JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Subscriber access provided by ISTANBUL TEKNIK UNIV

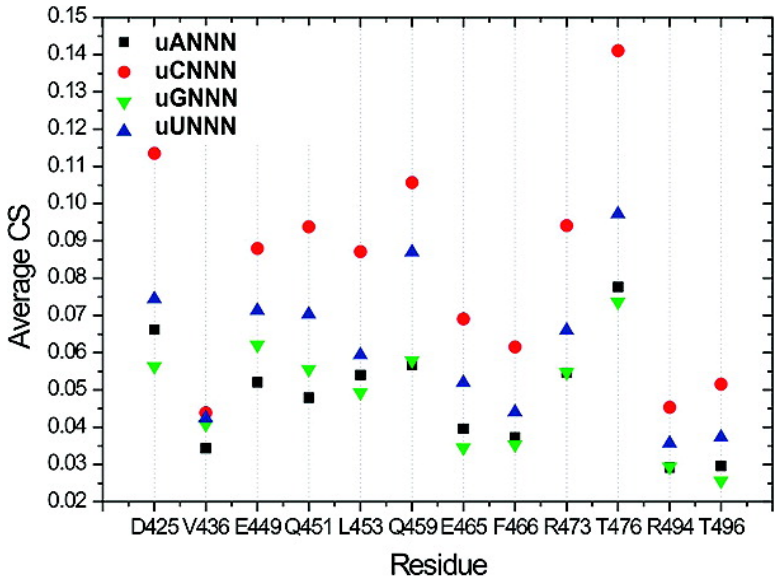
Article

Scaffold-Independent Analysis of RNA–Protein Interactions: The Nova-1 KH3–RNA Complex

Barbara Beuth, Mara Flor Garca-Mayoral, Ian A. Taylor, and Andres Ramos

J. Am. Chem. Soc., **2007**, 129 (33), 10205-10210• DOI: 10.1021/ja072365q • Publication Date (Web): 26 July 2007

Downloaded from http://pubs.acs.org on February 15, 2009



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 2 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

ACS Publications
High quality. High impact.

Journal of the American Chemical Society is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

J|A|C|S
ARTICLES
Published on Web 07/26/2007

Scaffold-Independent Analysis of RNA−Protein Interactions: The Nova-1 KH3−RNA Complex

Barbara Beuth, María Flor García-Mayoral, Ian A. Taylor,* and Andres Ramos*

Contribution from the Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K.

Received April 4, 2007; E-mail: aramos@nimr.mrc.ac.uk; itaylor@nimr.mrc.ac.uk

Abstract: We describe a method to analyze the sequence specificity of an RNA-binding domain. The method, which we have named scaffold-independent analysis, reports on the specificity for each nucleotide position within an RNA target, uncoupled from the surrounding structural and sequence context. We expect this information to improve our understanding of protein−RNA interfaces in ssRNA binding domains (e.g., KH or RRM domains) and to be useful to the design of novel protein−RNA recognition surfaces. Our NMR binding assays using the third KH domain of the Nova-1 protein provide a proof-of-principle for the method and novel information on the specificity of this domain for its RNA targets.

Introduction

Determining the range of sequences that can be accommodated by a nucleic acid binding protein is both an important goal and a challenging problem. While established methods such as the systematic evolution of ligands by exponential enrichment (SELEX) can identify an RNA sequence that binds with high affinity to a protein, it is also important to define the degree of selectivity for each of the nucleotides within the RNA recognition sequence, with disallowed nucleotides being as important as the preferred ones. Establishing selectivity at the single nucleotide level simplifies the understanding of sequence recognition and facilitates the design of proteins that bind to specific RNA sequences. Additionally, it facilitates the functional interpretation of high-resolution structural data. Studies of protein−peptide recognition have used oriented degenerate peptide libraries to establish the selectivity of a protein domain for amino acids at each position of a peptide target sequence.[1] In the field of RNA−protein interactions, equivalent methods are currently not available, and mutational studies represent a costly and lengthy alternative.

To probe the role of individual RNA bases in specifying the affinity of an RNA−protein interaction, we have developed an unbiased method based on the analysis of the interaction between an array of short randomized RNA sequences and the target protein. In this article, the interaction is examined using H^1-N^{15} heteronuclear NMR spectroscopy, and analysis of the chemical shift changes upon RNA binding is used to assess the degree of specificity at each position in the sequence. The use of NMR allows us to operate under equilibrium conditions and to detect interactions in the 0.1−100 μM K_d range. This is the affinity range relevant to the binding of most protein domains to short unstructured nucleic acid sequences (that is, the range that we need to explore if we want to uncouple the specificity of protein−RNA interactions from their structural context).

We have applied the method to the third K homology domain (KH3) of the Nova-1 protein. This domain is structurally well-characterized,[2,3] and its RNA binding specificity has been analyzed in depth using SELEX and electrophoretic mobility shift assay (EMSA),[4,5] allowing for a comparison of our data with the results of these assays. Our results provide a proof-of-principle for the method and confirm its applicability to domains that recognize short single-stranded DNA or RNA sequences, such as KH or RRM domains. The prevalence of KH domains in the human genome (190 in 84 *H. sapiens* proteins; 5743 in 3391 proteins, total database) (SMART database; http://smart.embl.de/)[6] and their association with important neurological diseases such as Fragile X syndrome (FMRP protein) and schizophrenia (QKI protein) make them attractive targets. Likewise, RRM domains have an even higher prevalence (745 in 446 *H. sapiens* proteins; 14 432 in 8767 proteins, total database) and provide another class of important disease-related targets. However, in principle the method is applicable to nucleic acid binding proteins with longer recognition sequences.

Results and Discussion

Published studies have shown that Nova-1 KH3 makes extensive contacts with five nucleotides within its SELEX-derived RNA target. Four of them are recognized with significant specificity, whereas the 5′ nucleotide makes nonspecific contacts with the protein via the 2′ OH and (bridging) phosphate group.[3] The RNA array we have used consists of 16 pools of

(1) Yaffe, M. B.; Cantley, L. C. *Methods Enzymol.* **2000**, *328*, 157−170.
(2) Lewis, H. A.; Chen, H.; Edo, C.; Buckanovich, R. J.; Yang, Y. Y.; Musunuru, K.; Zhong, R.; Darnell, R. B.; Burley, S. K. *Struct. Folding Des.* **1999**, *7*, 191−203.
(3) Lewis, H. A.; Musunuru, K.; Jensen, K. B.; Edo, C.; Chen, H.; Darnell, R. B.; Burley, S. K. *Cell* **2000**, *100*, 323−332.
(4) Jensen, K. B.; Musunuru, K.; Lewis, H. A.; Burley, S. K.; Darnell, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5740−5745.
(5) Musunuru, K.; Darnell, R. B. *Nucleic Acids Res.* **2004**, *32*, 4852−4861.
(6) Letunic, I.; Copley, R. R.; Pils, B.; Pinkert, S.; Schultz, J.; Bork, P. *Nucleic Acids Res.* **2006**, *34*, D257−260.

Table 1. RNA Pools

	position 1	position 2	position 3	position 4
(A) pools	uANNN	uNANN	uNNAN	uNNNA
(C) pools	uCNNN	uNCNN	uNNCN	uNNNC
(G) pools	uGNNN	uNGNN	uNNGN	uNNNG
(U) pools	uUNNN	uNUNN	uNNUN	uNNNU

RNA oligonucleotides. Each individual pool contains 64 pentameric RNAs that contain a nonvarying uridylate as the 5′ ribonucleotide (defined here as position 0 of the oligo). Three of the four other positions are occupied by a randomized mixture of the four bases (N). The fourth, which can be either position one, two, three, or four, is occupied by A, C, G, or U. Therefore, each of the 16 pools contains 64 different RNA sequences that have in common only a single nucleotide at a unique position.

The pool compositions are presented in Table 1, and the complete nucleotide sequences of two pools are shown in Supporting Information Table 1.

NMR can be employed to quantitate the strength of protein−RNA interactions.[7] Protein resonances in fast exchange on the chemical shift time scale will migrate from the free to the RNA-bound position in $^{15}N-^{1}H$ HSQC NMR spectra recorded during the course of an RNA titration. Because the position of a resonance during the titration is proportional to the molar fraction of bound protein, these data can be used to construct binding isotherms. However, in our scaffold-independent analysis (SIA) assay, $^{15}N-^{1}H$ HSQC NMR spectroscopy is used in a comparative fashion to evaluate the relative affinity of the protein for the 16 RNA pools described above. In each of the binding assays, we titrate the protein with a single 64-ribo-oligonucleotide pool (e.g., uNCNN) to a stoichiometry of ∼4:1, RNA−protein. At the final point of the titration, an individual RNA contributes to a small shift of the protein resonances toward the RNA-bound position for that RNA. The final position of each protein peak in an individual binding assay is simply the resultant of all the changes caused by the binding of the 64 RNAs in a pool. Interestingly, the shifts obtained with an ∼4:1 uNCNN−protein ratio are comparable to the ones obtained with a 1:1 uUCAU−protein ratio. This indicates that a significant proportion of the oligos within a pool is bound by Nova-1 (KH3) and that oligos within a pool span a continuous range of affinities. Figure 1 shows $^{15}N-^{1}H$ HSQC spectra recorded during titrations of Nova-1 (KH3) with the uNANN, uNCNN, uNGNN, and uNUNN pools. It is apparent that the same peaks are affected upon RNA(s) binding, and the direction of the shifts observed for these peaks is very similar in all titrations (Figure 1). This observation indicates that the variations that may exist in the bound peak position(s) for the different RNAs are, on average, not large, reflecting a similarity in the binding mode of the RNA oligos. Small differences are averaged out by the simultaneous evaluation of a set of peaks. We can then relate the shifts observed for a set of peaks in a fast regime of exchange to the molar fraction of bound protein. This relationship can be exploited to rank the preference of the protein for a ribonucleotide in a given position. Although it would be interesting to relate this preference with an actual affinity, this is not straightforward because individual binding assays can be biased by the coupling between neighboring nucleotides. It is important to point out that the assignment of the resonances is not a requirement for the application of SIA because the method needs only a set of peaks in the protein $^{15}N-^{1}H$ HSQC spectrum that are responsive to the addition of the RNA pools.

In principle, methyl resonances in $^{13}C-^{1}H$ HSQC spectra can be used for a similar purpose. However, they represent a less attractive alternative mainly because of spectral overlap. In our assays, 12 peaks from the $^{15}N-^{1}H$ HSQC spectrum were chosen where the chemical shift changes upon addition of the RNA pools were the largest and most consistent. In the case of Nova-1 (KH3), these resonances have been assigned to amides in the RNA binding groove and shown to be sensitive to nucleic acid binding.[8] However, SIA is applicable before backbone assignment and/or structure determination, providing valuable information in the early stages of a structural and/or biophysical study of a protein−RNA interaction. Also, the number of pools in the array (and therefore the number of assays we have to undertake) is related not to the total length of the binding sequence but to the number of positions that we want to examine (e.g., only 12 pools are necessary to examine three positions within a longer nucleotide sequence). The limit of SIA in analyzing long sequences depends on the relative contribution of one nucleotide to the overall binding affinity, and this cannot be predicted on the basis of the Nova-1 KH3 data only.

To evaluate the results of our SIA assays, we have displayed in four histograms the chemical shift changes observed upon addition of the 16 pools for each of the 12 residues selected from the Nova-1 (KH3) (Figure 2 and Supporting Information Figure 1). Each histogram contains the data for a single nucleotide position sampled with all four bases. In some cases, the result is particularly striking. For instance, it is clear that at position 2 the interaction with the uNCNN pool causes large chemical shift perturbations to all the resonances from the selected residues, while the other pools have little effect. This qualitative inspection of the data reveals that our analysis is in excellent agreement with previous gel retardation binding assays[4] that demonstrated that Nova-1 (KH3) has stringent requirement for a cytosine base at position 2. To rank the differences between pools in a simple format that allows an immediate understanding of the binding preference, we have first normalized the four shifts displayed in each histogram for each peak to the highest value and then averaged this value over all 12 peaks (Table 2). Comparison of this final score for the four nucleotides in the same position reports on the binding preference(s) of the protein. Inspection of the scores reveals the Nova-1 (KH3) nucleotide preference at each position and enables a consensus sequence to be built. At position 1, C is the preferred base followed by U. At position 2, C is clearly preferred to all other nucleotides, and in position 3, A is best while U is the less favored. In position 4, all nucleotides are tolerated although C less than A, G, or U. It is also apparent that the degree of specificity at the four different positions is quite different, being highest in position 2 and lowest in position 4. A comparison with the results of SELEX/EMSA experiments[4,5] shows that while the two sets of data agree on the sequence preference for the first three positions (YCA) in position 4 SIA allows a G that is excluded by the SELEX/EMSA analysis.

(7) Auweter, S. D.; Oberstrass, F. C.; Allain, F. H.-T. *J. Mol. Biol.* **2007**, *367*, 174−186.

(8) Ramos, A.; Hollingworth, D.; Major, S. A.; Adinolfi, S.; Kelly, G.; Muskett, F. W.; Pastore, A. *Biochemistry* **2002**, *41*, 4193−4201.

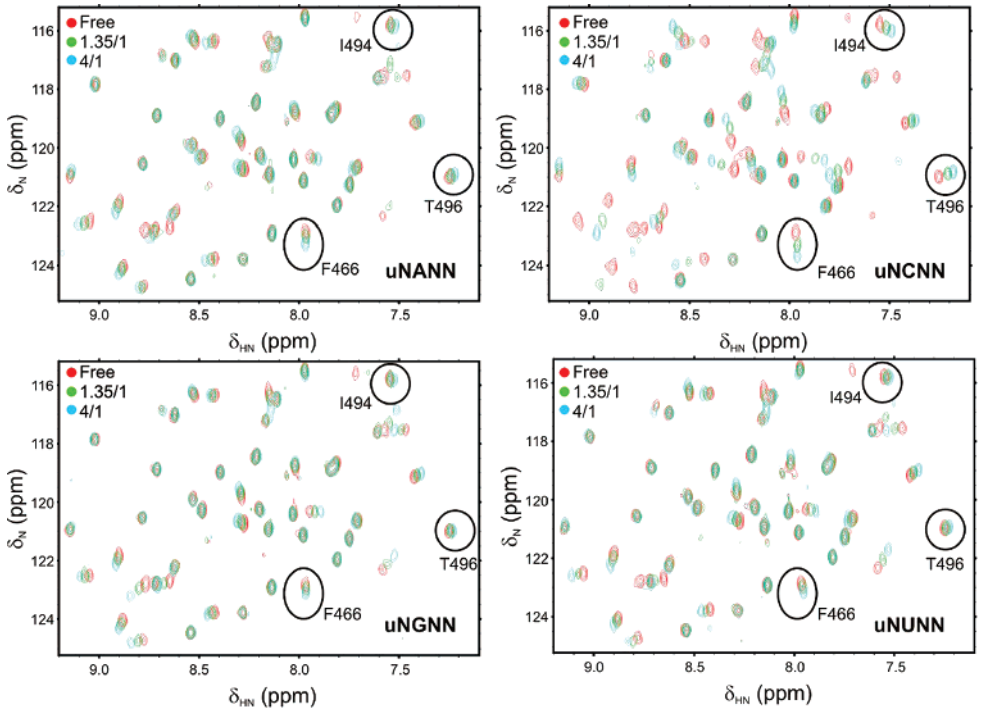


Figure 1. SIA analysis of Nova-1 (KH3). $^{15}N-^{1}H$ HSQC spectra recorded during four titrations of Nova-1 (KH3) with the uNANN, uNCNN, uNGNN, and uNUNN pools. Each panel displays an overlay of the three spectra recorded during the same titration. Free protein is in red, and 1.35:1 and 4:1 RNA−protein ratios are in green and blue, respectively. Three peaks are highlighted to show that the direction of the shifts is approximately the same in the different titrations.

To investigate this discrepancy, we have tested the binding between three short linear RNAs (uUCAU, uGUGU, and uUCAG) and Nova-1 (KH3) using NMR. EMSA and SIA both predict that the protein binds strongly to uUCAU and weakly to uGUGU. However, the two analyses disagree on uUCAG, predicted to bind to the protein 2 orders of magnitude more weakly than uUCAU by EMSA[4] but with similar affinity by SIA. Comparison of NMR binding assays for the three RNAs shows that uUCAU and uUCAG bind Nova-1 (KH3) with similar affinity, while the interaction of the protein with uGUGU is much weaker (Figure 3a), confirming the results of our SIA assays. Isothermal calorimetry was used to validate these NMR results and showed that KH3 has a similar affinity for the uUCAU and uUCAG RNAs (K_d values are 5.9 ± 1.2 μM and 6.2 ± 0.9 μM, respectively; Figure 3b). It is important to point out that our assays make use of short linear RNAs, while the reported EMSA assays were performed on the SELEX target, which presents the four target nucleotides within the loop of an RNA hairpin.

1D ^{1}H NMR spectra show that base pair stability in the SELEX hairpin is greatly increased by introducing a G at position 4 of the specific binding sequence (data not shown), and it is very much possible that the low affinity observed for position 4 is due to a reduced accessibility of the bound sequence.

We have demonstrated that SIA can reliably determine the sequence preference of a KH RNA binding domain. SIA complements the methods currently used to analyze the specificity of nucleic acid binding proteins, providing an assessment of the preference for a base at each position of the recognition sequence and leading to a better understanding of the sequence specificity of RNA binding domains. UV cross-linking immunoprecipitation analysis together with other biochemical and genetic data have shown that the UCAU sequence, a target originally identified by SELEX, is associated with functional Nova-1 targets at the splice site of premessenger RNAs.[9,10] The information now provided by SIA has extended the repertoire of sequences bound by Nova-1 (KH3) allowing for G at position 4 and should facilitate the identification of new in vivo targets of the protein.

The growing ensemble of high-resolution structures of protein−RNA complexes has allowed some tentative rules about the recognition of the core positions of an RNA sequence by a protein domain to be formulated.[11,12] However, specific information defining the contribution of single H-bonds, charge, and hydrophobic interactions to specificity is necessary to complement the structural data and is often not easy to obtain. SIA is an unbiased equilibrium method that requires little or no optimization and provides information on base selectivity of the protein for its RNA target. SIA information can be directly related to high-resolution structural knowledge and used within established methods for the de novo design of RNA binding modules.[13] Importantly, a single RNA pool can be used to test several protein mutants, dissecting the effect of single mutations. It can also be used to test the sequence specificity of domains of very similar sequences, in the same or different organisms, furthering the understanding of the complex networks of recognition at the basis of post-transcriptional regulation.

(9) Ule, J.; Jensen, K. B.; Ruggiu, M.; Mele, A.; Ule, A.; Darnell, R. B. *Science* **2003**, *302*, 1212−1215.
(10) Ule, J.; Stefani, G.; Mele, A.; Ruggiu, M.; Wang, X.; Taneri, B.; Gaasterland, T.; Blencowe, B. J.; Darnell, R. B. *Nature* **2006**, *444*, 580−586.
(11) Auweter, S. D.; Oberstrass, F. C.; Allain, F. H.-T. *Nucleic Acids Res.* **2006**, *34*, 4943−4959.
(12) Backe, P. H.; Messias, A. C.; Ravelli, R. B. G.; Sattler, M.; Cusack S. *Structure* **2006**, *13*, 1055−1067.
(13) Dobson, N.; Dantas, G.; Baker, D.; Varani, G. *Structure* **2006**, *14*, 847−856.

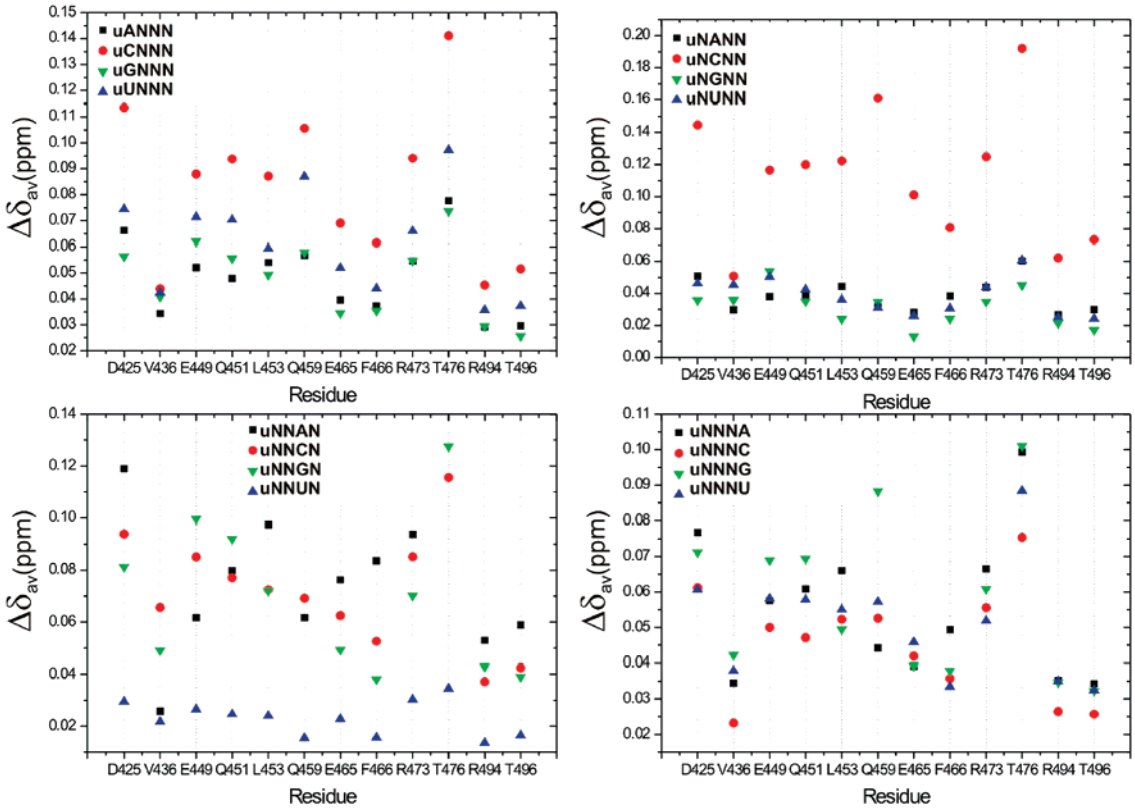


Figure 2. Chemical shift changes of 12 selected residues (see Experimental Section) in the 16 protein−RNA titrations. The data are grouped in four histograms, and in each histogram the 12 residues are displayed on the *x*-axis while the weighted chemical shift changes $\Delta\delta_{av}$ for each residue upon addition of RNA (4:1, RNA−protein ratio) are on the *y*-axis. For each of the 12 residues, four $\Delta\delta_{av}$ values are reported: top left are the values from the uANNN, uCNNN, uGNNN, and uUNNN titrations; top right the values from the uNANN, uNCNN, uNGNN, and uNUNN titrations; and so forth. Comparison of the values within each histogram on a residue basis indicates which nucleotide confers the largest shifts and therefore the greatest average affinity in a specific position.

Table 2. Sequence Selectivity Score[a]

	position 1	position 2	position 3	position 4
(A) pools	0.60	0.37	*0.91*	0.91
(C) pools	*1.00*	*1.00*	0.74	0.74
(G) pools	0.60	0.30	0.66	*0.93*
(U) pools	0.75	0.38	0.28	0.84

[a] Scoring of the results reported in Figure 2. First, the four shifts displayed in each histogram for each peak are normalized to the highest of the four. Then, for each pool, the normalized shift values are averaged over all 12 peaks, and a final score is obtained. This score reports the preference of the protein for a nucleotide in a specific position on a scale from 0 to 1.

Experimental Section

Preparation of Ribo-Oligonucleotides and Nova-1 (KH3). The 16 RNA pools (1 μmol synthesis) were purchased as 5′ biotinylated HPLC-purified oligos from CureVac, Germany. All the pools were reconstituted at an equivalent final concentration of 200 μM in RNAse free 10 mM Tris-HCl, pH 7.5. The three short RNA oligos (uUCAU, uGUGU, and uUCAG) together with the 10021 SELEX-derived hairpin (GCGGACCUAGAUCACCCCGC) and the C-to-G mutant (GCGGACCUAGAUCAGCCCGC) were also purchased from CureVac. The pentameric RNAs were redissolved at 200 μM in the same way as the pools, and the two hairpins were reconstituted directly in 10 mM KPi, pH 6.8, and 25 mM NaCl at 110 μM.

The Nova-1 (KH3) construct, which spans amino acids 421−496 of the Nova-1 sequence, was expressed in BL21(DE3) pLysS cells as a His-tagged protein and isolated using a two-step standard purification procedure as previously described.[8] Briefly, the ^{15}N-labeled protein sample was obtained by expressing the protein in cells growing in M9 minimal media supplemented with ^{15}N ammonium sulfate. After cell lysis, the protein was purified on a nickel-affinity column (Biorad) and by gel filtration (G75 High Load 16/60 Pharmacia). The final protein solution was dialyzed against potassium phosphate buffer, pH 6.8, and 150 mM NaCl and stored at −80 °C.

NMR. Titrations of 25 μM Nova-1 (KH3) samples with the different RNA pools (i.e., uANNN, uCNNN, uGNNN, etc.) were carried out at 25 °C in 10 mM KPi, pH 6.8, and 25 mM NaCl. ^{15}N−^{1}H HSQC spectra were recorded at each step of the titration (RNA−protein ratios of 0:1, 1.35:1, and 4:1). Titrations of 50 μM Nova-1 (KH3) samples with the uUCAU, uGUGU, and uUCAG oligos were carried out in the same buffer and experimental conditions. ^{15}N−^{1}H HSQC spectra were recorded at each step of the titrations (RNA−protein ratios of 0:16, 1:16, 3:16, 9:16, and 16:16). 1D ^{1}H spectra of the 10021 and 10021-mutant RNA molecules were recorded at both 5 and 25 °C.

All NMR spectra were recorded on a Varian INOVA spectrometer operating at 800 MHz ^{1}H frequency. Water suppression was achieved by the WATERGATE pulse sequence.[14,15] The spectra were processed and zero-filled to the next power of two using the NMRPIPE program.[16] Baseline correction was applied when necessary. The HSQC spectra were analyzed and superimposed using the program Sparky.[17] The center of each peak was defined by the Sparky peak-picking routine, and ^{15}N and ^{1}H chemical shift differences were measured manually within the same program. Weighted average values of ^{15}N and ^{1}H chemical shift variations have been calculated as follows: $\Delta\delta_{av} = [(\Delta\delta^{1}H)^2 + (\Delta\delta^{15}N)^2/10]^{1/2}$. The Origin 5.0 suite of programs (OriginLab) was used for data manipulation and display.

(14) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661−665.
(15) Sklenar, V. *Basic Life Sci.* **1990**, *56*, 63−84.
(16) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(17) Goddard, T. D.; Kneller, D. G. *SPARKY*, version 3.110; University of California: San Francisco, 2004.

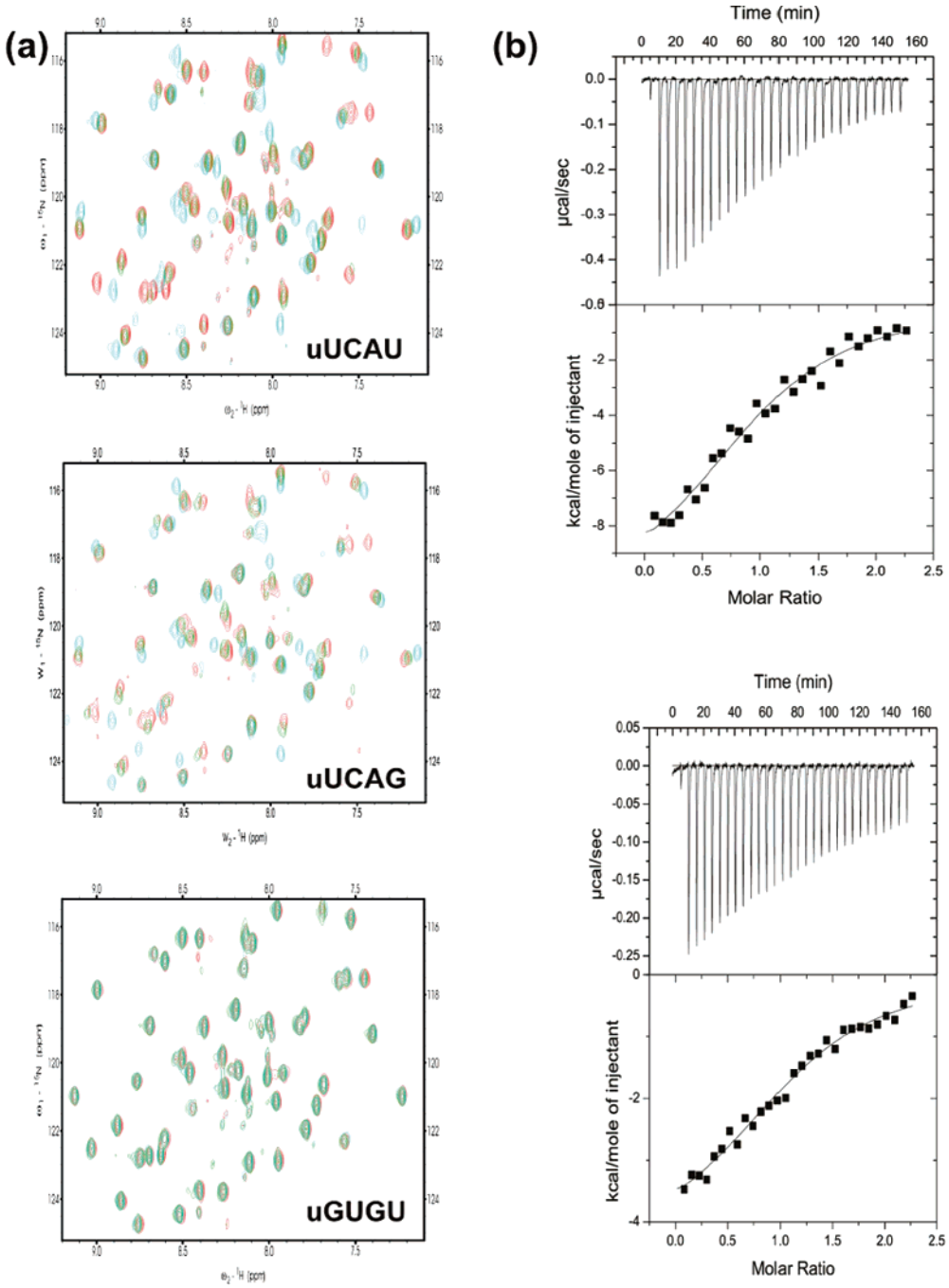


Figure 3. (a) $^{15}N-^{1}H$ HSQC spectra recorded during three titrations of Nova- 1 (KH3) with the uUCAU, uGUGU, and uUCAG RNAs. Each panel displays an overlay of the three spectra recorded during the same titration. The free protein spectrum is in red. The 1:1 and 3:1 RNA−protein spectra are green and blue, respectively. Peaks in the spectra of the uUCAU and uUCAG titrations show similar shifts at similar protein−RNA ratios, while titration with uGUGU results in no shift. (b) Nova-1 KH3−RNA interaction by ITC. Each panel contains a thermogram as a result of the titration (top) and the best fit to the data (bottom). Top panel reports on the uUCAU assay, bottom panel on the uUCAG assay.

Data Scoring and Analysis. To average out the differences that may exist for a single peak, we analyzed the RNA-mediated changes in the chemical shift of a subset (12) of the peaks affected by binding. These peaks, that were in fast or moderately fast regime of exchange, showed the largest chemical shift changes in the uNCNN titration. The same peaks could be clearly followed in all other 15 titrations, with two exceptions due to exchange broadening: T476 could not be followed in the uNNAN titration and Q459 could not be followed in the uUNGN one. Two other peaks in fast exchange (E452 and K462) were instead used in the analysis of the protein specificity for position 3. Full assignments of free and bound Nova KH3 backbone resonances were available from previous work;[8] this assignment shows that the 12 residues selected are scattered across the protein−RNA interface.

ITC. ITC was carried out using a VP-ITC calorimeter (MicroCal). Data was analyzed using the Origin-based software provided by the manufacturers. Briefly, protein and RNA samples were dialyzed in 10 mM KPi, pH 6.8, and 25 mM NaCl. A typical experiment involved 20 μM RNA in the sample cell and 200 μM protein in the injection syringe.

Acknowledgment. We thank Dr. G. Kelly for many useful discussions during the planning and the execution of the experiments. We also thank Dr. T. Frenkiel for critically reading the manuscript and for discussing the method and Drs. P.

Rosenthal and S. Smerdon for critically reading the manuscript. Finally, we thank Dr. A. Pastore for the gift of the Nova-1 (KH3) samples, David Hollingworth for useful discussions, critical reading of the manuscript, and help in the analysis of the samples, and Dr. Katrin Rittinger for help with the ITC measurements. All NMR spectra were recorded at the MRC Biomolecular NMR Centre, National Institute for Medical Research, London, U.K.

Supporting Information Available: Examples of a pool composition (Table 1) and $^{15}N-^{1}H$ HSQC spectra recorded during the titration of Nova-1 (KH3) with the 12 pools with fixed, first, third, and fourth positions (Figure 1a,b). This material is available free of charge via the Internet at http://pubs.acs.org.

JA072365Q