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

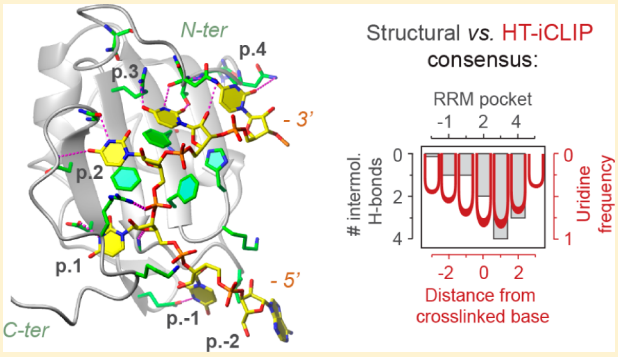
Structural and Mechanistic Insights into Poly(uridine) Tract Recognition by the hnRNP C RNA Recognition Motif

Zuzana Cieniková,[§] Fred F. Damberger,[§] Jonathan Hall,[†] Frédéric H.-T. Allain,*[,§] and Christophe Maris[§]

[§]Department of Biology, Institute of Molecular Biology and Biophysics, ETH Zürich, 8093 Zürich, Switzerland
[†]Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences, ETH Zürich, 8093 Zürich, Switzerland

Supporting Information

ABSTRACT: HnRNP C is a ubiquitous RNA regulatory factor and the principal constituent of the nuclear hnRNP core particle. The protein contains one amino-terminal RNA recognition motif (RRM) known to bind uridine (U)-rich sequences. This work provides a molecular and mechanistic understanding of this interaction. We solved the solution structures of the RRM in complex with poly(U) oligomers of five and seven nucleotides. The five binding pockets of RRM recognize uridines with an unusual 5′-to-3′ gradient of base selectivity. The target recognition is therefore strongly sensitive to base clustering, explaining the preference for contiguous uridine tracts. Using a novel approach integrating the structurally derived recognition consensus of the RRM with a thermodynamic description of its multi-register binding, we modeled the saturation of cellular uridine tracts by this protein. The binding pattern is remarkably consistent with the experimentally observed transcriptome-wide cross-link distribution of the full-length hnRNP C on short uridine tracts. This result re-establishes the RRM as the primary RNA-binding domain of the hnRNP C tetramer and provides a proof of concept for interpreting high-throughput interaction data using structural approaches.



1. INTRODUCTION

The heterogeneous nuclear ribonucleoprotein C1/C2 (hnRNP C) was first identified in the cell nucleus as the major component of the 40S hnRNP core particle.[1] The protein was linked to a range of mRNA-related biological processes. The hnRNP particle is known to package newly transcribed RNA,[2] serving as a molecular ruler that sorts transcripts longer than 230 bases into the mRNA maturation pathway.[3] It is also implicated in splicing regulation. HnRNP C protein is indeed one of the most important alternative exon silencers;[4] transcriptome-wide UV cross-linking studies showed that the hnRNP C particle clusters on pre-mRNA around alternatively spliced or cryptic exons,[5] where it competes with the core splicing factor U2AF65.[6] Outside the nucleus, hnRNP C has been found associated with internal ribosome entry sites of several eukaryotic mRNAs.[7,8]

Human hnRNP C forms in the cell stable tetramers composed of two isoforms, C1 and C2, in a 3:1 stoichiometry.[9,10] Five functionally and structurally distinct regions were identified within each monomer. An RNA recognition motif (RRM, also called RBD) is located in the amino-terminal region (residues 8−87). Its structure was solved in the free state in 2004, showing a canonical $\beta\alpha\beta\beta\alpha\beta$ fold with the characteristic, conserved aromatic residues of the so-called RNP1 and RNP2 consensus motif exposed on the β-sheet.[11] The RRM is followed by an uncharacterized segment of 53 (C1) or 66 (C2) residues. The downstream 39-residues-long segment is highly basic and is thought to be implicated in RNA binding.[12] It is followed by a helical domain of 27 residues, responsible for the tetramerization.[13] Finally, the acidic C-terminal domain appears to stabilize tetramer formation.

Several studies investigated the RNA binding by hnRNP C. Some studies suggested that the RNA binding is independent of the ribonucleotide sequence and is mainly defined by the central basic region.[12,14] In contrast, SELEX experiments on the full-length hnRNP C protein indicated a preference for binding uninterrupted uridine (U) tracts of at least five nucleotides, and localized this recognition within the RRM with flanking termini.[15,16] UV cross-linking data supported sequence-specific recognition of RNA by showing that poly(U) sites of four or more uridines are enriched among the transcriptome-wide hnRNP C targets.[5]

To obtain more detailed information on the interaction of hnRNP C with RNA and establish how the protein achieves sequence-specific recognition of uridine tracts, we determined the NMR structure and dynamics of the RRM in complex with uridine-rich single-stranded RNAs.

2. EXPERIMENTAL PROCEDURES

2.1. Sample Preparation. Human hnRNP C RRM (AA 2−106) was cloned in the pTYB11 vector and transformed into Codon+ cells

Received: July 28, 2014
Published: September 12, 2014

 © 2014 American Chemical Society dx.doi.org/10.1021/ja507690d | J. Am. Chem. Soc. 2014, 136, 14536−14544

which were grown either in LB medium or M9 minimal medium supplemented with [^{15}N]NH_4Cl and [^{13}C]D+glucose. Protein purification was performed as described in the IMPACT-CN protocol (NEB). Unlabeled RNAs were purchased from Dharmacon. Two heptaribonucleotides containing [$^{13}C_S$]-labeled ribose were synthesized using 2′-*O*-TOM-protected ^{13}C-labeled phosphoramidites under standard conditions:[17] A-U*-U-U*-U-U*-C and A-U-U*-U-U*-U-C, where U* is a $^{13}C_S$-labeled uridine.

2.2. ITC Measurements. Measurements were conducted in ITC buffer on a VP-ITC calorimeter (MicroCal) and analyzed with Origin 7.0 (OriginLab). The sample cell was loaded with 2 mL of RNA at 10 μM and the syringe with 400 μL of protein at 100 μM; for weak complexes, the measurement was repeated with increased concentrations.

2.3. NMR Spectroscopy and Structure Calculation. Samples containing 1 mM of protein and RNA diluted in the NMR buffer were measured at 293 K on Bruker AVIII-500 MHz, AVIII-600 MHz, AVIII-700 MHz, and Avance-900 MHz spectrometers equipped with cryoprobes. Standard triple-resonance and NOESY experiments[18] were recorded for the protein assignment and determination of intramolecular NOE restraints. RNA resonance assignment and intramolecular restraints, as well as intermolecular restraints, were obtained from 2D and 3D filtered/filtered and filtered/edited NOESY experiments.[19] Intra-protein NOE peak lists were generated by ATNOS[20] and assigned and calibrated by CYANA 3.0;[21,22] all other NOEs were assigned and calibrated manually. Additional restraints included hydrogen-bond distance restraints and ribose δ angle restraints. The torsion angle dynamics calculation with CYANA was started with 200 random molecules; 50 structures with the lowest target function were further refined in AMBER 9.[23]

2.4. NMR Dynamics. RRM backbone dynamics was investigated with the {^{1}H}^{15}N NOE experiment.[24] The kinetic data were extracted from 2D ^{13}C or ^{15}N ZZ-exchange spectra.[25] The macroscopic rate constants k_{mM}, k_{Mm} were obtained by fitting of ribose ^{13}C ZZ signal intensities to the exchange model described by Demers et al.,[26] which we will call model [1]. Similarly, the kinetic rate constants k'_{on} and k_{off} were determined by fitting the amide ^{15}N ZZ signal intensities to the exchange model [1] or to an extended model [2] taking into consideration a fraction of inactive free protein. The uncertainties (standard deviation, SD) in the binding rates were estimated by Monte Carlo simulations. The ^{15}N longitudinal relaxation rates R_1 of the protein resonances, derived independently from the exponential decay of the amide signal intensities in a 2D ^{1}H−^{15}N HSQC-based N_z relaxation experiment,[27] were used as restraints during exchange fitting.

2.5. Modeling of Multi-register Binding. A three-state system with species *F*, *M*, and *m* corresponding respectively to free RRM, RRM bound in the major RNA register, and RRM bound in the minor register, is described by the observable association constant of the protein-RNA complex *K* and register equilibrium constant K_E. The microscopic association constants K_M and K_m fulfill $K = K_M + K_m$ and $K_E = K_M/K_m$. An indirect register transfer model with two equilibria, $F \rightleftarrows M$ and $F \rightleftarrows m$, admits four microscopic kinetic rate constants, k_{FM}, k_{Fm}, k_{mF}, and k_{MF}, and two macroscopic rate constants, k_{MFm} and k_{mFM}. Assuming $k_{FM} = k_{Fm}$, we obtain $K_E = k_{mF}/k_{MF}$, $k_{off} = (K_E k_{MF} + k_{mF})/(K_E + 1)$, $k_{on} = k_{FM} + k_{Fm}$, $k_{MFm} = k_{MF}/2$, and $k_{mFM} = k_{mF}/2$.

The model describing uridine tract saturation by RRM ligand is based on the approach of Saroff.[28] Three levels of uridine selectivity were defined for the ligand motif: U, u, and N. The affinity penalty of a non-uridine in a U or a u pocket is respectively $1/w$ or $1/v$. The saturation *r* for tracts of length *m* can be expressed as

$$r(m, w, v) = \frac{B_1(m, w, v)K_S[F]}{1 + B_1(m, w, v)K_S[F]}$$

with $[F]$ the free ligand concentration, K_S the affinity of a register perfectly satisfying the motif requirements, and B_1 the weight factor determining the combined affinity of all 1:1 registers. The RRM saturation of cellular uridine tracts is $P(m) = r(m) \times D(m)$, with $D(m)$ the transcriptomic distribution of uridine *m*-mers.[5] $P(m)$ was fitted to hnRNP C iCLIP saturation data[5] for $m < 8$. The SD of the fitted parameters was estimated by Monte Carlo simulation.

More detailed methodological analysis as well as a description of protein and RNA sample preparation procedures and NMR experiments are provided in the Supporting Information (SI).

3. RESULTS

3.1. HnRNP C RRM Binds Uridine Tracts; RNA Sequence Degeneracy Causes Register Exchange. To study spectroscopically the binding of hnRNP C RRM to RNA, we purified a recombinant protein corresponding to residues 2−106 of human hnRNP C (Figure 1A), and performed an initial screen of measurement conditions and RNA targets. The spectra were recorded at a range of temperatures (293−313 K) and in the presence of single-stranded RNA oligomers with lengths of five and seven ribonucleotides and containing up to five contiguous uridines. At 293 K, the binding is in the slow to intermediate regime on the NMR time scale (SI, Figure S1A). We observe substantial chemical shift changes of the protein amides upon addition of RNA; the largest perturbations are observed for the amides at the surface of the β-sheet and in the N- and C-termini (Figure 1B). The increase in the {^{1}H}^{15}N NOE values of the backbone amides located in the N and C-

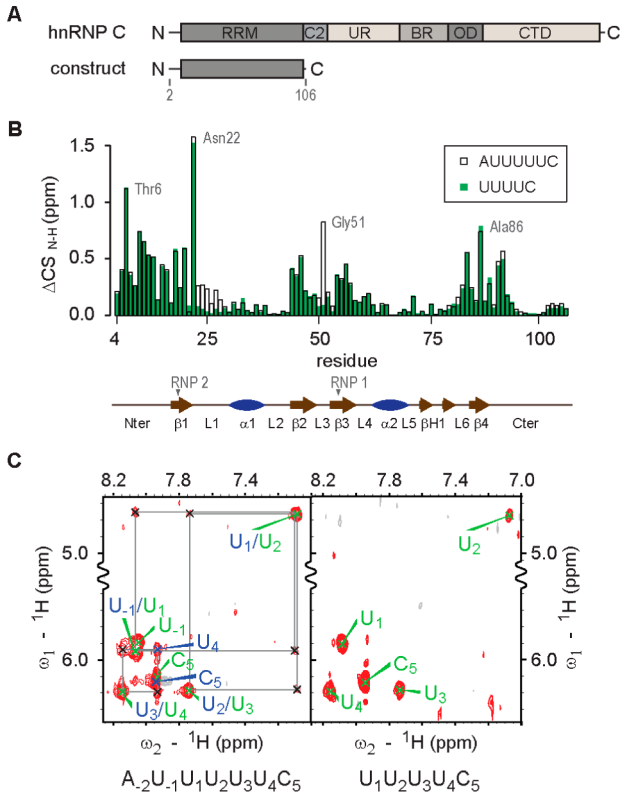


Figure 1. Both the AUUUUUC and the UUUUC oligomers bind the hnRNP C RRM. (A) Schematic representation of the hnRNP C monomer. Abbreviations: RRM, RNA recognition motif; C2, 13 amino-acid insertion in the C2 variant; UR, unstructured region; BR, basic region; OD, oligomerization domain; CTD, C-terminal domain. (B) Combined (^{1}H, ^{15}N) chemical shift perturbation of the RRM main-chain amides between the free and the RNA-bound forms. (C) RNA H5−H6 TOCSY cross-peaks at 1:1 protein:RNA ratio, 293 K. For AUUUUUC, the signals corresponding to the major bound form are labeled in green, those of the minor form in blue, exchange peaks in black. Resonances correlated through exchange are connected with gray lines.

dx.doi.org/10.1021/ja507690d | J. Am. Chem. Soc. 2014, 136, 14536−14544

terminal extremities of the RRM indicates that these protein regions become ordered when bound to RNA (Figure S1B).

While the protein resonances are clearly visible at 1:1 stoichiometry for all screened temperatures and RNA oligomers, the RNA resonances experience severe line broadening at 313 K, suggesting that the bound RNA undergoes chemical exchange (Figure S1C). Decreasing the temperature to 293 K and optimizing the RNA improved considerably the line-widths of RNA resonances. Two sequences, 5′-UUUUC-3′ and 5′-AUUUUUC-3′, were retained for further analysis.

The two RNA sequences bind in a similar manner, as suggested by the highly similar chemical shifts of their bound nucleotides (SI, Table S1). On the protein side, the chemical shift changes of the main-chain amides induced by UUUUC and AUUUUUC have comparable magnitudes—with the exception of the loops L1 and L3, affected only in the case of the longer oligomer (Figure 1B). However, the AUUUUUC target shows exchange cross-peaks between H5 and H6 resonances of neighboring nucleotides in $^{1}H-^{1}H$ TOCSY (Figure 1C, left), indicating the presence of a conformational exchange slow on the NMR time scale. In the case of the UUUUC RNA, this exchange is greatly reduced: in the 2D TOCSY, only one set of H5−H6 resonances is visible, no exchange peaks are observed, and in general, the resonance line-widths improve (Figure 1C, right). We therefore first determined the structure of hnRNP C RRM bound to UUUUC RNA.

3.2. Structure of the hnRNP C RRM in Complex with the 5′-UUUUC-3′ RNA. We calculated 20 low-energy structures of the RRM−UUUUC complex, using 2819 distance restraints (including 128 intermolecular NOE restraints collected at 293 K and nine collected at 278 K, as the U_2 H3 could be observed only at this temperature, SI, Figure S2A). A complete list of intermolecular upper distance restraints and the structural statistics are reported in SI, Tables S2 and S3, respectively.

The overall fold of the hnRNP C RRM ($\beta\alpha\beta\beta\alpha\beta$ with a β-hairpin preceding the fourth β-strand, Figure 2A) is preserved upon complex formation. All four uridines of the UUUUC RNA are ordered and make contacts with the protein (Figure 2B). These nucleotides adopt a C2′-*endo* ribose conformation with the 2′ hydroxyls and the H3′ protons accessible to solvent. The terminal cytidine makes no contacts with the protein and is disordered, in agreement with the distinctly narrower NMR lineshapes of its resonances.

The RRM of hnRNP C docks with the RNA substrate in a canonical fashion, with the β-sheet and the adjacent loops constituting the primary RNA-binding surface. The N- and C-terminal extensions of the RRM provide however the core of the specific contacts to the RNA. The 11 residues C-terminal to the fourth β-strand adopt an extended conformation and interact with the bases and the phosphate backbone. A new secondary structure element—a short α-helix, α0 (represented in light-blue in Figure 2A)—is observed in the segment N-terminal to the first RRM β-strand. This helix, interrupted by a kink induced by Pro11, then enables the arrangement of the ten upstream residues on the surface of the β-sheet, perpendicularly to the strands, in an extended conformation allowing specific recognition of uracil bases. This arrangement is stabilized by numerous hydrogen bonds between the amino-terminus and the RRM (SI, Table S4). In particular, we observe a ionic interaction between the positively charged main-chain amide of

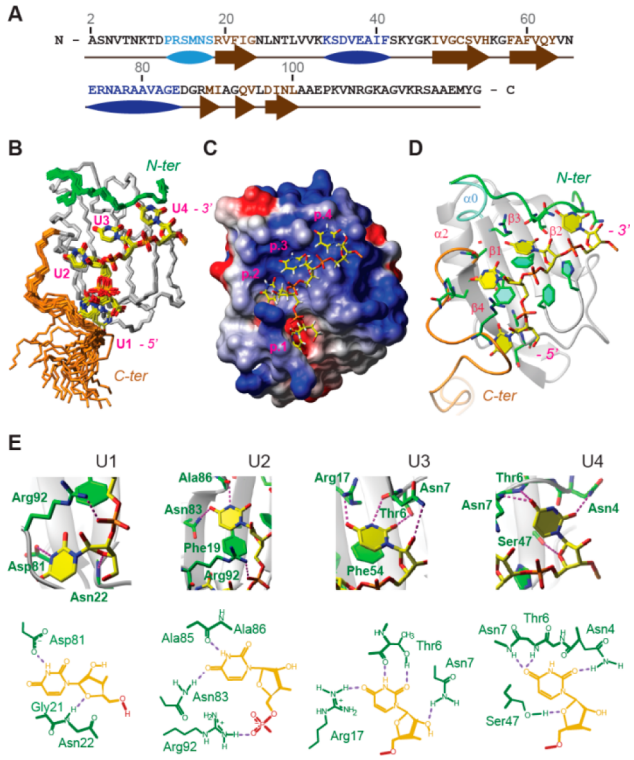


Figure 2. Structure of the hnRNP C RRM bound to the UUUUC oligomer. (A) Sequence of the RRM construct, with highlighted secondary structure elements (blue, α-helices; brown, β-strands). (B) Overlay of 20 energy-best structures of the major conformer of the RRM−UUUUC complex. The unbound cytidine is not shown. (C) Lowest-energy structure, with the RNA visualized as sticks and the protein in van der Waals surface representation with the heavy atoms colored according to their electrostatic potential (blue, positive charge; red, negative charge; white, neutral). Binding pockets are numbered in magenta. (D) Lowest-energy structure with RRM in ribbon view. The amino-terminal α0 helix is labeled in light blue, other secondary structure elements in red. (E) Detailed view of the uridines inside their binding pockets; important molecular contacts are featured in the schematic representation below. Well-supported intermolecular hydrogen bonds (see SI, Table S5 for reporting criteria) are rendered as purple dashed lines.

Ala2 (which is likely to be the N-terminal residue of hnRNP C *in vivo*[29,30]) and the Glu34 side-chain carboxyl.

The bound RNA is stabilized by a dense network of intermolecular hydrophobic packing contacts, hydrogen bonds, and electrostatic interactions. The bases of the second and third uridine stack on the aromatic rings of the conserved RNP motifs, Phe19 and Phe54, respectively. Additional sequence-independent contacts include a hydrogen bond formed between the *trans* amide proton of Asn7 and ribose O2′ of U_3, the main-chain amide of Asn22 stabilizing the U_1 ribose by a hydrogen bond to O4′, and the hydroxyl of Ser47 contacting O4′ of U_4. Finally, the positively charged side chain of Arg92 interacts with the phosphate bridging the first and the second nucleotide (Figure 2E).

An extensive hydrogen-bonding network provides sequence-selective readout of the uracil bases (SI, Table S5), in particular near the 3′ terminal uridines U_3 and U_4. The structure reveals a hydrogen bond between the side-chain amide of Asn4 and the carbonyl group O2 of U_4 base. A bifurcated hydrogen bond is also observed between the O4 carbonyl of U_4 and the backbone amides of Thr6 and Asn7. The existence of the hydrogen bond

dx.doi.org/10.1021/ja507690d | J. Am. Chem. Soc. 2014, 136, 14536−14544

to Thr6 is supported by the distinctive down-field chemical shift change of its amide.

Thr6 is furthermore implicated in the recognition of the downstream U_3 base: its side-chain hydroxyl reaches out toward U_3 O2 and the backbone carbonyl hydrogen-bonds to its H3 imino. The side-chain amide of Arg17 additionally hydrogen-bonds to U_3 O4, completing the recognition of this uracil.

The base of U_2 is very well defined too due to the extensive intermolecular NOE network involving its imino H3 proton (SI, Figure S2A). This imino is exchange protected, corroborating its involvement in a hydrogen bond to Ala85 carbonyl, located in the C-terminal segment following the RRM. Additionally, the side-chain amide of Asn83 possibly interacts with the carbonyl group O4 of U_2.

The recognition of the first uridine is less well supported. Only one structurally conserved hydrogen bond, from the imino of U_1 to the Asp81 carboxylate, is observed. The nucleotide is nevertheless quite well defined, due to numerous intermolecular NOEs originating from the β-sheet and from the C-terminal tail.

Overall, the analysis of the structural data of the hnRNP C RRM−UUUUC complex indicates that the RRM recognizes the consensus sequence 5′-uUUU-3′. The pocket accepting the 5′-end nucleotide indicates a weak preference for uridines: although the nucleobase is recognized through a single intermolecular hydrogen bond, the size of the pocket and the restrained position of the ribose suggest that the recognition of other bases, in particular large purines, is unfavorable. In contrast, the extensive hydrogen-bonding network ensures a strong uracil specificity of the remaining three nucleotide pockets.

3.3. Structure of the hnRNP C RRM in Complex with the 5′-AUUUUUC-3′ RNA. The hnRNP C RRM structure in complex with the UUUUC oligomer allows us to understand how three to four consecutive uridines are specifically discriminated. However, reports indicate that this protein preferentially binds tracts of five or more uridines.[5,15] We therefore decided to structurally investigate the RRM−AUUUUUC complex. We noted earlier that this RNA experiences conformational exchange. The close similarity of chemical shifts of H5−H6 resonances indicates that a nucleotide *i* of one bound conformer shares the same chemical environment (i.e., binding pocket) as the nucleotide *i*+1 of the other conformer, meaning that the AUUUUUC RNA has two binding registers differing by one nucleotide (Figure 1C, left). The population ratio K_E of the two forms is 3.3 ± 0.5 (mean ± SD), as estimated from integrated ribose peak intensities in $^1H-^{13}C$ HSQC spectra of two bound, selectively ^{13}C-labeled RNAs (SI, Figure S2B). Given that the population of the minor conformer is rather low (23%), its NMR signals were weak and could be disregarded. We focused solely on the resonances of the major form in our structural analysis.

A careful inspection of the spectra revealed that the first nucleotide of UUUUC RNA occupies the same position as the third nucleotide of AUUUUUC (in the major conformer), and so forth for the following nucleotides (Figures 1C and 3A). The NOE signal intensity is reduced for the exchanging AUUUUUC complex, but the pattern of the intermolecular NOEs is very similar between the two oligomers (SI, Figure S2C), indicating that the binding of the two RNAs is likely to be identical in the pockets 1−4. We therefore used the intermolecular restraints of the non-exchanging UUUUC complex as input for the structure determination in these

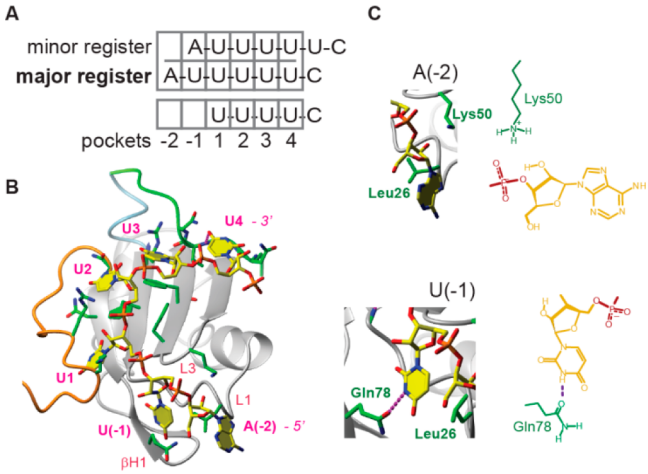


Figure 3. Structure of the hnRNP C RRM with the AUUUUUC RNA in the major conformation. (A) Schematic representation of the RNA binding registers, specifying the arrangement of the nucleotides in the binding pockets. (B) Lowest-energy structure of the complex in ribbon view. Protein secondary structure elements contacting the two 5′-end nucleotides are labeled in red. (C) Arrangement of the two 5′-end nucleotides.

pockets. For the first two nucleotides (AU) of this 7-mer RNA, we could observe 28 intermolecular NOE restraints (SI, Table S2).

The structure shows that the global binding mode of AUUUUUC RNA is very similar to that of UUUUC (Figure 3 and SI, Figure S3; the structural statistics and the observed intermolecular hydrogen bonds are reported in SI, Tables S6 and S5, respectively). The two additional nucleotides A_{-2} and U_{-1} are less precisely defined. In particular, the pocket accepting the first adenosine is rather shallow and does not provide any discernible recognition of the base. The arrangement of this nucleotide is mostly driven by the NOEs to the downstream uridine and the methyl group of Leu26. The base adopts a *syn* conformation and the ribose exposes the O4′ atom to the solvent. The neighboring uridine is partially recognized through its imino proton, which is seen contacting Gln78 side-chain carbonyl. The positively charged side chains of Lys50 and Lys94 possibly stabilize the RNA phosphate backbone; however, these ionic contacts were only observed in 30% and 5% of calculated conformers, respectively.

The structural analysis with the AUUUUUC RNA allows the consensus sequence recognized by the hnRNP C RRM to be extended to 5′-$(N)_U$uUUU-3′. The first nucleotide, (N), is only superficially and non-specifically bound. The second site seems to have a very weak uridine preference, compared to the neighboring downstream sites. Indeed, the Gln side-chain hydrogen-bonding to the U_{-1} imino could potentially act as a donor or an acceptor with any nucleobase. However, the 3:1 population ratio of the major versus the minor AUUUUUC binding register suggests some residual preference for a uracil over an adenine in this site. On the other hand, the binding register consisting of nucleotides 1−5 (UUUUC) bound in the five pockets from −1 to 4 is not observed, clearly indicating that pocket 4 is very important for the overall affinity and also strictly uridine-specific. With this second structure, the picture emerging is one of an RRM binding five nucleotides, with a gradient of uridine specificity increasing in the 5′-to-3′ direction. The strong sites cluster together, rationalizing the preference of hnRNP C for uninterrupted stretches of uridines.

dx.doi.org/10.1021/ja507690d | J. Am. Chem. Soc. 2014, 136, 14536−14544

3.4. Improving Our Understanding of the hnRNP C RRM Specificity Using Affinity Studies. To refine the NMR-derived binding consensus of the RRM, we conducted isothermal titration calorimetry (ITC) measurements with variants of the RRM domain and diverse ribonucleotide sequences (Table 1 and SI, Figure S4). We find that the RNP1/2 motif residues Phe19 and Phe54 are essential for RNA binding. The β-sheet residue His49 also contributes substantially to the overall affinity. Mutating Lys50 to Ala reduces the affinity by a factor of 3.5, which is consistent with this charged residue's favorable interaction with the RNA phosphate backbone.

Table 1. Probing the hnRNP C RRM Specificity Using Isothermal Titration Calorimetry

	K_D (μM)	affinity factor	N	T (K)
	AUUUUUC +			
RRM WT	2.0 ± 0.3[a]	1.0	0.8 ± 0.1	303
K50A	7.0 ± 0.7	3.5	1.0 ± 0.2	303
H49A	40 ± 6	20	1[b]	303
F19A	NA[c]	NA	NA	303
F54A	NA	NA	NA	303
	RRM WT +			
UUUUC	11 ± 2	5.5	0.9 ± 0.1	303
AUCUUUC[d]	23 ± 1	11	0.9 ± 0.1	303
AUUUC[d]	52 ± 5	26	0.9 ± 0.1	303
AUGUUGC	178 ± 12	89	1[b]	303
AUUUUUC[d]	0.44 ± 0.01	0.21	1.0 ± 0.1	293
AUUUUUC[d]	33 ± 6	16	1[b]	313

[a]Values are reported as means ± standard deviations (SD). [b]For these low-affinity conditions, a reasonable fit could be obtained only when setting the expected stoichiometry N to 1. [c]"NA" indicates affinity too low to be quantified. [d]One measurement per condition; the uncertainties on the fitted parameters were estimated from the data spread. All other conditions were recorded in duplicate and the reported fitted values are means and SD of these measurements.

ITC data with different RNAs show that the macroscopic affinity of the RRM for the register-exchanging AUUUUUC RNA is in the micromolar range. Comparison with the UUUUC sequence reveals that the first two nucleotides A_{-2} and U_{-1} (and possibly the register multiplicity) contribute a factor of 5.5 to the overall affinity of the longer RNA. Shortening the uridine tract to three consecutive uridines in the AUUUC RNA leads to an additional 5-fold drop in the binding affinity, confirming that the binding of a base other than uracil is less favorable in the protein pocket 1. A comparable effect is expected for AUCUUUC RNA compared to AUUUUUC. The actual reduction in affinity (11-fold) is larger, suggesting that there is an extra penalty for interrupting the uridine tract in longer oligomers due the positive contribution of the multi-register binding to the overall affinity. Finally, a comparison between the AUCUUUC and the very weakly binding AUGUUGC RNA (7.8-fold affinity decrease) supports our conclusion that three adjacent uridines constitute the minimal high affinity motif.

3.5. Investigating the Register Sliding—A Kinetic Study of the Complex. The RNA register exchange phenomenon observed in the NMR spectra for sequences of five or more uridines appears to be an intrinsic feature of the hnRNP C RRM–RNA interaction. We wondered whether the register shuttling could be a one-step mechanism bypassing the RRM/RNA dissociation step (direct transfer model, i.e., sliding). We therefore used NMR to characterize the kinetics of the RRM binding to AUUUUUC. To determine the macroscopic exchange kinetic constants between the major (M) and the minor (m) RNA-bound forms, we recorded a series of 2D ^{13}C ZZ exchange spectra on samples containing site-specifically ^{13}C-labeled RNA and unlabeled protein.[26] Taking advantage of a labeling scheme where every second nucleotide was ^{13}C-labeled, these experiments allowed observing the RNA magnetization transfer between the neighboring binding registers, giving rise to exchange cross-peaks connecting the two states (Figure 4A and SI, Figure S5A). Five ribose resonances, U_{-1} C–H1′ and C–H4′, U_3 C–H4′, U_4 C–H1′ and C–H4′, showed resolved auto and exchange signals. The exchange rates fitted independently on individual resonances were mutually consistent (SI, Table S7a). The two U_4 ribose resonances had the highest signal-to-noise ratio and presented the best fit quality (Figure 4B and Figure S5B). They give a forward rate k_{mM} = 11.7 ± 1.2 s^{-1} and a backward rate k_{Mm} = 3.6 ± 0.4 s^{-1}. The resulting equilibrium constant K_E = 3.3 ± 0.2 is equal to the population ratio of the two conformations determined independently from a ^{1}H–^{13}C HSQC spectrum (section 3.3).

In order to determine whether the measured exchange rates reflect a direct RNA register transfer process occurring without protein dissociation, we measured next the association and dissociation rates of the RRM. To determine the macroscopic dissociation rate k_{off}, we prepared ^{15}N-labeled protein and the AUUUUUC RNA in an approximate 2:1 stoichiometric ratio and recorded a set of 2D ^{15}N ZZ NMR spectra to observe the exchange of protein amides between the free and the bound state (Figure 4C and SI, Figure S6). Four residues (Phe52, Phe54, Leu84, and Gly93) gave rise to resolved auto and exchange cross-peak signal quartets. Since the magnetization decay curve of the free form resonance indicated a mixture of states, we fitted the data to an extended model [2] accounting for an inactive (but folded) *apo* protein population, decaying more slowly through pure longitudinal relaxation. This model describes the data significantly better than model [1], with an at least 10^4-fold improvement in relative model likelihood, based on the corrected Akaike information criterion (Figure 4D). Fitting the model [1] to a reduced data set excluding the free auto peak data points yields similar results, supporting the robustness of our analysis (SI, Table S7b). The rates obtained by independent fitting of the four residues are mutually consistent; a global fit over all these residues gives a dissociation rate of k_{off} = 18.7 ± 1.4 s^{-1}.

To make inferences about the register transfer mechanism, the macroscopic binding constants need to be translated into intrinsic, microscopic constants. We have sufficient information to do so for the indirect exchange model admitting four intrinsic kinetic rates. Under this model, the two microscopic association constants are $K_M = (7.8 \pm 0.4) \times 10^6$ M^{-1} and $K_m = (2.4 \pm 0.2) \times 10^6$ M^{-1}, knowing $K = 1.0 \times 10^7$ M^{-1} (Figure 4E) which is the sum of the intrinsic affinities and K_E = 3.3 their ratio. The measured macroscopic k_{off} is an average of the k_{MF} and k_{mF} microscopic rates, weighted by the population fraction between M and m, while the macroscopic k_{on} is the sum of the intrinsic on-rates of the two conformers, k_{FM} and k_{Fm}. The on-rates can be assumed to be identical since they involve the same partners and very similar binding surfaces, and then all kinetic constants defined within the indirect transfer model can be

dx.doi.org/10.1021/ja507690d | J. Am. Chem. Soc. 2014, 136, 14536–14544

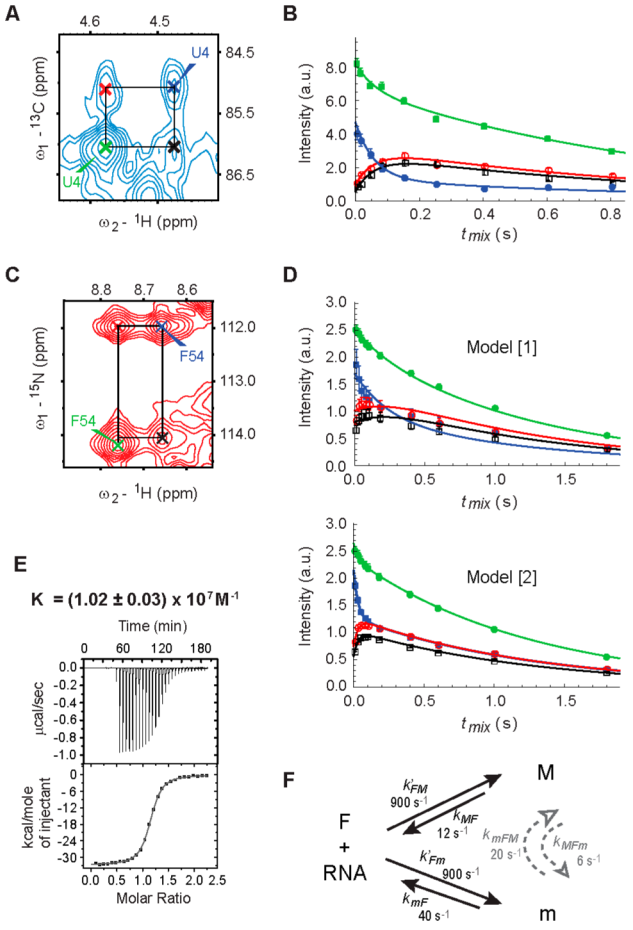


Figure 4. Kinetic rate determination of the hnRNP C RRM - AUUUUUC complex by NMR. (A) 2D $^{1}H-^{13}C$ ZZ exchange spectrum (recorded at t_{mix} = 80 ms) of the RRM-bound AUUUUUC RNA, ^{13}C-labeled at positions −1, 2, and 4. Representative U_4 C4′-H4′ resonance is shown. Signals originating from the major conformation (M) in green, minor (m) in blue, forward $m \rightarrow M$ exchange cross-peak in red and backward in black. (B) Evolution of the magnetization of this ribose resonance with increasing mixing time, and the corresponding exchange curves. The error bars (SD) on integrated magnetization intensities represent the average residual error of the fitted curves. (C) 2D ^{15}N ZZ exchange spectrum at t_{mix} = 80 ms, representative quartet of peaks of the Phe54 residue: free auto peak labeled in green, bound in blue, dissociation cross-peaks in red and association peak in black. (D) Intensities of the exchanging Phe54 peaks, fitted using either the model [1] or [2]. (E) ITC binding isotherm of the hnRNP C RRM titrated to the AUUUUUC RNA at NMR sample conditions. (F) Description of the binding states and transitions of the RRM−AUUUUUC system in the indirect transfer model. Microscopic pathways marked in black, macroscopic transition in gray.

determined (Figure 4F and SI, Table S8). In this model, the rates of the $M \rightleftarrows m$ transition (passing through the state F) are $k_{MFm} = 6.0 \pm 0.7\ s^{-1}$ and $k_{mFM} = 20.0 \pm 2.4\ s^{-1}$. These rates are comparable to or greater than the measured $M \rightleftarrows m$ exchange rates (3.6 ± 0.4 and $11.7 \pm 1.2\ s^{-1}$, respectively). The $M \rightleftarrows m$ exchange is therefore likely to occur as a two-step process; the existence of a faster direct register transfer pathway is not supported by our results. In conclusion, in the case of hnRNP C RRM exchanging between adjacent RNA registers, protein dissociation and re-association is more likely than protein sliding. The observed disparity between the experimental and model-dependent register exchange rates is possibly due to the presence of additional bound species not accounted for in the three-state model. Such complexes would deplete available molecules and decrease the rates of the equilibrium involving F, M, and m.

3.6. Recognition of Poly(U) Lattices. The hnRNP C RRM manifests a multi-register binding to continuous uridine stretches longer than the minimal, $(U)_5$ recognition motif. As seen in the simple case of two binding registers, each register contributes incrementally to the overall affinity (section 3.5). We wanted to understand how this binding behavior and our structural work relate to the recently determined poly(U) RNA interaction pattern of hnRNP C, observed genome-wide by UV cross-linking.[5] We therefore modeled the RRM binding using an approach developed originally for non-specific binders on infinite uniform lattices.[28] In this model, all binding registers have identical micro-affinities K_S and therefore the macroscopic affinity K increases linearly with the number of registers. Typical uridine tracts cross-linking with hnRNP C are, however, rather short (4−15 nt) and the assumption of an infinite lattice is hardly appropriate. The model was therefore extended to properly account for binding to short uridine tracts, by including the contribution of imperfect registers at the tract edges. Since the structural data show that one RRM ligand binds five uridines, with the sites strongly specific for uracils clustering at the 3′ end of the motif, we considered five models for ligand binding consensus: NNUUU, NUUUU, UUUUU, uUUUU, and uuUUU, where N, U, and u represent non-specific, strongly uracil-specific, and weakly uracil-specific binding pockets, respectively. The affinities of registers violating the consensus motif are penalized by a factor of $1/w$ or $1/v$ for each mismatched U or u pocket. Since the focus of this investigation was to understand the determinants governing the binding of a single hnRNP C RRM, we limited the analysis to short uridine tracts below eight nucleotides, and considered uniquely a 1:1 stoichiometric ratio. The ideal, $(U)_5$ register affinity K_S *in vivo* was estimated from ITC data of the RRM-AUUUUUC complex at 313 K as $3.3 \times 10^4\ M^{-1}$, and the free hnRNP C concentration $[F]$ as 10 μM.[15] The ligand saturation functions, representing the binding probabilities to a uridine tract of length m, increase monotonically from almost no RRM binding at tract lengths shorter than the minimal consensus, to equimolar saturation at infinite lengths (Figure 5A). The choice of the model for the ligand consensus impacts most the binding behavior at tract lengths between three and five uridines: while the NNUUU motif allows significant binding to tracts of three uridines, the saturation is negligible for motifs requiring four or more consecutive uridines. The comparison of the U-tract saturation models to the experimental cross-linking data requires taking into account the occurrence of the uridine m-mers in the ensemble of cellular transcripts. Their distribution was approximated by counting the number of uridine tracts in the annotated human transcriptome.[5] The ligand saturation of uridine tracts weighted by their occurrence in the cellular transcript pool exhibits then a maximum at uridine tract lengths of 3−5 uridines, depending on the RRM recognition model used (Figure 5B). These data compare very well with the hnRNP C iCLIP distribution.[5] The consensus model describing best the results of the experimental data—in particular the sharp increase of saturation between uridine tract lengths of three and four, and the maximum reached for uridine pentamers—is the motif sequence uUUUU (SI, Table S9). This consensus is in a very good agreement with the structures

dx.doi.org/10.1021/ja507690d | J. Am. Chem. Soc. 2014, 136, 14536−14544

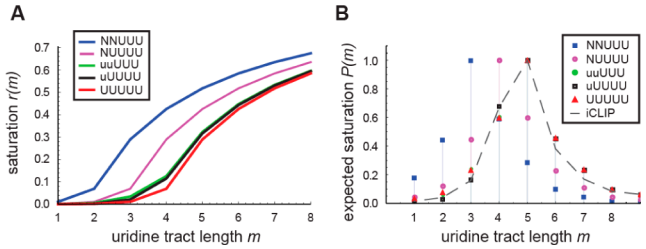


Figure 5. Modeling of multi-register binding of the hnRNP C RRM to poly(U) lattices. (A) Saturation of uridine tracts by the RRM, for different protein recognition motifs. Visualized curves correspond to nucleotide mismatch penalties $w = 10$ and $v = 4$, and $K_S[F] = 1/3$. Dependence of the saturation on these parameters is analyzed in more detail in SI, Figure S7. (B) Transcriptome-wide saturation of uridine tracts. Shown modeled saturation curves correspond to $K_S[F] = 1/3$ and mismatch weights minimized against the *in vivo* iCLIP data[5] for the uuUUU, uUUUU, and UUUUU motifs. The NNUUU and NUUUU motifs are unable to describe the experimental data and curves with $w = 14$ are shown.

of the UUUUC and AUUUUUC complexes. It confirms that five protein pockets indeed discriminate uridines, with the four consecutive downstream sites highly selective for uridines. The fit to the iCLIP data leads to affinity penalty factors for non-uracil bases $v = 5.1 \pm 1.1$ for the 5′-end base and $w = 19.6 \pm 3.7$ for each of the four following bases. These values are comparable with those deduced from ITC ($v \approx 5$ estimated from comparison of affinities of AUUUUUC vs UUUUC, and $w \approx 8$ estimated from the comparison of AUCUUUC and AUGUUGC). Additionally, the model predicts an affinity ratio between the major and minor RRM-binding registers of the AUUUUUC oligomer to be $K_E \approx v = 5.1$, which is close to the value 3.3 ± 0.5 observed in our NMR thermodynamic analysis. Altogether, these results show that the cellular RNA binding properties of hnRNP C can be very nicely explained by the thermodynamic and structural knowledge gained from our study of the isolated RRM domain *in vitro*, in the presence of short poly(U) oligomers.

4. DISCUSSION

4.1. Our Investigation Provides the Molecular Foundation Explaining the Preference of hnRNP C for Oligo(U) RNA. As early as 1992, Dreyfuss and co-workers[31] initiated a solution NMR investigation of $(U)_8$-bound hnRNP C RRM. They showed that the core β-sheet and the N- and C-terminal extensions of the RRM were the most affected by RNA binding. The closely similar amide chemical shift mapping, which we obtained with the AUUUUUC oligomer (with the exception of the Ser13 and Lys30 N−H spins, where our assignments differ), indicates that the RNA used in our study, albeit shorter, captured well the overall binding mode of the hnRNP C RRM to poly(U) RNA. The severe RNA NMR signal line-broadening detected with $(U)_8$ is in line with our observation that degenerated sequences of five or more consecutive uridines experience binding register exchange on the RRM surface. Such a dynamic behavior poses an important challenge for both NMR and X-ray crystallography. Limiting the register exchange by shortening the uridine tract was fundamental for a successful structure determination. A similar approach was previously applied in the case of the PTB RRM complex studied with solution NMR in the presence of the repetitive $(CU)_3$ oligomer.[32]

In an effort to narrow down the determinants of uridine specificity of hnRNP C, Görlach et al.[15] showed that an RRM preserving its flanking extensions can discriminate between uridine-rich and random RNA sequences. Our structure is consistent with this observation since we confirm that the critical features of uridine recognition are located within the N-terminal—and to a lesser extent the C-terminal—region of the isolated protein domain, outside of the RRM core fold. A point-mutation screen of the full-length hnRNP C[16] identified the residues Thr6 and Arg17 as essential for poly(U) recognition; our structure reveals that this is due to their extensive hydrogen-bonding to uracil bases. The Asp81 and Asn83 mutants manifested a moderate loss of binding and base specificity. These side chains surround the less selective binding pockets recognizing the central and 5′-end ribonucleotides; they are indeed involved in intermolecular hydrogen bonds which are however less well structurally conserved. These *in vitro* data are therefore in good agreement with the structurally derived poly(U) recognition consensus of hnRNP C RRM manifesting a 5′-to-3′ gradient of base selectivity.

4.2. Structural Comparison with Other Uridine-Binding RRMs. About a dozen structures of RRMs bound to uridine-rich sequences have been published to date. We examined the similarities and the differences in poly(U) RNA binding between these RRMs. In our comparison, we considered nine proteins (SI, Table S10 and references therein). On average, three to four nucleotides are recognized by an individual RRM. The two binding positions involving the RNP1 and RNP2 motifs (defined as pockets 2 and 3 for the hnRNP C RRM) form the highest number of intermolecular contacts and have the highest preference for uridines (Figure 6A). Regarding the 3′ terminal nucleotide binding pocket (position 4), we noted that the binding mode of hnRNP C is unique. HnRNP C RRM and U2AF65 RRM1 are the only domains recognizing a uracil base in this position sequence-specifically, through contacts provided by residues in the N-terminal region. In both proteins, the amino-terminal extensions form a short α-helix; however, while in the case of U2AF65 this helix directly contacts the uracil base, an elongated terminus folding back onto the β-sheet is only observed for hnRNP C (Figure 6B). Since these two RRMs are not particularly closely related by evolution (SI, Figure S8), the observed structural and RNA-binding similarities possibly stem from a functional convergence due to shared biological context. In fact, hnRNP C and U2AF65 are known competitors[6] for the binding to poly(pyrimidine) tracts (PPT) located directly upstream of 3′ splice sites.[35] The splice site loci, characterized by the highly conserved consensus (C)AG(G), are targeted by spliceosomal regulatory proteins.[36] Weak selectivity of the U2AF65 and hnRNP C pockets recognizing the 3′ end of PPT would potentially lead to a partial coverage of the splice site and disruption of splicing regulation. It appears therefore important that the U2AF65 RRM1 and hnRNP C 3′ nucleotide pockets must be strongly sequence-specific.

4.3. The *in Vivo* Pattern of Short Uridine Tract Recognition by the Full-Length hnRNP C Is Explained by the RNA-Binding Properties of Its RRM Subdomain. Experimental investigation of hnRNP C binding sites through SELEX or high-throughput iCLIP revealed that the protein preferentially interacts with low-complexity poly(U) tracts of highly variable length.[5,15] We have shown with our structural work that the RRM of hnRNP C, which combines a preference for uridine pentamers with a discriminating power of individual

dx.doi.org/10.1021/ja507690d | J. Am. Chem. Soc. 2014, 136, 14536−14544

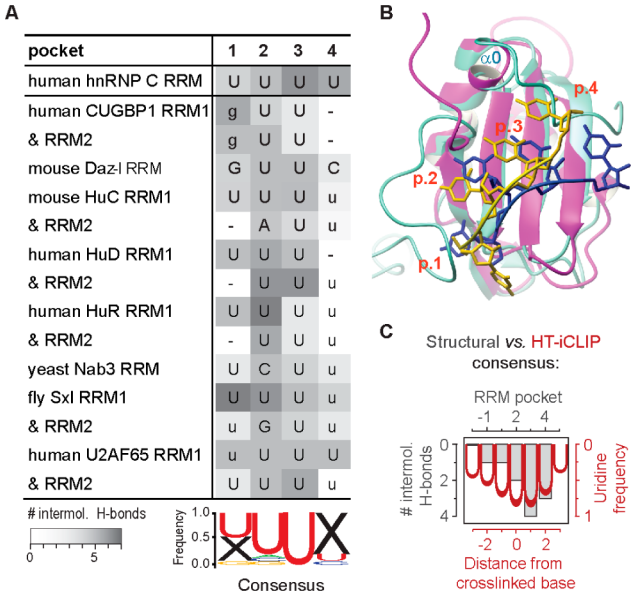


Figure 6. (A) Structural comparison of RRMs binding uridine rich sequences. Lower case indicates a base in a non-conserved arrangement (compared to hnRNP C); upper case stands for a nucleotide similarly oriented as in the hnRNP C structure; a dash is for empty pocket. The number of H-bonds between the nucleotide and the protein domain was quantified by referring to the original publications, with the exception of HuC complex where the quantification was performed using the geometry criteria applied for hnRNP C. The weblogo[33] summarizes the position-specific frequencies of the bases with conserved arrangement. Empty pockets or those occupied by a nucleotide in a non-conserved conformation are represented by ×. (B) Structural overlay of U2AF65 RRM1–poly(U) complex[34] (protein and RNA in magenta and yellow, respectively) and hnRNP C RRM–UUUUC complex (aquamarine and blue). (C) Comparison between the structurally derived uridine selectivity criteria of hnRNP C RRM (in gray) and the transcriptome-wide sequence conservation of full-length hnRNP C cross-linking sites[5] (in red). Number of hydrogen bonds between the uracil base and the protein pocket was derived from the UUUUC complex structure.

binding pockets increasing gradually in the 3′ direction, allows for an extensive multi-register binding on sequences containing four or more uridines. This behavior impacts favorably the overall binding affinity, counteracting the rather weak affinity of individual RRM–RNA interactions. Similar affinity enhancement arising from register multiplicity was previously described for other RNA-binding proteins recognizing nucleobase repeats.[37,38] In addition, the multivalent RRM binding modeled using our structural and thermodynamic data leads to short uridine tract saturation that is remarkably consistent with the full-length hnRNP C binding sites enrichment observed in transcriptome-wide iCLIP experiments.[5] Furthermore, the consensus motif derived from structural characterization of RRM binding to short poly(U) RNAs captures very well the cross-link site consensus motif,[5] with five uridines surrounding the cross-linked nucleotide showing a strong and, again, progressively 5′ to 3′ increasing conservation (Figure 6C). This observation corroborates further the idea that the hnRNP C *in vivo* cross-links originate from the RRM domain, more precisely from the pocket 2. The aromatic ring stacking contact between a uracil and a phenylalanine residue, seen in this position in our structure, is known to be prone to UV cross-linking.[39,40]

The reproducibility of the transcriptome-wide hnRNP C binding behavior using a structure based model of its RRM–RNA interaction reaffirms the (previously contested[12]) key role the RRM subdomain plays in RNA recognition, in the context of the full-length hnRNP C tetramer. It also highlights the applicability of structural biology studies for interpreting genome-wide protein–RNA interactions.

■ ASSOCIATED CONTENT

Supporting Information

Experimental procedures; additional figures and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

frederic.allain@mol.biol.ethz.ch

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank S. Pitsch for providing the ^{13}C-labeled phosphoramidites and M. Zimmermann for excellent assistance with the oligoribonucleotide synthesis. We thank M. Blatter for assistance with structure calculations, and Z. Žigová and C. Giese for helpful discussion of the kinetic analysis. We also thank K. Zarnack, J. König, and J. Ule for providing and discussing the hnRNP C iCLIP data. This research was supported by the Swiss National Science Foundation (SNF) grant numbers 3100A0-118118, 31003ab-133134 and 31003A-149921 to F.H.-T.A.

■ REFERENCES

(1) Choi, Y. D.; Dreyfuss, G. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 7471.

(2) Rech, J. E.; Huang, M. H.; LeStourgeon, W. M.; Flicker, P. F. *J. Struct. Biol.* **1995**, *114*, 84.

(3) McCloskey, A.; Taniguchi, I.; Shinmyozu, K.; Ohno, M. *Science* **2012**, *335*, 1643.

(4) Venables, J. P.; Koh, C. S.; Froehlich, U.; Lapointe, E.; Couture, S.; Inkel, L.; Bramard, A.; Paquet, E. R.; Watier, V.; Durand, M.; Lucier, J.-F.; Gervais-Bird, J.; Tremblay, K.; Prinos, P.; Klinck, R.; Elela, S. A.; Chabot, B. *Mol. Cell. Biol.* **2008**, *28*, 6033.

(5) König, J.; Zarnack, K.; Rot, G.; Curk, T.; Kayikci, M.; Zupan, B.; Turner, D.; Luscombe, N.; Ule, J. *Nat. Struct. Mol. Biol.* **2010**, *17*, 909.

(6) Zarnack, K.; König, J.; Tajnik, M.; Martincorena, I.; Eustermann, S.; Stévant, I.; Reyes, A.; Anders, S.; Luscombe, N.; Ule, J. *Cell* **2013**, *152*, 453.

(7) Sella, O.; Gerlitz, G.; Le, S.-Y.; Elroy-Stein, O. *Mol. Cell. Biol.* **1999**, *19*, 5429.

(8) Schepens, B.; Tinton, S. A.; Bruynooghe, Y.; Parthoens, E.; Haegman, M.; Beyaert, R.; Cornelis, S. *EMBO J.* **2007**, *26*, 158.

(9) Barnett, S. F.; Friedman, D. L.; LeStourgeon, W. M. *Mol. Cell. Biol.* **1989**, *9*, 492.

(10) Merrill, B. M.; Barnett, S. F.; LeStourgeon, W. M.; Williams, K. R. *Nucleic Acids Res.* **1989**, *17*, 8441.

(11) He, F.; Muto, Y.; Inoue, M.; Kigawa, T.; Shirouzu, M.; Terada, T.; Yokoyama, S., 2004; PDB ID 1WF2.

(12) McAfee, J. G.; Shahied-Milam, L.; Soltaninassab, S. R.; LeStourgeon, W. M. *RNA* **1996**, *2*, 1139.

(13) Whitson, S. R.; LeStourgeon, W. M.; Krezel, A. M. *J. Mol. Biol.* **2005**, *350*, 319.

(14) Soltaninassab, S. R.; McAfee, J. G.; Shahied-Milam, L.; LeStourgeon, W. M. *Nucleic Acids Res.* **1998**, *26*, 3410.

(15) Görlach, M.; Burd, C. G.; Dreyfuss, G. *J. Biol. Chem.* **1994**, *269*, 23074.
(16) Wan, L.; Kim, J. K.; Pollard, V. W.; Dreyfuss, G. *J. Biol. Chem.* **2001**, *276*, 7681.
(17) Wenter, P.; Reymond, L.; Auweter, S. D.; Allain, F. H. T.; Pitsch, S. *Nucleic Acids Res.* **2006**, *34*, e79.
(18) Sattler, M.; Schleucher, J.; Griesinger, C. *Prog. NMR Spectrosc.* **1999**, *34*, 93.
(19) Dominguez, C.; Schubert, M.; Duss, O.; Ravindranathan, S.; Allain, F. H. T. *Prog. NMR Spectrosc.* **2011**, *58*, 1.
(20) Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **2002**, *24*, 171.
(21) Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Mol. Biol.* **2002**, *319*, 209.
(22) Güntert, P. In *Protein NMR Techniques*; Downing, A. K., Ed.; Humana Press: Totowa, NJ, 2004; Vol. *278*, p 353.
(23) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.
(24) Grzesiek, S.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 12593.
(25) Farrow, N.; Zhang, O.; Forman-Kay, J.; Kay, L. E. *J. Biomol. NMR* **1994**, *4*, 727.
(26) Demers, J.-P.; Mittermaier, A. K. *J. Am. Chem. Soc.* **2009**, *131*, 4355.
(27) Kay, L. E.; Nicholson, L. K.; Delaglio, F.; Bax, A.; Torchia, D. A. *J. Magn. Reson.* **1992**, *97*, 359.
(28) Saroff, H. A. *Biopolymers* **1995**, *36*, 121.
(29) Wittekind, M.; Görlach, M.; Friedrichs, M.; Dreyfuss, G.; Mueller, L. *Biochemistry* **1992**, *31*, 6254.
(30) Bradshaw, R. A.; Hope, C. J.; Yi, E.; Walker, K. W. In *The Enzymes*; Rose, E. D., David, S. S., Eds.; Academic Press: San Diego, 2002; Vol. *22*, p 387.
(31) Görlach, M.; Wittekind, M.; Beckman, R. A.; Mueller, L.; Dreyfuss, G. *EMBO J.* **1992**, *11*, 3289.
(32) Auweter, S. D.; Oberstrass, F. C.; Allain, F. H. T. *J. Mol. Biol.* **2007**, *367*, 174.
(33) Crooks, G.; Hon, G.; Chandonia, J.-M.; Brenner, S. *Genome Res.* **2004**, *14*, 1188.
(34) Sickmier, E. A.; Frato, K. E.; Shen, H.; Paranawithana, S. R.; Green, M. R.; Kielkopf, C. L. *Mol. Cell* **2006**, *23*, 49.
(35) Senapathy, P.; Shapiro, M. B.; Harris, N. L. *Methods Enzymol.* **1990**, *183*, 252.
(36) Wahl, M. C.; Will, C. L.; Lührmann, R. *Cell* **2009**, *136*, 701.
(37) Banerjee, H.; Rahn, A.; Davis, W.; Singh, R. *RNA* **2003**, *9*, 88.
(38) Mackereth, C. D.; Sattler, M. *Curr. Opin. Struct. Biol.* **2012**, *22*, 287.
(39) Meisenheimer, K. M.; Meisenheimer, P. L.; Koch, T. H. *Methods Enzymol.* **2000**, *318*, 88.
(40) Chodosh, L. A. *Curr. Protoc. Mol. Biol.* **2001**, *36*, 12.5.1.