Published in final edited form as:

Immunogenetics. 2005 June ; 57(5): 374–383.

Genomic organization of the channel catfish CD45 functional gene and CD45 pseudogenes

Evgueni Kountikov, Melanie Wilson, Norman Miller, William Clem, and Eva Bengtén

Department of Microbiology, University of Mississippi Medical Center, 2500 North State Street, Jackson, MS, 39216-4505, USA e-mail: ebengtén@microbio.umsmed.edu Tel.: +1-601-9841739 Fax: +1-601-9841708

Sylvie Quiniou

Catfish Genetics Research Unit, USDA-Agricultural Research Service, Stoneville, MS, 38776, USA

Abstract

CD45 is a transmembrane protein tyrosine phosphatase, which in mammals plays an important role in T and B cell receptor and cytokine signaling. Recently, a catfish cDNA was shown to contain all characteristic CD45 features: an alternatively spliced amino-terminus, a cysteine-rich region, three fibronectin domains, a transmembrane region, and two phosphotyrosine phosphatase domains. However, analyses of CD45 cDNAs from various catfish lymphoid cell lines demonstrated that catfish CD45 is unique in that it contains a large number of alternatively spliced exons. Sequence analyses of cDNAs derived from the catfish clonal B cell line 3B11 indicated that this cell line expresses up to 13 alternatively spliced exons. Furthermore, sequence similarity among the alternatively spliced exons suggested duplication events. To establish the exact number and organization of alternatively spliced exons, a bacterial artificial chromosome library was screened, and the catfish functional CD45 gene plus six CD45 pseudogenes were sequenced. The catfish functional CD45 gene spans 37 kb and contains 49 exons. In comparison, the human and pufferfish CD45 genes consist of 34 and 30 exons, respectively. This difference in the otherwise structurally conserved catfish gene is due to the presence of 18 alternatively spliced exons that were likely derived through several duplication events. In addition, duplication events were also likely involved in generating the six pseudogenes, truncated at the 3' ends. A similarly 3' truncated CD45 pseudogene is also present in the pufferfish genome, suggesting that this specific CD45 gene duplication occurred before catfish and pufferfish diverged (~400 million years ago).

Keywords

CD45; Phosphatase; Bacterial artificial chromosome; Channel catfish; Gene duplication

Introduction

The transmembrane protein tyrosine phosphatase, CD45, is found on the surface of all nucleated cells of hematopoietic origin, and its expression is crucial in lymphocyte development and function (reviewed in Hermiston et al. 2003; Justement 1997; Penninger et al. 2001). In all species sequenced to date, the CD45 extracellular region consists of a heavily O-glycosylated linked variable amino-terminus, followed by a cysteine-rich region and three fibronectin type III (FNIII) domains. The transmembrane (TM) region is followed by a long cytoplasmic tail that contains the two tandem tyrosine phosphatase (PTP) domains (Hermiston et al. 2003). Briefly, CD45 is primarily responsible for dephosphorylating the negative regulatory site of Src family kinases, making them active. More recently, CD45 has been shown to be a negative regulator of the Jak/STAT activation pathway. Of the two PTP domains, only the first has intrinsic phosphatase activity; however, both are required for optimal function (Felberg and Johnson 1998, 2000; Hayami-Noumi et al. 2000).

In humans and mice, up to eight potential CD45 isoforms are produced by alternative splicing of exons 4, 5, and 6 (or A, B, and C), and isoform expression is dependent upon cell type and activation and differentiation states (Justement 1997). However, the functional roles of the various isoforms and their regulation are still poorly understood. Interestingly, the alternative splicing of amino-terminal exons to produce multiple isoforms is phylogenetically conserved. CD45 molecules have not only been identified in mammals (Hall et al. 1988; Saga et al. 1988) and in chickens (Fang et al. 1994) but also in various ectotherms. At the DNA/cDNA level, CD45 was found in the horned shark, *Heterodontus francisci* (Okumura et al. 1996), Pacific hagfish, *Eptatretus stoutii* (Nagata et al. 2002), sea lamprey, *Petromyzon marinus* (Uinuk-Ool et al. 2002), pufferfish, *Takifugu rubripes* (Diaz del Pozo et al. 2000), carp, *Cyprinus carpio* (Fujiki et al. 2000), and more recently, at both cDNA and protein levels in channel catfish, *Ictalurus punctatus* (Kountikov et al. 2004). Among the different species, the TM and intracellular region sequences are the most conserved; however, only the general structural features of the extracellular region sequences are conserved, i.e., the variable amino-terminus, cysteine-rich region, and three FNIII domains. In chickens and sharks, three alternatively spliced exons have been identified, in hagfish and pufferfish, two have been found, and in carp, only a single cDNA is available so the number cannot as yet be determined. In contrast, catfish can express up to 13 alternatively spliced exons at the mRNA level, and both large and small isoforms are observed at the protein level (Kountikov et al. 2004). The present study describes the isolation and sequencing of the channel catfish CD45 functional gene, the finding of 18 alternatively spliced exons, and the identification of putative promoter elements in the 5' UT region of the gene. In addition, six CD45 pseudogenes, which encode only truncated extracellular portions of the CD45 gene, have been mapped and sequenced.

Materials and methods

BAC library screening and sequencing

A channel catfish bacterial artificial chromosome (BAC) library, designated CCBL1, was screened by PCR using primer pairs specific for catfish CD45 (Table 1). This BAC library was prepared from a single homozygous gynogenetic catfish and provided 7.2× genome coverage. It was arrayed and stored by the Catfish Genetics Research Unit, USDA-ARS, Stoneville, MS (Quiniou et al. 2003). The 15 µl PCR reaction contained 10 mM Tris-HCl, pH 9.0, 50 mM KCl, 1 or 2 mM MgCl₂ depending on the primer pairs, 400 nM of each primer, 67 µM deoxynucleotides, 0.1% Triton X-100, and 1 unit of Taq (Promega Corporation, Madison, WI). The PCR cycling protocol was 95°C, 3 min; 40 cycles of 95°C, 1 min; 55–59°C (depending on the primer set), 30 s; 72°C, 1 min; and final extension at 72°C for 4 min. The products were separated on 1.0% agarose gels and visualized by ethidium bromide staining. A total of three CD45 BACs were sequenced; two, 22E12 and 11J16, contained pseudogenes, and one, 129I5, contained the full-length CD45 functional gene. The individual BACs were fragmented and cloned into plasmids using the shotgun method (Sulston et al. 1992). Briefly, DNA from BACs 22E12 and 11J16 was prepared using the Qiagen Large-Construct Kit (Qiagen Inc., Valencia, CA) according to the manufacturer's protocol. Fragments, 1–3 kb, were cloned into the pGEM-T vector (Promega). Plasmids were sequenced on both strands using universal forward and reverse primers by MWG Biotech Inc. (High Point, NC). A shotgun library for BAC 129I5 was prepared using the TOPO Shotgun Subcloning kit (Invitrogen Life Technologies, Carlsbad, CA). Briefly, 20 µg of BAC DNA was sheared for 30 s at 15 psi, blunt-end repaired, and dephosphorylated. DNA, 30–100 ng, was then ligated into pCR4Blunt-TOPO plasmid. Plasmids containing inserts of 1–2 kb were sequenced on both strands using universal forward and reverse primers by the USDA-ARS MSA Genomics Laboratory. Sequences were assembled into contigs using the Seqman II module of DNASTAR v5.05 (Lasergene, Madison, WI). When necessary, contigs were extended by primer walking using gene-specific primers. BAC 11J16, containing an ~192-kb insert, was assembled at 5.4× coverage; BAC 22E12,

containing an ~155-kb insert, was assembled at 4.6× coverage. Contig order and orientation were verified by PCR and by comparing restriction fragment sizes with the sizes predicted by sequencing. The catfish CD45 functional gene position and exon boundaries were defined by aligning BAC 129I5 with the CD45 cDNA sequence (AY366233) using the GeneQuest module of DNASTar v5.05 (Lasergene). The predicted acceptor and donor splice sites were also confirmed by comparison to the consensus splice sites cataloged by Mount (1982). The two BACs, 11J16 and 22E12, when assembled overlapped by ~48 kb with BAC 11J16 located 5' of BAC 22E12. BAC 129I5 contained an ~165-kb insert and was assembled at 7.1× coverage.

Reverse-transcription PCR

To detect if multiple CD45 transcription initiation sites occur in catfish, RT-PCR was performed on cDNA from the catfish clonal B cell line 3B11 (Wilson et al. 1997). Briefly, 1 µg of 3B11 total RNA was reversed transcribed using a primer, DC04 (Table 2), which is specific for the CD45 cytoplasmic region and SuperScript III Reverse Transcriptase (Invitrogen Life Technologies). Reaction conditions were according to the manufacturer's recommended protocol. One half microliter (2.5%) of the resulting cDNA was used as template for PCR reactions (50 µl total) that contained 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 3.0 mM MgCl₂, 150 µM each dATP, dCTP, dGTP, and dTTP, 250 ng each forward and reverse primer, and 1.25 U AmpliTaq (Applied Biosystems, Foster City, CA). A reverse primer (DP20) spanning the splice junction of CD45 exons 1 and 2 was used in conjunction with eight different forward primers specific to nucleotides upstream and downstream of the CD45 putative TATA box (Table 2). The cycling protocol was as follows: 2 min denaturation at 95°C, followed by 32 cycles of 25 s, 95°C; 25 s, 52°C; 35 s, 72°C; followed by 2 min, 72°C. Precautions used to avoid the amplification of contaminating genomic DNA included treatment of RNA with DNase (Invitrogen Life Technologies) using a reverse PCR primer that spanned CD45 exons 1 and 2. As a negative control, PCR was performed using DNase-treated RNA that had not been reverse transcribed.

Results

Previous cloning and sequence analyses revealed that the catfish CD45 homolog can contain up to 13 alternatively spliced exons (Kountikov et al. 2004). In order to map this potentially complicated locus, a catfish BAC library was initially screened with primer pairs that flank the CD45 alternatively spliced region. Nine positive BACs were identified; interestingly, these were shown by Southern blotting and PCR analyses to be overlapping and to contain CD45 pseudogenes that lacked the required PTP domains. Two overlapping BACs (11J16 and 22E12), which together contained six pseudogenes, were sequenced. To obtain the functional gene, the catfish BAC library was screened a second time with different primer pairs specific for the second PTP domain, the third fibronectin domain of the extracellular portion of the molecule, and the region directly 3' of the TM (see Table 1). This screening strategy identified BAC 129I5 which contained the functional CD45 gene. In addition, this BAC also contained a gene similar to the serine/threonine kinase, NEK7. In humans and mice, NEK7 is located ~600 kb 5' of the CD45 gene. Similarly, in catfish, NEK7 is located ~33 kb 5' of the CD45 gene, demonstrating conserved synteny of these genes. The linkage between the functional catfish CD45 gene and the pseudogenes is currently unknown.

Catfish CD45 gene organization

The catfish CD45 gene spans ~37 kb and consists of 49 exons (Fig. 1). Exon 1 and the first 39 nucleotides of exon 2 are untranslated; the remaining 3' portion of exon 2 encodes the leader peptide (LP), a situation reminiscent of the human CD45 gene (Hall et al. 1988). Similarly, exons 1 and 2 of the pufferfish CD45 gene are fused and the LP is encoded by its 3' portion (Diaz del Pozo et al. 2000). As described previously, catfish CD45 cDNAs are unusual in that

they were shown to contain up to ten alternatively spliced exons and RT-PCR approaches estimated that up to 13 may occur (Kountikov et al. 2004). Here, sequencing shows that 18 alternatively spliced exons are actually present. The first alternatively spliced exon (A) is encoded by exon 3 and separated from the other 17 alternatively spliced exons by a short 27-bp constant intervening exon (CI; exon 4). These 17 alternatively spliced exons (exons 5–21) are divided into four groups based on sequence similarity. The B group consists of exons #5 (B1), #9 (B2), #13 (B3), #17 (B4), and #21 (H or B5); C group consists of exons #6 (C1), #10 (C2), #14 (C3), and #18 (C4); D group consists of #7 (D1), #11 (D2), #15 (D3), and #19 (F or D4); E group consists of #8 (E1), #12 (E2), #16 (E3), and #20 (G or E4). The constant extracellular region of the molecule, which includes the cysteine-rich region and the three FNIII domains, is encoded by exons 22 through 31. Exon 32 encodes the TM region, and exons 33 through 49 encode the CD45 cytoplasmic region where two PTP domains are found. Exon 49 is the largest exon (1,436 nucleotides), albeit only the first 336 bp are translated.

With the exception of the alternatively spliced exons, the exon/intron structure of CD45 genes is conserved between catfish, pufferfish, and mammals (Hall et al. 1988; Saga et al. 1988), albeit the genes vary considerably in size. The human CD45 gene is 120 kb in length, the catfish CD45 is 37 kb, and in pufferfish, the CD45 gene is 12 kb. The pufferfish genome is very compact, ~7.5 times smaller than the human genome due to short introns (Venkatesh et al. 2000). In all three species, the CD45 genes contain a large intron between exon 2 (LP) and exon 3, ~50 kb in mammals, ~2.2 kb in pufferfish, and ~3.8 kb in catfish. Table 3 compares the exons that encode the conserved extracellular and intracellular regions of CD45 molecules in catfish, human, and pufferfish. The exon number and length of the intracellular and TM regions are remarkably consistent between these three species, which is not surprising since sequence identities in these regions are high, i.e., $\geq 51\%$ at the amino acid level for the intracellular region and $\geq 70\%$ for the TM (Kountikov et al. 2004). In contrast, the extracellular exons vary in length and exon fusions are found. Catfish exon 22, human exon 8, and pufferfish exon 4 represent the shortest CD45 exons in these species and are found directly 3' of the alternatively spliced exons. Although human exon 8 has been shown to be alternatively spliced at the mRNA level (Chang et al. 1991; Tsujikawa et al. 2000; Virts et al. 1998), catfish exon 22 and pufferfish exon 4 are found in all cDNA transcripts sequenced to date (Diaz del Pozo et al. 2000; Kountikov et al. 2004). Catfish exon 23 (225 bp) and pufferfish exon 5 (255 bp) correspond in location and size and exhibit 20% similarity with each other at the amino acid level. No equivalent exon is found in the human CD45 gene. Instead, human exon 9 (219 bp) is homologous to catfish exon 24 (180 bp) and pufferfish exon 6 (219 bp) with 32 and 31% similarity, respectively. These exons encode for the cysteine-rich region. The human CD45 exon 13 (159 bp) and pufferfish exon 10 (150 bp) encode for the second FNIII domain. In the catfish, this FNIII domain is encoded differently; exon 28 (21 bp) and exon 29 (132 bp) are split by a 102-bp intron. An additional, somewhat interesting, observation is the presence of the catfish CI exon. This exon separates the first alternatively spliced exon [exon 3(A)] from the other alternatively spliced exons and is encoded by exon 4 which is 27 bp in length. The corresponding exon in human CD45 is exon 3, which is also 27 bp in length and is always found expressed, i.e., it is not alternatively spliced.

The sequence 5' of catfish CD45 is quite different from mammalian and pufferfish 5' CD45 regions. For example, in mice and humans, there are multiple transcription initiation sites, but no TATA box sequences have been found (Saga et al. 1987, 1988). However, promoter sequences are found in the small intron 1 in human and mouse CD45 genes (Timon and Beverley 2001). There is no equivalent to this intron in pufferfish, and no potential promoter sequences in the pufferfish have been currently identified (Diaz del Pozo et al. 2000). In contrast, catfish CD45 has identifiable TATA and CAAT boxes. These are located 156 and 199 bp, respectively, 5' of the start of catfish CD45 cDNA (accession #AY366233). Of 18 full-length CD45 cDNAs identified by 5'RACE (Kountikov et al. 2004), all were found to initiate

between +110 and +140 bp (down-stream) of the TATA box suggesting that transcription initiation is TATA independent. To examine if additional CD45 initiation sites exist, PCR reactions using forward primers specific to nucleotides found -220 to +144 of the TATA sequence in combination with a reverse primer located 262 bp 3' of the TATA box (Table 2) were performed. Results show an intense band when the forward primer at +144 of the TATA sequence was used and weaker bands using forward primers at +23, +45, and +70 (Fig. 2). No, or very faint, bands were observed using forward primers at -220, -101, -25, or -5. Taken together, while the results are by no means quantitative, they suggest that there are at least two CD45 initiation sites, i.e., a TATA-independent site (at position +110 to +140) and another possibly TATA-dependent site (at approximately position +23).

Catfish CD45 alternative spliced exons

The high sequence identities between the alternatively spliced exons of catfish CD45 and their order in the genome suggest that they are likely the result of a series of gene duplications of a BCDE block (Fig. 3a). Within an individual group, exons are either identical or highly similar. For example, group C exon identities range from 98 to 100% and group E exon identities range from 77 to 100%. Such high identities prevent a minimum duplication unit or a history of duplications from being established, although it can be speculated that the more diverse exons in a group, exons H (B5), F (D4), and G (E4), have an earlier origin than the other members of their respective groups. Likewise, it could be that the group members showing high identity became homogenized through gene conversion events. The higher sequence identities between exon group D and E members (84–88%) infer a common origin. A common origin or duplication event(s) is also suggested by the sequence identities of the introns. For example, Fig. 3b shows a dot plot alignment of exons F (group D) and G (group E). These two exons and fragments of the neighboring introns (totaling 156 bp) exhibit 70% identity at the nucleotide level.

Catfish CD45 pseudogenes

A cluster of six CD45 pseudogenes spanning 230 kb is found at the 3' end of the scaffold for the overlapping catfish BACs, 11J16 and 22E12 (Fig. 4a). They appear to have evolved through a series of duplication events and are named according to their order in the scaffold. All six are truncated copies of the 5' end of the functional gene. ψ CD45-1 is the shortest pseudogene and consists of exon 1 through the CI (exon 4). The other ψ CD45 genes have either lost or partially deleted their alternatively spliced exons but kept most of the remaining extracellular exons intact, albeit with some variation (Fig. 4b). ψ CD45-2 contains only exons A and H, whereas ψ CD45-4 contains exons F and H and ψ CD45-6 contains exons A, B and H. ψ CD45-3 contains exons A, B, F, G and H, whereas ψ CD45-5 contains exons A, B, G, and H. As for the nonalternatively spliced exons, ψ CD45-4 lacks exon 32 (the TM exon) and ψ CD45-3 lacks exons 30 and 31 that encode for the third FNIII domain and exon 32 that encodes the TM.

The CD45 pseudogene duplication boundaries were determined by aligning the scaffold sequence with itself and with the functional CD45 gene. Basically, the pseudogenes are in two blocks of 121 and 105 kb. The first block consists of a long ~66-kb fragment containing ψ CD45-1, followed by two shorter fragments (~27 and ~28 kb) containing ψ CD45-2 and ψ CD45-3, respectively. The second block consists of the ψ CD45-4 fragment which is ~55 kb in length, followed by the two shorter ~28-kb ψ CD45-5 and ~22-kb ψ CD45-6 fragments. It seems likely that this block duplication is the more recent event as compared to the duplication event(s) that created the three pseudogenes within the blocks. This notion is supported by the observations that both long fragments share a 3' ~30-kb intron stretch that is 60% identical and that both ψ CD45-1 and ψ CD45-4 lack a *IpTc1*/Mariner transposable element present in the others. Deletions as well as insertions have also occurred in introns throughout the CD45 pseudogene region. For example, a 1.3-kb deletion is present between exons 3 and 4 in all six

pseudogenes implying that it was an early event before the duplications occurred. Likewise, a 0.3-kb insertion in intron 27 occurs in all pseudogenes except ψ CD45-3.

Since CD45 pseudogenes had not been previously reported in other species, the finding of six catfish pseudogenes is surprising. Thus, in order to determine if the catfish was unique in this aspect or if such pseudogenes exist in other teleosts, the available *Fugu* database was analyzed, i.e., the extracellular region of pufferfish CD45 was used in a BLAST search of the *Fugu rubripes* genomic assembly v.30 (<http://www.genome.jgi-psf.org/fugu6/fugu6.home.html>; Aparicio et al. 2002). This search identified a single scaffold, 2810, on the minus strand that contained a single CD45 pseudogene linked to the CD45 functional gene (data not shown). The scaffold consists of three contigs. Contig 1 is a 19-kb contig containing the 3' portion of the functional gene. Contig 2 is a 1.5-kb contig containing at its 3' end 109 bp of the 5' portion of the CD45 functional gene as reported by Diaz del Pozo et al. (2000). Contig 3 is a 5.8-kb contig containing the CD45 pseudogene. As with the catfish CD45 pseudogenes, the pufferfish CD45 pseudogene lacks the exons encoding the cytoplasmic portion of the molecule. It consists of exons 1 through 13 (TM) with all of the alternatively spliced exons (exons 2 and 3) and part of exon 12 (0.2 kb) deleted. Overall, the organization of the pufferfish pseudogene is very similar to that of the catfish homolog suggesting that the pseudogene duplication occurred before the two species diverged. When the CD45 pseudogenes and their corresponding functional genes are compared in a phylogenetic tree, the results indicate that the pseudogenes in the two species are more closely related to their functional counterparts than to each other (Fig. 4c). In addition, the branching of the tree supports the notion of the catfish CD45 pseudogene block duplication event discussed above, i.e., ψ CD45-3 branches with ψ CD45-6, ψ CD45-2 branches with ψ CD45-5, and ψ CD45-1 branches with ψ CD45-4.

Discussion

In conjunction with the previous report of channel catfish CD45 molecules, this study reveals an unexpected complexity of the catfish CD45 gene. There are 18 alternatively spliced exons in the CD45 gene, and 13 of these were identified earlier by cDNA sequencing (Kountikov et al. 2004). All the exons are in-frame and appear functional with viable splice acceptor and donor sites with the same splicing phase. It is hypothesized that these alternatively spliced exons are the result of a series of intragenomic duplications of a block of four exons. Furthermore, the presence of exons F (D4), G (E4), and H (B5), which are more diverse than their respective D, E, and B alternatively spliced exon counterparts, in the catfish pseudogenes suggests an early origin of these exons. In contrast, the CD45 genes of other species contain fewer alternatively spliced exons. Mammals have three, and there is evidence at the mRNA level that two more, exons 7 and 8, may be alternatively spliced (Chang et al. 1991; Tsujikawa et al. 2000; Virts et al. 1998). Similarly, in chickens and sharks, there are three alternative exons, whereas in pufferfish and hagfish, there are only two. The reason for such a high number of alternatively spliced exons in the catfish is unknown. One possibility may be that the higher original number of five alternatively spliced exons (exons A, B, F, G, and H plus the CI) in the catfish compensated for each exon's smaller size. For example, catfish alternatively spliced exons range in size from 66 to 90 bp with an average size of 80 bp. In comparison, the alternatively spliced exons A, B, and C in human CD45 range in size from 141 to 198 bp with an average of 161 bp, and in the pufferfish, exons are 127 and 99 bp. The smallest catfish CD45 isoform (AY366213) sequenced to date from the catfish clonal B cell line 3B11 contains three alternatively spliced exons plus the CI. In comparison, the smallest isoform from a clonal T cell line contains no alternatively spliced exons but does contain the CI. The longest cDNAs (AY366227, AY366220, AY366219, and AY366214) are from 3B11 B cells, and they contain ten exons plus the CI (Kountikov et al. 2004). Of the various functional models proposed for the different mammalian CD45 isoforms, one suggests that the size of CD45 plays a role in cell-cell interactions (reviewed in Justement 1997). Another model suggests that O-linked

glycosylations of the alternatively spliced exons regulate CD45 dimerization effecting phosphatase activity, i.e., extensive glycosylation creates a negative charge that physically impedes dimerization, so that one molecule cannot block the catalytic center of the other molecule, thereby inhibiting phosphatase activity (Xu and Weiss 2002). However, until more is known about CD45 function, it is difficult to determine how and if these models apply to the multiple long CD45 isoforms found expressed in the catfish. It should also be noted that the currently available anti-CD45 mAbs do not permit determining which of the catfish 18 alternatively spliced exons are expressed at the protein level.

When the remaining nonalternatively spliced exons of the catfish CD45 gene are compared to those of the human and pufferfish CD45 genes, very few structural differences are seen. Exon sizes are shorter in the catfish (Table 3) as the result of shifted splice sites (exons 23 and 24), and some exons have short deletions within the coding region. The other difference is the acquisition of the intron creating catfish exons 28 and 29. No such event occurred in the human and pufferfish CD45 genes or in the pufferfish CD45 pseudogene. These two exons and intron are also present in five of the catfish CD45 pseudogenes, suggesting that the intron may have been acquired after speciation of the catfish and pufferfish ~400 millions years ago.

In addition to the 18 alternatively spliced exons, other distinguishing features of catfish CD45 are the presence of the putative TATA and CAAT boxes upstream of the functional gene and the truncated six pseudogenes. No TATA boxes have been found upstream of the human (or mouse) and pufferfish CD45 genes, although a sequence with weak promoter activity has been identified within mouse intron one (DiMartino et al. 1994), and a sequence with strong activity but no tissue specificity has been identified in human intron one (Timon and Beverley 2001). Whether or not the putative promoter elements found 5' of the catfish CD45 gene are functional remains to be determined. The six sequenced catfish CD45 pseudogenes and the identification of the pufferfish pseudogene are the first reports of CD45 pseudogenes in any species. These pseudogenes are rather similar in their organization in that they lack exons encoding the intracellular domains, and the genomic region containing the alternatively spliced exons has undergone deletions. However, unlike the pufferfish pseudogene where the two alternatively spliced exons are both deleted, the catfish pseudogenes contain at least one. In addition to exon A, exons B, F, G, and H are found, although in different combinations. The level of structural conservation between the pseudogenes may suggest that the duplication creating a CD45 pseudogene occurred before catfish and pufferfish speciation. However, phylogenetic analyses indicate that the pseudogenes in the two species are more closely related to their corresponding functional gene than to each other; perhaps gene conversion events resulted in homogenization of the pseudogenes with their respective functional genes. Such events or processes have been documented to occur with other immune genes, and recently, gene conversion has been proposed as the mechanism responsible for the μ CH1-like δ CH1 exons in the IgD genes of artiodactyls (Zhao et al. 2002). Alternatively, pseudogenes lacking exons encoding the intracellular portion of CD45 may have emerged independently in the two fish lineages. The reason for the large number of CD45 pseudogenes in the catfish is unknown. Another example of multiple pseudogenes occurring in the catfish is found in the immunoglobulin heavy (IgH) chain locus. There are three copies of Ig δ constant H chain genes each linked to an Ig μ gene or pseudogene. Two of the δ genes have been completely sequenced and are expressed, and the third is truncated but the exons are intact. Tc1/mariner-like elements, which are highly represented in catfish (Henikoff 1992), are also present in the catfish IgH chain locus, and it has been suggested that they may have been involved in the duplications (Bengtén et al. 2002). The presence of multiple transposon sequences flanking the CD45 pseudogenes as well as three transposon sequences within the alternatively spliced region of the functional gene suggests that they may have contributed to the duplications.

In summary, although the organization and overall features of the catfish CD45 gene are conserved, the results presented here illustrate three unusual characteristics of the gene. The first is the presence of 18 functional alternatively spliced exons, the second is the existence of a probable functional TATA box, and the third is the occurrence of at least six CD45 pseudogenes in the catfish genome. Since previous studies have shown that catfish B cells express larger isoforms and catfish T cells express smaller isoforms, it is anticipated that future studies employing various catfish lymphoid cell lines, CD45 isoform primers, and mAbs will enable for a better understanding of the presumed strict regulation of alternative splicing in this complex gene.

Acknowledgements

This work was supported by grants from the National Institutes of Health (R01AI-19530), the National Science Foundation (MCB-0211785), and the US Department of Agriculture (2003-35205-12829). The pufferfish data were provided freely by the Fugu Genome Consortium for use in this publication only. The experiments performed comply with the current laws of the USA.

References

- Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 2002;297:1301–1310. [PubMed: 12142439]
- Bengtén E, Quiniou SM, Stuge TB, Katagiri T, Miller NW, Clem LW, Warr GW, Wilson M. The IgH locus of the channel catfish, *Ictalurus punctatus*, contains multiple constant region gene sequences: different genes encode heavy chains of membrane and secreted IgD. *J Immunol* 2002;169:2488–2497. [PubMed: 12193718]
- Chang HL, Lefrançois L, Zaroukian MH, Esselman WJ. Developmental expression of CD45 alternate exons in murine T cells. Evidence of additional alternate exon use. *J Immunol* 1991;147:1687–1693. [PubMed: 1831832]
- Díaz del Pozo E, Beverley PC, Timon M. Genomic structure and sequence of the leukocyte common antigen (CD45) from the pufferfish *Fugu rubripes* and comparison with its mammalian homologue. *Immunogenetics* 2000;51:838–846. [PubMed: 10970098]
- DiMartino JF, Hayes P, Saga Y, Lee JS. A novel initiator/promoter element within the CD45 upstream region. *Int Immunol* 1994;6:1279–1283. [PubMed: 7981155]
- Fang KS, Sabe H, Saito H, Hanafusa H. Comparative study of three protein-tyrosine phosphatases. Chicken protein-tyrosine phosphatase lambda dephosphorylates c-Src tyrosine 527. *J Biol Chem* 1994;269:20194–20200. [PubMed: 7519604]
- Felberg J, Johnson P. Characterization of recombinant CD45 cytoplasmic domain proteins. Evidence for intramolecular and intermolecular interactions. *J Biol Chem* 1998;273:17839–17845. [PubMed: 9651387]
- Felberg J, Johnson P. Stable interdomain interaction within the cytoplasmic domain of CD45 increases enzyme stability. *Biochem Biophys Res Commun* 2000;271:292–298. [PubMed: 10799290]
- Fujiki K, Shin DH, Nakao M, Yano T. Molecular cloning of carp (*Cyprinus carpio*) leucocyte cell-derived chemotaxin 2, glia maturation factor beta, CD45 and lysozyme C by use of suppression subtractive hybridisation. *Fish Shellfish Immunol* 2000;10:643–650. [PubMed: 11081442]
- Hall LR, Streuli M, Schlossman SF, Saito H. Complete exon–intron organization of the human leukocyte common antigen (CD45) gene. *J Immunol* 1988;141:2781–2787. [PubMed: 2971730]
- Hayami-Noumi K, Tsuchiya T, Moriyama Y, Noumi T. Intra- and intermolecular interactions of the catalytic domains of human CD45 protein tyrosine phosphatase. *FEBS Lett* 2000;468:68–72. [PubMed: 10683443]
- Henikoff S. Detection of Caenorhabditis transposon homologs in diverse organisms. *New Biol* 1992;4:382–388. [PubMed: 1320399]

- Hermiston ML, Xu Z, Weiss A. CD45: a critical regulator of signaling thresholds in immune cells. *Annu Rev Immunol* 2003;21:107–137. [PubMed: 12414720]
- Justement LB. The role of CD45 in signal transduction. *Adv Immunol* 1997;66:1–65. [PubMed: 9328639]
- Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980;16:111–120. [PubMed: 7463489]
- Kountikov E, Wilson M, Miller N, Clem W, Bengten E. Organization and expression of thirteen alternatively spliced exons in catfish CD45 homologs. *Dev Comp Immunol* 2004;28:1023–1035. [PubMed: 15236932]
- Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 2001;17:1244–1245. [PubMed: 11751241]
- Mount SM. A catalogue of splice junction sequences. *Nucleic Acids Res* 1982;10:459–472. [PubMed: 7063411]
- Nagata T, Suzuki T, Ohta Y, Flajnik MF, Kasahara M. The leukocyte common antigen (CD45) of the Pacific hagfish, *Eptatretus stoutii*: implications for the primordial function of CD45. *Immunogenetics* 2002;54:286–291. [PubMed: 12136341]
- Okumura M, Matthews RJ, Robb B, Litman GW, Bork P, Thomas ML. Comparison of CD45 extracellular domain sequences from divergent vertebrate species suggests the conservation of three fibronectin type III domains. *J Immunol* 1996;157:1569–1575. [PubMed: 8759740]
- Penninger JM, Irie-Sasaki J, Sasaki T, Oliveira-dos-Santos AJ. CD45: new jobs for an old acquaintance. *Nat Immunol* 2001;2:389–396. [PubMed: 11323691]
- Quiniou SM, Katagiri T, Miller NW, Wilson M, Wolters WR, Waldbieser GC. Construction and characterization of a BAC library from a gynogenetic channel catfish *Ictalurus punctatus*. *Genet Sel Evol* 2003;35:673–683. [PubMed: 14604514]
- Saga Y, Tung JS, Shen FW, Boyse EA. Alternative use of 5' exons in the specification of Ly-5 isoforms distinguishing hematopoietic cell lineages. *Proc Natl Acad Sci U S A* 1987;84:5364–5368. [PubMed: 3037546]
- Saga Y, Tung JS, Shen FW, Pancoast TC, Boyse EA. Organization of the Ly-5 gene. *Mol Cell Biol* 1988;8:4889–4895. [PubMed: 3211131]
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–425. [PubMed: 3447015]
- Sulston J, Du Z, Thomas K, Wilson R, Hillier L, Staden R, Halloran N, Green P, Thierry-Mieg J, Qiu L, et al. The *C. elegans* genome sequencing project: a beginning. *Nature* 1992;356:37–41. [PubMed: 1538779]
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]
- Timon M, Beverley PC. Structural and functional analysis of the human CD45 gene (PTPRC) upstream region: evidence for a functional promoter within the first intron of the gene. *Immunology* 2001;102:180–189. [PubMed: 11260323]
- Tsujikawa K, Uchino Y, Ichijo T, Furukawa T, Yamamoto H. Detection of CD45iota mRNA in murine Th1 but not Th2 clones. *Immunobiology* 2000;201:506–514. [PubMed: 10834309]
- Uinuk-Ool T, Mayer WE, Sato A, Dongak R, Cooper MD, Klein J. Lamprey lymphocyte-like cells express homologs of genes involved in immunologically relevant activities of mammalian lymphocytes. *Proc Natl Acad Sci U S A* 2002;99:14356–14361. [PubMed: 12391333]
- Venkatesh B, Gilligan P, Brenner S. Fugu: a compact vertebrate reference genome. *FEBS Lett* 2000;476:3–7. [PubMed: 10878239]
- Virts E, Barritt D, Raschke WC. Expression of CD45 isoforms lacking exons 7, 8 and 10. *Mol Immunol* 1998;35:167–176. [PubMed: 9694517]
- Wilson M, Bengten E, Miller NW, Clem LW, Du Pasquier L, Warr GW. A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD. *Proc Natl Acad Sci U S A* 1997;94:4593–4597. [PubMed: 9114035]
- Xu Z, Weiss A. Negative regulation of CD45 by differential homodimerization of the alternatively spliced isoforms. *Nat Immunol* 2002;3:764–771. [PubMed: 12134145]

Zhao Y, Kacs Kovics I, Pan Q, Liberles DA, Geli J, Davis SK, Rabbani H, Hammarstrom L. Artiodactyl IgD: the missing link. J Immunol 2002;169:4408–4416. [PubMed: 12370374]

Fig. 1.

The catfish CD45 gene organization. Map in center shows the 49 CD45 exons. *Vertical arrows* mark specific exons: LP(2)=exon 2 encodes leader peptide; A(3)=exon 3 encodes alternatively spliced exon A; CI(4)=exon 4 encodes constant intervening exon; TM(32)=exon 32 encodes transmembrane. A Gypsy-type transposable element is marked. Above the map is the sequence of the CD45 promoter region; the putative CAAT, TATA boxes, and 5' end of exon 1 are marked. UP24 and UP23 represent primers used to determine the start site of exon 1. Below the map are the alternatively spliced exons 5–21, which are found between the CI and exon 22. The letter names are shown and IpTc1/Mariner elements are marked. GenBank accession number AY775953

Fig. 2.

Preliminary mapping of CD45 transcription initiation sites. RT-PCR of RNA from catfish 3B11 cells utilizing unique forward primers found between -220 and $+144$ of the TATA box and a reverse primer (DP20) spanning the junction of CD45 exons 1 and 2. *Lanes* are labeled according to the forward primer used and its location relative to the TATA sequence. As a control, PCR was performed using the UP22 forward primer and DP20 reverse primer and DNase-treated 3B11 RNA

Fig. 3.

a Nucleotide and amino acid alignments of the catfish CD45 alternatively spliced exons. Exon splice acceptor and donor sites are indicated by (/); a second (/) indicates that two acceptor sites are found used. Inferred amino acid sequences are shown above the second base of each codon; *periods* indicate identity. **b** Dot plot alignment of alternatively spliced exons F (group D) and G (group E). Alignment includes sequences of the flanking introns. Settings were 65% over a window size of 30. The long and short *solid lines* indicate identity of 74.5 and 62.4%, respectively; *hatched lines* indicate the 100% identity observed when intron F–G is aligned with itself. Exon/intron schematic is depicted at top and at left

Fig. 4.

a Genomic organization of the catfish CD45 pseudogenes. Map in center shows the six CD45 pseudogenes; (x) marks the IpTc1/Mariner elements. The schematic of the two overlapping BACs, 11J16 and 22E12, is shown below with the boundaries of the duplicated fragments marked (^). The above schematics show the exon organization of ψ CD45-1 and ψ CD45-5; *triangles* indicate positions of insertions ($\blacktriangledown$) and deletions ($\blacktriangle$), respectively. GenBank accession number for the scaffold: AY775952. **b** Schematic comparing the CD45 pseudogene exons. ($\blacksquare$) indicates exons present, and ($\square$) indicates truncated exons. **c** Phylogenetic tree showing the relationships of CD45 pseudogenes (marked ps) and corresponding fragments of the functional CD45 genes from pufferfish, catfish, and human (accession #NT_004487, nucleotides 1–89353). Sequences were aligned using CLUSTALW (Thompson et al. 1994), and variability of the sequences was assessed with the help of the MEGA 2.1 program (Kumar et al. 2001) by using synonymous and nonsynonymous substitutions estimated by Kimura's two-parameter model (Kimura 1980). The phylogenetic tree was drawn by the neighbor-joining method (Saitou and Nei 1987) using the MEGA 2.1 program based on Kimura's two-parameter distances for nucleotide sequences. The topological stability of the tree was assessed by 500 bootstrap replications

Table 1

Primers used in BAC library screening

Name	Sequence(5'–3')	Location	Position on 129I5 BAC
UP09	AAACTGGGAAATTACTGCTCTGC	Upstream of alternatively spliced region; intron 3	(47644–47666)
DP10	TTACCTGGTCCATTTTGTGTCATCT	Upstream of alternatively spliced region; intron 4	(48104–48081)
UP08 ^a	ACAGAGACATCCCCAAGTG	Downstream of alternatively spliced region; exon 22	(60573–60592)
DP03	CGTTTTCTTCTCGAGGGTGTAG	Downstream of alternatively spliced region; exon 23	(60714–60692)
UP17 ^b	ATCGGAAAGAGAAGGCAGCAGAA	Intracellular; exon 40	(71446–71468)
DP17	TGACGGCGAAGTTTGACCACA	Intracellular; exon 41	(71816–71796)
2U04	CAAAATGGTAATGGGAACAGAACA	Extracellular; exon 31	(67086–67109)
2D04	AACAATGAGAAAAGGCCAGGAATC	TM; exon 32	(67310–67288)
UP02	GGAATCTTCTGCGCCTTGTGGA	PTP II; exon 48	(76598–76619)
DP02	CTCTGGCTCGCTGCTGGACTCA	PTP II; exon 49	(77578–77557)

^a Bases 11–12 of the primer are mismatched with bases TG on the template

^b Base 2 of the primer is mismatched with G on the template

Table 2
Primers used to map CD45 transcription initiation sites

Name	Sequence(5'–3')	Position to TATA box ^a	Position on 129I5 BAC
UP27	AAAAGGCTTATCAGCTTGGTTAG	–220	41372–41394
UP26	ACAGCAAATTAGCATGACTGACA	–101	41491–41513
UP25	CATGGACATGAGGATGTGTATT	–25	41567–41588
UP24	TTATAAATGCATAATGTTCTTACAGTG	–5	41587–41613
UP23	TGATTCTTAGCTTCAACAAGTGAG	+23	41615–41638
UP20	AGACAGAAGTACAAAAAGCGAGAG	+45	41637–41660
UP21	TTGTGGCGTAGCTGAAAATA	+70	41662–41681
UP22 ^b	CACTGCTCCGTCTGTCAAG	+144	41736–41754
DP20 ^b	TTCTAAGACACCATTAAGTC/GCTTTC	+262	41987–41969/41834–41829
DC04	GTCAGCCCTCTTCTTCT	+3216	67700–67616

^aTATA box positioned at 41594 of BAC 129I5

^bSplice site-specific primer site indicated by “/”

Table 3
Comparison of CD45 exons in catfish, human, and pufferfish

Catfish		Human		Fugu		
Exon	Length (bp)	Exon	Length (bp)	Exon	Length (bp)	Encoding region
1	223	1a	228			5' UTR
		1b	101			5' UTR
2	97	2	67	1	221	LP
3 (A)	90	-	-	-	-	
4	27	3	27	-	-	CI
5 (B)	90	4 (A)	198	-	-	
6 (C)	66	5 (B)	141	-	-	
7 (D)	81	6 (C)	144	2	147	
21 (H) ^a	90	7	75	3	99	
22	15	8	27	4	30	
23	225	-	-	5	255	
24	180	9	219	6	219	CR ^b
25	99	10	129	7	105	FNIII-1
26	120	11	138	8	132	
27	126	12	120	9	147	
28	21	-	-	-	-	FNIII-2
29	132	13	159	10	150	
30	212	14	209	11	209	FNIII-3
31	64	15	61	12	61	
32	109	16	109	13	109	TM
33	35	17	35	14	38	
34	101	18	110	15	110	
35	91	19	90	16	91	
36	65	20	78	17	77	
37	37	21	37	18	37	PTP DI
38	98	22	98	19	98	
39	126	23	126	20	126	
40	158	24	158	21	158	
41	136	25	136	22	136	
42	147	26	150	23	150	
43	88	27	91	24	91	
44	134	28	134	25	140	PTP DII
45	135	29	135	26	135	
46	123	30	123	27	117	
47	161	31	179	28	158	
48	136	32	136	29	136	
49	>1,443	33	590	30	618	

^a Catfish alternatively spliced exons 8–20 are not shown

^b Cysteine-rich region