

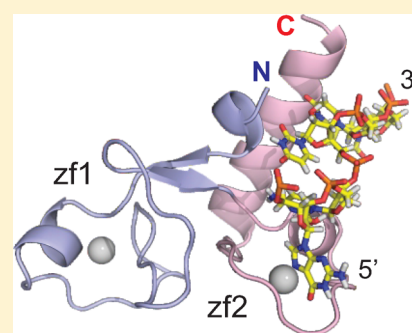
# Structural Basis for Interaction of the Tandem Zinc Finger Domains of Human Muscleblind with Cognate RNA from Human Cardiac Troponin T

Sangho Park, Priti Deka Phukan,<sup>†</sup> Markus Zeeb,<sup>‡</sup> Maria A. Martinez-Yamout, H. Jane Dyson,<sup>§</sup> and Peter E. Wright<sup>\*</sup>

Department of Integrative Structural and Computational Biology and Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States

## Supporting Information

**ABSTRACT:** The human muscleblind-like proteins (MBNL) regulate tissue-specific splicing by targeting cardiac troponin T and other pre-mRNAs; aberrant targeting of CUG and CCUG repeat expansions frequently accompanies the neuromuscular disease myotonic dystrophy. We show, using bilayer interferometry (Octet) and NMR spectroscopy, that the zinc finger domains of MBNL isoform 1 (MBNL1) are necessary and sufficient for binding CGCU sequences within the pre-mRNA of human cardiac troponin T. Protein constructs containing zinc fingers 1 and 2 (zf12) and zinc fingers 3 and 4 (zf34) of MBNL1 each fold into a compact globular tandem zinc finger structure that participates in RNA binding. NMR spectra show that the stoichiometry of the interaction between zf12 or zf34 and the CGCU sequence is 1:1, and that the RNA is single-stranded in the complex. The individual zinc fingers within zf12 or zf34 are nonequivalent: the primary RNA binding surface is formed in each pair by the second zinc finger (zf2 or zf4), which interacts with the CGCU RNA sequence. The NMR structure of the complex between zf12 and a 15-base RNA of sequence <sub>95</sub>GUCUCGCUUUUCCCC<sub>109</sub>, containing a single CGCU element, shows the single-stranded RNA wrapped around zf2 and extending to bind to the C-terminal helix. Bases C101, U102, and U103 make well-defined and highly ordered contacts with the protein, whereas neighboring bases are less well-ordered in the complex. Binding of the MBNL zinc fingers to cardiac troponin T pre-mRNA is specific and relatively simple, unlike the complex multiple dimer–trimer stoichiometries postulated in some previous studies.



## INTRODUCTION

The muscleblind-like (MBNL) proteins are a subfamily of CCCH zinc finger proteins that function as developmentally programmed regulators of alternative pre-mRNA splicing. Muscleblind was originally identified in *Drosophila*, where it regulates the terminal differentiation of photoreceptors and muscles.<sup>1,2</sup> Muscleblind proteins are exclusive to metazoans and are found in all genomes from *Caenorhabditis elegans* to *Homo sapiens*.<sup>3,4</sup> Vertebrate genomes contain 3 muscleblind homologues: MBNL1 (also named MBNL/EXP), MBNL2 (MBLL/MLP1), and MBNL3 (MBXL/CHCR). Each of these MBNL homologues contain two distantly spaced pairs of CCCH zinc fingers (tandem zinc fingers, TZF) that are absolutely required for RNA binding.<sup>3,5</sup> Fingers 1 and 3 have zinc ligand spacing CX<sub>7</sub>CX<sub>6</sub>CX<sub>3</sub>H, whereas fingers 2 and 4 are CX<sub>7</sub>CX<sub>4</sub>CX<sub>3</sub>H motifs; fingers 1 and 2 are separated by 14 residues, and fingers 3 and 4 are separated by 16 residues.<sup>6–8</sup> In vertebrate muscleblind, the two TZF domains, zf12 and zf34, are separated by an ~80 residue linker (Figure 1A) that has a biased amino acid composition and low sequence complexity (24% alanine, 18% proline) and is therefore likely disordered. The muscleblind tandem zinc finger sequences and their linkers are strongly conserved from *C. elegans* to humans (Figure 1B).

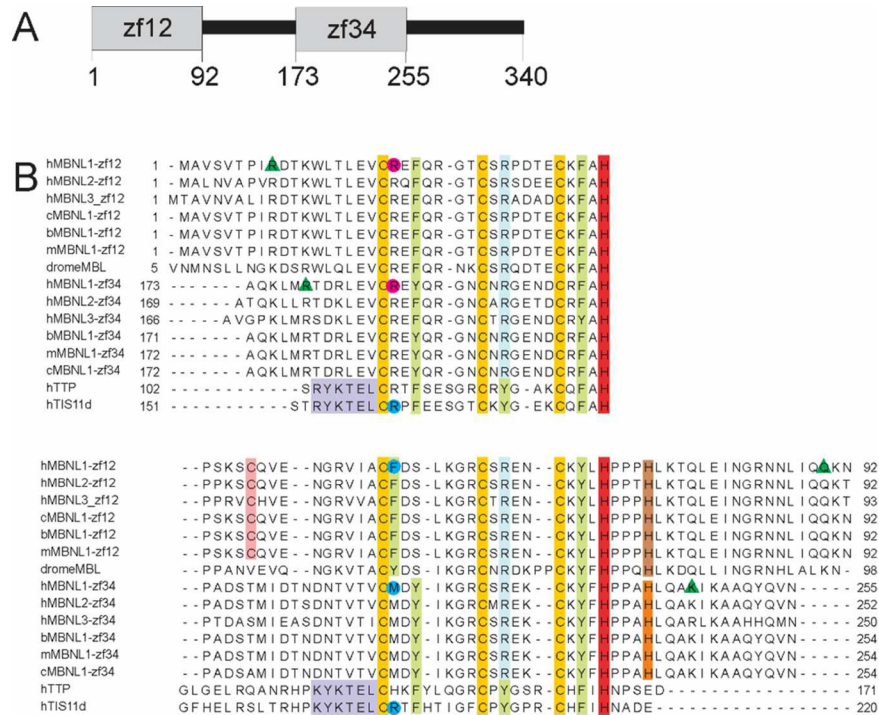
Human MBNL is a member of a large family of proteins containing tandem CCCH zinc finger domains that plays an important role in RNA metabolism. The best characterized CCCH zinc finger proteins are those of the TIS11 family, which control gene expression by modulation of mRNA stability. The TIS11 proteins bind via their zinc finger domains to a specific regulatory element, the AU-rich element (ARE) in the 3'-untranslated region (UTR) of short-lived cytokine, growth factor, and proto-oncogene mRNAs.<sup>9,10</sup> The structure of the tandem zinc finger domain of TIS11d has been determined in complex with its RNA recognition element.<sup>11</sup>

In concert with CUG-BP, Elav-like family (CELF) proteins, muscleblind functions as a key developmental switch that controls splicing of several genes including human cardiac troponin T (hcTnT), insulin receptor (IR), and muscle-specific chloride ion channel (CIC-1).<sup>12–15</sup> MBNL proteins bind directly to the hcTnT pre-mRNA at a site immediately adjacent to the alternative exon 5;<sup>14</sup> binding appears to involve recognition of the consensus RNA motif YGCY, where Y =

Received: May 19, 2017

Revised: July 14, 2017

Published: July 18, 2017



**Figure 1.** (A) Domain organization of human muscleblind (MBNL) zinc finger domains showing domain boundaries for the zf12 and zf34 constructs used in the present work. (B) Sequence alignment of the two tandem zinc fingers of the zinc finger pairs zf12 and zf34 for the three human (h) isoforms of MBNL (MBNL1, MBNL2, MBNL3) with MBNL proteins from chicken (c), bovine (b), mouse (m), and the MBL protein of *Drosophila* (drome) (zf12 only) and with prototype tandem CCCH zinc finger proteins human tristetraprolin (TTP) and TIS11d. The zinc ligands are outlined in yellow (cysteine) and red (histidine). Conserved aromatic residues are outlined in green, and arginines conserved in MBNL and replaced by tyrosine in TTP and TIS11d are outlined in blue. The characteristic “RYKTEL/KYKTEL” sequences of the latter two proteins are outlined in purple. Pink, brown, and orange outlines show conserved cysteine and histidine residues that we show by mutagenesis are not involved in zinc ligation. Blue circles indicate residues from zf2 or zf4 that show large downfield shifts in <sup>1</sup>H–<sup>15</sup>N HSQC spectra of MBNL1 (this work) or TIS11d<sup>11</sup> upon addition of RNA; red circles indicate the corresponding residues from zf1 and zf3 that disappear upon the addition of RNA. Green triangles indicate the start and end of the shorter constructs used in the published crystal structures.<sup>25</sup>

pyrimidine.<sup>16–19</sup> Berglund and co-workers have suggested that MBNL1 regulates the exclusion of exon 5 in the hcTnT mRNA by binding at the 3'-end of intron 4 in hcTnT pre-mRNA. If MBNL1 is absent, the CELF proteins and splicing machinery recognize this site, and the exon is included.<sup>20</sup> In *Drosophila*, muscleblind plays a direct role in regulating alternate splicing of the  $\alpha$ -actinin pre-mRNA.<sup>21</sup>

In addition to binding pre-mRNA, muscleblind binds tightly to expanded pathogenic CUG and CCUG repeats that accumulate in nuclear foci in myotonic dystrophy.<sup>5–7,22</sup> It has been proposed that, under pathogenic conditions, muscleblind proteins are aberrantly sequestered in the nuclear foci, thereby interfering with MBNL regulated pre-mRNA splicing activity, disrupting normal cellular developmental programs and resulting in myotonic dystrophy.<sup>14</sup> The muscleblind zinc finger domains are necessary and sufficient for high affinity binding to double-stranded CUG and CCUG repeat RNA. Deletion of the entire C-terminal region of human muscleblind has no effect on binding,<sup>5</sup> whereas deletion of zinc fingers from natural muscleblind isoforms results in abrogation of binding to RNA.<sup>6,22</sup> Isoforms lacking the long linker between fingers 2 and 3 are also active in RNA binding.<sup>6</sup> In contrast to the human protein, *Drosophila* muscleblind contains only one zinc finger pair, which is highly homologous to human fingers 1 and 2; however, as with human MBNL, CUG repeat transcripts sequester *Drosophila* muscleblind and interfere with its splicing activity.<sup>21</sup> Thus, a single pair of zinc fingers appears to be sufficient to impart the ability to bind both to cognate single-

stranded RNA and CUG repeat RNA. Mutagenesis studies have shown that the zinc finger pairs of MBNL1 are not equivalent with zf12 exhibiting higher RNA binding affinity and splicing activity than those of zf34.<sup>23,24</sup>

In this study, we report the determination of the 3-dimensional NMR structures of the two sets of MBNL1 TZFs, zf12 and zf34, and of the complex of MBNL1 zf12 with an RNA sequence derived from the pre-mRNA of cardiac troponin T. Both zf12 and zf34 adopt a structure in which the two zinc fingers are cooperatively folded to form a single globular unit with a stable hydrophobic core. We have characterized the RNA binding activity of MBNL1 by biolayer interferometry (Octet) and NMR-based titration experiments. In addition, we show through NMR titration and self-diffusion experiments that one pair of zinc fingers binds to one YCGCU(U/G)Y motif in the respective hcTnT pre-mRNA. We demonstrate that the hydrogen-bonded secondary structure, likely a stem-loop structure in a 32-nucleotide sequence at the 3' end of intron-4 in hcTnT, unfolds to single-stranded RNA as a consequence of MBNL1 zinc finger binding. We have calculated the NMR solution structure of the complex of this single-stranded RNA with zf12. During the course of this work, three-dimensional structures of free MBNL1<sup>25</sup> and MBNL2<sup>26</sup> zinc finger domains were reported. A crystal structure analysis of MBNL1 zf34 bound to RNA<sup>25</sup> used MBNL zinc finger fragments and RNA molecules that are significantly shorter than those reported here, which have been shown to compromise RNA binding.<sup>27</sup> Together with the

structural analysis of the free two-finger domains, we present both a comprehensive mapping of the RNA binding surfaces of the MBNL1 zinc finger domains and a solution structure of zf12 in complex with RNA. These results show that the preferred stoichiometry of the interaction between a zinc finger pair (zf12 or zf34) and a YGCY RNA motif is 1:1.

## ■ EXPERIMENTAL PROCEDURES

**Preparation of MBNL1 Protein Constructs.** The DNA sequences encoding human MBNL1-zf12 (residues 1–92), MBNL1-zf34 (residues 173–255), and MBNL1-zf1234 (residues 1–255) were isolated by PCR amplification from a full-length human liver cDNA library (Mobictec) and subsequently cloned into pET21a. Protein overexpression was performed at 37 °C in transformed *E. coli* BL21 (DE3) (DNAY) in minimal medium (48 mM Na<sub>2</sub>HPO<sub>4</sub>, 22 mM KH<sub>2</sub>PO<sub>4</sub>, 8.6 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.1 mM CaCl<sub>2</sub>, BME vitamin mix (Sigma), and trace metals) in the presence of 100 mg/mL of carbenicillin and 30 mg/mL of kanamycin. Stable isotope labeling was performed by adding 1 g/L of (<sup>15</sup>NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> as nitrogen source and either 4 g/L of D-glucose (for <sup>15</sup>N-labeled protein) or 2 g/L of <sup>13</sup>C D-glucose (for <sup>13</sup>C, <sup>15</sup>N labeled protein) as carbon source. Cells were grown to A<sub>600</sub> ≈ 0.8 at 37 °C supplemented with 0.150 mM ZnSO<sub>4</sub> and induced with IPTG to a final concentration of 0.5 mM. For expression of MBNL1-zf12 and MBNL1-zf1234, cells were allowed to grow for 4 h after induction at 37 °C; whereas for MBNL1-zf34, cells were grown overnight at 18 °C after induction. Cells were harvested by centrifugation and frozen at –20 °C. Frozen cell pellets were suspended in 20 mM Tris/HCl, 10 mM DTT pH 8.0 and lysed by sonication (Branson Sonifier 450) at 4 °C. MBNL1 protein constructs were mainly expressed in the soluble fraction and purified under native conditions. Cell debris was removed by centrifugation, and a streptomycin cut was performed to remove nucleic acids: 10 mL of a freshly prepared 10% solution of streptomycin sulfate was added to 100 mL of lysis supernatant, and the solution was stirred at room temperature for 30 min. The suspension was clarified by centrifugation, and the supernatant was dialyzed against 4 L 20 mM Tris pH 8, 0.5 mM DTT overnight at 4 °C. The supernatant was passed through a 5 mL HiTrap Q column to separate the protein of interest from remaining nucleic acids and other impurities and applied onto a 75 mL Sepharose SP FF column pre-equilibrated with 20 mM Tris/HCl, 50 mM NaCl, 2 mM DTT pH 8.0. Finally, fractions containing MBNL1 protein were pooled, concentrated, and purified on a Superdex 75 XK16 gel filtration column (20 mM Tris/HCl, 100 mM NaCl, 2 mM DTT pH 7.0). SDS-PAGE, analytical HPLC, and MALDI mass spectrometry were used to verify identity, purity, and isotope enrichment. Mutant proteins were produced following the QuikChange mutagenesis protocol. All protein variants were prepared and purified using the same protocol.

For all three protein constructs, MBNL1-zf12, MBNL1-zf34, and MBNL1-zf1234, samples were prepared for NMR by exchange into NMR buffer (20 mM Tris (pH 7.0), 50 mM NaCl, 2 mM DTT) All NMR experiments were performed at protein concentrations of 0.1 mM (for HSQC titration experiments and NMR diffusion experiments) and 0.4 mM (for assignments and structure determination).

**Preparation of MBNL1 Protein–RNA Complexes.** RNAs were purchased as PAGE-purified oligonucleotides from IDT DNA (San Diego). Samples were dissolved in water, desalted, and exchanged into NMR buffer (20 mM Tris,

pH 7.0, 50 mM NaCl, 2 mM DTT) with a NAP column. The concentration was determined by UV spectroscopy using absorbance at 260 nm. The product was freeze-dried until use. Before preparing the protein–RNA complex, RNA was dissolved in NMR buffer to a final concentration of ~2 mM, heated to 90 °C for 3 min, and then snap-cooled in ice for 10–15 min. The concentrated RNA was then added to fresh protein to prepare the complex.

**Biolayer Interferometry.** BLI experiments were carried out on an OCTET RED 96 instrument with streptavidin sensors. 5′ Biotinylated RNAs (purchased from IDT) were suspended in 20 mM Tris pH 7.2, 150 mM NaCl, and then RNA and protein stocks were further buffer exchanged on a NAP desalting column in the same buffer. Protein and RNA concentrations were determined by UV, and stocks were kept frozen at –80 °C. BLI experiments were carried out at 28 °C. Sensors were loaded at 0.5 μM RNA concentration, and kinetic experiments were performed at 40–200 nM protein concentration in a final buffer consisting of 20 mM Tris pH 7.2, 300 mM NaCl, 1 mM DTT, 0.2 mg/mL of BSA. Binding curves were fit to a 1:1 binding model. Several repeat runs were performed with each sensor, and experiments were performed at least in triplicate with different sets of sensors. Average K<sub>d</sub> values are reported.

**NMR Spectroscopy.** NMR spectra were recorded at 25 °C on Bruker Avance 500, 600, 800, and 900 MHz spectrometers. 1D spectra were processed using Bruker XWINNMR or topspin software, whereas multidimensional spectra were processed with NMRPipe/NMRDraw software.<sup>28</sup> The processed NMR spectra were analyzed using Sparky<sup>29</sup> or NMRView.<sup>30</sup> Backbone resonances were assigned using 3D HNCA, HNCACB, CBCA(CO)NH, and HNCO experiments<sup>31</sup> and correlated back to <sup>15</sup>N HSQC and <sup>13</sup>C CT-HSQC spectra. Dihedral angle restraints  $\phi$  and  $\psi$  were calculated with TALOS<sup>32,33</sup> using backbone chemical shifts (<sup>13</sup>CO, <sup>13</sup>C $\alpha$ , <sup>1</sup>H $\alpha$ , <sup>1</sup>HN, <sup>15</sup>NH). In addition,  $\phi$  was experimentally determined using coupling constants derived from an HNHA spectrum.<sup>34</sup> Aliphatic side chain assignments were compiled from (H)C(CO)NH-TOCSY, H(C)(CO)NH-TOCSY, and HCCH–COSY spectra,<sup>31</sup> and aromatic side chains were assigned by (HB)CB(CGCD)HD and (HB)CB-(CGDCE)HE experiments.<sup>35</sup> Distance restraints were obtained from a <sup>15</sup>N-edited NOESY-HSQC spectrum collected in water at 600 MHz with 175 ms mixing time and a <sup>13</sup>C-edited HMQC-NOESY spectrum in D<sub>2</sub>O at 900 MHz with 150 ms mixing time. Coupling constants from HNHB,<sup>36</sup> HACAHB-COSY,<sup>37</sup> <sup>13</sup>C-[<sup>13</sup>CO], and <sup>13</sup>C-[<sup>15</sup>N] spin–echo difference CT-HSQC spectra<sup>38,39</sup> were determined to provide stereospecific assignments for H $\beta$  protons and to generate  $\chi_1$  torsion angle restraints for structure calculations. Zinc-coordinating His residues and their tautomeric state were identified by <sup>1</sup>H–<sup>15</sup>N HMQC experiments<sup>40</sup>

The single-stranded RNA hcTnT15 was not isotopically labeled. The resonances of the portion of the RNA that were bound to the protein were assigned by <sup>1</sup>H 2D methods employing the filtered NOESY spectrum<sup>41</sup> of the complex of hcTnT15 with uniformly <sup>15</sup>N,<sup>13</sup>C-labeled MBNL1 zf12 where only NOEs between <sup>12</sup>C-attached <sup>1</sup>H resonances were detected. Because the protein was uniformly labeled, only the NOEs between the RNA protons were detected in this experiment. The majority of the RNA is unstructured such that the resonances significantly overlapped. Only the resonances of the bound nucleotides are shifted outside this envelope sufficiently

**Table 1. Experimental Restraints and Structure Statistics for the 20 Lowest Energy Structures of MBNL1 Proteins and RNA Complex**

RNA	MBNL1-zf12	MBNL1-zf34	MBNL-zf12/hcTnT15	
			protein	RNA
(A) Distance Constraints				
total unambiguous NOE restraints	1779	1095	1003	6
short-range ( $ i - j  \leq 1$ )	894	544	377	
medium-range ( $1 <  i - j  < 5$ )	333	225	254	
long-range ( $ i - j  \geq 5$ )	552	326	372	
intermolecular				71
total dihedral angle restraints <sup>a</sup>	74	102	128	79
$\Phi$	47	59	68	
$\Psi$	27	43	60	
hydrogen bond constraints	18	18	56	6
(B) Structure Statistics				
violations analysis				
maximum distance violation (Å)	0.12	0.18	0.30	
maximum angle violation (deg.)	0.00	0.84	0.82	
energies				
mean AMBER energy (kcal/mol)	-4825	-4097	-5887	
mean restraint violation energy (kcal/mol)	3.25	9.14	6.65	
average pairwise rmsd <sup>b</sup>				
backbone (Å)	0.55	0.60	0.37	
heavy (Å)	1.32	1.38	0.98	
(C) Ramachandran Statistics <sup>c</sup>				
allowed regions	81.1%	81.1%	83.1%	
additionally allowed regions	18.6%	18.8%	16.3%	
generously allowed regions	0.2%	0.1%	0.6%	
disallowed regions	0.1%	0.0%	0.0%	

<sup>a</sup>Obtained from C<sup>α</sup>, C<sup>β</sup>, and C chemical shifts using TALOS. <sup>b</sup>Pairwise rmsd was calculated among 20 refined structures for structured region residues 10–85 for MBNL1-zf12 free and in the RNA complex and residues 179–251 for MBNL1-zf34. <sup>c</sup>Obtained from PROCHECK.<sup>47</sup>

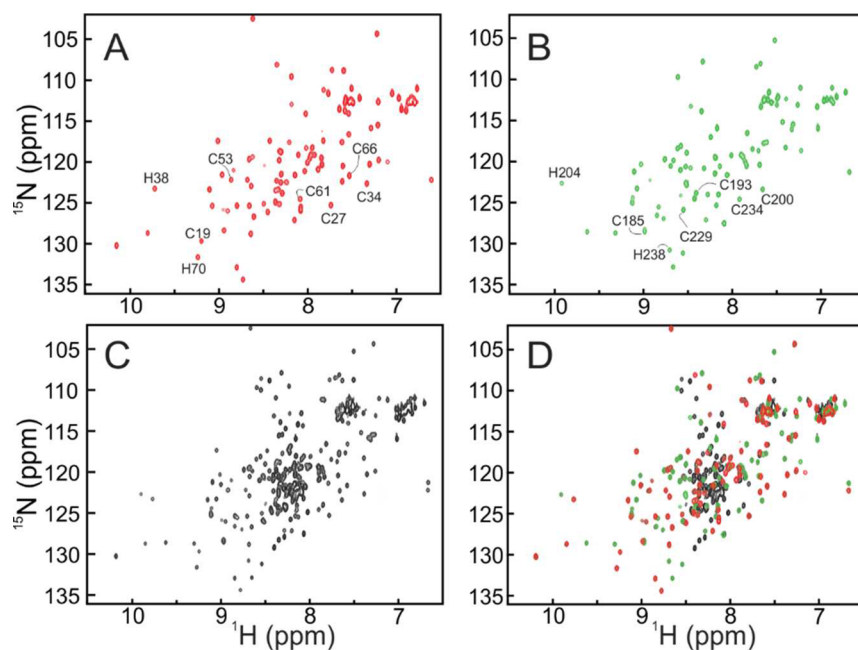
for sequential assignments to be made. Standard NOE connectivities were used to identify nearest-neighbor nucleotides, and the assignments were verified by the correlation of the intermolecular protein–RNA NOEs (listed in Table S1) obtained from an edit-filtered spectrum.<sup>42</sup>

1D diffusion experiments were recorded using the PG-SLED gradient echo pulse program<sup>43</sup> at 298 K on a Bruker 900 MHz spectrometer with 4096 complex data points. For each sample, a set of 12 experiments were recorded with varying gradient strengths; the last 5 experiments were repeated for consistency and error calculations. The length of the pulsed field gradient (PFG) was kept the same, and the amplitude was stepped up from 15 to 70% of the maximum gradient strength in steps of 5%. As the strength of the PFG pulse was increased, the signal amplitude was attenuated due to translational diffusion.

**Structure Calculation.** Protein structures were calculated in a semiautomated iterative manner by using CYANA version 2.1<sup>44</sup> with distance restraints derived from <sup>15</sup>N- and <sup>13</sup>C-edited NOESY spectra.<sup>42</sup> The NOESY peak list used as input for CYANA analysis was generated automatically using Sparky<sup>29</sup> based on the chemical shift list generated in the assignment process. Additional input for initial structure calculations included stereospecifically assigned H<sup>β</sup> and prochiral methyl protons derived from coupling constants as well as zinc ligand restraints. Backbone dihedral angle restraints were obtained from chemical shifts using TALOS and from a <sup>3</sup>J-HNHA experiment. After the first few rounds of calculations, the spectra were analyzed again to identify additional cross-peaks consistent with the structural model and to correct mis-assigned NOE restraints.

Slowly exchanging amides were identified by lyophilizing the protein from H<sub>2</sub>O and then dissolving it in 100% D<sub>2</sub>O; hydrogen bond donors were identified by the presence of an amide peak in the HSQC recorded after 30 min. The corresponding acceptors were identified from CYANA structures calculated without any hydrogen bonding constraint to identify carbonyl groups that were at a distance of approximately 2.0 Å from slowly exchanging amides. Hydrogen bonding constraints were then added at this stage. When the CYANA automated structure calculation gave an average target function of 0.7 Å, the distance and angle constraints were used to generate 100 starting structures by following the REDAC strategy of CYANA2.1. Refinement of these starting structures was accomplished by molecular dynamics calculations using AMBER 8 with reduced charges in vacuo. Zinc coordination was restrained to be tetrahedral, and zinc ions were covalently attached with the ligand residues. Iterative structure refinement was performed by modification of the distance restraint list by identification of ambiguous or incorrect assigned distance restraints using the original NOESY data. The latter was applied for final refinement in AMBER 8<sup>45</sup> incorporating a generalized Born solvation model.<sup>46</sup> The 20 best structures were selected to represent the ensemble of converged MBNL1-zf12 and MBNL1-zf34 structures. Experimental data and structural statistics are summarized in Table 1. The quality of the structure was evaluated using PROCHECK-NMR.<sup>47</sup>

Structure calculation for the complex of MBNL-zf12 with hcTnT15 RNA used a similar method with the RNA modeled as single-stranded. Two-hundred starting structures of MBNL1-zf12 were generated by using CYANA as described above and then were docked to RNA during the refinements using



**Figure 2.**  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectra of MBNL zinc finger constructs (A) MBNL1-zf12 (residues 1–92, red), (B) MBNL1-zf34 (residues 173–255, green), (C) MBNL1-zf1234 (residues 1–255, black), and (D) overlay of the three spectra in (A–C), showing that the cross peaks corresponding to the individual zinc finger pairs exactly superimpose on the corresponding cross peaks for the full-length construct. Cross peaks assigned to the zinc ligand residues are labeled in parts (A) and (B).

AMBER 11.<sup>48</sup> An RNA model was generated by using XLEAP, incorporating loose restraints for the RNA backbone torsion angles ( $-120^\circ$  to  $120^\circ$  for  $\alpha$ ,  $140^\circ$  to  $220^\circ$  for  $\beta$ ,  $20^\circ$  to  $100^\circ$  for  $\gamma$ ,  $55^\circ$  to  $190^\circ$  for  $\delta$ ,  $170^\circ$  to  $300^\circ$  for  $\epsilon$ , and  $-120^\circ$  to  $120^\circ$  for  $\zeta$ ). Because of the paucity of experimental RNA restraints, these torsion angle restraints were added to penalize the formation of energetically unfavorable RNA backbone conformations.<sup>49,50</sup> The modeled RNA was added to each starting protein structure at a random position  $\sim 30$  Å from the protein structure. The docking was achieved through seven cycles of simulated annealing in AMBER 11 using a generalized Born solvent model. Intermolecular restraints were not used for the first cycle while the system was heated to 1000 K for 6 ps and then annealed by cooling to 0 K for 14 ps. For the remaining cycles, intermolecular distant restraints were gradually decreased from 50 to 15 to 10 Å to final values obtained from experiments while the system was heated to 400 K for 6 ps and then annealed by cooling to 0 K for 28 ps. During these cycles, the restraint energy for intermolecular distances was increased from 0 to 1 to 10 to 20 to 30 kcal/mol Å<sup>2</sup>. The majority of intermolecular distances involved protein pseudotoms and long upper distance bounds. As a consequence, it was difficult to define the protein–RNA interface, and many cycles of restrained molecular dynamics were required to generate a converged family of structures that satisfied all of the experimental restraints. The F54 amide cross peak is strongly downfield shifted in the RNA complex, as observed for the equivalent residue in Tis11d, which is shifted by hydrogen bond formation with an adenine base.<sup>11</sup> The experimental NOE restraints were more uniformly satisfied and the structures converged better when an artificial distance restraint was imposed between the C101 N3 and the F54 NH. In the resulting subset of structures that best satisfied the restraints in this region, the N4 and O2 atoms of C101 were consistently within hydrogen binding distance to the A52 CO and D55 NH atoms. These hydrogen bonds were then introduced as

intermolecular restraints in the final cycles of simulated annealing. After the final cycle of restrained simulated annealing with all restraints active, the 20 structures with the lowest AMBER energy were selected to represent the ensemble of the complex of MBNL1-zf12 with hcTnT15. The structural statistics are summarized in Table 1.

## RESULTS

**Domain Structure and Zinc Binding.** Human muscleblind proteins contain four CCCH zinc finger motifs arranged as two tandem zinc fingers (TZFs) separated by an 80 residue linker (Figure 1A). Initial constructs of MBNL1-zf12 and MBNL1-zf34 containing residues 1–84 and 173–249, respectively, were designed on the basis of sequence alignment (Figure 1B) to coincide with the boundaries of the tandem zinc finger domain of the previously characterized prototype CCCH zinc finger protein TIS11d.<sup>11</sup> Constructs of a similar size (residues 9–90 for zf12 and 178–246 for zf34) were also used in the published X-ray crystallographic study.<sup>25</sup> NMR spectra of the 1–84 and 173–249 constructs were of poor quality, and analysis of the chemical shift index<sup>51</sup> and  $[\text{H}]-^{15}\text{N}$  heteronuclear NOE showed the presence of helical structure that extended to the C-terminal residues, 84 in zf2 and 249 in zf4. Extended constructs for MBNL1 zf12 (containing residues 1–92, 1–104, and 1–129) showed better solution behavior. Comparison of the CD and HSQC spectra of these constructs showed that residues beyond 92 were unstructured. For MBNL zf34, the C-terminus was extended to the end of the zinc finger domain (residues 173–255), which gave satisfactory NMR spectra. These experiments establish that secondary structure in both MBNL1 TZF pairs extends beyond the boundaries of the previously characterized CCCH TZF protein, TIS11d, and that the constructs used in the MBNL1 X-ray crystal structures<sup>25</sup> were likely too short. The MBNL1 constructs containing

residues 1–92 (MBNL1-zf12), and residues 173–255 (MBNL1-zf34) were used in all subsequent experiments.

A construct containing all four zinc fingers of MBNL1 (MBNL1-zf1234, residues 1–255) was also prepared. Comparison of the  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectra of the individual two-finger constructs MBNL1-zf12 (Figure 2A) and MBNL1-zf34 (Figure 2B) with that of the four-finger construct (Figure 2C) revealed an almost perfect overlap of common cross peaks (Figure 2D), indicating that the zf12 and zf34 motifs fold autonomously and that there are no long-range interactions between them. Cross peaks of residues in the long linker between the two tandem zinc finger domains (black cross peaks visible in Figure 2D) are sharp and appear in the random coil chemical shift region, indicating that they belong to an unstructured and flexible linker. Addition of EDTA induces unfolding of MBNL1, resulting in loss of cross peak dispersion in the  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectrum, demonstrating that zinc is essential for stabilizing the native fold of the MBNL1 CCCH zinc fingers.

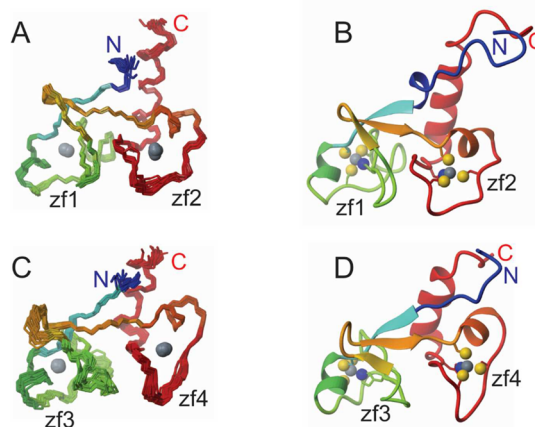
The sequence of the human muscleblind proteins is highly conserved with little variation in the zinc finger region and identical spacing between zinc ligands in zinc fingers 1 and 3 and zinc fingers 2 and 4, respectively (Figure 1B). However, an extra cysteine residue is conserved at position 43 between zf1 and zf2 (Figure 1B, pink). To rule out the possibility that this residue plays a role in zinc binding, the mutants C43A, C43S, and C43V were expressed. The  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectra of these variants showed only minor differences from the wild-type spectrum, confirming that Cys43 is not a zinc ligand.

The zinc-coordinating histidine residues and their tautomeric state were identified from  $^2\text{J}_{\text{HN}}$   $^1\text{H}$ – $^{15}\text{N}$  HMQC spectra using  $^{15}\text{N}$  chemical shift analysis and relative intensities of cross peaks arising from two- and three-bond couplings between nonexchangeable histidine ring protons  $\text{H}^{\delta 2}$  and  $\text{H}^{\epsilon 1}$ .<sup>40</sup> These spectra show that the zinc-binding histidines are His38, His70, His204, and His238, which bind zinc through  $\text{N}^{\epsilon 2}$  in zinc fingers 1–4, respectively, whereas His74 (end of finger 2, brown in Figure 1B) and His242 (end of finger 4, orange in Figure 1B) are not zinc ligands, consistent with the sequence alignment.

#### Structure of MBNL1-zf12 and MBNL1-zf34 Domains.

The structures of the free MBNL1-zf12 and MBNL1-zf34 domains were determined by multidimensional NMR spectroscopy. Backbone assignments were made using triple resonance experiments, and side chain assignments were completed using 3D heteronuclear-edited TOCSY and COSY experiments.<sup>52</sup> Chemical shift analysis of MBNL1 tandem zinc fingers showed that, unlike the TIS11d TZF,<sup>11</sup> the region between the zinc fingers in each pair contains secondary structure in the free state.

Automated NOE assignment and structure calculation were performed using several rounds of CYANA version 2.1<sup>44</sup> using 2895 NOEs (1779 unambiguous) in the case of MBNL1-zf12 and 1849 NOEs (1095 unambiguous) in the case of MBNL1-zf34 as well as dihedral angle restraints and hydrogen bond restraints. Once structural convergence was achieved and the CYANA target function was less than 1.0, 200 CYANA structures were generated and subsequently refined by restrained molecular dynamics-simulated annealing using the AMBER 8 software package.<sup>45</sup> The 20 lowest-energy NMR structures are shown in Figure 3, and experimental restraints and structural statistics are summarized in Table 1. The high quality of the structures was reflected in the favorable Ramachandran statistics as well as the tight convergence of

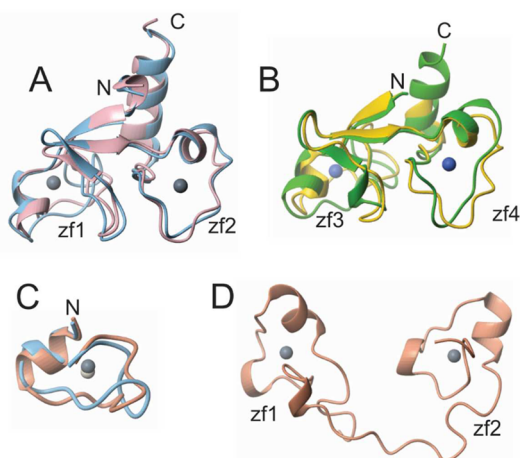


**Figure 3.** NMR solution structures of free MBNL1-zf12 (A, B) and MBNL1-zf34 (C, D). (A) Family of 20 lowest-energy structures of free MBNL1-zf12 superimposed on the well-structured region of the protein from residues Thr11 to Asn85. The disordered N- and C-termini have been omitted for clarity. (B) Ribbon representation of the lowest-energy structure of the ensemble shown in (A), showing the ligation of the two zinc centers. (C) Family of 20 lowest-energy structures of free MBNL1-zf34 superimposed on the well-structured region of the protein from residues Thr179 to Gln251. The disordered N- and C-termini have been omitted for clarity. (D) Ribbon representation of the lowest-energy structure of the ensemble shown in (C), showing the ligation of the two zinc centers (gray). In each panel, the backbone is colored from blue at the N-terminus to red at the C-terminus. In parts (B) and (D), yellow spheres represent the sulfur atoms of the ligating cysteines, and the blue sphere represents the  $\text{N}^{\epsilon}$  of the ligating histidine.

both structures with average backbone RMSD of 0.55 and 0.60 Å in the case of MBNL1-zf12 and MBNL1-zf34, respectively.

MBNL1-zf12 adopts a folded structure from Thr11 to Asn86, and MBNL1-zf34 is well-structured between residues Thr179 and Gln251; residues outside these regions are disordered. The ligation pattern for zinc fingers 1 and 3 is  $\text{C}(\text{X})_7\text{C}(\text{X})_6\text{C}(\text{X})_3\text{H}$ , extending from residue Cys19 to His38 and residues Cys185 to His204, respectively, and the pattern for zinc fingers 2 and 4 is  $\text{C}(\text{X})_7\text{C}(\text{X})_4\text{C}(\text{X})_3\text{H}$ , extending from residue Cys53 to His70 and residues Cys221 to His238, respectively. Each of the four zinc fingers is stabilized by binding of zinc through the  $\text{S}^{\gamma}$  atoms of three cysteine residues and the  $\text{N}^{\epsilon 2}$  atom of one histidine. The zinc-coordinating residues are Cys19, Cys27, Cys34, and His38 (zf1); Cys53, Cys61, Cys66, and His70 (zf2); Cys185, Cys193, Cys200, and His204 (zf3); and Cys221, Cys229, Cys234, and His238 (zf4).

Each of the four individual zinc fingers of MBNL1 is similar in structure with a short helix between the first and second cysteine of each finger. The structured portions of the individual zinc fingers superimpose well with the corresponding portions of the published X-ray crystal structures (Figure 4A, B), although the C-terminal helix of zf34 is truncated in the X-ray structure (Figure 4B). The individual zinc fingers of MBNL1 also superimpose well with the individual TIS11d zinc fingers<sup>11</sup> (Figure 4C), but whereas each pair (zf1-zf2; zf3-zf4) of MBNL1 zinc fingers forms a compact domain, the TIS11d zinc fingers are independent and separated by a long and flexible linker (Figure 4D). The linker between the MBNL1 zf1-zf2 zinc finger pairs forms a short  $\beta$ -hairpin (Cys43 to Ala52 in zf12), which together with a third  $\beta$ -strand (residues Thr15-Glu17 in zf12) participates in a three-stranded  $\beta$ -sheet stabilized by packing against the C-terminal  $\alpha$ -helix (Pro73 to



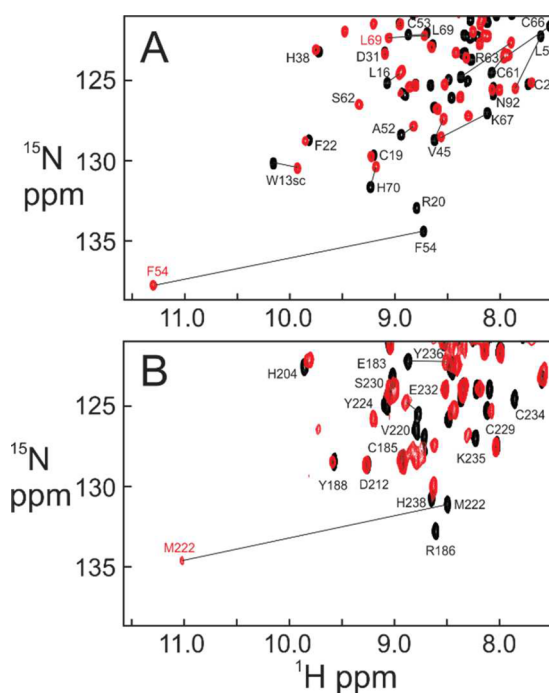
**Figure 4.** Comparison of the structures of MBNL1-zf12 and MBNL1-zf34 with similar published structures. (A) Overlay of the lowest-energy structure of MBNL1-zf12 (structured region, residues 11–85 only; blue) with the published crystal structure of free zf12 (residues 11–86; pink) (PDB ID: 3D2N).<sup>25</sup> (B) Overlay of the lowest-energy structure of MBNL1-zf34 (structured region, residues 180–252; green) with the published crystal structure of free zf34 (residues 180–245; gold) (PDB ID: 3D2Q).<sup>25</sup> (C) Overlay of MBNL1-zf1 (residues 18–38; blue) with zf1 of TIS11d (residues 158–178; coral).<sup>11</sup> (D) Ribbon representation of the backbone of TIS11d in complex with an RNA sequence representing the 3'-UTR of a target mRNA,<sup>11</sup> showing the separation of the two zinc finger domains.

Arg84 in zf12). Similar interactions are seen for the linker between MBNL1 zf3-zf4 ( $\beta$ -hairpin Ile211 to Thr219; third  $\beta$ -strand Arg181-Glu183; helix packing Ala241-Ala250). In contrast, the linker between fingers 1 and 2 of TIS11d is fully extended to allow binding of the individual fingers to their cognate RNA recognition motifs, and the TIS11d zinc fingers lack the C-terminal helix found in MBNL1.<sup>11</sup> The zf12 and zf34 domains of MBNL1 thus adopt an unusual fold in which the two CCCH zinc fingers are packed together to form a compact globular structure with an extensive hydrophobic core formed by the residues of the C-terminal  $\alpha$ -helix and of the  $\beta$ -sheet between the two zinc fingers.

**Identification of a Minimal RNA Binding Site.** Cross-linking experiments identified a common motif YGCU(U/G)Y for MBNL1 binding in the human cardiac troponin T pre-mRNA, where Y is a pyrimidine.<sup>14</sup> A schematic diagram of the intron 4-intron 5 region of hcTnT pre-mRNA, together with the full nucleotide sequence of the region that interacts with muscleblind protein, is shown in Figure S1. In hcTnT RNA, the pyrimidine 5' to the GCU consensus is a cytidine, making the RNA binding sequence CGCU(U/G)Y. There are two CGCU binding sites located in the human cTnT RNA in the 3' region of intron 4.<sup>20</sup> The binding of MBNL1 zf12 and zf34 constructs to RNA constructs of different lengths containing the CGCU(U/G)Y motif was tested by biolayer interferometry (Octet) and NMR.

A qualitative assessment of the strength of binding of various RNAs can be made using HSQC titrations of <sup>15</sup>N-labeled MBNL1 zf12 and zf34. Formation of a high affinity complex results in slow-exchange shifts of numerous cross peaks and relatively uniform peak intensities, and weaker binding leads to exchange broadening and smaller changes in the chemical shift for residues in the RNA binding sites. For MBNL1 zf1234, the addition of RNA caused extensive resonance broadening in the

<sup>15</sup>N HSQC spectra even at low RNA:protein ratios. This result indicates that the affinity of zf1234 for the RNA is high, but that the presence of two similar binding sites on both the protein and RNA likely results in intermediate exchange between alternative combinations of the protein and RNA sites. The 1:1 stoichiometry of the interaction between MBNL1 zf1234 and hcTnT32 is seen in the 1D spectra shown in Figure 7. Representative data for slow-exchange binding are shown in Figure 5 (complete HSQC spectra are shown in Figure S2), and the results are summarized in Table 2.



**Figure 5.** Binding of the 15 nucleotide RNA hcTnT15 (5'-GUCUCGCUUUUCCCC-3') to MBNL1 two-finger constructs, shown by superposition of a portion of the <sup>1</sup>H-<sup>15</sup>N HSQC spectra of (A) MBNL1-zf12 (black) with that following the addition of a 1:1 mol ratio of hcTnT15 (red) and (B) MBNL1-zf34 (black) with that following the addition of a 1:1 mol ratio of hcTnT15 (red).

**Table 2. RNA Constructs and Binding Affinity<sup>a</sup>**

5'-agacuaacc <u>gucgcuuuu</u> ccccucCGCUgcgcccauccc-3'	NMR			Octet
	zf12	zf34	zf1234	zf12
hcTnT32	+++	+++	+++	nd
hcTnT27	+++	++	+++	nd
hcTnT23	+++	++	++(+)	nd
hcTnT15	+++	+(+)	++(+)	130 ± 20 nM
hcTnT15m	nd	nd	nd	500 ± 70 nM
hcTnT10	++(+)	+(+)	nd	1100 ± 300 nM
hcTnT14m	nd	nd	nd	400 ± 70 nM

<sup>a</sup>The nucleotide sequence is from the intron 4 sequence of human cardiac troponin T (hcTnT) (Figure S1). The shaded box denotes the sequence used in the solution structure determination of the RNA complex of zf12. The numbers correspond to the nucleotide labels in Figure 9. Affinities estimated by NMR are denoted +++ for binding in slow exchange, ++ for binding in intermediate exchange, and + for binding in fast exchange. Additional + signs in brackets indicate that the spectrum shows a mixture of cross peaks with different exchange behaviors. nd: not determined. hcTnT15m is identical to hcTnT15 except that U103 is replaced with G. hcTnT14m has the same sequence as hcTnT10 with the addition of four extra nucleotides.

The NMR experiments clearly show that the tandem zinc finger domains bind to the CGCU motifs and that high affinity binding requires a minimum of 9–10 nucleotides. Shorter RNA constructs, such as CGCUG, bind only very weakly, showing that nucleotides outside the CGCU motif also interact with the protein. Substitution of the G in the CGCU motif by C severely impairs zf12 binding, showing that this nucleotide makes critical interactions with the protein. The NMR spectra in Figure 5 show large changes in chemical shift in the presence of RNA for resonances that belong to zf2 and zf4. Initial experiments with hcTnT10 and hcTnT15, representing the 3' and 5' CGCU sequences of the hcTnT RNA, showed additional broadening and disappearance of resonances belonging to zf1 and zf3, suggesting the presence of a secondary interaction site in this region. This effect is especially notable for the amide resonances of Arg20 and Arg160 in zf1 and zf3, respectively. However, in experiments performed using zf12, the Arg20 resonance reappeared close to its position in the free protein spectrum when the temperature was increased from 25 to 37 °C. Furthermore, no broadening of the Arg20 cross peak was observed upon binding of zf12 to a truncated RNA (hcTnT13), lacking the two 5' nucleotides of hcTnT15. These results imply that the interaction that causes the broadening and disappearance of the zf1 resonances (and probably those of zf3 in the zf34 complexes) is a secondary interaction involving nucleotides on the 5'-side of the CGCU motif.

NMR titration experiments show that a single tandem zinc finger domain (zf12 or zf34) is sufficient for RNA binding activity. Titration of MBNL1-zf12 with hcTnT10 and hcTnT15, each of which contains a single CGCU sequence showing 1:1 binding stoichiometry with no further chemical shift changes in the <sup>1</sup>H–<sup>15</sup>N HSQC spectra at protein:RNA ratios higher than 1:1. On the other hand, hcTnT27 and hcTnT32, which contain two CGCU motifs, bind two mole equivalents of zf12. Titration of MBNL1-zf1234 protein with hcTnT32 RNA also shows a 1:1 binding stoichiometry, confirming that each pair of zinc fingers binds to one CGCU unit. Previous studies have shown that the RNA binding affinity is enhanced in constructs containing both zf12 and zf34.<sup>23,27,53</sup>

Biolayer interferometry was used to obtain estimates of binding affinity for MBNL1-zf12 and showed that bases flanking the CGCU motif influence the specificity of the interaction. The Octet measurements (traces shown in Figure S3) gave *K<sub>d</sub>* values of 130 nM for binding of MBNL1-zf12 to the 5' CGCU motif in hcTnT15 (5'-GUCUCGCUUUUUC-CCC-3') and 500 nM for binding to a mutant RNA sequence where the U following the CGCU sequence is changed to G (5'-GUCUCGCGUUUUCCCC-3'). These results demonstrate that, in this system, the consensus RNA sequence YGCU(U/G)Y<sup>14</sup> shows a preference for CGCUuY over CGCUgY.

**Confirmation of the Binding Stoichiometry by NMR Diffusion Measurements.** During the course of this work, a structural study of the MBNL zinc fingers<sup>25</sup> reported apparent trimer or dimer structures for free MBNL1-zf12 and MBNL-zf34 (shorter constructs than ours). The structure of a complex between MBNL1-zf34 and the RNA sequence 5'-CGCUGU-3' based on hcTnT mRNA showed MBNL1-zf34 protein as a dimer with a complex interaction between three RNA strands and three zinc fingers and with a fourth zinc finger involved in protein–protein contacts.<sup>25</sup> Because our studies showed the free zf12 and zf34 proteins to be monomeric and our titration

experiments indicated that one CGCU motif binds to one pair of zinc fingers, it became essential for us to establish the oligomerization state of the free MBNL1 protein constructs and that of the protein–RNA complexes. Further information on the state of the proteins in solution and the stoichiometry of RNA binding was provided by pulsed field gradient (PFG) NMR diffusion experiments<sup>43</sup> carried out under the same buffer conditions, temperature, and sample concentrations as the structure determination and RNA titrations. The results of the NMR diffusion experiments are shown in Table 3. The

**Table 3. Diffusion Coefficients and Molecular Weight Calculated by NMR PFG Diffusion Experiments**

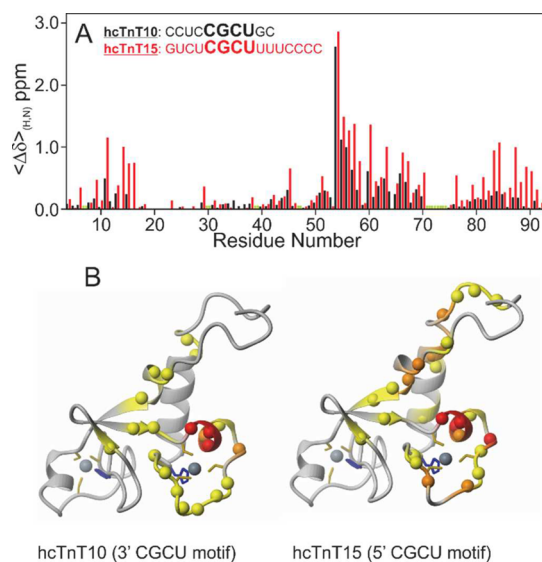
	diffusion coefficient ( $\times 10^{-7}$ cm <sup>2</sup> /s)	molecular weight (kDa)	theoretical molecular weight (kDa)
MBNL1-zf12	11.7 $\pm$ 0.5	9.2 $\pm$ 0.4	10.55
MBNL1-zf34	11.6 $\pm$ 0.4	9.5 $\pm$ 0.3	9.58
MBNL1-zf12 + hcTnT15	10.3 $\pm$ 0.5	13.3 $\pm$ 0.7	15.16
MBNL1-zf34 + hcTnT15	10.3 $\pm$ 0.3	13.5 $\pm$ 0.4	14.18
MBNL1-zf34 + hcTnT32	8.4 $\pm$ 0.3	25.3 $\pm$ 0.9	29.14

molecular weights obtained for the free MBNL1-zf12 and zf34 proteins are very close to the theoretical values expected for monomers. Similarly, the molecular weights calculated for complexes of MBNL1-zf12 + hcTnT15 and MBNL1-zf34 + hcTnT15 confirm the 1:1 stoichiometric ratio inferred from the NMR titration and Octet results. Also consistent with this finding, the molecular weight of the MBNL1-zf34 + hcTnT32 RNA complex was calculated from diffusion measurements as 25.3  $\pm$  0.9 kD, which is in good agreement with the theoretical molecular weight of two MBNL1-zf34 proteins binding to one hcTnT32 RNA molecule.

#### Mapping the Interaction of MBNL1-zf12 with RNA.

The binding sites of the RNA constructs hcTnT15 (containing the 5' CGCU motif) and hcTnT10 (containing the 3' CGCU motif) were mapped onto the structure of MBNL1-zf12 using chemical shift differences between <sup>1</sup>H–<sup>15</sup>N HSQC spectra of free and RNA-bound protein. Small differences in the RNA sequence outside the central CGCU motif in each oligonucleotide appear to have little effect on the contact site, but the length of the RNA construct and the number of nucleotides on the immediate 3' and 5' sides of the CGCU motif are crucial. Both hcTnT10 and hcTnT15 contain four nucleotides on the 5' side of the CGCU motif, and our NMR experiments using hcTnT13 indicate that this sequence should ideally be truncated by two nucleotides to avoid resonance broadening due to secondary (probably “fuzzy”<sup>54</sup>) interactions between the RNA and zf1. hcTnT10 has two nucleotides 3' of the CGCU, and hcTnT15 has seven. The length of this segment, though likely not its base composition (compare hcTnT10 and hcTnT14m in Table 2), provides increased affinity of the RNA for the TZFs. This is illustrated in Figure 6A, which shows increased RNA contact with the zf12 C-terminal helix and the nearby N-terminal helical turn for the complex of hcTnT15 with MBNL1-zf12 compared to the hcTnT10-zf12 complex, probably due to the increased length of the RNA, which enhances the affinity for MBNL1 zf12 (Table 2). The largest chemical shift changes are observed for zf2 residues Phe54–Leu69 together with residues in the region between zf1





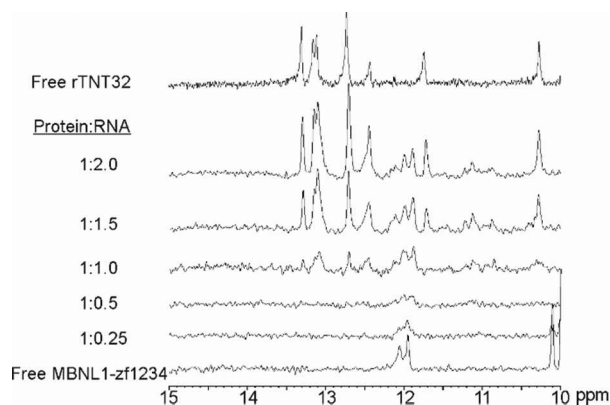
**Figure 6.** Mapping of the RNA binding site of MBNL1-zf12. (A) Chemical shift perturbation ( $\langle \Delta\delta \rangle_{\text{H,N}} = 1/2\sqrt{[(\Delta\delta_{\text{HN}})^2 + 0.2(\Delta\delta_{\text{N}})^2]}$ ) in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of MBNL1-zf12 upon addition of hcTnT10 (filled black bars) and hcTnT15 (filled red bars). Small green bars indicate missing cross peaks (including prolines) in both the free and complex spectra. (B) Chemical shift changes  $\langle \Delta\delta \rangle_{\text{H,N}}$  upon addition of (left) hcTnT10 or (right) hcTnT15 mapped onto the lowest-energy structure of MBNL1-zf12. The backbone ribbon and the backbone N atom of residues where  $\langle \Delta\delta \rangle_{\text{H,N}}$  takes values  $>2\times$  the standard deviation (SD) greater than the mean value are colored red, where  $2\times \text{SD} < \text{value} < 1\times \text{SD}$  are shown in orange, and where  $1\times \text{SD} < \text{value} < \text{mean}$  are shown in yellow. Zinc atoms are shown in gray, and the side chains of zinc ligands are shown as sticks, yellow for cysteine, and blue for histidine.

and zf2. Significant shifts are also observed for the cross peaks of Trp13 and Leu14 and several residues in the  $\beta$ -strands (Leu16, Glu17, Val45, Ile51, Ala52) and in the C-terminal  $\alpha$ -helix (Glu80, Ile81, Gly83, Asn85). The changes in the HSQC spectrum of MBNL-zf12 upon binding of hcTnT10 and hcTnT15 are mapped onto the structure in Figure 6B.

Several of the aromatic amino acids that interact with RNA in the TIS11d complex<sup>11</sup> are conserved in the MBNL zinc fingers, including Phe22, Phe36, and Tyr68 (Figure 1B). To assess the role of these MBNL1 aromatic residues in binding of the hcTnT RNA, each was mutated to alanine, and the interaction with hcTnT10 was examined using HSQC titrations. Mutation of Phe36 to Ala makes little difference for binding to hcTnT10 but appears to hinder the nonspecific interaction that occurs for hcTnT10 and hcTnT15 at zf1, consonant with the position of this Phe between two of the zf1 zinc ligands. Substitution of Tyr68 with alanine seriously impairs hcTnT10 binding, such that the cross peaks that characterize high affinity RNA binding to finger 2 are broadened and are not observed in the HSQC spectra. Thus, it is clear that, as for TIS11d, exposed aromatic residues in MBNL1-zf12 play an important role in RNA binding.

**Role of RNA Secondary Structure.** It has been reported that the MBNL zinc fingers bind specifically to RNA hairpins.<sup>20,55</sup> To assess this possibility, we performed NMR titrations with longer RNA constructs (hcTnT32 and hcTnT27 containing 32 and 27 nucleotides, respectively), which have the potential to form hairpins between the CGCU motifs located in the stem.<sup>20</sup> A 1D WATERGATE NMR spectrum of free

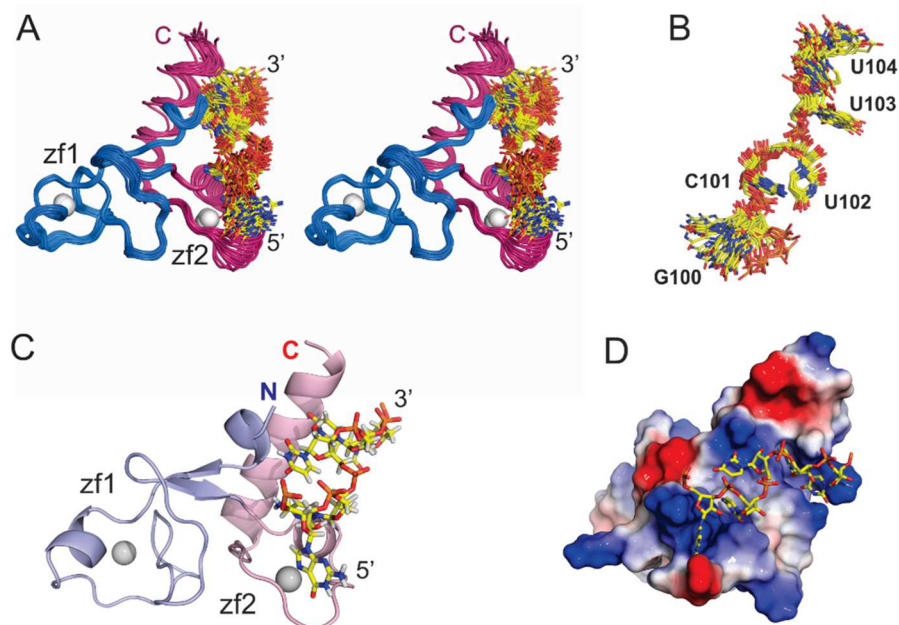
hcTnT32 reveals distinct imino proton resonances, confirming the presence of RNA secondary structure (Figure 7). As



**Figure 7.** One-dimensional spectrum of the region 10–15 ppm containing the imino resonances of hydrogen-bonded RNA. Free hcTnT32 contains several imino peaks (top spectrum). No imino resonances are present in the spectrum of the free protein (bottom spectrum) or in the presence of less than 1:1 mol ratio of hcTnT32 RNA. Above 1:1, the imino resonances of the RNA are present in the spectrum, indicating that the excess RNA contains hydrogen-bonded secondary structure.

hcTnT32 RNA was titrated into an MBNL1234 protein sample, no peaks appeared in the imino proton region until the protein:RNA ratio reached 1:1. The absence of imino peaks at substoichiometric RNA concentrations suggests that hcTnT32 RNA is single-stranded when it is bound to the MBNL zinc fingers. At protein:RNA ratios greater than 1:1, imino resonances start to appear at the same chemical shifts as for the free RNA, suggesting that the excess RNA does not participate in binding to protein and exhibits the same secondary structure as in the free RNA hairpin. Thus, although the free hcTnT32 RNA adopts secondary structure, potentially similar to that predicted by Warf and Berglund,<sup>20</sup> MBNL1-zf1234 binding disrupts the secondary structure. That zf1234 binds preferentially to single-stranded RNA rather than a hairpin stem is confirmed by the similar binding constants and nearly identical chemical shifts observed for key residues in the RNA binding site in HSQC spectra of zf1234 complexes with hcTnT32 and with shorter RNA constructs that are unable to form a hairpin (Table 2). Higher affinity of MBNL zinc fingers for single-stranded RNA, compared to double-stranded forms, has also been demonstrated for other RNA sequences.<sup>53</sup>

We repeated our experiments by adding MBNL1-zf12 or MBNL1-zf34 to hcTnT32. The results were similar to those obtained for binding to MBNL1-zf1234. Consistent with our NMR titration experiments, which showed each pair of zinc fingers binding to one CGCU recognition site in the hcTnT RNA; no imino peaks were observed until a protein:hcTnT32 ratio of 2:1 was reached, indicating once again that the hcTnT32 RNA is single stranded when bound to MBNL. Beyond the 2:1 ratio, the excess RNA adopts hydrogen-bonded secondary structure as evidenced by the presence of imino proton peaks between 10 and 15 ppm. Lower affinities were observed for the binding of zf12 and zf34 to similar RNA constructs compared with the four-finger zf1234.<sup>53</sup> The lower affinity of the two-finger constructs was ascribed to the presence of C-G base pairing between RNA strands, which could be successfully disrupted to give a single-stranded RNA



**Figure 8.** Solution structures of the complex of MBNL1 zf12 and hcTnT15 single-stranded RNA. (A) Stereo ribbon representation of the family of structures. The backbones of zf1 and zf2 are shown in blue and pink, respectively. Zinc atoms are shown as white spheres. The non-hydrogen atoms of the RNA (nucleotides G100–U104) are colored yellow for carbon, blue for nitrogen, red for oxygen, and orange for phosphorus. (B) The family of RNA structures from G100–C101–U102–U103–U104 are colored as in (A). The structures have been rotated  $\sim 180^\circ$  from their positions in (A). (C) Cartoon ribbon representation of the lowest-energy structure of MBNL1 zf12 in complex with hcTnT15 RNA. Nucleotides G100–C101–U102–U103–U104 are shown as sticks and colored as in (A) with RNA hydrogen atoms shown in white. (D) Vacuum electrostatic surface representation of the protein portion of the structure shown in (C) calculated within the program PyMol (The PyMOL Molecular Graphics System, version 1.8 Schrödinger, LLC). Areas of negative and positive charge are shown in red and blue, respectively. The structure has been rotated  $\sim 90^\circ$  both in the plane of the paper and around the vertical axis compared to the cartoon in (C). The non-hydrogen atoms of the RNA (G100–U104) are colored as in (A).

by zf1234, but which could not be disrupted by the two-finger constructs.<sup>53</sup>

**Structure of the Complex of MBNL1-zf12 with hcTnT15 RNA.** Having established that each pair of MBNL1 TZF domains (zf12 and zf34) interacts with single-stranded RNA containing a single CGCU sequence, we proceeded to determine the solution structure of MBNL-zf12 (residues 1–92) bound to the minimal cognate RNA oligonucleotide hcTnT15 with sequence 5′-GUCUCGCUUUUCCCC-3′. Resonance assignments for  $^{15}\text{N}$ ,  $^{13}\text{C}$ -labeled zf12 in complex with unlabeled RNA were made using a standard suite of triple resonance spectra, as described above for the free protein. Resonance assignments for the unlabeled RNA in the complex were more problematic, as resonance overlap was severe, particularly in the regions of the RNA that are not intimately associated with the protein. Partial  $^1\text{H}$  assignments were obtained from resolved cross peaks in a two-dimensional  $^{13}\text{C}$ -filtered spectrum.<sup>56</sup> NOEs between zf12 and hcTnT15 were obtained from a 3D  $^{13}\text{C}$  filter-edit spectrum<sup>42</sup> yielding 71 intermolecular NOEs.

NOE assignment and structure calculation were performed using CYANA version 2.1,<sup>44</sup> and the structures were refined using AMBER 11<sup>48</sup> using a similar method to that described above for the free zf12 and zf34 proteins. The 20 lowest-energy NMR structures of the complex are shown in Figure 8A, and experimental restraints and structural statistics are included in Table 1. The structure is well-defined between Asp10 and Asn85 with a pairwise rms deviation of 0.37 Å for the polypeptide backbone. Residues 1–9 and 89–92 are disordered, and the last turn of the C-terminal helix (Arg84–

Ile88) is frayed in the structural ensemble. The RNA structure and MBNL1 interactions are well-defined only for nucleotides C101–U104 (Figure 8B). Although some intermolecular NOEs are observed to the first two nucleotides in the CGCU motif, C99 and G100, they involve long protein side chains and long upper distance bounds; as a consequence, the RNA structure is poorly defined in this region. The orientation of the bases and protein interactions are well-defined only for C101, U102, and U103; the U104 base adopts two orientations in the ensemble of NMR structures that cannot be distinguished on the basis of the observed NOEs. The structure of MBNL1 zf12 shows very little change between free and bound states. The RNA-binding surface of MBNL1 zf12 indicated by the changes in  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts (Figure 6) corresponds well with the binding site observed in the calculated solution structures (Figure 8C).

## DISCUSSION

**Single-Stranded RNA is the Preferred Binding Partner for MBNL.** The affinity of muscleblind zinc fingers for RNA has received a great deal of attention. Although early studies appeared to show that MBNL zinc fingers bound to RNA hairpins,<sup>20,55</sup> it is clear from our studies and from literature reports that have appeared during the course of this work<sup>18,53,57</sup> that single-stranded RNA binds preferentially to the MBNL zinc fingers. The presence of stable base-paired double-stranded RNA lowers the affinity of GST-bound single zinc finger pairs (zf12 or zf34), but the GST-bound tandem construct zf1234 was able to destabilize this structure and bind to the resulting single-stranded RNA.<sup>53</sup> Despite this evidence that the full-

length MBNL has additional RNA affinity over the two-finger constructs (which may well arise from the entropic gain from intramolecular bivalency), our NMR data suggest that the 80 residue linker between the two pairs of tandem zinc fingers (TZF) is largely disordered in the free protein. The  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of the four-finger construct (MBNL1 1–255) superimpose exactly with those of the individual TZF domains (zf12, residues 1–92, and zf34, residues 173–255) (Figure 2), leaving only a region in the center of the spectrum containing the random-coil cross peaks of the disordered linker. These observations validate our experimental strategy of characterizing the two TZF domains independently.

#### Comparison of the Free Two-Finger Constructs.

Consistent with the amino acid sequence homology between zf12 and zf34 (Figure 1B), the solution structures calculated for zf12 and zf34 are very similar (Figure 3). Within each TZF, the two zinc centers are folded together to form a single structure. Residues between the individual CCCH zinc fingers fold to form a three strand  $\beta$ -sheet, which together with the C-terminal  $\alpha$ -helix forms a compact globular structure with an extensive hydrophobic core. During the course of this work, several structures were reported for MBNL molecules similar to those reported here. Our structures are very similar both in backbone as well as side chain orientation to the solution structures of MBNL isoform 2 calculated by the RIKEN structural genomics initiative (PDB ID: 2RPP corresponding to our zf12 and 2ESS corresponding to our zf34), consistent with the high degree of amino acid sequence homology (Figure 1). On the other hand, there are several differences between our tandem zinc finger domain structures and the X-ray structures reported by Teplova and Patel (PDB ID: 3D2N for zf12 and 3D2Q for zf34).<sup>25</sup> The most obvious difference is that, in the crystal structures, free MBNL1-zf12 is a trimer whereas free MBNL1-zf34 is a dimer. We note that both of the protein constructs used for the X-ray structures are shorter than ours (residues 9–90 for zf12 with coordinates reported for residues 11–86, and residues 178–246 for zf34, with coordinates reported for residues 180–245). By contrast, our NMR structures for the corresponding TZFs were calculated using data for residues 1–92 (zf12) and 173–255 (zf34). The NMR structures show that only residues 11–85 (zf12) and 179–251 (zf34) are well-structured in the free state (Figure 3). The differences between X-ray and NMR in the extent of the structured core of each TZF are rather minor for zf12, but the difference at the C-terminus of zf34 is significant with the C-terminal helix being truncated by two helical turns in the crystal structure.<sup>25</sup> In solution, we observe only monomeric structures for the two TZF constructs, as confirmed by the PFG NMR results (Table 3). We suggest that the oligomers observed in the X-ray structures might have arisen either due to truncation of the C-terminal helix in the case of zf34 or due to crystal contacts.

**Comparison of MBNL1 and TIS11d Zinc Finger Structures.** The structure of a CCCH tandem zinc finger (TZF) domain showed the zinc fingers of TIS11d in complex with an AU-rich RNA from the 3'-untranslated region of TNF- $\alpha$  mRNA (5'-UUAUUUAUU-3').<sup>11</sup> The two CCCH zinc centers of TIS11d form two independently folded zinc fingers connected by a flexible amino acid chain of 18 residues.<sup>11</sup> An alignment of structures of the individual zinc fingers of TIS11d with that of MBNL1 (Figure 4C) shows that their structures are very similar with the presence of a short  $\alpha$ -helix between the first and second zinc-binding cysteines. However, the overall folds of these proteins are quite different: the C-terminal  $\alpha$ -

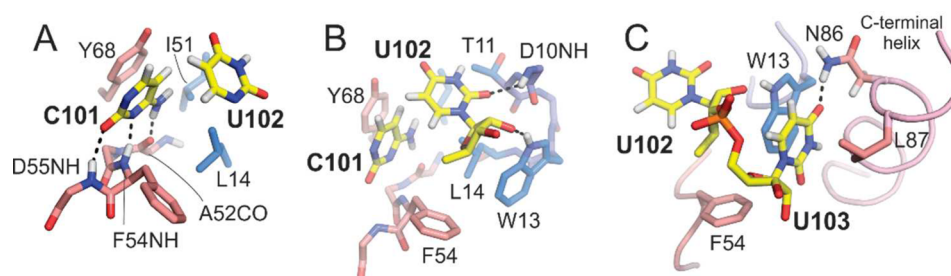
helix and three-strand  $\beta$ -sheet observed in the MBNL1 TZFs are absent, and the two zinc fingers of TIS11d behave like independent beads on a flexible string instead of forming a single globular structure as seen for MBNL1.

An interesting feature of the zinc fingers of TIS11d is the presence of a highly conserved (R/K)YKTEL motif immediately preceding the first cysteine in each zinc finger (Figure 1). The structure of the TIS11d-RNA complex shows that this motif plays a critical role in RNA recognition by forming two walls of a deep pocket that accommodates the uracil bases of the recognition element. Sequence alignment of TIS11d with MBNL1 (Figure 1) shows that the RYKTEL motif of TIS11d finger 1 is replaced by WLTLEV in MBNL1 zf1 and TDRLEV in MBNL1 zf3, whereas the KYKTEL motif is replaced by NGRVIA in MBNL1 zf2 and DNTVTV in MBNL1 zf4. An extensive sequence comparison between muscleblind family members<sup>3</sup> showed that the LEV box and to a lesser extent the NGR box are highly conserved between widely divergent species (Figure 1). These sequences participate in the formation of the three-stranded  $\beta$ -sheet, which forms part of the unique fold of the MBNL TZFs.

**RNA Recognition by MBNL.** The two TZF constructs MBNL1-zf12 and MBNL-zf34 bind to RNA sequences from cardiac troponin T with comparable  $K_d$  values in the 100–200 nM range; the affinity of full length MBNL1-zf1234 is higher<sup>20</sup> (Table 2). The NMR titrations establish that the hcTnT RNAs bind preferentially in the vicinity of the second zinc site of each TZF with the largest chemical shift changes upon addition of RNA occurring for residues 54–60 in zf2 of MBNL1-zf12 and for the corresponding residues 222–228 of zf4 in MBNL-zf34. Resonance broadening for Arg20 and Arg186, at equivalent sites in zf1 and zf3 (circled in red in Figure 1B), in the presence of RNA constructs containing four nucleotides on the 5' end of the CGCU interaction motif is attributed to secondary interactions. These interactions are abrogated by 5'-truncation of the RNA (in hcTnT13) or by raising the temperature from 25 to 37 °C. We conclude that the primary RNA binding site is zf2 in MBNL-zf12 and zf4 in MBNL-zf34 and that the secondary interactions with zf1 and zf3 are weak. Although zf1 does present an RNA binding surface, the potential interactions are limited compared to those of zf2, which functions in conjunction with side chains in the N-terminal region and C-terminal helix to form an extensive binding site and provide a network of base-specific RNA contacts.

It is interesting to note the unusually large downfield chemical shift of the amide cross peaks of Phe54 (zf2;  $\Delta\delta = 3.7$  ppm in  $^{15}\text{N}$  and  $\Delta\delta = 2.48$  ppm in  $^1\text{H}$ ) and Met222 (zf4;  $\Delta\delta = 3.25$  ppm in  $^{15}\text{N}$  and  $\Delta\delta = 2.52$  ppm in  $^1\text{H}$ ) induced by RNA binding (Figure 5; residues circled in Figure 1B). The resonances of the equivalent residues in TIS11d, Arg160 in zf1, and Arg198 in zf2 (circled in Figure 1B) also undergo substantial downfield shifts upon RNA binding.<sup>11</sup> The large downfield shift of the Arg160 and Arg198 amide resonances of TIS11d is due to formation of hydrogen bonds between these amide protons and the N1 atom of adenines A3 and A7. The large downfield shifts for the Phe54 and Met222 resonances of MBNL1 suggest that their amide protons are similarly involved in hydrogen-bonding interactions with hcTnT RNA.

Our NMR structure of MBNL1 zf12 bound to the 15 nucleotide hcTnT RNA are complementary to the published crystal structure of zf34 bound to CGCUGU.<sup>25</sup> Together, the NMR and X-ray structures provide detailed new insights into the mechanism of RNA recognition by the muscleblind zinc



**Figure 9.** Protein–RNA contacts in the complex between MBNL1 zf12 and hcTnT15. (A) Hydrogen-bonding interactions between C101 and the backbone amide protons of Phe54 and Asp55, and the CO of Ala52, and hydrophobic interactions between U102 and the side chains of Leu14 and Ile51, and between Cyt101 and the side chain of Tyr68. (B) Hydrogen-bonding interaction between U102 and the backbone amide proton of Asp10 observed in many of the NMR structures. (C) Hydrogen-bonding interaction between U103 and one of the Asp86 side chain amide protons. The U103 base is stacked between the hydrophobic side chains of Trp13 and Leu 87.

fingers. In the solution structure, the RNA binds in a positively charged groove formed between residues Arg9–Thr15, zf2, and the C-terminal helix (Figure 8D). The first six nucleotides at the 5' end of the hcTnT15 RNA, including the first two bases of the CGCU binding motif, are structurally disordered. Although G100 clearly binds to zf2, the details of the interaction cannot be discerned in the NMR structures because of this disorder. Fortunately, the guanosine interactions are clear in the X-ray structure of the zf34 complex,<sup>25</sup> where it is packed between the conserved Arg231, Lys235, and Tyr236 side chains (equivalent to Arg63, Lys67, and Tyr78 in zf12) and forms a hydrogen bond to the Lys amide. It is likely that this binding mode is conserved for zf12. Although the paucity of NOE restraints prevented resolution of the interactions between G100 and zf2, the amide <sup>1</sup>H and <sup>15</sup>N resonances of K67 are shifted strongly downfield (Figure 5) upon binding to RNA, consistent with hydrogen bond formation to the guanosine base. Only one weak intermolecular NOE is observed for C99, suggesting that its interactions with zf12 are transient and nonspecific. This conclusion is supported by the X-ray structure, where the equivalent cytosine from the CGCU motif makes no contact with zf4 and is only observed in the electron density for a subset of RNA molecules in which it is bound to zf3 of a neighboring protein molecule in the crystal lattice.<sup>25</sup> This interaction may be a crystallization artifact because the 5' cytosine base of the CGCU motif cannot reach the zf3 binding site observed in the X-ray structure, which is more than 30 Å from the guanosine binding site on zf4.

The conformation of C101 is well-defined by the NMR restraints. The cytosine ring projects into a deep pocket formed by the side chains of Ile51, Phe54, and Tyr68 on zf2 and Leu13 in the N-terminal region of the zf12 construct, where it can hydrogen bond with the polypeptide backbone (N4–A52 CO, N3–F54 NH, O2–D55 NH) (Figure 9A). The same pattern of hydrogen bonding is observed between the corresponding cytosine and backbone amides of V220, M222, and D223 in the X-ray structure of the zf34 complex.<sup>25</sup>

The protein interactions made by bases on the 3' side of the GC element differ substantially in the NMR and X-ray structures, most likely due to the truncation of the C-terminal helix of zf4 in the construct used for crystallization<sup>25</sup> as well as to the presence of a G in place of U103 (see below). The uracil moiety of U102 packs against the edge of the C101 base and the hydrophobic side chains of T11, W13, and L14 in the N-terminal region of zf12, and the O2 atom is within hydrogen bonding distance of the Asp10 backbone NH in many of the NMR structures (Figure 9B). The uracil at this position could

be replaced by cytosine without perturbing the hydrogen bonding or hydrophobic interactions with zf12; indeed, substitution of uridine by cytidine has no effect on binding affinity but replacement by a purine strongly impairs binding.<sup>18</sup> The conformation of U103 is also well-defined by the NMR restraints and is positioned with its uracil ring packed in a groove between the face of the Trp13 indole ring and the side chain of Leu87 on the C-terminal helix of zf2 (Figure 9C). Interestingly, the equivalent nucleotide in the X-ray crystal structure<sup>25</sup> is a G, which we show by biolayer interferometry corresponds to a lower-affinity complex. The uracil O4 atom forms a hydrogen bond to the Nδ of Asn86 in the majority of the NMR structures, and there is potentially an additional hydrogen bond between the uracil N3 and the backbone CO of Gly83 (depending on the torsion angle about the glycosidic bond). The base at this position appears to contribute weakly to specificity, and substitution by guanine decreases the binding affinity approximately 4-fold (Table 1). The uracil ring of U104 is conformationally disordered, adopting two approximately equally populated conformations in the ensemble of NMR structures, both of which pack against the C-terminal helix.

In contrast to the NMR structures of the zf12 complex, the nucleotides 3' to the GC undergo very different contacts with zf34 in the crystal. The first U forms similar interactions with the protein in the zf12 (NMR) and zf34 (X-ray) structures. However, subsequent nucleotides (G and U) bind to the zf3 finger in a neighboring molecule in the crystal lattice,<sup>25</sup> whereas the corresponding bases in the NMR structure (U104 and U105) interact with the C-terminal helix of zf3. These differences could be a consequence of crystal packing or, and perhaps more likely, due to truncation of the zf34 construct used for crystallization, leading to loss of the entire C-terminal helix and critical contact residues required for binding of these 3' nucleotides.

The NMR and X-ray structures in combination substantially advance our understanding of RNA recognition by the MBLN1 tandem zinc finger domains. Previous studies have shown that MBLN1 binds preferentially to the sequence motifs YGCU(U/G)Y or YGCY (where Y is a pyrimidine).<sup>14,16</sup> The central GC dinucleotide forms highly specific interactions with the zf12 and zf34 domains, and substitution of these nucleotides drastically decreases the binding affinity.<sup>16</sup> Although the molecular basis for recognition of the guanosine is not clear from the NMR data, the X-ray structure shows hydrogen bonding interactions to one of the cysteine ligands and a backbone NH in zf4 that is specific to a guanosine base.<sup>25</sup> The solution and crystal structures show that specific recognition of the cytosine by

zf2 and zf4 is mediated by a pattern of hydrogen bonding to backbone amides that is unique to the cytosine base. In addition, the cytosine projects into a confined hydrophobic pocket that could not accommodate a purine ring. The NMR structures provide new insights into the interactions involving nucleotides on the 3' side of the GC motif, showing the molecular interactions that favor a pyrimidine immediately following the cytidine, and the weak preference for uridine at the following position.<sup>18</sup> Neither the X-ray nor NMR structures provide insights into the preference for a pyrimidine immediately 5' of the GC dinucleotide.

**Effect of the Length of the RNA Sequence on Binding.** MBNL TZF domains require a minimum of approximately 10 nucleotides to bind RNA in a stable complex. The NMR titrations of MBNL-zf12 with RNAs of various lengths derived from the troponin sequence showed that zf2 binds to the CGCU recognition site, whereas nucleotides four or more bases 5' to that interact with zf1 but more weakly. Similarly, for MBNL-zf34, the CGCU sequence binds strongly to zf4, whereas 5' nucleotides interact weakly with zf3. The human cTNT pre-mRNA contains two CGCU sites separated by a 9 nucleotide spacer in the 3' end region of intron 4 where MBNL is predicted to bind.<sup>20</sup> Our results suggest that each of these CGCU units is recognized by a single tandem zinc finger domain, i.e., zf12 binds to one CGCU motif whereas zf34 binds to the other. This mode of binding is consistent with the results of our diffusion experiments, which showed that the hcTnT10 and hcTnT15 RNAs, each containing a single CGCU unit, bind to MBNL1-zf12 and MBNL1-zf34 in a 1:1 ratio, whereas hcTnT32 RNA, which contains two CGCU subunits, binds two molecules of MBNL1-zf34. Such a binding mode is contrary to the model proposed by Teplova and Patel<sup>25</sup> in which the hcTnT RNA forms a hairpin and the CGCU motifs interact with zf3 and zf4 on the same tandem zinc finger domain.

**Protein Data Bank Accession Code.** The coordinates for MBNL1-zf12 and zf34 and the zf12-RNA structure have been deposited with the Protein Data Bank with accession codes 5U6H, 5U6L, and 5U9B, respectively. The NMR resonance assignments have been deposited in the Biological Magnetic Resonance Data Bank with accession numbers 30207, 30208, and 30210, respectively.

## ■ ASSOCIATED CONTENT

### Supporting Information

A table listing . The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.7b00484.

Intermolecular protein–RNA NOEs used in the structure calculation of the MBNL zf12–RNA complex, RNA sequences, complete NMR <sup>1</sup>H–<sup>15</sup>N HSQC spectra of MBNL1 tandem zinc fingers and their complexes, and biolayer interferometry traces for MBNL zf12 with RNA (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Tel.: +1 858 784 9721. Fax: +1 858 784 9822. E-mail: wright@scripps.edu.

### ORCID

H. Jane Dyson: 0000-0001-6855-3398

## Present Addresses

<sup>†</sup>P.D.P.: Archer Norris, 777 South Figueroa Street, Los Angeles, CA 90017

<sup>‡</sup>M.Z.: Department of Medicinal Chemistry, Boehringer Ingelheim Pharma GmbH & Co. KG, D-88397 Biberach/Riss, Germany

## Funding

This work was supported by Grant GM36643 from the National Institutes of Health and by the Skaggs Institute for Chemical Biology. P.D.P. acknowledges receipt of a grant from the Muscular Dystrophy Association. M.Z. acknowledges receipt of a DFG postdoctoral research fellowship (ZE 809/1-1).

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Gerard Kroon for assistance with NMR experiments and Euvel Manlapaz and Linda Tennant for technical assistance.

## ■ REFERENCES

- (1) Begemann, G., Paricio, N., Artero, R., Kiss, I., Perez-Alonso, M., and Mlodzik, M. (1997) Muscleblind, a gene required for photoreceptor differentiation in *Drosophila*, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. *Development* 124, 4321–4331.
- (2) Artero, R., Prokop, A., Paricio, N., Begemann, G., Pueyo, I., Mlodzik, M., Perez-Alonso, M., and Baylies, M. K. (1998) The muscleblind gene participates in the organization of Z-bands and epidermal attachments of *Drosophila* muscles and is regulated by Dmef2. *Dev. Biol.* 195, 131–143.
- (3) Pascual, M., Vicente, M., Monferrer, L., and Artero, R. (2006) The Muscleblind family of proteins: An emerging class of regulators of developmentally programmed alternative splicing. *Differentiation* 74, 65–80.
- (4) Oddo, J. C., Saxena, T., McConnell, O. L., Berglund, J. A., and Wang, E. T. (2016) Conservation of context-dependent splicing activity in distant Muscleblind homologs. *Nucleic Acids Res.* 44, 8352–8362.
- (5) Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N., and Ishiura, S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.* 13, 495–507.
- (6) Miller, J. W., Urbinati, C. R., Teng-Ummuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., and Swanson, M. S. (2000) Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.* 19, 4439–4448.
- (7) Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshere, M. G., Harper, P. S., and Brook, J. D. (2002) 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.* 11, 805–814.
- (8) Squillace, R. M., Chenault, D. M., and Wang, E. H. (2002) Inhibition of muscle differentiation by the novel muscleblind-related protein CHCR. *Dev. Biol.* 250, 218–230.
- (9) Shaw, G., and Kamen, R. (1986) A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* 46, 659–667.
- (10) Chen, C. Y., and Shyu, A. B. (1995) AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem. Sci.* 20, 465–470.
- (11) Hudson, B. P., Martinez-Yamout, M. A., Dyson, H. J., and Wright, P. E. (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat. Struct. Mol. Biol.* 11, 257–264.
- (12) Cooper, T. A., and Ordahl, C. P. (1985) A single cardiac troponin T gene generates embryonic and adult isoforms via

developmentally regulated alternate splicing. *J. Biol. Chem.* 260, 11140–11148.

(13) Savkur, R. S., Philips, A. V., and Cooper, T. A. (2001) Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat. Genet.* 29, 40–47.

(14) Ho, T. H., Cooper, T. A., Charlet, B., Singh, G., Poulos, M. G., and Swanson, M. S. (2004) Muscleblind proteins regulate alternative splicing. *EMBO J.* 23, 3103–3112.

(15) Ladd, A. N., Stenberg, M. G., Swanson, M. S., and Cooper, T. A. (2005) Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development. *Dev. Dyn.* 233, 783–793.

(16) Goers, E. S., Purcell, J., Voelker, R. B., Gates, D. P., and Berglund, J. A. (2010) MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res.* 38, 2467–2484.

(17) Wang, E. T., Cody, N. A., Jog, S., Biancolella, M., Wang, T. T., Treacy, D. J., Luo, S., Schroth, G. P., Housman, D. E., Reddy, S., Lécuyer, E., and Burge, C. B. (2012) Transcriptome-wide Regulation of Pre-mRNA Splicing and mRNA Localization by Muscleblind Proteins. *Cell* 150, 710–724.

(18) Cass, D., Hotchkro, R., Barber, P., Jones, K., Gates, D. P., and Berglund, J. A. (2011) The four Zn fingers of MBNL1 provide a flexible platform for recognition of its RNA binding elements. *BMC Mol. Biol.* 12, 20.

(19) Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P. A., and Burge, C. B. (2014) RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol. Cell* 54, 887–900.

(20) Warf, M. B., and Berglund, J. A. (2007) MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA* 13, 2238–2251.

(21) Vicente, M., Monferrer, L., Poulos, M. G., Houseley, J., Monckton, D. G., O'Dell, K. M. C., Swanson, M. S., and Artero, R. D. (2007) Muscleblind isoforms are functionally distinct and regulate alpha-actinin splicing. *Differentiation* 75, 427–440.

(22) Mankodi, A., Urbinati, C. R., Yuan, Q. P., Moxley, R. T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M. S., and Thornton, C. A. (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.* 10, 2165–2170.

(23) Purcell, J., Oddo, J. C., Wang, E. T., and Berglund, J. A. (2012) Combinatorial Mutagenesis of MBNL1 Zinc Fingers Elucidates Distinct Classes of Regulatory Events. *Mol. Cell Biol.* 32, 4155–4167.

(24) Konieczny, P., Stepniak-Konieczna, E., and Sobczak, K. (2014) MBNL proteins and their target RNAs, interaction and splicing regulation. *Nucleic Acids Res.* 42, 10873–10887.

(25) Teplova, M., and Patel, D. J. (2008) Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat. Struct. Mol. Biol.* 15, 1343–1351.

(26) He, F., Dang, W., Abe, C., Tsuda, K., Inoue, M., Watanabe, S., Kobayashi, N., Kigawa, T., Matsuda, T., Yabuki, T., Aoki, M., Seki, E., Harada, T., Tomabechi, Y., Terada, T., Shirouzu, M., Tanaka, A., Guntert, P., Muto, Y., and Yokoyama, S. (2009) Solution structure of the RNA binding domain in the human muscleblind-like protein 2. *Protein Sci.* 18, 80–91.

(27) Edge, C., Gooding, C., and Smith, C. W. (2013) Dissecting domains necessary for activation and repression of splicing by muscleblind-like protein 1. *BMC Mol. Biol.* 14, 29.

(28) Delaglio, F., Grzesiek, S., Vuister, G. W., Guang, Z., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293.

(29) Goddard, T. D., and Kneller, D. G. (2006) SPARKY 3, University of California, San Francisco.

(30) Johnson, B. A., and Blevins, R. A. (1994) NMRView: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* 4, 603–614.

(31) Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure

determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 34, 93–158.

(32) Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* 13, 289–302.

(33) Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44, 213–223.

(34) Vuister, G. W., and Bax, A. (1993) Quantitative J correlation: A new approach for measuring homonuclear three-bond  $J(\text{H}^{\text{N}}\text{H}^{\text{A}})$  coupling constants in  $^{15}\text{N}$ -enriched proteins. *J. Am. Chem. Soc.* 115, 7772–7777.

(35) Yamazaki, T., Forman-Kay, J. D., and Kay, L. E. (1993) Two-dimensional NMR experiments for correlating  $^{13}\text{C}_\beta$  and  $^1\text{H}_\alpha/e$  chemical shifts of aromatic residues in  $^{13}\text{C}$ -labeled proteins via scalar couplings. *J. Am. Chem. Soc.* 115, 11054–11055.

(36) Archer, S. J., Ikura, M., Torchia, D. A., and Bax, A. (1991) An alternative 3D NMR technique for correlating backbone  $^{15}\text{N}$  with side chain H $\beta$  resonances in larger proteins. *J. Magn. Reson.* 95, 636–641.

(37) Grzesiek, S., Kuboniwa, H., Hinck, A. P., and Bax, A. (1995) Multiple-quantum line narrowing for measurement of Ha-H $\beta$  J coupling in isotopically enriched proteins. *J. Am. Chem. Soc.* 117, 5312–5315.

(38) Grzesiek, S., Vuister, G. W., and Bax, A. (1993) A simple and sensitive experiment for measurement of  $J_{\text{CC}}$  couplings between backbone carbonyl and methyl carbons in isotopically enriched proteins. *J. Biomol. NMR* 3, 487–493.

(39) Vuister, G. W., Wang, A. C., and Bax, A. (1993) Measurement of three-bond nitrogen-carbon J couplings in proteins uniformly enriched in  $^{15}\text{N}$  and  $^{13}\text{C}$ . *J. Am. Chem. Soc.* 115, 5334–5335.

(40) Pelton, J. G., Torchia, D. A., Meadow, N. D., and Roseman, S. (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated III<sup>Glc</sup>, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques. *Protein Sci.* 2, 543–558.

(41) Ogura, K., Terasawa, H., and Inagaki, F. (1996) An improved double-tuned and isotope-filtered pulse scheme based on a pulsed field gradient and a wide-band inversion shaped pulse. *J. Biomol. NMR* 8, 492–498.

(42) Zwahlen, C., Legault, P., Vincent, S. J. F., Greenblatt, J., Konrat, R., and Kay, L. E. (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *J. Am. Chem. Soc.* 119, 6711–6721.

(43) Altieri, A. S., Hinton, D. P., and Byrd, R. A. (1995) Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *J. Am. Chem. Soc.* 117, 7566–7567.

(44) Güntert, P. (2004) Automated protein structure calculation with CYANA. *Methods Mol. Biol.* 278, 353–378.

(45) Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S., and Kollman, P. A. (2004) AMBER 8, University of California, San Francisco.

(46) Xia, B., Tsui, V., Case, D. A., Dyson, H. J., and Wright, P. E. (2002) Comparison of solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model and with explicit water. *J. Biomol. NMR* 22, 317–331.

(47) Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486.

(48) Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B. S., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossvary, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui,

C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2010) *AMBER 11*, University of California, San Francisco.

(49) Lee, B. M., Xu, J., Clarkson, B. K., Martinez-Yamout, M. A., Dyson, H. J., Case, D. A., Gottesfeld, J. M., and Wright, P. E. (2006) Induced fit and "Lock and Key" recognition of 5 S RNA by zinc fingers of transcription factor IIIA. *J. Mol. Biol.* 357, 275–291.

(50) Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

(51) Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* 31, 1647–1651.

(52) Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, Rance, M., and Skelton, N. J. (2007) *Protein NMR Spectroscopy: Principles and Practice*, 2nd ed., Elsevier Academic Press, Burlington, MA.

(53) Fu, Y., Ramisetty, S. R., Hussain, N., and Baranger, A. M. (2012) MBNL1-RNA recognition: contributions of MBNL1 sequence and RNA conformation. *ChemBioChem* 13, 112–119.

(54) Fuxreiter, M., and Tompa, P. (2009) Fuzzy interactome: the limitations of models in molecular biology. *Trends Biochem. Sci.* 34, 3–3.

(55) Yuan, Y., Compton, S. A., Sobczak, K., Stenberg, M. G., Thornton, C. A., Griffith, J. D., and Swanson, M. S. (2007) Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.* 35, 5474–5486.

(56) Ogura, K., Terasawa, H., and Inagaki, F. (1996) Fully <sup>13</sup>C-refocused multidimensional <sup>13</sup>C-edited pulse schemes using broadband shaped inversion and refocusing pulses. *J. Magn. Reson., Ser. B* 112, 63–68.

(57) Grammatikakis, I., Goo, Y. H., Echeverria, G. V., and Cooper, T. A. (2011) Identification of MBNL1 and MBNL3 domains required for splicing activation and repression. *Nucleic Acids Res.* 39, 2769–2780.