Phylogenetic Distribution and Evolution of the Linked RNA-Binding and NOT1-Binding Domains in the Tristetraprolin Family of Tandem CCCH Zinc Finger Proteins

Perry J. Blackshear¹ and Lalith Perera²

In humans, the tristetraprolin or TTP family of CCCH tandem zinc finger (TZF) proteins comprises 3 members, encoded by the genes *ZFP36*, *ZFP36L1*, and *ZFP36L2*. These proteins have direct orthologues in essentially all vertebrates studied, with the exception of birds, which appear to lack a version of *ZFP36*. Additional family members are found in rodents, amphibians, and fish. In general, the encoded proteins contain 2 critical macromolecular interaction domains: the CCCH TZF domain, which is necessary for high-affinity binding to AU-rich elements in mRNA; and an extreme C-terminal domain that, in the case of TTP, interacts with NOT1, the scaffold of a large multi-protein complex that contains deadenylases. TTP and its related proteins act by first binding to AU-rich elements in mRNA, and then recruiting deadenylases to the mRNA, where they can processively remove the adenosine residues from the poly(A) tail. Highly conserved TZF domains have been found in unicellular eukaryotes such as yeasts, and these domains can bind AU-rich elements that resemble those bound by the mammalian proteins. However, certain fungi appear to lack proteins with intact TZF domains, and the TTP family proteins that are expressed in other fungi often lack the characteristic C-terminal NOT1 binding domain found in the mammalian proteins. For these reasons, we investigated the phylogenetic distribution of the relevant sequences in available databases. Both domains are present in family member proteins from most lineages of eukaryotes, suggesting their mutual presence in a common ancestor. However, the vertebrate type of NOT1-binding domain is missing in most fungi, and the TZF domain itself has disappeared or degenerated in recently evolved fungi. Nonetheless, both domains are present together in the proteins from several unicellular eukaryotes, including at least 1 fungus, and they seem to have remained together during the evolution of metazoans.

Introduction

POST-TRANSCRIPTIONAL REGULATION of mRNA stability is an important aspect of gene expression. Many recent advances in our understanding of this control locus have come from studies of mRNA-binding proteins that can promote or inhibit the decay of their “target” mRNAs. The tristetraprolin (TTP) family of mRNA-binding proteins represents 1 group of such *trans*-acting factors. As exemplified by TTP itself, these proteins bind to AU-rich elements in mRNAs with low nanomolar affinity, then appear to recruit deadenylases to increase poly(A) tail removal or deadenylation. Deadenylation is considered the initial and probably rate-limiting step in mRNA decay in eukaryotes. TTP activity can be regulated in many ways, including at the level of gene transcription, mRNA stability, protein phosphorylation, and nucleo-cytoplasmic shuttling [recently re-

viewed in Ross and others (2012), Sanduja and others (2012), Brooks and Blackshear (2013), Ciaia and others (2013)].

It has been known for some time that TTP can promote mRNA deadenylation and decay. The effects on mRNA decay were first worked out as the result of experiments with TTP knockout mice and cells derived from them, which identified the tumor necrosis factor (TNF) alpha mRNA as the first physiological TTP target transcript (Taylor and others 1996; Carballo and others 1997, 1998). These studies determined that the highly conserved tandem zinc finger (TZF) domain of TTP bound directly to the AU-rich regions of target mRNAs. However, it was the evaluation of a second target transcript, encoding granulocyte-macrophage colony stimulating factor (CSF2), that permitted the conclusion that one of TTP's physiological activities was to promote poly(A) tail removal (Carballo and others 2000).

Laboratories of ¹Signal Transduction and ²Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina.

This stimulated a search for associated deadenylases, a search that is ongoing. Studies from our laboratory indicated that TTP could “effectively” activate poly(A) ribonuclease, or PARN, but we were unable to demonstrate direct interactions of the 2 proteins (Lai and others 2003). Other laboratories described evidence of direct association with other deadenylases, particularly involving the CCR4–NOT1 complex (Sandler and others 2011). These authors demonstrated that the C-terminus of TTP interacted directly with the NOT1 protein, which forms the central scaffold for the large CCR4–NOT1 complex of proteins. They concluded that “Not1 is required for TTP-mediated mRNA deadenylation and decay” (Sandler and others 2011). Very recently, Fabian and others (2013) demonstrated that a small, conserved sequence motif at the extreme C-terminus of human TTP could bind directly to an interior sequence of the NOT1 protein, and solved a crystal structure of this interacting complex. In cell transfection studies, removal of this small domain from TTP severely impeded its ability to stimulate mRNA decay, but did not abrogate it entirely, suggesting that other interactions may still be involved. Figure 1 illustrates several aspects of this proposed interaction. In Fig. 1A, modified from Collart and Panasenko (2012), the central position of NOT1 as the “scaffold” of a large complex is shown, along with the known protein interactors, at least 2 of which are deadenylases that are capable of acting on the poly(A) tail. Figure 1B and C, taken from Fabian and others (2013), show the sites of interaction between human TTP and NOT1 (Fig. 1B) and the proposed organization of the TTP–NOT1 complex interacting with a target AU-rich element in an mRNA (Fig. 1C).

It has been known for many years that TZF domain-containing members of the TTP family of proteins are widely distributed, and can be found in many fungi as well

as in insects and vertebrates. It was also known that there was a highly conserved sequence at the extreme C-terminus in many of these species. Its identification as a NOT1-binding domain in human TTP has led us to ask whether both domains have been present in tandem in the same protein throughout the evolution of eukaryotes. The present overview is an attempt to survey the phylogenetic distribution of the 2 linked binding activities. We do not intend to review the literature of these interactions, but rather, the focus will be on publicly available sequence data from a broad spectrum of organisms. The data presented here indicate that the 2 binding activities are linked in many lineages of eukaryotes, suggesting that they have evolved together from a common ancestor dating back to more than a billion years. However, fungi in particular may have dispensed with one or both activities during their dramatic evolutionary expansion.

Viruses and Prokaryotes

The TZF domain being discussed here is extremely stereotyped in sequence. There are 2 similar CCCH zinc fingers, with the internal spacing C-X8-C-X5-C-X3-H (where X can be a variety of amino acids). These are separated by 18 amino acids. There is a short lead-in sequence to each finger, which also very similar in the 2 fingers, that contains aromatic residues that are critical for stacking interactions with the RNA bases (Hudson and others 2004).

When this domain from human TTP (GenBank accession number NP_003398.2, amino acids 109–172) was used to search bacterial and archaeal sequences in GenBank, there were no apparently similar domains in these prokaryotes. However, we identified a single virus, lymphocystis disease virus 1, that expresses a protein containing 4 potential

FIG. 1. Interaction of human tristetraprolin (TTP) with NOT1. In (A) is shown a schematic representation of NOT1 as a scaffolding protein for a large, multiprotein complex, containing at least 2 deadenylases. This was modified from Collart and Panasenko (2012), with permission. (B) Shows the organization of the NOT1-binding domain in human TTP, and the TTP-binding domain of human NOT1, as identified by the indicated deletions and truncations of the respective fusion proteins. (C) Shows the proposed organization of the TTP mRNA binding and deadenylating activities, with TTP binding to the AU-rich region of mRNAs through its tandem zinc finger (TZF) domain, and binding to NOT1 through its C-terminal domain. NOT1 then brings into play its attached deadenylases, promoting deadenylation and accelerated destruction of the mRNA. (B, C) are taken from Fabian and others (2013), with permission.

CCCH zinc fingers in 2 tandem pairs. Two different isolates of this virus contained DNA sequences encoding similar proteins with 4 aligned potential CCCH zinc fingers (GenBank accession numbers NP_078696.1 and YP_073645.1). The 2 pairs of fingers are somewhat analogous to the prototype TTP TZF domain. However, although the 4 potential zinc fingers contain the requisite CCCH residues, they differ from the conventional sequence in their intra- and inter-finger spacing. An interaction with an AU-rich mRNA sequence is possible, but no experimental data are available to support this possibility, and the virus appears to be difficult to work with. It is intriguing that this virus also expresses at least 1 mRNA, encoding an apparent protease (NP_078647.1), that contains a TTP-like target sequence in the presumed 3'UTR within about 10bp of the stop codon: ATATTTATATTTATATTTATATT.

To our knowledge, this is the only evidence to date that a TZF-domain-like sequence is present in viruses. It is curious that only a single virus has been shown to contain sequences encoding a protein of this type, given all the viral genomic sequences that have been completed to date. One intriguing possibility is that the virus acquired this sequence horizontally from elsewhere, for example, its preferred environment, the skin of fish. It is probably coincidental that fish and amphibians express a fourth TTP family member, not yet found in other vertebrates, that also contains 4 CCCH zinc fingers (see next). Unfortunately for this theory, the 4 viral zinc fingers do not align very well with the fish sequences, but, of course, they could have evolved from the initial hypothetical horizontal gene transfer.

Eukaryotes

Vertebrates and their relatives

TZF domain-containing proteins have been identified in many eukaryotes. We recently described the high level of conservation among the 3 human family members and their direct orthologues in birds, reptiles, amphibians, and other mammals (Lai and others 2013). All 3 are present in all of these lineages, with the apparent exception of birds, which appear to lack TTP. In addition, rodents express a fourth family member, ZFP36L3, which is not present in other mammals (Blackshear and others 2005; Frederick and others 2008), and *Xenopus laevis* expresses a different fourth family member, C3H-4 (De and others 1999). The situation in fish is more complex, as several relatives of the 3 canonical proteins appear to be present, as well as a fish version of the *Xenopus* protein C3H-4, resulting in a total of 7 distinct proteins in *Danio rerio* (AAI39895.1, NP_571014.1, NP_001025418.1, NP_001070621.1, NP_955943.1, NP_996938.1, and XP_002665468.1). The C3H-4 protein from amphibians and fish is discussed in greater detail next. One of the interesting aspects of this analysis was the extreme sequence conservation of not only the TZF domain but also potential C-terminal NOT1-binding domains in all of these vertebrates. These 2 binding activities, therefore, appear to have been linked in a common ancestor of amphibians and humans. In Fig. 2A, we have shown the 4 mouse TTP family members as proxies for the orthologous proteins from all vertebrates, that is, from the subphylum Vertebrata of the phylum Chordata, with the exception of the C3H-4 proteins of *Xenopus* and fish, as discussed next.

Figure 2A shows the sequences of the TZF domains and their linked C-termini, containing the putative NOT1 binding sites, from representatives of major eukaryotic groups, in roughly descending order of evolutionary complexity. Figure 2B shows the approximate evolutionary timeline for the divergence of these major groups as seen from the perspective of the fungi, as described more fully next.

We then searched for these linked domains in cartilaginous and jawless fish, and in representatives of “pre-vertebrate” chordates, as well as in invertebrates, insects, amoebae, choanoflagellates, and plants. As shown in Fig. 2A, both the TZF domain and a putative NOT1-binding domain were conserved and present in 1 of the 2 TTP family members that we identified in the little skate, *Leucoraja erinacea*, a representative of sharks and rays. Among the jawless fish or Agnatha, the freshwater Arctic lamprey, *Lethenteron camtschaticum*, contains at least 3 sequences encoding separate TTP family members in its genome, 1 of which is also found in the sea lamprey *Petromyzon marinus*; the latter sequence is illustrated in Fig. 2A.

We searched the other 2 subphyla of Chordata, the Tunicata and Cephalochordata, and found single proteins in at least 1 species from each subphylum. In the case of tunicates, a TTP family member protein could be found in at least 2 species, and the 2 binding domains from *Ciona intestinalis* are shown in Fig. 2A. A similar, single protein sequence containing both domains was also found in the Florida lancelet or amphioxus, *Branchiostoma floridae*, a cephalochordate whose lineage was thought to have diverged from the vertebrate lineage about 500 million years ago (Fig. 2A).

Proteins with both domains were found in species from phyla related to Chordata, including the echinoderm phylum, represented by a single protein in the sea urchin, *Strongylocentrotus purpuratus* (Fig. 2A). Similarly, a single protein with both domains was found in a representative of the phylum Hemichordata, the acorn worm *Saccoglossus kowalevskii*. This sequence is not represented in Fig. 2A, but contains a typical TZF domain and a typical C-terminal potential NOT1-binding site, RLPIFSRLSIDE-. As in the case of the lancelet protein sequence, the acorn worm sequence could be assembled entirely from expressed sequence tags (ESTs), documenting that these transcripts are expressed in these species; strikingly, there was only a single TZF domain-containing protein sequence in these EST collections.

These findings suggest that tunicates, cephalochordates, echinoderms, and acorn worms express single TTP family members, each of which has both TZF- and NOT1-binding domains (Fig. 2A). These proteins are most closely related in amino-acid sequence to the human protein ZFP36L1, which is probably the closest thing to an ancestral protein expressed in humans. As one approaches bony vertebrates, in the case of the lampreys, there are apparently 3 separate protein species, compatible with at least 1 of the 2 “waves” of gene duplication that have been thought to occur during vertebrate evolution (Beutler and Moresco 2008).

One aspect of this analysis is that, in all cases mentioned earlier, the TZF domain is in the “middle” of the proteins, and the putative NOT1-binding site is at the extreme C-terminus in all cases. It should be emphasized that we are predicting NOT1 binding based on sequence conservation alone; to our knowledge, none of these proteins has been demonstrated to interact directly with NOT1 with the exception of human TTP.

FIG. 2. Linked TZF domains and NOT1-binding domains of TTP family members in eukaryotes. In (A) are shown the TZF domains and their respective putative NOT1-binding domains of TTP family member proteins from various eukaryotic species in approximately descending order of complexity, starting from the 4 mouse proteins. In general, common names have been used in this figure, but specific species names can be found in the text in each case. Since this linked binding domain arrangement is much the same in all known vertebrate proteins (except for frog and fish C3H-4—see text), the mouse is the only bony vertebrate shown, with the next level that of the cartilaginous fishes, represented by the little skate. The TZF domains are shown separated by gaps consisting of various numbers of amino acids from the extreme C-terminal putative NOT1-binding domains from the same protein. The C-terminal stop codons are indicated by the dashes to the right of the protein sequence. Alignments of both domains were by ClustalW2, with its usual consensus conventions. Amino-acid display used Boxshade. In (B) is shown a tree showing the approximate time scale of evolutionary divergence of the major eukaryotic groups, from the perspective of the fungi. This was modified from Stajich and others (2009), with permission. All of the major groups at the top of the figure, the Plantae, Amoebozoa, Choanozoa, and Metazoa, contain proteins with linked TZF domains and C-terminal NOT1-binding domains. However, although most fungi contain TZF domains, as indicated, the only species that has been found to date to contain a protein with a linked TZF domain and typical NOT1-binding domains is the chytrid fungus, *Spizellomyces punctatus*, as indicated by the asterisk. The dashed line indicates the uncertainty about the position of the Microsporidia, as discussed in Stajich and others (2009). See the original reference for further details.

Insects and arachnids

Insects generally appear to express a single TTP family member per species, although these can exist in multiple isoforms, as in the case of the *Drosophila melanogaster* Tis11A, B, and C isoforms. Similar to the other metazoans described earlier, the insect species investigated have both TZF domains that are highly conserved with those of the mammalian proteins, as well as putative C-terminal NOT1-binding domains that are conserved with the mammalian sequences as well (Fig. 2A) (Fabian and others 2013). In recent experiments, we have demonstrated that removal of this putative NOT1-binding domain from the *Drosophila* TIS11 protein adversely affected its function to promote the

decay of a TNF-based AU-rich element (ARE) mRNA in transfected mammalian cells, raising the possibility that the fly protein could directly interact with mammalian NOT1 (W.S. Lai and P.J. Blackshear, Unpublished data). A simulation structure strongly supports this possibility, but we should emphasize that binding of the *Drosophila* C-terminal sequence to a *Drosophila* NOT1 sequence has not been demonstrated directly, to our knowledge.

Fewer sequences are available among arachnids, but an interesting example of a recently sequenced genome is in the case of the western predatory mite, *Metaseiulus occidentalis*. GenBank searches identified several TZF domain-containing protein sequences in this species, all with typical internal spacing, lead-in sequences, and so on. However,

only 2 of these proteins have typical, C-terminal putative NOT1 binding sites, of which one is shown in Fig. 2A. This is evidence that the joining of the 2 binding domains in a single protein was likely already present at the divergence of insects and arachnids from early Bilateria.

Other Ecdysozoa and Lophotrochozoa

The nematode *Caenorhabditis elegans* is a representative of another branch of the Ecdysozoa from arthropods, and its genome sequence has been known for some time. It contains a variety of proteins with TTP-like TZF domains, but many of them have atypical internal spacing compared with the sequences discussed here. However, 1 protein from *C. elegans* contains a typical TTP-like TZF domain, and that is CCCH-1 (NP_505926.2) (Fig. 2A). Isoforms a and b for this protein differ only at the extreme amino terminus. Strikingly, both isoforms also have an extreme C-terminal consensus NOT1-binding site (Fig. 2A), as noted by Fabian and others (2013), and it seems likely that this protein is a typical representative of the linked binding site concept that we are describing here. As a representative of the Lophotrochozoa, the Atlantic oyster (*Crassostrea virginica*) also has at least 1 protein with linked TZF- and NOT1-binding domains (Fig. 2A).

Plants

A comprehensive review of plant TZF domain-containing proteins is beyond the scope of this survey, but an examination of plant protein sequences in GenBank suggests that while there are proteins in many species that contain intact predicted TZF domains, most of these are not obviously linked to a predicted C-terminal NOT1 binding domain of the TTP type. Most of the proteins identified in a TZF domain-based search reach their C-terminal end shortly after the completion of the TZF domain, suggesting that, similar to most fungi (see next), they may have “lost” the typical C-terminal NOT1-binding domain presumed to be present in a common ancestor with other eukaryotes. Much further work will be necessary to test this idea.

However, we identified 3 plant species that contain genes encoding probable proteins with both TZF domains and C-terminal NOT1-binding domains. These species are the Christmas bush, *Chromolaena odorata* (translation of Unigene mRNA sequence GACH01022939.1), and 2 mosses, *Selaginella moellendorffii* (XP_002980683.1) and *Physcomitrella patens* (XP_001779415.1). Although it is possible that these genes emerged in plants as a consequence of horizontal gene transfer, the fact that 3 different species from at least 2 major plant groups express proteins with both linked binding activities suggests that these were present in a common ancestor of plants as well as of animals.

The relevant sequences from the Christmas bush, as a representative of the plants, are shown in Fig. 2A. We have modeled the NOT1-binding domain of the Christmas bush protein interacting with the relevant NOT1 domain in Fig. 3B, based on the original human complex structure [Fig. 3A (Fabian and others 2013)]. The human NOT1 TTP C-terminal peptide shown in Fig. 2A forms an amphipathic, 2 turn alpha helix that contains 4 hydrophobic residues which are inserted into the hydrophobic groove of NOT1 (Fig. 3A); the peptide is RLPIFNRISVSE, where the 4

hydrophobic residues are underlined. Other critical residues include the initial arginine and the first serine, which are involved in salt-bridge and hydrogen bonding interactions with the NOT1 protein. An examination of Fig. 2A shows that these key residues are highly conserved in the species shown, including *Chromolaena odorata*. In keeping with this sequence conservation, the solution structure model for the plant peptide shows a similar 2 turn alpha helix with similar hydrophobic and other interactions (Fig. 3B).

Single-celled eukaryotes

Amoebae and choanoflagellates. From the earlier discussion, it appears that many, if not most, multicellular animals, and at least some plants, contain genes that express proteins containing the linked pair of binding activities, in every case described so far with the putative NOT1 binding site at the extreme C-termini of the proteins, and the TZF domain further toward the amino termini. Since, with the single exception of the lymphocystis disease virus, there have been no identified TZF domain-containing proteins in prokaryotes or viruses, we searched single-celled eukaryotes for clues as to the evolutionary origins of the 2 linked domains.

Relatively few complete genomes are available for non-fungal single-celled organisms, but representatives of the major groups shown in Fig. 2B have been sequenced as a part of a quest to understand the origins of Metazoa. From the Amoebozoa, the freshwater amoeba *Acanthamoeba castellanii* expresses a single TZF domain-containing protein that also contains a potential C-terminal NOT1-binding site (Fig. 2A) (conceptual translation of AHJ101001149.1, reverse complement of 26270–24672). A predicted structure of the putative NOT1-binding domain in complex with NOT1 from this organism is shown in Fig. 3C. A similar pair of sequences was found in another single cell organism from the Choanozoa, representing another major branch point during metazoan evolution (Fig. 2B). The potential binding sequences of 1 such organism, the choanoflagellate *Monosiga brevicollis* (XP_001746753.1), are shown in Fig. 2A, and the proposed structure of the TTP family member binding to NOT1 from this species is shown in Fig. 3D. Both the TZF RNA-binding domain and the C-terminal NOT1-binding domains coexist in a single protein in these single-celled eukaryotes, suggesting that their linkage was present in a common ancestor very early in metazoan evolution.

Similarly, *Dictostelium discoideum* contains a protein with both activities, as shown in Fig. 2A and as modeled in Fig. 3F.

Others. We also searched sequences from a number of organisms that have been considered among the most primitive eukaryotes. For example, *Toxoplasma gondii* is an important human intracellular parasitic pathogen. It expresses at least 3 proteins that contain TZF domains (GenBank accession numbers EPT30178, EPR61632, and EEE30927), all of which also contain potential NOT1-binding sites in their C-termini. The TZF domains of these proteins are slightly different in internal spacing from the vertebrate proteins, and it remains to be seen whether their affinity and specificity for RNA targets are affected by these spacing changes. *Giardia lamblia* or *G. intestinalis* is another important human pathogen that has been described as a possible “basal” eukaryote, or at least a very early diverging one (discussed in Morrison and others 2007). This

FIG. 3. Models of TTP family member putative NOT1-binding sites associated with the NOT1 protein of the same species. These solution structure models are based on the original coordinates of Fabian and others (2013); a view of the human complex discussed in that paper is shown in (A). The other models are based on the sequences of the predicted NOT1 protein orthologues in those species, and the predicted TTP family member protein C-terminal domains, discussed in the text and in Fig. 2A (Fabian and others 2013). The initial structures were obtained by homology modeling with appropriate mutations on the X-ray crystal structure of the C-terminal segment of human TTP bound to human NOT1. These initial models were then solvated in water, followed by a series of equilibration trajectories, and were finally subjected to lengthy molecular dynamics simulations over 30 ns using standard molecular dynamics protocols at constant temperature and constant volume. The hydrophobic TTP family protein residues that are in contact with NOT1 residues are shown in yellow, and the polar and charged TTP family protein residues which are in contact with NOT1 residues are in red. Residue numbers were omitted for clarity. (A) *Homo sapiens*; (B) *Chromolaena odorata* (Christmas bush); (C) *Acanthamoeba castellanii*; (D) *Monosiga brevicollis*; (E) *Spizellomyces punctatus*; (F) *Dictyostelium discoideum*.

organism has at least 4 closely related sequences that encode TTP-like TZF domains, with normal internal spacing (eg, conceptual translation of GenBank accession number ACVC01000123.1, 58686–59264). However, these putative proteins do not include a typical NOT1-binding motif of the type discussed here, although there are several conserved C-terminal domains that could fall into a looser consensus.

Fungi

Fungi have a complex pattern of expression of TZF domain-containing proteins. To our knowledge, the first identifications of fungal proteins of this type were the initial discoveries of CTH1 and CTH2 in *Saccharomyces cerevisiae* (Ma and Herschman 1995; Thompson and others 1996). These proteins were later shown by the Thiele and Puig groups to be important regulators of a group of genes in-

involved in iron metabolism in this species, by regulating post-transcriptional mRNA stability (Puig and others 2005; Martinez-Pastor and others 2013). More recently, we have studied the single TZF domain-containing protein in *S. pombe*, known as Zfs1p, and have described its importance in preventing abnormally increased cell-cell interactions, leading to flocculation (Cuthbertson and others 2008; Wells and others 2012). In this species, Zfs1p promotes mRNA decay of AU-rich element-containing mRNAs, and strains deleted in *zfs1* accumulate abnormal levels of these target transcripts. However, the TTP family members expressed in these species do not contain obvious homologous sequences to the characteristic NOT1-binding domains described earlier. In fact, the protein sequences in many yeasts and fungi come to an end very soon after the C-terminal end of the TZF domain. This is exemplified in *Meyerozyma guilliermondii*, whose TTP family protein is only 156 amino

acids long (conceptual translation of AAFM01000068.1, 20179–20646), and the C-terminus of the protein contains the TZF domain plus only a single additional residue before the stop codon. The *S. pombe* protein, Zfs1p, only has 15 additional amino acids after the TZF domain and before the stop codon (NP_596453.1).

It also appears that many species of fungi have “lost” the expression of an intact TTP family member altogether. Figure 2B shows the relationship between the major fungal divisions and the presence of TTP family members. As best we can tell, in all the major fungal divisions, from the most primitive Microsporidia and Chytridiomycota, through the more advanced Saccharomycotina and Schizosaccharomycetes, sequences corresponding to intact TZF domains can be readily found in most, if not all, sequenced species. However, of the 137 genomes currently represented in GenBank from the most recently evolved fungi, the Pezizomycotina, there was not a single instance of a predicted protein with an intact TZF domain. Some organisms had TZF domain-like regions that were degenerate in terms of internal spacing or contained other variations. For example, a sequence in *Tuber melanosporum* was roughly similar to a canonical TZF domain, but was missing a single amino-acid residue in the C-terminal C-x8-C interval, and the final H was modified to an R (conceptual translation of CABJ01000136.1, 35694–36077). We cannot exclude cryptic introns, sequencing errors, or other factors in this comparison. Other fungi in this group contained similar sequences with additional “missing” amino acids compared with the consensus. Although structure-function studies of this domain from human TTP have suggested that some amino-acid deletions can occur with no loss of affinity to typical ARE binding sites, it is not clear what effect the major changes seen in these fungi would have on binding site affinity or specificity. It seems likely that these sequences ultimately derived from intact TZF domains of the TTP type, but they have evolved into degenerate forms of unclear function.

As noted earlier, the other 2 major divisions within Ascomycota are full of predicted proteins with intact TZF domains. Within the Saccharomycotina, essentially all of the species surveyed contained at least 1 TZF domain-containing predicted protein, including the 2 mentioned earlier in *S. cerevisiae*. Similarly, in the Schizosaccharomycetes, all members of the genus checked so far contain the TTP family member Zfs1p-like proteins. Within the Basidiomycota, including the 57 genomes currently in GenBank, there were several obviously orthologous sequences. However, the evaluation of many of the others is complicated by the apparent presence of one or more introns within sequences encoding the TZF domains. This has been definitively sorted out in the case of *Pneumocystis murina*, in which both genomic and mRNA sequences have been determined, and a classical TZF domain sequence is present in the protein translated from the mRNA sequence (see locus PNEG_01914.1 in http://broadinstitute.org/annotation/genome/Pneumocystis_group/MultiHome.html).

We also searched the fungal genomes available in GenBank and elsewhere for predicted protein sequences that contained both TTP-like TZF domains and C-terminal NOT1-binding sequences. In only a single species, we found such a pair of sequences, in the chytrid fungus, *Spizellomyces punctatus* (Fig. 2A) (conceptual translation of

ACOE01000143.1, reverse complement of 45550–43592). The predicted sequence of the NOT1-binding domain from this protein (Fig. 2A), complexed with the relevant NOT1 domain from this species, are modeled in Fig. 3E. The chytrid fungi are generally considered among the most primitive fungal forms (Fig. 2B), and they branched off from a common ancestor near the beginning of fungal evolution. Although sequences from other *Spizellomyces* species are not available, the sequence of another chytrid fungus, *Batrachochytrium dendrobatidis*, is available. The TZF domain-containing protein from this species (from a conceptual translation of ADAR01000269.1, reverse complement of 146325–144061) does not have an obvious predicted C-terminal NOT1-binding site, with the open reading frame terminating before reaching the putative NOT1-binding site in *S. punctatus*. However, the TZF domain sequence from *B. dendrobatidis* is quite similar to that of *S. punctatus*.

One possibility is that *S. punctatus* represents an ancestral form that reflects the early organization of a TTP family member which contains both a TZF domain RNA-binding sequence and a C-terminal NOT1-binding site, and that this then disappeared during the evolution of more advanced fungi. This organism is famous for having an unusual form of RNA editing that is shared by a completely different single-celled eukaryote, the amoeba *A. castellanii*. Perhaps coincidentally, this organism also has a TTP family member with both a TZF domain and a predicted NOT1-binding site (Fig. 2A; also earlier). As in the case of RNA editing, the possibility exists that *S. punctatus* acquired the bi-functional TTP family member through some form of horizontal gene transfer, although a search of non-fungal eukaryotes with the *S. punctatus* protein sequence does not reveal an obvious source. It will be very interesting to determine whether other *Spizellomyces* species contain similar 2-domain proteins. Our current knowledge does not allow us to distinguish between these 2 possibilities. By whatever mechanism the 2 binding functions arose in *S. punctatus*, it seems obvious that the typical C-terminal NOT1-binding domain has been lost from most other fungal species.

These considerations lead to a couple of predictions that should be experimentally testable. One is that the TZF domain-containing proteins of fungi and yeast could interact with NOT1 through a sequence which is atypical compared with the sequences shown in Fig. 2A. A second possibility is that these proteins bind to a separate NOT1 binding site-containing protein to reassemble the equivalent of an intact TTP family member as a 2 protein complex. Third, the fungal proteins could have evolved another mechanism for associating with cellular deadenylase activities, one that may be separate from the mammalian-style NOT1 interactions. This putative mechanism could also be shared in mammalian cells, or may be unique.

A final possibility is that the proteins without obvious NOT1-binding domains actually do not promote deadenylation and mRNA decay. This remains a possibility, but the examples of *S. cerevisiae* and *S. pombe* prove that at least some family member proteins without obvious C-terminal NOT1-binding sites can function to promote mRNA decay in their respective species. We hope that protein-binding partner studies in various fungal species will soon answer some of these questions.

FIG. 4. Alignment of C3H-4 sequences from 2 species of frogs and 5 species of fish. Protein alignments from the apparent C3H-4 orthologues from the listed species were aligned with ClustalW2, using their defaults and labeling conventions. The GenBank accession numbers from these species were as follows: NP_571014.1 (*Danio rerio*); CAA71245.2 (*Cyprinus carpio*); XP_003458454.1 (*Oreochromis niloticus*); XP_004574367.1 (*Maylandia zebra*); XP_003966860.1 (*Takifugu rubripes*); NP_001108269.2 (*Xenopus laevis*); and NP_001039082.1 (*Silurana tropicalis*). The solid underline indicates the position of the TZF domain; the dotted over- and underlines indicate the third and fourth zinc fingers, which do not align in all species by this method. The position of the highly conserved C-terminal domain is also indicated, which we speculate may be an unusual NOT1-binding domain in this protein. See the text for further details.

The Special Case of the *Xenopus* C3H-4 Protein

The earlier discussion suggests that TTP-like proteins in most, if not all, eukaryotes evolved from an ancestral protein which contained both RNA- and NOT1-binding sites. Given the early separation of plants and animals illustrated in Fig. 2B, this putative ancestral bifunctional protein may have been present in an ancestral organism as long ago as 1.5 billion years. During subsequent evolution, as in the cases of some proteins from the predatory mite, and from most plants and fungi, related proteins are still expressed that contain the TZF domain but have “lost” the typical C-terminal NOT1-binding domain.

For this analysis, we have relied on a consensus sequence for the NOT1-binding site from human TTP, exemplified in Fig. 2A. However, other sequences could well interact with the same region of NOT1, or even different regions. An

interesting situation exists in the case of the *Xenopus* protein C3H-4 and its apparent orthologues in *S. tropicalis* and various fish. As mentioned earlier, amphibians and fish express orthologues of the 3 common vertebrate TTP proteins, but the C3H-4 protein from *Xenopus* does not have a direct mammalian orthologue. This protein contains a typical TZF domain as well as 2 somewhat degenerate zinc fingers (Fig. 4). A similar protein is expressed in fish (Fig. 4). An alignment of the orthologous proteins from 2 frog species and 5 fish species shows that there are no obvious NOT1-binding sequences at the extreme C-terminus, in contrast to the other TTP family members from these species. In *Xenopus*, we initially showed that this protein was limited to maturing oocytes, eggs, and very early embryos (De and others 1999). More recently, it was demonstrated to accumulate in the first meiotic metaphase; when it was ablated, meiotic arrest ensued (Belloc and Mendez 2008). These

FIG. 5. Model of the hypothetical NOT1-binding site from *D. rerio* C3H-4 associated with NOT1 from the same species. The sequence of the postulated atypical NOT1-binding site from *D. rerio* is indicated in Fig. 4. The NOT1 sequence shown is from NP_001073420.1, amino acids 817–990. The solution simulation structure was constructed by the methods described in the legend to Fig. 3. The hydrophobic (in yellow) and polar (in red) TTP residues that are in contact with NOT1 residues (in gray) are shown. The residue numbers are from the *D. rerio* C3H-4 protein, NP_571014, as follows: L²⁹¹LL²⁹³PL²⁹⁵A²⁹⁶L²⁹⁷RL²⁹⁹Q³⁰⁰.

authors demonstrated a physical interaction between this protein and the CCR4 deadenylase in cell extracts, but did not look for direct or indirect binding to NOT1. The techniques used in that paper, as best we can tell, could not distinguish between the direct binding of C3H-4 to CCR4, on the one hand, and the binding of C3H-4 to CCR4 via NOT1, on the other hand. In either case, the authors suggested that C3H-4 could promote mRNA deadenylation during exit from metaphase, thus permitting meiotic progression.

At first glance, this concept does not jibe with our failure here to find a typical NOT1-binding sequence at the extreme C-terminus. However, an examination of the alignment of the orthologous protein sequences from the available frog and fish species demonstrates a highly conserved domain fairly near the C-terminus of all of these proteins (Fig. 4). This is quite different from the typical TTP C-terminal type of NOT1 binding site shown in Fig. 2A, but it, nonetheless, contains amino-acid residues that can be modeled to form a NOT1-binding site, based on the original human TTP-NOT1 crystal structure (Fabian and others 2013) (Fig. 5). A direct interaction between this domain and frog or fish NOT1 remains to be demonstrated, but such a demonstration could lead to a broadening of the consensus sequence requirements for this domain in other organisms.

Conclusions

In this brief discussion, we have surveyed the “tree of life” for the existence of 2 key linked domains that characterize the vertebrate TTP family of proteins: the RNA-binding TZF domain and the C-terminal NOT1-binding domain. With the exception of a single virus, the lympho-

cystis disease virus 1 of fish, which contains 4 apparent but degenerate CCCH zinc fingers, TTP family proteins appear to be confined to eukaryotes. Within eukaryotes, in general, both the TZF domain and the predicted NOT1-binding domains are present together in many single-celled organisms of very different lineages, suggesting that the coexistence of these 2 binding domains is an ancient phenomenon which dates to a common ancestor more than a billion years ago. Since that time, many things have happened to this presumed ancestral protein, including apparent loss of the C-terminal NOT1-binding sequence in many plants and most fungi; the degeneration or loss of the TZF domain itself in modern fungi of the Pezizomycotina; and more modern events, such as the development of the multiple related proteins in vertebrates, especially fish, and the apparent loss of TTP in modern birds. Within this framework, the wealth of currently available and rapidly accumulating sequence information about many of these disparate organisms should allow for various types of structure–function studies. Much slower will be the elucidation of the physiological roles of these proteins in their host organisms, although data are gradually accumulating on their key roles in model organisms such as *Mus musculus*, *D. melanogaster*, *X. laevis*, *S. cerevisiae*, and *S. pombe*. It will be fascinating to determine whether organisms such as some fungi and plants which appear to have “lost” the ancestral C-terminal NOT1-binding domain have reconstituted that domain by binding to a second protein; have substituted another NOT1-binding sequence; have used another unrelated deadenylase-linked binding sequence; or, in some cases, have lost TTP-like activity altogether.

Acknowledgments

The authors are grateful to the members of their laboratories for helpful discussions, and to Guang Hu and Dori Germolec for useful comments on this article. They thank Marc Fabian for the coordinates of the original TTP-NOT1 crystal structure, Sebastian Shimeld for helpful discussions, and Melissa Wells for insights on *Pneumocystis*. This study was supported by the Intramural Research Program of the NIEHS, NIH.

Author Disclosure Statement

No competing financial interests exist.

References

- Belloc E, Mendez R. 2008. A deadenylation negative feedback mechanism governs meiotic metaphase arrest. *Nature* 452:1017–1021.
- Beutler B, Moresco EM. 2008. The forward genetic dissection of afferent innate immunity. *Curr Top Microbiol Immunol* 321:3–26.
- Blackshear PJ, Phillips RS, Ghosh S, Ramos SB, Richfield EK, Lai WS. 2005. Zfp3613, a rodent X chromosome gene encoding a placenta-specific member of the Tristetraprolin family of CCCH tandem zinc finger proteins. *Biol Reprod* 73:297–307.
- Brooks SA, Blackshear PJ. 2013. Tristetraprolin (TTP): Interactions with mRNA and proteins, and current thoughts on mechanisms of action. *Biochim Biophys Acta* 1829: 666–679.

- Carballo E, Gilkeson GS, Blackshear PJ. 1997. Bone marrow transplantation reproduces the tristetraprolin-deficiency syndrome in recombination activating gene-2 (-/-) mice. Evidence that monocyte/macrophage progenitors may be responsible for TNF α overproduction. *J Clin Invest* 100: 986–995.
- Carballo E, Lai WS, Blackshear PJ. 1998. Feedback inhibition of macrophage tumor necrosis factor- α production by tristetraprolin. *Science* 281:1001–1005.
- Carballo E, Lai WS, Blackshear PJ. 2000. Evidence that tristetraprolin is a physiological regulator of granulocyte-macrophage colony-stimulating factor messenger RNA deadenylation and stability. *Blood* 95:1891–1899.
- Ciais D, Cherradi N, Feige JJ. 2013. Multiple functions of tristetraprolin/TIS11 RNA-binding proteins in the regulation of mRNA biogenesis and degradation. *Cell Mol Life Sci* 70:2031–2044.
- Collart MA, Panasenko OO. 2012. The Ccr4—not complex. *Gene* 492:42–53.
- Cuthbertson BJ, Liao Y, Birnbaumer L, Blackshear PJ. 2008. Characterization of zfs1 as an mRNA-binding and -destabilizing protein in *Schizosaccharomyces pombe*. *J Biol Chem* 283:2586–2594.
- De J, Lai WS, Thorn JM, Goldsworthy SM, Liu X, Blackwell TK, Blackshear PJ. 1999. Identification of four CCCH zinc finger proteins in *Xenopus*, including a novel vertebrate protein with four zinc fingers and severely restricted expression. *Gene* 228:133–145.
- Fabian MR, Frank F, Rouya C, Siddiqui N, Lai WS, Karetnikov A, Blackshear PJ, Nagar B, Sonenberg N. 2013. Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. *Nat Struct Mol Biol* 20: 735–739.
- Frederick ED, Ramos SB, Blackshear PJ. 2008. A unique C-terminal repeat domain maintains the cytosolic localization of the placenta-specific tristetraprolin family member ZFP36L3. *J Biol Chem* 283:14792–14800.
- Hudson BP, Martinez-Yamout MA, Dyson HJ, Wright PE. 2004. Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat Struct Mol Biol* 11:257–264.
- Lai WS, Kennington EA, Blackshear PJ. 2003. Tristetraprolin and its family members can promote the cell-free deadenylation of AU-rich element-containing mRNAs by poly(A) ribonuclease. *Mol Cell Biol* 23:3798–3812.
- Lai WS, Stumpo DJ, Kennington EA, Burkholder AB, Ward JM, Fargo DL, Blackshear PJ. 2013. Life without TTP: Apparent absence of an important anti-inflammatory protein in birds. *Am J Physiol Regul Integr Comp Physiol* 305: R689–R700.
- Ma Q, Herschman HR. 1995. The yeast homologue YTIS11, of the mammalian TIS11 gene family is a non-essential, glucose repressible gene. *Oncogene* 10:487–494.
- Martinez-Pastor M, Vergara SV, Puig S, Thiele DJ. 2013. Negative feedback regulation of the yeast CTH1 and CTH2 mRNA binding proteins is required for adaptation to iron deficiency and iron supplementation. *Mol Cell Biol* 33:2178–2187.
- Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, Davids BJ, Dawson SC, Elmendorf HG, Hehl AB, Holder ME, Huse SM, Kim UU, Lasek-Nesselquist E, Manning G, Nigam A, Nixon JE, Palm D, Passamaneck NE, Prabhu A, Reich CI, Reiner DS, Samuelson J, Svard SG, Sogin ML. 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* 317:1921–1926.
- Puig S, Askeland E, Thiele DJ. 2005. Coordinated remodeling of cellular metabolism during iron deficiency through targeted mRNA degradation. *Cell* 120:99–110.
- Ross CR, Brennan-Laun SE, Wilson GM. 2012. Tristetraprolin: roles in cancer and senescence. *Ageing Res Rev* 11:473–484.
- Sandler H, Kreth J, Timmers HT, Stoecklin G. 2011. Not1 mediates recruitment of the deadenylase Caf1 to mRNAs targeted for degradation by tristetraprolin. *Nucleic Acids Res* 39:4373–4386.
- Sanduja S, Blanco FF, Young LE, Kaza V, Dixon DA. 2012. The role of tristetraprolin in cancer and inflammation. *Front Biosci (Landmark Ed)* 17:174–188.
- Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, Taylor JW. 2009. The fungi. *Curr Biol* 19: R840–R845.
- Taylor GA, Carballo E, Lee DM, Lai WS, Thompson MJ, Patel DD, Schenkman DI, Gilkeson GS, Broxmeyer HE, Haynes BF, Blackshear PJ. 1996. A pathogenetic role for TNF α in the syndrome of cachexia, arthritis, and autoimmunity resulting from tristetraprolin (TTP) deficiency. *Immunity* 4: 445–454.
- Thompson MJ, Lai WS, Taylor GA, Blackshear PJ. 1996. Cloning and characterization of two yeast genes encoding members of the CCCH class of zinc finger proteins: zinc finger-mediated impairment of cell growth. *Gene* 174:225–233.
- Wells ML, Huang W, Li L, Gerrish KE, Fargo DC, Ozsolak F, Blackshear PJ. 2012. Posttranscriptional regulation of cell-cell interaction protein-encoding transcripts by Zfs1p in *Schizosaccharomyces pombe*. *Mol Cell Biol* 32:4206–4214.

Address correspondence to:

Dr. Perry J. Blackshear

Laboratory of Signal Transduction, F1–13

National Institute of Environmental Health Sciences

111 Alexander Drive

Research Triangle Park, NC 27709

E-mail: black009@niehs.nih.gov

Received 19 December 2013/Accepted 20 December 2013