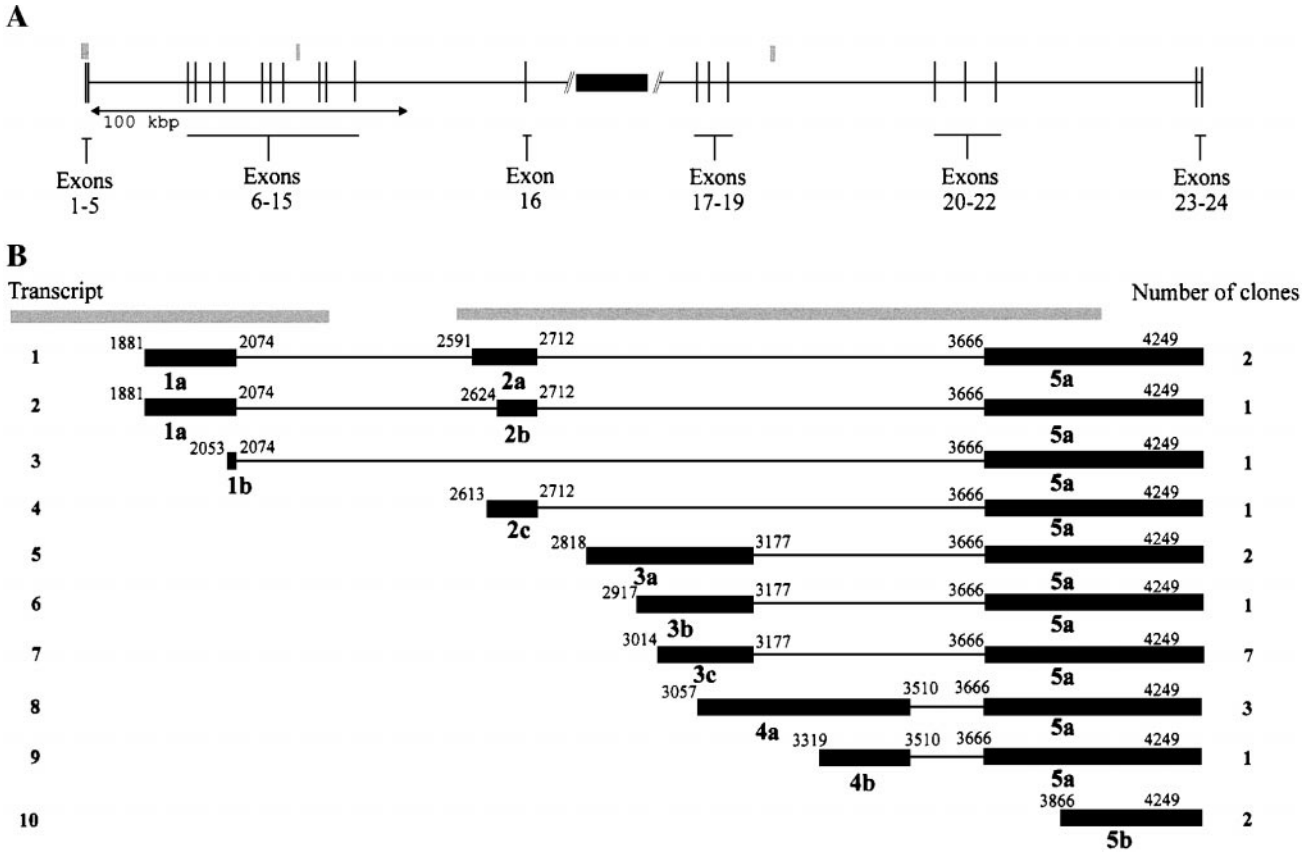

FIG. 2. Structure of the TrkB gene and its 5′ UTR. (A) The structure of the TrkB gene is drawn to scale. Exons are indicated as vertical lines, introns are shown as horizontal lines. The ChAB4 repeat is indicated by a black box. The gap in the sequence surrounding the ChAB4 repeat is indicated by a double tilted line. CpG islands are indicated as gray boxes on top. (B) Structure of the 5′ end of the TrkB gene. The first five exons are indicated as black boxes. Numbers indicate the exon positions in the genomic sequence (accession number AF410902). The CpG islands are indicated with gray boxes on the top.

Two of them were found to be false positives and one contained an internal recombination. Four clones (RPCIP704J22583, RPCIP704P04728, RPCIP704N1042 and RPCIP704F1647) covering the entire known mRNA sequence were obtained from the German resource center in Berlin and characterized further. The clones were sequenced by primer walking starting from the known exons. The list of primers can be obtained under http://www.stamms-lab.net/pub_data/trkB_sequencing_primers.txt.

Structure of the Gene

Sequencing of the gene shows that it contains 24 exons and 23 introns (Fig. 2A). One remarkable feature of the gene is its large size. The gene spans at least 590 kbp. The exons are located in five clusters. The first cluster contains the putative promoter region and the exons 1–5. The start codon is present in exon 5. Approximately 30 kbp downstream of this exon is the second exon cluster comprised of exons 6–15. These exons encode the extracellular part of the protein, the transmembrane region and part of the juxtamembrane domain. Exon 16 which is used in the truncated receptor (TrkB-T1) is located 50 kbp downstream. A large intron of at least 300 kbp separates this exon from the third cluster. The third cluster is composed of exons 17–19. Exons 17 and 18 encode the intracellular juxtamembrane part of the receptor. Exon 19 is an alternative terminating exon which is used in a new truncated receptor form (TrkB-T-Shc). The fourth cluster containing the exons coding for part of the kinase domain is located approximately 60 kbp downstream. It is separated by a 50 kbp intron from the last two exons that compose the carboxy terminal part of the kinase domain and the phospho lipase C gamma binding site. The exons vary in size from 36 to more than 6000 nt with median size around 100 nt. The introns range from 115 nt to more than 300 kbp. Another unusual feature of the TrkB gene is the presence of a ChAB4 repeat between exon 16 and 17. This 250 kbp large repeat is found about 50 times in the human genome (28). The translational phase of the exons changes within the gene: three exons are in first, eight exons in second phase and seven in the third phase (Fig. 4). The

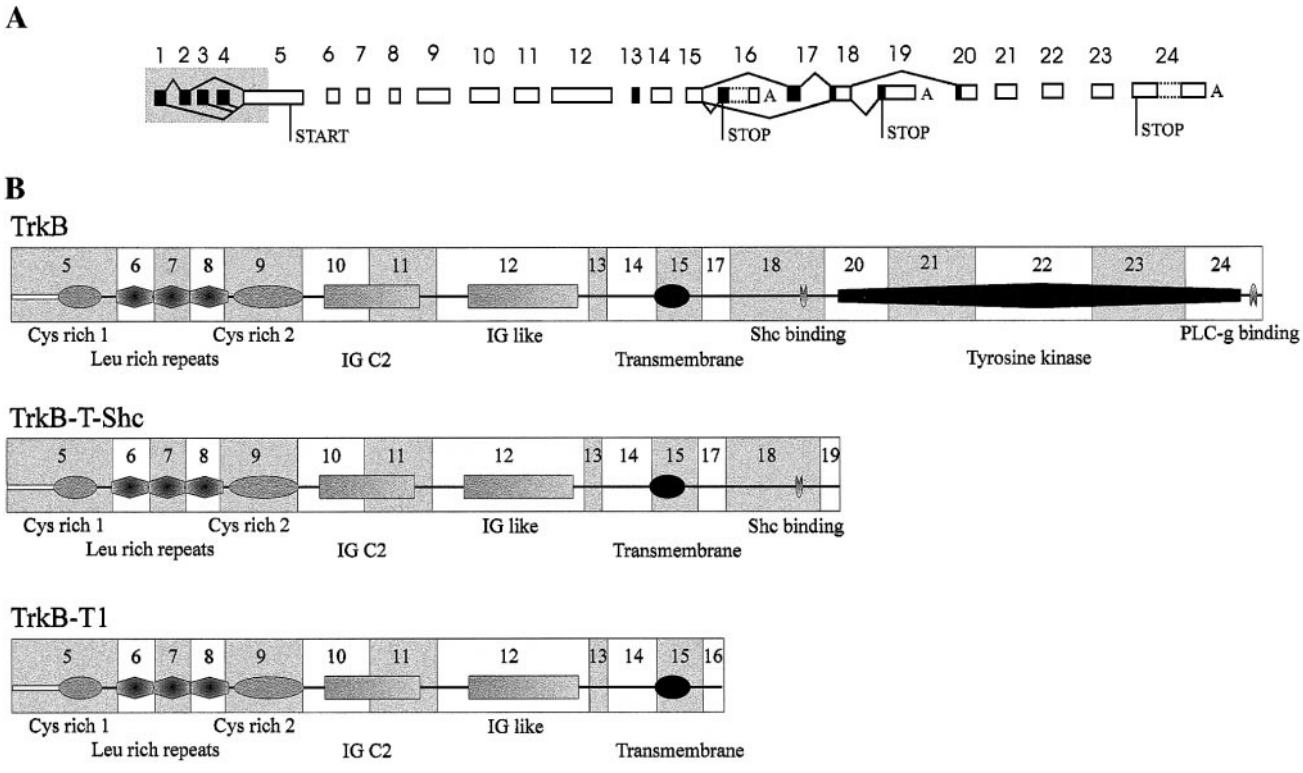

FIG. 3. Splicing pattern of the TrkB gene and structure of the generated proteins. (A) Alternative splicing patterns of the TrkB gene. Exons are shown as boxes and numbered. Constitutive exons are shown as white boxes, alternative exons are shown as black boxes. Alternative splicing patterns are indicated by lines. The area marked in gray is enlarged in Fig. 2B. Dashed boxes indicate incomplete exon sequence. (B) Structure of the TrkB proteins. The domain structure of the TrkB isoforms is drawn to scale. The domain types are indicated below. Alternating shading shows the positions of the exons.

Grail program (29) detects four CpG islands. Two of them are located in the promoter region. Two more CpG islands are found in introns 12 and 19 (Figs. 2A and 2B).

Structure of the 5′ UTR

In order to analyze the 5′ untranslated region of the gene, we amplified capped RNA using the generator method (Invitrogen). This technique uses subsequent dephosphorylation, decaping and RNA oligonucleotide ligation to amplify only capped pre-mRNA molecules, which results in the detection of putative transcription start sites (30–32). We employed this method on human RNA derived from SH-SY5Y cells. Sequencing of the clones revealed that the TrkB gene is transcribed from at least nine different start sites. Based on the number of clones, the transcriptional start at position 3014 (Fig. 2B) is the most common one. The putative promoter region shows no apparent initiator elements (i.e. TATA box) and is highly G/C rich. Computer prediction indicated the presence of two CpG islands in the promoter region (Fig. 2B). We conclude that the TrkB promoter shows hallmarks of housekeeping genes and uses several transcriptional start sites.

The 5′ UTR Contains Several Alternative Exons, One with Unusual Splice Sites

In addition to alternative transcription start site usage, diversity at the 5′ UTR is generated by alternative splicing. Exon 2 can be alternatively spliced, using two different 3′ splice sites. Furthermore, exons 3 and 4 do not contain 3′ splice sites and therefore are not joined to exons 1 or 2. A remarkable feature of the 5′ UTR is the presence of a noncanonical exon. Exon 4 uses a GC rather then a GT dinucleotide in the 5′ splice site of its downstream intron. We searched the annotated genes in the human genome at NCBI for more introns containing such splice site. The search resulted in 832 examples which is 1% of all annotated introns. This introns can be found under ftp://stamms-lab.net/pub/EID.

Alternative Splicing Patterns

We systematically searched the EST database with cDNA and genomic sequences form the TrkB locus and used RT-PCR and RACE to identify novel exons. We identified exons 1, 2, 3, 4, 17 and 19 that have not been previously described (Figs. 2B and 3A). Two of them

Exon	Size	3' splice site /*Transcriptional start site*		5' splice site		Intron size
		Site	Score	Site	Score	
1a	194	*ccccgcagccATCATTTAACT....*		....ACGCGCAAG/gtacgg	8.4	517 to Exon 2a
						550 to Exon 2b
1b	22	*cgcacaccctAGCACACATG*				1592 to Exon 5a
2a	122	cgcatttttcag/AGCTGAACCA....	6.5	....CATCTTAGAG/gtacct	5.2	1536
2b	89	catttcaaaaag/GGAGACAGCC....	1.1	....CATCTTAGAG/gtacct		
2c	100	*cacggtttccATTTCAAAAA*		....CATCTTAGAG/gtacct		
3a	360	*gtgaactcccACATGCTGCT....*		....CGGAGCCCGG/gtgagc	7.9	584
3b	261	*tttgtctggaGGGTGTTATG....*		....CGGAGCCCGG/gtgagc		
3c	165	*cagccctcacGTCACTTCGC....*		....CGGAGCCCGG/gtgagc		
4a	454	*gcggctcccgGAATTGGGTT....*		....CCATCGCGGG/gcaagt	−1.3	251
4b	192	*attctgcagcATCATTCGGG....*		....CCATCGCGGG/gcaagt		
5a	584	gttttgtcctcag/CCTCGAGGTG....	7.0	**....TC ACC GAA AT/gtgagt**	8.9	31 198
5b	351	*ctgccggaacACTCTTCGCT....*		**....TC ACC GAA AT/gtgagt**		
6	75	gctttgttacag/**T TTC ATC GCA**....	5.9	....**TG AGA AAT CT**/gtgagt	5.4	115
7	72	gtgttttcacag/**G ACA ATT GTG**....	8.5	....**TG CAG CAC AT**/gtaagt	9.0	5 431
8	69	gattcctttcag/**C AAT TTT ACC**....	6.3	....**TG TCT GAA CT**/gtaagt	7.0	2 728
9	155	tgttccctgtag/**G ATC CTG GTG**....	10.0	....**CCC AAT TGT G**/gtaatt	6.0	12 782
10	137	cttgttccatag/**GT TTG CCA TC**....	9.3	....**CC AAA CAA TG**/gtaagg	9.1	513
11	133	attcatttgtag/**A AAT GAA ACA**....	4.4	....**ACT GTG CAT T**/gtacgt	4.4	3 296
12	306	ttttcaatttag/**TT GCA CCA AC**....	4.2	....**ATT GAC GAT G**/gtgagt	10.4	13 923
13	36	tttgccttttag/**GT GCA AAC CC**....	9.4	....**AAT TTA TAA G**/gtagct	5.9	3 046
14	101	tgcccacttaag/**AT TAT GGA AC**....	2.4	....**A CAT CTC TCG**/gtgagt	8.9	6 913
15	100	tcctttctctag/**GTC TAT GCT G**....	11.2	...**GGC ATG AAA G**/gtaaga	11.3	56 400 to Exon 16
						>110 584 to Exon 17
16	~6000	tgtttcttttag/**GT TTT GTT TT**....	9.4	Putative terminating exon		
17	48	tgtggttttcag/**AT TTC TCA TG**....	10.2	....**CAA GGT GTT G**/gtaagt	9.2	6 156
18	189	ttccatctccag/**GC CCA GCC TC**....	6.1	....**CCA GAC ACA T**/gtaagt	9.0	4 358 to Exon 19
						66 731 to Exon 20
19	6 109	actttatgcag/**GG CCC AGA GG**....	7.5	Terminating exon		
20	131	tttgttttgcag/**TT GTT CAG CA**....	8.6	....**TG GCA GTA AG**/gtaaga	11.3	14 170
21	174	cacccatccccag/**A ACC CTG AA**....	5.7	....**AG TTC CTC AG**/gtacag	4.8	6 649
22	234	gtccttccccag/**G GCA CAC GG**....	10.0	....**C TAC TAC AGG**/gtgagt	10.2	64 689
23	159	ctttctccccag/**GTC GGT GGC C**....	13.3	....**A AAC AAT GAG**/gtgtgc	6.7	888
24	~6000	atctccatccag/**GTG ATA GAG T**....	7.9	Putative terminating exon		

FIG. 4. Compilation of exon–intron boundaries. Protein coding exon sequences are indicated by boldface upper-case letters, untranslated exon sequences by upper-case letters, and intronic sequences by lower case letters. Alternative exons are underlined. The amino acid usage is indicated by spaces between the triplets. Exon and intron sizes are indicated in nt. 5′ and 3′ scores (21) are given for each splice site. A perfect match to U1 snRNA would give a theoretical maximum 5′ score of 12.6 and the average score of constitutive exons is 8.0. Likewise, the theoretical maximum 3′ score is 14.2 and the average score for constitutive exons is 9.4. Transcriptional start sites are in italic letters; nucleotides found in the cDNA clones are indicated in capital italic letters.

are cassette exons (exons 2 and 17), one has alternative 3′ splice site (exons 2a and 2b), one exon is terminating with alternative polyadenylation site (exon 19) and three exons are generated from alternative transcription start sites (exons 1, 3 and 4). When the nine different start sites are taken into account, the TrkB gene could generate 100 different mRNA isoforms which code for 10 different proteins. The protein domain composition is often reflected in the gene structure. In the human trkB gene, several protein domains

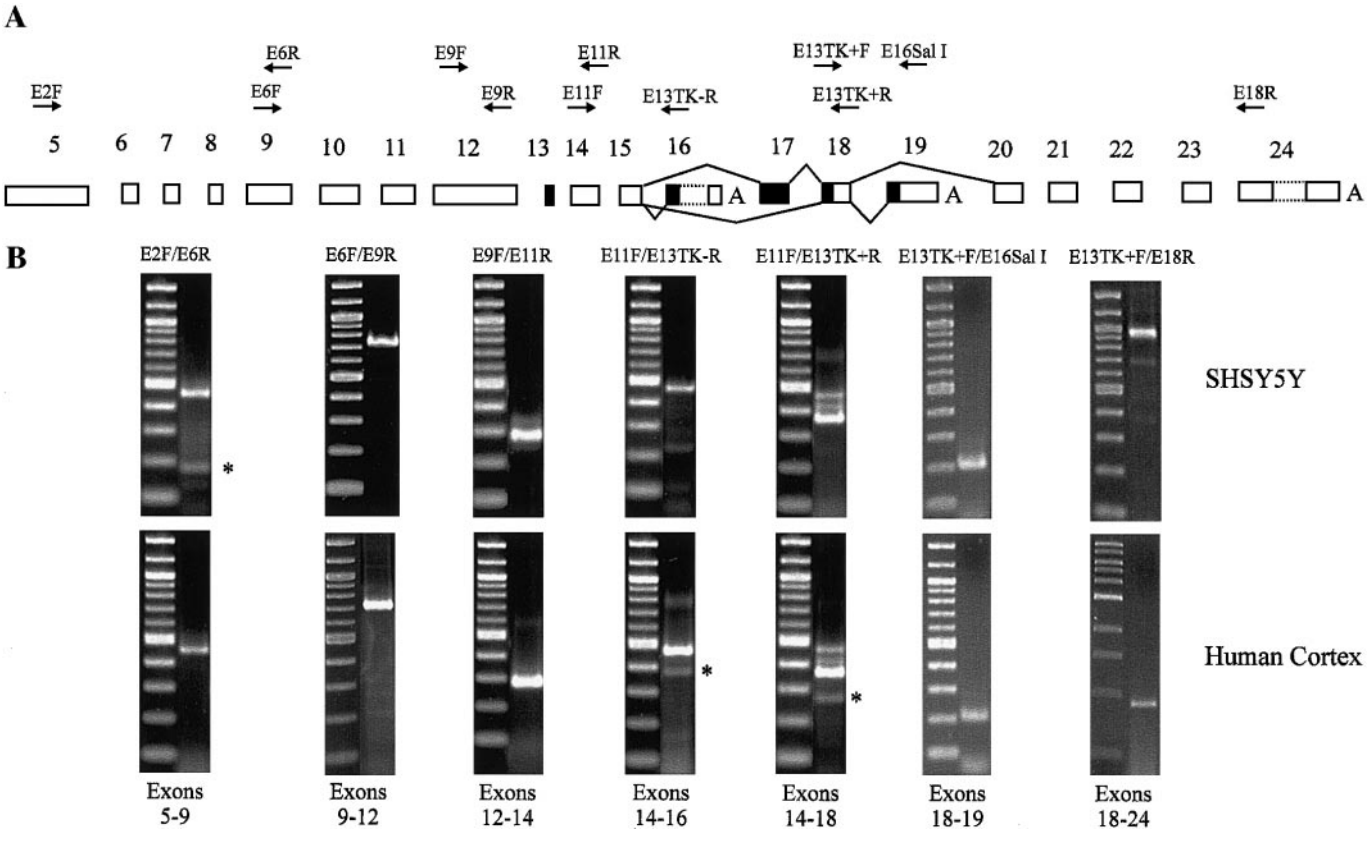

FIG. 5. Expression of the alternative splice variants in SHSY5Y cells and human cortex. (A) Partial gene structure of TrkB with primer pairs used for RT-PR indicated on top. (B) RT-PCR with cDNA derived from human cortex and SH-SY5Y cells. All PCR products were subcloned and verified by sequencing, which showed that the minor bands indicated with stars were artificial. The marker represents a 100 bp ladder in all experiments, only in human cortex using primers E13TK+F/E18R a 1 kb ladder was used. The amplified exons are shown at the bottom.

are encoded by a single exon. Exon 5 encodes the cysteine rich domain and each of exons 6 to 8 encodes a leucine rich motif. The IG C2 and IG like domains are formed by exons 10 and 11, well as exon 12, respectively. The transmembrane domain corresponds to exon 15 and the tyrosine kinase domain is composed of exons 20 to 24. The isoforms with different polyadenylation sites are formed by usage of the alternative exons 16 and 19 (Fig. 3B). This shows that the overall protein structure is reflected in the gene structure.

Alternative exon usage is often a result of suboptimal splice sites. In order to assess the quality of the splice sites of the alternative exons, we calculated their score. The score expresses how well the splice site follows the consensus sequence and binds to the components of the spliceosome (Fig. 4). With the exception of exon 4a, no strong deviations from the consensus were found in alternative exons. Exon 4a is described by a weak score, because it uses a noncanonical GC rather than a GT. The good match of the alternative splice sites with the mammalian consensus sequence indicates that their regulation might involve repressing, rather than activating splicing regulatory factors.

Expression of the Splicing Variants

To address the question how abundantly the individual splice variants are expressed, we amplified regions of the pre-mRNA containing alternative exons. As RNA template, we used RNA derived from human cortex and from the SH-SY5Y human neuroblastoma cell line (33). As shown in Fig. 5, the newly identified exon 17 is used in both human cortex and SH-SY5Y cells. In contrast, all other cassette exons represent minor forms. We conclude that part of the TrkB juxtamembrane region that is encoded by exon 17 is regulated by alternative splicing.

Isoform Diversity Is Generated by Usage of Different Polyadenylation Sites

Next, we asked how strong the exons with alternative polyadenylation sites are used. We investigated the presence of exon 24, 16 and 19 containing messages using Northern blot analysis of human RNA. We found that the TrkB gene generated three major isoforms, the full length tyrosine receptor kinase (TrkB), a form lacking the catalytic domain (TrkB-T1) and a form

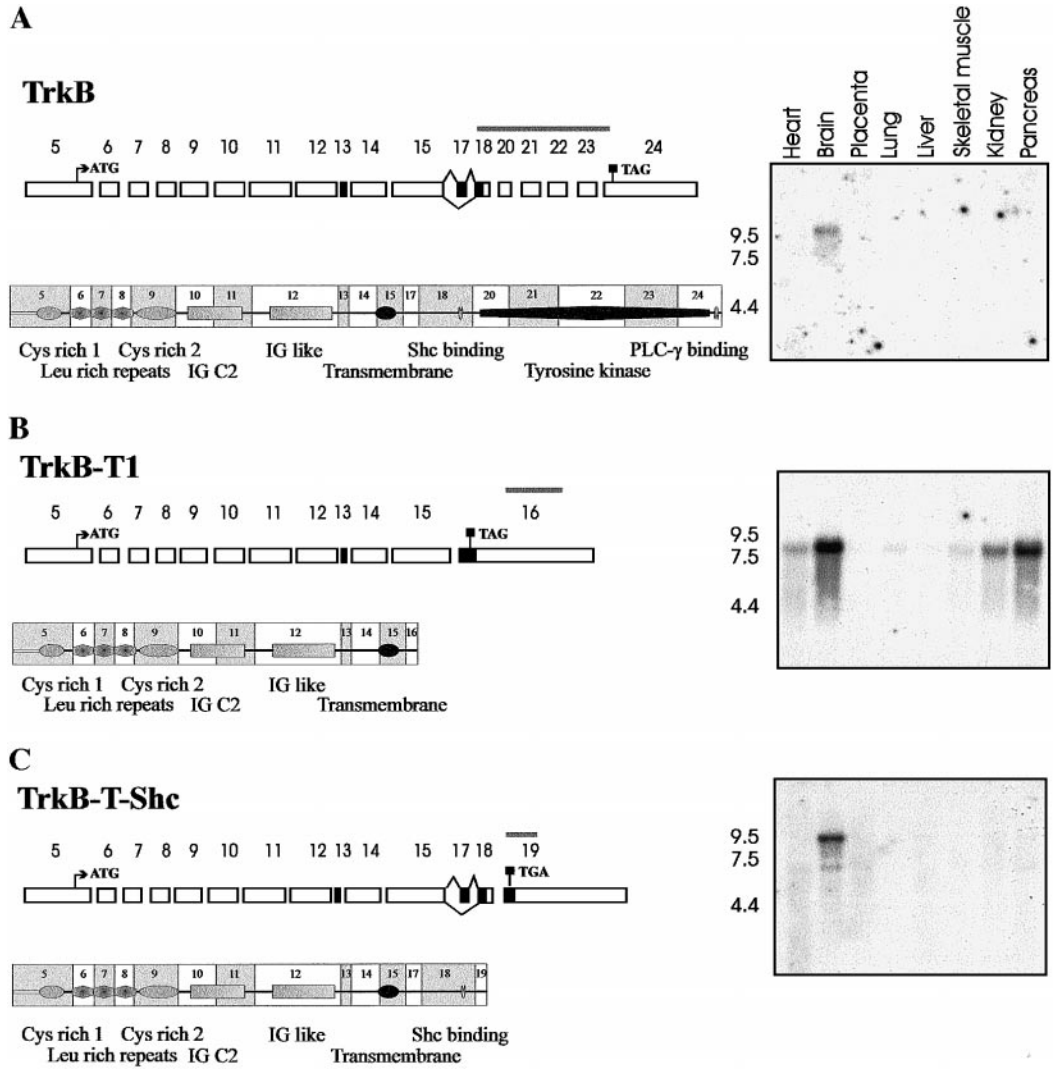

FIG. 6. Protein structure and RNA expression pattern of the major TrkB variants. The structure of the coding mRNA part is indicated on the left and the corresponding domain structure of the resulting protein is indicated below the RNA. The expression pattern of the mRNA in humans was determined by Northern blot analysis and is shown on the right for each major TrkB variant. The cDNA probe used to detect the expression pattern is shown as a gray bar above the RNA structure.

lacking the catalytic domain but possessing a Shc site (TrkB-T-Shc). The full length TrkB form uses the stop codon and polyadenylation site present in exon 24 and is predominantly expressed in brain (Fig. 6A). The previously described TrkB-T1 form is expressed in multiple tissues. However, expression is highest in brain, pancreas, kidney and heart (Fig. 6B).

TrkB-T-Shc, a Novel TrkB Isoform Predominantly Expressed in Brain

When we searched the EST database with each individual exon sequence, we detected an EST (accession number AL134306.1) containing a novel alternative spliced exon (exon 19). We obtained the EST clone from the German Resource center in Berlin and subsequent sequencing showed that it contains part of exon 18 and a large single exon (exon 19) ending with a poly A tail (accession number AF410898). Since exon 19 contains a stop codon and a polyadenylation site, its usage leads to a truncated protein (accession number AF410901 and accession number AF410900), similar to the TrkB-T1 isoform. However, in contrast to the TrkB-T1 isoform, this isoform results in a protein that retains the Shc binding site (Fig. 6C). Northern blot analysis reveals that this new isoform is predominantly expressed in the brain (Fig. 6C). Therefore, there are two different classes of tyrosine kinase deficient TrkB receptors in human brain. The TrkB-T1 form which is also found in other tissues can not signal to Shc and the TrkB-T-Shc which is brain specific and could bind to Shc.

TrkB-T-Shc Is Present in Mouse

To determine whether the newly found exons 17 and 19 are present in rodents we screened the ENSEMBL trace repository (34). Several traces were detected with exon 19 sequence and were assembled using the Staden Package (http://www.mrc-lmb.cam.ac.uk/pubseq/). The alignment with the human TrkB sequence shows high similarity, which is not only confined to the coding sequence but extends into the 3′ UTR and the upstream intron (Fig. 7A). RT-PCR analysis showed brain specific expression of exon 19 (Fig. 7B). Interestingly it is not detectable in primary astrocytes but is present in cultured neurons from embryonic cortex (Fig. 7B), mesencephalon and superior cervical ganglia (data not shown). The RT-PCR and the trace repository screens do not show any evidence for a mouse exon that will correspond to the human exon 17.

TrkB-T-Shc Is Not Phosphorylated by Full Length TrkB

We wondered whether the novel TrkB-T-Shc isoform could be part of the established TrkB signaling pathway. As a first step we determined whether it interacts with the full length TrkB isoform. We therefore cotransfected Flag-tagged full length TrkB with EGFP tagged TrkB-T-Shc. Immunoprecipitation with anti-Flag-antibody followed by Western blot analysis with anti-EGFP antibody revealed that both forms coimmunoprecipitate (Fig. 8A). As the Shc site of the full length receptor is phosphorylated upon its activation we were interested if the Shc site of TrkB-T-Shc can be phosphorylated by the catalytic variant. We therefore cotransfected human full length TrkB receptor cDNA with TrkB-T-Shc expression constructs. The expressed protein was immunoprecipitated with anti-EGFP antibody and the immunoprecipitates were analyzed for tyrosine phosphorylation. Whereas the full length receptor was tyrosine phosphorylated, no phosphorylation of the TrkB-T-Shc variant is detectable (Fig. 8B).

TrkB-T-Shc Localizes to the Plasma Membrane

Since TrkB-T-Shc differs in its phosphorylation properties form TrkB, we next wished to determine its intracellular distribution. As shown in Fig. 9, EGFP tagged TrkB-T-Shc localizes predominantly to the membrane and is also present inside the cell. We conclude that TrkB-T-Shc is present in the same cellular compartment as the full length receptor and could therefore interfere with its function.

DISCUSSION

General Structure of the TrkB Gene

Here we describe the characterization of the human TrkB gene. We sequenced approximately 300 kbp of the gene and the latest assembly of the human genome estimates the gene size to be about 350 kbp. However, this sequence contains two clone gaps. Our results show the presence of a ChAB4 repeat at the gap location (Fig. 2). ChAB4 repeats are characteristic for humans and chimpanzee. Only a few copies of these repeats are found in gorilla and orangutan. ChAB4 repeats have been estimated to be about 240 kbp in size. This suggest a gene size of at least 590 kbp. One large intron, located between exons 16 and 17 accounts for at least half its size. This is well above the average for the human genes that have a median of 14 kbp and a mean of 27 kbp in their size distribution. The presence of the ChAB4 repeat is another unusual feature, as there are only about 50 such repeats present on ten different human chromosomes (35). The TrkB gene consists of 24 exons which can be grouped into six clusters. Our analysis of the 5′ UTR showed that the gene uses at least nine different start sites. The promoter is located in a CpG island. No apparent initiator elements could be detected. The combination of alternative exon and promoter usage generates TrkB mRNAs with considerable variation in their 5′ UTRs. The alternative promoters do not create protein isoforms. The exact function of this divergence remains to be determined, but alterations in the 5′ UTRs are generally associated with differences in translation efficiency (36). The overall gene structure of the human TrkB gene is similar to the gene structure of TrkC (37). Furthermore, similar to the mouse TrkB gene (38) and the human TrkC gene (37) it uses TATA-less promoters located in CpG islands.

Alternative Splicing Variants

A combination of alternative splicing, alternative polyadenylation and alternative promoter usage leads to 100 isoforms that can encode 10 different proteins. Our RT-PCR analysis reveals that most of these alternative exons are not abundantly used, suggesting that they might fulfill specialized roles only in small cell populations or during certain developmental stages. Transcription and splicing occur simultaneously in a large complex, termed transcriptosome or “RNA factory” (39–41). Most alternative spliced exons of the TrkB gene are flanked by splice sites following the human consensus, indicating that their alternative use is due to repression mechanisms that remain to be elucidated. The only exon with suboptimal splice sites is exon 4 that uses a GC instead of the canonical GT at its 5′ splice site. A recent survey found 162 GC-AG introns derived from 145 genes (42). Our own survey found 832 GC-AG introns in the human genome which corresponds to about 1% of all annotated introns. The exact splicing mechanism of those introns remains to be determined. It could involve either optimized base pairing at the GC 5′ splice site to U1 snRNA, or different mechanism, such as changing GC to GU by an editing

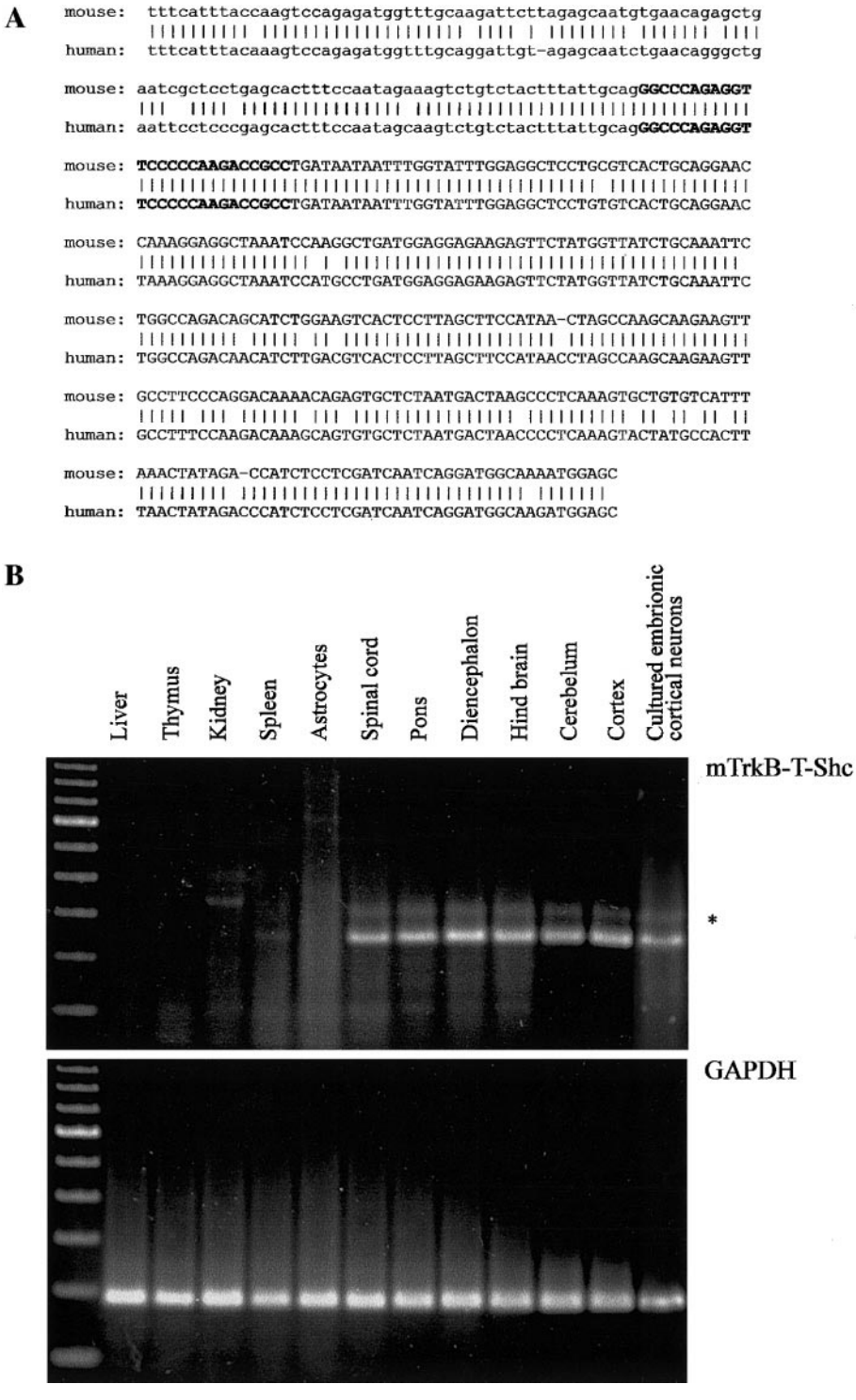

FIG. 7. TrkB-T-Shc is present in mouse. (A) Alignment between mouse and human sequences in the beginning of exon 19. The intron sequence is in lower-case, the exon sequence is in upper-case, and the coding part of the exon is shown in bold. (B) RT-PCR analysis of TrkB-T-Shc expression in mouse (top). The isoform is expressed in primary neurons from embryonic cortex and in cortex, cerebellum, hind brain, diencephalon, pons and spinal cord. It is absent from liver, thymus, kidney, spleen and cultured primary astrocytes. GAPDH was amplified from the same RNA samples to show that approximately equal amounts of RNA were used for the RT-PCR (bottom). The band marked with an asterix is a PCR artifact.

event. The large size of the TrkB introns makes it likely that introns will be removed stepwise, as found for the ultrabithorax genes (43). It could also be used to regulate the alternative splicing pattern of the gene, as three of the largest introns (15, 16 and 19) separate the alternative exons that generate the three major isoforms.

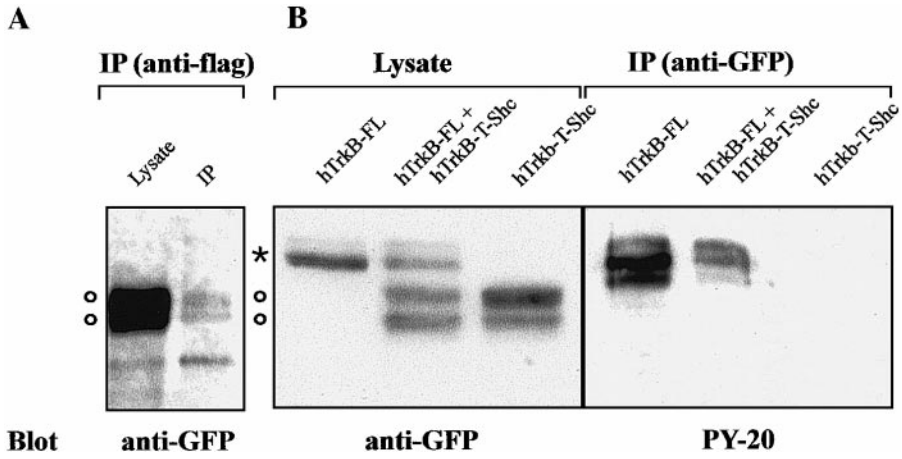

FIG. 8. The TrkB-T-Shk isoform is not tyrosine phosphorylated. TrkB-T-Shc expression constructs were cotransfected into 293 cells in the presence or absence of full length TrkB expression clones. (A) The full length TrkB receptor binds to TrkB-T-Shc. Expression clones for flag tagged TrkB-full length and EGFP tagged TrkB-T-Shc were cotransfected in 293 cells and proteins were immunoprecipitated using anti-flag. Using anti EGFP antibody, the presence of TrkB-T-Shc in lysates (left) and immunoprecipitates (right) can be detected. (B) EGFP tagged TrkB full length protein and EGFP tagged TrkB-T-Shc were used. TrkB variants were immunoprecipitated with anti-EGFP antibody and tyrosine phosphorylation was determined with the anti-phospho-tyrosine antibody (right). The full length and TrkB-T-Shc forms can be identified by their migration behavior and are indicated with a star (TrkB) and a circle (TrkB-T-Shc). Proper expression of the forms was detected by probing the lysates (left) with anti-EGFP antibody.

Novel Isoform

Similarly to TrkC (44–46), the TrkB gene generates two types of noncatalytic isoforms, which are conserved in human and rodents. The usage of exon 19 results in a new truncated TrkB isoform TrkB-T-Shc. In contrast to the previously described truncated isoforms TrkB-T1 and TrkB-T2 this variant has a Shc binding site. Another difference is its expression pattern. The formerly described TrkB-T1 variant is expressed in most tissues, but the novel TrkB-T-Shc isoform is expressed only in brain and is neuron specific. Sequence analysis shows that this isoform is highly conserved between humans and mouse. This indicates that it could play an important role in TrkB signal transduction. We therefore determined its ability to be phosphorylated by the full length TrkB receptor. To our surprise, the full length receptor could bind to the TrkB-T-Shc isoform, but was unable to phosphorylate it. Overexpressed TrkB-T-Shc is present in the plasma membrane and similar to the full length TrkB receptor accumulates inside the cell. Since the TrkB-T-Shc isoform is present only in neurons, it might regulate the function of the catalytic form of the TrkB receptor. It could act as a dominant negative form in regards to the full length receptor the same way the TrkB-T1 isoform does. Alternatively it could have a specific function in

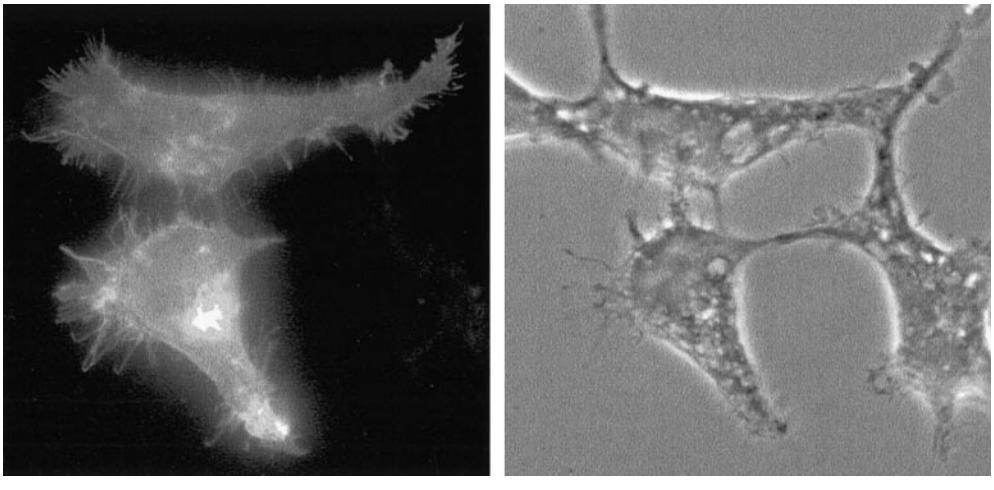

FIG. 9. TrkB-T-Shc localizes to the cell membrane. EGFP tagged TrkB-T-Shc was cotransfected in 293 cells and cells were analyzed by fluorescence microscopy. Protein can be detected on the cell membrane and in the cytoplasm. Left: EGFP fluorescence; right: phase contrast.

brain allowing for regulation of the TrkB receptor by other kinases. This hypothesis is supported by the finding that TrkC paralogues of TrkB-T-Shc can promote neuronal differentiation in cooperation with $p75^{NTR}$ (47). Furthermore, the existence of a brain specific truncated isoform raises the question about the function of the TrkB-T1 isoform that is present in most tissues.

ACKNOWLEDGMENTS

This work was supported by grants from the European Union (ERBIO4CT980259) and the Deutsche Forschungsgemeinschaft (Sta3993-1 and SFB473/C8). We are grateful to Joost Verhaagen and Daniela Glockenhammer for helpful discussion and reagents.

REFERENCES

1. Robinson, D. R., Wu, Y.-M., and Lin, S.-F. (2000). The protein tyrosine kinase family of the human genome. *Oncogene* **19,** 5548–5557.
2. Blume-Jensen, P., and Hunter, T. (2001). Oncogenic kinase signalling. *Nature* **411,** 355–365.
3. Thoenen, H. (1995). Neurotrophins and neuronal plasticity. *Science* **270,** 593–598.
4. Lewin, G. R., and Barde, Y. A. (1996). Physiology of the neurotrophins. *Ann. Rev. Neurosci.* **19,** 289–317.
5. Patapoutian, A., and Reichardt, L. F. (2001). Trk receptors: Mediators of neurotrophin action. *Curr. Opin. Neurobiol.* **11,** 272–280.
6. Bibel, M., and Barde, Y. A. (2000). Neurotrophins: Key regulators of cell fate and cell shape in the vertebrate nervous system. *Genes Dev.* **14,** 2919–2937.
7. Middlemas, D. S., Lindberg, R. A., and Hunter, T. (1991). trkB, a neural receptor protein-tyrosine kinase: Evidence for a full-length and two truncated receptors. *Mol. Cell. Biol.* **11,** 143–153.
8. Allen, S. J., Dawbarn, D., Eckford, S. D., Wilcock, G. K., Ashcroft, M., Colebrook, S. M., Feeney, R., and MacGowan, S. H. (1994). Cloning of a non-catalytic form of human trkB and distribution of messenger RNA for trkB in human brain. *Neuroscience* **60,** 825–834.
9. Shelton, D. L., Sutherland, J., Gripp, J., Camerato, T., Armanini, M. P., Phillips, H. S., Carroll, K., Spencer, S. D., and Levinson, A. D. (1995). Human trks: Molecular cloning, tissue distribution, and expression of extracellular domain immunoadhesins. *J. Neurosci.* **15,** 477–491.
10. Schneider, R., and Schweiger, M. (1991). A novel modular mosaic of cell adhesion motifs in the extracellular domains of the neurogenic trk and trkB tyrosine kinase receptors. *Oncogene* **6,** 1807–1811.
11. Nakagawara, A., Liu, X. G., Ikegaki, N., White, P. S., Yamashiro, D. J., Nycum, L. M., Biegel, J. A., and Brodeur, G. M. (1995). Cloning and chromosomal localization of the human TRK-B tyrosine kinase receptor gene (NTRK2). *Genomics* **25,** 538–546.
12. Klein, R., Conway, D., Parada, L. F., and Barbacid, M. (1990). The trkB tyrosine protein kinase gene codes for a second neurogenic receptor that lacks the catalytic kinase domain. *Cell* **61,** 647–656.
13. Biffo, S., Offenhauser, N., Carter, B. D., and Barde, Y. A. (1995). Selective binding and internalisation by truncated receptors restrict the availability of BDNF during development. *Development* **121,** 2461–2470.
14. Haapasalo, A., Koponen, E., Hoppe, E., Wong, G., and Castren, E. (2001). Truncated trkB.T1 is dominant negative inhibitor of trkB.TK+-mediated cell survival. *Biochem. Biophys. Res. Commun.* **280,** 1352–1358.
15. Eide, F. F., Vining, E. R., Eide, B. L., Zang, K., Wang, X. Y., and Reichardt, L. F. (1996). Naturally occurring truncated trkB receptors have dominant inhibitory effects on brain-derived neurotrophic factor signaling. *J. Neurosci.* **16,** 3123–3129.
16. Ninkina, N., Grashchuck, M., Buchman, V. L., and Davies, A. M. (1997). TrkB variants with deletions in the leucine-rich motifs of the extracellular domain. *J. Biol. Chem.* **272,** 13019–13025.
17. Strohmaier, C., Carter, B. D., Urfer, R., Barde, Y. A., and Dechant, G. (1996). A splice variant of the neurotrophin receptor trkB with increased specificity for brain-derived neurotrophic factor. *EMBO J.* **15,** 3332–3337.
18. Garner, A. S., Menegay, H. J., Boeshore, K. L., Xie, X.-Y., Voci, J. M., Johnson, J. E., and Large, T. H. (1996). Expression of TrkB receptor isoforms in the developing avian visaul system. *J. Neurosci.* **16,** 1740–1752.
19. Hackett, S. F., Friedman, Z., Freund, J., Schoenfeld, C., Curtis, R., Di Stefano, P. S., and Campochiaro, P. A. (1998). A splice variant of trkB and brain-derived neurotrophic factor are co-expressed in retinal pigmented epithelial cells and promote differentiated characteristics. *Brain Res.* **789,** 201–212.
20. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921.
21. Stamm, S., Zhu, J., Nakai, K., Stoilov, P., Stoss, O., and Zhang, M. Q. (2000). An alternative-exon database and its statistical analysis. *DNA Cell Biol.* **19,** 739–756.
22. Tupler, R., Perini, G., and Green, M. R. (2001). Expressing the human genome. *Nature* **409,** 832–833.
23. Stoss, O., Olbrich, M., Hartmann, A. M., Konig, H., Memmott, J., Andreadis, A., and Stamm, S. (2001). The STAR/GSG family protein rSLM-2 regulates the selection of alternative splice sites. *J. Biol. Chem.* **276,** 8665–8673.
24. Daoud, R., Da Penha Berzaghi, M., Siedler, F., Hubener, M., and Stamm, S. (1999). Activity-dependent regulation of alternative splicing patterns in the rat brain. *Eur. J. Neurosci.* **11,** 788–802.
25. Womble, D. D. (2000). GCG: The Wisconsin Package of sequence analysis programs. *Meth. Mol. Biol.* **132,** 3–22.
26. Stamm, S., Zhang, M. Q., Marr, T. G., and Helfman, D. M. (1994). A sequence compilation and comparison of exons that are alternatively spliced in neurons. *Nucl. Acids Res.* **22,** 1515–1526.
27. Hartmann, A. M., Nayler, O., Schwaiger, F. W., Obermeier, A., and Stamm, S. (1999). The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in nuclear

dots is regulated by the Src family kinase p59(fyn). *Mol. Biol. Cell.* **10,** 3909–3926.

28. Wohr, G., Fink, T., and Assum, G. (1996). A palindromic structure in the pericentromeric region of various human chromosomes. *Genome Res.* **6,** 267–279.
29. Xu, Y., and Uberbacher, E. C. (1997). Automated gene identification in large-scale genomic sequences. *J. Comput. Biol.* **4,** 325–338.
30. Schaefer, B. C. (1995). Revolutions in rapid amplification of cDNA ends: New strategies for polymerase chain reaction cloning of full-length cDNA ends. *Anal. Biochem.* **227,** 255–273.
31. Maruyama, K., and Sugano, S. (1994). Oligo-capping: A simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138,** 171–174.
32. Volloch, V., Schweitzer, B., and Rits, S. (1994). Ligation-mediated amplification of RNA from murine erythroid cells reveals a novel class of beta globin mRNA with an extended 5′-untranslated region. *Nucl. Acids Res.* **22,** 2507–2511.
33. Biedler, J. L., Roffler-Tarlov, S., Schachner, M., and Freedman, L. S. (1978). Multiple neurotransmitter synthesis by human neuroblastoma cell lines and clones. *Cancer Res.* **38,** 3751–3757.
34. ENSEMBL traces repository. Available at http://trace.ensembl.org.
35. Kehrer-Sawatzki, H., Wohr, G., Schempp, W., Eisenbarth, I., Barbi, G., and Assum, G. (1998). Mapping of members of the low-copy-number repetitive DNA sequence family chAB4 within the p arms of human acrocentric chromosomes: Characterization of Robertsonian translocations. *Chromosome Res.* **6,** 429–435.
36. Gray, N. K., and Wickens, M. (1998). Control of translation initiation in animals. *Ann. Rev. Cell Dev. Biol.* **14,** 399–458.
37. Ichaso, N., Rodriguez, R. E., Martin-Zanca, D., and Gonzalez-Sarmiento, R. (1998). Genomic characterization of the human trkC gene. *Oncogene* **17,** 1871–1875.
38. Barettino, D., Pombo, P. M. G., Espliguero, G., and Rodriguez-Pena, A. (1999). The mouse neurotrophin receptor trkB gene is transcribed from two different promotors. *Biochim. Biophys. Acta* **1446,** 24–34.
39. Misteli, T., and Spector, D. L. (1998). The cellular organization of gene expression. *Curr. Opn. Cell Biol.* **10,** 323–331.
40. McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., and Bentley, D. L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385,** 357–361.
41. Corden, J. L., and Patturajan, M. (1997). A CTD function linking transcription to splicing. *Trends Biochem. Sci.,* 413–419.
42. Thanaraj, T. A., and Clark, F. (2001). Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucl. Acids Res.* **29,** 2581–2593.
43. Hatton, A. R., Subramaniam, V., and Lopez, A. J. (1998). Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon–exon junctions. *Mol. Cell.* **2,** 787–796.
44. Tsoulfas, P., Soppet, D., Escandon, E., Tessarollo, L., Mendoza-Ramirez, J. L., Rosenthal, A., Nikolics, K., and Parada, L. F. (1993). The rat trkC locus encodes multiple neurogenic receptors that exhibit differential response to neurotrophin-3 in PC12 cells. *Neuron* **10,** 975–990.
45. Menn, B., Timsit, S., Calothy, G., and Lamballe, F. (1998). Differential expression of TrkC catalytic and noncatalytic isoforms suggests that they act independently or in association. *J. Comp. Neurol.* **401,** 47–64.
46. Lamballe, F., Klein, R., and Barbacid, M. (1991). trkC, a new member of the trk family of tyrosine protein kinases, is a receptor for neurotrophin-3. *Cell* **66,** 967–979.
47. Hapner, S. J., Boeshore, K. L., Large, T. H., and Lefcort, F. (1998). Neural differentiation promoted by truncated trkC receptors in collaboration with p75(NTR). *Dev. Biol.* **201,** 90–100.