Published in final edited form as:
Biochemistry. 2008 July 8; 47(27): 7284–7294. doi:10.1021/bi702252d.

RNA Binding Specificity of *Drosophila* Muscleblind†

Emily S. Goers, **Rodger B. Voelker**, **Devika P. Gates**, and **J. Andrew Berglund***
Department of Chemistry and Institute of Molecular Biology, 1229 University of Oregon, Eugene, OR 97403

Abstract

Members of the muscleblind family of RNA binding proteins found in *Drosophila* and mammals are key players in both the human disease myotonic dystrophy and the regulation of alternative splicing. Recently, the mammalian muscleblind-like protein, MBNL1, has been shown to have interesting RNA binding properties with both endogenous and disease-related RNA targets. Here we report the characterization of RNA binding properties of the *Drosophila* muscleblind protein Mbl. Mutagenesis of double-stranded CUG repeats demonstrated that Mbl requires pyrimidine-pyrimidine mismatches for binding and that the identity and location of the C-G and G-C base pairs within the repeats are essential for Mbl binding. Systematic evolution of ligands by exponential enrichment (SELEX) was used to identify RNA sequences that bind Mbl with much higher affinity than CUG repeats. The RNA sequences identified by SELEX are structured and contain a five-nucleotide consensus sequence of 5′-AGUCU-3′. RNase footprinting of one of the SELEX RNA sequences with Mbl showed that Mbl binds both double-stranded and single-stranded regions of the RNA. Three guanosines show the strongest footprint in the presence of Mbl; mutation of any of these three guanosines eliminates Mbl binding. It was also found that Mbl specifically bound a human MBNL1 RNA target, demonstrating the conservation of the muscleblind proteins in recognizing RNA targets. Our results reveal that Mbl recognizes complex RNA secondary structures.

Drosophila muscleblind (Mbl) is an RNA binding protein important in differentiation and development. Mbl is known to have roles in photoreceptor differentiation as well as terminal muscle differentiation (1,2). In addition to roles in differentiation, Mbl is necessary for early development as Mbl null flies die at late embryo stages or early larval stages (2). Mbl has recently been shown to regulate the alternative splicing of two genes involved in the cytoskeleton, α-*actinin* and *tun* (3,4). The human orthologues of Mbl are the muscleblind-like proteins (MBNL1-3) and these proteins also regulate alternative splicing of transcripts important in development (5,6). Alternative splicing of the Mbl transcript leads to the production of at least four different Mbl protein isoforms ranging in size from 84 to 316 amino acids. Three Mbl isoforms presumably confer RNA binding through two $CX_7CX_6CX_3H$ zinc fingers; the shortest Mbl isoform contains only one zinc finger.

The MBNL proteins have been shown to play an important role in myotonic dystrophy (DM) in humans. There are two subtypes of DM, both involving repeat expansions that result in toxic RNAs. Myotonic dystrophy type 1 (DM1) is caused by an expansion of CTG repeats in the 3′-UTR of the *DMPK* gene (7-9), while myotonic dystrophy type 2 (DM2) is caused by an expansion of CCTG repeats in the first intron of the *ZNF9* gene (10). These two types of expansions appear to function through a common disease mechanism: upon transcription, both the CUG and CCUG repeats fold into double-stranded stem-loops with the helices adopting a primarily A-form structure (11-13). MBNL proteins bind these stem-loop structures leading

†This work was supported by NIH grant (AR053903) to J.A.B.

*Email: aberglund@molbio.uoregon.edu Phone: 541-346-5097 Fax: 541-346-5891

to the sequestration of these important splicing factors in nuclear foci (14-16). MBNL sequestration results in mis-splicing of various transcripts, several of which have been linked to symptoms observed in DM patients (reviewed in refs 5 and 6).

Approximately 20 different human pre-mRNAs have been shown to be mis-spliced in DM (6). The most well-characterized example of mis-splicing in DM is the cardiac troponin T (cTNT) pre-mRNA in which MBNL1 has been shown to promote exclusion of an alternatively spliced exon (17). cTNT is developmentally regulated, and the presence or absence of this exon determines if the expressed protein is the fetal or adult isoform, respectively (18-21). Incorrect cTNT splicing in DM1 patients may be linked to cardiac abnormalities, one of the symptoms of DM1 (5). The binding site for MBNL1 in the cTNT pre-mRNA has been mapped to a region upstream of the skipped exon (17). Recently, we have shown that this section of cTNT RNA folds into a stem-loop with mismatches and adopts a structure similar to the CUG repeat RNA which contains U-U mismatches (22). Yuan and colleagues have also shown that MBNL1 binds a stem-loop containing a pyrimidine-pyrimidine mismatch in an intron near the Tnnt3 fetal exon (23), suggesting MBNL recognizes structured RNA elements in its pre-mRNA targets.

Drosophila has been used as a model to investigate several aspects of DM1 pathology. *Drosophila* overexpressing 162 CUG repeats have nuclear foci that contain localized Mbl and CUG repeats but do not display any abnormal phenotypes (24). Longer, interrupted CUG repeats (480 repeats) result in nuclear foci and cause the flies to have muscle, eye, and other degeneration phenotypes, consistent with the Mbl null flies (1,25). Overexpression of human MBNL1 can rescue these mutant phenotypes (26), indicating MBNL1 and Mbl are functionally conserved. Since *Drosophila* has one *mbl* gene with four protein isoforms (while humans have three muscleblind-like genes, each with multiple isoforms) experimental design and data analysis can be simplified when using *Drosophila* as a model system. Characterization of the RNA binding properties of *Drosophila* Mbl will lead to a better understanding of its role in alternative splicing and the role of MBNL in DM. In this study we show that recombinant Mbl (amino acids 1-105) is a highly specific RNA binding protein with a strong preference for structured RNA containing specifically placed guanosines.

MATERIALS AND METHODS

Protein Construct Design and Protein Purification

Mbl cDNA (amino acids 1-105) was amplified by PCR from the *Drosophila* cDNA library DGC release 2 using primers: 5′-GCGAATTCGCCAACGTTGTCAATATGAACAGC-3′ and 5′-ATAGTTATGCGGCCGCTCAGCCCATCTGTTGCATCAGTGC-3′ and restriction digested with EcoR1 and Not1. The PCR product was ligated into pGEX6P-1 (Amersham) to create the Mbl (1-105) construct, which has a N-terminal glutathione S-transferase (GST) tag. This construct was expressed in *Escherichia coli* Rosetta cells and grown in LB medium with 50 μM $ZnCl_2$, 50 μg/mL Amp, 25 μg/mL chloramphenicol, and 2% glucose at 37°C. Cells were induced at an OD_{600} of 0.5-1.0 with 0.4 mM IPTG at 30°C for 3 h. Cells were then pelleted and stored at −80°C. The pellet was dissolved in lysis buffer (500 mM NaCl, 25 mM Tris pH 8.0, 10 mM BME and 5% glycerol), sonicated, and centrifuged at 31,000 × g for 15 min. The supernatant was bound to glutathione-agarose beads for 30 min to 1h at 4°C. The protein was eluted with 10 mM reduced glutathione, diluted to 60 mM NaCl, and run over a Source 30Q anion-exchange column (Amersham). Protein impurities bound to the column, but Mbl did not; therefore, flow-through was collected, concentrated, and dialyzed into protein storage buffer (50% glycerol, 25 mM Tris pH 7.5, 5 mM BME, 0.1 % triton x-100 and 500 mM NaCl) overnight at 4°C. The presence of the GST tag does not appear to alter RNA binding because removing GST did not alter the RNA binding of Mbl (data not shown).

Transcription and Kinase Reactions

RNA oligonucleotides were transcribed with T7 RNA polymerase and (α-^{32}P) CTP from linearized plasmids, PCR products or fully complementary DNA oligos. The transcription reaction mixtures were incubated at 37°C for 3 h or 25°C overnight.

RNAs were kinased using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P] ATP. Both transcribed and kinased RNAs were gel purified with 8% denaturing polyacrylamide gels.

Gel Mobility Shift Assay

Radiolabeled RNA was heated at 95°C for 3 min and then placed directly on ice for 10 min in binding buffer (20 mM Tris pH 7.5, 5 mM $MgCl_2$ and 100 mM NaCl). RNA was equilibrated to 25°C for 5 min. Heparin (final concentration of 0.5 mg/mL) and loading dye were added to the RNA. Protein was serially diluted into protein storage buffer on ice and aliquoted into 2 μL per binding reaction mixture at 25°C. RNA in binding buffer (8 μL) was added to the aliquoted protein and allowed to bind for 15-30 min at 25°C. The RNA concentration was equal to or less than the dissociation constant (K_d) value for each reaction. RNA-protein complexes (2 μl per binding reaction mixture) were resolved with a native polyacrylamide gel (6% 37.5:1 acrylamide/bisacrylamide mixture and 0.5X TB) run at 4°C for 30 min at ~150 V. The gels were dried and exposed to a phosphoimager screen (Molecular Dynamics) overnight at 25°C.

K_d values were determined on the basis of quantitation of the bound and unbound RNA fractions using ImagQuant (Molecular Dynamics). K_d values were calculated as described by Warf and Berglund (22).

SELEX

The template for the SELEX (systematic evolution of ligands by exponential enrichment) (27) experiment was the 80-mer DNA oligonucleotide 5′-GGGAATGGATCCACATCTACGAATTCN$_{30}$TTCACTGCAGACTTGACGAAGCTT-3′. The forward primer for SELEX was 5′-GATAATACGACTCACTATAGGGAATGGATCCACATCTACGA-3′ and the reverse primer was 5′-AAGCTTCGTCAAGTCTGCAGTGAA-3′. The initial PCRs were amplified by eight rounds of PCR and calculated to a concentration of 1×10^{14} molecules. The transcription reactions were conducted under the following conditions: 100-2000 ng of template, 40 mM Tris pH7.9, 26 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, nucleotides (5 mM each), 0.1 μg of yeast pyrophosphatase/100 μL, ~2 mg/mL T7 polymerase, and 40 mM DTT. The transcription reactions were performed with small amounts of [α-^{32}P] CTP for monitoring purposes. After transcription, reactions mixtures were treated with DNase for 1 hr at 37°C, gel purified, ethanol precipitated and run over a Bio-Spin-6 size exclusion column (Bio-Rad). For each round, the appropriate concentration of Mbl (1-105) was bound to 20 μL of glutathione-agarose beads at 4°C for 15 min, then washed three times with 200 μL of SELEX binding buffer (160 mM NaCl, 25 mM Tris pH 7.5, 5 mM $MgCl_2$, 0.002% triton x-100 and 1 mM BME added the day of use). This amount of beads should be enough to completely bind all of the added Mbl. All washes were conducted in a similar manner. The RNA was heated in SELEX binding buffer at 95°C for 3 min and placed immediately on ice for 10 min. tRNA (0.1 mg/mL) was added, as a nonspecific inhibitor, to the RNA (in all rounds except round 3). Negative selection was carried out on glutathione-agarose beads prior to binding to Mbl-bound beads by prebinding the RNA to the glutathione-agarose beads. The Mbl-bound beads were incubated with the RNA at 25°C for 20 min. The beads were washed three times in 200 μL of SELEX binding buffer and RNA was released and collected from the beads by a phenol/chloroform extraction and subsequent ethanol precipitation. The collected RNA was reverse transcribed using AMV reverse transcriptase, AMV buffer, 8.8 μM reverse primer, and two-thirds of the RNA isolated after the round for 1 hr at 42°C. After the reverse transcription was

complete, the DNA was amplified by 11 PCR cycles. The DNA was transcribed and the SELEX cycle repeated six times with the concentration of RNA and Mbl varying according to Table 1. After the rounds were completed, individual clones were isolated by TOPO cloning and then sequenced.

Structure Probing and Footprinting

RNA was transcribed, and the terminal triphosphate was removed using shrimp alkaline phosphatase (USB) and then kinased using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]ATP to 5′-end label the RNA. The RNA was gel purified and ethanol precipitated. Purified RNA was incubated with either RNase T1, RNase V1, or RNase 1 or under alkaline hydrolysis conditions (Ambion) in footprinting buffer and protein storage buffer (20% glycerol, 0.04% triton X-100, 2 mM BME, 20 mM Tris pH 7.5, 200 mM NaCl, 100 mM KCl, 0.5 mg/mL tRNA and 10 mM $MgCl_2$) to a final volume of 10 μL for 3 min. Reactions were quenched with phenol-chloroform and extracted. In the reaction mixtures containing Mbl, Mbl was added in protein buffer to a final concentration of 4 or 12 μM and incubated with RNA for 10 min at 25°C before RNases were added. After the phenol extraction, the RNA was ethanol precipitated and cleavage products were resolved with a denaturing 15% polyacrylamide gel for ~2 h at 35 W. The gel was dried and exposed to a phosphoimager screen overnight at 25°C.

Determination of n-mer Bias in the SELEX Sequences

SELEX was initiated with a population of RNAs containing a uniformly random sequence; however, the random sequence was flanked by two constant regions that are necessary for amplification. It is possible that the nonrandom sequences can inadvertently influence the selection process (e.g., by directly acting as a binding site or by interacting with the nonrandom region to form secondary structures that influence binding). Therefore, we developed a method for calculating enrichment that attempts to account for the constant regions. To calculate the degree to which *n*-mers were enriched during the SELEX, we first performed a sliding-window count of all *n*-mers (3 to 5 nt) within the non-redundant sample of 25 selected sequences. This count included the entire constant and random regions. To calculate the maximum likelihood probabilities for expected occurrences we also performed a similar sliding-window count on a set of 500,000 computer generated sequences containing a uniformly random 30 nt region flanked by the same constant regions used in the experiment. For both of these data sets the counts were transformed to probabilities and the enrichment was determined according to the binomial confidence interval method previously described (28).

RESULTS

A Minimal Mbl Protein Binds Specifically to Expanded CUG and Short CUG Repeats

To characterize the RNA binding specificity and affinity of Mbl, we have designed a minimal protein – RNA system. Initially we began our RNA binding studies with a full-length version of Mbl (amino acids 1-243), isoform C of the four different alternative splice forms of Mbl. However, due to its low level of solubility and instability, Mbl (1-243) proved difficult for use in binding studies. The first 100 amino acids are highly conserved and contain the zinc finger domains. We therefore truncated Mbl by removing the C-terminus and fused the N-terminal 105 amino acids to an N-terminal GST tag, creating a more stable protein. This construct will be referred to as Mbl and all experiments were performed with Mbl unless otherwise indicated.

Mbl bound to fifty four CUG repeats (CUG_{54})with a K_d of 500 nM (Figure 1A) while a minimal RNA containing only four CUG repeats with a UUCG cap (CUG_4) bound Mbl with a K_d of 430 nM. (Figure 1A). The UUCG cap is an ultra-stable loop that forces stems to form and maintain stability (29). Normally 10 or more CUG repeats are necessary for duplex formation

and human MBNL1 binding (16). This minimal $(CUG)_4$ RNA facilitates the dissection of the RNA binding specificity of Mbl by replacing the U-U mismatches and/or C-G base pairs with other bases. Using this strategy we measured the binding affinities of twelve $(CUG)_4$ mutants. We found that when the U-U mismatches were replaced with C-U or C-C mismatches, Mbl bound with only a 2-fold decrease in affinity compared to $(CUG)_4$ (Figure 1B). However, mutating the U-U mismatches to G-U wobble base pairs or purine-purine mismatches such as A-G, A-A or G-G abolished binding (Figure 1B). Replacing the U-U mismatches with Watson-Crick base pairs also disrupted binding by Mbl (RNA #7 and #8, Figure 1B). Furthermore, changes to the C-G and G-C Watson-Crick base pairs including reversing the orientation of the RNA (5′-GUC-3′) and making one strand all pyrimidines and the other all purines eliminated binding by Mbl (Figure 1B). Finally, an RNA with different Watson-Crick base pairs at all of the positions didn't bind Mbl (RNA #10, Figure 1B). These results indicate that all of the positions of the CUG helix are important for binding, and that the pyrimidine-pyrimidine mismatch is essential for binding.

Identification of RNA Sequences That Bind with High Affinity to Mbl

SELEX was performed using Mbl and a random pool of 10^{14} RNA oligonucleotides. Each RNA consisted of a 30-nucleotide randomized region flanked by shorter constant regions, creating an 80-nucleotide RNA. Twenty-five unique RNA sequences were identified (five sequences were found multiple times) after six rounds of SELEX. The sequences were classified into three groups according to the presence of enriched motifs, binding affinities to Mbl and predicted secondary structures (Table 2 and Figure 2).

An *n*-mer analysis comparing the expected occurrences in a sample of uniformly random RNA sequences to the 25 SELEX sequences identified several highly enriched *n*-mers (Table 2). The *n*-mers were found predominately in groups I and II and rarely in group III with AGUCU identified as the most highly enriched 5-mer. Most of the high scoring *n*-mers were located within two main regions of the SELEX sequences; UGUG, GUGCG and CGGUA were 5′ of the AGUCU sequence in Group I while AGUC was found within the AGUCU sequence. Almost all group I SELEX sequences fold into a common structure (similar to RNA 20, figure 2B) of two stems connected by a linker of one to three nucleotides. The location within the predicted secondary structures for the *n*-mers found 5′ of AGUCU was generally in the linker between stem 1 and stem 2 or the base of stem 2. It is possible that these sequences were selected during SELEX to promote specific secondary structures.

Group I and II RNA sequences bound Mbl with high affinity (K_d values ranged from 0.050 to 1.0 nM, Figure 3), while group III RNA sequences bound Mbl weakly or not at all (Figure 3). Group I and II RNAs both contained the highly enriched sequence, 5′-AGUCU-3′, in the random region that base pairs to a region in the 3′-constant region. In group I, this consensus sequence was located at the 3′-terminus of the randomized region, while in group II the sequence was located at the 5′-terminus of the randomized region (Figure 2A). Occasionally, the consensus sequence spanned the junction of the randomized region and the 3′-constant region (for the group I sequences). For group I sequences, the reverse complement of the consensus sequence (5′-AGUCU-3′) was found six to nine nucleotides downstream from the consensus sequence. For sequences in Group II, the reverse complement of the consensus sequence was found 32 nucleotides downstream in the constant region. Regardless of the distance between the consensus sequence and its reverse complement, the minimal free energy structures (according to mfold) predicted that the 5′-AGUCU-3′/5′-AGACU-3′ sequences were paired. The double-stranded consensus region (5′-AGUCU-3′/5′-AGACU-3′) was found in all SELEX RNA sequences that Mbl bound tightly and corresponded to regions of common structural aspects across Group I and II (Figure 3A,B), suggesting this motif is a core component for Mbl binding.

The Secondary Structure and Mbl Binding Site on a SELEX Group I RNA

Structure probing and footprinting were performed to determine if the secondary structure of the group I RNA sequences (predicted by mfold 30) was correct and to identify the Mbl binding site. RNA 20 was chosen for this characterization because, of the group I RNAs tested, it bound the tightest (Figures 2A and 3A). The structure probing with RNase T1, RNase V1 and RNase 1 revealed that the mfold structure prediction was consistent with the experimentally-determined structure (compare Figures 2B and 4C). The results of the RNase T1 cleavage, which is specific for single-stranded guanosines, are shown in panels A and B of Figure 4 (lane 3). Cleavage sites were identified at G1, G2, G8, G21, G39, G40, G43, G45, G47, G48, G51, G53, G63, G66, G71, G74, and G77. Lane 6 (Figure 4A,B) shows the results of RNase V1, which preferentially cleaves in double-stranded regions or regions adjacent to double-stranded RNA. The cleavages by RNase V1 were seen in stem 1A (nucleotides 11-22), stem 1B (nucleotides 31, 36, 39, 40, and 42), the base of stem 2A (nucleotides 46 and 47), and stems 2A and 2B (nucleotides between positions 53 and 78 display cleavage; however, the resolution of the gel does not allow discrete nucleotides to be identified). RNase 1, which preferentially cleaves in single-stranded regions, cleaved the 5′-tail (nucleotides 1-4), loop 1 (nucleotides 22-27), the base of stem 1 (nucleotide 40), the linker between stem 1B and stem 2A (nucleotides 41-43), loop 2 (nucleotides 54-62) and a few nucleotides in the 5′-tail (nucleotides 76 -78 although these are difficult to distinguish). Several nucleotides (including U42, G43, G47, and G48) are cleaved by more than one RNase, indicating either that region of the RNA is not statically single- or double-stranded or that several structural isoforms exist. Furthermore, it is not obvious why different levels of cleavage by RNase V1 in stem 1A and stem 1B were observed; it is possible that the tertiary structure of RNA 20 allows some regions of stem 1 to be more accessible to cleavage than other regions.

The addition of Mbl significantly altered the cleavage pattern at three positions. All three nucleotides are guanosines, which were strongly protected from RNase T1, RNase V1, and RNase 1 activity. G53 was protected from RNase V1 by Mbl at both concentrations of Mbl, 4 μM and 12 μM (Figure 4B, lanes 7 and 8). G63 was protected from RNase T1 cleavage at both Mbl concentrations as well. G77 was protected from RNase T1 cleavage but only slightly protected from RNase V1 and RNase 1 by Mbl. This protection indicates Mbl binds RNA 20 within stem-loop 2 and the 3′-tail (Figure 4C). No protection was observed in stem 1 or the 5′-tail, which is consistent with the binding information used to group the SELEX RNA sequences. G53 is in the 5′-AGUCU-3′ consensus motif, implying Mbl directly binds this particular guanosine in stem 2.

In the SELEX sequences, guanosines made up 33% of all bases while adenosine, uridine, and cytosine made up 22, 27, and 18%, respectively, of the base composition of the randomized region after six rounds of SELEX. These results suggest Mbl selected for guanosine-rich sequences and against cytosine-rich sequences while adenosines and uridines were neutral. The selection against cytosines is surprising because Mbl and MBNL bind CUG and CCUG sequences. The bias toward guanosines in the SELEX sequences could be the result of selecting for structured, stable RNAs within a relatively short randomized region due to the fact that guanosine can bind both cytosine and uridine.

Truncation and Mutational Analysis Identifies Features of SELEX RNA 20 That Are Essential for Binding by Mbl

To determine a minimal binding site within SELEX RNA 20, binding studies were conducted with the 5′- and 3′-halves of the RNA, dividing it between nucleotides 44 and 45 to maintain stem 1 (stem-loop 1 half) and stem 2 (stem-loop 2 half). The stem-loop 2 half RNA contained an additional three guanosines at the 5′-end that was necessary for transcription using T7 RNA polymerase. No binding was observed for stem-loop 1 half RNA and weaker binding was

observed for stem-loop 2 half RNA (~1000-fold lower affinity than that of full length RNA 20) (Figure 5). This is consistent with the Mbl footprinting to stem 2, but the significant loss of binding with just stem-loop 2 half compared to the full-length RNA indicates that at least a portion of stem-loop 1 half aids Mbl binding. Many of the highly enriched *n*-mers (Table 2) are located at the base of stem 1 and the linker between stem 1 and stem 2 which were disrupted in the stem-loop 1 half and stem-loop 2 half truncations. To identify the 5′-region that aids in Mbl binding, we began deleting sections of the RNA and found that removing the majority of the eight nucleotide 5′-tail abolished Mbl binding, indicating that the 5′-tail is required for binding (truncated 5′-tail, Figure 5). When a different RNA tail (six of eight nucleotides in the 5′-tail were mutated) was added back, binding was partially restored (K_d value of 140 nM for mutated 5′-tail RNA compared to 0.23 nM for full-length RNA 20, Figure 5), indicating that both the presence of the tail and its sequence are important for Mbl binding.

Not surprisingly, deleting the six nucleotide 3′-tail abolished Mbl binding; one of the guanosines that Mbl protected is within this region (Figure 5). To determine if the 5′- and 3′-tails from the SELEX RNA 20 could generally enhance Mbl binding, we added these tails to the $(CUG)_4$ RNA and tested Mbl binding. Apparently, the tails are specific to the SELEX RNA because the addition of the tails to the CUG repeats did not increase the level of binding (data not shown). Next we reduced the length of stem 1, maintaining the first four base pairs at the base of this stem and capping it with an ultrastable UUCG sequence. Tight binding was maintained with no loss in affinity compared to the full-length RNA 20 (shortened stem-loop 1 RNA, Figure 5). These binding studies indicate that both the 3′ and 5′ tails are required for binding by Mbl, but that a minimal stem can replace the extended stem 1 and maintain high-affinity binding. This suggests that the 5′-tail, 3′-tail and stem 1 are important for the overall structure of the RNA and Mbl binding.

To test the importance of the guanosines protected by Mbl, each was mutated individually to adenosine. Each G to A mutation severely inhibited or eliminated Mbl binding (Figure 6). G63 and G77 were in single-stranded regions that, when mutated, did not affect the predicted structure. G53 and C68 were both mutated to form a new U-A base pair to maintain the secondary structure of RNA 20; however mfold predicted a change in structure. To determine if nucleotides near these guanosines are important for binding, we introduced two additional single mutations. C59 was mutated from a cytosine to adenosine (data not shown) and A75 from an adenosine to a guanosine (Figure 6, bottom right panel). Neither mutation affected Mbl binding. Our results suggest that G53, G63, and G77 are positioned within the secondary structure in RNA 20 in such a way as to facilitate high-affinity binding by Mbl.

Although the G53A mutation suggested this specific nucleotide was important for Mbl binding, we wanted to determine if Mbl would still bind with high affinity if the structure of stem 2 was maintained but the sequence altered. This was accomplished by changing the 6 bp stem 5′-G_{51}AGUCU-3′ / 5′-AGACUU-3′ to 5′-C_{51}CACAC-3′ / 5′-GUGUGG-3′ to maintain the structure but not the sequence (2A/2B-6, Figure 6). The affinity did not decrease compared to that of RNA 20, indicating that the stem structure was selected for and not necessarily the sequence, although Mbl could be contacting a different guanosine in this new sequence.

The lack of obvious pyrimidine-pyrimidine mismatches in SELEX RNA 20 was initially a puzzling result. We hypothesized that loop 2 might contain one or more pyrimidine-pyrimidine mismatches and function in a manner analogous to that of the mismatches found in the MBNL1 binding sites. To determine if the U57–C64 mismatch was important for Mbl binding, U57 was mutated to a guanosine (U57G) to form a G-C base pair (G57-C64). The level of binding to U57G was reduced 30-fold compared to that of RNA 20 suggesting that this mismatch plays a role in Mbl binding (Figure 6).

Mbl Recognizes an MBNL1 Binding Site in the cTNT Pre-mRNA with High Affinity

MBNL1 binds a stem-loop in the cTNT pre-mRNA and regulates the splicing of the adjacent exon through this interaction (17,22). To determine if Mbl binds this MBNL1 pre-mRNA target, we performed binding studies with this cTNT RNA. A 50-nucleotide region from within intron 4 of the cTNT pre-mRNA was found to form a stem-loop that contains several pyrimidine-pyrimidine mismatches bordered by G-C and G-U base pairs (Figure 7A). Mbl binds this RNA with a K_d value of 0.67 nM, an affinity similar to that of the Mbl SELEX RNA sequences (Figure 7B). Mutating four of the guanosines to pyrimidines eliminates binding at the Mbl concentrations that were tested (Figure 7B). These mutations reduce the level of MBNL1 binding by 100-fold (22), while Mbl appears to display even more specificity (>500-fold). This result shows that Mbl is capable of recognizing human MBNL1 sites and suggests Mbl may recognize similar RNA motifs during splicing regulation in *Drosophila*.

DISCUSSION

This first characterization of the RNA binding properties of Mbl shows that the protein is a very specific and high-affinity RNA binding protein. Mbl binds CUG repeats as expected (Figure 1) on the basis of the findings that, like MBNL1, Mbl colocalizes with CUG repeats in *Drosophila* (24). The high specificity of Mbl for the CUG repeat RNA is surprising. We previously found that MBNL1 tolerated many different mutations to the CUG repeats (22), while Mbl only tolerated replacement of the U-U mismatch with other pyrimidine-pyrimidine mismatches (Figure 1). This indicates that Mbl is a more specific RNA binding protein than MBNL1.

Using SELEX, we identified RNAs that bound Mbl with K_d values ranging from the picomolar to the low nanomolar range. All of these RNAs contain the consensus motif of 5′-AGUCU-3′ (Figure 2). The footprinting studies suggest Mbl interacts directly with the guanosine in this sequence, and mutating this nucleotide eliminated Mbl binding (Figures 4 and 6); however, the importance of this sequence is partly or fully due to the fact that it base pairs with a region (5′-AGACU-3′) in the constant region. This was demonstrated by changing the sequence of the stem 2 and finding that the affinity of this RNA for Mbl was the same as that for RNA 20 (Figure 6). It is possible that Mbl interacts with a different guanosine in this new sequence, although it would be in a different location in the helix. Interestingly, the 5′-AGUCU-3′ sequence and the putative MBNL1 consensus sequence (YGCUUY) (17) are similar in that both contain a guanosine followed by a run of pyrimidines. The major difference between the motifs is the presence of the adenosine in the SELEX motif. Also importantly, the SELEX 5′-AGUCU-3′ motif is involved in all Watson-Crick base pairs, while the YGCUUY motifs in cTNT, when base paired, contain both Watson-Crick base pairs and pyrimidinepyrimidine mismatches (22). However, in the SELEX RNA 20, a pyrimidine-pyrimidine mismatch 3′ to the 5′-AGUCU-3′ motif (U57-C64) appears to be important for binding because if the mismatch is replaced with a G-C base pair, the level of binding is reduced 30-fold (Figures 6 and 8).

Although stem-loop 2 is clearly important for Mbl binding, the 5′- and 3′-single-stranded tails are also important for binding (Figure 5). These tails could be involved in tertiary interactions with other regions of the RNA that are difficult to monitor with gel mobility shift assays and footprinting. It is possible that the G-rich regions of the 5′-tail interact with loop 1 or the 3′-tail and aid in loading of Mbl onto the RNA or proper folding of the RNA, and upon removal, these interactions are lost. Apparently, the tails are specific to the SELEX RNA because the addition of the tails to the CUG repeats did not enhance binding (data not shown). The 50-nucleotide cTNT RNA also contains 5′-and 3′-tails that could be aiding in the formation of the correct RNA structure recognized by Mbl.

A comparison of the SELEX RNA 20 with the human cTNT RNA reveals a site that the two RNAs may share. As shown in Figure 8, the loop and upper portion of stem-loop 2 of RNA 20 and the upper portion of the cTNT RNA are quite similar in sequence and structure, assuming RNA 20 and cTNT adopt similar three dimensional structures. Both of these RNAs contain two pyrimidine-pyrimidine mismatches (positions 1 and 3) with a potential U-G wobble base pair between them (position 2). Below the position 3 mismatches are two additional base pairs, with position 5 being a C-G base pair in both RNA 20 and the cTNT structures. MBNL1 has also been shown to bind a structured RNA site within the mouse Tnnt3 pre-mRNA and regulate splicing through this site (23). Mutational analysis revealed the recognition site for MBNLI as being a stem-loop structure within a larger RNA structure. This stem-loop also contains a pyrimidinepyrimidine mismatch but lacks the sequence similarity to SELEX RNA 20 and the cTNT RNA (Figure 8). The comparison of these three RNAs indicates Mbl and MBNL1 bind structured RNA stem-loops and that there is significant flexibility in the composition of the sequences, which makes it challenging to identify a consensus sequence within the context of a stem or stem-loop.

We speculate that stem-loop structures containing correctly positioned guanosines and pyrimidine-pyrimidine mismatches will be found as the binding sites in Mbl's endogenous targets such as *tun* and α-*actinin* pre-mRNA. These transcripts were recently shown to be mis-spliced in Mbl mutant flies (3). Vicente and colleagues (4) proposed a CUG-rich region downstream of a regulated exon in the *α-actinin* pre-mRNA as a potential binding site for Mbl. This region can potentially fold into a stem-loop containing pyrimidine-pyrimidine mismatches (4). In preliminary studies, we have found a 115-nucleotide RNA, containing this CUG-rich region, binds Mbl with weaker affinity compared to CUG repeats, SELEX RNAs, and the cTNT RNA (data not shown). Although our study has improved our understanding of how Mbl binds RNA, it is clear that additional investigations of Mbl- and MBNL-RNA interactions are required to reach a point at which binding sites can be predicted.

Zinc finger proteins, other than muscleblind proteins, are capable of recognizing complex RNA structures. There are several classes of zinc finger proteins, including C_2H_2, CCCH, and CCHC, that bind RNA (31). Some zinc finger proteins are promiscuous and bind double-stranded DNA and single- and double-stranded RNA (32), while others, like Mbl, bind specifically to double-stranded RNA. An example of a zinc finger protein that binds RNA is TIS11d. TIS11d binds single-stranded RNA with a 5′-UAUU-3′ motif by making specific stacking interactions between aromatic side chains and the RNA bases (32). This mode of binding by TIS11d is probably different than Mbl's mode of binding due to its double-stranded RNA targets. However, the HIV nucleocapsid protein (NC) contains two zinc fingers of the $CX_2CX_4HX_4C$ type and has been found to recognize a variety of stem-loop structures from the HIV RNA (33,34). There are three exposed guanosines in the loop portion of stem-loop 3 from the HIV ψ-RNA recognition element that directly interact with multiple amino acids positioned by the zinc fingers of the HIV NC protein (35). The zinc fingers of Mbl may be performing a role similar to the role of the HIV NC zinc fingers in that the amino acids are positioned for specific recognition of the guanosines in the SELEX RNA sequences and endogenous RNA targets of Mbl.

The mutagenesis with the CUG repeats and the cTNT RNA revealed that Mbl is a specific RNA binding protein and is apparently more specific than MBNL1 since it displays a greater degree of specificity for both the CUG repeats and the cTNT RNA substrates. The primary difference between these two muscleblind proteins is that Mbl contains only two zinc fingers while MBNL1 contains four zinc fingers. One possibility is that the two additional zinc fingers of MBNL1 are nonspecific RNA binding domains, resulting in a less discriminate RNA binding protein. The third and fourth zinc fingers in MBNL1 appear to be related to the first two of Mbl and MBNL1 in sequence and could be the result of a duplication event (36). In future

studies, it will be interesting to determine if muscleblind proteins with two zinc fingers are generally more specific RNA binding proteins compared to muscleblind proteins with four zinc fingers. This minimal Mbl protein construct should prove useful for structural studies and in further dissecting the RNA binding abilities of this important protein as well as the identification of its binding sites in *Drosophila* RNAs.

ACKNOWLEDGEMENTS

We thank Amy Mahady for her assistance in analyzing the SELEX data. We also thank Bryan Warf for help with the CUG mutant screen and useful comparisons of binding data. We are grateful to Eric Johnson for the *Drosophila* cDNA and appreciate useful advice on the manuscript from Maurice Swanson, Mary Scanlan, and members of the Berglund laboratory.

ABBREVIATIONS

Mbl, muscleblind
DM, myotonic dystrophy
SELEX, systematic evolution of ligands by exponential enrichment
K_d, dissociation constant
Drosophila, *Drosophila melanogaster*
GST, glutathione s-transferase
G, guanosine
A, adenosine
C, cytosine
U, uridine
cTNT, cardiac troponin T
n-mer, oligonucleotide of length n

REFERENCES

1. Begemann G, Paricio N, Artero R, Kiss I, Perez-Alonso M, Mlodzik M. Muscleblind, a gene required for photoreceptor differentiation in Drosophila, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. Development 1997;124:4321–4331. [PubMed: 9334280]
2. Artero R, Prokop A, Paricio N, Begemann G, Pueyo I, Mlodzik M, Perez-Alonso M, Baylies MK. The muscleblind gene participates in the organization of Z-bands and epidermal attachments of Drosophila muscles and is regulated by Dmef2. Dev. Biol 1998;195:131–143. [PubMed: 9520330]
3. Machuca-Tzili L, Thorpe H, Robinson TE, Sewry C, Brook JD. Flies deficient in Muscleblind protein model features of myotonic dystrophy with altered splice forms of Z-band associated transcripts. Hum. Genet 2006;120:487–499. [PubMed: 16927100]
4. Vicente M, Monferrer L, Poulos MG, Houseley J, Monckton DG, O'Dell K,M, Swanson MS, Artero RD. Muscleblind isoforms are functionally distinct and regulate alpha-actinin splicing. Differentiation 2007;75:427–440. [PubMed: 17309604]
5. Ranum LP, Cooper TA. RNA-Mediated Neuromuscular Disorders. Annu Rev Neurosci. 2006
6. Osborne RJ, Thornton CA. RNA-dominant diseases. Hum Mol Genet 2006;15(Spec No 2):R162–169. [PubMed: 16987879]
7. Brook JD, Shaw DJ, Meredith L, Bruns GA, Harper PS. Localisation of genetic markers and orientation of the linkage group on chromosome 19. Hum Genet 1984;68:282–285. [PubMed: 6595199]
8. Fu YH, Pizzuti A, Fenwick RG Jr. King J, Rajnarayan S, Dunne PW, Dubel J, Nasser GA, Ashizawa T, de Jong P, et al. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science 1992;255:1256–1258. [PubMed: 1546326]
9. Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barcelo J, O'Hoy K, et al. Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. Science 1992;255:1253–1255. [PubMed: 1546325]

10. Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, Ranum LP. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science 2001;293:864–867. [PubMed: 11486088]
11. Napierala M, Krzyzosiak WJ. CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. J Biol Chem 1997;272:31079–31085. [PubMed: 9388259]
12. Dere R, Napierala M, Ranum LP, Wells RD. Hairpin structure-forming propensity of the (CCTG.CAGG) tetranucleotide repeats contributes to the genetic instability associated with myotonic dystrophy type 2. J Biol Chem 2004;279:41715–41726. [PubMed: 15292165]
13. Mooers BH, Logue JS, Berglund JA. The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. Proc. Natl. Acad. Sci. U S A 2005;102:16626–16631. [PubMed: 16269545]
14. Fardaei M, Rogers MT, Thorpe HM, Larkin K, Hamshere MG, Harper PS, Brook JD. 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]
15. Mankodi A, Urbinati CR, Yuan QP, Moxley RT, Sansone V, Krym M, Henderson D, Schalling M, Swanson MS, Thornton CA. Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. Hum. Mol. Genet 2001;10:2165–2170. [PubMed: 11590133]
16. Miller JW, Urbinati CR, Teng-Umnuay P, Stenberg MG, Byrne BJ, Thornton CA, Swanson MS. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. Embo J 2000;19:4439–4448. [PubMed: 10970838]
17. Ho TH, Charlet BN, Poulos MG, Singh G, Swanson MS, Cooper TA. Muscleblind proteins regulate alternative splicing. Embo. J 2004;23:3103–3112. [PubMed: 15257297]
18. Cooper TA, Ordahl CP. A single troponin T gene regulated by different programs in cardiac and skeletal muscle development. Science 1984;226:979–982. [PubMed: 6095446]
19. Briggs MM, Schachat F. Origin of fetal troponin T: developmentally regulated splicing of a new exon in the fast troponin T gene. Dev Biol 1993;158:503–509. [PubMed: 8344466]
20. 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]
21. Ryan KJ, Cooper TA. Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle. Mol Cell Biol 1996;16:4014–4023. [PubMed: 8754799]
22. 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 Oct 2007;17Epub ahead of print
23. Yuan Y, Compton SA, Sobczak K, Stenberg MG, Thornton CA, Griffith JD, Swanson MS. Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. Nucleic Acids Res 2007;35:5474–5486. [PubMed: 17702765]
24. Houseley JM, Wang Z, Brock GJ, Soloway J, Artero R, Perez-Alonso M, O'Dell KM, Monckton DG. Myotonic dystrophy associated expanded CUG repeat muscleblind positive ribonuclear foci are not toxic to Drosophila. Hum. Mol. Genet 2005;14:873–883. [PubMed: 15703191]
25. de Haro M, Al-Ramahi I, De Gouyon B, Ukani L, Rosa A, Faustino NA, Ashizawa T, Cooper TA, Botas J. MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1. Hum. Mol. Genet 2006;15:2138–2145. [PubMed: 16723374]
26. Monferrer L, Artero R. An interspecific functional complementation test in Drosophila for introductory genetics laboratory courses. J. Hered 2006;97:67–73. [PubMed: 16394256]
27. Conrad, RC.; Bruck, FM.; Bell, S.; Ellington, AD. In vitro selection of nucleic acid ligands. In: Smith, CWJ., editor. RNA:Protein Interactions. Oxford University Press; New York: 1998. p. 285-325.
28. Voelker RB, Berglund JA. A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. Genome Res 2007;17:1023–1033. [PubMed: 17525134]
29. Tuerk C, Gauss P, Thermes C, Groebe DR, Gayle M, Guild N, Stormo G, d'Aubenton-Carafa Y, Uhlenbeck OC, Tinoco I Jr. et al. CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. Proc Natl Acad Sci U S A 1988;85:1364–1368. [PubMed: 2449689]
30. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003;31:3406–3415. [PubMed: 12824337]

31. Hall TM. Multiple modes of RNA recognition by zinc finger proteins. Curr Opin Struct Biol 2005;15:367–373. [PubMed: 15963892]

32. Brown RS. Zinc finger proteins: getting a grip on RNA. Curr Opin Struct Biol 2005;15:94–98. [PubMed: 15718139]

33. Hagan NA, Fabris D. Dissecting the protein-RNA and RNA-RNA interactions in the nucleocapsid-mediated dimerization and isomerization of HIV-1 stemloop 1. J Mol Biol 2007;365:396–410. [PubMed: 17070549]

34. Lawrence DC, Stover CC, Noznitsky J, Wu Z, Summers MF. Structure of the intact stem and bulge of HIV-1 Psi-RNA stem-loop SL1. J Mol Biol 2003;326:529–542. [PubMed: 12559920]

35. De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. Science 1998;279:384–388. [PubMed: 9430589]

36. 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]

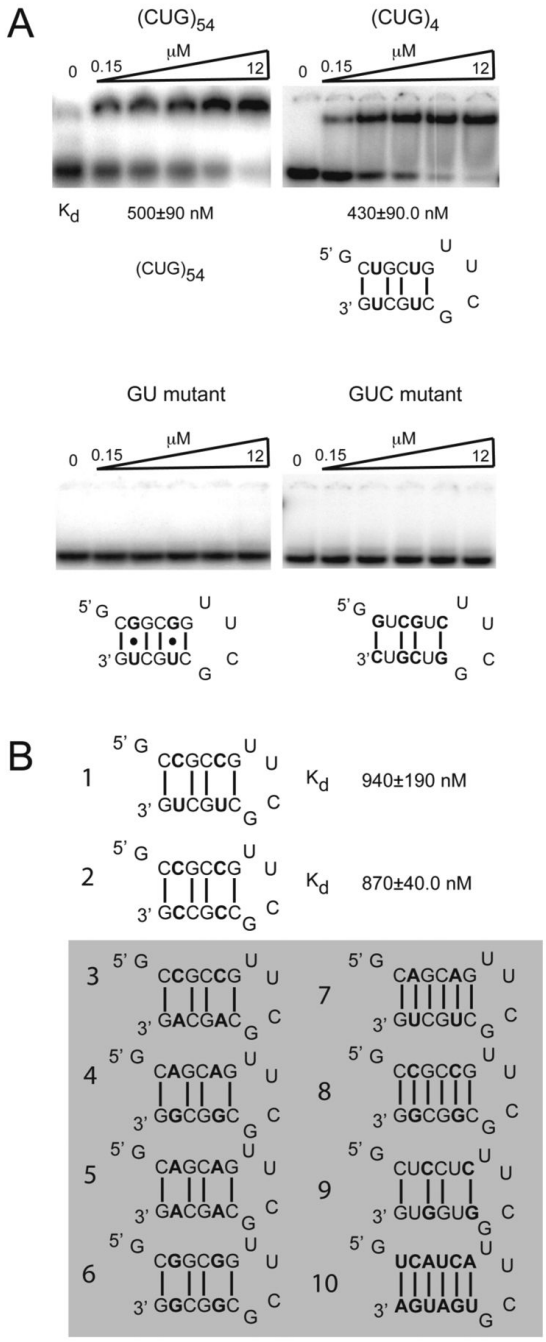

Figure 1.

Mbl binds CUG repeats with a high degree of specificity. (A) The four gels are gel mobility shift assays with the name of the RNA above each gel, the structure of the RNA below each gel, and increasing concentrations of Mbl along the top. The GU mutant gel contains $(CUG)_4$ with the U-U mismatches replaced with G-U base pairs. The GUC mutant has reversed 5′ to 3′ orientation compared to $(CUG)_4$. The nucleotides mutated relative to $(CUG)_4$ are shown in bold. The Mbl concentrations in panel A are 0.15 μM, 0.45 μM, 1.4 μM, 4.1 μM, and 12 μM. The K_d values refer to the RNA oligonucleotides shown in the gel above. (B) The secondary structures of 10 other CUG mutant RNA oligonucleotides that were tested for binding by Mbl. Only the C-U and CC mutants (1 and 2) bound to Mbl, while the other eight

RNA oligonucleotides (numbers 3-10, shaded in gray) showed no binding with up to 12 μM Mbl. The nucleotides mutated relative to $(CUG)_4$ are shown in bold. The K_d values were determined on the basis of a minimum of three independent gel mobility shift experiments.

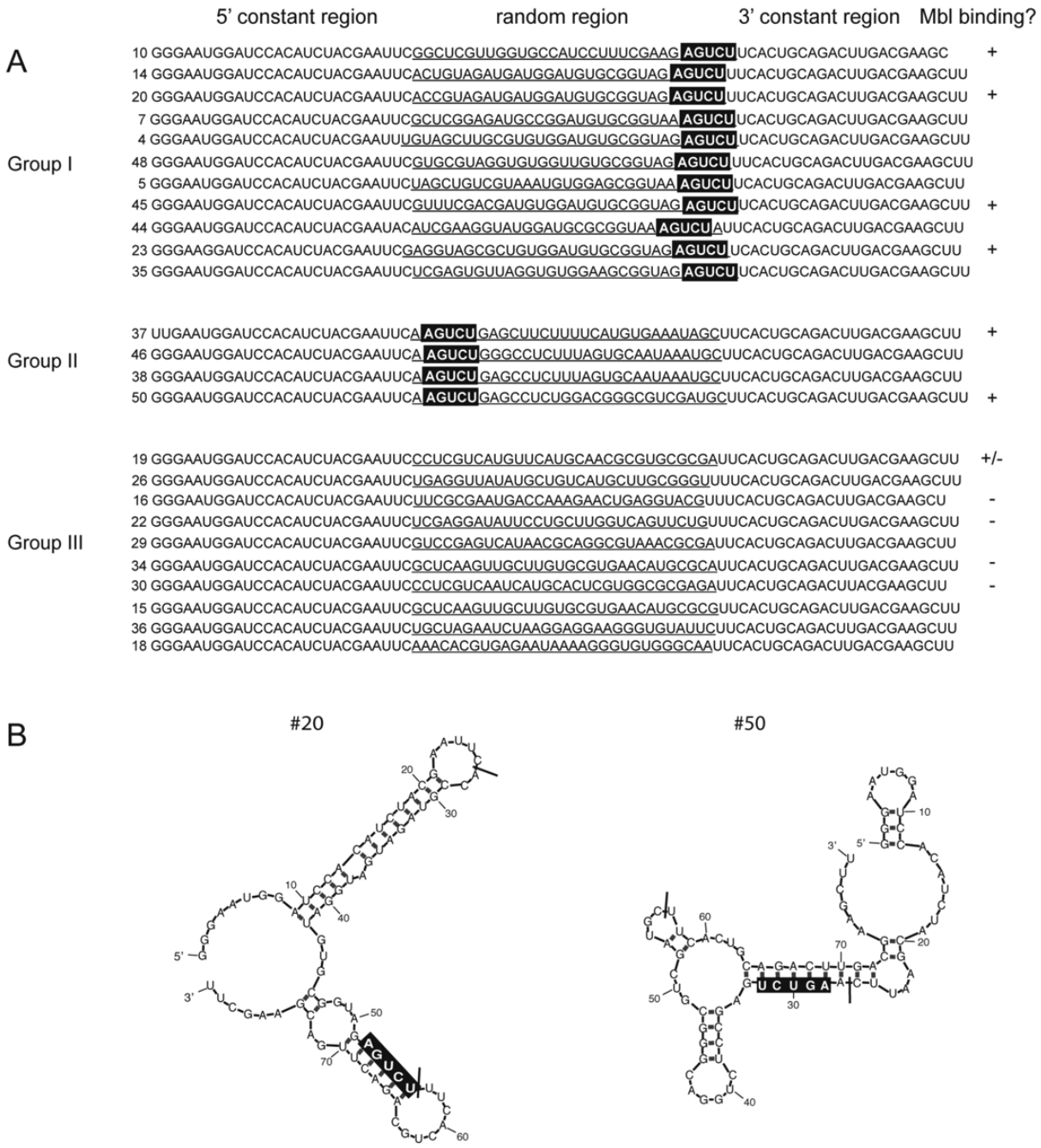

Figure 2.

High-affinity RNA ligands for Mbl identified by SELEX. (A) The 25 SELEX RNA sequences were separated into three groups on the basis of whether they contained 5′-AGUCU-3′ and then further categorized by the location of the 5′-AGUCU-3′ sequence within the random region. The location of the 5′-AGUCU-3′ sequence dictated the fold of the RNA, which also contributed to the grouping. Group I RNA sequences contain 5′-AGUCU-3′ (black boxes) near the 3′-end of the random region while Group II RNA sequences contain 5′-AGUCU-3′ near the 5′-end of the random region. Both Group I and II RNA sequences bound Mbl with K_d values of ≤50 nM. The + indicates a K_d value of <50 nM, the +/− indicates a K_d value of ~ 300 nM and the – indicates a K_d value of >300 nM. Group III RNA sequences did not contain

5′-AGUCU-3′, did not fold into a common structure, and had K_d values of >300 nM. (B) The 5′-AGUCU-3′ sequences of both group I and II (black boxes) form a double-stranded region with 5′-AGACU-3′ found in the 3′-constant region, giving rise to characteristic, predicted secondary structures within each group. RNA 20 represents the common fold of group I RNA sequences, and RNA 50 represents the common fold of group II RNA sequences. The hatch marks indicate the divisions between the constant regions and the randomized region.

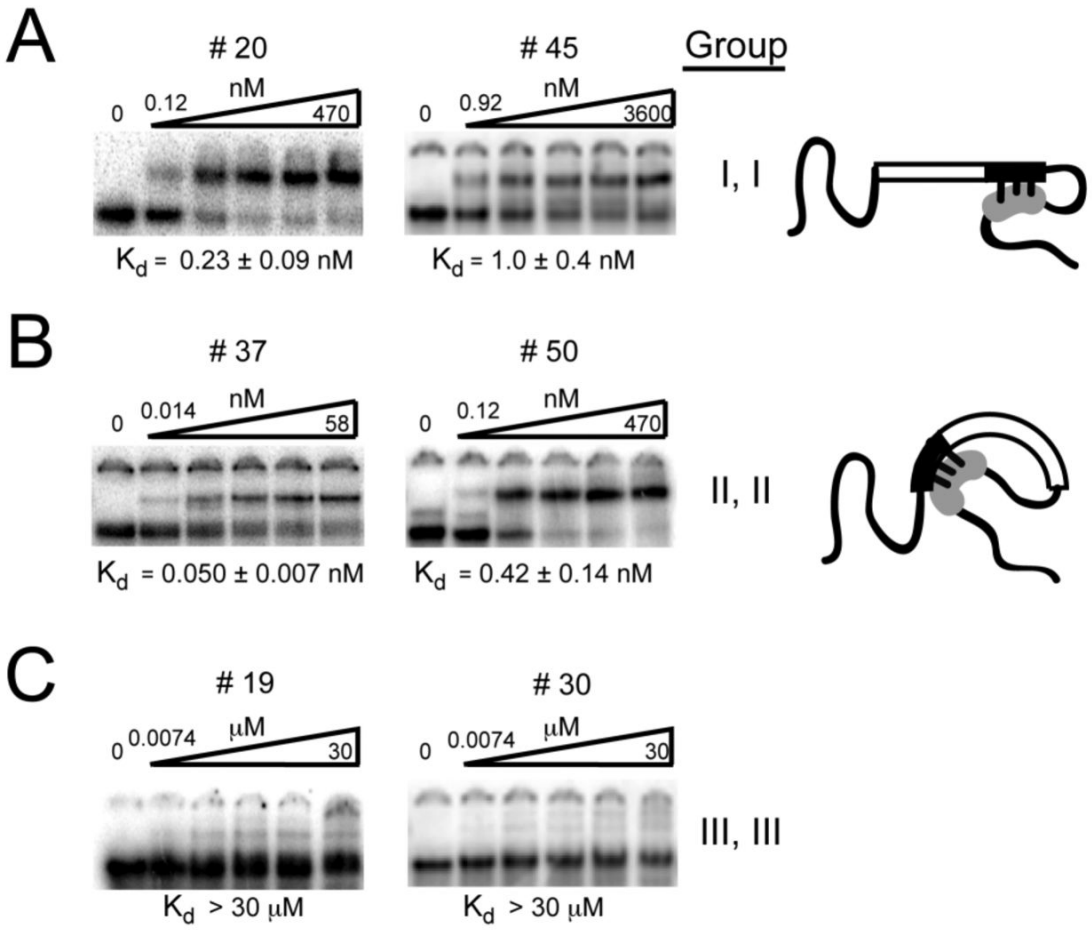

Figure 3.

RNA sequences from Group I and II bind Mbl with high affinity. (A) Gel mobility shift assays with RNAs 20 and 45 from Group I are shown. A schematic representation of their predicted secondary structures is shown to the right of the gels. The box represents the random region, and the black portion represents the 5′-AGUCU-3′ sequence and the gray section the 5′-AGACU-3′ sequence in the 3′-constant region. Mbl concentrations for the RNA 20 gel were 0.12 nM, 0.92 nM, 7.3 nM, 58 nM, and 470 nM. Mbl concentrations for the RNA 45 gel were 0.92 nM, 7.3 nM, 58 nM, 470 nM and 3600 nM. (B) Gel mobility shift assays with RNAs 37 and 50 from Group II and a schematic representation of their predicted secondary structures. Mbl concentrations for the RNA 37 gel were 0.014 nM, 0.12 nM, 0.92 nM, 7.3 nM and 58 nM. Mbl concentrations for the RNA 50 gel were 0.12 nM, 0.92 nM, 7.3 nM, 58 nM, and 470 nM. (C) RNAs 19 and 30 from Group III bind Mbl with a K_d value >30 μM. Mbl concentrations for RNA 19 and 30 gels were 0.0073 μM, 0.058 μM, 0.47 μM, 3.6 μM, and 30 μM. The K_d values were determined on the basis of a minimum of three independent gel mobility shift experiments.

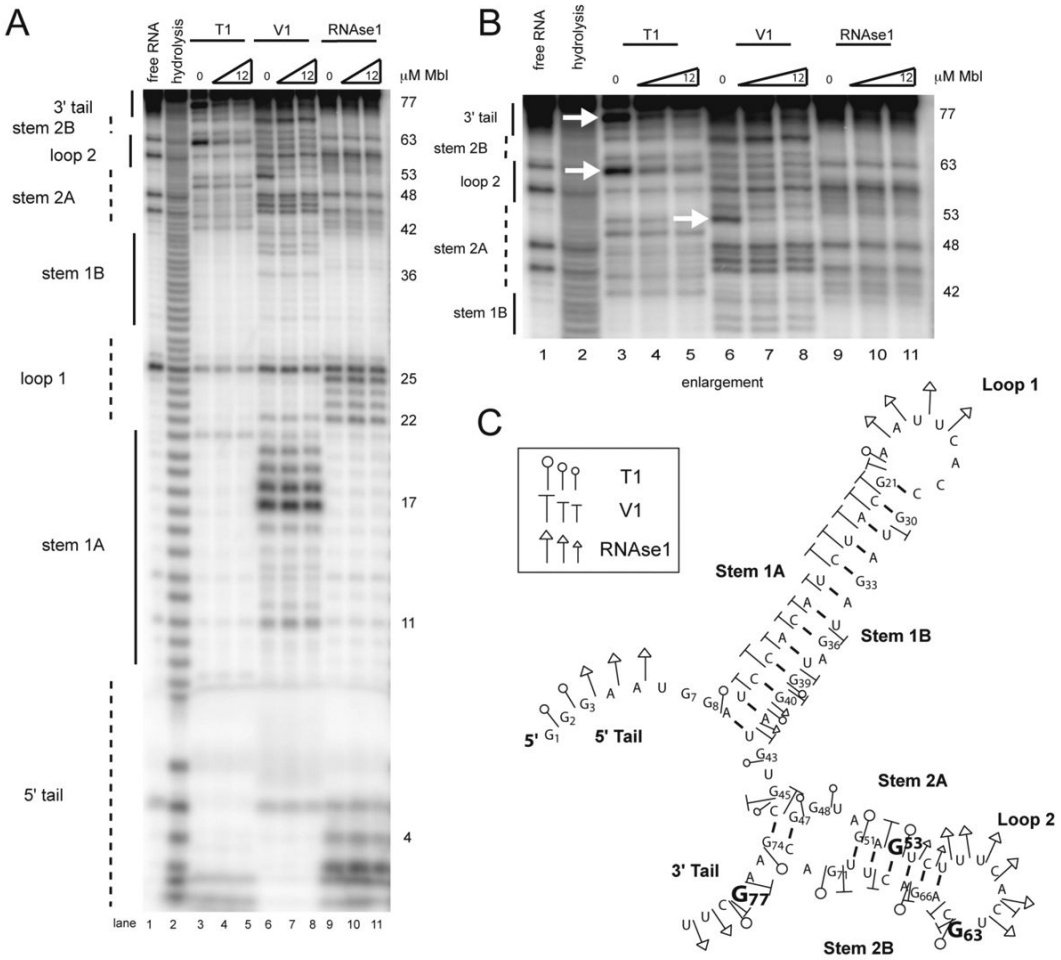

Figure 4.

Secondary structure determination and identification of the Mbl binding site on RNA 20. (A) Structure probing was performed on RNA 20 with RNase T1, RNase V1, and RNase 1 in the absence and presence of Mbl. Lane 1, free RNA, shows some degradation that was considered as background cleavage. Alkaline hydrolysis of RNA 20 was performed and run in lane 2. The concentration of Mbl ranged from 4 to 12 μM. Solid and dashed lines to the left of the gel indicate different structural elements. The numbers on the right of the gel correspond to residue numbers. (B) This is an enlargement of the upper region of the gel shown in panel A. The white arrows highlight protection of three guanosines by Mbl: G53, G63, and G77. (C) Predicted secondary structure of RNA 20 from mfold with cleavages from the three RNases (30). The three sizes of symbols indicate the intensity of the cleavage. Pins with circles indicate RNase T1 cleavages, T symbols RNA V1 cleavages, and triangles RNase 1 cleavages. The large, bold guanosines indicate bases protected by Mbl.

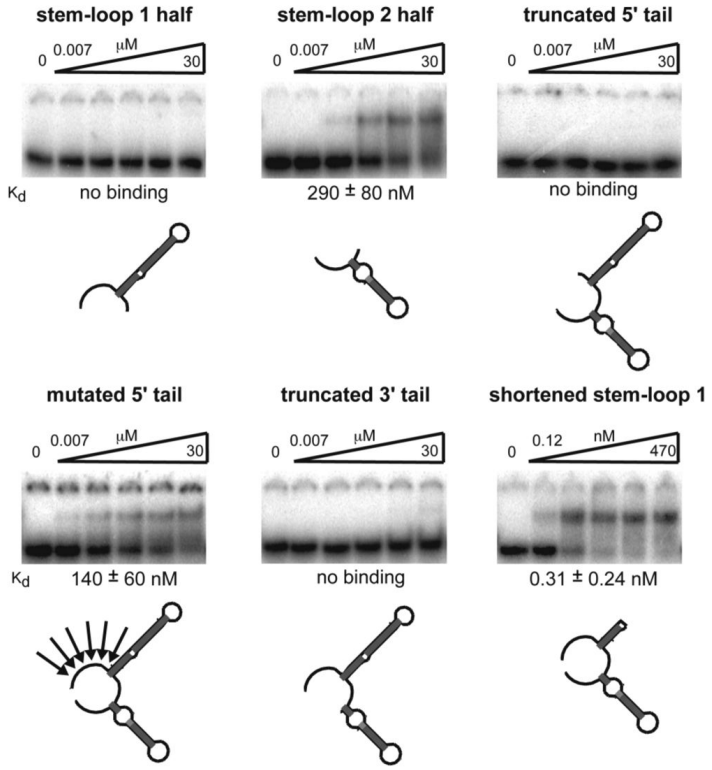

Figure 5.

Mbl requires 5′- and 3′-single-stranded tails but not full-length stem-loop 1 for high-affinity RNA binding. Various truncations and mutations of RNA 20 were bound to Mbl and analyzed using gel mobility shift assays. Stem-loop 1 half and stem-loop 2 half RNAs were cleaved between residues 44 and 45 with three guanosines added 5′ to G45 in stem-loop 2 half RNA. Truncated 5′ tail RNA was deleted up to residue 6, and U6 was replaced with a guanosine. Six residues were changed in mutated 5′ tail RNA as indicated by the arrows. The residues were mutated from 5′-G_3AAUGG-3′ to 5′-A_3CUCCA-3′. Truncated 3′-tail RNA was deleted up to residue 74. The shortened stem-loop 1 RNA was truncated up to the 12-39 base pair and replaced with a stable UUCG cap. Mbl concentrations were 0.0073 μM, 0.059 μM, 0.47 μM, 3.8 μM, and 30 μM for stem-loop 1 half, stem-loop 2 half, truncated 5′-tail, mutated 5′-tail and truncated 3′-tail RNAs. Mbl concentrations for shortened stem-loop 1 were 0.12 nM, 0.93 nM, 7.3 nM, 59 nM and 470 nM. The secondary structures shown below each gel mark the location of the truncation or mutation. The K_d values were determined on the basis of a minimum of three independent gel mobility shift experiments.

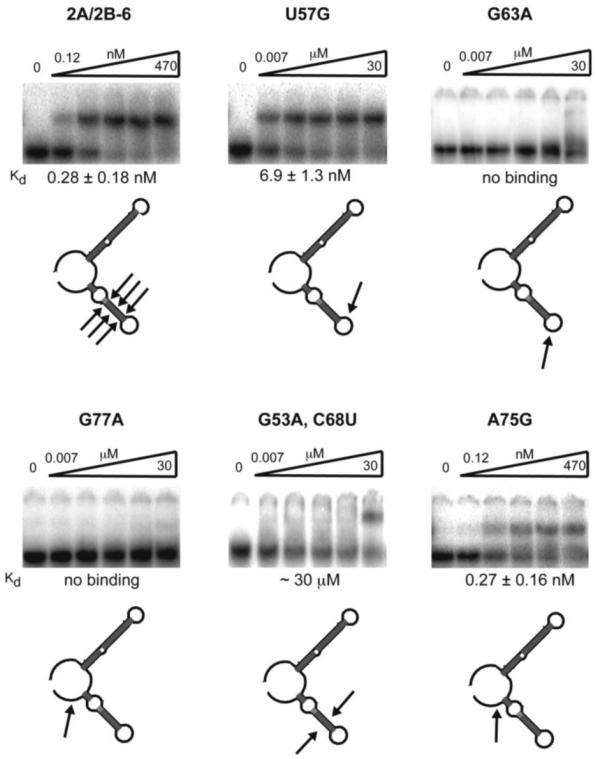

Figure 6.

Mbl binds specific structural elements and sequences. In 2A/2B-6 mutant RNA, six base pairs in stem 2 (containing the AGUCU consensus region) were mutated to maintain the structure but change the sequence. In U57G mutant RNA, a potential U-C pyrimidine-pyrimidine mismatch was mutated to a G-C base pair. The three guanosines (G53, G63, and G77) that exhibited strong protection in the footprinting assay were individually mutated to adenosines. The arrows indicate the residue or group of residues that was mutated. A75G mutant RNA was tested for binding to Mbl as a control. Mbl concentrations for U57G, G63A, G77A, and G53A/C68U mutant gels were 0.0073 μM, 0.059 μM, .47 μM, 3.8 μM, and 30 μM. Mbl concentrations for 2A/2B-6 and A75G mutant gels were 0.12 nM, 0.92 nM, 7.3 nM, 59 nM, and 470 nM. The K_d values were determined on the basis of a minimum of three independent gel mobility shift experiments.

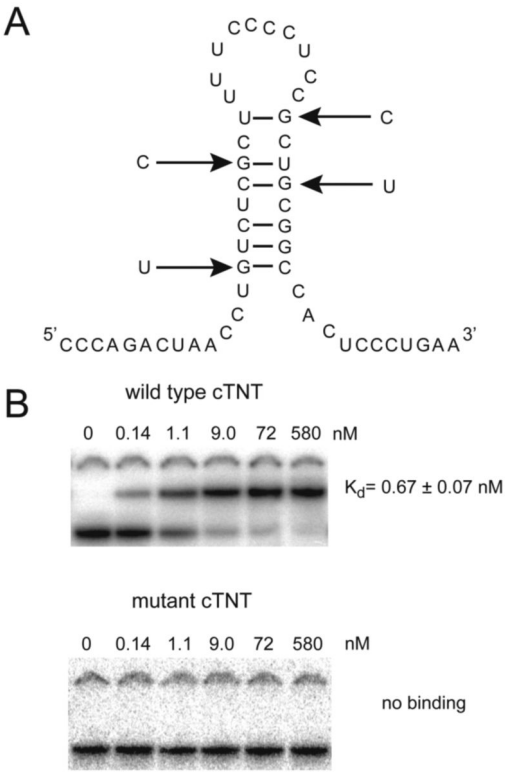

Figure 7.

Mbl binds a human MBNL1 binding site in cTNT pre-mRNA with high affinity and specificity. (A) Secondary structure of the cTNT RNA based on structure probing by Warf and Berglund (22). The arrows indicate the four guanosine to pyrimidine mutations (mutant cTNT) that abolish the stem structure. (B) Gel mobility shift assays of the 50 nucleotide region of cTNT RNA as the wild-type (top gel) or mutated (bottom gel). The K_d values were determined on the basis of a minimum of three independent gel mobility shift experiments.

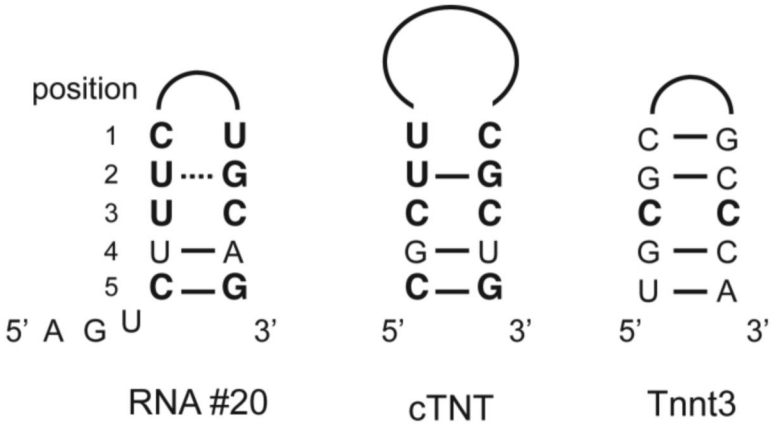

Figure 8.

Comparison of three stem-loop binding sites for the muscleblind family of proteins. From the SELEX RNA 20, a portion of stem-loop 2 is shown beginning with the highly enriched 5′-AGUCU-3′ sequence (RNA 20). The upper portion of the cTNT stem-loop structure (cTNT) is shown. The upper portion of the mouse Tnnt3 stem-loop structure (Tnnt3) is shown (23). The position and numbers refer to base pairs or mismatches in the structures to the right. The bold nucleotides highlight the sequence similarity between the structures. The dashed line indicates a potential G-U wobble base pair in RNA 20.

Table 1

Concentration of RNA and Protein used in SELEX.

round	RNA (μM)	Protein (μM)
1	20	5
2	1.2	0.12
3	4.3	0.43
4	2.4	0.24
5	5.6	0.56
6	0.56	0.056

Table 2

Highly enriched *n*-mers found in SELEX sequences.

n **-mer**	**observed**[a]	**expected**[b]	***P*** **value**
AGUCU	15	0.740438	1.03E-61
GUGCG	11	0.625438	2.54E-39
CGGUA	10	0.717688	5.82E-28
UGUG	17	2.63369	7.81E-19
UGCG	16	2.61488	1.06E-16
AGUC	16	2.65144	2.07E-16
GUG	32	10.9052	1.33E-10
GCG	29	10.9132	3.68E-08
GUC	23	10.9851	2.56E-04

[a] Observed is the total occurrences of the *n* -mer within the sample of 25 SELEX sequences.

[b] Expected is the number of occurrences according to the maximum likelihood estimate in a random pool of RNA of the same size as the 25 SELEX sequences.