The YTH Domain Is a Novel RNA Binding Domain*

Zhaiyi Zhang†, Dominik Theler§, Katarzyna H. Kaminska∥∥, Michael Hiller‡‡, Pierre de la Grange¶¶, Rainer Pudimat***, Ilona Rafalska¶, Bettina Heinrich‡, Janusz M. Bujnicki***, Frédéric H.-T. Allain†*, and Stefan Stamm‡

From the Institute for Biochemistry, Universität Erlangen-Nuremberg, Fahrstrasse 17, 91054 Erlangen, Germany, the Institute of Computer Science, Bioinformatics Group, Albert-Ludwigs-University Freiburg, Georges-Koehler-Allee 106, 79110 Freiburg, Germany, the Institut Universitaire d’Hematologie, Saint-Louis Hospital, 75010 Paris, France, the Institut fur Molekularbiologie und Biophysik, ETH Zurich, HPK G18, CH-8093 Zurich, Switzerland, the Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Trojdena 4, PL-02-109 Warsaw, Poland, and the Department of Molecular and Cellular Biochemistry, University of Kentucky, College of Medicine, Lexington, Kentucky 40536.

The YTH (YT521-B homology) domain was identified by sequence comparison and is found in 174 different proteins expressed in eukaryotes. It is characterized by 14 invariant residues within an α-helix/β-sheet structure. Here we show that the YTH domain is a novel RNA binding domain that binds to a short, degenerated, single-stranded RNA sequence motif. The presence of the binding motif in alternative exons is necessary for YT521-B to directly influence splice site selection in vivo.

Array analyses demonstrate that YT521-B predominantly regulates vertebrate-specific exons. An NMR titration experiment identified the binding surface for single-stranded RNA on the YTH domain. Structural analyses indicate that the YTH domain is related to the pseudouridine synthase and archaeosine transglycosylase (PUA) domain. Our data show that the YTH domain conveys RNA binding ability to a new class of proteins that are found in all eukaryotic organisms.

The binding of proteins to RNA is a fundamental aspect of biology that interferes with most aspects of gene expression and cellular functions. The presence of various binding motifs defines the group of RNA binding proteins (1). Commonly found RNA binding domains include the RNA recognition motif (RRM), the double-stranded RNA binding domain, the Piwi Argonaut and Zwille domain, and the heterogeneous ribonucleoprotein K homology domain. The most prominent RNA binding domain is the RRM that is found in ~2% of human proteins (2). The RRM is composed of two consensus sequences RNP2 and RNP1 that contain aromatic residues important for RNA binding. In other RNA binding motifs, such as the PUA (pseudouridine synthase and archaeosine transglycosylase) and OB-fold (oligonucleotide/oligosaccaride binding fold), the RNA interacts with the β-sheets that form pseudobarns (3). The general composition of the PUA domain is reminiscent of the OB-fold, a nucleic acid binding motif that displays only a low degree of sequence similarity between its members. The OB-fold consists of two three-stranded antiparallel β-sheets, where strand 1 is shared by both sheets. The individual β-sheets can be separated by protein parts of different length, which makes the identification based on primary structure difficult (4). The β-sheets in the PUA and OB-folds form a ligand binding surface that can bind to nucleic acids through aromatic stacking, hydrogen bonding, as well as polar and hydrophobic interactions. The so-far unexplained RNA binding activities of proteins such as apoptic (5) demonstrate that not all RNA binding domains have been described.

One of the potentially new RNA binding domains is the YTH (YT521 homology) domain. The YTH domain is highly conserved during evolution and was identified by comparing all known protein sequences with the splicing factor YT521-B (6). The domain is found only in eukaryotes and is abundant in plants. The YTH domain can be between 100 and 150 amino acids in size and is characterized by 14 invariant and 19 highly conserved residues. It is predicted to contain four α helices and six β strands. The conservation of aromatic residues within the β stands of the YTH domain is reminiscent of the RRM, PUA, and OB-fold structures. The presence of the YTH domain defines a new class of proteins from which currently 174 members have been identified in various eukaryotic species. YT521-B is the founding member of the YTH-containing proteins. It is a nuclear protein that interacts with other proteins implicated in RNA metabolism, such as Sam68/p62, rSLM-1, rSLM-2, hnRNPG, rSAF-B, and emerin (7–9). The protein is localized in a novel nuclear compartment, the YT bodies, that overlaps with transcriptional start sites (10). The protein regulates alternative splice site selection in a concentration-dependent manner and is only present in vertebrate species (11). The other molecularly characterized YTH-family protein is the
nuclear yeast RNA binding protein Mmi1 (meiotic mRNA interception) that is part of a yeast mRNA-destruction system that eliminates meiotic-specific mRNAs (12, 13). The biological function of the remaining YTH-family members that share no common sequence motifs outside the YTH domain is unknown. A structure of the YTH domain has been solved by NMR and was deposited without published analysis in the protein data base by the Japanese RIKEN Structural Genomics/Proteomics Initiative (accession number 2yud). Here, we demonstrate that the YTH domain is a novel RNA binding domain, which suggests a molecular function for the uncharacterized proteins of the YTH protein family.

EXPERIMENTAL PROCEDURES

Recombinant YT521-B—Recombinant YT521-B was generated by cloning full-length YT521-B (7) and a YTH domain deletion mutant (YTHdel) into the vector pFastBac HTa (Invitrogen). Recombinant baculovirus and N-terminally His-tagged YT521-B or YTHdel was generated in Sf9 cells according to the manufacturer’s instructions. YT521-B does not express in bacterial systems.

SELEX—SELEX was performed as described previously (14) using recombinant YT521-B and a pool of N20-mers. The Matrix consensus sequence was determined using the CONSENSUS algorithm (15), searching for 6-mers on the sense RNA strand.

Minigene Analysis—Minigene analysis was essentially performed as described (16). The primers for the SXN reporter minigenes were: forward primer, CCATTGACCATTCACCCAC; reverse primer, CACTCTGTAGCTGTATGTT. With the exception of siRNA, all transfections were done using the calcium phosphate method, described in a previous study (16).

siRNA Analysis of Minigenes—siRNA analysis of minigenes was performed in HEK293 cells. The day before transfection, 40,000 cells were seeded in 0.5 ml of medium. On the day of transfection, 100 ng of minigenes and 34.5 ng of siRNA were diluted in 25 μl of HBS buffer (20 mM HEPES and 150 mM NaCl, pH 7.4). 3 μl of Hyperfect Transfection Reagent (Qiagen) was added to 22 μl of HBS buffer. After incubating for 5 min at room temperature, the two dilutions were mixed and incubated for another 5 min at room temperature. The transfection complexes were added onto the cells. 42 h after transfection cells were harvested for RNA and protein isolation. The siRNA used was UGGAUUUGCGAGCCGUAUU.

Array Analysis—Array analysis was performed using Affymetrix Splicearrays. 1 μg of RNA was labeled using the RibonMinus Kit (Invitrogen), and hybridization was performed according to the GeneChip Whole Transcript Sense Target Labeling Assay Manual (Affymetrix). For microarray analysis, the Partek software (Partek Inc., St. Louis, MO) was employed. The algorithms used to analyze microarray data is an alternative splicing analysis of variance model implemented in Partek Genomics Suite. A filter is also used to select for probe sets showing a significant alternative splicing score, which is determined at a 5% false discovery rate. Only transcripts that have a high alternative splicing score but show no significant differential expression at the gene level were included in the candidate list that was verified by RT-PCR.

YT521-B Binding Sites—YT521-B binding sites were identified using a previously described scoring-matrix based algorithm (17). Briefly, the overall score is calculated by adding the score of each position: each position score is based on frequency of a given base at this position and on the amount of information present at this position, including the number of sequences involved to design the matrix. In our case, the motif score is calculated for each 6-mer of sequences to be tested. All 6-mers having a score above a threshold of larger than 3.04 were predicted to be potential YTH binding sites.

Protein Fold Recognition Analysis—Protein fold recognition analysis to detect relationship between the YTH sequence and the solved protein structures was done via the GeneSilico meta-server (18).

Cloning, Expression, and Purification of the YTH Domain for NMR Study—The YTH domain (amino acids 346–502) of Rattus norvegicus YT521-B protein was cloned into pTYB11 vector (New England Biolabs). This vector was transformed into Escherichia coli BL21(DE3) Codon + RIL (Strategene). Cells were grown in minimal medium M9 containing 1 g/liter 15NH4Cl and 4 g/liter glucose (for 15N-labeled proteins) or 1 g/liter 15NH4Cl and 2 g/liter 13C-glucose (for 13C/15N-labeled proteins). Cell cultures were induced at A600 ~ 0.6 by 1 mM isopropyl β-D-thiogalactopyranoside at 18 °C and harvested after 30 h. Purification was performed according to the manufacturer’s instructions on a Chitin column (New England Biolabs). Dialysis was performed using a 3.5-kDa cutoff membrane (Spectrum) against 50 mM HNa2PO4, 50 mM NaCl, 10 mM β-mercaptoethanol, pH 7.0. The protein was concentrated to ~0.8 mM for measurements of the free protein. RNA (5′-GCAUAC-3′) was purchased from Dharmacon Research, deprotected according to the manufacturer’s instructions, desalted using a G-15 size-exclusion column (Amersham Biosciences), lyophilized, and resuspended in H2O (final concentration 1.3 mM). For titration experiments, RNA was stepwise added to the protein (0.2 mM concentration) in the NMR tube.

NMR Spectra—NMR spectra were acquired on AvanceIII-600 and Avance-900 Bruker spectrometers equipped with triple resonance probes. Data were measured at 20 °C. NMR spectra were processed with Topspin1.3 (Bruker) and analyzed with SPARKY 3 (T. D. Goddard and D. G. Kneller, University of California, San Francisco). The 1H, 13C, and 15N chemical shifts of the free protein backbone were assigned by standard methods (19). Figures of the NMR structure of the YTH domain were prepared using the program MOLMOL (20).

RESULTS

YT521-B Binds to a Degenerate RNA Sequence—Based on the predicted mixed α-helix β-sheet fold and conserved aromatic residues, we predicted that the YTH domain could bind to nucleic acids (6). To test this hypothesis, we expressed YT521-B protein in Drosophila cells using the baculovirus system and investigated nucleic acid binding properties of the protein. A protein lacking the YTH domain (YTHdel) served as a control (Fig. 1A). Filter binding and pulldown assays indicated that YT521-B binds to single-stranded RNA, but not single- or dou-
ble-stranded DNA sequences or double-stranded RNA (data not shown). To identify the sequence requirements of the RNA sequences that bind to YT521-B, we performed in vitro SELEX experiments using baculovirus-generated YT521-B protein and a pool of N20-mer RNA. After seven rounds of SELEX, we isolated RNA sequences that bind to recombinant YT521-B, which are shown in Fig. 1B. Similar to other RNA binding proteins, YT521-B binds to highly degenerate sequences that can be only described by a weight matrix, which is shown in Fig. 1C.

None of the identified sequences contained stable RNA secondary structures as predicted by the Zucker algorithm (21). We conclude that, similar to other RNA binding proteins (22), YT521-B binds to short RNA motifs that loosely follow a consensus sequence.

Next, we performed gel-retardation assays to test the ability of YT521-B to bind to RNA in a different experimental system. We used radioactively labeled probes containing the YTH domain binding motif and a non-related RNA probe lacking the motif (control). These probes were incubated in vitro with recombinant YT521-B, and a change in the probe’s mobility was detected using non-denaturing polyacrylamide gel electrophoresis. As a negative control, we used the YTH protein (YTHdel, Fig. 1A) that lacks the YTH domain, but contains all other protein parts. HeLa nuclear extract was used as a positive control. As shown in Fig. 2A, we observed a retardation in probe mobility in the presence of full-length YT521-B protein. We did not see a shift in mobility when the YTHdel protein, lacking the YTH domain, was used. A control RNA lacking the YTH binding motif did not show any RNA binding to YT521-B (Fig. 2B). Finally, we did not observe a mobility shift when a single-stranded DNA probe, with its sequence corresponding to the RNA probe, was used in the assay (Fig. 2C).

Next, we determined the affinity between the YT521-B protein and RNA by titrating the amount of oligonucleotide used (supplemental Fig. S1A). Quantification showed that the $K_d$ was 26 ± 8.5 μM. This affinity compared to a $K_d$ of 20 μM and 1 μM for the RRM containing splicing factors SRp20 and PTB (23–25) and a low micromolar affinity of K homology domains (26). However, other splicing factors, such as Fox-1, show a $K_d$ in the subnanomolar range (27). The data indicate that the YTH domain binds RNA with an affinity comparable to other splicing factors.

NMR Studies of the YTH Domain in Complex with GCAUAC—Based on the SELEX data (Fig. 1, B and C) we designed a hexanucleotide GCAUAC and tested its binding to the YTH domain in an NMR titration experiment (Fig. 3A). The spectra show that the domain binds the sequence in the fast to intermediate exchange regime on the NMR timescale (milliseconds) and that the YTH domain is folded in the free as well as in the bound state. The construct used for NMR experiments (R. nori-
A Novel RNA Binding Domain, YTH

Next, we compared the spectra of the free YTH domain and the YTH domain bound to the GCAUAC RNA hexanucleotide. The observed shifts between the spectra (Fig. 3A) allowed us to assign the RNA binding surface of the YTH domain to the deposited structure of the human YTH domain. From this mapping, we observe a clear RNA binding surface that involves five positively charged residues, because it contains seven lysines, three arginines, and one histidine (Fig. 3C). Among the residues on this surface, Lys-364, Trp-380, and Arg-478 are absolutely conserved among YTH domains, suggesting that RNA binding might be general for YTH domains. These data show that the YTH domain used positively charged residues to interact with bound RNA.

The YTH Binding Motif Causes Dependence of Alternative Exons on YT521-B Concentration—Previously, it was shown that increasing the YT521-B concentration influences alternative splice site selection (7). We therefore used reporter minigenes to determine whether the binding between YT521-B and RNA sequences has an effect on alternative splice site selection in vivo. 20-mer sequences containing either one or two hexamer YT521-B binding sites were introduced into the alternatively spliced exon of the SXN-minigene (28) (Fig. 4A), which has been widely used to analyze the impact of sequences on pre-mRNA processing. We tested several SELEX winners in the minigenes, including the sequences we analyzed in gel-retardation assays (Fig. 2). In all experiments, the amount of YT521-B was changed by cotransfecting an increasing amount of expression constructs with the reporter minigene construct. An identical construct lacking the YTH domain (YTHdel, Fig. 1A) was used as a negative control. As can be seen in Fig. 4 (B and C), YT521-B promoted inclusion of the alternative exon in a concentration-dependent manner. In contrast, no effect could be detected when comparable amounts of YT521-B lacking the YTH domain were transfected (YTHdel, Fig. 4, B, C, and E), or when a reporter minigene was used that contained no YTH binding sequence in its alternative exon (Fig. 4D).

Next, we used the same test system to investigate the effect of YT521-B depletion on alternative splice site selection and removed YT521-B by siRNA (Fig. 5A). siRNA was cotransfected with reporter minigenes containing one or two YT521-B binding sites, and the effect on splice site selection was determined by RT-PCR. In contrast to Fig. 4, where we transfected cells with calcium phosphate, we used a mixture of cationic and neutral lipids and longer times for these studies, which slightly changed the amount of basal exon inclusion. As shown in Fig. 5B, removal of YT521-B caused exon skipping in the reporter minigenes when the YT521-B binding site was present. In contrast, YT521-B did not affect exons lacking its binding site (Fig. 5C). Together, these data indicate that binding between YT521-B and pre-mRNA is necessary to influence alternative splice site selection in a YT521-B concentration-dependent manner.

FIGURE 2. Interaction of YT521-B with RNA. A, gel mobility shift analysis of YT521-B using RNA probes containing the YTH binding motif. 20 μM of recombinant, baculovirus-generated YT521-B and YT521-B lacking the YTH domain (YTHdel) were incubated with 15 μM YTH binding motif RNA probe. HeLa NE, HeLa nuclear extract serving as a positive control. Probe RNA: GCAUGC. B, gel mobility shift analysis of YT521-B using RNA probes lacking the YTH binding motif. 20 μM of recombinant, baculovirus-generated YT521-B and YT521-B lacking the YTH domain (YTHdel) were incubated with 15 μM control RNA probe (ctlRNA) which is a CUAUCU sequence lacking YTH binding motif. C, gel mobility shift analysis of YT521-B using DNA probes. 20 μM of recombinant, baculovirus-generated YT521-B and YT521-B lacking the YTH domain (YTHdel) were incubated with 15 μM single-stranded DNA probe. Probe DNA: GCATGC.
Three of the Conserved Residues in the YTH Domain Are Necessary for Splice Site Selection—A striking feature of the YTH domain is the full evolutionary conservation of 14 residues between its members (Fig. 6A) (6). To investigate their function for splice site selection, we mutated each residue and tested the mutants in cotransfection experiments using the MG-YT2 minigene (Fig. 4C). The regulation of its alternative exon by YT521-B requires the presence of the YTH domain (Fig. 4C), indicating that YT521-B exerts its effect by binding to the pre-mRNA and not by sequestering other splicing factors using its C terminus. The mutational analysis showed that changing the conserved residues Trp-380, Phe-412, and Gly-414 abolished the effect on splice site selection (Fig. 6B, circled in Fig. 6A). All mutants expressed similar amounts of protein of the expected size when compared with endogenous YT521-B (Fig. 6C), indicating that the change in activity is due to changes caused by the mutation. We confirmed the effect of the three point mutations on splice site selection by assaying them over a wider protein concentration range. As shown in Fig. 6D, even at higher concentrations that resulted in an increase of the mutant protein (Fig. 6E) none of the mutants was able to influence splice-site selection. A possible explanation for the loss of function could be a defect in RNA binding caused by the mutations. We tested this hypothesis by gel-shift analysis and found that mutating the three conserved mutants did completely abolish RNA binding (supplemental Fig. S1B). Because the residues are part of the hydrophobic core, it is likely that they are important for the structural integrity of the YTH domain. Their mutation could lead to a collapse of the domain and loss of RNA binding ability. In summary, these data show that three evolutionary conserved residues of the YTH domain are necessary for its ability to bind to RNA and modulate splice site selection.

The YTH Domain Does Not Dictate Protein Localization—YT521-B is localized in a novel nuclear compartment, the YT body. To test whether the YTH domain influences the localization of other YT proteins, we determined the intracellular localization of all known human YT protein family members. As shown in Fig. 5F, the deletion of the YTH domain did not influence the localization of YT521-B in YT bodies. Furthermore, the three other human YTH domain-containing proteins, HGRG8/YTHDF2, DACA1/YTHDF1, and YTHDF3 are all localized in the cytosol. Together, these data indicate that the YTH domain does not dictate the localization of the proteins in which it resides.
Array Analysis Shows That YT521-B Regulates Endogenous Alternative Exons That Contain Clusters of YTH Binding Signatures—So far, the influence of YT521-B on splice site selection was limited to model constructs. To determine the effect of YT521-B on splice site selection of endogenous targets we performed genome-wide Affymetrix Splicearray analysis. We compared the splicing patterns of all known human genes in the presence or absence of overexpressed EGFP-YT521-B. Previous analyses revealed several examples where the presence of the YTH domain is not needed for the ability of YT521-B to influence splice-site selection. In these cases YT521-B most likely acts by sequestering other splicing factors that its binds to (7, 11). This sequestration effect of splicing factors has been observed in other cases (29). To discriminate between direct and sequestration effects, we transfected either YT521-B or YT521-B lacking the YTH domain (YTHdel) into HEK293 cells and compared their mRNA isoforms with untransfected cells using Affymetrix Splicearrays. High scoring events were verified by RT-PCR using the primers indicated in A, \( p < 0.001 \). C, cotransfection assay using a minigene MG-YT2 with the seq2 sequence that contains two YTH binding motifs (seq2: GATGCATGCAATGGATGCGG), \( p < 0.01 \). D, cotransfection assay using a minigene containing a control sequence seq3 lacking the YTH binding motif (seq3: GGCGATAATGTGTAAATGCC). E, Western blot detecting expression of transfected YT521-B and YT521-B (YTHdel) in cell lysates of the transfection assays. An antibody against EGFP was used for detection. The relative levels of endogenous and transfected YT521-B are shown in Fig. 6C, lane 1.

**Figure 4.** Influence of YT521-B binding motifs on alternative splicing in vivo. A, structure of the SXN minigene (28). The construct consists of globin exon 1 and 2 (shaded) that flank a central artificial exon. The RNA sequence is introduced in the central exon, indicated by “motif.” Arrows indicate the localization of the primers used for RT-PCR. At least three independent experiments were evaluated using the Student’s \( t \) test. B, cotransfection assay using a minigene MG-YT1 with the seq1 sequence that contains one YT521-B binding motif, which is underlined (seq1: AGAGTCACTCTGTAGCTCA) sequence. A minigene containing this sequence was cotransfected with an increasing amount of YT521-B or YT521-B (YTHdel). The resulting RNA was analyzed by RT-PCR using the primers indicated in A, \( p < 0.001 \). C, cotransfection assay using a minigene MG-YT2 with the seq2 sequence that contains two YTH binding motifs (seq2: GATGCATGCAATGGATGCGG), \( p < 0.01 \). D, cotransfection assay using a minigene containing a control sequence seq3 lacking the YTH binding motif (seq3: GGCGATAATGTGTAAATGCC). E, Western blot detecting expression of transfected YT521-B and YT521-B (YTHdel) in cell lysates of the transfection assays. An antibody against EGFP was used for detection. The relative levels of endogenous and transfected YT521-B are shown in Fig. 6C, lane 1.

Array Analysis Shows That YT521-B Regulates Endogenous Alternative Exons That Contain Clusters of YTH Binding Signatures—So far, the influence of YT521-B on splice site selection was limited to model constructs. To determine the effect of YT521-B on splice site selection of endogenous targets we performed genome-wide Affymetrix Splicearray analysis. We compared the splicing patterns of all known human genes in the presence or absence of overexpressed EGFP-YT521-B. Previous analyses revealed several examples where the presence of the YTH domain is not needed for the ability of YT521-B to influence splice-site selection. In these cases YT521-B most likely acts by sequestering other splicing factors that its binds to (7, 11). This sequestration effect of splicing factors has been observed in other cases (29). To discriminate between direct and sequestration effects, we transfected either YT521-B or YT521-B lacking the YTH domain (YTHdel) into HEK293 cells and compared their mRNA isoforms with untransfected cells using Affymetrix Splicearrays. High scoring events were verified by RT-PCR using the primers indicated in A, \( p < 0.001 \). C, cotransfection assay using a minigene MG-YT2 with the seq2 sequence that contains two YTH binding motifs (seq2: GATGCATGCAATGGATGCGG), \( p < 0.01 \). D, cotransfection assay using a minigene containing a control sequence seq3 lacking the YTH binding motif (seq3: GGCGATAATGTGTAAATGCC). E, Western blot detecting expression of transfected YT521-B and YT521-B (YTHdel) in cell lysates of the transfection assays. An antibody against EGFP was used for detection. The relative levels of endogenous and transfected YT521-B are shown in Fig. 6C, lane 1.

**Figure 4.** Influence of YT521-B binding motifs on alternative splicing in vivo. A, structure of the SXN minigene (28). The construct consists of globin exon 1 and 2 (shaded) that flank a central artificial exon. The RNA sequence is introduced in the central exon, indicated by “motif.” Arrows indicate the localization of the primers used for RT-PCR. At least three independent experiments were evaluated using the Student’s \( t \) test. B, cotransfection assay using a minigene MG-YT1 with the seq1 sequence that contains one YT521-B binding motif, which is underlined (seq1: AGAGTCACTCTGTAGCTCA) sequence. A minigene containing this sequence was cotransfected with an increasing amount of YT521-B or YT521-B (YTHdel). The resulting RNA was analyzed by RT-PCR using the primers indicated in A, \( p < 0.001 \). C, cotransfection assay using a minigene MG-YT2 with the seq2 sequence that contains two YTH binding motifs (seq2: GATGCATGCAATGGATGCGG), \( p < 0.01 \). D, cotransfection assay using a minigene containing a control sequence seq3 lacking the YTH binding motif (seq3: GGCGATAATGTGTAAATGCC). E, Western blot detecting expression of transfected YT521-B and YT521-B (YTHdel) in cell lysates of the transfection assays. An antibody against EGFP was used for detection. The relative levels of endogenous and transfected YT521-B are shown in Fig. 6C, lane 1.
YT521-B regulates vertebrate-specific RNAs, which it binds to via its YTH domain.

**DISCUSSION**

*The YTH Domain Is a Novel RNA Binding Domain*—We analyzed the role of the YTH domain in the founding member of the YTH domain family of proteins, YT521-B. Our data show that the function of the YTH domain is to bind to single-stranded, non-structured RNA. The relatively weak binding affinity of 26 ± 8.5 μM is comparable to other splicing regulators, such as SRp20, PTB, and K homology domain-containing proteins (23–26). However, because the binding is weak compared with other splicing factors such as Fox-1, it is likely that YT521-B binding to RNA is aided by interacting proteins that stabilize its RNA binding, which is common among splicing factors.

Another feature of RNA sequences recognized by YT521-B and several splicing factors is their high degeneracy. This degeneracy of the RNAs binding to the YTH domain was also observed for the other biochemically characterized member of the YTH domain proteins, Mmi1 (12). Mmi1 binds to RNA of at least four yeast target genes, which contain a cis-acting element called determinant of selective removal. No strong consensus sequence could be identified in these determinants of selective removal, suggesting that the YTH domain of Mmi1 also recognizes highly degenerate RNA sequences. The Splicearray analysis shows that the YTH domain regulates distinct pre-mRNAs, which shows high specificity in vivo. Because all in vivo targets of YT521-B contain clusters of the binding motif, the high specificity observed in vivo is most likely due to binding of multiple YT521-B molecules to a single RNA. The formation of these complexes is facilitated by the ability of YT521-B C termini to bind to each other (7). It is therefore likely that the high specificity of YT521-B in vivo is achieved by combining interactions between different proteins and binding of their YTH domains to several degenerate motifs on the RNA.

A further similarity of YT521-B and the YTH domain-containing protein Mmi1 (12) is their localization in dynamic nuclear foci. YT521-B is localized in a novel nuclear compartment, the YT body, whose formation is regulated by phosphorylation (11). Mmi1 is found in several scattered foci during mitosis but, due to its interaction with Mei2, converges into one single dot in meiosis (12). The comparison of all YTH domain
The YTH family members shows that in addition to the YTH domain they contain numerous other protein domains that mostly enable protein interaction. Yeast two-hybrid analysis of YT521-B demonstrated that the YTH domain is not necessary for binding to other proteins (7). A possible general function of the YTH domain could therefore be to add RNA binding ability to proteins containing other interaction domains.

The YTH domain is not a cellular targeting signal, as we demonstrated that the human YTH protein family members reside both in the nucleus and cytosol. It is therefore likely that YTH protein family members function in different aspects of cellular RNA metabolism.

The YTH Domain Is Related to the PUA-fold—New PSI-BLAST searches of databases for YTH homologs revealed no obvious relationships to sequences other than the previously described members of the YTH family (6). To identify related structures, we performed a fold recognition analysis using the GeneSilico metaserver (18). We found a significant relationship of the YTH domain (residues 358–495 of YT521-B) to a group of closely related protein structures from the uncharacterized protein family DUF55. The images underneath show the statistical evaluation of three independent experiments and the increase of expression of each mutant, which was detected by Western blot of cellular lysates using an antibody against EGFP. E, localization of YTH-domain containing proteins. Proteins that were tagged with EGFP at the N terminus were expressed in HEK293 cells. The images show representative cells. YT521-B and YTHdel are localized in the nucleus. The images are enlarged to show the nuclear substructure.
even though they exhibit different elaborations in the C terminus: α3-β7-α4-α5 in YTH but β7 in PUA. The YTH domain also exhibits an additional N-terminal α helix that is missing from PUA and DUF55 domains. Supplemental Fig. S2 illustrates the common topology of the polypeptide chain of YTH/DUF55 and PUA structures, local structural differences notwithstanding. In support of our analysis the DUF55 family has been recently included in the PUA superfamily in the SCOP data base. Together, our data indicate that the YTH domain is closely related to the DUF55 family and more remotely related to the PUA domain that has been previously implicated in RNA binding.

**Features of RNA Binding by the YTH Domain**—NMR titration experiment of the YTH domain in complex with the GCAUAC oligonucleotide showed that the RNA interacts with a protein surface containing eleven positively charged amino acids. To determine the role of the evolutionary conserved residues we performed a mutational analysis. We could mutate most conserved residues without changing the ability of the mutants to change splice site selection. The only exceptions were residues Trp-380, Phe-412, and Gly-414. Mutating these residues abolished RNA binding of YT521-B. Mapping of the mutated residues onto the 2yud structure revealed that the invariant residues Trp-380 and Phe-412 are in the hydrophobic core and are therefore most likely essential for the YTH domain stability. Gly-414 is located within a buried β-strand, where any bigger side chain would interfere with the backbone of a neighboring β-strand. Phe-412 is adjacent to this residue in the same β-strand and is probably needed for proper folding. Again, this is reminiscent to a glycine residue that is conserved due to steric reason both in PUA domains (32) and in the structurally related classic OB-fold (4).

However, the remaining nine invariant residues YTH domain can be mutated without interfering with splice site selection. These residues map on the surface of the YTH OB-fold (31). For example 2yud/YTH exhibits root mean square deviation of 3.7 Å over 84 superimposable residues of 2ane/PUA structure (DALI Z-score 3.3). They share the core pattern of secondary structures: β1-α1-β2-α2-β3-β4-β5-β6,
domain (Fig. 3) and with the exception of Lys-364 and Arg-478 are not located on the RNA binding surface. The substitution of individual residues that belong to this group may locally influence the structure, but it is unlikely to disrupt the global fold. It is possible that these residues are evolutionary conserved to form a scaffold that allows the arrangement of positively charged residues that bind to RNA. Because these positively charged residues are less conserved among YTH domain family members, it is likely that individual members have different RNA binding properties.

**YT521-B Is Found Only in Vertebrate Genomes**—The YTH domain is found in all eukaryotes. Data base analysis indicates that its founding member, YT521-B, is present only in vertebrate genomes. Similar to other splicing factors like SF2/ASF, its founding member, YT521-B, is present only in vertebrate genomes. Data base analysis indicates that its founding member, YT521-B, is present only in vertebrate genomes. Because these positively charged residues are less conserved among YTH domain family members, it is likely that individual members have different RNA binding properties.

**YT521-B Can Influence the Generation of Proteins from Different Genes**. YT521-B can influence the generation of proteins from different genes. A detailed analysis of NOVA target genes demonstrated that this splicing regulatory protein influences a set of functionally related genes (33). The example of YT521-B shows that a splicing factor cannot only regulate functionally related pre-mRNAs, but also evolutionary related RNAs. YT521-B target genes share an array of YTH binding sites on the RNA, which is probably recognized by multiple YT521-B molecules that also bind to each other. The formation of this vertebrate-specific complex generates then a vertebrate-specific readout of the genome.

**REFERENCES**

Supplementary Figure 1: Gel mobility shift analysis

A. Gel mobility shift analysis of a set of equilibrium binding reactions including variable concentrations of YTH domain protein and 2.6µM of RNA probe containing one YTH-binding motif. Probe RNA: GCAUGC. RNA probe was uniformly labeled with 32P. The concentration of YTH domain protein ranges from 0.86µM to 12.9µM. After electrophoresis, the gel was dried and scanned on a Typhoon Scanner. Bound and free RNA were quantified in each, using ImageQuant software. The Kd calculated from the plot ln([RNA_{bound}]/[RNA_{free}]) vs ln[Protein_{free}] is 26µM with a standard deviation of 8.5µM.

Measuring equilibrium binding constant (K_d) using gel retardation assay.

When analyzing RNA:Protein complexes, these complexes coexist with free RNA and protein in the binding reagent. This dynamic balance can be described by the equilibrium equation:

\[ [\text{RNA:Protein}_{\text{bound}}] \rightleftharpoons [\text{RNA}_{\text{free}}] + [\text{Protein}_{\text{free}}] \]

The equilibrium binding constant K_d, which describes the RNA and protein interaction is formalized by the equation: K_d = [RNA_{free}] [Protein_{free}]/[RNA:Protein_{bound}], as the concentration of RNA:Protein complexes is equal to the concentration of bound RNA, the equation can be rearranged to \((1/K_d)[Protein_{free}] = [RNA_{bound}]/[RNA_{free}]\). After natural logarithm is applied, the equation became linear: \(\ln([RNA_{bound}]/[RNA_{free}])=\ln[Protein_{free}] + \ln(1/K_d)\). By plotting \(\ln([RNA_{bound}]/[RNA_{free}])\) vs \(\ln[Protein_{free}]\), the K_d can be obtained when the concentration of bound RNA is equal to the concentration of free RNA.

B. Gel mobility shift analysis of YTH domain mutants (W380D, F412D and G414I) using RNA probes containing the YTH-binding motif. 20µM of recombinant, bacteria-expressed YTH domain and YTH domain mutants were incubated with 15µM YTH-binding motif RNA probe. Probe RNA: GCAUGC.

Supplementary figure 2: Comparision between the YTH domain and the PUA/DUF55 domain

A-C. Structural similarities between the YTH, DUF55 and PUA domains. A. YTH domain, (PDB: 2yud), B. DUF55 domain (PDB 2ar1), C. PUA domain PDB: 2ane (N-terminal PUA domain of E. coli Lon protease). The secondary structure elements are colored from N-terminus (blue) to C-terminus (red) to emphasize the similarity.

D. Localization of the conserved residues on the YTH domain. Numbers are indicated as in Figure 5A.
Supplemental Fig. 1

A

YTH domain protein

0.86 µM
2.16 µM
6.47 µM
10.8 µM
12.9 µM
0 µM

B

RNA probe
YTH domain
YTHW380D
YTHF412D
YTHG414I

\[
y = 0.5264x + 5.4472
\]

\[
\ln[\text{protein}_{\text{free}}]
\]

\[
\ln[\text{RNA}_{\text{bound}}/\text{RNA}_{\text{free}}]
\]