

Communication

Enhanced Binding Affinity for an i-Motif DNA Substrate Exhibited by a Protein Containing Nucleobase Amino Acids

Xiaoguang Bai, Poulami Talukder, Sasha M Daskalova, Basab Roy, Shengxi Chen, Zhongxian Li, Larisa M Dedkova, and Sidney M. Hecht

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/jacs.6b11825 • Publication Date (Web): 06 Mar 2017

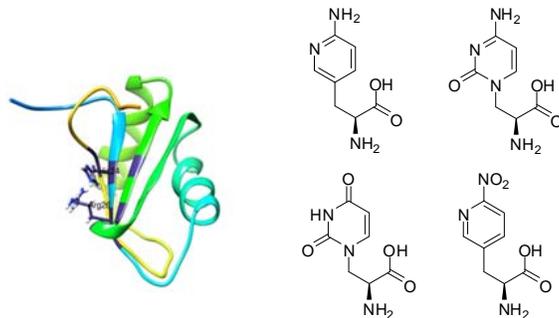
Downloaded from <http://pubs.acs.org> on March 7, 2017

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.



TOC Graphic



Enhanced Binding Affinity for an i-Motif DNA Substrate Exhibited by a Protein Containing Nucleobase Amino Acids

Xiaoguang Bai,[†] Poulami Talukder,[†] Sasha M. Daskalova, Basab Roy, Shengxi Chen, Zhongxian Li,[‡] Larisa M. Dedkova,^{*} and Sidney M. Hecht^{*}

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

Supporting Information Placeholder

ABSTRACT: Several variants of a nucleic acid binding motif (RRM1) of putative transcription factor hnRNP LL containing nucleobase amino acids at specific positions have been prepared and used to study binding affinity for the *BCL2* i-motif DNA. Molecular modeling suggested a number of amino acids in RRM1 likely to be involved in interaction with the i-motif DNA, and His24 and Arg26 were chosen for modification based on their potential ability to interact with G14 of the i-motif DNA. Four nucleobase amino acids were introduced into RRM1 at one or both of positions 24 and 26. The introduction of cytosine nucleobase **2** into position 24 of RRM1 increased the affinity of the modified protein for the i-motif DNA, consistent with the possible Watson-Crick interaction of **2** and G14. In comparison, the introduction of uracil nucleobase **3** had a minimal effect on DNA affinity. Two structurally simplified nucleobase analogues (**1** and **4**) lacking both the N-1 and the 2-oxo substituents were also introduced in lieu of His24. Again, the RRM1 analogue containing **1** exhibited enhanced affinity for the i-motif DNA, while the protein analogue containing **4** bound less tightly to the DNA substrate. Finally, the modified protein containing **1** in lieu of Arg26 also bound to the i-motif DNA more strongly than the wild-type protein, but a protein containing **1** both at positions 24 and 26 bound to the DNA less strongly than wild type. The results support the idea of using nucleobase amino acids as protein constituents for controlling and enhancing DNA-protein interaction. Finally, modification of the i-motif DNA at G14 diminished RRM1-DNA interaction, as well as the ability of nucleobase amino acid **1** to stabilize RRM1-DNA interaction.

There are numerous proteins which play important roles in cellular functions by selective interaction with DNA or RNA. These proteins include those involved in transcriptional regulation, DNA replication and repair of DNA damage. DNA interactive proteins include topoisomerases, gyrases, recombinases, transcription factors and DNA polymerases. Proteins which recognize RNA selectively include aminoacyl-tRNA synthetases, the initiation, elongation and termination factors which participate in protein translation, ribosomal proteins, RNA polymerases and proteins involved in RNA splicing complexes. Many of these proteins take part in essential cellular functions, such that modifying their selective targeting could in principle modulate (aberrant) cellular function.

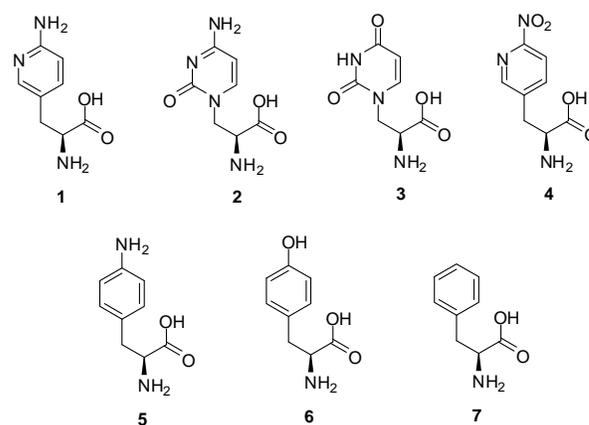


Figure 1. Nucleobase amino acids (**1-4**), and amino acids (**5-7**), prepared and used to activate a suppressor tRNA_{CUA} and incorporated into RRM1 via *in vitro* ribosomal synthesis.

While a number of nucleic acid-protein binding motifs have been documented, including the helix-turn-helix¹ and leucine zipper² DNA binding motifs, the *de novo* design of new/modified protein motifs to enable selective targeting of predetermined DNA and RNA structures is still quite challenging. As all nucleic acid structures are assembled using only a small number of nucleobases, and only a limited number of base pairing/association schemes are operative, a novel strategy for nucleic acid recognition might involve the use of nucleobases as amino acid side chains, enabling nucleic acid recognition by familiar and well understood interactions such as base pairing and stacking.

Although there is as yet no report of any protein which recognizes a nucleic acid target selectively by employing a nucleobase incorporated within the protein, the basic concept is implicit within a number of related studies involving peptides. Of particular note are the efforts of the Mihara laboratory in preparing a series of peptides containing L-amino- γ -nucleobase butyric acids for the purpose of RNA binding.³ For the synthesis of larger protein mimetics, peptides containing a nucleobase amino acid were synthesized chemically and condensed by means of native chemical ligation.^{3b} A number of interesting results were obtained in this study. For example, when three nucleobase acid residues were included within arginine-rich α -helical peptides targeted to a hairpin RNA, the af-

finities varied with the nature of the nucleobases present.^{3a} The introduction of nucleobase acids within the dimerization domain of HIV-1 protease enhanced affinity, but lowered the catalytic activity to some extent.^{3b}

The nucleobase amino acids employed for this study are of a type first described by the Diederichsen laboratory⁴ for the preparation of nucleobase functionalized peptide nucleic acids^{4a} and β -peptides.^{4b} These contain a single methylene group between the nucleobase and α -carbon atom of the derived amino acid, i.e. the same spacing as in the four proteinogenic aromatic amino acids (histidine, phenylalanine, tyrosine and tryptophan). We have recently described⁵ the synthesis of the amino acids corresponding to the five canonical nucleobases in RNA and DNA (uracil, thymine, cytosine, adenine and guanine), two of which (**2** and **3**, Figure 1) are employed in the present study. While these amino acids were previously incorporated into two proteins, the potentially altered functions of the derived proteins were not explored.⁵

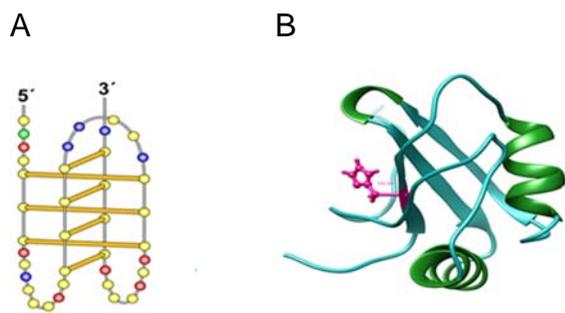


Figure 2. (A) Structure of the *BCL2* i-motif DNA.⁶ The cytidine rich i-motif stem is shown in yellow with two of the lateral loops and the central loop. (B) Human hnRNP LL RRM1 domain, including His24, generated by I-Tasser software using the structure of *Mus musculus* RRM domain of BAB28521 protein (pdb 1WEX) as a template.¹⁵

Presently, we prepared several constructs of RRM1, one of the nucleic acid recognition motifs of transcription factor hnRNP LL which has recently been shown to bind to the lateral loops of the *BCL2* i-motif DNA and to unfold the DNA.⁶ We identified His24 of RRM1 as a potential binding partner for G14 of the i-motif DNA and report that the modified RRM1 containing cytosine nucleobase **2** in lieu of His24 binds to the i-motif DNA more strongly than wild-type RRM1, while the introduction of uracil nucleobase **3** produces no increase in binding. We also report the syntheses of two

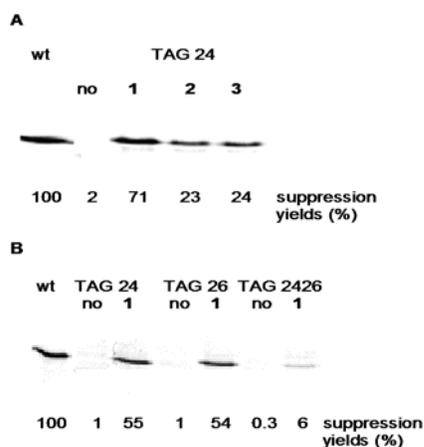
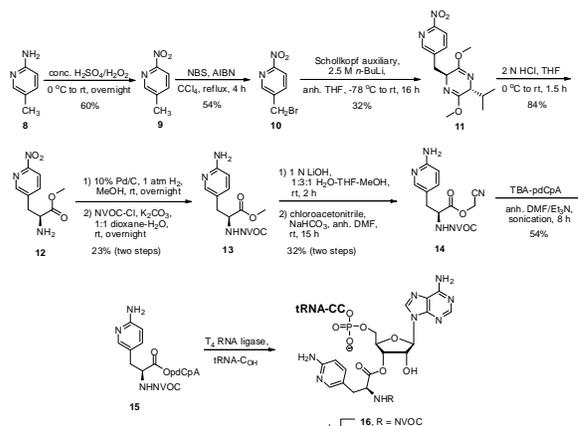


Figure 3. SDS-Polyacrylamide gel showing (A) incorporation of nucleobase amino acids **1**, **2** and **3** into position 24 of RRM1, (B) incorporation of amino acid **1** into position 24, position 26 or both positions of RRM1. Suppression yields are shown. No, absence of activated suppressor tRNA.

new pyrimidine-like nucleobases (**1** and **4**, Figure 1) which lack both the 2-oxo and N-1 substituents of pyrimidines. In spite of its structural differences from **2**, nucleobase **1** seemed to function analogously to **2** in its DNA binding properties when introduced into position 24 of RRM1. As anticipated, nucleobase **4** diminished RRM1–i-motif DNA interaction when introduced into position 24. Also included to facilitate interpretation of the results obtained with the nucleobase amino acids were amino acids *p*-NH₂Phe (**5**), tyrosine (**6**) and phenylalanine (**7**), all of which have been incorporated into proteins previously by nonsense codon suppression.⁷

The chemical synthesis of derivatives of **1** began from commercially available 5-methyl-2-aminopyridine (**8**) (Scheme 1).^{8a} The amino group was oxidized to afford the respective 2-nitropyridine (**9**) in 60% yield, the latter of which was treated with *N*-bromosuccinimide^{8b} in CCl₄ to provide benzylic bromide **10**. Regioselective lithiation of the Schöllkopf chiral auxiliary with *n*-BuLi,^{8c} followed by admixture with **10**, gave **11** which was isolated as a single diastereomer.^{8d} Mild hydrolysis using 2 N HCl then afforded the α -substituted amino acid methyl ester **12** in 84% yield. Following hydrogenation of the nitro group over Pd/C, and NVOC protection of amino acid **13** as the activated cyanomethyl ester **14**, treatment with the dinucleotide pdCpA⁹ provided NVOC-**1** as its pdCpA ester (**15**) in 54% yield. The aminoacylated dinucleotide was then ligated to an abbreviated suppressor tRNA transcript (tRNA_{CUA}-COH) using T₄ RNA ligase, and the NVOC protecting group was removed by photolysis prior to its use for the synthesis of analogues of RRM1. Closely analogous transformations employing synthetic intermediate **12** were used to prepare a suppressor tRNA activated with **4** (SI, Scheme S1).

Scheme 1. Synthesis of the pdCpA and tRNA_{CUA} Esters of Nucleobase Amino Acid **1**



As a model system for studying the interaction of nucleobase-containing proteins with DNA, we chose the RRM1 from hnRNP LL, which has been expressed in a cell free system and shown to bind to the lateral loops of the i-motif DNA formed in the promoter region of *BCL2*.⁶ Based on an analysis of the likely molecular mode of hnRNP LL–DNA interaction using iTASSER,¹⁰ DP-bind,¹¹ and NCBI CDD¹² as well as a knowledge of the important RNA binding roles of the invariant residues His105 and Arg107 in the closely related protein hnRNP L,^{13,14} we chose the analogous residues (His24 and Arg26) in RRM1 for modification. The structures of the i-motif DNA and RRM1 are shown in Figure 2.

Three RRM1 DNA constructs were created, enabling the introduction of different nucleobase amino acids in lieu of His24 or Arg26, or at both of these positions. The yields of modified RRM1s obtained using a suppressor tRNA activated with **1**, **2** or **3** were studied (Figure 3A). As shown in the Figure, the suppression yield of modified RRM1 containing **1** at position 24 was ~3-fold greater than for **2** or **3** (71% vs 23 and 24%).¹⁶ Also studied comparatively

were the efficiencies of incorporation of **1** into positions 24 (55%) and 26 (54%), and into both positions (6%) (Figure 3B). While the yield of the modified RRM1 containing two amino acid **1** residues was lower as anticipated, it was still sufficient to provide material enabling biochemical studies, a critical finding since the implementation of a strategy to use nucleobase side chains in proteins to achieve selective nucleic acid recognition will clearly require the use of two or more closely spaced nucleobase amino acids. Also prepared to aid in interpretation of i-motif DNA binding studies were RRM1s containing nucleobase amino acid **4**, *p*-NH₂Phe (**5**), Tyr (**6**) or Phe (**7**) at position 24.

Table 1. Quantification of the Binding of Modified RRM1s Containing **1 to *BCL2* i-Motif DNA**

Protein	BC ₅₀ value (nM)
RRM1 His24 Arg26 (wt)	96 ± 13
RRM1 1 -24 Arg26	53 ± 12
RRM1 His24 1 -26	63 ± 11
RRM1 1 -24 1 -26	130 ± 28

Table 2. Quantification of the Binding of Modified RRM1s Containing **2 - **7** to *BCL2* i-Motif DNA**

Protein	BC ₅₀ value (nM)
RRM1 His24 Arg26 (wt)	80 ± 10
RRM1 2 -24 Arg26	54 ± 11
RRM1 3 -24 Arg26	82 ± 16
RRM1 4 -24 Arg26	110 ± 10
RRM1 5 -24 Arg26	92 ± 9
RRM1 6 -24 Arg26	141 ± 5
RRM1 7 -24 Arg26	165 ± 4

Four RRM1 samples prepared by *in vitro* protein synthesis (wild-type and three modified proteins containing **1**) were purified by Strep-Tactin chromatography, and analyzed by denaturing SDS-polyacrylamide gel electrophoresis (Figure S3A), verifying the electrophoretic homogeneity of the samples. RRM1 containing **1** at position 24 was analyzed by MALDI-MS/MS (Figure S4), verifying the presence and position of this nucleobase amino acid.

Binding of the *BCL2* i-motif DNA to each of the RRM1s was studied using an electrophoretic mobility shift assay (EMSA).^{17,18} The results are shown in Table 1 and SI, Figure S7A. The RRM1s containing amino acid **1** at position 24 (BC₅₀ 53 ± 12 nM) or 26 (BC₅₀ 63 ± 11 nM) exhibited increased affinity for the *BCL2* i-motif DNA relative to wild-type RRM1 (BC₅₀ 96 ± 13 nM). Thus replacement of either His24 or Arg26 by **1** resulted in energetically improved contacts with the target DNA, relative to the normal amino acids. The selective nature of the interactions with i-motif DNA mediated by **1** at position 24 or 26 can be appreciated from the observation that the simultaneous introduction of **1** into both positions actually substantially diminished the affinity of the resulting RRM1 for the DNA (BC₅₀ 130 ± 28 nM).

Analysis of two other mutant proteins RRM1-24-**2** and RRM1-24-**3**, having cytosine and uracil-based amino acids in position 24, prepared analogously (Figure S3B), provides mechanistic insights (Table 2, SI, Figure S7B). The presence of cytosine nucleobase **2** at RRM1 position 24 also increased DNA affinity relative to wild type (BC₅₀ 54 ± 11 nM), providing affinity similar to that of RRM1 containing **1** at position 24, while RRM1 having uracil nucleobase **3** at that position had an affinity comparable to that of wild-type RRM1 (BC₅₀ 82 ± 16 vs 80 ± 10 nM). While not definitive, the foregoing observations are consistent with the thesis that H-bonding interactions between the nucleobase moieties of amino acids **1** - **3** and the nucleobase of G14 play a key role in enhanced binding of the modified RRM1s to the i-motif DNA. While the enhanced binding observed for RRM1 containing **1** would be consistent with the formation of Watson-Crick like H-bonds between **1** and G14, one might have expected a diminution of RRM1-DNA binding for the RRM1 containing **3**. Plausibly, the RRM1 containing **3** may

bind to G14 via a wobble-type interaction (SI, Figure S8), explaining its more modest affinity.¹⁹ Less surprising was the finding that an RRM1 having nucleobase amino acid **4** at position 24 exhibited greatly diminished DNA affinity (BC₅₀ 110 ± 10 nM) (Table 2). This seems consistent with what might have been expected from the introduction of a nitro substituent with its dipolar covalent bond in lieu of a H-bond donor/acceptor.

Also studied were RRM1 analogues having *p*-NH₂Phe (**5**), Tyr (**6**) or Phe (**7**) at position 24 (Table 2 and SI, Figure 7C). RRM1 containing **5** had a BC₅₀ value of 92 ± 9 nM. This loss of affinity relative to the RRM1 containing **1** at position 24 (BC₅₀ 53 ± 12 nM), argues that the ring N atom must function as a H-bond acceptor, and suggests that enhanced binding in this system is not due to enhanced base stacking. The lesser affinity of RRM1 **6**-24 Arg (BC₅₀ 141 ± 5 nM) reinforces the role of the exocyclic N atom in **1** as a source of enhanced affinity, logically through H-bonding. Introduction of Phe into RRM1 **7**-24 Arg further reduced DNA binding.

Table 3. Quantification of the Binding of Modified RRM1s Containing **1 to wild-type (G14) and Modified (A14) *BCL2* i-Motif DNAs**

Protein	BC ₅₀ value (nM)	
	G14 i-motif	A14 i-motif
RRM1 His24 Arg26 (wt)	84 ± 3	99 ± 8
RRM1 1 -24 Arg26	47 ± 5	81 ± 9
RRM1 His24 1 -26	55 ± 10	78 ± 11

We also studied the effect of altering the i-motif DNA binding partner for RRM1. The binding interaction involves the lateral loops of the i-motif, where G14 is located.⁶ Replacement of G14 with A14 diminished RRM1 affinity for the i-motif (Table 3). Interestingly, while the introduction of **1** at position 24 or 26 resulted in stronger binding of RRM1 to the modified i-motif, the improvement in affinity was much less pronounced than for the native RRM1 sequence (Table 3, SI, Figure S9).

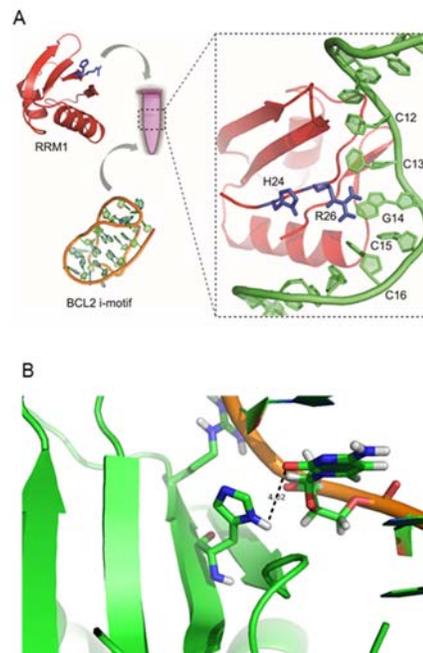


Figure 4. Computationally docked models showing the binding of RRM1 to the i-motif DNA via His24 and Arg26 interaction. The modeling was done using SWISS-MODEL²⁰ homology modeling (A) and HADDOCK²¹ web server was used for data-driven docking (B). A scoring algorithm used factors such as backbone flexibility, surface residues and specific amino acids to rank order the docked structures. The structure shown is a high scoring model illustrating the potential distance between His24 and G14.

We have used computational docking to attempt to understand the molecular basis for the observation that introduction of nucleobase **1** into position 24 or 26 of RRM1 enhanced DNA binding but the presence of **1** in both positions significantly diminished binding. Figure 4A shows one of the most energetically favorable docked structures. The Arg26 guanidine moiety is within H-bonding distance to O⁶ and N² of G14 of the i-motif DNA (as well as O² of C-13), while the His24 imidazole N is more than 7 Å from all DNA bases (and ~8 Å from N² of G14). Another very favorable docked structure is shown in Figure 4B. The His24 imidazole N is ~4 Å from O⁶ of G14, while Arg26 has been displaced close to the DNA backbone. Thus, in these energetically favorable models, either His24 or Arg26 (but not necessarily both) can be H-bonded to G14. This is consistent with the experimental observations noted in Table 1 when **1** was introduced in lieu of one or both of these amino acids.

It seems logical to anticipate that optimizing selective protein–nucleic acid interactions will require the use of multiple nucleobase amino acids in a single protein. In this regard a few observations seem worthy of note. First, because modified nucleobase amino acids such as **1** and **4**¹⁶ appear to be incorporated with greater efficiency than “canonical” nucleobase amino acids such as **2** and **3**, once fully optimized they should facilitate the introduction of multiple nucleobase amino acids into a protein. Second, our recent development of modified ribosomes capable of incorporating dipeptides into proteins²² may enable the introduction of two contiguous nucleobase amino acids in a single ribosomal event. The finding (Table 1) that incorporation of amino acid **1** both into positions 24 and 26 of RRM1 substantially reduced the affinity of this protein for the i-motif (BC₅₀ 130 ± 28 vs 96 ± 13 nM), as compared with the increases in affinity observed following either of the same substitutions made singly, indicates that developing nucleic acid binding proteins that function with favorable affinity and selectivity will demand creative new strategies, and that the full repertoire of interactions noted for nucleic acid interactions, including stacking, van der Waals and electrostatic interactions, may find utility in addition to the apparent H-bonding putatively involved in the current study. Finally, while the increases in protein affinity observed in the present study were not large, they may arguably be sufficient in many cases to alter the specificity for their nucleic acid target. The introduction of multiple nucleobase amino acids into a single protein should also conduce to that outcome. Presumably, increasing affinity would at some point negatively impact biological function.

ASSOCIATED CONTENT

Supporting Information

Experimental procedures for the synthesis and characterization of the pdCpA derivatives of **1** and **4**, for the synthesis and biochemical evaluation of modified RRM1s, and the EMSA data for the protein–DNA binding experiments.

AUTHOR INFORMATION

Corresponding Authors

*Email: sidney.hecht@asu.edu. Phone: (480) 965-6625. Fax: (480) 965-0038.

*Email: larisa.dedkova@asu.edu. Phone: (480) 965-8248. Fax: (480) 965-0038

Present Addresses

‡Z.L.: High & New Technology Research Center, Henan Academy of Sciences, Zhengzhou, 450002, P. R. China.

Author Contributions

†X.B. and P.T. contributed equally.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

This study was supported by Grant GM 103861 from the National Institute of General Medical Sciences, National Institutes of Health. Z. L. thanks the NFSC-joint fund for talent cultivation in Henan Province for a fellowship (U1404206).

REFERENCES

- (1) Brennan, R. G.; Matthews, B. W. *J. Biol. Chem.* **1989**, *264*, 1903.
- (2) (a) Landschulz, W. H.; Johnson, P. F.; McKnight, S. L. *Science* **1988**, *240*, 1759. (b) Vinson, C. R.; Sigler, P. B.; McKnight, S. L. *Science* **1989**, *246*, 911. (c) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809.
- (3) (a) Miyanishi, H.; Takahashi, T.; Mihara, H. *Bioconjugate Chem.* **2004**, *15*, 694. (b) Takahashi, T.; Yana, D.; Mihara, H. *ChemBioChem* **2006**, *7*, 729. (c) Watanabe, S.; Tomizaki, K.; Takahashi, T.; Usui, K.; Kajikawa, K.; Mihara, H. *Peptide Sci.* **2007**, *88*, 131.
- (4) (a) Diederichsen, U. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 445. (b) Chakraborty, P.; Diederichsen, U. *Chem. Eur. J.* **2005**, *11*, 3207.
- (5) Talukder, P.; Dedkova, L. M.; Ellington, A. D.; Yakovchuk, P.; Lim, J.; Ansllyn, E. V. Hecht, S. M. *Bioorg. Med. Chem.* **2016**, *24*, 4177. For the purposes of this study, we define nucleobase amino acids as species having at least one N atom within the aromatic substituent.
- (6) (a) Kang, H.-J.; Kendrick, S.; Hecht, S. M.; Hurler, L. H. *J. Am. Chem. Soc.* **2014**, *136*, 4172. (b) 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. (c) Roy, B.; Talukder, P.; Kang, H.-J.; Tsuen, S. S.; Alam, M. P.; Hurler, L. H.; Hecht, S. M. *J. Am. Chem. Soc.* **2016**, *138*, 10950.
- (7) Gao, R.; Zhang, Y.; Dedkova, L.; Choudhury, A. K.; Rahier, N. J.; Hecht, S. M. *Biochemistry* **2006**, *45*, 8402.
- (8) (a) Zhang, F.; Zhang, S.; Duan, X. F. *Org. Lett.* **2012**, *14*, 5618. (b) Jones, P. A.; Wilson, I.; Morisson-Ive, V.; Jones, C.; Woodcraft, J.; Jackson, A.; Wynn, D. International Patent WO 106564 A2, 2009. (c) Schöllkopf, U.; Groth, U.; Westphalen, K. O.; Deng, C. Z. *Synthesis* **1981**, 969. (d) Talukder, P.; Chen, S. X.; Arce, P. M.; Hecht, S. M. *Org. Lett.* **2014**, *16*, 556.
- (9) Robertson, S. A.; Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Nucleic Acids Res.* **1989**, *17*, 9649.
- (10) Roy, A.; Kucukural, A.; Zhang, Y. *Nat. Prot.* **2010**, *5*, 725.
- (11) Hwang, S.; Gou, Z.; Kuznetsov, I. B. *Bioinformatics* **2007**, *23*, 634.
- (12) Marchler-Bauer, A.; Derbyshire, M. K.; Gonzales, N. R.; Lu, S.; Chitsaz, F.; Geer, L. Y.; Geer, R. C.; He, J.; Gwadz, M.; Hurwitz, D. I.; Lanczycki, C. J.; Lu, F.; Marchler, G. H.; Song, J. S.; Thanki, N.; Wang, Z.; Yamashita, R. A.; Zhang, D.; Zheng, C.; Bryant, S. H. *Nucleic Acids Res.* **2015**, *43*, (database issue), D222.
- (13) Zhang, W.; Zeng, F.; Liu, Y.; Zhao, Y.; Lv, H.; Niu, L.; Teng, M.; Li, X. *J. Biol. Chem.* **2013**, *288*, 22636.
- (14) These two residues correspond to His 74 and Arg 76 in hnRNP LL. A summary of potential DNA interactive residues in RRM1 is summarized in the Supporting Information (SI) Figure S1.
- (15) Zhang, Y. *BMC Bioinformatics* **2008**, *9*, 40.
- (16) The incorporation of nucleobase amino acid **4** into positions 24 and 26 was carried out in a separate experiment; the suppression yields were 61 and 48%, respectively (SI, Figure S2).
- (17) Kendrick, S.; Akiyama, Y.; Hecht, S.M.; Hurler, L.H. *J. Am. Chem. Soc.* **2009**, *131*, 17667.
- (18) The assay was optimized using wild-type RRM1 elaborated in *E. coli* (SI, Figures S5 and S6). Wild-type RRM1 prepared *in vitro* had an affinity for the i-motif DNA quite similar to the sample isolated from *E. coli* (SI, Figure S6).
- (19) For another possible reason for the behavior of RRM1 1-24 Arg26 see: Hildbrand, S.; Blaser, A.; Parel, S. P.; Leumann, C. J. *J. Am. Chem. Soc.* **1997**, *119*, 5499.
- (20) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. *Bioinformatics* **2006**, *22*, 195.
- (21) Dominguez, C.; Boelens, R.; Bonvin, A. M. *J. Am. Chem. Soc.* **2003**, *125*, 1731.
- (22) Maini, R.; Dedkova, L. M.; Paul, R.; Madathil, M. M.; Roy Chowdhury, S.; Chen, S.; Hecht, S. M. *J. Am. Chem. Soc.* **2015**, *137*, 11206.

