Evolutionary conservation of a molecular machinery for export and expression of mRNAs with retained introns

BAOMIN WANG, DAVID REKOSH, and MARIE-LOUISE HAMMARSKJOLD

Myles H. Thaler Center for AIDS and Human Retrovirus Research, Department of Microbiology, Immunology, and Cancer Biology, University of Virginia School of Medicine, Charlottesville, Virginia 22908, USA

ABSTRACT

Intron retention is one of the least studied forms of alternative splicing. Through the use of retrovirus and other model systems, it was established many years ago that mRNAs with retained introns are subject to restriction both at the level of nucleocytoplasmic export and cytoplasmic expression. It was also demonstrated that specific *cis*-acting elements in the mRNA could serve to bypass these restrictions. Here we show that one of these elements, the constitutive transport element (CTE), first identified in the retrovirus MPMV and subsequently in the human *NXF1* gene, is a highly conserved element. Using GERP analysis, CTEs with strong primary sequence homology, predicted to display identical secondary structure, were identified in *NXF* genes from >30 mammalian species. CTEs were also identified in the predicted *NXF1* genes of zebrafish and coelacanths. The CTE from the zebrafish *NXF1* was shown to function efficiently to achieve expression of mRNA with a retained intron in human cells in conjunction with zebrafish Nxf1 and cofactor Nxt proteins. This demonstrates that all essential functional components for expression of mRNA with retained introns have been conserved from fish to man.

Keywords: CTE; conserved RNA element; intron retention; post-transcriptional gene regulation; RNA export

INTRODUCTION

It is now well established that a majority of human genes are subject to alternative splicing (AS) and this is likely also true for many other mammalian species (Pan et al. 2008; Wang et al. 2008). In mammals, the most common form of AS is exon skipping and intron retention was once thought to be one of the rarest. However, deep sequencing and other recent advances have revealed that this form of AS is more common than previously appreciated (Ner-Gaon et al. 2004; Yap et al. 2012; Wong et al. 2013). In *Arabidopsis* and other plants, retained introns are a prominent feature in AS and have been proposed to be important in adaptive regulation (Ner-Gaon et al. 2004; Remy et al. 2014; Vitulo et al. 2014).

Intron retention is often coupled to both nucleocytoplasmic export restrictions and nonsense mediated decay (NMD), since the retained intron frequently contains premature stop codons (Lejeune and Maquat 2005). For this reason, it has been hypothesized that the vast majority of mRNAs with retained introns are never stably expressed in the cytoplasm or translated into proteins (Yap et al. 2012; Wong et al. 2013; Ge and Porse 2014). However, there are now many examples in the literature of mammalian genes that express mRNAs with retained introns that are translated and where the resulting proteins can be readily detected. Examples include *NXF1*, *ERBB2* (Herstatin), *ID3*, *PASD1*, *CEACAM6*, Kallikrein genes, *TRP-2*, *ANIA-6* (*CCNL1*), *PRX*, and several others (Dytrych et al. 1998; Lupetti et al. 1998; Doherty et al. 1999; Berke et al. 2001; Forrest et al. 2004; Liggins et al. 2004; Michael et al. 2005; Li et al. 2006; Kurio et al. 2008).

The evolutionarily conserved *NXF1* gene encodes a protein that is thought to serve as a major mRNA export receptor in mammalian and many other species (for recent reviews, see Siddiqui and Borden 2012; Müller-McNicoll and Neugebauer 2013; Natalizio and Wente 2013). The human Nxf1 protein (then known as Tap), was originally identified because of its direct interaction with the Mason-Pfizer Monkey Virus (MPMV) CTE, a *cis*-acting RNA element necessary for the nucleocytoplasmic export and translation of the unspliced MPMV genome mRNA that contains a retained intron (Bray et al. 1994; Ernst et al. 1997a,b; Pasquinelli et al. 1997; Grüter et al. 1998). In a previous study, we showed that intron 10 of the human *NXF1* gene contains a sequence with striking primary sequence and secondary structure ho-

Corresponding author: mh7g@virginia.edu

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.048520.114.

mology with the MPMV CTE (Li et al. 2006). We also showed that many human cells express an alternatively spliced *NXF1* mRNA that retains this intron. The nucleocytoplasmic export and expression of this mRNA was demonstrated to require a direct interaction of the *NFX1* CTE with the Nxf1 protein and its cofactor Nxt1. Furthermore, we showed that the mRNA with the retained intron was translated into an alternative short Nxf1 protein with unknown function.

Nxt1 and Nxf1 proteins in *trans* and the CTE in *cis* can also functionally replace the HIV Rev protein and its *cis*-acting HIV RNA element, known as the RRE, in HIV replication (Bray et al. 1994; Zolotukhin et al. 1994). Many studies have demonstrated that Rev/RRE function is essential for the nucleocytoplasmic export of unspliced HIV mRNA, as well as for the export of several subgenomic HIV mRNAs that retain introns (Felber et al. 1989; Hammarskjöld et al. 1989; Malim et al. 1989). Taken together, our previous studies have clearly demonstrated a role for Nxf1 and the CTE in the export of both viral and cellular mRNAs with retained introns.

Here we show that the *NXF* CTE sequence and predicted secondary structure is conserved in many mammalian *NXF* genes and that CTEs can also be identified in the genomes of both *Danio rerio* (zebrafish) and *Latimeria chalumnae* (a coelacanth). In zebrafish, the CTE is present in the same intron as in the human *NXF1* gene (intron 10) and we demonstrate that this intron is retained in an alternative *NXF1* mRNA. Furthermore, when transplanted into a HIV reporter construct, which allows quantification of export and expression of mRNA with a retained intron, the zebrafish CTE functions very efficiently in human cells in conjunction with the zebrafish Nxf1 and Nxt2 proteins.

RESULTS

NXF1 CTE sequences are conserved within *NXF1* genes from many different mammalian species

Previously, we showed that the mouse *NXF1* gene contains a sequence with almost perfect homology with the CTE in human *NXF1* (Li et al. 2006). In both species, the CTE maps to intron 10 and has been shown to be present in alternative *NXF1* mRNA transcripts retaining this intron. Using genomic evolutionary rate profiling (GERP) analysis (Cooper et al. 2005) and the Ensembl genome browser to examine intron 10 in other mammalian species, we identified a conserved region of 94 nt in intron 10 in multiple mammalian *NXF1* genes. Further analysis showed that this region comprises the CTE and surrounding nucleotides. The alignments of this region in 36 mammalian species are shown in Figure 1. As can be seen in the figure, there is remarkable conservation of this intronic region across most of the species, with no nucleotide changes in the "inner loop" known to bind Nxf1 and surrounding sequences (for a folded RNA structure of the human *NXF1* CTE, see Fig. 2C; Ernst et al. 1997b). The conservation of this "noncoding" intronic region in many diverse mammalian species is consistent with it having a conserved, potentially important biological function. The conservation was less pronounced in a few mammalian species: (*Sorex araneus* [common shrew], *Dasypus novemcinctus* [nine-banded armadillo], and *Tursiops Truncatus* [bottlenose dolphin]).

Intron 10 of the predicted *NXF1* genes in both *Danio rerio* and *Latimeria chalumnae* contain sequences predicted to function as CTEs

We next performed complete BLAST searches to identify sequences in other mammalian genes with significant primary sequence homology with the *NXF1* CTE. However, these searches failed to identify CTEs outside of the *NXF* gene family.

To explore if the genomes of nonmammalian species contain DNA sequences with homology to mammalian CTEs, we performed BLAST searches using nonmammalian databases. Although these searches did not find such sequences in birds or reptiles or in the model organisms *Caenorhabditis elegans* and *Drosophila*, a potential CTE with extensive primary sequence homology with the human CTE was found in a predicted *NXF* gene in the *Danio rerio* (zebrafish) genome.

The zebrafish genome contains two regions with homology with human *NXF1*. However, the CTE homology region is only present in one of these (ZFIN:ZDB-GENE-030131-2585), located on chromosome 21 (see Fig. 2A). This gene has been recently officially designated as the *Danio rerio* (*Dr*) *NXF1* gene in the National Center for Biotechnology Information database (NCBI). Like the human gene, the *DrNXF1* gene contains 21 exons and 20 introns and the putative CTE is present in intron 10 (Fig. 2B). However, *DrNXF1* intron 10 is significantly larger than the human *NXF1* intron 10 (3992 nt, compared with 1801 nt for *huNXF1*). Outside of the CTE region there is little conservation of intronic sequences between the human and zebrafish genes, as evidenced by BLAST searches using intron sequences (data not shown). In contrast, the exons show significant conservation and the human and the predicted *DrNXF1* large protein isoforms (619 and 642 amino acids) show ~60% identity at the amino acid level. This is not surprising, as important functions of Nxf1 in post-transcriptional mRNA regulation appear to be conserved in evolution. A second putative *NXF* ortholog in *Danio rerio* is located on chromosome 14 (ZDB-GENE-060825-299). We have been unable to identify a putative CTE in this gene, the function of which remains unknown. It is predicted to encode a 610 amino acid protein with 54% sequence identity to the large human Nxf1 protein and a similar domain structure. Figure 2C shows the putative folding of the zebrafish CTE based on the known

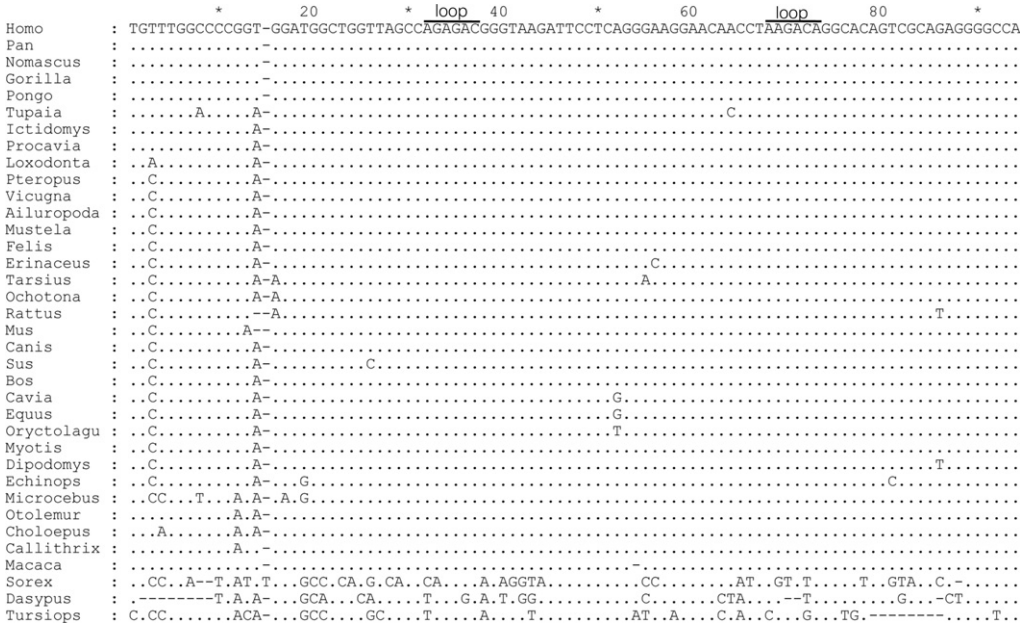

FIGURE 1. Alignment of human *NXF1* CTE homologs from different mammalian species. The *top* line shows the sequence of human *NXF1* CTE. A dot (.) shows that the nucleotide is the same as the one in the *top* line. A dash (-) shows that there is a gap in the alignment. The position of the inner loop formed by nucleotides 32–37 and 68–73 is indicated *above* the human sequence. Species abbreviations: (Ailuropoda) *Ailuropoda melanoleuca*; (Bos) *Bos taurus*; (Callithrix) *Callithrix jacchus*; (Canis) *Canis lupus familiaris*; (Cavia) *Cavia porcellus*; (Choloepus) *Choloepus hoffmanni*; (Dasypus) *Dasypus novemcinctus*; (Dipodomys) *Dipodomys ordii*; (Echinops) *Echinops telfairi*; (Equus) *Equus caballus*; (Erinaceus) *Erinaceus europaeus*; (Felis) *Felis catus*; (Gorilla) *Gorilla gorilla gorilla*; (Homo) *Homo sapiens*; (Ictidomys) *Ictidomys tridecemlineatus*; (Loxodonta) *Loxodonta africana*; (Macaca) *Macaca mulatta*; (Microcebus) *Microcebus murinus*; (Mus) *Mus musculus*; (Mustela) *Mustela putorius furo*; (Myotis) *Myotis lucifugus*; (Nomascus) *Nomascus leucogenys*; (Ochotona) *Ochotona princeps*; (Oryctolagus) *Oryctolagus cuniculus*; (Otolemur) *Otolemur garnettii*; (Pan) *Pan troglodytes*; (Pongo) *Pongo abelii*; (Procavia) *Procavia capensis*; (Pteropus) *Pteropus vampyrus*; (Rattus) *Rattus norvegicus*; (Sorex) *Sorex araneus*; (Sus) *Sus scrofa*; (Tarsius) *Tarsius syrichta*; (Tupaia) *Tupaia belangeri*; (Tursiops) *Tursiops truncatus*; (Vicugna) *Vicugna pacos*.

secondary structure of the MPMV CTE. The human *NXF1* CTE is also shown for comparison.

Recently, sequencing of the genome of *Latimeria chalumnae* (a coelacanth, also known as "living fossil" fish) was completed and deposited in Ensembl and other genome browsers (Nikaido et al. 2013). The coelacanths are lobe-finned fish that appear to be more closely related to lungfish, reptiles, and mammals than to the common ray-finned fishes. A BLAST search revealed that the *Latimeria* genome also contains a sequence with remarkable homology with the human *NXF1* CTE (Fig. 2A,C). As can be seen in Figure 2C, the predicted stem–loop structure of the 53-nt "core" region is identical for all four CTEs. In the predicted *Latimeria* CTE core structure, there are only 3 nt differences compared with the human *NXF1* CTE, whereas there are seven differences between the zebrafish and human CTEs and six differences between the *Latimeria* and zebrafish CTEs.

The predicted *NXF1* gene in *Latimeria chalumnae* (ENSLACG00000001232) (Fig. 2B) is 30.35 kbp, quite a bit larger than both the human and zebrafish genes, mainly because of the increased size of many of the introns. Nevertheless, as in the case of the other *NXF* genes, intron 10 (3426 nt) contains the predicted CTE. For all of these *NXF1* genes, the position of the predicted CTE is closer to the 5′SS than the 3′′SS in intron 10. The *Latimeria NXF1* gene is predicted to express a large Nxf1 protein of 628 amino acids with a similar domain structure as other Nxf proteins (RNA binding domain, NTF2 like domain and a nucleoporin complex [NPC] binding C-terminal domain). It displays a 71% identity to the human Nxf1 protein.

The zebrafish CTE is functional in human cells

We have previously shown that both the MPMV and human *NXF1* CTEs function to enable export and translation of mRNAs with retained introns in mammalian cells (Ernst et al. 1997a; Guzik et al. 2001; Jin et al. 2003; Li et al. 2006; Swartz et al. 2007). The ability of a CTE to do this can be assessed by inserting it into a reporter plasmid containing the GagPol ORF from HIV1, where this ORF is present within an intron. Using this system, we have shown in several previous studies that expression of HIV Gag and GagPol proteins requires the nucleocytoplasmic export and translation of the GagPol mRNA with the retained intron (Bray et al. 1994; Guzik et al. 2001; Jin et al. 2003). When the Gag and GagPol proteins are produced in the cell, they are assembled into virus-like particle and secreted into the medium. The amount of protein in the medium of transfected cells can

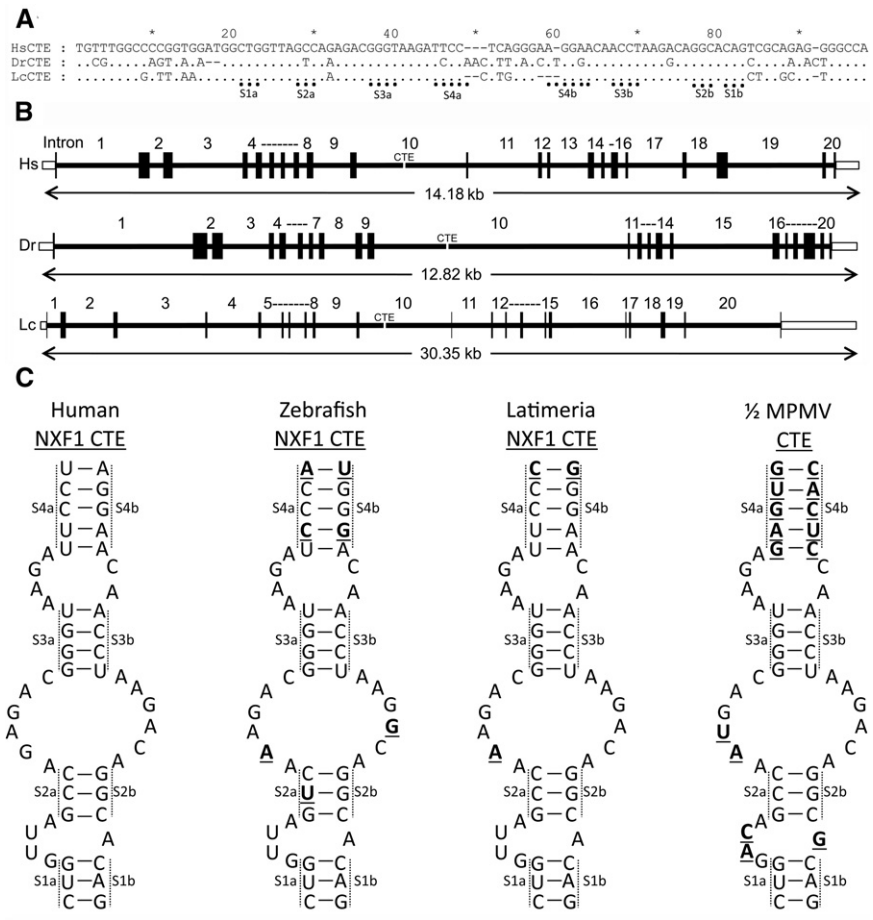

FIGURE 2. (*A*) Alignment of CTE homologs in intron 10 of the human (*Hs*) *NXF1*, the zebrafish (*Dr*) *NXF1*, and the *Latimeria* (Lc) *NXF1* genes. The *top* line shows the sequence of human *NXF1* CTE. A dot (.) shows that the nucleotide is same as the one in the *top* line. A dash (-) shows that there is a gap in the alignment. Regions indicated *below* the sequences correspond to CTE secondary structures involved in stem formation as shown in *C*. (*B*) Schematic representation of the *NXF1* genes from human (ENSG00000162231 11:62559595-62573774:-1), zebrafish (ENSDARG00000055076 21:26058501-26071321:-1), and *Latimeria* (ENSLACG00000001232 JH126593.1:10100-40448:-1). Open boxes represent the 5′ and 3′ untranslated regions. Vertical bars represent exons. Each intron is numbered and the respective length of each gene in kilobases is shown. Intron 10, containing the CTE, is 1801 nt long in the human gene, 3995 nt long in zebrafish, and 3426 nt long in *Latimeria*. (*C*) Secondary structures of CTE core sequences from human, zebrafish, *Latimeria*, and MPMV. Nucleotide variations from the human CTE are shown in bold and underlined.

be quantified by an ELISA targeted to the p24 domain present in Gag and GagPol (Li et al. 2006). Thus the measured level of p24 is a measure of CTE function.

To analyze if the putative DrCTE could function in human cells, we extracted genomic DNA from whole adult zebrafish and used PCR with specific primers to amplify and clone the "CTE" region. We confirmed that the cloned sequence was identical to the *DrNXF1* (chromosome 21) intron 10- sequence deposited in the ZFIN database (data not shown). The DrCTE was then cloned into the Gag-Pol reporter plasmid either as a single copy (DrCTE(1X)) or as a tandem direct repeat containing two copies (DrCTE(2X)). The plasmid with the duplicated sequence serves to mimic the CTEs that are present in type D retroviruses. These CTEs contain two nearly identical single stranded loop regions that both provide binding sites for the Nxf1/Nxt1 proteins, rather than the single loop present in the *NXF1* CTEs. We have previously shown that a similar duplication of the human *NXF1* CTE in the context of our HIV reporter vectors is essential for high level p24 expression from the reporter (Li et al. 2006). We hypothesize that the viral CTE mechanism that uses two binding sites for Nxf1/Nxt1 evolved to allow efficient competition for these host cell proteins during viral replication.

To test DrCTE function, we first transfected either the DrCTE1X or 2X constructs into 293T cells and measured p24 levels in the supernatants. Since we have previously shown that coexpression of additional mammalian Nxf1 and Nxt1 proteins serves to significantly enhance CTE function in these cells, we also performed cotransfections with plasmids expressing mammalian versions of these proteins (Guzik et al. 2001; Jin et al. 2003; Li et al. 2006). As can be seen in Figure 3A, a very low amount of p24 was detected with the 1X CTE construct alone, but the levels were still above the background levels normally obtained without a CTE (Li et al. 2006). However, when plasmids expressing Nxf1 and Nxt1 were added to the transfections, much higher levels of p24 were produced. This is consistent with our previously published results using the human *NXF1* CTE (Li et al. 2006).

As expected, based on our previous results, higher levels of p24 were observed with the 2X CTE construct and these levels were further greatly enhanced in cotransfections with mammalian Nxf1 and Nxt1. As shown in Figure 3B, the amount of p24 increased in a dose-dependent manner, when increasing amounts of the plasmid expressing Nxt1 were transfected with a fixed amount of plasmid expressing Nxf1. These results replicate our previously published results using the human *NXF1* CTE. Taken together, these experiments clearly demonstrated that the zebrafish *NXF1* gene contains a CTE that can function in human cells to enhance expression from an mRNA with a retained intron. Furthermore, the results demonstrate that the mammalian Nxf1 and Nxt1 proteins can functionally interact with the DrCTE to enhance expression of mRNA with a retained intron.

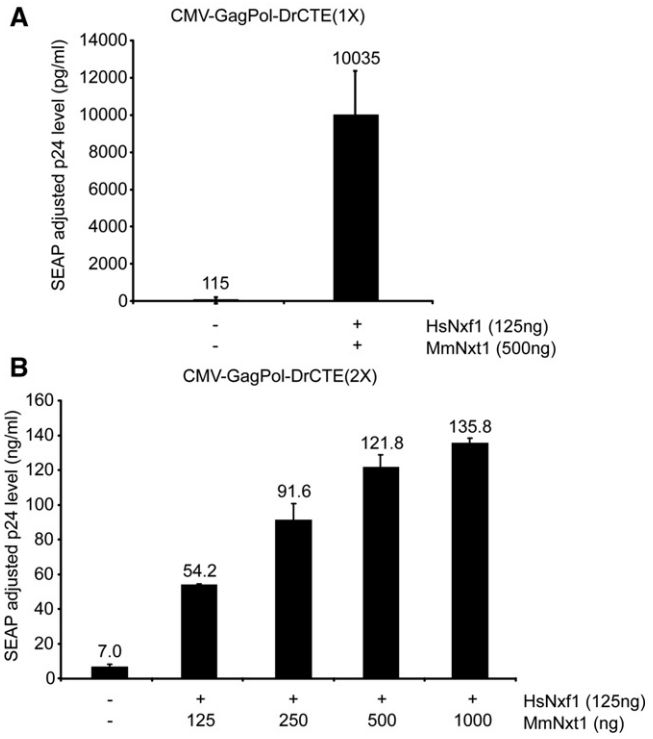

FIGURE 3. Functional analysis of the zebrafish CTE. (*A*) Expression of p24 from CTE reporter plasmid pCMVGagPol-DrCTE(1X) (HR4402) in the presence of human Nxf1 (HsNxf1) and mouse Nxt1 (MmNxt1). 293T cells (1.2×10^6) were transfected with 1.8 µg pCMVGagPol-DrCTE (HR4402) and 90 ng of a plasmid expressing secreted alkaline phosphatase (pCMV-SEAP [HR1831]) to control for transfection efficiency, with or without plasmids expressing *Hs*Nxf1 and *Mm*Nxt1 as indicated in the figure. At 60 h post-transfection, supernatants were collected and analyzed for p24 levels and SEAP activity. p24 values shown in the figure have been normalized for SEAP activity. The data represent the average from two independent transfections. (*B*) Titration analysis of the Nxt1 protein on the function of the duplicated zebrafish CTE in the GagPol reporter construct. 293T cells (1.2×10^6) were transfected with 1.8 µg pCMVGagPol-DrCTE(2X) (HR4466), 90 ng of pCMV-SEAP(HR1831), 125 ng pCMV-HsNxf1 (HR3704), and increasing amounts of pCMV-MmNxt1 (HR2415), as indicated in the figure. Transfection, p24, and SEAP analyses are the same as described in *A*.

The predicted DrNxf1 and DrNxt proteins function in human cells in conjunction with the Dr CTE, as well as with the MPMV and human *NXF1* CTEs

As mentioned above, the *Danio rerio* genome contains two genes predicted to encode Nxf proteins. A sequence comparison between the full-length human Nxf1 protein and the zebrafish Nxf1 protein that is predicted to be expressed from the gene on chromosome 21 containing the CTE, shows that the two proteins display a relatively high degree of homology in most regions (Fig. 4A). However, the region between amino acid 100 and amino acid 129 of the zebrafish protein shows little homology with the corresponding region in the human protein. In this region, the zebrafish protein is distinguished by a large number of glycine residues. This divergent region is amino-terminal to the RNA recognition motif (RRM). The zebrafish protein also has a second shorter polyglycine "repeat" closer to the NH_2-terminus (amino acids 45–55). This is flanked by two apparently redundant nuclear localization signals (R[K/N]KKGRGPFR and RYNPY), conserved in both proteins (Zhang et al. 2011). This region of the human Nxf1 protein has also been shown to be involved in protein dimerization, which appears to be important for function (Matzat et al. 2008).

To be able to test if the zebrafish Nxf1 protein could function in conjunction with CTEs to promote mRNA export and expression in mammalian cells, we made full-length cDNA from total RNA isolated from zebrafish using RT-PCR and specific primers. The complete coding region for the predicted 642-amino acid zebrafish Nxf1 protein was then cloned into a mammalian CMV expression vector containing a Flag-tag, to put an NH_2-terminal Flag epitope in the expressed protein.

Since Nxf1 functions in conjunction with Nxt1 in mammalian cells, we also created a vector that expressed the zebrafish ortholog of Nxt1. In zebrafish, there is only one gene with significant homology with the two mammalian *NXT* genes. The gene was given the name *DrNXT2*, since it appeared more related to human *NXT2* than human *NXT1* (Huang et al. 2005). Several different *DrNXT2* splicing isoforms were described. The mRNA isoform that was of most interest to us was predicted to express a protein of 143 amino acids, which is similar to the size of the human and mouse Nxt1 proteins. The alignment between the two mammalian Nxt1 proteins and the predicted zebrafish protein is shown in Figure 4B. As can be seen in this figure, the human and mouse Nxt1 proteins are identical with the exception of two conservative amino acid changes, whereas they show an overall homology of ~70% to the zebrafish Nxt2 protein. To express the zebrafish protein, we obtained a cDNA containing the 143 amino acid ORF from Thermo Scientific Open Biosystems and cloned it into a mammalian CMV expression vector, in frame with an NH_2-terminal Flag-tag.

The Nxf/Nxt plasmids were then tested for their ability to enhance the expression from pCMVGagPol-DrCTE(2X). As can be seen in Figure 5, expression of DrNxf1 alone enhanced p24 levels almost sevenfold, similar to the enhancement seen with the human Nxf1 protein. In contrast, the DrNxt2 protein alone did not enhance function. However, a combination of DrNxt2 and DrNxf1 resulted in very high levels of p24 expression (nearly a 1300-fold increase), which were about fivefold higher than the levels obtained with the mammalian proteins. The combination of HsNxf1 and DrNxt2 worked almost as well as HsNxf1 and MmNxt1, whereas DrNxf1 and MmNxt1 gave lower levels. Figure 5B is a Western blot with a Flag antibody that shows expression of all of the proteins with the exception of MmNxt1 that lacks a Flag tag. However, this protein was also expressed well as judged by the increase in p24 that was obtained in conjunction with Nxf1. The MmNxt1 protein is expressed from the same plasmid used in many of our previous publications (Guzik et al. 2001; Jin et al. 2003; Li et al. 2006).

```
A
                *       20        *       40        *       60        *       80        *      100        *
HsNxf1 : ~~MADEGKSYSEHDDERVNFPQRKKKGRGPFRWKYG----------EGNRRSGRGGSGIRSSRLEEDDGDVAMSDAQ-DGPRVRYNPYTTRPNRRGDTWHDRDR---
DrNxf1 : MAT..DS.Y.N....RVGGPKH.N........APLYSDQMSRPRHRGGHS.GGGG.G..GPGPR...HD.....T...IPQ.SSQR.....GRQQ...D.RNRKGG.GGG

              120       *       140       *       160       *       180       *       200       *       220
HsNxf1 : ---IHVTVRRDRAPPERGGAGTSQDGTSKNWFKITIPYGRKYDKAWLLSMIQSKCSVPFTPIEFHYENTRAQFFVEDASTASALKAVNYKILDRENRRISIIINSSAPPH
DrNxf1 : GNDDRGGN.FGGRGGGG..G.GPS..R.................D....S...V.P...HA....LDGN........S...N..FK.SR..T..DGY..A.....CP..T

                *       240       *       260       *       280       *       300       *       320       *
HsNxf1 : TILNELKPEQVEQLKLIMSKRYDGSQQALDLKGLRSDPDLVAQNIDVVLNRRSCMAATLRIIEENIPELLSLNLSNNRLYRLDDMSSIVQKAPNLKILNLSGNELKSERE
DrNxf1 : L.HSD...AD..H..QC.A............NN..V.....S...E.T....NS.C.V............C..............VD..H.V.........H........

              340       *       360       *       380       *       400       *       420       *       440
HsNxf1 : LDKIKGLKLEELWLDGNSLCDTFRDQSTYISAIRERFPKLLRLDGHELPPPIAFDVEAP-TTLPPCKGSYFGTENLKSLVLHFLQQYYAIYDSGDRQGLLDAYHDGACCS
DrNxf1 : .........VD.S.E..P..HH..IHAD..................D.....G...DDGP.....S.A...C..E..N...C......N........P.........TF.

                *       460       *       480       *       500       *       520       *       540       *
HsNxf1 : LSIPFIPQNPARSSLAEYFKDSRNVKKLKDPTLRFRLLKHTRLNVVAFLNELPKTQHDVNSFVVDISAQTSTLLCFSVNGVFKEVDGKSRDSLRAFTRTFIAVPASNSGL
DrNxf1 : ......SL..S....G..Q.............T.........................A......NTY.A...A...C...........E.Y.....V......GS...

              560       *       580       *       600       *       620       *       640
HsNxf1 : CIVNDELFVRNASSEEIQRAFAMPAPTPSSSPVPTLSPEQQEMLQAFSTQSGMNLEWSQKCLQDNNWDYTRSAQAFTHLKAKGEIPEVAFMK
DrNxf1 : ................R...VA..............AH.....A...LK...................E.A..I..G...Q.K..D.....

B
                *       20        *       40        *       60        *       80        *      100        *
MmNxt1 : M-ASVDFKTYVDQACRAAEEFVNVYYTTMDKRRRLLSRLYMGTATLVWNGNAVSGQESLSEFFEMLPSSEFQISVVDCQPVHDDATPSQTTVLVVICGTVKFEGNKQRDF
HsNxt1 : .-.....................................................................E.............S...........
DrNxt2 : .SLT...R.Q...S..YS...I.I..EC...K..N.K...LDK.........T...A.G....S.......VQTL......EQ..QG...L...AA.S.........F.

              120       *       140
MmNxt1 : NQNFILTAQASPSN--TVWKIASDCFRFQDWAS
HsNxt1 : .............--.................
DrNxt2 : ....L.......NSDQP..............N
```

FIGURE 4. (*A*) Sequence alignment of the human Nxf1 (HsNxf1; GenBank accession no. Q9UBU9) and the zebrafish Nxf1 (DrNxf1; GenBank accession no. XM_001923926) proteins. (*B*) Alignment of the mouse Nxt1 (MmNxt1; GenBank accession no. NP_001103629), the human Nxt1 (HsNxt1; GenBank accession no. NP_037380), and the zebrafish Nxt2 (DrNxt2; GenBank accession no. DT866593) proteins. In both *A* and *B* identical amino acids are marked with dots. Dashes indicate gaps in the alignment.

Most importantly these results clearly show that all the known basic components of the CTE-mediated pathway necessary for export and expression of mRNA with retained introns (the CTE and the Nxf1 and Nxt proteins) can be successfully transplanted from zebrafish to human cells. The results also clearly demonstrate that no other zebrafish proteins are essential for efficient export and expression through the CTE in mammalian cells. This indicates that the zebrafish Nxf1/Nxt2 complex can interact efficiently with the human machinery involved in export and expression of CTE-containing mRNA.

To analyze if the zebrafish proteins could also function in conjunction with the MPMV and the human *NXF1* CTEs, we performed similar experiments with reporters containing these elements (Supplemental Figs. S1, S2). These experiments showed that the DrNxf1 and DrNxt2 proteins function equally well as the mammalian proteins in conjunction with both of these CTEs, further demonstrating the functional conservation.

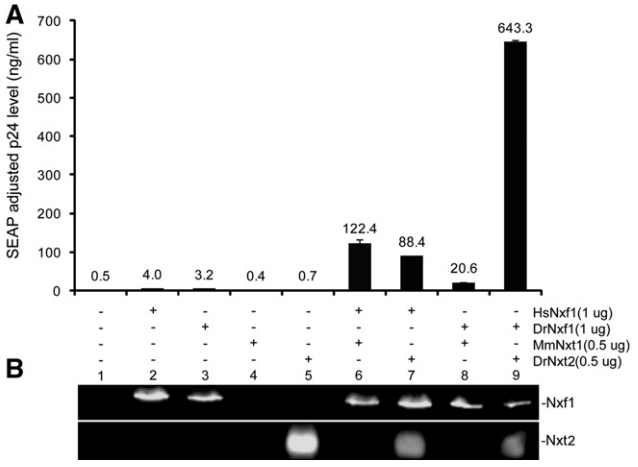

FIGURE 5. Comparison of zebrafish CTE function in 293T cells with added Nxf1 and Nxt from zebrafish and mammals. (*A*) 293T cells (1.2×10^6) were transfected with 1.8 µg of the CTE-reporter plasmid pCMVGagPol-DrCTE(2X) (HR4466), 90 ng of pCMV-SEAP (HR1931), and plasmids expressing zebrafish or mammalian Nxf1 and Nxt proteins as indicated in the figure. The HsNxf1, DrNxf1, and DrNxt2 proteins contain an amino-terminal Flag-epitope tag. The MmNxt1 protein is not Flag-tagged. DNA transfections, p24 and SEAP activity analyses were performed as described in the legend of Figure 3A. (*B*) Western blot analysis of protein expression in the transfected 293T cells. Lysates from the transfected cells were separated on denaturing SDS-PAGE gels and analyzed by Western blotting. Blots were probed with an M2 monoclonal anti-Flag antibody (Sigma) and detected with a goat anti-mouse IgG antibody-IRDye 800. Blots were visualized and analyzed using the LI-COR Odyssey infrared imaging system.

The *Danio rerio* CTE promotes nucleocytoplasmic export of intron-containing RNA and higher levels of cytoplasmic RNA are observed when the *Dr* Nxf1 and Nxt2 proteins are coexpressed

Previous studies have shown that mammalian Nxf1/Nxt1 proteins work in conjunction with a CTE to enhance

expression from mRNA with retained introns both at the level of nucleocytoplasmic RNA export and protein translation in the cytoplasm (Bray et al. 1994; Ernst et al. 1997a; Jin et al. 2003). To further investigate how the *DrNXF1* CTE and *Dr* Nxf1/Nxt2 proteins function to promote p24 protein expression, we performed a Northern blot analysis. In these experiments, we compared levels of cytoplasmic and total RNA, from cells transfected with the GagPol plasmid containing the zebrafish *NXF1* CTE, in the absence and presence of cotransfected DrNxf1/Nxt2. As controls in these experiments, we used RNA from cells transfected with an HIV GagPol-RRE reporter plasmid, in the presence and absence of the HIV Rev protein, since Rev and the RRE are known to efficiently mediate both the export and translation of the mRNA produced from this reporter (Coyle et al. 2003). All cells were also transfected with a plasmid expressing SEAP as a normalization control.

The results of these experiments (Fig. 6) showed that in the absence of cotransfected Nxf1/Nxt2, only low levels of cytoplasmic GagPol-CTE RNA were seen with the DrCTE-containing reporter plasmid (panel B). These levels increased 48-fold when the DrNxf1 and Nxt2 proteins were coexpressed. To determine if the increase in cytoplasmic GagPol-DrCTE RNA was due to an increase in stability or an increase in export, we also examined total RNA (nuclear + cytoplasmic). In total RNA, coexpression of Nxf1/Nxt2 (panel A) also resulted in higher levels of GagPol-DrCTE RNA, but the increase was only 6.4-fold. This mimics results previously obtained with the human Nxf1/Nxt1 proteins and suggests that the increase of GagPol-DrCTE RNA in total mRNA is a reflection of the fact that the RNA is more stable, when it is translated in the cytoplasm, compared with when it is sequestered in the nucleus (Bray et al. 1994; Jin et al. 2003). Together these results indicate a significant 7.5-fold (48/6.4) specific increase in mRNA export in the presence of the zebrafish Nxf1/Nxt2 proteins.

The control experiment with Rev/RRE gave the expected result (panels C and D). In this case, 12-fold higher levels of GagPol-RRE RNA were observed in the cytoplasm in the presence of Rev, whereas there was a 2.6-fold increase in total RNA. Thus Rev and the HIV RRE increased RNA export ~4.5-fold.

In summary, these results clearly show that the zebrafish CTE promotes export of mRNA with retained introns in a manner that is significantly enhanced by the expression of the cognate Nxf1 and Nxt2 proteins in human cells. However, the 48-fold increase in cytoplasmic RNA when zebrafish Nxf1/Nxt2 were coexpressed (panel B) is still

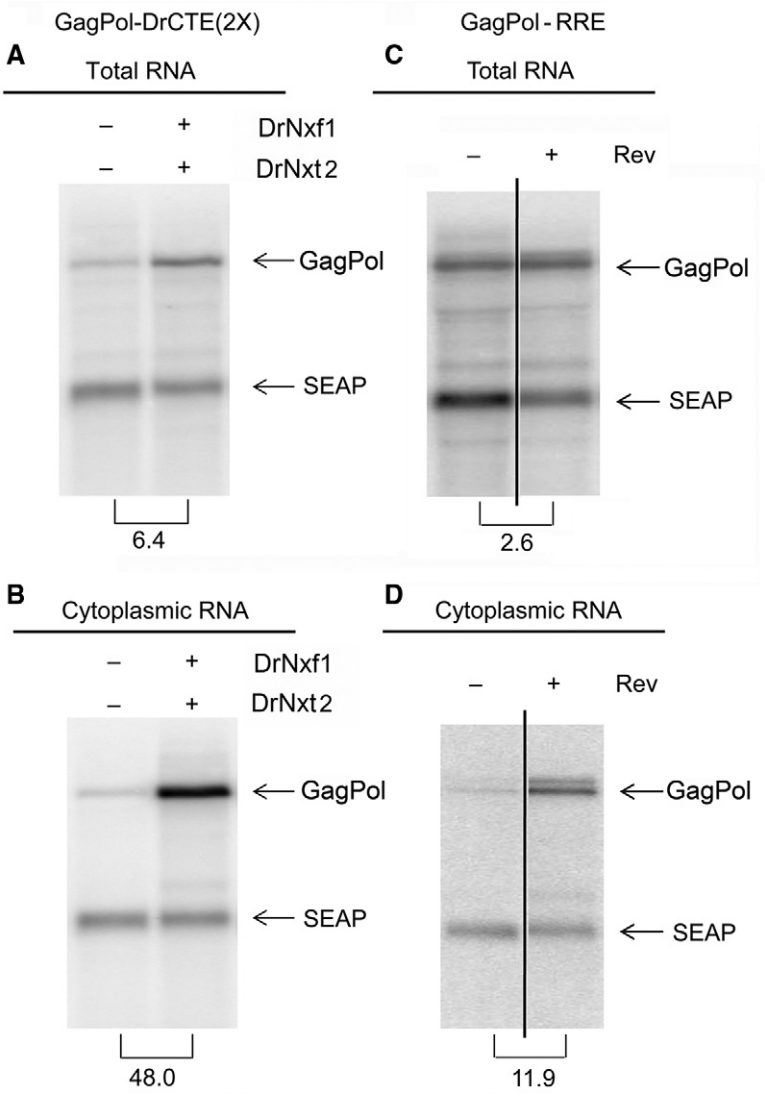

FIGURE 6. Northern blot analysis of total and cytoplasmic GagPol-DrCTE or GagPol-RRE mRNA from transfected 293T cells. (*A*,*B*) 293T cells (1×10^7 in a 15-cm culture dish) were transfected with 15 µg pCMVGagPol-DrCTE(2X) (HR4466) and 1 µg pCMV-SEAP (HR1831) with or without plasmids that expressed zebrafish Nxf1 (5 µg) and zebrafish Nxt2 (2.5 µg) proteins. (*C*,*D*) 293T cells (1×10^7 in a 15-cm culture dish) were transfected with 15 µg pCMVGagPol-RRE (pHR354) and 1 µg pCMV-SEAP (pHR1831) with or without plasmid pHR30 (5 µg) that expresses the HIV Rev protein. Fifty-five hours post-transfection, total (*A*,*C*) and cytoplasmic (*B*, *D*) mRNAs were isolated from the transfected cells as described in Materials and Methods. Blots containing mRNAs (5 µg/lane) were hybridized with ^{32}P-labeled GagPol and SEAP probes. Brackets show the fold difference in the levels of the GagPol mRNA between the indicated lanes after normalization for SEAP RNA levels. In *C* and *D* the individual lanes ±Rev were put together electronically from different parts of the original gel.

much lower than the 1300-fold increase in p24 protein expression (0.5 versus 643 ng/mL) (see Fig. 5). Thus the zebrafish Nxf1 and Nxt2 proteins appear able to also promote expression after export to the cytoplasm, as was demonstrated previously for the human proteins (Jin et al. 2003).

Intron retention in alternatively spliced *DrNXF1* transcripts

Since the human *NXF1* CTE has been shown to regulate the expression of an alternatively spliced *NXF1* mRNA that retains the CTE-containing intron 10, we next decided to analyze whether this intron is also retained in alternatively spliced *NXF1* mRNA in zebrafish. Several EST entries support this notion and there is a predicted alternative mRNA isoform of 6461 nt that retains intron 10 (OTTDART 00000049512).

To further investigate this, we extracted total RNA from adult zebrafish and isolated mRNA using poly(A) selection. The RNA was then subjected to Northern blot analysis using two different *DrNXF1* probes. One probe was an "exon" probe complementary to a 580 nt sequence at the 5′ end of the predicted completely spliced *DrNXF1* mRNA (exons 1, 2, and part of exon 3). This probe would be expected to detect all predicted mRNA isoforms expressed from the *DrNXF1* gene. The second probe was complementary to a 1.6 kb sequence in *DrNXF1* intron 10, containing the DrCTE and surrounding sequences.

Northern blot analysis using the "exon" probe showed multiple bands, suggesting the presence of multiple *NXF1* alternative mRNA isoforms (Fig. 7). The predominant band was ~2.6 kb, corresponding to the expected size of the completely spliced *DrNXF1* mRNA (GenBank: XM_00192 3926) that expresses the "large" Nxf1 protein. Two larger bands were also detected. Both of these were also detected using the intron probe, suggesting that both of these mRNAs retain intron 10. The smaller of the two bands corresponds to an mRNA of ~6.6 kb, the expected size of an mRNA that retains only intron 10 (in addition to all of the exons present in the ~2.6 kb). The larger band may represent a species that retains additional introns. There is also some reactivity with smaller mRNAs, potentially indicating alternatively spliced forms retaining all or part of intron 10. In summary, the Northern blot analysis demonstrated that adult zebrafish expresses mRNA that retains intron 10, suggesting that CTE regulation of *NXF1* expression operates also in zebrafish. Just as in the case of the mammalian *NXF1* genes, the mRNA that retains intron 10 would be expected to express a short Nxf1 protein, with a few C-terminal amino acids encoded by intron 10 sequences, before a stop codon is encountered. This short protein would contain the two nuclear localization signals and most of the predicted RNA-binding regions, but would lack the predicted regions involved in interactions with Nxt and nuclear pore components.

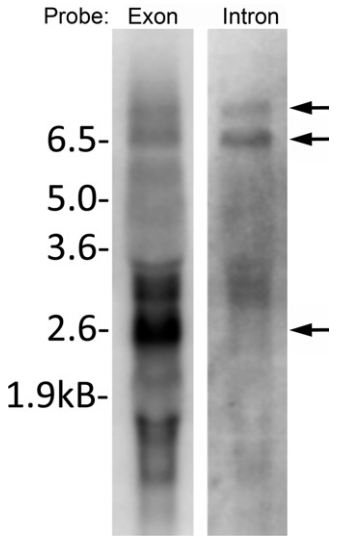

FIGURE 7. Northern blot analysis of the zebrafish *NXF1* mRNAs. Total RNA was extracted from adult zebrafish and selected with oligo(dT)-cellulose and analyzed on Northern blots as described in Materials and Methods. Blots containing mRNAs (5 μg/lane) were hybridized with either a zebrafish *NXF1* exon-specific probe (*left* lane) or an intron 10-specific probe (*right* lane). The positions of molecular weight markers running in a parallel lane are indicated. The arrows point to the major bands discussed in the text.

DISCUSSION

The data presented here clearly demonstrate that CTEs have been highly conserved in mammalian evolution and that a remarkable primary sequence and secondary structure conservation extends beyond mammals to teleost fish. When the CTE was first discovered, little was known about mRNA export and the CTE was thought to be a virus-specific element (Bray et al. 1994). At the time, one of the most studied mRNA export systems was in HIV, where the viral Rev protein binds to the viral RRE to promote export of mRNAs that retain introns. The fact that MPMV was shown to use a *cis*-acting RNA element to achieve export of unspliced mRNA, in the absence of a viral *trans*-acting factor, led to the hypothesis that a host cell protein must recognize the CTE, and perform a function similar to Rev, even though there was little evidence at the time that specific proteins were involved in cellular mRNA export.

Groundbreaking work from Grüter et al. (1998) subsequently showed that the cellular Nxf1 protein (then known as Tap), bound directly to two internal loop regions in the MPMV CTE. Further work from many groups has suggested that the Nxf1 protein functions as an important mRNA export receptor for a large number of cellular mRNAs (Natalizio and Wente 2013). However, since CTE-like elements are not found in most mRNAs, Nxf1 has been proposed to function in conjunction with RNA binding adaptor proteins, rather than through direct RNA binding (Huang et al. 2004; Hautbergue et al. 2008). It was suggested

that the CTE evolved in MPMV to "mimic" these protein adaptors and directly recruit Nxf1 to the mRNA, perhaps allowing it to "lock in" on cellular mRNA without the help of protein adaptors (Hautbergue et al. 2008). However, our discovery of a highly related CTE in the human *NXF1* gene led us to hypothesize that the CTE is of host cell, rather than viral, origin (Li et al. 2006).

Before this study, it remained a formal possibility that the *NXF1* gene might have acquired the CTE from an integrating retrovirus. Our data showing strong primary sequence conservation of the CTE in evolution, and the functional conservation of the Nxf1/CTE RNA export mechanism, now provide strong evidence in support of the hypothesis that the CTE was acquired from a cell by an ancestral retrovirus to promote efficient export of its intron-retaining RNA genome in infected host cells. Since the MMPV CTE contains two Nxf1 binding loops, but all of the cellular *NXF* CTEs identified to date contain only one, it seems highly likely that the original cellular element was subsequently duplicated in the virus for better function and competition with cellular mRNAs.

Whereas exon conservation in the *NXF* genes is expected because of the important conserved roles that the Nxf proteins play in post-transcriptional gene regulation, intron sequences in protein coding genes generally show little conservation. The striking primary CTE sequence conservation in multiple species, not only in the internal loop sequence that binds Nxf/Nxt, but also in surrounding regions, was thus surprising. Specifically, there is a conserved 5′ splice site sequence close to the internal loop. While it was previously shown that this site could be used for splicing in the context of a reporter plasmid (Pasquinelli et al. 1997), there is no evidence that it is utilized in the context of any of the *NXF1* genes. One possibility is that it may serve as a binding site for U1 snRNP, which may be needed to stabilize the intron 10-containing mRNA, in much the same way that U1 snRNP has been shown to stabilize the HIV mRNA containing the envelope protein intron (Lu et al. 1990; Kammler et al. 2001). Specific mutagenesis experiments will be necessary to further test this hypothesis.

Since no zebrafish components are required for efficient CTE function in mammalian cells beyond Nxf1/Nxt, our results indicate that these proteins are sufficiently functionally conserved to allow interactions with all of the other components of the human cell machinery (such as NPC components) required for export and expression of CTE-containing mRNA with retained introns. In fact, using all three components from zebrafish in human cells, we observed even higher expression than when using mammalian Nxf1/Nxt. This might indicate that human cells contain modulating "negative" factors that are not able to interact with the zebrafish components.

Taken together, the demonstrated conservation of the CTE and our other data suggest that export and expression of the alternative *NXF1* mRNA that retains intron 10 serves an important cellular function. In human cells, we previously showed that the *NXF1* mRNA that retains intron 10 is translated into a "small" Nxf1 protein (Li et al. 2006). We have obtained preliminary data that indicate that "small" Nxf1 may function in neuronal RNA trafficking, since it is highly expressed in hippocampal and cortical regions of the brain in rats and colocalizes with Staufen proteins in cytoplasmic granules in mouse neuronal cells. Other preliminary data show that the small Nxf1 protein can interact with the "large" Nxf1 protein to function as an alternative partner to Nxt1 in mRNA trafficking. Although we do not yet know if a similar small Nxf1 protein is expressed in zebrafish, the fact we have detected mRNAs that retain the CTE containing intron 10 suggests that this may be the case.

Very little is known about the function of the *NXF1* and *NXT* genes in zebrafish. Defects in the *NXF1* gene have been shown to lead to early embryonic death, whereas a defect in the *NXT* gene has been associated with heart malformation (Huang et al. 2005). The gene identified was referred to as *NXT2* because of slightly higher homology with the human *NXT2* gene. However, we have not been able to find another *NXT* gene in zebrafish and therefore *NXT2* may be a misnomer. Using CRISPR/Cas technology, it should now be possible to further study *NXF1*/*NXT* function in this species, as well as consequences of deleting the CTE.

Although a region with strong homology with the CTE can be detected in most mammalian species, there is less conservation in a few of the organisms sequenced to date. Although there is still enough homology for detection by GERP analysis, the potential CTE sequences detected in *Sorex araneus* (common shrew), *Dasypus novimcinctus* (Nine-Banded Armadillo), and *Tursiops truncatus* (Common Bottle-nosed Dolphin) are much less conserved in sequence and predicted secondary structure. In spite of this, it is possible that CTE function may be conserved. Using bioinformatic tools, we have not been able to find any genes outside of the *NXF* family that possess CTEs based on sequence homology. However, the genomes of several mammalian and endogenous retroviruses have been shown to contain non-homologous elements that function as CTEs in conjunction with Nxf1/Nxt (Ogert and Beemon 1995; Yang and Cullen 1999; Nappi et al. 2001; Legiewicz et al. 2010; Sakuma et al. 2014). In addition, we have previously used retroviral vector "trap" strategies to identify several host cell CTEs that lack apparent homology with the MPMV/*NXF1* CTE, but which still function in conjunction with Nxf1/Nxt1 (Bor et al. 2006). However, more experimentation will be needed to determine how commonly CTE and/or CTE-like elements are used to regulate expression of mRNA with retained introns.

MATERIALS AND METHODS

Zebrafish maintenance, DNA and RNA extraction

Wild-type (AB strain) zebrafish (*Danio rerio*) were maintained on a 14-h light/10-h dark cycle. Adult zebrafish were transferred into a

5 μM nicotine solution (N3876; Sigma), flash-frozen in liquid nitrogen, and stored at −80°C for future use. Zebrafish genome DNA was extracted using the Phenol:Chloroform method (Westerfield 2007). For zebrafish total RNA extraction, frozen zebrafish were ground with a chilled mortar and pestle in the presence of liquid nitrogen. The powdered tissue was then mixed with Tri Reagent (MRCgene), and RNA was extracted following the manufacturer's instructions.

Bioinformatics

Analysis of sequence conservation in the *NXF1* gene intron 10 of different mammalian species was facilitated using genomic evolutionary rate profiling (GERP) (http://mendel.stanford.edu/SidowLab/downloads/gerp/index.html) and the Ensembl genome browser (http://www.ensembl.org/index.html). CTE sequence BLAST searches against other *NXF1* genes were performed using National Center for Biotechnology BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and with the Ensembl genome browser (http://www.ensembl.org/Multi/Tools/Blast). RNA secondary structure analyses were performed using Mfold (http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form). Sequence alignments were created using the EMMA program of the EMBOSS. (EMBOSS: The European Molecular Biology Open Software Suite [2000]) available at http://emboss.sourceforge.net/interfaces/#jemboss and manipulated using GeneDoc (http://www.nrbsc.org/gfx/genedoc/ebinet.htm) (Rice et al. 2000).

Plasmids

Each plasmid described below is referred to by both name and number (in the form of pHRXXXX), to facilitate identification. Some of the plasmids used in this study have been described in previous publications. Plasmid pCMV-HsNxf1 (pHR3704) expresses human Nxf1 (Li et al. 2006). pCMV-SEAP (pHR1831) expresses secreted alkaline phosphatase (SEAP) (Berger et al. 1988). pCMV-MmNxt1 (pHR2415) expresses mouse Nxt1 (Black et al. 1999) and pCMV-Rev(pHR30) expresses HIV Rev (Smith et al. 1993). The subgenomic HIV-1 reporter vector pCMVGagPol-U2 (pHR2739) contains an XhoI–U2RNA–BamHI fragment that was replaced with CTE elements (Li et al. 2006). pCMVGagPol-MPMV-CTE (pHR1361) contains the MPMV-CTE (Srinivasakumar et al. 1997) and pCMVGagPol-HsCTE(2X) (pHR3406) contains two tandem copies of the human *NXF1* CTE (Li et al. 2006).The reporter plasmid pCMVGagPol-RRE (pHR354) contains the HIV1 RRE instead of U2 (Srinivasakumar et al. 1997).

Some of the plasmids used in this study were newly constructed. These are as follows: pCMVGagPol-DrCTE(1X) (pHR4402), pCMVGagPol-DrCTE(2X), (pHR4466)pCMV-DrNxf1 (pHR4488) and pCMV-DrNxt2 (pHR4525). Details of their construction are given in the Supplemental Material.

293T cell maintenance and transfection

293T cells were maintained in Iscove's minimal essential medium supplemented with 10% bovine calf serum and 0.1% gentamicin. 293T cell transfections were performed by using the calcium phosphate method as described previously (Graham and van der Eb 1973). Transfections were performed with the amounts of plasmids described in the figure legends. Plasmid DNAs were propagated in *Escherichia coli* and prepared with *Qiagen Maxiprep* kit (Qiagen).

p24 ELISA and secreted alkaline phosphatase (SEAP) assays

Supernatants from transfected cells were collected at the indicated times post-transfection and subjected to a short spin in a microcentrifuge to remove residual cells and debris. p24 ELISA and SEAP assays were performed as previously described (Coyle et al. 2011).

Northern blot analysis

Transfected 293T cells were harvested at 55 h post-transfection for total and cytoplasmic RNA extraction and Northern blot analysis as previously described (Hammarskjöld et al. 1986, 1989, 1994). The GagPol DNA fragment used to probe the blots corresponded to nucleotides 682–2093 of the HIV-1 BH10 clone, and the SEAP probe corresponded to nucleotides 213–1698 of the human SEAP cDNA.

For the zebrafish *NXF1* mRNA Northern blot, a 1.6-kb cloned zebrafish *NXF1* intron 10 DNA fragment was used as an intron probe and a DNA fragment corresponding to nucleotides 1–580 of the zebrafish *NXF1* cDNA (NCBI: XM_001923926) was used as the exon probe.

Western blot analysis

For Western blot analyses, lysates of transfected 293T cells were separated on a 13% SDS-polyacrylamide gel and transferred to an Immobilon-FL membrane (Millipore). For detection of Flag-tagged proteins, the blots were probed with a 1:5000 dilution of M2 anti-Flag monoclonal antibody (Sigma), followed by incubation with a goat anti-mouse IgG antibody-IRDye 800 (1:25,000 dilution; LI-COR). The blots were scanned on the Odyssey infrared imager and the protein levels were quantified with the Odyssey software.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Dr. Bernard Thisse and Dr. Christine Thisse for generously providing zebrafish and the DNA fragment containing 500 bp of the *DrNXF1* coding region. We also thank Dr. Yeou-cherng Bor for technical advice. This work was supported by National Institutes of Health (NIH) R01 GM087651 and funds from the Myles H. Thaler Center research Endowment Fund at the University of Virginia. Salary support for D.R. and M.L.H. was provided by the Myles H. Thaler and Charles H. Ross Jr. Endowments at the University of Virginia.

Received October 9, 2014; accepted December 9, 2014.

REFERENCES

Berger J, Hauber J, Hauber R, Geiger R, Cullen BR. 1988. Secreted placental alkaline phosphatase: a powerful new quantitative indicator of gene expression in eukaryotic cells. *Gene* **66:** 1–10.

Berke JD, Sgambato V, Zhu PP, Lavoie B, Vincent M, Krause M, Hyman SE. 2001. Dopamine and glutamate induce distinct striatal splice forms of Ania-6, an RNA polymerase II-associated cyclin. *Neuron* **32:** 277–287.

Black BE, Lévesque L, Holaska JM, Wood TC, Paschal BM. 1999. Identification of an NTF2-related factor that binds Ran-GTP and regulates nuclear protein export. *Mol Cell Biol* **19:** 8616–8624.

Bor YC, Swartz J, Morrison A, Rekosh D, Ladomery M, Hammarskjöld ML. 2006. The Wilms' tumor 1 (WT1) gene (+KTS isoform) functions with a CTE to enhance translation from an unspliced RNA with a retained intron. *Genes Dev* **20:** 1597–1608.

Bray M, Prasad S, Dubay JW, Hunter E, Jeang KT, Rekosh D, Hammarskjöld ML. 1994. A small element from the Mason–Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proc Natl Acad Sci* **91:** 1256–1260.

Cooper GM, Stone EA, Asimenos G; NISC Comparative Sequencing Program, Green ED, Batzoglou S, Sidow A. 2005. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* **15:** 901–913.

Coyle JH, Guzik BW, Bor YC, Jin L, Eisner-Smerage L, Taylor SJ, Rekosh D, Hammarskjöld ML. 2003. Sam68 enhances the cytoplasmic utilization of intron-containing RNA and is functionally regulated by the nuclear kinase Sik/BRK. *Mol Cell Biol* **23:** 92–103.

Coyle JH, Bor YC, Rekosh D, Hammarskjöld ML. 2011. The Tpr protein regulates export of mRNAs with retained introns that traffic through the Nxf1 pathway. *RNA* **17:** 1344–1356.

Doherty JK, Bond C, Jardim A, Adelman JP, Clinton GM. 1999. The HER-2/neu receptor tyrosine kinase gene encodes a secreted autoinhibitor. *Proc Natl Acad Sci* **96:** 10869–10874.

Dytrych L, Sherman DL, Gillespie CS, Brophy PJ. 1998. Two PDZ domain proteins encoded by the murine periaxin gene are the result of alternative intron retention and are differentially targeted in Schwann cells. *J Biol Chem* **273:** 5794–5800.

Ernst RK, Bray M, Rekosh D, Hammarskjöld ML. 1997a. A structured retroviral RNA element that mediates nucleocytoplasmic export of intron-containing RNA. *Mol Cell Biol* **17:** 135–144.

Ernst RK, Bray M, Rekosh D, Hammarskjöld ML. 1997b. Secondary structure and mutational analysis of the Mason-Pfizer monkey virus RNA constitutive transport element. *RNA* **3:** 210–222.

Felber BK, Hadzopoulou-Cladaras M, Cladaras C, Copeland T, Pavlakis GN. 1989. rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc Natl Acad Sci* **86:** 1495–1499.

Forrest ST, Barringhaus KG, Perlegas D, Hammarskjöld ML, McNamara CA. 2004. Intron retention generates a novel Id3 isoform that inhibits vascular lesion formation. *J Biol Chem* **279:** 32897–32903.

Ge Y, Porse BT. 2014. The functional consequences of intron retention: alternative splicing coupled to NMD as a regulator of gene expression. *Bioessays* **36:** 236–243.

Graham FL, van der Eb AJ. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52:** 456–467.

Grüter P, Tabernero C, von Kobbe C, Schmitt C, Saavedra C, Bachi A, Wilm M, Felber BK, Izaurralde E. 1998. TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol Cell* **1:** 649–659.

Guzik BW, Levesque L, Prasad S, Bor YC, Black BE, Paschal BM, Rekosh D, Hammarskjöld ML. 2001. NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of intron-containing RNA in mammalian cells. *Mol Cell Biol* **21:** 2545–2554.

Hammarskjöld ML, Wang SC, Klein G. 1986. High-level expression of the Epstein-Barr virus EBNA1 protein in CV1 cells and human lymphoid cells using a SV40 late replacement vector. *Gene* **43:** 41–50.

Hammarskjöld ML, Heimer J, Hammarskjöld B, Sangwan I, Albert L, Rekosh D. 1989. Regulation of human immunodeficiency virus env expression by the rev gene product. *J Virol* **63:** 1959–1966.

Hammarskjöld ML, Li H, Rekosh D, Prasad S. 1994. Human immunodeficiency virus env expression becomes Rev-independent if the env region is not defined as an intron. *J Virol* **68:** 951–958.

Hautbergue GM, Hung ML, Golovanov AP, Lian LY, Wilson SA. 2008. Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. *Proc Natl Acad Sci* **105:** 5154–5159.

Huang Y, Yario TA, Steitz JA. 2004. A molecular link between SR protein dephosphorylation and mRNA export. *Proc Natl Acad Sci* **101:** 9666–9670.

Huang H, Zhang B, Hartenstein PA, Chen JN, Lin S. 2005. NXT2 is required for embryonic heart development in zebrafish. *BMC Dev Biol* **5:** 7.

Jin L, Guzik BW, Bor YC, Rekosh D, Hammarskjöld ML. 2003. Tap and NXT promote translation of unspliced mRNA. *Genes Dev* **17:** 3075–3086.

Kammler S, Leurs C, Freund M, Krummheuer J, Seidel K, Tange TO, Lund MK, Kjems J, Scheid A, Schaal H. 2001. The sequence complementarity between HIV-1 5′ splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-mRNA. *RNA* **7:** 421–434.

Kurio H, Murayama E, Kaneko T, Shibata Y, Inai T, Iida H. 2008. Intron retention generates a novel isoform of CEACAM6 that may act as an adhesion molecule in the ectoplasmic specialization structures between spermatids and sertoli cells in rat testis. *Biol Reprod* **79:** 1062–1073.

Legiewicz M, Zolotukhin AS, Pilkington GR, Purzycka KJ, Mitchell M, Uranishi H, Bear J, Pavlakis GN, Le Grice SF, Felber BK. 2010. The RNA transport element of the murine musD retrotransposon requires long-range intramolecular interactions for function. *J Biol Chem* **285:** 42097–42104.

Lejeune F, Maquat LE. 2005. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr Opin Cell Biol* **17:** 309–315.

Li Y, Bor YC, Misawa Y, Xue Y, Rekosh D, Hammarskjöld ML. 2006. An intron with a constitutive transport element is retained in a Tap messenger RNA. *Nature* **443:** 234–237.

Liggins AP, Brown PJ, Asker K, Pulford K, Banham AH. 2004. A novel diffuse large B-cell lymphoma-associated cancer testis antigen encoding a PAS domain protein. *Br J Cancer* **91:** 141–149.

Lu XB, Heimer J, Rekosh D, Hammarskjöld ML. 1990. U1 small nuclear RNA plays a direct role in the formation of a rev-regulated human immunodeficiency virus env mRNA that remains unspliced. *Proc Natl Acad Sci* **87:** 7598–7602.

Lupetti R, Pisarra P, Verrecchia A, Farina C, Nicolini G, Anichini A, Bordignon C, Sensi M, Parmiani G, Traversari C. 1998. Translation of a retained intron in tyrosinase-related protein (TRP) 2 mRNA generates a new cytotoxic T lymphocyte (CTL)-defined and shared human melanoma antigen not expressed in normal cells of the melanocytic lineage. *J Exp Med* **188:** 1005–1016.

Malim MH, Hauber J, Le SV, Maizel JV, Cullen BR. 1989. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338:** 254–257.

Matzat LH, Berberoglu S, Levesque L. 2008. Formation of a Tap/NXF1 homotypic complex is mediated through the amino-terminal domain of Tap and enhances interaction with nucleoporins. *Mol Biol Cell* **19:** 327–338.

Michael IP, Kurlender L, Memari N, Yousef GM, Du D, Grass L, Stephan C, Jung K, Diamandis EP. 2005. Intron retention: a common splicing event within the human kallikrein gene family. *Clin Chem* **51:** 506–515.

Müller-McNicoll M, Neugebauer KM. 2013. How cells get the message: dynamic assembly and function of mRNA–protein complexes. *Nat Rev Genet* **14:** 275–287.

Nappi F, Schneider R, Zolotukhin A, Smulevitch S, Michalowski D, Bear J, Felber BK, Pavlakis GN. 2001. Identification of a novel post-transcriptional regulatory element by using a rev- and RRE-mutated

human immunodeficiency virus type 1 DNA proviral clone as a molecular trap. *J Virol* **75:** 4558–4569.

Natalizio BJ, Wente SR. 2013. Postage for the messenger: designating routes for nuclear mRNA export. *Trends Cell Biol* **23:** 365–373.

Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, Rubin E, Ophir R, Fluhr R. 2004. Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. *Plant J* **39:** 877–885.

Nikaido M, Noguchi H, Nishihara H, Toyoda A, Suzuki Y, Kajitani R, Suzuki H, Okuno M, Aibara M, Ngatunga BP, et al. 2013. Coelacanth genomes reveal signatures for evolutionary transition from water to land. *Genome Res* **23:** 1740–1748.

Ogert RA, Beemon KL. 1995. Identification of a constitutive transport element in the 3′ untranslated region of the Rous Sarcoma virus RNA. In *Retroviruses* (ed. Fan H, Rein A), p. 178. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40:** 1413–1415.

Pasquinelli AE, Ernst RK, Lund E, Grimm C, Zapp ML, Rekosh D, Hammarskjöld ML, Dahlberg JE. 1997. The constitutive transport element (CTE) of Mason–Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway. *EMBO J* **16:** 7500–7510.

Remy E, Cabrito TR, Batista RA, Hussein MA, Teixeira MC, Athanasiadis A, Sa-Correia I, Duque P. 2014. Intron retention in the 5′UTR of the novel ZIF2 transporter enhances translation to promote zinc tolerance in *Arabidopsis*. *PLoS Genet* **10:** e1004375.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet* **16:** 276–277.

Sakuma T, Davila JI, Malcolm JA, Kocher JP, Tonne JM, Ikeda Y. 2014. Murine leukemia virus uses NXF1 for nuclear export of spliced and unspliced viral transcripts. *J Virol* **88:** 4069–4082.

Siddiqui N, Borden KL. 2012. mRNA export and cancer. *Wiley Interdiscip Rev RNA* **3:** 13–25.

Smith AJ, Srinivasakumar N, Hammarskjöld ML, Rekosh D. 1993. Requirements for incorporation of Pr160gag-pol from human immunodeficiency virus type 1 into virus-like particles. *J Virol* **67:** 2266–2275.

Srinivasakumar N, Chazal N, Helga-Maria MC, Prasad S, Hammarskjöld ML, Rekosh D. 1997. The effect of viral regulatory protein expression on gene delivery by human immunodeficiency virus type 1 vectors produced in stable packaging cell lines. *J Virol* **71:** 5841–5848.

Swartz JE, Bor YC, Misawa Y, Rekosh D, Hammarskjöld ML. 2007. The Shuttling SR protein 9G8 plays a role in translation of unspliced mRNA containing a constitutive transport element. *J Biol Chem* **282:** 19844–19853.

Vitulo N, Forcato C, Carpinelli EC, Telatin A, Campagna D, D'Angelo M, Zimbello R, Corso M, Vannozzi A, Bonghi C, et al. 2014. A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. *BMC Plant Biol* **14:** 99.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470–476.

Westerfield M. 2007. *The zebrafish book*, 5th ed. University of Oregon Press, Eugene.

Wong JJ, Ritchie W, Ebner OA, Selbach M, Wong JW, Huang Y, Gao D, Pinello N, Gonzalez M, Baidya K, et al. 2013. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* **154:** 583–595.

Yang J, Cullen BR. 1999. Structural and functional analysis of the avian leukemia virus constitutive transport element. *RNA* **5:** 1645–1655.

Yap K, Lim ZQ, Khandelia P, Friedman B, Makeyev EV. 2012. Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev* **26:** 1209–1223.

Zhang ZC, Satterly N, Fontoura BM, Chook YM. 2011. Evolutionary development of redundant nuclear localization signals in the mRNA export factor NXF1. *Mol Biol Cell* **22:** 4657–4668.

Zolotukhin AS, Valentin A, Pavlakis GN, Felber BK. 1994. Continuous propagation of RRE(-) and Rev(-)RRE(-) human immunodeficiency virus type 1 molecular clones containing a *cis*-acting element of simian retrovirus type 1 in human peripheral blood lymphocytes. *J Virol* **68:** 7944–7952.