JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Interaction of Individual Structural Domains of hnRNP LL with the *BCL2* Promoter i-Motif DNA

Basab Roy,[†] Poulami Talukder,[†] Hyun-Jin Kang,[‡] Shujian S. Tsuen,[†] Mohammad P. Alam,[†] Laurence H. Hurley,[‡,§] and Sidney M. Hecht*[,†]

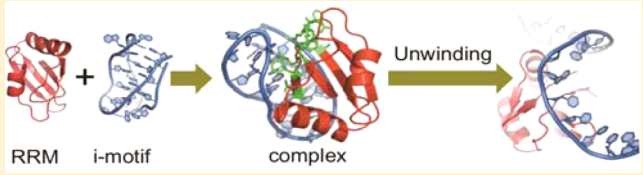
[†]Biodesign Center for BioEnergetics, and School of Molecular Sciences, Arizona State University, Tempe, Arizona 85287, United States

[‡]College of Pharmacy, University of Arizona, Tucson, Arizona 85721, United States

[§]Arizona Cancer Center and BIO5 Institute, University of Arizona, Tucson, Arizona 85721, United States

Supporting Information

ABSTRACT: The recently discovered role of the *BCL2* (B-cell lymphoma 2 gene) promoter i-motif DNA in modulation of gene expression via interaction with the ribonucleoprotein hnRNP L-like (hnRNP LL) has prompted a more detailed study of the nature of this protein−DNA interaction. The RNA recognition motifs (RRMs) of hnRNP LL were expressed individually, and both RRM1 and RRM2 were found to bind efficiently to the *BCL2* i-motif DNA, as well as being critical for transcriptional activation, whereas RRM3−4 bound only weakly to this DNA. Binding was followed by unfolding of the DNA as monitored by changes in the CD spectrum. Mutational analysis of the i-motif DNA revealed that binding involved primarily the lateral loops of the i-motif. The kinetics of binding of the DNA with RRM1 was explored by recording CD spectra at predetermined times following admixture of the protein and DNA. The change in molar ellipticity was readily apparent after 30 s and largely complete within 1 min. A more detailed view of protein−DNA interaction was obtained by introducing the fluorescence donor 6-CNTrp in RRM1 at position 137, and the acceptor 4-aminobenzo[g]quinazoline-2-one (C_f) in lieu of cytidine$_{22}$ in the i-motif DNA. The course of binding of the two species was monitored by FRET, which reflected a steady increase in energy transfer over a period of several minutes. The FRET signal could be diminished by the further addition of (unlabeled) RRM2, no doubt reflecting competition for binding to the i-motif DNA. These experiments using the individual RRM domains from hnRNP LL confirm the role of this transcription factor in activation of *BCL2* transcription via the i-motif in the promoter element.



INTRODUCTION

Overexpression of the proto-oncogene *BCL2* (B-cell lymphoma 2 gene) contributes to the resistance of cancer cells to apoptosis by suppressing the ability of the proapoptotic protein Bax to mediate cell death.[1−5] Overexpressed Bcl-2 is also associated with chemoresistance,[6] particularly for lymphocytic cancers.[2,7−9] Perhaps unsurprisingly, *BCL2* has become a target for antitumor therapy, and the approaches pursued have included reduction of mRNA levels with antisense oligonucleotides,[10] as well as the use of small molecules to disrupt protein−protein interactions.[10,11]

Transcriptional control of *BCL2* expression represents another potential strategy for therapeutic intervention. G-quadruplex structures are associated with transcriptional start sites in the promoter and 5′-UTR regions of many genes,[12,13] and may form as a consequence of the negative supercoiling that results from local DNA unwinding during transcription.[14,15] By treating the formed G-quadruplex structures as structural elements that may permit transcription to be controlled,[16−18] the G-quadruplex promoter elements of a number of genes have been targeted with the goal of altering gene expression.[16,18−27]

The presence of G-rich regions on one strand of B-form DNA in regulatory regions of numerous genes dictates that there will also be C-rich regions on the complementary DNA strand. C-rich elements in DNA can form a secondary structure known as an i-motif (Figure 1).[28] The i-motif DNA structure is more dynamic than the G-quadruplex, and has been shown to exhibit increased stability at low pH[29] and under conditions of negative supercoiling.[14] More detailed studies have revealed that i-motif stability also varies according to parameters such as loop size[30,31] and environment; molecular crowding conditions led to an i-motif found to be stable even at pH 8.0.[32] More recently, it has been proposed that the i-motif may also play a role in transcriptional regulation.[14,30,31,33,34] In this regard it may be noted that the i-motif is in equilibrium with a flexible hairpin,[35,36] and that in at least under some circumstances the folding of the i-motif and G-quadruplex may be mutually exclusive.[37,38]

In a recent study, we showed that the dynamic equilibrium between the *BCL2* promoter i-motif and flexible hairpin

Received: May 16, 2016
Published: August 2, 2016

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

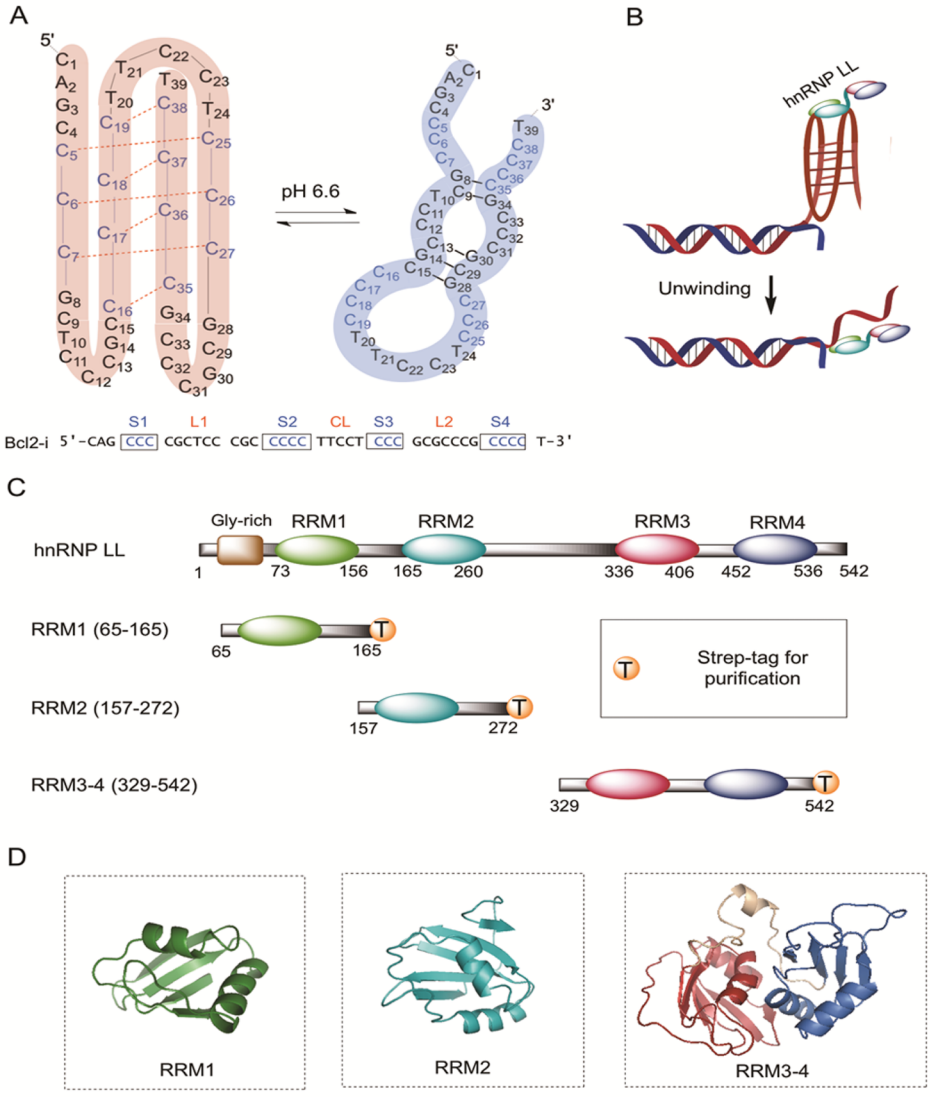


Figure 1. *BCL2* promoter i-motif sequence and structures of individual domains of hnRNP LL. (A) Interchangeable i-motif and hairpin structure of *BCL2* promoter element. (B) Proposed mechanism of *BCL2* gene regulation by the interaction of the DNA i-motif structure and transcription factor hnRNP LL. The binding interaction of hnRNP-LL domains RRM1 and RRM2 with i-motif lateral loops 1 and 2 unfolds the i-motif structure and facilitates downstream transcription events. (C) Sequence of hnRNP LL with four interactive domains and constructs containing RRM1, RRM2 and RRM3−4. (D) Predicted structures of RRM1, RRM2 and RRM3−4 based on sequence homology with other members of that protein family.

structure could be controlled by the use of small molecules, providing a mechanism for control of the expression of *BCL2*.[35] Further, we identified hnRNP L-like (hnRNP LL) as a transcription factor which can activate transcription by binding to the *BCL2* i-motif.[36] We demonstrated that hnRNP LL binds to the two i-motif lateral loops via its RNA recognition motifs (RRMs), leading to unwinding of the i-motif structure by hnRNP LL.[36] Presently we extend the analysis of hnRNP LL−i-motif binding to the level of the individual RRMs. RRM1 and RRM2 are both shown to bind to the *BCL2* i-motif; binding is mediated via the i-motif lateral loops as proposed earlier. Further, we monitored the unfolding of the i-motif by the use of a FRET experiment involving a construct of RRM1 containing the fluorescence donor 6-cyanotryptophan at position 137, and a tricyclic cytidine nucleoside acceptor at position 22 of the lateral loop of the *BCL2* i-motif. The results strongly support the model of i-motif binding and unwinding proposed previously[36] and provide a more detailed molecular understanding of the interaction.

RESULTS

Expression and Purification of the RNA Recognition Motifs of hnRNP LL, and Modeling of their Interaction with the i-Motif DNA. In order to understand the individual interactions of the four RNA recognition motif (RRM) domains of transcription factor hnRNP LL with the *BCL2* i-motif DNA (Figures 1A and B), the domains were cloned in pET28a vectors and expressed separately using *E. coli* BL21-DE3 cells. Domains RRM1, RRM2 and RRM3−4 were constructed with an additional 15 to 20 amino acids on both the C and N termini of the constructs, flanking the DNA sequence optimized for expression in *E. coli* (Figure 1C). The genes were equipped with a Strep-tag at their C-termini to facilitate Strep-Tactin mediated protein purification. Polyacrylamide gels illustrating the expressed RRM1, RRM2 and

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

RRM3−4 domains before and after protein purification are shown in Figure S1 of the Supporting Information. The modeled domain structures, as illustrated in Figure 1D, were determined by sequence homology interpretation employing homologous proteins of the family.

The three-dimensional structure of RRM1 was modeled with reference to the solution structure of N-terminus mouse protein BAB28521 (PDB ID: 1WEX). Similarly, the structure of RRM2 was predicted with reference to the solution structure of mouse hnRNP LL RRM2 (PDB ID: 2E5I), whereas the structure of RRM3−4 was modeled based on the crystal structure of human hnRNP L (PDB ID: 3TO8).

Also recorded were the CD spectra and thermal denaturation profiles of RRM1, RRM2 and RRM3−4 (Figure S2 of the Supporting Information). These data indicated that all of the constructs had Tm values well above room temperature.

Determination of the Critical RRM Domains of hnRNP LL Required for the Transcription of *BCL2*. It is proposed that the lateral loops of the *BCL2* i-motif are initially recognized by adjacent RRM1 and RRM2 and then the i-motif structure is unfolded following rearranged binding of the individual RRM domains.[36] To demonstrate the relative roles of each RRM domain of hnRNP LL on the transcriptional activation of *BCL2*, the promoter activities of *BCL2* were determined following cotransfection of RRM1, RRM2, RRM1−2 and RRM3−4 using hnRNP LL as a positive control. The domains with codons optimized for expression in *E. coli* were cloned into pCDNA3.1 with a FLAG-tag sequence at the N-terminal for expression in mammalian cells. HEK293TT cells were transfected with the pGL3-BCL2 construct, including the i-motif-forming sequence of the *BCL2* promoter,[36] the pCDNA3.1/FLAG-RRM domain and pRL-TK for normalization and then dual luciferase assay was conducted after 24 h. The expression of FLAG-tagged RRM domains in HEK293TT, a modified human embryonic kidney cell line, was confirmed by immunoblot analysis using a FLAG-antibody (Figure S3 of the Supporting Information). As shown in Figure 2, RRM1, RRM2, and RRM3−4 exhibited luciferase activity but reduced by ∼30% compared to the full-length hnRNP LL. The luciferase activity with RRM1−2 was not significantly reduced relative to hnRNP LL, while that of RRM3−4 was clearly less than full length hnRNP LL. This suggests that RRM1 and RRM2 of hnRNP LL play more critical roles in the transcriptional regulation of *BCL2* than RRM3−4. To further define the role of each RRM domain, especially RRM1 and RRM2, in the activation of BCL2 transcription through binding to the i-motif various experimental approaches, such as circular dichroism (CD), electrophoretic mobility shift assay (EMSA), bromine footprinting and Förster resonance energy transfer (FRET) were subsequently used.

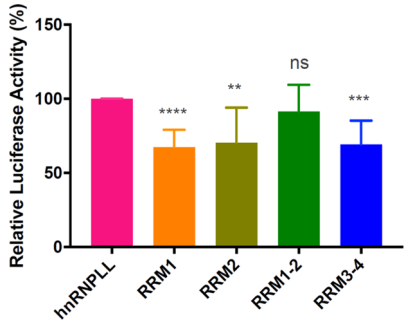


Figure 2. Effect of each RRM domain on the promoter activity of *BCL2*. HEK293TT cells were transfected with three different constructs, pGL3-*BCL2* WT, pCDNA3.1/Flag-RRM domains and pRL-TK and luciferase assays were conducted after 24 h. The ratio of firefly to renilla luciferase for each RRM domain was normalized to that for full-length hnRNP LL. Mean ± SEM are shown in the graph. *P* values (***P* < 0.01, ****P* < 0.001, *****P* < 0.0001, ns: not significant) were determined by two-tailed *t* test. Each of the assays was repeated a minimum of six times.

Binding of the RRMs to the *BCL2* i-Motif DNA. The binding interactions between each of the three constructs, containing the four domains, of hnRNP LL and the *BCL2* i-motif DNA were studied by employing a gel shift assay (Figure 3). As postulated in our previous study,[36] RRM2 and RRM1 exhibited strong binding to the i-motif DNA (Figure 3A and 3B) when the DNA:RRM domains molar ratios were varied

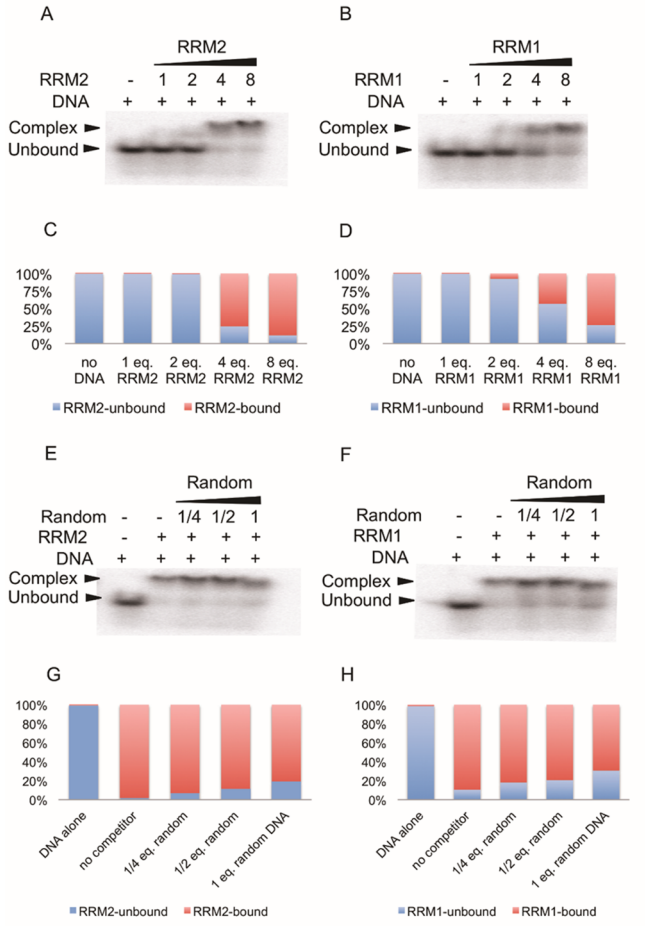


Figure 3. Binding affinity study of RRM1 and RRM2 with the i-motif DNA at pH 6.6. (A and B) Binding affinity of RRM2 and RRM1 (1, 2, 4, and 8 mol equiv) to the i-motif DNA, respectively. (C and D) Histogram showing percentage of DNA−protein complex formation with varying concentrations of RRM2 and RRM1, respectively. (E and F) RRM2 and RRM1 binding competition with i-motif and varying concentrations (0.25, 0.5, and 1 mol equiv of nucleotide concentrations) of sheared calf thymus DNA ("random" DNA), respectively. (G and H) Histogram representing percentage of DNA−protein complex formation while radiolabeled i-motif and varying concentrations (0.25, 0.5, and 1 mol equiv of nucleotide concentrations) of calf thymus DNA were incubated with RRM2 and RRM1, respectively.

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

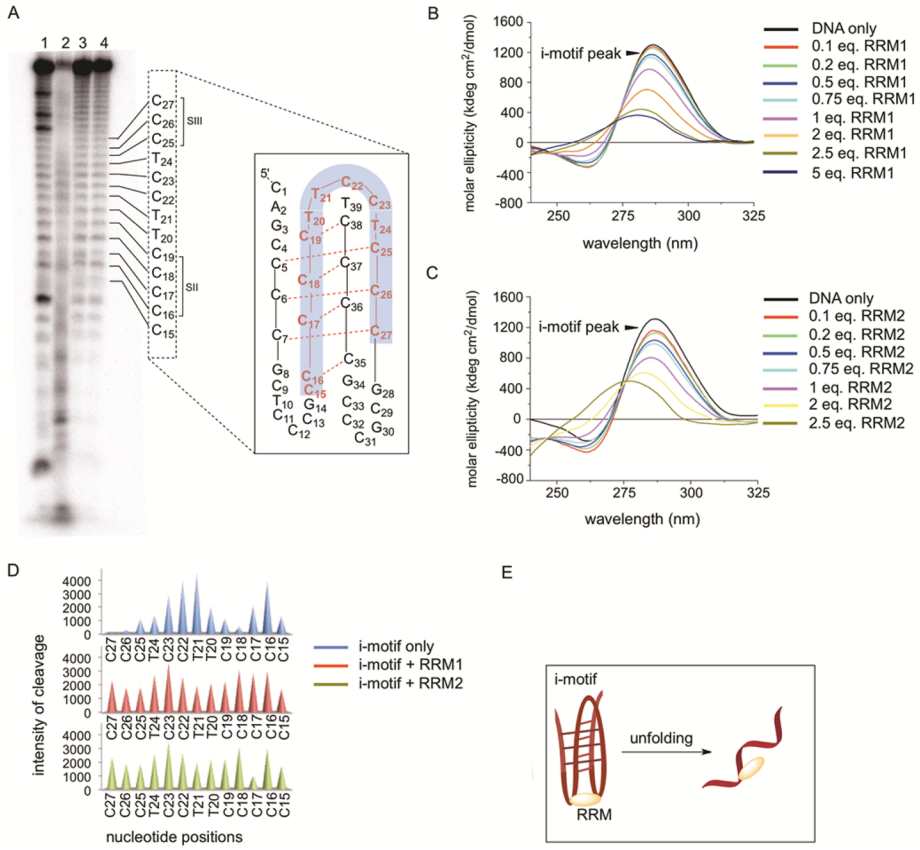


Figure 4. Analysis of structural change in the *BCL2* i-motif as a result of binding with RRM1 and RRM2. (A) Structural change of i-motif observed in a bromine footprinting assay at pH 6.6. Lane 1, Maxam−Gilbert G+A lane; lane 2, 10 pmol DNA alone; lane 3, 10 pmol DNA + 80 pmol RRM1; lane 4, 10 pmol DNA + 80 pmol RRM2; i-motif structure highlighting C-rich stems (SII, SIII) and loop CL in red, which shows significant changes in the bromine footprinting assay. (B) CD spectra of i-motif DNA at pH 6.6 with varying concentrations of RRM1. (C) CD spectra of i-motif DNA at pH 6.6 with varying concentration of RRM2. (D) Histogram resulting from bromine footprinting assay shows structural changes in the region C_{15} to C_{27}, encompassing C-rich stems SII, SIII and loop CL of the i-motif due to the binding interaction with RRM1 and RRM2. (E) Proposed mechanism of RRM1 and RRM2 interaction with i-motif DNA and subsequent unfolding of the complex structure.

from 1 to 8 equiv. Further, RRM1 and RRM2 domains were shown to interact with the i-motif DNA with good affinities. The histograms in Figure 3C and 3D illustrate that when present at 8 mol equiv relative to the i-motif DNA, both domains bound to the *BCL2* i-motif DNA to the extent of >50% at room temperature. In comparison, it was clear that RRM2 bound to the i-motif DNA more effectively when present at 4 mol equiv than did RRM1 (cf. Figures 3C and 3D). Interestingly, RRM3−4 (Figure 1D) proved to be a much weaker binder of the i-motif DNA (Figure S4 of the Supporting Information). Varying concentrations of RRM3−4 (2.5 to 100 mol equiv), when incubated with 1 equiv of the i-motif DNA, resulted in less than 50% DNA binding at protein levels <50 mol equiv. The gel also revealed the presence of two protein−DNA complexes, labeled as complex 1 and complex 2 (Figure S4A), suggesting a more complicated mode(s) of interaction between RRM domains 3−4 of hnRNP LL and the *BCL2* i-motif DNA.

To obtain additional insight into the affinity and specificity of RRM1 and RRM2 for the i-motif DNA, a competition experiment was carried out using calf thymus DNA. Domains RRM1 and RRM2 were both shown to have a stronger affinity toward the i-motif DNA compared to calf thymus DNA on a nucleotide basis. The binding interaction was studied using varying concentrations of calf thymus DNA (0.25 to 1 molar nucleotide equivalent) in mixture with an equivalent of an i-motif DNA complex with RRM1 or RRM2 (Figure 3E and 3F). In the presence of the equivalent nucleotide concentration of calf thymus DNA, both of the domains showed more than 50% binding interaction when incubated with equimolar substrate, i.e., the *BCL2* i-motif DNA (Figure 3G and 3H). Further, RRM2 was a more effective competitor than RRM1, in agreement with the data in Figure 3C and 3D.

Structural Change in the *BCL2* i-Motif DNA Following Binding to RRMs. A previous investigation suggested that the interaction between hnRNP LL and the i-motif DNA is likely to unwind the complex i-motif secondary structure, which may result in transcription initiation from the promoter region of the *BCL2* gene.[36] In the present study, the structural change of the *BCL2* i-motif has been studied following admixture with the RRM1 and RRM2 domains of hnRNP LL, by employing circular dichroism as well as a bromine-footprinting technique.[35,36,39] The i-motif shows unwinding of the complex structure at higher pH, with a transition pH at 6.6.[30] As observed in Figure 4A (lane 2), the i-motif DNA shows a differential rate of bromination on cytidine nucleotides in the SII, CL and SIII regions of the i-motif (shaded area in Figure 4A), followed by cleavage of DNA at brominated nucleotides at pH 6.6. The variation in the extent of bromination is attributed to the three-dimensional structure of the i-motif, which restricts the accessibility of some cytidines for bromination.[35,36,39] It is evident from Figure 4A (lanes 3 and 4) that when the i-motif

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

DNA was treated with RRM1 and RRM2, the accessibility of nucleotides for bromination changed dramatically, consistent with the putative conversion of the i-motif structure to a less complex DNA structure (Figures 1B and 4E). A similar experiment carried out at pH 7.5, which had a higher proportion of linear DNA within the region potentially involved in i-motif formation, showed a similar, yet lesser structural change upon protein binding.

The summary of the susceptibility of nucleotides, spanning C15 to C27, to bromine-induced cleavage, upon incubation with RRM1 and RRM2 at pH 6.6, revealed that stems SII and SIII of the i-motif become more accessible to bromination after the binding of the hnRNP LL domains, which must have resulted from at least partial relaxation of the folded i-motif structure (Figure 4D and 4E).

The i-motif structural change was also studied by analyzing the CD spectra of the DNA in the presence of RRM1 and RRM2 (Figure 4B and 4C). As noted previously,[30,35,36] the *BCL2* i-motif has a characteristic CD spectrum with a distinct peak at 286 nm. This peak decreases when the CD spectrum is monitored at higher pH, suggesting a shift of equilibrium between the i-motif structure and one or more alternative forms of the DNA (Figure 1A).[30,35] As established in our previous studies, the folded *BCL2* i-motif structure is the predominant form at lower pH values, and decreases gradually with increasing pH.[30] As shown in Figure 4B and 4C, the peak at 286 nm decreased with the addition of increasing amounts of RRM1 and RRM2, suggesting a significant structural change in the i-motif when it interacted with the RRM1 and RRM2 domains of hnRNP LL. The reductions in the i-motif peak intensity in the presence of RRM1 or RRM2 were compared (Figure S5). The binding curves suggested a slightly stronger binding interaction between RRM2 and i-motif. The change in the i-motif CD spectrum was less pronounced when the interaction of RRM1 and RRM2 with the i-motif DNA was examined at a higher pH (pH 7.5) (Figure S6 of the Supporting Information). The lesser reduction of the CD peak at pH 7.5 is attributed to the presence of the DNA in structural forms in addition to the i-motif at the higher pH value. It may be noted that for RRM2, but not RRM1, admixture of one equivalent of protein to the DNA resulted in a spectral shift to shorter wavelength. Plausibly, this may reflect the role of RRM2 as the primary DNA binding domain of hnRNP LL.

The rate at which unfolding of the i-motif DNA was effected by RRM1 was studied by measuring the change in the CD spectrum of the i-motif DNA following admixture of a single equivalent of RRM1. As shown in Figure 5, there was a decrease in the intensity of the peak at 286 nm over a period of 5 min, undoubtedly as a consquence of i-motif binding with RRM1 and subsequent unwinding of the i-motif DNA structure (Figure 1B). The change in intensity of the CD peak following admixture of RRM1 was substantial after 30 s, largely complete after 1 min, and complete after 5 min, defining the rate of i-motif DNA unfolding.

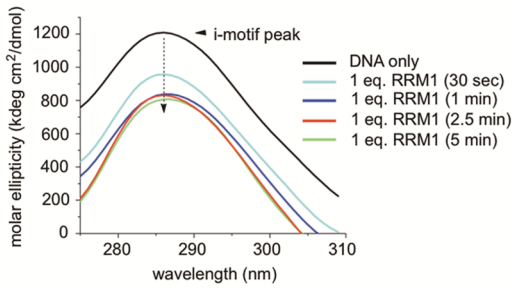


Figure 5. CD spectra of i-motif DNA showing time-dependent unwinding at 10 °C, following the addition of one equivalent of RRM1.

Mutational Analysis of *BCL2* i-Motif DNA Binding to RRM1 and RRM2. In previous studies, lateral loops 1 and 2 of the i-motif DNA (Figure 1A) were proposed to interact with RRM1 and RRM2.[36] The foregoing experiments verified the binding of RRM1 and RRM2 to the *BCL2* i-motif DNA, but not the structural elements in the DNA involved in the binding interaction. In order to determine definitively which region(s) of the i-motif DNA interacts with RRM1 and RRM2, several modified i-motif DNAs were synthesized (Figure 6C). Their structures were characterized by circular dichroism (Figure 6A and 6B) and their binding interactions with RRM1 and RRM2 were studied using an electrophoretic mobility shift assay (EMSA) (Figure 7).

The CD spectra of the modified i-motifs, recorded at pH 6.6, indicated that mutations in the C-rich i-motif stem did not eliminate the i-motif structures as significant contributors to the equilibrium mixture present (Figure 1A), which is evident from the limited reduction in the i-motif peak at 286 (Figure 6A). Interestingly, mutations in the loop structures of the i-motif DNA indicated that changes in the nucleotide sequences within the L1, L2 and CL loops could have a greater and more varied effect on the i-motif structure, although (central) loop CL appeared to be somewhat less important in the stabilization of the i-motif (Figure 6B).

As shown in Figure 7, mutations involving L1 and L2 sharply decreased the binding of the i-motif DNAs to RRM1 and RRM2 (Figure 7B and 7C), whereas changes in the C-rich stem region of the i-motif DNA had much less effect on the binding interactions (Figure 7A and 7C). In full agreement with our previous report,[36] this suggested that the lateral loops of the i-motif DNA structure are important for RRM1 and RRM2 binding.

FRET between the *BCL2* i-Motif DNA and RRM1. A binding interaction study and kinetic analysis were envisioned, employing Förster resonance energy transfer (FRET) as a tool. In a recent paper we reported the synthesis and photophysical characterization of 6-cyanotryptophan (6-CNTrp) and the use of a cytidine analogue, 4-aminobenzo[g]quinazoline-2-one (C_f),[40] as a FRET acceptor for 6-CNTrp (Figure 8A).[41] The absorption of C_f (Figure 8A) at 360 nm overlaps with the emission of 6-CNTrp at 370 nm, resulting in acceptor emission at 440 nm. The critical Förster distance (R_0) was determined experimentally to be 24.1 Å, which is suitable for monitoring DNA−protein binding events. In the current study, we have used this FRET pair to demonstrate a specific association of the RRM1 domain of hnRNP LL with its i-motif DNA substrate.

Previous studies have demonstrated that the C-rich upstream promoter region of *BCL2* forms a highly dynamic i-motif structure (Figure 8B), which can act as a transcriptional switch for the expression of the *BCL2* gene.[35,36] Recent findings further reveal that the *BCL2* i-motif can be a potential target for small molecules and, more interestingly, for putative transcription factors such as hnRNP LL.[36] However, the details of the i-motif DNA interaction with hnRNP LL have yet to be fully defined. According to the hypothesis set forth in previous studies, two of the hnRNP domains, RRM1 and RRM2 (Figure 1D), are most likely to recognize and bind the *BCL2* i-motif

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

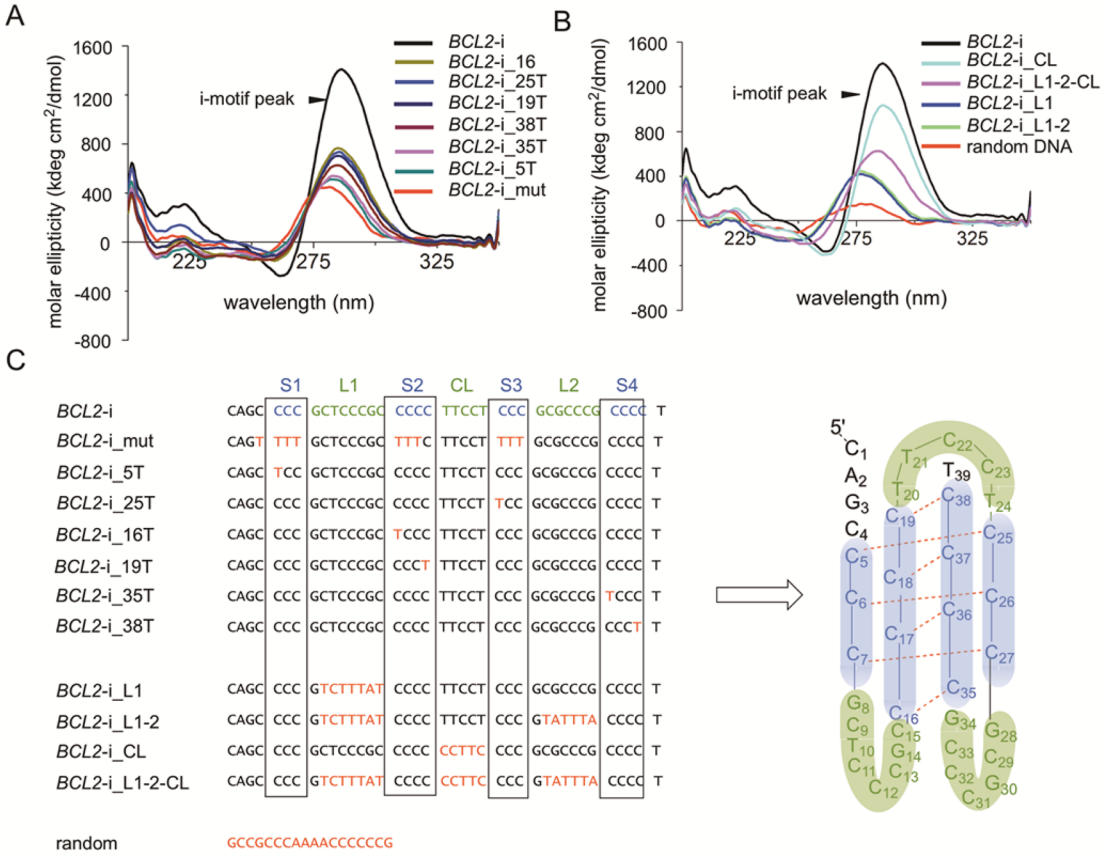


Figure 6. Circular dichroism analysis of *BCL2* i-motif variants. (A and B) Change in CD spectra of *BCL2* i-motif stem and loop mutants, respectively. (C) Table illustrating sequences of the mutant i-motifs used in the binding study. These experiments were carried out at 25 °C in 50 mM Tris-HCl, pH 6.6, containing 100 mM NaCl and 1 mM EDTA.

DNA. Here, we have studied FRET between RRM1, substituted with 6-CNTrp, and the i-motif DNA containing the fluorescent nucleotide C_f.

A computer simulation based on homologous protein hnRNP L,[42] predicted the amino acids in RRM1 which may take part in substrate (i-motif) binding. The domain structure was determined by sequence homology interpretation based on other homologous proteins of the family. The DNA binding residues in RRM1 were found using DP-bind web-tool.[43,44] The predicted participating residues were R105, G106, F134, K135, R136, Q137, Q165, F168, N170, Y171. The binding pockets on RRM1 were scanned by Patchdock web-tool for favorable binding interactions,[45,46] which was then used in the rigid body docking, using the Hex-docking algorithm[47] and the Haddock docking algorithm,[48−50] using participating residues as "ambiguous-binding restraints". Similarly, the residues of the i-motif DNA involved in binding were found to be in loops L1 and L2 (lateral loops), whereas the central loop appears to be least important for protein binding. On the basis of the DNA−protein docking (Figure 8B), two tyrosine residues, Tyr137 and Tyr104, were chosen for substitution by 6-CNTrp. Rigid body docking also revealed that the distances between the central loop of the i-motif and Tyr137 or Tyr104 were 24.5 and 29.7 Å, respectively, both of which were comparable to the critical Förster distance R_0 between C_f and 6-CNTrp (24.1 Å).

The RRM1 domain of hnRNP LL protein was cloned in pET28a and expressed in *E. coli* BL21-DE3 cells. In order to investigate the interaction of RRM1 and the i-motif, two modified proteins were synthesized, each of which contained a 6-CNTrp residue in lieu of the native tyrosine resides at positions 104 or 137 (cf. Figure 8). These were prepared by in vitro protein synthesis, using 6-cyanotryptophanyl-$tRNA_{CUA}$ and plasmid DNAs containing the RRM1 coding region with a TAG codon at Tyr137 or Tyr104. This afforded proteins RRM1−137-CNTrp and RRM1−104-CNTrp, respectively (Figure S7 of the Supporting Information). The purified proteins were used for FRET analysis with DNA substrate *BCL2*-i22C, i.e., the i-motif DNA modified at the central loop (CL) with cytidine analogue C_f (Figure 8).[40] RRM1−137-CNTrp (5 μM concentration) was incubated with 0.05−0.20 μM *BCL2*-i22C for 10 min prior to excitation at 310 nm (Figure 9A). Likewise, RRM1−104-CNTrp (5 μM concentration) was incubated with 0.10−0.20 μM *BCL2*-i22C for 10 min before excitation at 310 nm (Figure 9B). As illustrated in Figure 9, increasing concentrations of the i-motif DNA increased the C_f peak at 440 nm as a result of FRET while decreasing the 6-CNTrp peak at 360 nm, strongly suggestive of a binding interaction between RRM1 and the i-motif DNA. In contrast, the wild-type RRM1 protein lacking 6-CNTrp did not display FRET with the i-motif DNA (Figure S8 of the Supporting Information).

From the FRET experiments in Figure 9, the emission profiles of the donors (RRM1−137-CNTrp, RRM1−104-CNTrp) and the donor−acceptor complexes were normalized, integrated and fitted using the equation, $E = I_A/(I_A + I_D)$, where E is the FRET efficiency; I_A and I_D are the integrated intensities of the donor in unbound and bound forms, respectively (Table 1). The greater FRET efficiency involving RRM1-*137-CNTrp*

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

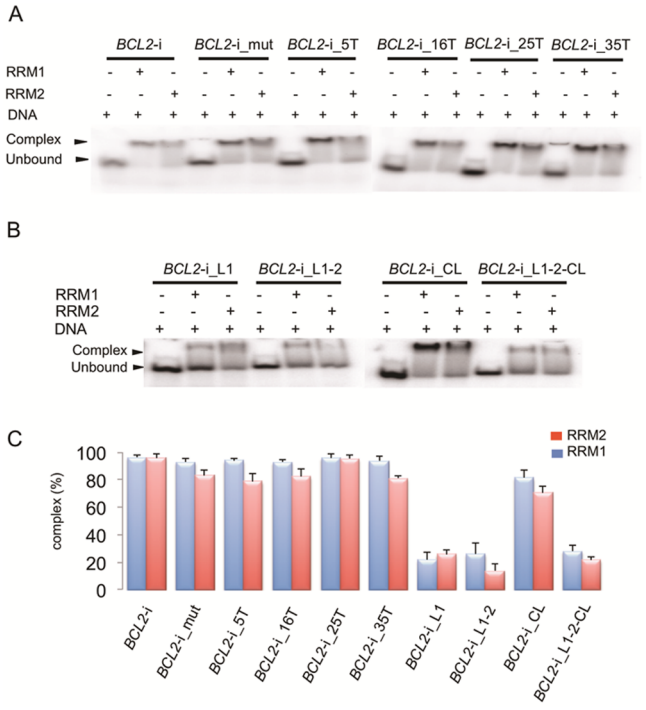


Figure 7. Binding interaction study of mutated *BCL2* i-motifs with RRM1 and RRM2. (A and B) Binding affinity of RRM1 and RRM2 to five of the *BCL2* i-motifs with specific modifications at the C-rich region and four of the modified *BCL2* i-motifs with specific modifications at loops flanking the C-rich region, respectively. (C) Histogram summarizing percentage formation of complexes from RRM1, RRM2 and mutant i-motifs. The experiments were carried out in 50 mM Tris-HCl, pH 6.6, containing 100 mM NaCl and 1 mM EDTA. The protein−DNA complexes were incubated at room temperature and then analyzed on a 12% polyacrylamide gel, which was run at 80 V for 45 min at 4 °C.

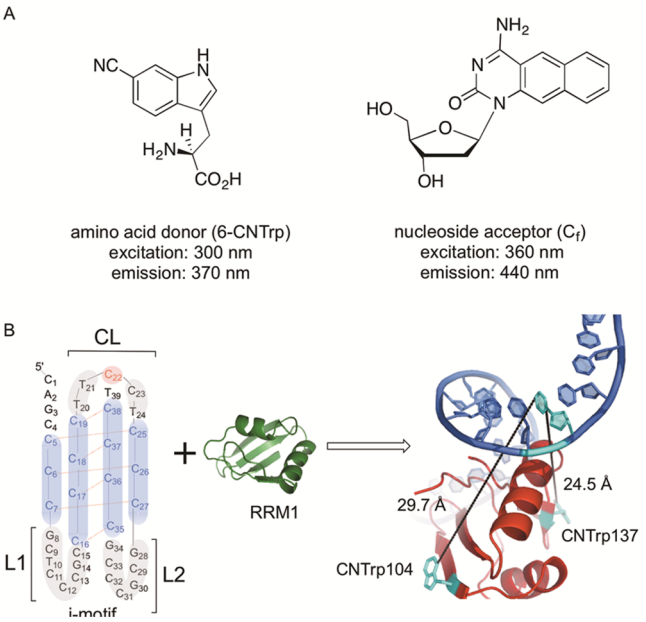


Figure 8. Structures of the fluorescent probes used in this study, the *BCL2* promoter i-motif DNA and the RRM1 structural domain of hnRNP LL. (A) 6-Cyanotryptophan (6-CNTrp) and the deoxyribonucleoside derivative of 4-aminobenzo[g]quinazoline-2-one (C_f). (B) *BCL2* promoter i-motif DNA and the RRM1 structural domain of hnRNP LL. The i-motif stem is shown in blue with two of the lateral loops (L1 and L2) and the central loop (CL). The red nucleotide (C_{22}) indicates the position of substitution by C_f. The three-dimensional structure of RRM1 was modeled with reference to the solution structure of the RRM domain of mouse protein BAB28521 (PDB ID: 1WEX).

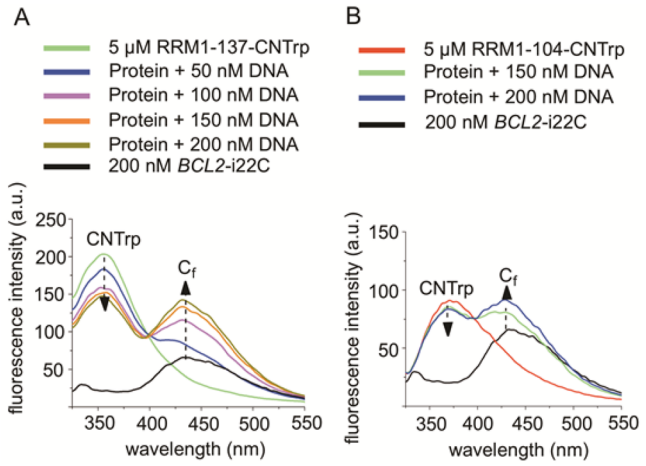


Figure 9. FRET between *BCL2*i-22C and RRM1−137-CNTrp or RRM1−104-CNTrp. (A) RRM1−137-CNTrp (5 μM) was allowed to interact with *BCL2*i-22C (50, 100, 150, and 200 nM) and the fluorescence emission spectra were obtained after 10 min of incubation at pH 6.6 and 25 °C, and excitation at 310 nm. (B) RRM1−104-CNTrp (5 μM) was allowed to interact with *BCL2*i-22C (100 and 200 nM) and the fluorescence emission spectra were obtained after 10 min of incubation at pH 6.6 and 25 °C, and excitation at 310 nm. In both cases, the experiments were carried out in 50 mM Tris-HCl, pH 6.6, containing 100 mM NaCl and 1 mM EDTA.

and *BCL2*-i22C was anticipated based on the shorter distance between Tyr137 and C_{22} of the i-motif, and made this construct suitable for further experimentation.

The time required for the optimum binding between RRM1 and the i-motif DNA was studied by means of a FRET experiment involving RRM1−137-CNTrp and *BCL2*-i22C. Figure 10A demonstrates that when 0.5 μM RRM1−137-CNTrp was incubated with 0.2 μM *BCL2*-i22C at pH 6.6 and 25 °C, there was a gradual increase in the intensity of the emission peak of C_f at 445 nm over the time period ranging from 1 to 10 min. This experiment suggested that the optimum binding between RRM1 and the i-motif was achieved within about 10 min. This time scale was roughly the same as that recorded for the unfolding of the i-motif by RRM1, but may have failed to record the most dynamic change by failing to record time points at time less than 1 min (cf. Figure 5).

A competition experiment was carried out to determine whether RRM1 can bind to the i-motif in the presence of other domains of hnRNP LL such as RRM2. A solution containing RRM1−137-CNTrp (0.2 μM concentration) was preincubated with equimolar i-motif DNA for 10 min prior to the addition of RRM2. In comparison with the C_f emission peak resulting from admixture of RRM1−137-CNTrp and *BCL2*-i22C (green line), the presence of 0.2 μM wild-type RRM2 caused the peak due to the C_f emission to decrease (blue line), and this peak was further decreased in the presence of 0.4 μM wild-type RRM2 (cyan line) (Figure 10B). This suggests that RRM2 may have displaced RRM1 from its binding site on the i-motif. It also suggests that RRM1 and RRM2 may target the same binding sequence in the i-motif DNA as constituents of hnRNP LL.

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

Table 1. FRET Efficiencies of the Donor−Acceptor Complexes[a]

donor	acceptor	FRET efficiency (%)
RRM1−137-CNTrp	*BCL2*i-22C	31 ± 3
RRM1−104-CNTrp	*BCL2*i-22C	12 ± 4

[a]Average FRET efficiencies calculated from three independent experiments, maintaining the same conditions.

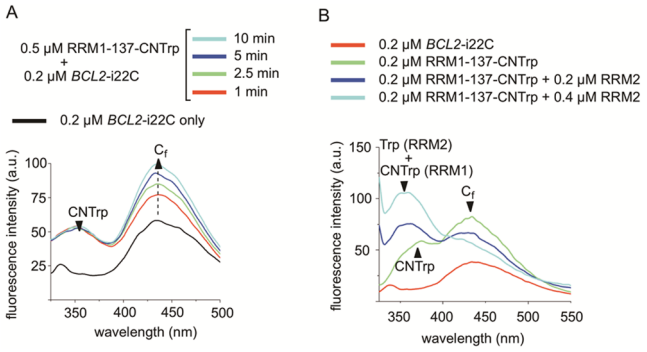


Figure 10. (A) Time-resolved FRET between RRM1−137-CNTrp and *BCL2*i-22C. RRM1−137-CNTrp (0.5 μM) was incubated with 0.2 μM *BCL2*i-22C. The fluorescence emission spectra were measured at four different time points. (B) FRET between RRM1−137-CNTrp (0.2 μM) and equimolar *BCL2*i-22C in the presence or absence of 0.2 μM (blue trace) or 0.4 μM (cyan trace) wild-type RRM2. For the experiments in panels A and B, the fluorescence emission spectra were obtained after incubation 25 °C in 50 mM Tris-HCl, pH 6.6, containing 100 mM NaCl and 1 mM EDTA, following excitation at 310 nm. For the experiment in panel B, RRM1 and the i-motif DNA were preincubated for 10 min prior to the addition of RRM2. The combined solution was then incubated for an additional 10 min before excitation at 310 nm.

DISCUSSION

In a recent study, we demonstrated that the i-motif DNA structure in the *BCL2* promoter region was in equilibrium with a flexible hairpin structure, and that each of these forms could bind selectively to a different steroid identified in the NCI Diversity Library.[35] Cholestane derivative NSC 138948 bound selectively to the i-motif, while pregnanol derivative NSC 59276 bound selectively to the flexible hairpin. Thus, the presence of either of the two steroids shifted the hairpin DNA−i-motif DNA equilibrium (Figure 1) toward the bound form. The finding that NSC 59276 diminished the transcription of *BCL2*, while cholestane derivative NSC 138948 increased *BCL2* transcription, was consistent with a model in which the i-motif participated in expression of the *BCL2* gene.[35] Because overexpressed Bcl-2 is associated with chemoresistance,[6] it was anticipated that NSC 59276 might overcome such resistance. In fact, NSC 59276 was found to induce chemosensitivity to etoposide in a chemoresistant SCID mouse model.[35]

Also investigated as part of the study was the mechanism by which the i-motif participates in the control of gene expression. Transcriptional factor hnRNP LL was shown to bind to the *BCL2* i-motif DNA and activate *BCL2* transcription.[36] Mutation of the i-motif DNA in different defined regions indicated that its binding to hnRNP LL was most strongly dependent on the DNA sequence of the two lateral loops. Bromine footprinting of the i-motif DNA 5 min following admixture of hnRNP LL revealed unwinding of the i-motif, an observation also confirmed using a FRET assay.[36]

In the present study, we have focused on the individual structural domains in hnRNP LL believed to mediate binding to the *BCL2* i-motif DNA. Specifically, domains RRM1, RRM2 and RRM3−4 were cloned in pET28a vectors and expressed separately in *E. coli* (Figure 1C), with 15−20 additional amino acids at both the N- and C-termini of each construct. The three-dimensional structure of each of the RRM domains was modeled based on the solution or X-ray crystal structure of a homologous domain (Figure 1D). Initially, gel shift assays were used to study the binding of the three constructs to the *BCL2* i-motif (Figure 3). RRM1 and RRM2 were both found to bind to the i-motif DNA; the binding of RRM2 was slightly stronger than that of RRM 1. In comparison, RRM3−4 bound only very weakly to the i-motif DNA. These results mimicked those found when the different RRM domains were tranfected into a cellular system and the luciferase activity was measured with the exception that the magnitude of the effect of the RRM3−4 in the luciferase assay was somewhat greater than that seen in the binding assay. A possible explanation for this discrepancy could be due to the presence of other proteins such as helicases in the cellular system.

The types of experiments carried out to characterize the nature of hnRNP LL binding to the i-motif DNA were repeated using the individual structural domains. As shown in Figure 4A, admixture of RRM1 or RRM2 dramatically altered the susceptibility of the i-motif structure to reaction with bromine, undoubtedly reflecting relaxation of the folded i-motif structure by RRM1 and RRM2. For both RRM1 and RRM2, admixture of increasing amounts of either protein to the i-motif DNA caused concentration dependent diminution of the amplitude of the CD spectrum (Figure 4B and 4C), as had also been noted in our earlier study upon admixture of hnRNP LL.[36] Finally, mutational analysis of the i-motif DNA led to the conclusion that the lateral loops L1 and L2 must represent the site of binding for RRM1 and RRM2 (Figures 6 and 7), as had also been found previously for hnRNP LL. It should be noted that our study of the binding of individual RRM domains to the i-motif DNA assumes that the mechanism by which these domains bind to the DNA bears relevance to the mechanism of binding by hnRNP LL itself. While the exact mechanisms of binding by the individual domains are not expected to be exactly the same as that by hnRNP LL, given the DNA affinities of RRM1 and RRM2 established in the present study, it is anticipated that these domains within hnRNP LL initiate i-motif DNA binding. The very similar results obtained for the mutational analysis of i-motif DNA binding by hnRNP LL in comparison with RRM1 and RRM2 supports the assertion that the two mechanisms of binding occur by what may be regarded as analogous processes. Nonetheless, it must be acknowledged that the binding data obtained for RRM1 and RRM2 alone in the present study do not fully support the proposed mechanism of unwinding that requires both domains to bind to independent sites.

As noted above, the unwinding of the i-motif DNA by hnRNP LL was verified 5 min after admixture of the two species by a FRET measurement using an i-motif DNA labeled with the fluorophores fluorescein (FAM) and TAMRA. In the present study, the kinetics of i-motif unfolding following the addition of RRM1 was measured using two complementary strategies. In the first, the unfolding of the i-motif DNA was monitored by the change in the CD spectrum beginning 30 s

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

after the addition of 1 equiv of RRM1 to the DNA. A large decrease occurred within 30 s; after 1 min minimal additional changes in the CD spectrum were observed (Figure 5).

A second approach involved the introduction of a fluorescence donor (6-cyanotryptophan) into RRM1 in lieu of tryptophan residues normally present. These amino acids were chosen based on a computer simulation involving protein hnRNP L, which has been characterized structurally[42] and is homologous to hnRNP LL. The predicted DNA binding residues in RRM1 were identified, enabling the mode of DNA−protein docking to be modeled. On the basis of this analysis, two RRM1 constructs were prepared, one containing 6-CNTrp at position 137, and the other at position 104. These were predicted to be 24.5 and 29.7 Å, respectively, from cytidine nucleotide C22 in the central loop of the i-motif DNA, the latter of which was replaced with the fluorescent tricyclic cytidine nucleotide shown in Figure 8. The Forster distance R_0 between this acceptor and the 6-CNTrp137 donor was calculated as 24.1 Å.[41] Admixture of the fluorescently labeled i-motif DNA to each of the proteins resulted in FRET, which increased with increasing DNA concentration. A stronger FRET response was noted for the RRM1 containing 6-CNTrp at position 137 (Figure 9A).

The RRM1 construct containing 6-CNTrp at position 137 was employed for two additional experiments. The first involved a time-resolved FRET experiment in which FRET was measured at fixed times following admixture of the fluorescently labeled i-motif DNA to RRM1. There was an increase in FRET from 1 to 10 min, consistent with the interpretation that the protein−DNA complex may continue to adjust to a final conformation after the i-motif DNA has been unwound (cf. Figures 5 and 10A). Additionally, a competition experiment involving the addition of (unlabeled) RRM2 to the preformed complex between fluorescently labeled RRM1 and i-motif DNA resulted in a diminution of the FRET signal after an additional 10 min incubation, indicating that RRM2 may be capable of displacing RRM1 from the i-motif DNA. Given that RRM2 has a greater affinity and is more competitive for binding to the i-motif DNA than is RRM1, it may be inferred that RRM2 of hnRNP LL initially targets and binds to one of the lateral loops of the BCL2 i-motif, leading to the subsequent binding of RRM1 to the other lateral loop to unfold the DNA structure for transcriptional activation.

Because of the inherent conformational flexibility of the i-motif which is evidenced by its dynamic equilibrium between the fully folded i-motif and the flexible hairpin,[36] it is difficult to define precisely when hnRNP LL unfolds the i-motif structure leading to the final thermodynamically stable species. This ultimate active binary complex which leads to transcriptional activation involves binding of the lateral loop sequences through RRMs 1 and 2 with additional involvement of a third RRM from the 3/4 binding pair with the furthest 3′-GCCC consensus binding sequence. We do know is that there is a 13-nucleotide optimum distance between the two consensus binding sites found between the lateral loops[36] and preorganization of the lateral loops provided by the structured i-motif which confers entropic and kinetic advantages for the initial hnRNP LL binding event. The subsequent steps are less well-defined. As a consequence of the dynamic nature of the i-motif, an unfolded form becomes available at some stage so that the presumably less constrained 3′-GCCC sequence is in proximity to the third RRM domain which is part of the RRM 3/4 pair binding site. This leads to the thermodynamically stable form of the binary complex between hnRNP LL and the unfolded i-motif sequence. Our studies reported here provide a more detailed view of the protein−DNA interactions leading to this complex.

Finally, we wish to discuss how the initial recognition of the two lateral loops and subsequent unfolding by hnRNP LL of the *BCL2* i-motif fits into the more complete picture of a potential switching mechanism for turning on and off *BCL2* gene expression by the G-quadruplex forming sequence on the guanine rich strand, and the i-motif forming sequence on the cytosine rich strand. The three overlapping G-quadruplexes that can form from the six runs of guanine are in dynamic equilibrium and when stabilized lead to inhibition of gene expression.[51] This suggests that these two alternative secondary DNA structures formed on the opposite strands, which are accessed from the duplex DNA under negative supercoiling, provide a potential on/off switching mechanism for control of BCL2 expression. While there is evidence from population molecular dynamics using laser tweezer experiments that both structures can exist at the same time (Cui et al.),[52] it is not clear at this point how the binding and unwinding of the i-motif by hnRNP LL demonstrated in this contribution affects this dynamic equilibrium and whether this will lead to mutual exclusivity of the two secondary DNA structures.

In conclusion, we have provided direct evidence for the structural elements in hnRNP LL which interact with the BCL2 i-motif DNA, modeled the interaction between one of the protein binding elements and DNA, and provided experimental FRET data consistent with the model. The kinetics of protein−DNA binding and structural reorganization of the complex have been studied using complementary techniques. These results have added significance because they confirm the role of hnRNP LL in the activation of *BCL2* transcription via the i-motif in the promoter element, and this is the first system in which the transcriptional activation role of a naturally occurring i-motif has been demonstrated.

EXPERIMENTAL SECTION

General Experimental Procedures. All i-motif DNAs were purchased from Integrated DNA Technologies (IDT).

Synthesis of Fluorophores. The synthesis of 6-cyanotryptophan (6-CNTrp) has been described.[41] The preparation of the cytidine analogue 4-aminobenzo[g]quinazoline-2-one (C_f) and its incorporation into a DNA i-motif have been described.[40,41]

Biochemical Procedures. *5′-^{32}P-End-labeling of Oligonucleotides.* *BCL2* i-motif DNA was [5′-^{32}P]-end labeled with γ-^{32}P ATP + T4 polynucleotide kinase enzyme. Ten pmol of DNA was 5′-^{32}P end labeled by incubation with 20 units of T4 polynucleotide kinase and 0.06 mCi [γ-^{32}P]ATP (specific activity 6000 Ci (222 TBq)/mmol) in 50 μL (total volume) of 70 mM Tris-HCl buffer, pH 7.6, containing 10 mM $MgCl_2$ and 5 mM DTT. The reaction mixture was incubated at 37 °C for 1 h followed by purification of DNA by 16% polyacrylamide gel electrophoresis at 1800 V for 2.5 h.

Maxam−Gilbert Sequencing Reaction.[53] Ten μL of a solution of 5′-^{32}P-end labeled DNA (~50 000 cpm) was treated with 25 μL of formic acid and incubated at 25 °C for 4−5 min. The reaction was stopped by treatment with 200 μL of 0.3 M sodium acetate, pH 7.0, mixed with 0.1 mM EDTA and 25 μg/mL tRNA. The resulting solution was treated with 700 μL of ethanol and the DNA was precipitated. The DNA pellet was purified by ethanol precipitation and resuspended in 75 μL of 10% aqueous piperidine. The reaction mixture was incubated at 90 °C for 30 min, and the cooled supernatant was removed under diminished pressure. The DNA pellet was washed and mixed with denaturing loading buffer containing 80% formamide, 2 mM EDTA, 1% bromophenol blue and 1% xylene cyanol, then heated at 90 °C for 10 min and stored at −20 °C. The samples were

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

resolved in a denaturing 16% polyacrylamide gel containing 7 M urea. The gel was run at 2000 V for 2 h.

Design and Synthesis of RRM Constructs. The cDNA of hnRNP LL protein (542 amino acids) was obtained from gene bank (Ac: NP_612403). Two of the constructs, RRM1 and RRM2, were designed with the individual RNA recognition motifs of hnRNP LL, and a third construct was designed which incorporated both RRM3 and RRM4. The constructs were designed with extra amino acids flanking the RRM regions on both termini, and codon optimized for expression in *E. coli*. A 28-nucleotide (TGGTCTCACCCGCAGTTC-GAAAAA) sequence was attached at the 3′-end of the constructs for encoding an additional octapeptide (WSHPQFEK) (Strep-tag) to facilitate purification. In vivo protein expression was done by adding IPTG to an *E. coli* BL-21 (DE-3) cell culture, containing pET28a vector with cDNAs of RRM1, RRM2 or RRM3−4, at an OD_{600} value of 0.6. The cultures were incubated for 6 h at 37 °C before lysing the cells by sonication. The separation of proteins from the cellular debris was performed by centrifugation at 4000 rpm. The supernatant was filtered and loaded directly on a Strep-Tactin column. The protein was eluted using a buffer solution consisting of 100 mM Tris-HCl, pH 8.0, containing150 mM NaCl, 2.5 mM desthiobiotin and 1 mM EDTA. RRM1 and RRM2 were visualized by 15% Tris-glycine SDS-PAGE, whereas RRM3−4 was visualized by12% Tris-glycine SDS-PAGE. Protein concentration was measured by two different methods, one involving UV absorbance, and the other BCA protein quantification. According to the first method, the absorbance of protein samples was measured at two different wavelengths. The protein concentration was determined by use of the equation $C = (1.55 \times A_{280}) - (0.76 \times A_{260})$, where C is the concentration of the protein in mg/mL, A_{280} is the absorbance at 280 nm and A_{260} is the absorbance at 260 nm.[54,55] The protein concentration obtained by this method was comparable with that determined by BCA protein quantification, using a kit provided by ThermoFisher Scientific.

Construction of Plasmids of FLAG-Tagged RRM Domains, Expression in Mammalian Cells, and Dual Luciferase Assay. The RRM domains in the pET28a were amplified by PCR using a pair of primers with restriction sites for *Bam*H1 in the forward primer and *Xho*I in the reverse primer to subclone into pCDNA3.1 with a FLAG-tag sequence at the N-terminal for expression in mammalian cells. The sequences of primers are provided in Figure S9 of the Supporting Information. The constructs were confirmed by sequencing analysis. To test the mammalian expression of the RRM domains with codons optimized for expression in *E. coli,* HEK293TT cells in 6-well plates were transfected with 1 μg of the pCDNA3.1/Flag-RRM constructs (RRM1, RRM2, RRM1−2, and RRM3−4) and full-length hnRNP LL as well as by Fugene HD transfection reagent, and then subjected to immunoblot analysis as described below. After 24−48 h transfection, cells were washed and collected with cold DPBS, lysed with RIPA buffer with protease inhibitor cocktail (Sigma) and 1 mM PMSF. Cell lysates were centrifuged at 14 000 rpm for 20 min to obtain supernatant, and the concentration of total protein was determined by Bradford assay. Proteins (100 μg/well) were separated by 12% SDS-PAGE and then transferred to PVDF membrane with 0.2 μm pore size in 20% MeOH/1X Tris-glycine for 30−40 min. The membrane was incubated in a blocking buffer with 5% BSA and 10% skim milk in TBS-T (0.1% Tween 20) for 90 min at room temperature and then washed three times with TBST buffer prior to overnight incubation with FLAG antibody (1:200, Santa Cruz) in TBST buffer with 5% BSA at 4 °C. For detection, the membrane was washed with TBST buffer and then incubated with secondary antibody, goat antirabbit IgG(H+L) Dylight 800 (1:7,500) in 5% nonfat milk (TBST) for 90 min at room temperature. The fluorescent immune complex bands were detected by LI-COR. In addition, the expressions of RRM domains with cotransfection of 1 μg of pGL3-BCL2 was also confirmed in the same way as shown in Figure S3 of the Supporting Information.

For the dual luciferase assay with RRM domains, HEK293TT cells were transfected with 500 ng of pGL3-*BCL2*,[36] 500 ng of pCDNA3.1/FLAG-RRM domain (RRM1, RRM2, RRM1−2, RRM3−4, or full-length hnRNP LL) and 5 ng of pRL-TK for normalization. After 24 h incubation, cells were lysed with passive lysis buffer and then subjected to dual luciferase assay (Promega) using FB11 luminometer. The ratio of firefly to renilla luciferase of the RRM domains was normalized to that of hnRNP LL to obtain the relative luciferase activity.

Mutagenesis and Subcloning. Two PCR (Polymerase Chain Reaction) site-directed mutagenesis procedures were carried out using a New England Biolabs Q_5 Mutagenesis kit. The forward and reverse primers for each PCR mutagenesis are listed in Figure S10 of the Supporting Information. NEB base changer web-tool was used to design the primers, in order to incorporate an amber codon (TAG) at either Y137 or Y104. PCR reactions were performed in 25-μL reaction mixtures, following the procedure described in the kit manual. Each reaction mixture contained 25 ng of wild-type template RRM1, encoded in a pET28a vector, 125 ng of forward and reverse primers, 10 nmol of dNTPs, 2.5 units of DNA polymerase in 35 mM Tris-HCl, pH 8.0, containing 12 mM potassium acetate, 5 mM DTT and 0.05% Triton X-100 in 0.05 mM EDTA. A combination of the forward and reverse primers, 137-Forward and 137-Reverse or 104-Forward and 104-Reverse were used for the synthesis of RRM1−137TAG and RRM1−104TAG, respectively. The products were ligated with the Q_5 ligase master mix and transformed in DH5α *E. coli* cells. Purified plasmids were verified by sequencing. The product plasmids were the RRM1 constructs with a TAG codon at position 137 or 104.

Preparation of Misacylated tRNAs. Crude NVOC-protected aminoacylated pdCpA samples were dissolved in DMSO. The concentration of the resulting solution was checked by measuring the absorbance of the dinucleotide at 260 nm, and was found to be 96 A_{260} units. The pdCpA solutions were used for tRNA ligation to a final concentration of 5.0 A_{260} units.

The plasmid pYRNA8, which encodes a 74-nucleotide abbreviated tRNA transcript (tRNA-C_{OH}), was transformed into DH5α cells and a single colony was picked for overnight growth in 500 mL LB media supplemented with 100 μg/mL ampicillin. The harvested cells were used for large-scale plasmid extraction following the Promega maxiprep-kit protocol. The plasmid concentration was measured by checking the absorbance at 260 nm and a stock of 1 μg/mL solution was prepared by dissolving the plasmid in water.

Plasmid pYRNA8 was linearized for subsequent transcription using T7 RNA polymerase. The plasmid was digested using *Fok*I restriction endonuclease. The linearized DNA was separated from protein impurities and purified by phenol−chloroform extraction, followed by ethanol precipitation. The digested product was run in a 1% agarose gel along with undigested plasmid as a control. The fastest migrating band was collected and purified. The DNA fragment containing the 74-nucleotide tRNA, expressed under the control of a T7 promoter, was subjected to in vitro transcription. The conventional Ampliscribe T7 transcription protocol (Illumina) was followed; a good yield of 74-nucleotide RNA was obtained by including an additional 500 μg of GMP in the transcription buffer.

Following transcription, the reaction mixture was precipitated by the addition of 40 μL of 3 M NaOAc, pH 5.0, and 1.3 mL of cold EtOH, and was then washed with 70% EtOH. The product was air-dried and dissolved in 300 μL of 0.1 M NaOAc, pH 5.0. The solution was loaded onto 800 μL of DEAE-Sepharose CL-6B resin, equilibrated with the same buffer, and eluted with a [0−0.9 M] step gradient of NaCl. The appropriate fractions were purified and analyzed by electrophoresis on an 8% polyacrylamide−urea gel using 89 mM Tris-borate buffer, pH 8.3, for 1 h at 100 V.

Suppressor tRNA aminoacylation was carried out in 100 mL (total volume) of 100 mM Hepes buffer (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.5, containing 2.0 mM ATP, 15 mM $MgCl_2$, 100 μg of suppressor tRNA-C_{OH}, 5.0 A_{260} units of NVOC-protected 6-cyanotryptophanyl-pdCpA,[41] 15% DMSO and 200 units of T4 RNA ligase. After incubation at 37 °C for 1 h, the reaction was quenched by the addition of 10 μL of 3 M NaOAc, pH 5.2, followed by 300 μL of ethanol. The reaction mixture was incubated at −20 °C for 30 min, then centrifuged at 15 000*g* at 4 °C for 30 min. The supernatant was decanted carefully and the tRNA pellet was washed with 100 μL of 70% EtOH and then dissolved in 30 μL of RNase free H_2O. The NVOC-protected 6-cyanotryptophanyl-tRNA was cooled to 2 °C and

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

irradiated with a 500 W mercury−xenon lamp for 5 min to effect deprotection of the NVOC protecting group.

In Vitro Protein Synthesis. S-30 systems were prepared from *E. coli* BL-21 (DE-3) containing rrnB genes as described.[56,57] The cells were grown at 31 °C in LB medium, supplemented with ampicillin (100 μg/L), until the optical density at 600 nm was 1.0. The cell solution was then diluted 10 times with LB medium and IPTG (500 μg) and growth was continued until the optical density was 2.0. The cells were harvested and lysed using egg lysozyme (150 μg/g of cells). Finally, the S-30 extract was dialyzed, aliquoted in 100 μL microcentrifuge tubes and stored at −80 °C.

Circular plasmid DNAs containing the gene of interest were transcribed and translated in a single reaction in the presence of T7 RNA polymerase. Reactions were carried out following a published procedure[54] in 10-μL reaction mixtures, with 600 ng of plasmid DNA in 35 mM Tris acetate, pH 7.4, containing 190 mM potassium glutamate, 30 mM ammonium acetate, 2 mM DTT, 0.2 μg/μL aminoacyl-tRNA, 3.5% polyethylene glycol (PEG)-6000, 20 μg/μL folinic acid, 20 mM ATP and GTP, 5 mM CTP and UTP, 100 mM mixture of amino acids, 0.1−0.4 μL/μL of S-30 preparation and 100 μM (0.5 μCi/μL)[35] S-methionine.

Aminoacylated tRNAs were added to the reaction mixture to a concentration of 0.3 μg/μL. The reactions were incubated at 37 °C for 1 h. Purification of the expressed protein was done following the Strep-Tactin manufacturer's protocol (Iba-Lifesciences). Reaction yields were analyzed by visualizing the protein band in a 15% Tris-glycine SDS-PAGE gel; quantification of protein was done using a phosphorimager.

In vivo protein expression was done by adding IPTG to an *E. coli* BL-21 (DE-3) cell culture containing RRM1, RRM2 or RRM3−4 plasmids after 3 h. The cultures were incubated overnight at 37 °C, followed by the addition of rifampicin (1 μg/mL) and incubation for another 12−16 h before lysing the cells by sonication. The separation of proteins from the cellular debris was performed by centrifugation at 4000 rpm. The supernatant was filtered and loaded directly on a Strep-Tactin column. The protein was eluted with 50 mM PBS buffer.

Bromine Footprinting Assay. The bromine footprinting assay was adapted from a published procedure.[30,39] Ten pmol of purified, 5′-^{32}P-end labeled *BCL2* i-motif DNA was incubated with bromine which had been formed in situ by admixture of a 50 mM aq solution of KBr and $KHSO_5$ for 20 min at room temperature in a total volume of 20 μL. The reaction was terminated by the addition of 200 μL of aq 0.3 M NaOAc containing 10 mg/mL of calf thymus DNA and 700 μL of cold EtOH. The reaction mixture was incubated at −80 °C for 20 min and the DNA pellet was purified by additional EtOH precipitation steps. The DNA pellet was resuspended in 70 μL of 10% aq piperidine. The reaction mixture was incubated at 90 °C for 30 min, and the cooled supernatant was concentrated under diminished pressure. The DNA pellet was washed with water and mixed with denaturing loading buffer containing 80% formamide, 2 mM EDTA, 1% bromophenol blue and 1% xylene cyanol, then heated at 90 °C for 10 min and resolved in a 16% denaturing polyacrylamide gel containing 7 M urea. The gel was run at 2000 V for 2 h.

Circular Dichroism Measurements. CD experiments were performed based on a published protocol.[30] The spectral data were collected using a Jasco-810 spectropolarimeter and a quartz cell of 1.0 mm path length. The oligonucleotides were dissolved in 50 mM Tris buffer, pH 6.6 or 7.5, containing 100 mM NaCl at a final oligonucleotide concentration of 5 μM. For binding interaction studies, CD spectra of i-motif DNA and DNA−protein mixtures were obtained over a wavelength range of 200−350 nm after incubating DNA and protein at room temperature for 15 min.

The time sensitive CD experiment was carried out at 10 °C. The i-motif DNA sample was treated with one equivalent of RRM1 in the same buffer, then mixed briefly and used for CD measurements for wavelengths between 275 and 310 nm after 0.5, 1.0, 2.5, and 5.0 min. The CD spectra of DNAs were obtained at a scan rate of 100 nm/min and a 1 s response time. The raw data was smoothed and baseline corrected to eliminate the signal contribution from buffer.

The CD spectra of the proteins were obtained at a scan rate of 50 nm/min. Three scans for each sample were performed at 25 °C over a wavelength range of 195−260 nm, and the average means were calculated. The melting points (Tm) were estimated on the basis of CD spectral data obtained at different temperatures (from 4 to 94 °C with 5 °C intervals). The signal intensities at 222 nm were plotted against the corresponding temperature value and smoothed with a nonlinear sigmoidal data fit program.

The molar ellipticity $[\theta]$ was calculated by following the equation, $[\theta] = \theta/(C \times l \times 1000)$ where $[\theta]$ is represented in Kdeg cm^2/dmol, θ is the ellipticity represented in degrees, C is the molar concentration of DNA or protein, and l is the path length in millimeters. The binding of i-motif DNA and individual RRM domains were compared by plotting the differences of the signal intensities of DNA−protein complexes from the i-motif DNA signal at 286 nm with increasing concentration of RRM1 or RRM2. The data was fitted with a nonlinear binding equation (Hill slope).

Electrophoretic Mobility Shift Assay. The *BCL2* i-motif DNA and each of the mutants described in Figure 7 were [5′-^{32}P]-end labeled. These constructs were prepared in 50 mM Tris buffer, pH 7.0, containing 100 mM NaCl. One μL (10 000 cpm) of each of the DNAs was mixed with 10 μM of the same nonradiolabeled oligonucleotide in a 100-fold excess of 50 mM Tris-HCl, pH 6.6, containing 100 mM NaCl and 1 mM EDTA. Ten pmol of each of the DNAs was then incubated with 2.5, 5, 7.5, or 10 pmol of RRM1 or RRM2 for 15 min at room temperature. The DNA−protein mixtures were subjected to 12% non denaturing gel electrophoresis at 80 V for 1.5 h. The resulting bands were visualized using a phosphorimager.

FRET Experiments. The FRET between donor and acceptor was measured experimentally by incubating 5 μM of either RRM1−137-CNTrp or RRM1−104-CNTrp in the presence or absence of 50−200 nM *BCL2*-i22C in 50 μL buffer containing 50 mM Tris-HCl, pH 6.6, 100 mM NaCl and 1 mM EDTA. The proteins were incubated with the DNA for 15 min and then excited at 310 nm. Emission was measured immediately in the range of 325−550 nm.

The competition assay was done by incubating 200 nM of RRM1−137-CNTrp with 200 nM *BCL2*-i22C, in the presence or absence of 200 or 400 nM RRM2, in 50 μL of 50 mM Tris-HCl, pH 6.6, containing 100 mM NaCl and 1 mM EDTA. The proteins were incubated with the DNA for 15 min and then excited at 310 nm. Emission was measured immediately in the range of 325−550 nm. The emission profiles of the donors (RRM1−137-CNTrp, RRM1−104-CNTrp) and the donor−acceptor complexes were normalized, integrated and fitted using the equation, $E = I_A/(I_A + I_D)$, where E is the FRET efficiency; I_A and I_D are the integrated intensities of the donor in unbound and bound forms, respectively.[58]

Graphical Modeling and Simulation of DNA−Protein Interaction. The structure of RRM1 was modeled based on homology modeling using Modeler.[59,60] A pdb file for the 39-nt i-motif was generated using the Make-NA algorithm.[61] The DNA binding residues in RRM1 were found using DP-bind web-tool.[43,44] The predicted active residues were R105, G106, F134, K135, R136, Q137, Q165, F168, N170, Y171. Based on the current study and the previously reported results,[36] the residues of the i-motif involved in binding to RRM1 were found to be in lateral loops L1 and L2. The binding pockets on RRM1 were scanned by Patchdock web-tool for favorable binding interactions,[45,46] which was followed by rigid body docking using Hex-docking algorithm.[47] The data was further analyzed by using the Haddock docking algorithm, using participating residues as "ambiguous-binding restraints".[48−50] Further, the dynamic docking was performed using Haddock expert interface, considering flexible regions of protein and DNA, solvent factors, polar−nonpolar interactions, energy minimization, hydrogen bonding and backbone conformational restraints. The molecules were visualized and virtual mutations were done using PyMol[62] software.

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b05036.

Data illustrating purification of RRM1, RRM2 and RRM3−4, binding of RRM3−4 with the i-motif DNA, bromine footprinting and CD spectra of the i-motif−RRM complexes at pH 7.5, thermal denaturation of the RRM constructs, elaboration of RRM1 containing 6-CNTrp, FRET analysis of wild-type RRM1 in the presence of fluorescently labeled i-motif DNA and the structures of the forward and reverse primers used for PCR mutagenesis. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*sidney.hecht@asu.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health to LHH (GM085585 and CA153821).

■ REFERENCES

(1) Oltvai, Z. N.; Milliman, C. L.; Korsmeyer, S. J. *Cell* **1993**, *74*, 609.
(2) Vaux, D. L.; Cory, S.; Adams, J. M. *Nature* **1988**, *335*, 440.
(3) Joensuu, H.; Pylkkanen, L.; Toikkanen, S. *Am. J. Pathol.* **1994**, *145*, 1191.
(4) Baretton, G. B.; Diebold, J.; Christoforis, G.; Vogt, M.; Muller, C.; Dopfer, K.; Schneiderbanger, K.; Schmidt, M.; Lohrs, U. *Cancer* **1996**, *77*, 255.
(5) Tjalma, W.; De Cuyper, E.; Weyler, J.; Van Marck, E.; De Pooter, C.; Albertyn, G.; van Dam, P. *Am. J. Obstet. Gynecol.* **1998**, *178*, 113.
(6) Reed, J. C.; Miyashita, T.; Takayama, S.; Wang, H.-G.; Sato, T.; Krajewski, S.; Aimé-Sempé, C.; Bodrug, S.; Kitada, S.; Hanada, M. J. *J. Cell. Biochem.* **1996**, *60*, 23.
(7) Reed, J. C.; Kitada, S.; Takayama, S.; Miyashita, T. *Ann. Oncol.* **1994**, *5*, S61.
(8) Schmitt, C. A.; Lowe, S. W. *Blood Cells, Mol., Dis.* **2001**, *27*, 206.
(9) Tracey, L.; Pèrez-Rosado, A.; Artiga, M. J.; Camacho, F. I.; Rodríguez, A.; Martínez, N.; Ruiz-Ballesteros, E.; Mollejo, M.; Martinez, B.; Cuadros, M.; Garcia, J. F.; Lawler, M.; Piris, M. Á. *J. Pathol.* **2005**, *206*, 123.
(10) Reed, J. C.; Pellecchia, M. *Blood* **2005**, *106*, 408.
(11) Fulda, S.; Galluzzi, L.; Kroemer, G. *Nat. Rev. Drug Discovery* **2010**, *9*, 447.
(12) Huppert, J. L.; Bugaut, A.; Kumari, S.; Balasubramanian, S. *Nucleic Acids Res.* **2008**, *36*, 6260.
(13) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908.
(14) Sun, D.; Hurley, L. H. *J. Med. Chem.* **2009**, *52*, 2863.
(15) Kouzine, F.; Gupta, A.; Baranello, L.; Wojtowicz, D.; Ben-Aissa, K.; Liu, J.; Przytycka, T. M.; Levens, D. *Nat. Struct. Mol. Biol.* **2013**, *20*, 396.
(16) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11593.
(17) Brooks, T. A.; Hurley, L. H. *Nat. Rev. Cancer* **2009**, *9*, 849.
(18) Brooks, T. A.; Hurley, L. H. *Genes Cancer* **2010**, *1*, 641.
(19) De Armond, R.; Wood, S.; Sun, D.; Hurley, L. H.; Ebbinghaus, S. W. *Biochemistry* **2005**, *44*, 16341.
(20) Cogoi, S.; Xodo, L. E. *Nucleic Acids Res.* **2006**, *34*, 2536.
(21) Bejugam, M.; Sewitz, S.; Shirude, P. S.; Rodriguez, R.; Shahid, R.; Balasubramanian, S. *J. Am. Chem. Soc.* **2007**, *129*, 12926.
(22) Qin, Y.; Rezler, E. M.; Gokhale, V.; Sun, D.; Hurley, L. H. *Nucleic Acids Res.* **2007**, *35*, 7698.
(23) Sun, D.; Liu, W. J.; Guo, K.; Rusche, J. J.; Ebbinghaus, S.; Gokhale, V.; Hurley, L. H. *Mol. Cancer Ther.* **2008**, *7*, 880.
(24) Palumbo, S. L.; Ebbinghaus, S. W.; Hurley, L. H. *J. Am. Chem. Soc.* **2009**, *131*, 10878.
(25) Drygin, D.; Siddiqui-Jain, A.; O'Brien, S.; Schwaebe, M.; Lin, A.; Bliesath, J.; Ho, C. B.; Proffitt, C.; Trent, K.; Whitten, J. P.; Lim, J. K.; Von Hoff, D.; Anderes, K.; Rice, W. G. *Cancer Res.* **2009**, *69*, 7653.
(26) Qin, Y.; Fortin, J. S.; Tye, D.; Gleason-Guzman, M.; Brooks, T. A.; Hurley, L. H. *Biochemistry* **2010**, *49*, 4208.
(27) Miyazaki, T.; Pan, Y.; Joshi, K.; Purohit, D.; Hu, B.; Demir, H.; Mazumder, S.; Okabe, S.; Yamori, T.; Viapiano, M.; Shin-ya, K.; Seimiya, H.; Nakano, I. *Clin. Cancer Res.* **2012**, *18*, 1268.
(28) Gehring, K.; Leroy, J.-L.; Guéron, M. *Nature* **1993**, *363*, 561.
(29) See, for example: (a) Xu, Y.; Sugiyama, H. *Nucleic Acids Res.* **2006**, *34*, 949. (b) Guo, K.; Gokhale, V.; Hurley, L. H.; Sun, D. *Nucleic Acids Res.* **2008**, *36*, 4598.
(30) Kendrick, S.; Akiyama, Y.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2009**, *131*, 17667.
(31) Brooks, T. A.; Kendrick, S.; Hurley, L. *FEBS J.* **2010**, *277*, 3459.
(32) Li, X.; Peng, Y.; Ren, J.; Qu, X. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 19658.
(33) Choi, J.; Kim, S.; Tachikawa, T.; Fujitsuka, M.; Majima, T. *J. Am. Chem. Soc.* **2011**, *133*, 16146.
(34) Brazier, J. A.; Shah, A.; Brown, G. D. *Chem. Commun.* **2012**, *48*, 10739.
(35) Kendrick, S.; Kang, H.-J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2014**, *136*, 4161.
(36) Kang, H.-J.; Kendrick, S.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2014**, *136*, 4172.
(37) Yu, Z.; Gaerig, V.; Cui, Y.; Kang, H.; Gokhale, V.; Zhao, Y.; Hurley, L. H.; Mao, H. *J. Am. Chem. Soc.* **2012**, *134*, 5157.
(38) Dhakal, S.; Yu, Z.; Konik, R.; Cui, Y.; Koirala, D.; Mao, H. *Biophys. J.* **2012**, *102*, 2575.
(39) Ross, S.; Burrows, C. *Nucleic Acids Res.* **1996**, *24*, 5062.
(40) Godde, F.; Toulmé, J.-J.; Moreau, S. *Biochemistry* **1998**, *37*, 13765.
(41) Talukder, P.; Chen, S.; Roy, B.; Yakovchuk, P.; Spiering, M. M.; Alam, M. P.; Madathil, M. M.; Bhattacharya, C.; Benkovic, S. J.; Hecht, S. M. *Biochemistry* **2015**, *54*, 7457.
(42) Zhang, W.; Zeng, F.; Liu, Y.; Zhao, Y.; Lv, H.; Niu, L.; Teng, M.; Li, X. J. *J. Biol. Chem.* **2013**, *288*, 22636.
(43) Kuznetsov, I. B.; Gou, Z.; Li, R.; Hwang, S. *Proteins: Struct., Funct., Genet.* **2006**, *64*, 19.
(44) Hwang, S.; Gou, Z.; Kuznetsov, I. B. *Bioinformatics* **2007**, *23*, 634.
(45) Duhovny, D.; Nussinov, R.; Wolfson, H. J. *Proceedings of the 2nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science* **2002**, *2452*, 185.
(46) Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. *Nucleic Acids Res.* **2005**, *33*, 363.
(47) Ritchie, D. W.; Venkatraman, V. *Bioinformatics* **2010**, *26*, 2398.
(48) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. *J. Am. Chem. Soc.* **2003**, *125*, 1731.
(49) De Vries, S. J.; Van Dijk, A. D. J.; Krzeminski, M.; Van-Dijk, M.; Thureau, A.; Hsu, V.; Wassenaar, T.; Bonvin, A. M. J. J. *Proteins: Struct., Funct., Genet.* **2007**, *69*, 726.
(50) De Vries, S. J.; Van Dijk, M.; Bonvin, A. M. J. J. *Nat. Protoc.* **2010**, *5*, 883.
(51) Dexheimer, T. S.; Sun, D.; Hurley, L. H. *J. Am. Chem. Soc.* **2006**, *128*, 5404.
(52) Cui, Y.; Kong, D.; Ghimire, C.; Xu, C.; Mao, H. *Biochemistry* **2016**, *55*, 2291.
(53) Maxam, A. M.; Gilbert, W. *Methods Enzymol.* **1980**, *65*, 499.
(54) Layne, E. *Methods Enzymol.* **1957**, *3*, 447.
(55) Stoscheck, C. M. *Methods Enzymol.* **1990**, *182*, 50.

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962

(56) Short, G. F., III; Golovine, S. Y.; Hecht, S. M. *Biochemistry* **1999**, *38*, 8808.
(57) Dedkova, L. M.; Fahmi, N. E.; Golovine, S. Y.; Hecht, S. M. *Biochemistry* **2006**, *45*, 15541.
(58) Jameson, D. M. *Introduction to Fluorescence*; CRC Press, Taylor and Francis Group: Boca Raton, FL, 2014; Vol. *8*, p 145.
(59) Eswar, N.; Marti-Renom, M. A.; Webb, B.; Madhusudhan, M. S.; Shen, D. E. M.; Pieper, U.; Sali, A. *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc., 2006; Vol. *15*, p 1.
(60) Marti-Renom, M. A.; Stuart, A.; Fiser, A.; Sánchez, R.; Melo, F.; Sali, A. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291.
(61) Arnott, S.; Hukins, D. W. L.; Dover, S. D.; Fuller, W.; Hodgson, A. R. *J. Mol. Biol.* **1973**, *81*, 107.
(62) *The PyMOL Molecular Graphics System*, Version 1.2r3pre; Schrödinger, LLC.

DOI: 10.1021/jacs.6b05036
J. Am. Chem. Soc. 2016, 138, 10950−10962