Published in final edited form as:

Nat Struct Mol Biol. 2008 December ; 15(12): 1343–1351. doi:10.1038/nsmb.1519.

Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1

Marianna Teplova and Dinshaw J. Patel

Structural Biology Program, Memorial Sloan-Kettering Cancer Center, 439 East 67th Street, New York, NY 10021, USA

Abstract

Muscleblind-like (MBNL) proteins, regulators of developmentally programmed alternative splicing, harbor tandem CCCH zinc-finger (ZnF) domains that target pre-mRNAs containing YGCU(U/G)Y sequence elements (where Y is a pyrimidine). In myotonic dystrophy, reduced levels of MBNL proteins lead to aberrant alternative splicing of a subset of pre-mRNAs. The crystal structure of MBNL1 ZnF3/4 bound to r(CGCUGU) establishes that both ZnF3 and ZnF4 target GC steps, with site-specific recognition mediated by a network of hydrogen bonds formed primarily with main chain groups of the protein. The relative alignment of ZnF3 and ZnF4 domains is dictated by the topology of the interdomain linker, with a resulting antiparallel orientation of bound GC elements, supportive of a chain-reversal loop trajectory for MBNL1-bound pre-mRNA targets. We anticipate that MBNL1-mediated targeting of looped RNA segments proximal to splice-site junctions could contribute to pre-mRNA alternative-splicing regulation.

The generation of functionally diverse proteins required for cell growth and differentiation in metazoan organisms is critically dependent on alternative splicing of pre-mRNAs. Alternative-splicing regulators control the expression of tissue-specific or developmental stage-specific protein isoforms through binding either to splice sites directly or to other sequences in the pre-mRNA, thereby enhancing or repressing the inclusion of alternative exons¹.

The proteins of the MBNL family have been identified as important tissue-specific alternative-splicing regulators that have a key role in terminal muscle differentiation². MBNL proteins promote splice-site choice appropriate for adult muscle tissues, and they act as either activators or repressors of splicing on different transcripts³. Postnatal activation of MBNL proteins leads to repression of fetal splicing patterns in numerous terminal muscle-

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

Correspondence should be addressed to D.J.P. (pateld@mskcc.org).

Accession codes. Protein Data Bank: Coordinates for MBNL1 ZnF1/2, MBNL1 ZnF3/4 and MBNL1 ZnF3/4-r(CGCUGU) have been deposited under the accession codes 3D2N, 3D2Q and 3D2S, respectively.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

AUTHOR CONTRIBUTIONS

The biochemical and structural research was undertaken by M.T. under the supervision of D.J.P. Both M.T. and D.J.P. contributed to the writing of the paper.

differentiation transcripts and to the expression of adult protein isoforms⁴. This normal splicing pattern is altered specifically in the neuromuscular disease myotonic dystrophy (DM), in part owing to inactivation of MBNL. Several genes, including the insulin receptor (*INSR*) and skeletal muscle ion-transporter chloride channel 1 (*CLCN1*), follow aberrant alternative mRNA splicing in myotonic dystrophy tissues, ultimately causing the clinical symptoms^{5,6}.

The genetic basis of myotonic dystrophy is an expansion of CTG trinucleotide repeats in the 3' untranslated region of the dystrophin myotonic protein kinase (*DMPK*) gene (DM1)⁷, or a large CCTG repeat expansion in intron 1 of the zinc finger 9 (*ZNF9*) gene (DM2)⁸. MBNL proteins have been shown to accumulate with both CUG- and CCUG-expanded repeat-bearing transcripts in nuclear foci and to bind CHG and CHHG repeats (where H is A, U or C) in a length-dependent manner^{9,10}.

According to the MBNL-based loss-of-function hypothesis for myotonic dystrophy pathogenesis supported by mouse models, sequestration of MBNL proteins associated with RNAs containing large CUG- and CCUG-repeat expansions in ribonuclear foci leads to misregulated alternative splicing^{9,11–13}. Although MBNL proteins are key targets in RNA-mediated pathogenesis, the additional factors CUG triplet repeat, RNA binding protein 1 (CUGBP1) and Heterogeneous nuclear ribonucleoprotein H (hnRNP H) have also been shown to contribute independently to aberrant splicing in myotonic dystrophy cells^{14–16}.

There are three MBNL genes in humans: MBNL1, MBNL2 and MBNL3. Of these, MBNL1, which is expressed at high levels in muscle, seems to promote muscle differentiation and has a pivotal role in the pathogenesis of muscle disease in myotonic dystrophy. MBNL3, which is expressed predominantly in the placenta and not in muscles, seems to function in an opposing manner by inhibiting the expression of muscle-differentiation markers^{11,17}. MBNL2, found at similar levels in all tissues, has a major role in cytoplasmic mRNA localization¹⁸.

Each of the three human MBNL proteins contain four highly conserved ZnF RNA binding domains of the CCCH type, arranged in tandem pairs that are positioned toward the N-terminal segment (ZnF1 and ZnF2) and in the middle (ZnF3 and ZnF4) of the polypeptide chain (Fig. 1a,b). ZnF1 and ZnF3 show CX₇CX₆CX₃H spacing between the zinc-coordinated residues, whereas ZnF2 and ZnF4 share the CX₇CX₄CX₃H sequence (Fig. 1c). Within each tandem segment, vertebrate proteins are 99% identical; the most distantly related muscleblind proteins, which contain only one ZnF pair, share 67% identity with human MBNL1 (Fig. 1c). Three linkers of 14 amino acids, 110 amino acids and 16 amino acids separate the ZnF domains in MBNL1 (Fig. 1a).

Tandem CCCH ZnF domains were first structurally characterized in the TIS11d protein, a member of the tristetrapolin (TTP) protein family, which is involved in the control of the inflammatory response and required for the rapid degradation of mRNAs that contain an AU-rich element (ARE) in their 3' untranslated region. The NMR structure of TIS11d bound to a model AU-rich RNA element shows that each of two ZnFs recognizes a UAUU repeat through stacking interactions and hydrogen bonds to the Watson-Crick edges of the bases¹⁹.

Unexpectedly, the sequence specificity is achieved primarily through interactions mediated by the protein backbone.

Full-length MBNL proteins containing four ZnF domains have been shown to regulate alternative-splicing events within the cardiac troponin T (*cTNT*) pre-mRNA, where the RNA target contains two YGCU(G/U)Y intronic sequence motifs near the alternative exon³. To elucidate the principles of YGCU(G/U)Y element recognition by the CCCH ZnF motifs of MBNL proteins, we solved the crystal structures of two tandem ZnF domains (ZnF1/2 and ZnF3/4) of human MBNL1 in the free form, as well as the crystal structure of second tandem ZnF3/4 domain bound to an r(CGCUGU) oligonucleotide.

Both tandem ZnF domains adopt a previously uncharacterized symmetric overall fold, in which the two ZnFs are connected by a structured linker that orients their RNA binding surfaces away from each other. This arrangement results in an antiparallel orientation of the two single-stranded RNA chains bound by ZnF3 and ZnF4 in the complex. The structure of the complex highlights how ZnF3 binds GC and ZnF4 binds GCU sequences, thereby revealing a common GC sequence-recognition motif. Our structural results and binding studies indicate that the tandem ZnF domains of MBNL1 target two separated GC(U) sites within a longer mRNA, thereby inducing a chain-reversal trajectory in the bound RNA.

RESULTS

Choice of protein and RNA constructs

Both ZnF1/2 and ZnF3/4 constructs (Fig. 1b) showed well-dispersed ¹⁵N-HSQC NMR spectra, with ZnF3/4 having the higher-quality spectrum. Each of eight RNA sequence variants with the consensus YGCU(G/U)Y 6-nucleotide (nt) sequence caused similar chemical shift changes in several ZnF3/4 residues (data not shown), consistent with protein-RNA complex formation.

Crystallization and structure determination

We successfully grew diffraction-quality crystals of individual ZnF1/2 (2.7- Å resolution, space group *P*3₁) and ZnF3/4 (1.5 Å, *P*2₁) constructs in the free form, and also of ZnF3/4 bound to an r(CGCUGU) sequence (1.7 Å, *P*2₁). We determined the structures of free ZnF1/2 and ZnF3/4 by SAD phasing on zinc atoms (Methods) and solved the structure of the complex by molecular replacement, using the refined X-ray structure of the free ZnF3/4 as a search model. The structure contains four protein molecules, each containing ZnF3 and ZnF4 motifs, and 16 RNA nucleotides originating from four 6-nt RNA strands in the asymmetric unit. A difference electron-density omit map for the bound RNA segment of the complex is shown in Supplementary Figure 1 online.

Structure of MBNL1 tandem ZnF1/2 and ZnF3/4 domains

The ZnF3/4 domain in both the free (Fig. 1d) and r(CGCUGU)-bound states, as well as ZnF1/2 in the free state, adopt folds in which structurally similar ZnFs are related by a two-fold rotation axis. Three cysteine residues and one histidine residue coordinate the zinc atom in each finger (Fig. 1c,d). In all ZnFs, there is a short α-helix between the first and the

second coordinated cysteines. A 3_{10} -helix element is found between the second and the third coordinated cysteines in ZnF1 and ZnF3, which have spacing of 2 residues longer than that found in ZnF2 and ZnF4. Otherwise, the overall fold of all four MBNL1 ZnFs is virtually the same.

A three-stranded β -sheet positioned between the tandem ZnF domains determines their antiparallel orientation relative to each other for both ZnF1/2 and ZnF3/4 (Fig. 1d). ZnF1/2 contains an extended α -helix at the C terminus (Supplementary Fig. 2a online) comprising residues Pro73 to Gly83 and has an interdomain linker 2 residues shorter than that found in ZnF3/4 (Fig. 1c). Other than these differences, the peptide backbone topology of ZnF1/2 is essentially the same as that of ZnF3/4 (pairwise C α r.m.s. deviation of 1.17 Å for 62 matching residues; Fig. 1c,d).

The 16-residue (Pro205 to Val220) linker connecting the ZnF3 and ZnF4 domains participates in the formation of a central β -sheet (Fig. 1d) stabilized by six NH-O main chain hydrogen bonds involving the ZnF3 N-terminal segment (Asp180 to Val184) and the interdomain linker sequence (Asp212 to Val220) (Fig. 1e). The solvent-exposed face of the β -sheet projects a hydrophilic surface, whereas its buried face is lined with conserved hydrophobic residues.

Crystal packing and stoichiometry of the complex

Different packing interactions are observed in the crystal lattices of free ZnF1/2 (Supplementary Fig. 2a) and ZnF3/4 (Supplementary Fig. 2b,c). The integrated structural unit of the complex shown in Figure 2 contains two ZnF3/4 domains bound to four RNA strands and includes protein molecules A and B and RNA molecules E and F within the asymmetric unit; it also contains two symmetry-related RNA molecules that interact with protein molecule A, namely E_s and the C1 nucleotide of molecule F_s. Three out of four ZnF domains in this complex (ZnF3 and ZnF4 of molecule A and ZnF3 of molecule B) bind RNA, with the remaining ZnF (ZnF4 of molecule B) involved in protein-protein contacts (Fig. 2b). The two ZnF3/4 molecules in the crystals of both the free (Fig. 2a) and RNA-bound ZnF3/4 (Fig. 2b) interact in such a way that the RNA binding surface of one of the ZnF4 domains engages in protein-protein interactions with the ZnF4 of another molecule and therefore becomes inaccessible to the RNA.

The superimposed MBNL1 ZnF3 (Val184 to Pro205, excluding the Gly196 to Asn197 insertion) and ZnF4 (Val220 to Pro239) domains in the RNA bound state (stereo view in Supplementary Figure 3 online) show a pairwise C α r.m.s. deviation of 0.53 Å. Both ZnF domains of ZnF3/4 retain their conformation upon RNA binding (pairwise C α r.m.s. deviation of 1.1 Å).

It is conceivable that the dimerization of ZnF3/4 observed in crystals of both the free (Fig. 2a) and RNA-bound (Fig. 2b) states of MBNL1 could originate from crystal lattice contacts, and it remains to be demonstrated whether such dimerization also occurs in solution.

RNA binding surfaces as protein-protein interacting domains

The same surface of ZnF4 with highly conserved interfacial residues is involved in protein-RNA (Fig. 2d, ZnF4 molecule A) and protein-protein (Fig. 2c, ZnF4 molecule B) interactions in the crystal structure of the MBNL1 ZnF3/4-r(CGCUGU) complex (Fig. 2b). The concept of an RNA binding domain also being used as a protein-protein interaction domain has been observed for other splicing regulators, such as U1A²⁰ and U2AF65 (refs. 21,22).

The ZnF4-ZnF4 interface in the crystal, which buries 1,320 Å² of protein surface, is defined by hydrophobic and hydrogen-bonding contacts involving residues within ZnF4 motifs of both molecules, as well as residues from the N terminus (Arg178 and Arg181) of molecule B and from the C terminus (Ala241, His242 and Gln244) of molecule A (Fig. 2c). The protein-protein interface blocks the RNA binding surface of the ZnF4 domain of molecule B, but leaves the same surface of molecule A available for RNA binding (Fig. 2d).

Base-specific adjoining binding pockets in complex

There are three sets of protein-RNA interactions observed in the crystal structure of the ZnF3/4 complex with C1-G2-C3-U4-G5-U6 single-stranded RNA (Fig. 2b). First, ZnF3 of molecule A contacts G5 and C1 nucleotides originating from two separate RNA strands, E and F_s. Second, ZnF4 of molecule A contacts the G2-C3-U4 segment of molecule E_s. Third, ZnF3 of molecule B binds the G2-C3 dinucleotide step of molecule F. In all three ZnF-RNA recognition events, the guanine and cytosine bases in a 5'-GC-3' context insert into adjoining base-specific binding pockets, resulting in the RNA backbone pointing outward from the protein surface. The RNA strands bound to ZnF3 of molecule A are aligned antiparallel to the RNA bound to ZnF4 of the same molecule, as shown in ribbon (Fig. 3a) and surface (Fig. 3b) views of the protein in the complex. The RNA binding pockets for G2 and C3 within ZnF3 and ZnF4 adopt the same peptide backbone topology and include one aromatic and one or two conserved arginine side chains. These are highlighted in a ribbon view of the intermolecular contacts involving the G2-C3-U4 segment (Fig. 3c) and the corresponding surface view of G2 and C3 inserted into adjoining binding pockets (Fig. 3d) for ZnF4 of molecule A.

Stacking and hydrogen-bonding contributions to recognition

A characteristic feature of the guanine binding pocket in the MBNL1 complex is a conserved subset of aromatic amino acids in the middle of the protein—Phe202 in ZnF3 (Fig. 4) and Tyr236 in ZnF4 (Fig. 4b)—which are inserted between the guanine and cytosine bases. The aromatic ring of Phe202 forms stacking interactions with C1 (Fig. 4d) and C3 (Fig. 4f) in ZnF3 complexes.

The guanine base in the GC step stacks over the guanidinium group of invariant Arg195 in ZnF3 (Fig. 4a,c) and Arg231 in ZnF4 (Fig. 4b) complexes, with these arginines being part of the loop segment between the second and the third cysteines of the ZnF3 and ZnF4 domains. The guanine N² and N¹ atoms are hydrogen-bonded to sulfur atoms of the first and the third cysteines coordinated to the zinc atom, whereas the guanine O⁶ forms a hydrogen bond with

the backbone amide group of the residue following the third cysteine of the ZnF domains (Fig. 4a–c).

The cytosine base in the GC step forms three intermolecular hydrogen bonds through its Watson-Crick edge with the backbone amide and carbonyl groups in each of the three recognition events (Fig. 4d–f) and an additional fourth hydrogen bond involving the Glu183 side chain oxygen of ZnF3 (Fig. 4d,f). Conserved Arg186, whose guanidinium group is packed on top of the cytosine ring (Fig. 4d,f), is the third side chain within the ZnF3 GC binding pocket involved in a stacking interaction with a base. The position of Arg186 is occupied by U4 of G2-C3-U4 bound to ZnF4 (Fig. 4e), which is anchored in place through a single base-specific hydrogen bond between its N³ atom and the side chain of conserved Asp180.

The complex also shows a network of intermolecular hydrogen-bonding interactions between the RNA sugar-phosphate backbone and the side chains of the ZnF domains (Fig. 4).

ZnF3/4 preferentially targets two separated GCU steps

In view of our structural results suggesting a chain-reversal loop trajectory for the target pre-mRNA bound to ZnF3/4, we next turned to RNA sequences containing two GCU trinucleotides separated by an intervening sequence as a potential ligand of the MBNL1 ZnF3/4 domain. Previous studies have identified two MBNL1 binding motifs in both human and chicken *cTNT* intron regions adjacent to alternative exons³. In human *cTNT* (*hcTNT*) pre-mRNA the two GCU triplets, which constitute two binding sites, are separated by 10 nt (Supplementary Fig. 4a online), whereas in chicken *cTNT* pre-mRNA they are separated by 25 nt³. We performed binding studies of ZnF3/4 with a *hcTNT* 5'-

UCGCUUUUCCCCUCCGCU-3' 18-nt pre-mRNA fragment, which spans residues 23–40 upstream of exon 5 (Supplementary Fig. 4a). ZnF3/4 binds this 18-nt RNA with 1:1 stoichiometry as determined by an electrophoretic mobility shift (gel-shift) assay (EMSA; Fig. 5a).

For a quantitative estimation of binding efficiency between MBNL1 ZnF3/4 and the 18-nt *hcTNT* RNA, protein fluorescence intensity was measured as a function of RNA concentration (Fig. 5b). Addition of 18-nt *hcTNT* RNA results in 90% quenching of the intrinsic tyrosine fluorescence of ZnF3/4. The shorter RNA sequences CGCUGU, used for crystallographic studies, and UGCU also affect protein fluorescence, causing approximately 90% and 70% quenching, respectively, at the highest concentrations used (3 μ M). The fluorescence titration data yielded a K_d of 40 ± 10 nM for ZnF3/4 binding to 18-nt *hcTNT* RNA, indicative of strong interactions between ZnF3/4 and the RNA ligand containing two GCU binding sites separated by a 10-nt sequence.

Effect of spacing between the two GCU sites

We have compared ZnF3/4 binding affinities for the unmodified *hcTNT* 18-nt fragment containing a 10-nt spacer between GCU sites (S10), against modified sequences with decreased (S5) and increased (S15) spacer elements, as part of an effort toward validation of

the proposed model involving a chain-reversal loop RNA trajectory induced by MBNL1 complex formation. Similar to S10, ZnF3/4 binds S15 with 1:1 stoichiometry as determined by gel-shift assays (Fig. 5c). No clear band shift was observed for S5 upon adding the equivalent concentration of ZnF3/4. However, smearing of the S5 band at a 3:1 protein:RNA ratio may be indicative of a weak protein-RNA interaction (Fig. 5c). The similar binding affinities measured for S10 ($K_d = 40 \pm 10$ nM) and S15 ($K_d = 50 \pm 20$ nM), together with the approximately 3.5-fold lower binding affinity measured for S5 ($K_d = 140 \pm 30$ nM) (Fig. 5d), as measured from fluorescence-monitored titration studies, are consistent with an optimal spacer length of between 10 and 15 nt.

DISCUSSION

We outline below the key features associated with GC-specific recognition by the ZnF domains of MBNL1 and how chain-reversal of the RNA trajectory upon binding the bipartite ZnF3/4 architecture of this splicing regulator could facilitate regulation of alternative splicing of pre-mRNAs. We anticipate that such MBNL1-mediated sequestration of splice-site junctions could prevent access to the splicing machinery, thereby facilitating exon skipping and the formation of alternatively spliced constructs.

MBNL1 ZnF3/4 domains primarily target GC steps

Molecular recognition between MBNL1 ZnF3 and ZnF4 with the GC-containing steps of its RNA target is dominated by intermolecular stacking and hydrogen-bonding interactions, such that ZnF3 specifically targets the GC dinucleotide step (Fig. 4a,c,d,f) and ZnF4 targets the GCU trinucleotide step (Fig. 4b,e). The guanine and cytosine bases of the GC step are inserted into adjoining binding pockets of ZnF3/4 (Supplementary Fig. 5 online), which are composed of conserved aromatic, arginine and lysine side chains. The Watson-Crick base edges of guanine (Fig. 4a–c) and cytosine (Fig. 4d–f) within the GC step form intermolecular hydrogen bonds primarily with backbone amide and carbonyl groups as well as the two zinc-coordinated cysteines that line the bottom of the adjoining binding pockets.

Somewhat similar interactions have been reported for extruded guanines that insert into binding pockets in RNA complexes of CCHC zinc knuckle domains of retroviral²³ and HIV-1 (ref. 24) nucleocapsid proteins, as well as in the complex of the fourth C₂H₂ zinc finger of TFIIIA with loop E of 5S rRNA^{25,26}.

The GC sequence specificity of MBNL1 seems to have biological relevance because simultaneous mutation of four guanines within the two MBNL1 binding sites of a *hcTNT* minigene substantially (100-fold) reduces the ability of MBNL1 to cross-link with this RNA, as well as its ability to regulate splicing of *hcTNT* exon 5 (ref. 3). Similar observations have also been reported in independent binding studies using a *hcTNT* 50-nt RNA construct (residues 8–58 upstream of exon 5) and MBNL1 (residues 1–260)²⁷.

Significance of antiparallel aligned tandem ZnF domains

The antiparallel orientation of the RNA strands bound by ZnF3 and ZnF4 in the crystal structure of the MBNL1 ZnF3/4–r(CGCUGU) complex has not been reported previously for other tandem CCHC ZnF domain proteins¹⁹ and is indicative of a chain-reversal loop

trajectory for the bound MBNL1 RNA target, given the separation between the two ZnF binding sites. A working model involving the binding of all four ZnF domains of MBNL1 to a long RNA is shown in Supplementary Figure 4b.

A similar antiparallel alignment topology has been reported previously for the tandem RNA binding RRM3 and RRM4 domains of the polypyrimidine tract-specific splicing regulator PTB²⁸. The RRM3 and RRM4 domains of PTB form a heterodimer mediated by an extensive hydrophobic interface, whereas their RNA binding surfaces point away from each other, with the bound 6-nt RNAs aligned in an antiparallel orientation. For such an alignment of RNA binding sites, the PTB RRM3/4 heterodimer has the capacity to bring two remote RNA pyrimidine tracts within ~30 Å of each other²⁸. This distance is comparable to the ~20 Å separating the two GC RNA binding sites in the MBNL1 ZnF3/4 complex. Thus, MBNL1 and PTB, both alternative-splicing regulators, share similar antiparallel arrangements of their RNA binding sites on opposing faces of the heterodimer, thereby achieving optimal binding of their RNA targets found in multiple copies within intronic regions near the alternatively spliced exons and in the regulated exons themselves.

Model of MBNL1-mediated pre-mRNA splicing repression

A possible model for the repression of exon 5 in the *hcTNT* pre-mRNA by MBNL1 could involve direct MBNL1 binding upstream of the regulated exon, thereby inhibiting recognition of the 3' splice site by the splicing machinery (Supplementary Fig. 4c), culminating in exon skipping²⁷. The crystal structure of ZnF3/4 bound to two antiparallel RNAs supports the induction of a stable RNA chain-reversal conformation as a result of MBNL1 binding to two distant GC/GCU motifs. In support of this hypothesis, mutations that induce stem-loop formation by residues 18–43 upstream of *hcTNT* exon 5 strongly repressed exon 5 inclusion, independently of MBNL1 (ref. 27). The 18-nt *hcTNT* sequence (residues 23–40 upstream of exon 5) contains high-affinity MBNL1-substrate GCU steps separated by a putative pyrimidine tract (Supplementary Fig. 4a), which, following binding by MBNL1, could become less accessible for recognition, thus additionally contributing to splicing repression.

Comparison of RNA-bound MBNL1 and TIS11d complexes

The crystal structure of the MBNL1 ZnF3/4-r(CGCUGU) complex reported in this study can be compared with the previously reported NMR-based solution structure of the TIS11d protein composed of tandem CCCH ZnF domains bound to an AU-rich RNA element¹⁹. There are notable similarities in the structure-based sequence alignment of the first and second ZnF domains within the tandem ZnF repeats of MBNL1 ZnF3/4, MBNL1 ZnF1/2 and TIS11d ZnF1/2, as outlined in Figure 6a. Superpositioning of MBNL1 ZnF3 (Glu183 to His204) and TIS11d ZnF1 (Glu157 to His178) domains in the RNA-bound state (Fig. 6a–c) results in a pairwise Cα r.m.s. deviation of 1.2 Å, establishing that individual ZnF domains adopt similar folds in MBNL1 and TIS11d proteins.

The main difference between the structures of the two complexes originates with the linker regions connecting the tandem ZnF domains in each of the proteins. Whereas the linker between MBNL1 ZnF3 and ZnF4, both in the free and RNA-bound state, is highly

structured through its participation in β -sheet formation (Fig. 6a,b), the corresponding linker connecting TIS11d ZnF1 and ZnF2 adopts an extended conformation in the complex, except for a 3_{10} -helix at the beginning of the linker (Fig. 6a,c). As a result, the relative orientations of the tandem ZnFs are dramatically different between the MBNL1 (Fig. 6b) and TIS11d (Fig. 6c) complexes. MBNL1 ZnF3/4, in which the tandem ZnF domains are aligned in an antiparallel orientation such that their RNA binding surfaces face away from each other, cannot bind two GCU recognition motifs unless they are separated by a linker sequence (Fig. 6b). By contrast, the RNA binding surfaces of the tandem ZnF1/2 domains of TIS11d are in register on the same face of the protein (Fig. 6c), thereby facilitating binding to adjacent UAUU sequences within the 9-nt RNA target.

To compare sequence-specific protein-RNA interactions, we superimposed ZnF3 (molecule B) of the MBNL1 ZnF3/4 complex and ZnF1 of the TIS11d ZnF1/2 complex. The G2 and C3 bases bound by MBNL1 ZnF3 (Fig. 6d) seem to occupy positions equivalent to the U8 and A7 bases, respectively, bound by TIS11d ZnF1 (Fig. 6e). The conserved aromatic side chains Tyr170 and Phe176 are involved in stacking interactions with U8 and A7, respectively, in the TIS11d complex (Fig. 6e). The equivalent two conserved residues in the MBNL1 complex are Arg195 and Phe202, and these residues stack in an equivalent manner with G2 and C3, respectively (Fig. 6d). The flanking RNA bases U6 and U9 of the U6-A7-U8-U9 tetranucleotide bound by TIS11d ZnF1 contribute to binding through formation of U6-Phe176-A7 and U8-Tyr170-U9 stacks.

Notably, sequence specificity of MBNL1 for the 5'-GC-3' step (Fig. 6d) and that of TIS11d for the 3'-UA-5' step (Fig. 6e) is achieved by a similar network of intermolecular hydrogen bonds with the RNA bases. Suffice to highlight that G2 in the MBNL1 complex (Fig. 6d) occupies a similar position to U8 in the TIS11d complex (Fig. 6e), and both use overlapping sets of backbone and side chains for RNA recognition. Similarly, an overlapping set of RNA-recognition elements target C3 in the MBNL1 complex (Fig. 6d) and the identically positioned A8 in the TIS11d complex (Fig. 6e). Notably, the RNA chains run in opposite directions relative to the ZnF scaffold in the MBNL1 (Fig. 6d) and TIS11d (Fig. 6e) complexes, implying that sequence specificity of RNA recognition by tandem CCCH ZnF domains seems to depend on the directionality of the bound RNA.

5'-GC-3' versus 3'-UA-5' recognition specificity

Structural similarities of the MBNL1-GC (Fig. 6d) and TIS11d-AU (Fig. 6e) binding pockets suggest that MBNL1 could perhaps bind a substrate containing a 3'-UA-5' instead of a 5'-GC-3' dinucleotide. To test this hypothesis, we have used EMSAs to monitor the binding of ZnF3/4 to the reverse sequence of a *hcTNT* 18-nt RNA substrate in which the two 5'-GC-3' steps are replaced by 3'-UA-5' steps (referred as GC/AU in Figure 5c). We detected no binding of this AU-containing RNA substrate to MBNL1 by EMSA.

We have tested the contribution to RNA binding of conserved acidic MBNL1 residues Glu187 in ZnF3 (Fig. 4d) and Asp223 in ZnF4 (Fig. 4e), which are involved in intermolecular recognition of the 2'-OH group of cytosine. The E187P D223T double mutant, in which the two conserved acidic residues of MBNL1 are replaced by their TIS11d counterparts (Fig. 6a), showed undetectable binding for both unmodified S10 and GC/AU

18-nt substrates (Fig. 5c). Thus, disruption of backbone interactions in the E187P D223T mutant of MBNL1 resulted in a substantial loss in binding affinity for the unmodified S10 substrate, but did not enhance the binding affinity for the GC/AU sequence.

Dual role of MBNL as a splicing repressor or enhancer

Our efforts in this paper have focused on a mechanistic understanding of MBNL1's role as a splicing repressor. Nevertheless, MBNL1, like some other splicing regulators, can act as either a splicing repressor or as an enhancer, depending on gene context. The location of MBNL1 binding sites relative to the regulated exon may be an important determinant of the probable alternative-splicing outcome, as has been shown for the brain-specific splicing regulator NOVA²⁹. One possibility accounting for its enhancer capabilities is that MBNL could target GC steps in the pre-mRNA to bring the 5' splice site and the branch point into close proximity to stimulate splicing, as has been proposed for hnRNP A/B and hnRNP H/F splicing factors³⁰.

METHODS

Protein and RNA preparation

We cloned the PCR-amplified cDNA fragments encoding human MBNL1 ZnF1/2 (9–90) and ZnF3/4 (178–246) into a modified pET28b vector that adds a Ulp1 protease-cleavable His₆-Smt3-tag at the N terminus. The recombinant proteins were expressed in *Escherichia coli* strain BL21-CodonPlus(DE3)-RIL (Stratagene) in LB medium containing 0.1 mM ZnCl₂. The proteins were purified from the soluble fraction by a nickel-chelating affinity column, followed by His₆-Smt3 tag cleavage with Ulp1 protease and additional purification by sequential chromatography on monoQ (to bind His₆-Smt3) and Superdex 75 columns (Amersham). ¹⁵N-labeled MBNL1 ZnF1/2 and ZnF3/4 for NMR studies were expressed as described³¹. Unlabeled and ¹⁵N-labeled proteins were purified in a similar fashion.

RNA oligonucleotides were commercially synthesized (Dharmacon Research), deprotected and purified by anion-exchange chromatography, followed by desalting.

Crystallization and data collection

We crystallized each of two tandem zinc-finger domains of human MBNL1 protein (isoform 2, 370 amino acids; Gen Bank NM_207292) using ZnF1/2 and ZnF3/4 constructs (Fig. 1b). Crystals of MBNL1 ZnF1/2 and ZnF3/4, as well as the complex of ZnF3/4 with bound 6-nt RNA, were grown by hanging drop vapor diffusion. Crystallization conditions were determined with sparse-matrix screens (Hampton Research). ZnF1/2 was crystallized by mixing 1 µl of 1.0 mM protein solution in 0.2 M NaCl, 25 mM Tris-HCl, pH 7.2, and 1 mM DTT with 1 µl of reservoir solution containing 20% (w/v) PEG4000, 20% (v/v) isopropanol and 0.1 M sodium citrate. Free ZnF3/4 was crystallized by mixing 1 µl of 1.2 mM protein solution in the same protein buffer with 1 µl of reservoir solution containing 18% (w/v) PEG3350, 0.01 M magnesium acetate, 0.2 M NaF, and 0.1 M HEPES, pH 6.8. Droplets were equilibrated against 1.0 ml reservoirs at 4 °C. To obtain MBNL1 ZnF3/4-r(CGCUGU) crystals, equal volumes of protein-RNA complex (1:2 ratio; protein concentration was 0.4

mM in 0.1 M NaCl, 25 mM Tris-HCl, pH 7.2, and 1 mM DTT) and reservoir solution (25% (w/v) PEG3350, and 0.1 M Bis-Tris, pH 5.5) were mixed and equilibrated at 4 °C.

For data collection, crystals were flash frozen (100 K) in the above reservoir solutions supplemented with 20% (v/v) ethylene glycol. SAD data sets were collected on the MBNL1 ZnF1/2 crystals (2.7- Å resolution) and on the MBNL1 ZnF3/4 crystals (1.5- Å resolution) at wavelength 1.2827 Å, corresponding to the zinc absorbance edge. A native (1.7- Å resolution) data set was collected on crystals of the MBNL1 ZnF3/4-r(CGCUGU) complex at wavelength 0.9795 Å. All diffraction data were collected on the 24-ID beamline at the Advanced Photon Source (APS). The data were processed by HKL2000 (ref. 32). The crystals of MBNL1 ZnF1/2 belonged to space group $P3_1$, with one protein molecule per asymmetric unit. The crystals of free and RNA-bound MBNL1 ZnF3/4 belonged to space group $P2_1$, with four protein molecules or protein-RNA complexes per asymmetric unit. Crystal and diffraction data characteristics are summarized in Table 1.

Structure determination and refinement

The program SHELEXD³³ was used to locate two and eight zinc sites in the crystallographic asymmetric units of the free ZnF1/2 and ZnF3/4 crystals, respectively. The program SHARP³⁴ was used to calculate SAD phases. The phases were further improved by density modification and solvent flipping, using the SOLOMON program³⁵. The electron-density map was calculated in the CCP4 suite³⁶. Automatic ZnF3/4 protein model building was performed with WARP³⁷. The resulting protein models were completed manually using TURBO-FRODO³⁸. The structure of the MBNL1 ZnF3/4-r(CGCUGU) complex was determined by molecular replacement with MOLREP in the CCP4 suite³⁶, using the free protein structure as the search model. The RNA model was built manually using one strand of an idealized 6-nt RNA duplex. The final models were refined using REFMAC³⁹. All protein residues are in allowed regions of the Ramachandran plot as evaluated using PROCHECK from CCP4. Refinement statistics are given in Table 1 and a portion of the electron-density map corresponding to the GCU segment of the complex shown in Supplementary Fig. 1.

Gel electrophoretic mobility shift binding assays

Protein-RNA binding interactions were evaluated by EMSA. MBNL1 ZnF3/4 and RNA were mixed in 50 mM Tris-acetate, 10 mM potassium acetate, 50 mM NaCl, 0.1 mM zinc acetate, 10% (v/v) glycerol and 1 mM DTT, pH 8.0, for a final sample volume of 10 µl. The concentration of RNA was 50 µM, whereas that of the MBNL1 ZnF3/4 construct ranged from 0 µM to 250 µM. After a 20-min incubation at 20 °C, samples were separated by 9% nondenaturing PAGE carried out at a constant voltage of 6 V cm⁻¹ at 4 °C in 50 mM Tris-acetate, pH 8.0, containing 10 mM potassium acetate and 0.1 mM zinc acetate. RNA and protein were visualized with Toluidine and Coomassie, respectively.

Fluorescence measurements

The MBNL1 ZnF3/4 protein contains conserved tyrosines at positions 188, 224 and 236 (Fig. 1c), rendering it suitable for fluorescence studies⁴⁰. Tyr188 and Tyr224 stack against the side chains of zinc-coordinated His204 and His238, and Tyr236 intercalates between the

guanine and the cytosine bases in the complex. Steady-state intrinsic tyrosine-fluorescence emission spectra of the MBNL1 ZnF3/4 protein were recorded over the wavelength range 295–400 nm, with an excitation wavelength of 280 nm on a Fluoromax-2 spectrofluorometer. Titrations of the protein at a concentration of 0.3 μ M in buffer containing 50 mM NaCl, 25 mM Tris-HCl, pH 7.2, and 1 mM DTT with RNA oligonucleotides were performed while monitoring the change of the tyrosine fluorescence of MBNL1 ZnF3/4 at 310 nm. The temperature of the cell compartment was kept at 20 ± 0.1 °C. All spectra were corrected for blank solutions. The spectral bandpass was 5 nm. No inner filter corrections were needed because the contribution to the absorbance at 280 nm by the oligonucleotides was below 0.1 at the highest concentration used.

The apparent dissociation constant K_d and the fluorescence intensity f at saturating concentrations of RNA were determined graphically by plotting the relative fluorescence intensity I/I_o versus the concentration of RNA⁴¹. The theoretical curves were fitted to the experimental data points by nonlinear least-squares analysis using the equation

$$I/I_o = 1 + (f-1) \frac{(P_o + N_o + K_D) - \sqrt{(P_o + N_o + K_D)^2 - 4P_o N_o}}{2P_o} \quad (1)$$

where I_o is the fluorescence intensity of the RNA-free protein and N_o and P_o are total RNA and protein concentrations, respectively. Fluorescence measurements were repeated to verify the dissociation constants.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the US National Institutes of Health grant CA049982 to D.J.P. We thank S. Ilin for help with NMR data collection and O. Rechkoblit for help with X-ray synchrotron data collection of one of the data sets. We would like to thank the staff of NE-CAT beam line at the Advanced Photon Source, Argonne National Laboratory, supported by the US Department of Energy, for assistance with data collection.

References

1. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem.* 2003; 72:291–336. [PubMed: 12626338]
2. Pascual M, Vicente M, Monferrer L, Artero R. The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation.* 2006; 74:65–80. [PubMed: 16533306]
3. Ho TH, et al. Muscleblind proteins regulate alternative splicing. *EMBO J.* 2004; 23:3103–3112. [PubMed: 15257297]
4. Lin X, et al. Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum Mol Genet.* 2006; 15:2087–2097. [PubMed: 16717059]
5. Osborne RJ, Thornton CA. RNA-dominant diseases. *Hum Mol Genet.* 2006; 15:R162–R169. [PubMed: 16987879]
6. Ranum LP, Cooper TA. RNA-mediated neuromuscular disorders. *Annu Rev Neurosci.* 2006; 29:259–277. [PubMed: 16776586]

7. Brook JD, et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*. 1992; 68:799–808. [PubMed: 1310900]
8. Liquori CL, et al. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science*. 2001; 293:864–867. [PubMed: 11486088]
9. Fardaei M, et al. Three proteins, MBNL, MBLL and MBXL, co-localize *in vivo* with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum Mol Genet*. 2002; 11:805–814. [PubMed: 11929853]
10. Kino Y, et al. Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum Mol Genet*. 2004; 13:495–507. [PubMed: 14722159]
11. Miller JW, et al. Recruitment of human muscleblind proteins to (CUG)_n expansions associated with myotonic dystrophy. *EMBO J*. 2000; 19:4439–4448. [PubMed: 10970838]
12. Kanadia RN, et al. A muscleblind knockout model for myotonic dystrophy. *Science*. 2003; 302:1978–1980. [PubMed: 14671308]
13. Kanadia RN, et al. Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. *Proc Natl Acad Sci USA*. 2006; 103:11748–11753. [PubMed: 16864772]
14. Paul S, et al. Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. *EMBO J*. 2006; 25:4271–4283. [PubMed: 16946708]
15. Philips AV, Timchenko LT, Cooper TA. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science*. 1998; 280:737–741. [PubMed: 9563950]
16. Savkur RS, Philips AV, Cooper TA. Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat Genet*. 2001; 29:40–47. [PubMed: 11528389]
17. Squillace RM, Chenault DM, Wang EH. Inhibition of muscle differentiation by the novel muscleblind-related protein CHCR. *Dev Biol*. 2002; 250:218–230. [PubMed: 12297108]
18. Adereth Y, Dammai V, Kose N, Li R, Hsu T. RNA-dependent integrin $\alpha 3$ protein localization regulated by the Muscleblind-like protein MLP1. *Nat Cell Biol*. 2005; 7:1240–1247. [PubMed: 16273094]
19. Hudson BP, Martinez-Yamout MA, Dyson HJ, Wright PE. Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat Struct Mol Biol*. 2004; 11:257–264. [PubMed: 14981510]
20. Varani L, et al. The NMR structure of the 38 kDa U1A protein – PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nat Struct Biol*. 2000; 7:329–335. [PubMed: 10742179]
21. Kielkopf CL, Rodionova NA, Green MR, Burley SK. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell*. 2001; 106:595–605. [PubMed: 11551507]
22. Selenko P, et al. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell*. 2003; 11:965–976. [PubMed: 12718882]
23. D'Souza V, Summers MF. Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature*. 2004; 431:586–590. [PubMed: 15457265]
24. De Guzman RN, et al. Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ -RNA recognition element. *Science*. 1998; 279:384–388. [PubMed: 9430589]
25. Lee BM, et al. Induced fit and “lock and key” recognition of 5S RNA by zinc fingers of transcription factor IIIA. *J Mol Biol*. 2006; 357:275–291. [PubMed: 16405997]
26. Lu D, Searles MA, Klug A. Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature*. 2003; 426:96–100. [PubMed: 14603324]
27. Warf MB, Berglund JA. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA*. 2007; 13:2238–2251. [PubMed: 17942744]
28. Oberstrass FC, et al. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*. 2005; 309:2054–2057. [PubMed: 16179478]

29. Ule J, et al. An RNA map predicting Nova-dependent splicing regulation. *Nature*. 2006; 444:580–586. [PubMed: 17065982]
30. Martinez-Contreras R, et al. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol*. 2006; 4:e21. [PubMed: 16396608]
31. Marley J, Lu M, Bracken C. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR*. 2001; 20:71–75. [PubMed: 11430757]
32. Otwinowski Z, Minor W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 1997; 276:307–326.
33. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr*. 2002; 58:1772–1779. [PubMed: 12351820]
34. de La Fortelle E, Bricogne G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous and multiwavelength anomalous diffraction methods. *Methods Enzymol*. 1997; 276:472–494.
35. Abrahams JP, Leslie AG. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr*. 1996; 52:30–42. [PubMed: 15299723]
36. Collaborative Computational Project, Number 4. Collaborative computational project number 4. The CCP4 suite: programmes for protein crystallography. *Acta Crystallogr D Biol Crystallogr*. 1994; 50:760–763. [PubMed: 15299374]
37. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol*. 1999; 6:458–463. [PubMed: 10331874]
38. Cambillau, C.; Roussel, A. Turbo Frodo, Version OpenGL.1. 1997. (<http://www.afmb.univ-mrs.fr/-TURBO->)
39. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 1997; 53:240–255. [PubMed: 15299926]
40. Poveda JA, Prieto M, Encinar JA, Gonzalez-Ros JM, Mateo CR. Intrinsic tyrosine fluorescence as a tool to study the interaction of the shaker B “ball” peptide with anionic membranes. *Biochemistry*. 2003; 42:7124–7132. [PubMed: 12795608]
41. Kurganov BI, Sugrobova NP, Yakovlev VA. Estimation of dissociation constant of enzyme-ligand complex from fluorometric data by “difference” method. *FEBS Lett*. 1972; 19:308–310. [PubMed: 11946238]

Figure 1.

Domain architecture, sequence alignment and structure of the tandem ZnF domains of human MBNL1. (a) Schematic of four ZnF domains and intervening linkers in human MBNL1. Boxed numbers indicate the length of linker segments. (b) Domain boundaries of the MBNL1 ZnF1/2 and ZnF3/4 constructs used in this study. (c) Sequence alignment of ZnF1/2 and ZnF3/4 tandem zinc finger domains from human (Hs; GenBank BAA24858) and *Gallus gallus* (Gg; GenBank CAG31624.1) MBNL1, *Drosophila melanogaster* ZnF1/2 (Dm; GenBank AAC01949.1) and *Caenorhabditis elegans* Mbl (Ce; GenBank NP_510746.1). Cysteine and histidine residues coordinated to zinc are highlighted in blue. Secondary-structure alignment of human MBNL1 ZnF1/2 and ZnF3/4 is shown above the sequences. Numbering above and below the sequences corresponds to human MBNL1 ZnF1/2 and ZnF3/4, respectively. Asterisks denote residues that form hydrogen bonds with RNA bases via their backbone functional groups; triangles indicate aromatic and arginine residues involved in base stacking with RNA bases; dots mark acidic residues interacting with RNA. (d) Overall fold of tandem ZnF3/4 domain. Ribbon view of the face of the β -sheet plane. Cysteine and histidine side chains coordinated to zinc atoms are numbered and

shown in ball-and-stick representation. **(e)** The close-up view of the ZnF3/4 β -sheet in a similar orientation as in **d** in ball-and-stick representation. Hydrogen-bonding interactions between backbone groups are marked as broken gray lines, and those mediated by protein side chains are in green. Water molecule-mediating hydrogen bonds are shown as red spheres.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2.

Crystal packing interactions of the MBNL1 ZnF3/4 in the free state and bound to the r(C1-G2-C3-U4-G5-U6) sequence. **(a)** Ribbon representation of the two MBNL1 ZnF3/4 domains. Molecules A (blue) and B (green) interact via their ZnF4 domains and constitute one half of the crystallographic asymmetric unit of the free protein crystal. Coordinated zinc atoms are shown as purple spheres. **(b)** Ribbon and stick representation of the complex containing four RNA molecules bound to two MBNL1 ZnF3/4 domains. The protein molecules A (blue) and B (green) and the RNA molecules E (pink) and F (light orange) constitute one half of the crystallographic asymmetric unit of the crystal. RNA molecule F interacts with ZnF3 of molecule B. RNA molecule E and the 5'-C1 of symmetry-related RNA molecule F_s interact with ZnF3 of molecule A. Symmetry-related RNA molecule E_s interacts with ZnF4 of molecule A. **(c)** Close-up view of the ZnF4-ZnF4 interface between molecule B (green) and molecule A (light blue) in the crystal of the complex. **(d)** The same view of the ZnF4-RNA interface between molecule A (light blue) and molecule E_s (pink) in the crystal of the complex. RNA and side chains of residues involved in interfacial contacts in **c** and **d** are shown in stick representation.

Figure 3.

Protein-RNA interactions in the molecule A complex. **(a)** Ribbon and stick representation of the complex containing one MBNL1 ZnF3/4 domain (molecule A) and three RNAs bound by ZnF3 and ZnF4. The ZnF3/4 molecule A is colored light blue. Cysteine and histidine residues coordinating zinc atoms are shown in ball-and-stick representation. The two symmetry-related RNA strands are colored pink and the third RNA strand is colored light orange. The G5 and 5'-C1 nucleotides originating from two RNA strands are bound to ZnF3, whereas the G2-C3-U4 segment of the third strand is bound to the ZnF4 motif. **(b)** An electrostatic view of the molecule A complex generated using the GRASP and PyMol programs. Basic and acidic regions appear in blue and red, with the intensity of the color

proportional to the local potential. (c) View of the protein-RNA interface highlighting intermolecular contacts between G2-C3-U4 segment and MBNL1 ZnF4. Stacking interactions of G2 and C3 bases with arginine and aromatic residues, as well as hydrogen-bonding contacts involving G2, C3 and U4, are highlighted. (d) Electrostatic surface representation of the protein-RNA interface in the same view as in c.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.

Details of intermolecular protein-RNA recognition contacts in the complex. (a) Hydrogen-bonding of the N² and N¹ of G5 with the side chain thiols of zinc-bound Cys185 and Cys200 and the O⁶ of G5 with the backbone amide of Arg201 of the molecule A ZnF3 motif. G5 stacks with Arg195 of the ZnF3 motif. Arg195 side chain N^{η1} is hydrogen-bonded to the G5 O^{5'} and U4 O^{2'} of the sugar-phosphate backbone linking the U4-G5 step. (b) Hydrogen-bonding of G2 with the side chain and backbone of molecule A ZnF4 motif. (c) Hydrogen-bonding of G2 with the side chain and backbone of molecule B ZnF3 motif. (d) Hydrogen-bonding of the O², N³ and N⁴ of 5'-C1 with the backbone amides of Glu187 and Arg186, and carbonyl of Val184 of the molecule A ZnF3 motif. C1 O² is also hydrogen-bonded to the side chain carboxyl of Glu183. C1 stacks with Phe202 and Arg186 side chains. The 2'-OH of C1 is hydrogen-bonded to the side chain of Glu187, and the two oxygens of the phosphate linking the C1-G2 step are hydrogen-bonded to the side chains of Arg186 and Arg190. (e) Hydrogen-bonding of C3 with the backbone of molecule A ZnF4 motif. (f) Hydrogen-bonding of C3 with the backbone of molecule B ZnF3 motif.

Figure 5.

Binding of GCU-containing RNAs to MBNL1 ZnF3/4. **(a)** EMSA data for binding of r(UCGCUUUUCCCCUCCGCU) to MBNL1 ZnF3/4, establishing a 1:1 stoichiometry of the complex. **(b)** Plots of relative fluorescence intensity for MBNL1 ZnF3/4 as a function of oligonucleotide concentration. Black, red and blue dots represent mean $\pm$ s.d. for at least two independent measurements for 18-nt r(UCGCUUUUCCCCUCCGCU), 6-nt r(CGCGUGU) and 4-nt r(UGCU) oligonucleotides, respectively. The GC steps in the sequences are underlined. The solid line for the 18-nt RNA indicates nonlinear least-squares fit according to equation (1) in Methods. **(c)** EMSA data for binding of S10, S5, S15 and GC/AU to wild-type (WT) ZnF3/4 and for binding of S10 and GC/AU to the E187P D223T mutant, in which Glu187 and Asp223 have been replaced by the residues that occupy equivalent positions in TIS11d. The original *hcTNT* 18-nt RNA fragment (S10) and modified sequences (S5 and S15) contain 10-nt, 5-nt and 15-nt spacers, respectively, between the two GCU cites. The spacer regions in the sequences are underlined. GC/AU is the reverse of S10, with 5'-GC-3' to 3'-AU-5' replacements. **(d)** Plots of relative fluorescence intensity for MBNL1 ZnF3/4 as a function of oligonucleotide concentration. Black, blue and red dots represent mean $\pm$ s.d. for at least two independent measurements for S10, S5 and S15, respectively. Solid lines indicate nonlinear least-squares fit according to equation (1) in Methods. Apparent dissociation constants determined by fluorescence quenching are listed together with $\pm$ fitting error.

Figure 6.

Comparison of MBNL1 ZnF3/4 and TIS11d¹⁹ ZnF1/2 complexes with their respective RNA targets. **(a)** Structure-based sequence alignment of MBNL1 and TIS11d (PDB 1RGO) tandem ZnF domains. The numbering listed above and below sequences refers to the MBNL1 ZnF3/4 and TIS11d ZnF1/2, respectively. Cysteines and histidines coordinated to zinc are highlighted in blue. Residues that form hydrogen bonds with RNA bases via their backbone functional groups in both structures are indicated by asterisks; those involved in base-stacking with RNA bases are indicated by triangles; common acidic residues interacting with the RNA are indicated by dots. **(b)** Overall structure of MBNL1 ZnF3/4–r(CGCUGU) complex (molecule A). Protein and RNA molecules are shown in ribbon representation. G5 and C1 nucleotides, which interact with ZnF3, and G2–C3–U4 segment, which interacts with ZnF4, are in stick representation. **(c)** Ribbon representation of the first structure from the 20 solution structures (PDB 1RGO) of TIS11d–RNA complex. A3, U4, A7 and U8 nucleotides, which occupy positions equivalent to C3, G2, C1 and G5, respectively, in **b** are shown in stick representation. The views in **b** and **c** are positioned for comparison following superposition of MBNL1 ZnF3 and TIS11d ZnF1 domains. Protein

and RNA are colored light blue and pink, respectively. The linker segment connecting the two ZnF domains in **b** and **c** is shown in green. **(d)** Protein-RNA interface highlighting intermolecular contacts between G2-C3 and the MBNL1 ZnF3 in molecule B. Stacking interactions between bases and arginine and aromatic residues, as well as hydrogen-bonding contacts involving G2 and C3, are highlighted. **(e)** Protein-RNA interface highlighting intermolecular contacts between A7-U8 and the TIS11d ZnF1. The views in **d** and **e** are positioned for comparison following superposition of MBNL1 ZnF3 (molecule B) and TIS11d ZnF1 domains.

Table 1

Data collection and refinement statistics

	Free form		Complex
	ZnF1/2	ZnF3/4	ZnF3/4
Data collection	SAD	SAD	Native
Space group	$P3_1$	$P2_1$	$P2_1$
Cell dimensions			
a, b, c (Å)	49.3, 49.3, 36.0	29.2, 80.5, 55.4	56.4, 56.9, 58.2
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 101.8, 90.0	90.0, 108.8, 90.0
Resolution (Å)	50–2.7	50–1.5	50–1.7
R_{sym}	6.0 (27.4) ^a	8.9 (48.2)	7.4 (44.1)
$I / \sigma I$	21.3 (4.6)	17.8 (3.6)	19.8 (3.4)
Completeness (%)	98.7	97.7	99.9
Redundancy	2.9	3.2	4.9
Refinement			
Resolution (Å)	20–2.7	20–1.5	20–1.7
No. reflections	2,553	36,934	36,684
$R_{\text{work}} / R_{\text{free}}$	23.0 / 27.9	20.2 / 22.5	22.4 / 26.0
No. residues			
Protein	76	257	270
RNA/Zn	–/2	–/8	22/8
Water	6	316	406
<i>B</i> -factors			
Protein	49.4	16.8	16.6
RNA/Zn	–/54.0	–/18.8	20.4/11.1
Water	52.3	23.5	34.4
R.m.s. deviations			
Bond lengths (Å)	0.009	0.009	0.010
Bond angles (°)	1.6	1.2	1.6

^aValues in parentheses are for highest-resolution shell.