JACS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

Cite This: J. Am. Chem. Soc. 2018, 140, 11661−11673

pubs.acs.org/JACS

Differential Conformational Dynamics Encoded by the Linker between Quasi RNA Recognition Motifs of Heterogeneous Nuclear Ribonucleoprotein H

Srinivasa R. Penumutchu,[‡,†] Liang-Yuan Chiu,[‡,†] Jennifer L. Meagher,[§] Alexandar L. Hansen,[⊥] Jeanne A. Stuckey,[§] and Blanton S. Tolbert*[,‡]

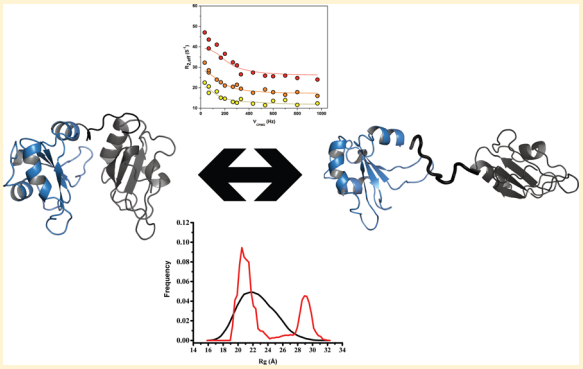
[‡]Department of Chemistry, Case Western Reserve University, Cleveland, Ohio 44106, United States

[§]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109, United States

[⊥]Campus Chemical Instrument Center, The Ohio State University, Columbus, Ohio 43210, United States

Supporting Information

ABSTRACT: Members of the heterogeneous nuclear ribonucleoprotein (hnRNP) F/H family are multipurpose RNA binding proteins that participate in most stages of RNA metabolism. Despite having similar RNA sequence preferences, hnRNP F/H proteins function in overlapping and, in some cases, distinct cellular processes. The domain organization of hnRNP F/H proteins is modular, consisting of N-terminal tandem quasi-RNA recognition motifs (F/HqRRM1,2) and a third C-terminal qRRM3 embedded between glycine-rich repeats. The tandem qRRMs are connected through a 10-residue linker, with several amino acids strictly conserved between hnRNP H and F. A significant difference occurs at position 105 of the linker, where hnRNP H contains a proline and hnRNP F an alanine. To investigate the influence of P105 on the conformational properties of hnRNP H, we probed the structural dynamics of its HqRRM1,2 domain with X-ray crystallography, NMR spectroscopy, and small-angle X-ray scattering. The collective results best describe that HqRRM1,2 exists in a conformational equilibrium between compact and extended structures. The compact structure displays an electropositive surface formed at the qRRM1−qRRM2 interface. Comparison of NMR relaxation parameters, including Carr−Purcell−Meiboom−Gill (CPMG) relaxation dispersion, between HqRRM1,2 and FqRRM1,2 indicates that FqRRM1,2 primarily adopts a more extended and flexible conformation. Introducing the P105A mutation into HqRRM1,2 alters its conformational dynamics to favor an extended structure. Thus, our work demonstrates that the linker compositions confer different structural properties between hnRNP F/H family members that might contribute to their functional diversity.



1. INTRODUCTION

RNA processing requires numerous and faithful interactions between *cis* sequence elements and RNA binding proteins (RBPs). In eukaryotes, members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family represent a large group of RBPs that engage RNA at nearly every stage of a transcript's life cycle. HnRNP proteins are most recognized as modulators of pre-mRNA splicing, yet owing to their abundance, modular domain organization, and often tissue-specific expression patterns, hnRNPs act as general regulators of cellular RNA metabolism under both normal and pathological conditions.[1−3]

The hnRNP F/H proteins constitute a subclass of hnRNPs that consists of five mammalian homologs (F, H, H′, GRSF1, and 2H9) whose biological roles overlap, but individual members often demonstrate context-dependent functional differences.[1,3] The domain organization of hnRNP F/H proteins comprises two or three quasi RNA recognition motifs (qRRMs) and glycine-rich auxiliary domains (Figure 1A). RNA biochemical studies demonstrate that hnRNP F/H proteins specifically recognize G-rich sequences, typically consisting of three or more consecutive guanosines.[4,5] Consensus sequence motifs derived from global transcriptome cross-linking (CLIP-seq) are consistent with the *in vitro* experiments, although slightly different homologue-specific patterns are observed for hnRNP H (G-rich with interspersed A's) and hnRNP F (G-rich with interspersed U's/A's).[6]

Solution NMR structures of the three isolated qRRM domains of hnRNP F (FqRRMs) in complex with G-tract RNAs indicate that each domain specifically binds consecutive guanosines using identical surfaces.[5] The observed modes of recognition are very distinct from those of canonical RRM-RNA complexes, however. The FqRRMs adopt 3D folds reminiscent of RRMs, whereby two α helices buttress a four-

Received: May 22, 2018
Published: August 18, 2018

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

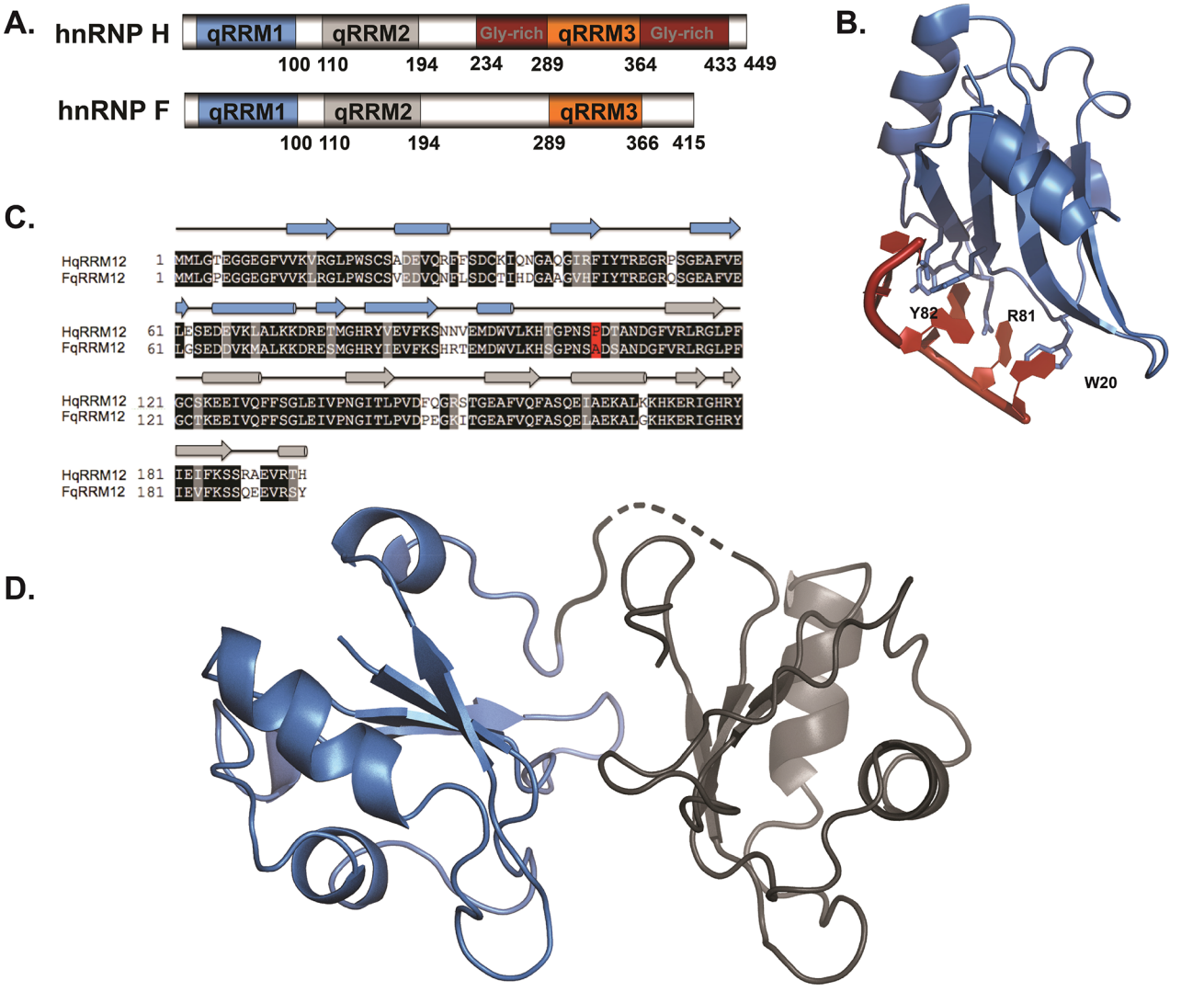


Figure 1. Structural overview of hnRNP F/H proteins. (A) Domain organization of human hnRNP H and hnRNP F, showing the three quasi RNA recognition motifs (qRRM1, blue; qRRM2, gray; and qRRM3, orange) along with the C-terminal glycine-rich domains (red). (B) NMR structure of the FqRRM1-AGGGAU complex (PDB entry 2KFY) solved by Dominguez et al.[5] The structure shows that the RNA-binding surface involves conserved loop residues, depicted as sticks. Mutating W20 to an alanine leads to a >1000-fold reduction in binding affinity for the AGGGAU oligo.[5] (C) Sequence alignment between HqRRM1,2 and FqRRM12. Identical residues are shown in black, similar residues are gray, and red represents position 105, where HqRRM1,2 has a proline and FqRRM1,2 an alanine. (D) Ribbon representation of the compact HqRRM1,2 crystal structure solved here, with missing linker residues depicted as a dashed line.

strand antiparallel β sheet.[7,8] Unlike canonical RRMs that interact with RNA through the antiparallel β sheet, FqRRMs encage consecutive guanosines using conserved residues located in two loops (Figure 1B). Individual FqRRMs bind G-tract RNAs with comparable micromolar affinities, and the binding strength is not greatly enhanced with the tandem N-terminal construct (FqRRM1,2).[5]

The qRRMs of hnRNP F and hnRNP H are highly conserved, which partly accounts for the nearly identical RNA sequence specificities.[9] Indeed, mutations of conserved FqRRM loop residues greatly diminish binding affinity for G-rich RNAs.[5] The sequence conservation extends to the linker connecting qRRM1 to qRRM2 (hnRNP F, H, H′, and GRSF1); however, an intriguing difference occurs at position 105 (hnRNP H numbering), where a proline is located in hnRNP H/H′ and an alanine in hnRNP F (Figure 1C). The proline at linker position 105 results in hnRNP H and H′ having the ubiquitous PXXP motif, which is known to influence the conformational and recognition properties of multi-domain proteins.[10,11] Since the residue composition of linkers affects the conformational dynamics of multi-domain proteins, it is plausible that the linker compositions differentially modulate the structures of hnRNP F and H.

Here, we performed a comprehensive study of the structure and dynamics of the N-terminal tandem qRRMs of hnRNP H (HqRRM1,2) to test the hypothesis that P105 influences the conformational properties of the dual-domain protein. The crystal structure of HqRRM1,2 solved here shows that the dual domain adopts a compact conformation, unlike that of hnRNP F. We further probed the solution behavior of HqRRM1,2 using NMR spectroscopy, wherein we combined residual dipolar couplings (RDCs) and paramagnetic relaxation enhancements (PREs) to demonstrate that the dual domain also populates the compact structure in solution. Ensemble analysis by small-angle X-ray scattering (SAXS) further revealed that HqRRM1,2 undergoes conformational sampling

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

between compact and extended conformers, with the compact structure predominating. Moreover, we show by NMR relaxation dispersion experiments that the HqRRM1,2 linker and interface loop residues undergo slow (milliseconds) motions. Some of the slowly exchanging loop residues coincide with the G-tract RNA binding surface. Despite having ~80% sequence identity to HqRRM1,2, only a small subset of FqRRM1,2 residues exhibit similar millisecond exchange behavior. Collectively, this study provides valuable conformational insights into an important multi-domain RBP, and it opens the possibility that differences in linker compositions modulate hnRNP F/H members.

2. RESULTS AND DISCUSSION

Crystal Structure of HqRRM1,2 Reveals a Compact Conformation for the Dual-Domain Protein. We solved the crystal structure of HqRRM1,2 to 3.5 Å resolution (Table S1). Although the resolution is low, domain placement and domain designation of HqRRM1 or HqRRM2 were unambiguous, as the structure was initially phased from the two seleno-methionine residues present in HqRRM1 (Figure 1C). The structure contains four molecules in the asymmetric unit (Figure S1). The HqRRM1,2 molecules in the asymmetric unit fold in a similar compact conformation, with an RMSD on Cα atoms between 0.58 and 1.58 Å.[12] Each of the qRRM domains displays the classical RRM architecture, containing the canonical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold. The linker region between the two HqRRM domains, residues 100−110, is disordered in the structure, with the exception of the B and C chains. In these chains, we see density for residues T100, G101, and P102 but are unable to see the rest of the linker region (Figure 1D). Due to the low resolution of the HqRRM1,2 crystal structure, we are not able to see side-chain density for some of the residues in the interface of the two domains.

A comparison of the NMR structure of FqRRM1 bound to RNA (PDB entry 2KFY) with the structure of HqRRM1 shows that the protein backbones are very similar, with an RMSD between Cα atoms of 1.82 Å (Figure S2). The HqRRM2 domain in HqRRM1,2 is partially blocking the binding site of the RNA, as seen in the FqRRM1 (Figure S2).[5]

Solution Properties of HqRRM1, HqRRM2, and HqRRM1,2. To determine if the compact HqRRM1,2 conformation observed in the crystal exists in solution, we performed NMR studies of the individual HqRRM1, HqRRM2, and HqRRM1,2 domains. The $^{1}H-^{15}N$ HSQC spectra show dispersed signals for all three constructs, indicating that the proteins adopt stable folds in solution (Figure S3). Overlay of the spectra of HqRRM1 and HqRRM2 with the spectrum of HqRRM1,2 shows that most of the resonances superimpose; however, several peaks experience detectable chemical shift perturbations (CSPs), likely indicating that the individual HqRRMs transiently associate within the context of the dual-domain protein (Figure S3).

To further assess the solution behavior, we determined the overall dynamics by measuring ^{15}N relaxation data. T_1, T_2, and $-^{15}N$ hetNOE relaxation parameters were acquired for HqRRM1, HqRRM2, and HqRRM1,2 (Figure 2). For the resonances that could be uniquely evaluated, the relaxation parameters confirm that the core regions of HqRRM1 (residues 10−100) and HqRRM2 (residues 111−194) are stable, with $-^{15}N$ hetNOE values between 0.7 and 0.9. The N- and C-termini of both qRRMs display increased flexibility, however (Figure 2). In particular, the C-terminus of HqRRM1

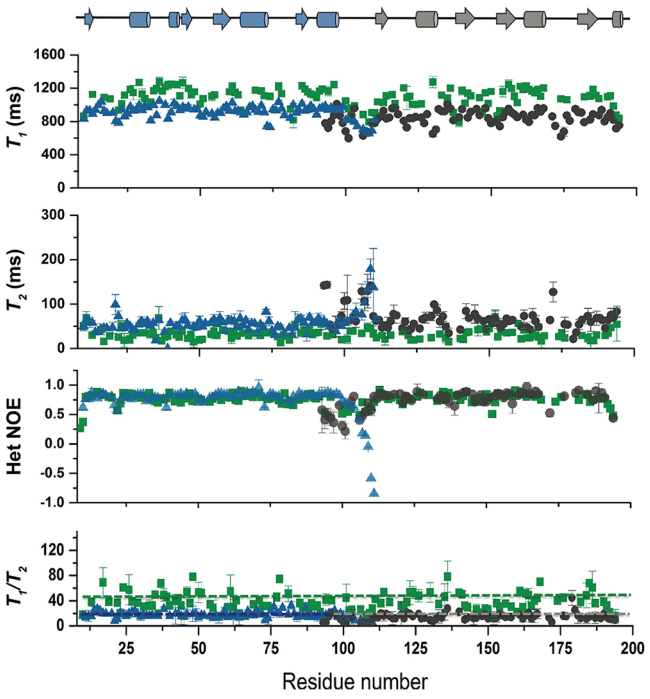


Figure 2. ^{15}N relaxation studies indicate that HqRRM1,2 tumbles as a single unit in solution. Measurement of T_1, T_2, and $\{^{1}H\}-^{15}N$ heteronuclear NOEs for HqRRM1,2 (green), isolated HqRRM1 (blue), and isolated HqRRM2 (gray). Of note, the hetNOE data indicate that the inter-qRRM linker of HqRRM1,2 is relatively rigid on the ps-ns time scale. Estimates of the rotational correlation times for HqRRM1,2 (14.5 ns), HqRRM1 (9.6 ns), and HqRRM2 (8.3 ns) were obtained by T_1/T_2 and further reveal that HqRRM1,2 tumbles as one larger unit in solution. The secondary structure elements of HqRRM1,2 are displayed above.

is very mobile, with negative $-^{15}N$ hetNOEs. A similar relaxation profile across the core regions of HqRRM1,2 was also observed (Figure 2). Interestingly, however, the $\{^{1}H\}-^{15}N$ hetNOE values for the inter-qRRM linker (IQL, residues 100−111) ranged from 0.5 to 0.72, with an average of 0.60 ± 0.08, indicating that at least part of the linker backbone is rigid. By comparison, reported $\{^{1}H\}-^{15}N$ hetNOEs for the IQL of FqRRM1,2 ranged from 0.19 to 0.5.[8]

To determine if the individual HqRRM domains tumble independently or collectively within the context of the dual-domain protein, we obtained estimates of the rotational correlation times (τ_c) for HqRRM1, HqRRM2, and HqRRM1,2. The average τ_c values for HqRRM1 and HqRRM2 are 9.6 and 8.3 ns, respectively. By contrast, the average rotational correlation time of HqRRM1,2 is 14.5 ns, and estimates of τ_c for HqRRM1 and HqRRM2 within the context of the dual-domain protein are 14.0 and 14.7 ns, respectively. The significantly larger and comparable rotational correlation times of HqRRM1 and HqRRM2 indicate that they tumble as part of a larger unit within the context of the dual domain.

As a proxy of the temperature dependence of inter-domain motions,[13] we measured global (^{15}N)-T_1 relaxation parameters for HqRRM1,2, FqRRM1,2, and a P105A mutant of HqRRM1,2 ($HqRRM1,2^{P105A}$). HqRRM1,2 shares >80% sequence similarity with FqRRM1,2; however, a significant difference occurs at position 105, where a proline is located in HqRRM1,2 and an alanine in FqRRM1,2. The proline at

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

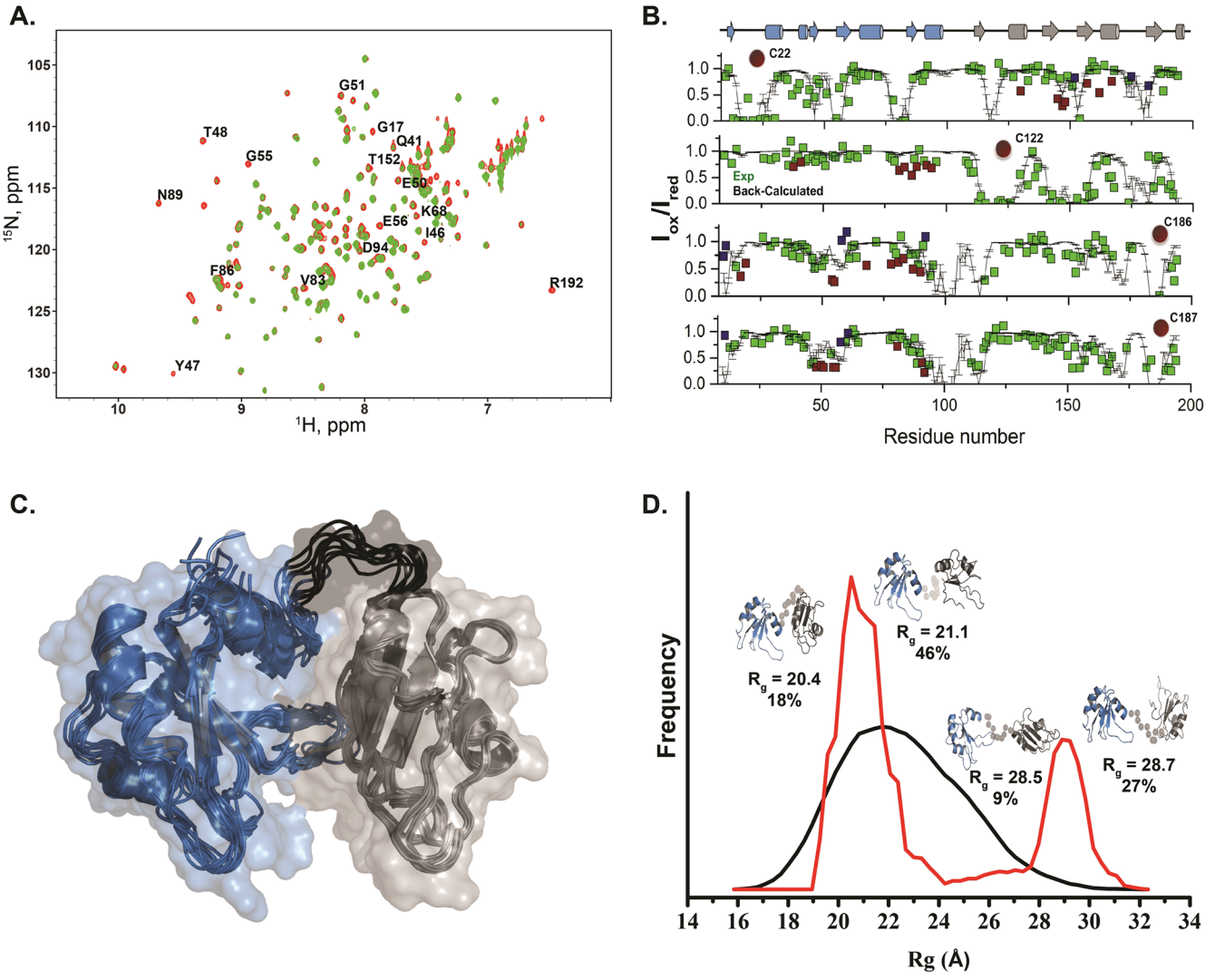


Figure 3. Characterization of the HqRRM1,2 structure in solution. (A) Overlay of ^{1}H−^{15}N HSQC spectra of MTSL-labeled (S187C) HqRRM1,2 at 800 MHz. Red and green correlation peaks represent the diamagnetic and paramagnetic forms of MTSL-labeled HqRRM1,2, respectively. Residues that experience significant peak broadening are labeled. (B) Paramagnetic enhancement to nuclear spin relaxation for HqRRM1,2 with MTSL labeled at positions 22, 122, 186, and 187. Experimentally measured (colored squares) amide proton PRE effect plotted as I_{para}/I_{dia} for HqRRM1,2 for each rigid (HetNOE > 0.6) and well-resolved residue. Back-calculated (solid black line) values with standard deviations were determined from the structure ensemble shown in panel C. Red and blue squares correspond to minor conformations with alternative inter-qRRM orientations. (C) Cartoon representation of the HqRRM1,2 structural model solved using PRE and RDC restraints. The PRE-derived structural model indicates that HqRRM1,2 primarily adopts a compact structure in solution. (D) The radius of gyration distribution (R_g) of HqRRM1,2 conformers calculated using the ensemble optimization method (EOM). The solid black line corresponds to the initial pool of 10 000 unbiased HqRRM1,2 conformers. The solid red line represents the bimodal distribution of EOM-selected HqRRM1,2 conformers and indicates that the protein fluctuates between an ensemble of compact and extended inter-qRRM orientations. Representative HqRRM1,2 models from the EOM calculation are also shown with the corresponding R_g values and fraction of occurrence.

position 105 of HqRRM1,2 is part of a PXXP motif, which is known to influence the conformational properties of linkers.[11] Therefore, it is plausible that P105 differentially modulates the overall dynamics of the IQL of HqRRM1,2. Figure S4 shows that the global (^{15}N)-T_1 values of FqRRM1,2 decrease linearly with increasing temperature. Conversely, the temperature dependence of the global (^{15}N)-T_{1s} of HqRRM1,2 shows a sharp transition between 305 and 308 K. We interpret the differential temperature dependence of the (^{15}N)-T_{1s} as manifestations of distinct conformational properties, resulting from the intrinsic linker compositions of HqRRM1,2 and FqRRM1,2. Indeed, the (^{15}N)-T_1 versus temperature profile of HqRRM1,2^{P105A} is more similar to that of FqRRM1,2 (Figure S4). To further assess the solution properties of HqRRM1,2^{P105A}, we collected T_1 and T_2 relaxation parameters to obtain an estimate of the rotational correlation time for this construct (Figure S5). The average τ_c for HqRRM1,2^{P105A} is 8.5 ns, which is close to the τ_c values measured for the isolated HqRRM domains and the value reported previously for FqRRM1,2 (Table S3). When the results are taken together, the ^{15}N relaxation study indicates that the individual qRRMs of HqRRM1,2 stably associate in solution and that P105 distinguishes the conformational properties of HqRRM1,2 from those of FqRRM1,2.

HqRRM1,2 Adopts a Compact yet Dynamic Conformation in Solution. Since our initial NMR study indicates that HqRRM1,2 populates a compact conformation, we proceeded to determine its solution structure. The

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

coordinates of HqRRM1 are already deposited in the Protein Database (PDB entry 2LXU);[14] therefore, we solved the structure of HqRRM2 and calculated structural models of HqRRM1,2.

The ensemble of the 10 lowest energy structures of HqRRM2 is shown in Figure S6, and structural statistics are provided in Table S2. As expected, HqRRM2 adopts the canonical RRM fold consisting of the $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology. Comparison of the isolated HqRRM1/HqRRM2 structure with that of FqRRM1/FqRRM2 shows that the isolated domains are very similar, with backbone Cα RMSDs of 0.66 and 1.59 Å, respectively. Additionally, the solution NMR structures of HqRRM1/HqRRM2 domains agree favorably with the structures of the sub-domains identified in the crystal (HqRRM1 Cα = 0.52 Å and HqRRM2 Cα = 1.47 Å).

To assess the solution conformation of HqRRM1,2, we acquired RDCs and prepared a series of mutants for paramagnetic resonance enhancement (PRE) measurements (see Materials and Methods). RDCs were measured with a single alignment medium consisting of a hexanol/PEG mixture, since attempts to use *pf1* bacteriophage led to a severe deterioration of spectral quality. Using a HqRRM1,2 construct where two of the three native cysteines were differentially changed to serines, PREs were obtained through the conjugation of an MTSL spin label at native positions C22 (C122S) and C122 (C22S) and mutated positions S186C (C22S/C122S) and S187C (C22S/C122S). Analysis of the $^{1}H-^{15}N$ HSQC spectra for each construct confirmed that the mutations do not grossly affect the folding of HqRRM1,2 (Figure S7). Incorporation of the spin labels led to a distance-dependent line broadening of the NMR signals, whereby long-range (~20 Å) distances can be reliably determined.[15,16] Therefore, the combination of RDCs and PREs allows the spatial positioning of different protein domains, even in the absence of inter-domain NOEs.[15−17] Figure 3 shows PRE profiles derived from four HqRRM1,2 constructs. As expected, local intra-domain PREs are observed within the vicinity of the MTSL spin label. Of significance, attachment of the spin label at positions C22, C186, and C187 produced detectable long-range and inter-qRRM PREs, indicating that the two domains are proximal in solution (Figure 3B).

We proceeded to calculate a structural model of HqRRM1,2, given the evidence that the dual domain populates a compact conformation in solution. A complete description of the structure calculation routine is provided in the Materials and Methods. In brief, a fully extended random-coiled conformer of HqRRM1,2 was subjected to simulated annealing in Aria wherein NOE, hydrogen bonding, and Φ/Ψ dihedral angle restraints were applied consistent with the structures of isolated HqRRM1 and HqRRM2; no inter-domain NOE restraints were measured. Based on backbone chemical shifts, Φ/Ψ dihedral angle restraints were also used to restrain the IQL (residues 100−110). Ten structures with low overall penalty functions were selected for conjoined refinement in XPLOR/CNS, where RDCs and PREs were included to define the relative orientation of the dual-domain protein. A report of the total restraints and structural statistics is provided in Table S2.

Figure 3C shows the 10 structural models of HqRRM1,2 that converged with a backbone RMSD of 1.57 Å. The back-calculated RDCs and PREs of the 10 lowest energy models agree well with the experimental data, with a global RDC RMS value of 0.11 and PRE *Q*-factor of 0.36 ± 0.03. Inclusion of the PRE and RDC restraints into the structure calculation routine did not distort the local folding of the HqRRMs, as judged by the favorable agreement with the NMR structures of the isolated domains (backbone RMSDs of 1.35 and 1.65, respectively). Similar to the structure of the isolated HqRRM1, residues 90−98 within the dual-domain fold into an α helix that packs against the β sheet surface of HqRRM1. The backbone reverses its direction at the first position (residue 100) of the IQL. Although the IQL does not adopt detectable secondary structure, its position is relatively rigid in each of the models, consistent with the $\{^{1}H\}$-^{15}N hetNOE values measured for this region. The sharp reversal of the backbone at the start of the linker brings the qRRMs within proximity such that HqRRM1,2 adopts a compact conformation in solution wherein the β sheet surface of each qRRM faces inward (Figure 3C). Interactions that stabilize the interface of the compact structure are not determined due to missing short-range distance restraints.

Comparison of the solution and crystal structures of HqRRM1,2 reveals that the relative orientations of the qRRMs are different (Figure S8). This difference likely reflects conformational dynamics whereby the qRRMs sample multiple inter-domain orientations. Evidence for potential conformational dynamics is observed when comparing experimental PREs to values back-calculated from the NMR structures (Figure 3B). Several of the experimental PREs differ from those back-calculated by more than ±0.2, indicating that HqRRM1,2 adopts other minor conformations in solution, including conformers with extended inter-domain geometries. To gain additional insights into the solution properties of HqRRM1,2, inline size exclusion chromatography with small-angle X-ray scattering (SEC-SAXS) data were acquired. Guinier analysis of the SAXS data confirms that HqRRM1,2 is monodispersed, with a radius of gyration (R_g) of 23.04 ± 0.366 Å, and the dimensionless Kratky plot has the characteristic inverted shape of a well-folded protein (Figure S9). Nevertheless, attempts to fit either the NMR or crystal structure of HqRRM1,2 to the experimental scattering intensities resulted in poor agreement (Figure S9 and Table S3). Moreover, the pairwise distribution function $P(r)$ of HqRRM1,2 has a shoulder at ~42 Å, and the function gradually tails off, with a maximum dimension (D_{max}) of 84 Å, larger than expected from the NMR or crystal structure (Figure S9). The bimodal shape and overall dimensions of the $P(r)$ function are compatible with HqRRM1,2 existing as a conformational ensemble between compact and extended conformers. We also acquired SEC-SAXS data on HqRRM1,2^{P105A} and FqRRM1,2 (Figure S10 and Table S3). The pairwise distribution functions $P(r)$ show that both proteins adopt more extended conformations in solution, as determined by their larger R_g (24.2 ± 0.4 Å for HqRRM1,2^{P105A} and 26.39 ± 0.22 Å for FqRRM1,2) and D_{max} values (98 Å for HqRRM1,2^{P105A} and 105 Å for FqRRM1,2). The more extended FqRRM1,2 structure is consistent with previous NMR relaxation studies that determined the qRRMs are non-interacting,[8] whereas the results for HqRRM1,2^{P105A} support the hypothesis that the P105A mutation increases the flexibility of the IQL.

In an attempt to account for potential HqRRM1,2 conformational dynamics, we proceeded to analyze the SEC-SAXS data using the Ensemble Optimization Method (EOM).[18,19] Conformational fluctuations that occur during the time scale (milliseconds to minutes) of a SAXS

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

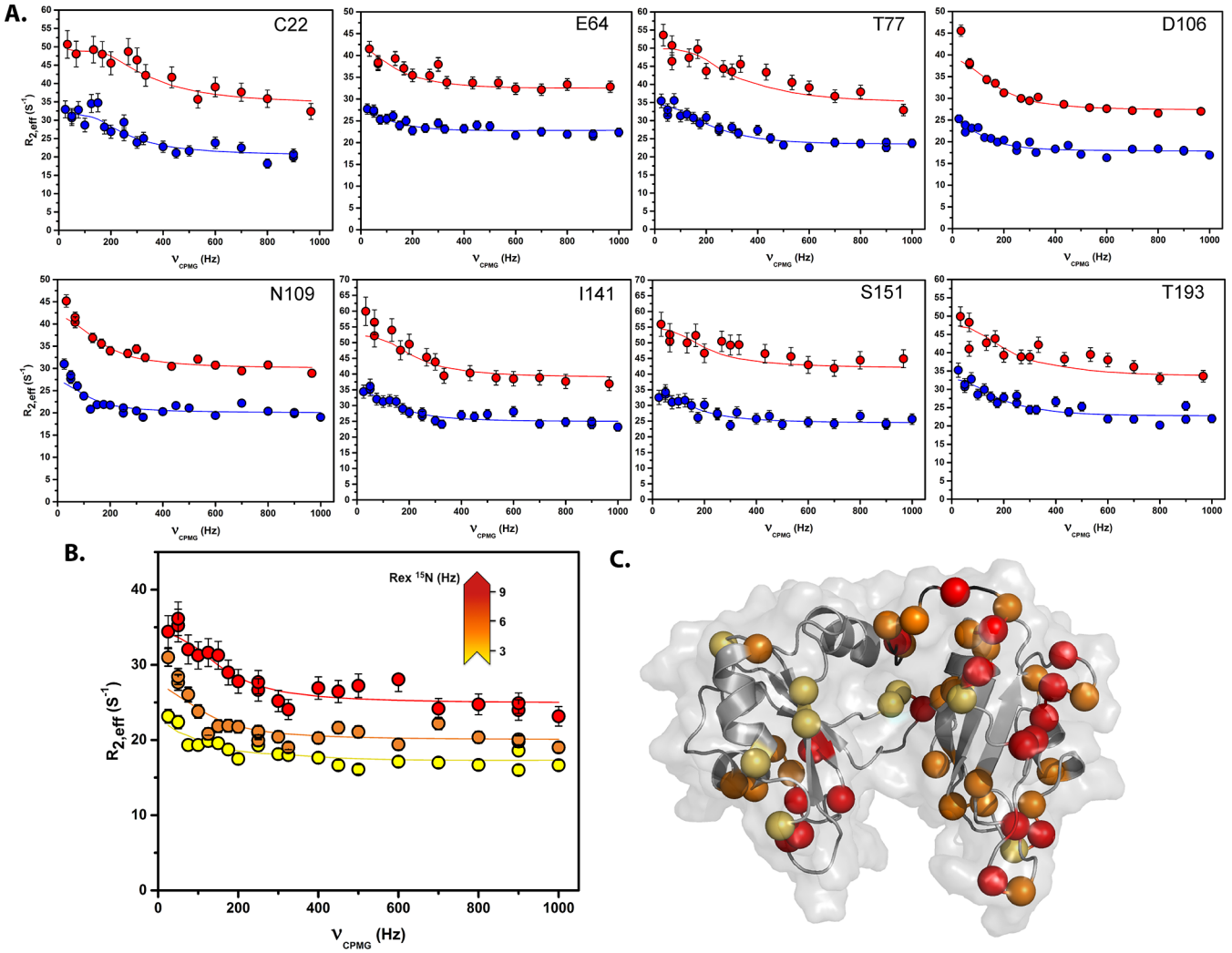


Figure 4. HqRRM1,2 undergoes microsecond-to-millisecond conformational dynamics. (A) Representative backbone ^{15}N-CPMG relaxation dispersion profiles for select residues located within HqRRM1 (C22, E64, and T77), the inter-qRRM linker (D106, N109), and HqRRM2 (I141, S151, and T193). The experimental CPMG R_{ex} profiles were recorded at 14.1 T (blue) and 20.0 T (red). Solid blue and red lines correspond to fits of the CPMG R_{ex} data to a two-site exchange model (see Materials and Methods). (B) Magnitude of R_{ex} values recorded at 14.1 T shown as color-coded spheres on the structures of HqRRM1,2: yellow color represents the R_{ex} values (A108) between 3 and 6 s^{-1}, orange represents the R_{ex} values (N109) between 6 and 9 s^{-1}, and red represents the R_{ex} values (I141) larger than 9 s^{-1}. (C) Mapping the R_{ex} values onto the compact HqRRM1,2 structure reveals that residues that undergo μs-ms exchange primarily localize to the interface of the dual qRRMs and the IQL.

measurement are encoded in the experimental scattering intensities, and as such the EOM approach attempts to deconvolute the population of conformers that contribute to the scattering signal.[18,19] An initial pool of 10 000 unbiased HqRRM1,2 models was built by attaching *ab initio* linkers to the qRRM domains with randomized geometries. Subsequent ensemble optimization resulted in pools of 50 conformers that fit the experimental SAXS data with significantly improved χ^2 values (1.27) compared to either the NMR or crystal structure. Figure 3D shows the comparison of the distribution of R_g values for the selected conformers against the initial pool of 10 000. The R_g distribution of the initial pool has one peak centered at ~23 Å, which agrees with the experimental value derived from Guinier fits of the scattering data (R_g = 23.04 ± 0.366 Å). Conversely, the EOM-selected conformers have a bimodal distribution, with major and minor peaks centered at approximately 21 and 29 Å, respectively. The EOM-selected distribution is consistent with a dynamic ensemble of compact and extended conformers. The positions of four structural ensembles that are most representative of the EOM-selected bimodal R_g distribution are also shown in Figure 3D. The most frequently occurring conformers (64%) have low overall R_g values and are illustrative of the compact HqRRM1,2 conformation observed by X-ray crystallography and solution NMR. The remaining EOM-selected distribution (36%) comprises models with more extended structures and significantly larger R_g values (Figure 3D). Thus, the SAXS data indicate that HqRRM1,2 exists as a dynamic equilibrium between compact and extended conformations, albeit with different relative inter-qRRM geometries. Of note, EOM analysis of SEC-SAXS data from HqRRM1,2^{P105A} and FqRRM1,2 showed bimodal R_g distributions, although both peaks were shifted to higher dimensions (not shown).

HqRRM1,2 Undergoes Slow Conformational Dynamics. The collective NMR and SAXS data indicate that HqRRM1,2 exists as a dynamic ensemble of compact and extended conformers. Such large-scale conformational rearrangements of the HqRRMs likely occur slowly on the μs-ms

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

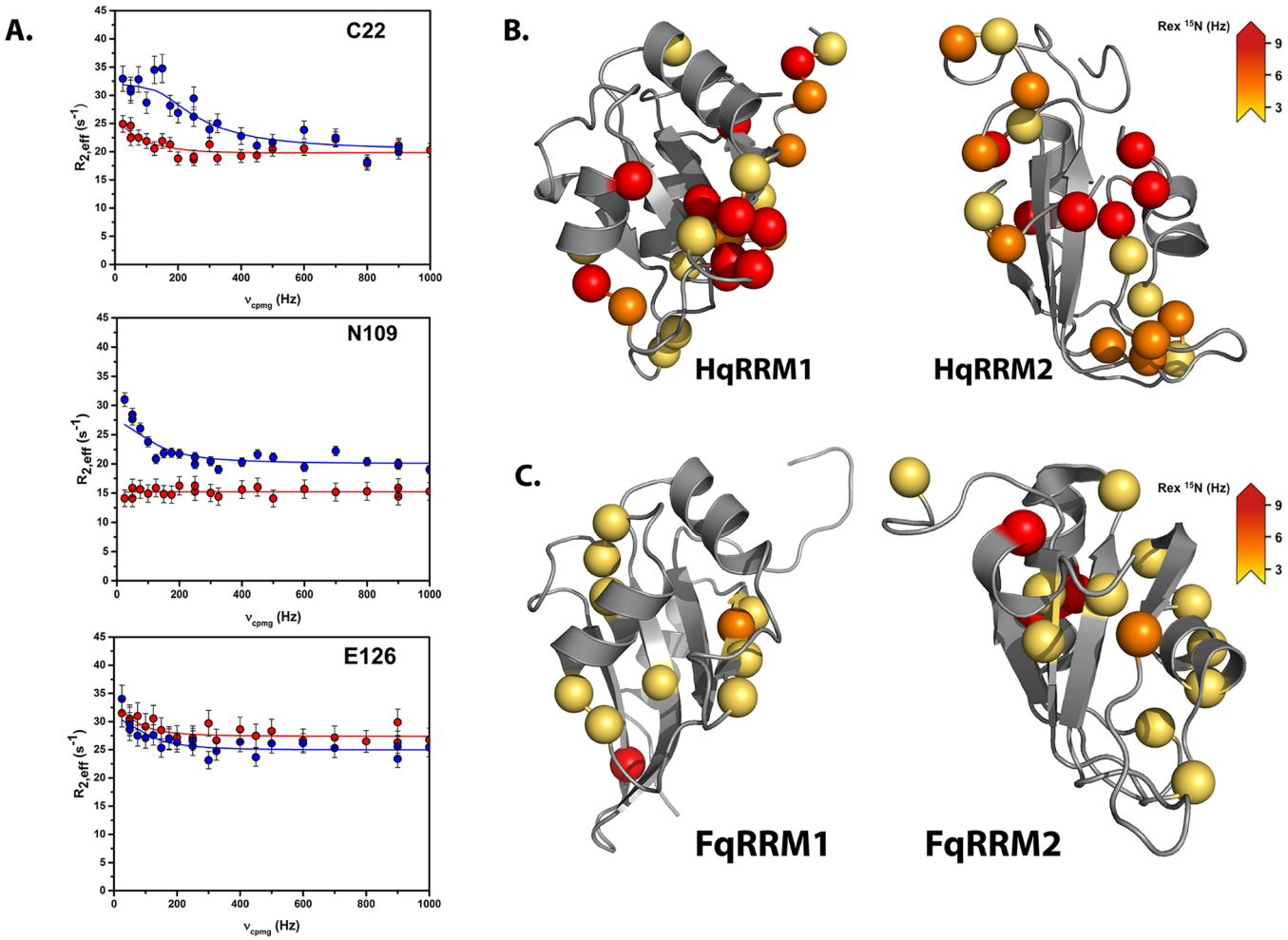


Figure 5. Microsecond-to-millisecond dynamics intrinsic to HqRRM1,2 are not conserved in FqRRM1,2. (A) Representative backbone ^{15}N-CPMG relaxation dispersion profiles shown for select residues of HqRRM1,2 (blue) and FqRRM1,2 (red). Solid lines correspond to fits of the CPMG R_{ex} data to a two-site exchange model. Magnitudes of R_{ex} values recorded at 14.1 T are shown as color-coded spheres on the structures of (B) HqRRM1,2 and (C) FqRRM1,2. The inter-qRRM linkers have been removed since there is currently no structure of the dual-domain FqRRM1,2 protein. ^{15}N-CPMG R_{ex} data were collected on the intact dual-domain proteins, however.

time scale. To probe for slow HqRRM1,2 motions, we performed backbone ^{15}N Carr−Purcell−Meiboom−Gill (CPMG) relaxation dispersion experiments. ^{15}N-CPMG relaxation dispersion provides site-specific information on the contribution of dynamic processes to the effective transverse relaxation rate constant ($R_{2,\mathrm{eff}} = R_2^\circ + R_{ex}$). Fitting of the R_{ex} dispersion curves to a two-site exchange model gives insight on the kinetics (k_{ex}) and thermodynamics (populations of major and minor states, p_a and p_b) of the interconverting species.[20−24] Pilot ^{15}N-CPMG studies showed that the R_{ex} is temperature dependent and more pronounced at lower temperatures; therefore, dispersion profiles were measured at 288 K and at two NMR fields (600 and 850 MHz). Analysis of the ^{15}N relaxation dispersion data shows that many residues across HqRRM1,2 experience μs-ms motions (Figure 4). Residues with the largest R_{ex} localize to the interface of the compact structure and the IQL (Figure 4). Several residues located in the IQL that are relatively rigid on the ps-ns time scale experience slow conformational dynamics; these include N103, S104, D106, A108, N109, and D110. Interestingly, some residues conserved with hnRNP F that form its RNA binding surface experience slow dynamics. These residues include R16 and Y82, which are located in β1 and loop 5. Global fitting of the relaxation dispersion data to a simple two-site exchange model (χ^2 = 1.28) reveals that HqRRM1,2 fluctuates between major and minor conformers (p_b = 3.04 ± 0.11%), with an exchange rate constant k_{ex} = 512 ± 49 s^{-1} (Table S4). Interestingly, we also observed R_{ex} behavior (derived from Modelfree analysis of R_1, R_2, and NOE collected at 800 MHz) for some residues within the isolated HqRRM1 and HqRRM2 domains; however, the magnitude and degree of the μs-ms exchange were significantly quenched by comparison to HqRRM1,2 (Figure S11). We therefore reason that the IQL is the dominant contributor to the overall conformational dynamics of HqRRM1,2, but the individual qRRM domains retain intrinsic μs-ms motions that primarily localize to the loops (Figure S11).

To explore if slow conformational dynamics are conserved in FqRRM1,2, we performed ^{15}N-CPMG experiments on a construct that was used in previous NMR studies.[8] Analysis of the ^{15}N-CPMG data acquired on FqRRM1,2 shows that some residues undergo μs-ms exchange, albeit to a far lesser degree and extent than those observed for HqRRM1,2 (Figure 5). Notably, residues that are conserved between the two proteins and which comprise their respective RNA binding surfaces show differential μs-ms conformational dynamics

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

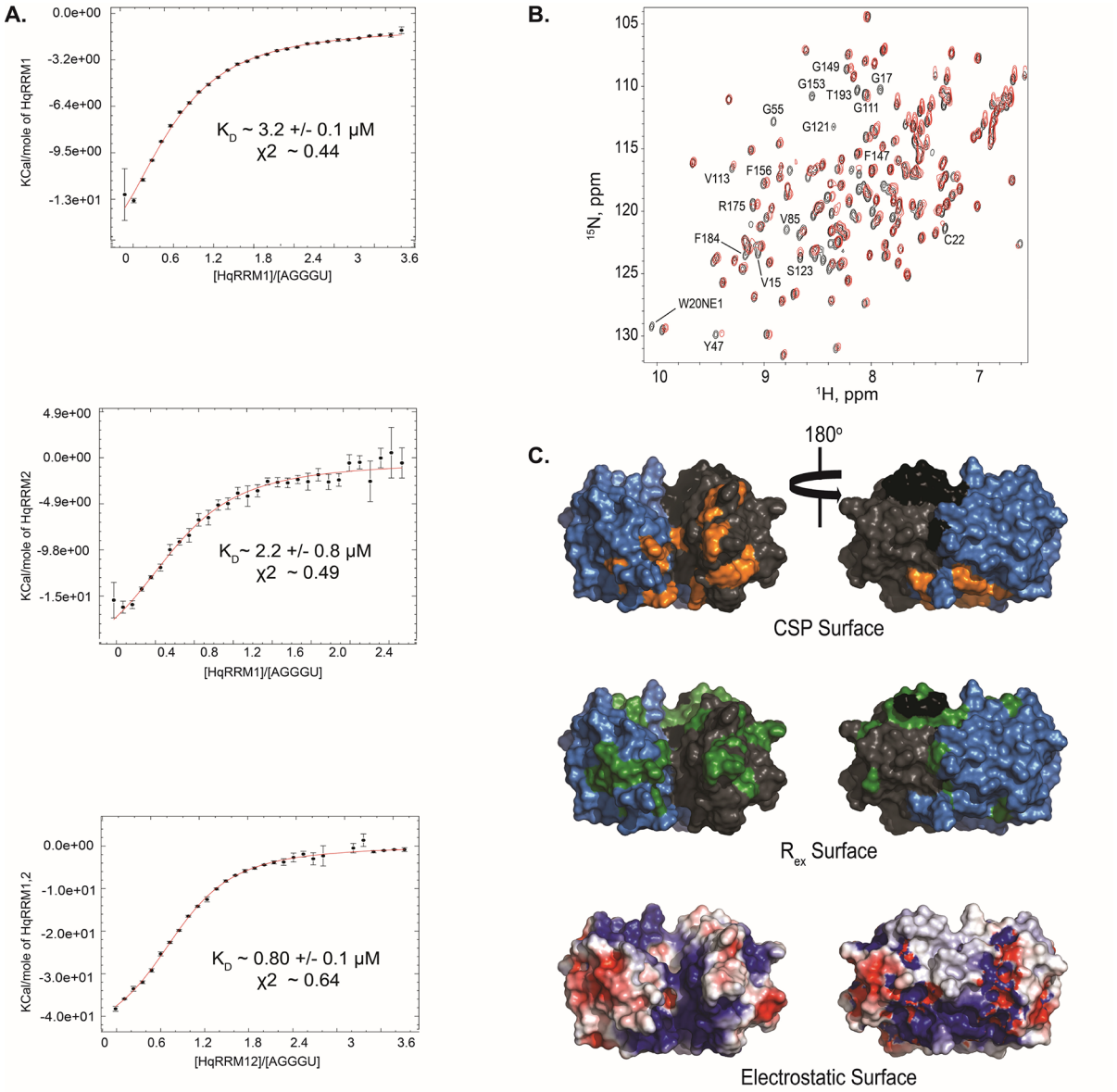


Figure 6. HqRRM1,2 uses a unique surface to recognize a single G-tract with 1:1 stoichiometry. (A) Representative calorimetric titration profiles of HqRRM1 (top), HqRRM2 (middle), and HqRRM1,2 (bottom) titrated into a model AGGGU oligomer. The titrations were performed at 298 K and in 20 mM sodium phosphate (pH 6.2), 20 mM NaCl, 4 mM TCEP. All titration data were processed and analyzed using Affinimeter. The processed thermograms were fit to a 1:1 stoichiometric binding model. Values of the binding dissociation constants (K_D) and corresponding standard deviations are from triplicate experiments. Goodness of fits (χ^2) of the experimental data to the 1:1 binding model are reported for each titration. (B) Overlay of $^1H-^{15}N$ HSQC spectra of free HqRRM1,2 (red) and AGGGU-bound HqRRM1,2 (black) at a 4:1 molar ratio. Residues that completely disappear in the presence of saturating amounts of AGGGU are labeled. (C) Surface representations of HqRRM1,2 color-coded by residues that disappear in the presence of saturating (4:1 molar ratio) amounts of AGGGU (top), by residues that experience significant ^{15}N relaxation dispersion (middle), and by the overall electrostatic potential surface (bottom).

(Figure 5). In general, the magnitude of R_{ex} observed for HqRRM1,2 is ~2-fold higher compared to those of identical residues with non-zero R_{ex} values in FqRRM1,2 (Figure 5). Thus, the data presented here indicate that HqRRM1,2 undergoes millisecond conformational fluctuations not conserved by FqRRM1,2, despite their overall high sequence similarity. Combining these findings with the results of HqRRM1,2^{P105A}, we conclude that the differential conformational dynamics of native HqRRM1,2 are encoded in the composition of its IQL.

Implications of HqRRM1,2 Structure on G-Tract Recognition. HnRNP H specifically recognizes G-rich RNA sequences composed of at least three consecutive guanosines, colloquially referred to as G-tracts.[5] To test if the observed conformational dynamics influences HqRRM1,2-RNA recognition, we carried out calorimetric and NMR titrations with the isolated HqRRMs, the dual domain, and a model 5′-AGGGU-3′ oligo. Figure 6A shows representative calorimetric thermograms of the HqRRM1,2 constructs titrated into 5′-AGGGU-3′. Global fits of the processed data to a 1:1 isotherm using the KinITC routines available in Affinimeter[25,26] show that HqRRM1 and HqRRM2 each bind 5′-AGGGU-3′ with comparable micromolar affinities (K_D = 3.2 ± 0.1 and 2.2 ± 0.8 μM, respectively). Interestingly, HqRRM1,2 also binds 5′-AGGGU-3′ with a 1:1 stoichiometry; however, the binding affinity is ~4-fold stronger (K_D = 0.80 ± 0.1 μM). We also

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

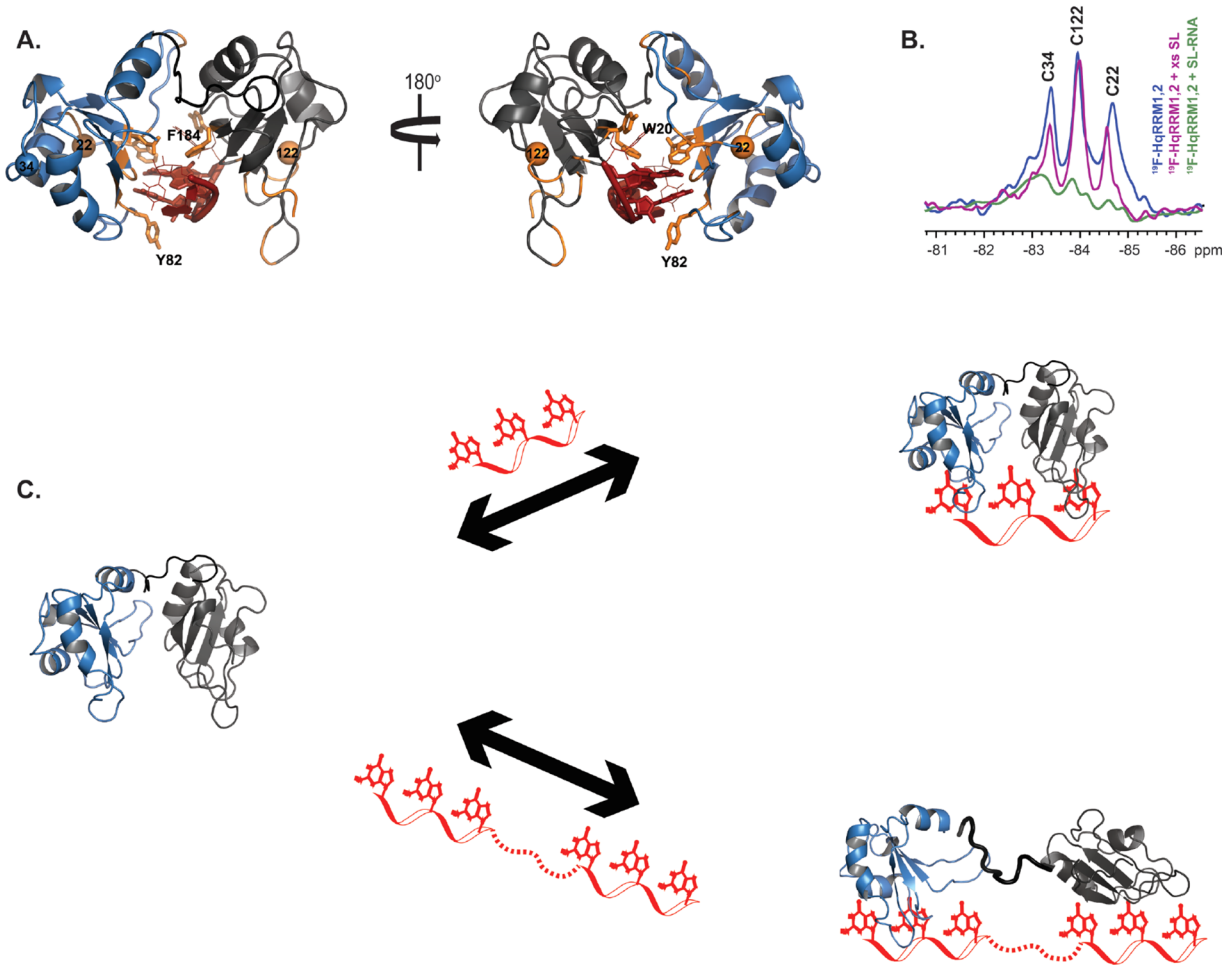


Figure 7. HqRRM1,2 uses a dynamic mechanism to select RNA targets with different numbers of G-tracts. (A) Superimposition of FqRRM1,2-AGGGAU complex (Figure 1C) onto HqRRM1,2 shows that the compact structure can easily accommodate a single G-tract element. The FqRRM1 structure has been removed for clarity; however, the orientation of the bound RNA (red) is as observed in the complex. The central GGG element is shown as filled rings, and the remaining nucleobases are shown as sticks. Select amino acids whose $^{1}H-^{15}N$ correlation peaks disappear at saturating amounts of unlabeled AGGGU are shown as orange sticks. (B) ^{19}F-detected PREs on a HqRRM1,2 construct labeled with BTFA at cytosine positions 22, 34, and 122 (shown as Cα spheres in panel A) provide evidence that the compact structure accommodates a single G-tract RNA. The ^{19}F 1D spectra correspond to (blue) free BTFA-labeled HqRRM1,2, (purple) free BTFA-labeled HqRRM1,2 with excess IAM-PROXYAL nitroxide spin label, and (green) a 1:1 complex of BTFA-labeled HqRRM1,2 with AGGGA*U modified at a specific internal phosphorothioate (*) position with IAM-PROXYAL. The significantly reduced intensity observed in the ^{19}F spectrum of the HqRRM1,2-AGGGA*U complex indicates that both qRRM domains are within close proximity to the bound RNA. (C) HqRRM12 undergoes conformational exchange between closed and extended conformers, with the compact structure as the major conformation to recognize a single isolated G-tract element. The extended structure can recognize multiple G-tracts connected via linker nucleotides.

probed the binding interface by performing HSQC titrations with ^{15}N-labeled HqRRM1,2 and unlabeled 5′-AGGGU-3′. Figure 6B shows that correlation peaks from residues located in both HqRRMs broaden beyond detection at saturating amounts of 5′-AGGGU-3′ (4:1 molar ratio). Mapping the perturbations to the compact structure of HqRRM1,2 reveals that the binding interface is extensive and localizes to one surface of the protein (Figure 6C). Coincidently, the binding surface overlaps with residues that experience μs-ms conformational dynamics and with a unique electropositive cavity formed at the interface of the HqRRMs (Figure 6C).

To explore if the compact structure can accommodate a single G-tract, we superimposed the FqRRM1-AGGGUA co-structure[8] (Figure 1B) onto HqRRM1,2 (Figure 7A). The superimposition shows that HqRRM1,2 easily accommodates a single G-tract, provided that the qRRMs slightly adjust their relative orientations to relieve steric clashes. Such inter-domain movements are consistent with the μs-ms conformational dynamics detected by CPMG relaxation dispersion (Figure 4). Interestingly, several of the correlation peaks that disappear in the HSQC titration correspond to residues that are within proximity of the RNA (Figure 7A).

To test the feasibility of the HqRRM1,2-AGGGUA docked model (Figure 7A), we prepared a construct where each native cysteine (22, 34, and 122) was chemically modified with the ^{19}F NMR-active BTFA probe and used this construct to detect PRE enhancements from a bound AGGGA*U oligo that was internally labeled at a specific phosphorothioate (*) position with the IAM-PROXYAL spin label (see Materials and Methods). We decided to use this approach, since many of the ^{15}N signals of HqRRM1,2 are broadened beyond detection within the complex (Figure 6B), thus precluding detection of intermolecular NOEs.

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

Analysis of the 1D ^{19}F NMR spectrum of BTFA-labeled HqRRM1,2 reveals three well-resolved peaks (Figure 7B), which were assigned using singly labeled HqRRM1,2 constructs. Addition of excess IAM-PROXYAL to BTFA-labeled HqRRM1,2 resulted in very minor perturbations to the 1D ^{19}F NMR spectrum; however, the signals were significantly attenuated in the presence of an equimolar amount of AGGGA*U modified with IAM-PROXYAL (Figure 7B). Since the ^{19}F signals from both qRRMs were equally attenuated, we therefore conclude that the two qRRMs of the compact HqRRM1,2 structure share the responsibility for binding a single G-tract.

3. CONCLUSION

Members of the hnRNP F/H family are important RNA binding proteins that function in overlapping and, in some cases, non-redundant biological processes.[1,3] Global CLIP-seq reveals that hnRNP F and H share a preference for poly-G stretches, although subtle differences in their consensus motifs are observed, with hnRNP F showing enrichment for UA flanking sequences and hnRNP H an interspersion of adenosines.[6] In a separate high-throughput study, hnRNP H was found to preferentially interact with UGGG tetrameric sequences located within introns.[27] These apparent differences in RNA preferences between two highly homologous proteins reflect complexities of protein−RNA interactions within the cellular environment; however, it is also conceivable that minor evolutionary alterations in their respective amino acid sequences modulate specificity.

Here, we integrated X-ray crystallography, NMR spectroscopy, and SAXS to provide a comprehensive description of the structural dynamics of the N-terminal tandem RNA binding domain of hnRNP H (HqRRM1,2). The significant observation is that HqRRM1,2 primarily adopts a compact structure, as determined by X-ray crystallography and NMR spectroscopy; however, the protein undergoes millisecond dynamics, likely to a more extended conformation. The magnitude and degree of the μs-ms motions intrinsic to HqRRM1,2 are not conserved in FqRRM1,2. Therefore, we reason that the differential inter-qRRM dynamics provide a mechanism by which hnRNP F/H members interact with distinct classes of RNA transcripts. The compact conformation of HqRRM1,2 achieves RNA recognition through mutual engagement of both qRRMs with a single G-tract, whereas presumably the extended conformation can bind two independent G-tracts similar to hnRNP F (Figure 7C). Supportive of this premise of plasticity in RNA recognition, the Drosophila homolog of hnRNP F (Glorund) was shown to bind structured UA regions using a surface distinct from its G-tract recognition site.[28]

Our results indicate that the IQL is a critical determinant in defining the solution properties of hnRNP H. Indeed, we demonstrated that the linker of HqRRM1,2 is relatively rigid (on the ps-ns time scale) by comparison to FqRRM1,2. The linkers from both proteins are nearly identical, with the notable exception at position 105, where hnRNP F has an alanine and hnRNP H a proline. Interestingly, position 105 (hnRNP H numbering) is variable across the hnRNP F/H family, suggesting that the identity of the residue at 105 plays a general role in modulating the conformational dynamics of this family of proteins. To that end, the P105A mutant of HqRRM1,2 behaves more like FqRRM1,2, as determined by its larger radius of gyration and overall NMR spin relaxation profile. Linker prolines are known to tune the conformational dynamics of multi-domain proteins by conferring structural rigidity or by acting as hinge points to allow inter-domain movement.[11,29] The Src tyrosine kinase family is a paradigmatic system for evaluating the influence of linker composition on the function and conformational properties of multi-domain proteins.[30] The Hck tyrosine kinase contains a 14-residue linker, with a PXXP motif, that connects its SH2 and kinase domains. Substitution of prolines 225 and 228 with alanines relieves an autoinhibitory interaction between the SH3 domain and a polyproline type II helix within the linker, resulting in deregulated kinase activity.[31] Moreover, SAXS ensemble analysis of the related Bruton's protein tyrosine kinase (BTK) demonstrated that the proline-rich linker connecting its PH-SH3 domains contributes to a dynamic interconversion between open and closed states.[32] By comparison, the work presented here indicates that the IQL of HqRRM1,2, with its PXXP motif, differentially modulates the conformational dynamics of hnRNP H such that the protein fluctuates between compact and extended structures.

4. MATERIALS AND METHODS

Cloning, Expression, Mutagenesis, and Purification of hnRNP F/H Sub-domains. The PCR-amplified cDNA encoding the qRRM1 (residues 10−111), qRRM2 (residues 94−194), and qRRM1,2 (residues 10−194) domains of hnRNP H was cloned into a bacterial expression pMCSG7 vector. The recombinant proteins were over-expressed in BL21(DE3) as host cells. Cells were grown at 37 °C to OD_{600} = 0.8, and then adjusted to 20 °C for 30 min before induction. Cells were induced with 1.0 mM IPTG and allowed to express for 16 h. Cells were lysed by sonication at 4 °C in 20 mM Na_2HPO_4, 20 mM imidazole, 500 mM NaCl, and 4 mM TCEP (tris(2-carboxyethyl)phosphine) at pH 8.0. Clarified lysate was filtered and was initially purified on His-Select resin equilibrated in the lysis buffer and washed with 20 mM Na_2HPO_4, 40 mM imidazole, 500 mM NaCl, and 4 mM TCEP at pH 8.0. The protein was eluted with lysis buffer containing 500 mM imidazole. Protein samples were concentrated and buffer exchanged into the lysis buffer. His_6 tag was removed by TEV cleavage (1−2 units per mg of protein) incubated at room temperature for 16 h. The uncleaved His-tagged protein and TEV were removed using a complete His-Tag purification column (Roche). The cleaved protein was further purified by size exclusion chromatography on a Superdex 75 column (GE Healthcare Life Sciences) pre-equilibrated with 20 mM HEPES, pH 6.5, 100 mM sodium chloride, and 2 mM TCEP. The purity of the protein was estimated to be >95% by SDS-PAGE.

A codon-optimized gene block encoding the qRRM1,2 (residues 1−194) domain of hnRNP F was purchased from IDT and cloned into the pMCSG7 vector between *Nde*I and *Eco*RI restriction sites. A protocol similar to that described for HqRRM1,2 was followed to express and purify the FqRRM1,2.

To prepare HqRRM1,2 constructs for PRE studies, site-directed mutagenesis was carried out by PCR, amplifying the wild-type HqRRM1,2 cDNA with Phusion polymerase (NEB) in the presence of the corresponding forward and reverse mutation primer sets. The amplified PCR products were digested by *Dpn*I at 37 °C overnight and transformed into *E. coli* NEB5-α cells.

Crystallization and X-ray Structure Determination of HqRRM1,2. For crystallization, HqRRM1,2 (residues 10−194) was concentrated to 16 mg/mL in buffer containing 20 mM HEPES, pH 6.5, 100 mM sodium chloride, and 2 mM TCEP. Crystals of HqRRM1,2 grew at 20 °C from drops containing equal volumes of protein and well solution (30−50% polyethylene glycol 400 and 0.1 M phosphate-citrate, pH 4.2). Prior to data collection, crystals were flash frozen in liquid nitrogen. Selenomethionine-incorporated HqRRM1,2 was expressed in Rosetta^2 cells in M9 minimal medium supplemented by an amino acid mixture containing selenomethonine as previously described[33] and purified the same as native protein.

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

Crystals of selenomethionine-incorporated HqRRM1,2 grew under similar conditions.

Data were collected at Advanced Photon Source at Argonne National Laboratory on LS-CAT beamline 21-ID-F at a wavelength of 0.9787 Å and processed with HKL2000.[34] The HqRRM12 crystallized the in the space group $P6_422$, with a unit cell of a = 204.668, b = 204.668, and c = 123.792 Å, $\alpha = \beta = 90°$, and $\gamma = 120°$. There are four molecules in the asymmetric unit, with a solvent content of 72.8%. We initially attempted to solve the structure of native HqRRM1,2 by molecular replacement using both Molrep and Phaser using various models of RRM domains, with no success. We were able to grow selenomethionine-derived crystals, and phases were determined by single-wavelength anomalous X-ray scattering from the selenium atoms using AutoSol in Phenix.[35] HqRRM1,2 contains two methionine residues, both in RRM1, making the correct solution unambiguous. A higher resolution data set (to 3.5 Å) was later collected, and the structure was solved by molecular replacement using the previously solved structure as a model. The structure was iteratively fit in Coot[36] and refined in Buster.[37] The structure was validated using Molprobity.[38] Data refinement and statistics are given in Table S1.

NMR Experiments. The resonance assignments were obtained using standard 2D and 3D heteronuclear NMR experiments performed on a uniformly double (^{15}N and ^{13}C)-labeled sample. All NMR experiments were acquired at 305 K on Bruker 800 and 900 MHz spectrometers equipped with triple-resonance cryoprobes. The protein sample concentrations used for all NMR experiments were in the 0.6−1.0 mM range. The ^{15}N-labeled and ^{15}N/^{13}C-labeled samples for NMR experiments were buffer exchanged in 20 mM sodium phosphate, 20 mM NaCl, 4 mM TCEP at pH 6.2. The 3D triple-resonance experiments used for qRRM2 and HqRRM1,2 involved HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO experiments. The side-chain assignments for qRRM2 were obtained via the HBHA(CO)NH, (H)CCCH-TOCSY, and H(C)CCH-TOCSY experiments, with a TOCSY mixing time using 25 ms. 3D NOESY-(^{13}C,^{1}H)-HSQC and 3D NOESY-(^{15}N,^{1}H)-HSQC spectra were recorded with a mixing time of 150 ms. The NMR data were processed using NMRPipe[39] and analyzed by PINE-SPARKY.[40]

The 8% polyethylene glycol-alkyl ether (PEG) bicelles were prepared by adding 50 μL of C12E5 (pentaethylene glycol monododecyl ether), 16 μL of hexanol, and 250 μL of buffer containing 20 mM sodium phosphate, 20 mM NaCl, 4 mM TCEP, and 10% D_2O at pH 6.2. The NMR samples were prepared by adding protein and PEG in a 1:1 ratio. The sample was placed in the NMR magnet, and ^{2}H splitting about 29 Hz was measured after 30 min.

For NMR titrations, the uniformly ^{15}N-labeled HqRRM1,2 samples were prepared at a concentration of 90 μM in 20 mM sodium phosphate, 20 mM NaCl, 4 mM TCEP, and 10% D_2O at pH 6.2. The unlabeled 5′-AGGGU-3′ oligonucleotide was added to uniformly ^{15}N-labeled HqRRM1,2 at molar ratios of 1:0:33, 1:0:66, 1:1, 1:2, and 1:4. All spectra were processed with NMRPipe/DRAW and analyzed using Sparky.[41]

NMR Backbone Dynamics of qRRM1, qRRM2, and qRRM1,2. All the T_1, T_2 relaxation and the {^{1}H}-^{15}N nuclear Overhauser effect (NOE) data were measured on a Bruker 800 MHz spectrometer at 305 K for HRRM1,2. Temperatures of 298, 303, 308, and 313 K were selected for temperature dependence studies. T_1 delays of 50, 100, 150, 200, 300, 400, 600, 800, 1000, 1200, and 1500 ms were used, with repeated 300 and 800 ms. T_2 delays of 0, 16, 33, 49, 66, 98, 115, 148, 197, and 246 ms were used, with repeated 0, 66, and 98 ms. The spectral data were processed through NMRPipe and Sparky. The relaxation values of ^{15}N R_1 ($1/T_1$) and ^{15}N R_2 ($1/T_2$) were analyzed by fitting the series of peak intensities with an exponential decay curve in Sparky software. The NOE values were derived from the ratio of peak intensities between saturated and unsaturated NOE spectra. PDB Inertia was used to transform the original coordinates into the protein's center of mass. Quadric was applied to estimate rotational diffusion tensors, including Dratio and θ/Φ angle by using R_2/R_1 Diffusion programs. ModelFree 4.2[42] was utilized to derive R_{ex} for the isolated domains HqRRM1 and HqRRM2 by fitting R_1, R_2, and NOE with error values. The default values of the N−H bond lengths and ^{15}N chemical shift anisotropy were 1.02 Å and −160 ppm, respectively. The correlation time was initially set to 8 ns during the 25 loops of calculations to fit the five models using model-free formalism. Overall rotational correlation times (τ_c) were estimated from the T_1/T_2 ratio with the amide residues that have non-overlapping peaks in the HSQC spectrum.

Paramagnetic Relaxation Enhancement (PRE). For each spin-labeled sample of HqRRM1,2, paramagnetic samples were prepared with an excess of *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1*H*-pyrrol-3-yl)methylmethanesulfonothioate (MTSL) by reaching the molar ratio of 5:1 (MTSL:protein = 5:1 ratio). The impact of spin labeling on the structure of HqRRM1,2 was evaluated by overlapping the HSQC spectrum with that of the non-labeling sample; only the mutant spectrum without significant change after spin labeling was used for further analysis (Supporting Information). The 2D ^{1}H−^{15}N HSQC spectrum was used to measure the PRE by recording the sample/parameter matched pair for each spin-labeling sample in diamagnetic and paramagnetic states. NMRPipe was used to process the spectra, and the resonance intensities were measured in SPARKY to determine the intensity ratio of the paramagnetic state vs the diamagnetic state. Based on the intensity ratio, ^{1}H transverse rate ($R_{2,pre}$) was calculated. $R_{2,pre}$ was used to determine the distance between the nitroxide and amide proton. Intensity ratios less than 0.2 normally were classified as close, and the distance restraint was set as 12 Å, with an upper limit of +4 Å. The cross peaks that were unaffected in the presence of MTSL (intensity ratios higher than 0.85) were restrained to >25 Å. Those resonances with intensities between 0.2 and 0.85 were converted into distances.[17,43,44] The grid search was applied to optimize the lowest Q value by including the local motion of spin label (τ_{pre}). The 10 lowest energy structures were calculated, and the distance between the average positions of MTSL and amide proton was back-calculated. The standard deviations of back-calculated distances were converted into the PRE intensity ratios and compared with the experimental results.

Structure Calculation of HqRRM2 and HqRRM1,2. Structures of HqRRM2 were calculated using ARIA2.3/CNS.[45,46] The distance restraints of isolated HqRRM2 were accessed from ^{15}N-edited and ^{13}C-edited HSQC-NOESY spectra (mixing time 150 ms), and NOE assignments were automatically selected using ARIA2.3. The backbone dihedral torsion angles' restraints (ψ,φ) were obtained by using TALOS based on chemical shift assignments (HN, HA, CA, CB, CO, N) of HqRRM2. The hydrogen bond restraints were included as determined by using chemical shift index 2.0.[47]

The initial extended structures of HqRRM1,2 were calculated using ARIA2.3/CNS with distance restraints obtained from isolated domains. Backbone dihedrals and hydrogen bond restraints were obtained from TALOS and CSI, respectively. The final 10 lowest energy structures were taken on to XPLOR/CNS[48] calculations. We used a simulated annealing protocol (refine.py script) that includes NOE, PRE, and RDC restraints. The sequence of HqRRM1,2 was modified to add MTSL-labeled Cys residues where PRE measurements were available. Using an extended starting structure, a total of 800 structures were calculated with a simulated annealing protocol in which the bath temperature was lowered from 3000 to 300 K. During the cooling stage, the van der Waals interactions were increased by varying the force constant of the repel function from 0.003 to 4 kcal·mol^{-1} Å^{-4} while the van der Waals radii were decreased from 0.9 to 0.75. A force constant of 200 kcal·mol^{-1}·rad^{-1} was used for the dihedral angle restraints. Force constants for NOE, hydrogen bond restraints, and PRE were fixed at 25, 25, and 25 kcal·mol^{-1}·Å^{-2} respectively, with flat-well harmonic potentials, and other parameters were set as default values. The 10 lowest energy structures were selected from the structural ensemble for further structure calculations. The final structures were refined using XPLOR water refinement scripts with default parameters. The ensemble was further analyzed with PROCHECK-NMR.[49]

SEC-SAXS Collection and Processing. Inline SEC-SAXS data were collected at BioCAT (beamline 18-ID; Advanced Photon

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

Source). All the HqRRM1,2 protein samples were buffer exchanged in 20 mM HEPES, 20 mM NaCl, 4 mM TCEP, pH 6.2, using a SEC column before SEC-SAXS experiment. A 200 μL concentrated sample of HqRRM1,2 (6−10 mg/mL) was loaded on the SEC column, and scattering data were acquired every 2 s of the exposure during the SEC run. The data points of single peaks in the UV and scattering intensity of the same radius of gyration (R_g) were considered for further analysis.

The PRIMUS module from ATSAS[50] and SCATTER[51] programs were used to analyze the scattering data. The scattering intensity $I(0)$, radius of gyration (R_g), particle distance distributions $P(r)$, and maximum particle dimensions (D_{max}) for all the fragments were calculated using the PRIMUS[52] and GNOM[53] modules for molecule reconstruction. Ensemble optimization method (EOM2.0)[19] was employed to calculate the average theoretical scattering density from the pool that fits with experimental SAX density. CRYSOL[54] was used to report the chi values of the fit the models to the experimental data.

Relaxation Dispersion. Experiments were performed at two spectrometer frequencies of 600 and 850 MHz (for HqRRM1,2) or 600 MHz (for FqRRM1,2) at 288 K. For the 600 MHz/850 MHz measurements, pseudo-3D data sets were collected using 23/22 CPMG field strengths ranging from 25 to 1000/2000 Hz, and 40/30 ms was set for the constant-time relaxation period. The NMR data were processed using NMRpipe and NMRFAM-SPARKY. Peak intensities were extracted with nlinLS and further analyzed by numerical simulation of the pulse sequence using ChemEx software version 0.6.1.[24] Those residues that exhibit R_{ex} differences in their effective relaxation rates at low and high CPMG field strength larger than 3 s^{-1} were fitted simultaneously with a two-state exchange model. The Bloch−McConnell equation was applied to fit the dispersion profiles and derive the k_{ex} between a major state and an excited state as well as the populations of each state (p_A and p_B). To obtain accurate global fits for k_{ex} and p_B, dispersion profiles were first fitted on a per-residue basis, and then residues were selected for determining k_{ex} and p_B.

Isothermal Titration Calorimetry Experiments. The binding affinities of HqRRM1, HqRRM2, and HqRRM1,2 with the G-tract RNA (AGGGU) oligonucleotides were characterized by measuring heat changes on titrating protein domains into each G-tract RNA oligonucleotide solution using a Microcal VP-ITC calorimeter. Protein and RNA solutions were buffer exchanged to 20 mM sodium phosphate, 20 mM NaCl, and 4 mM TCEP at pH 6.2, centrifuged, and degassed under vacuum before use. All titrations were performed at 25 °C, and the data were analyzed using the KinITC routines supplied with Affinimeter.[25]

Detection of ^{19}F PREs between HqRRM1,2 and AGGGA*U. The ^{19}F NMR active probe bromotrifluoroacetanilide (BTFA) was chemically ligated to HqRRM1,2 by resuspending cell pellets in buffer containing 20 mM Na_2HPO_4 (pH 8), 20 mM imidazole, 500 mM NaCl, and 25 μL of BTFA (8 M stock solution in acetonitrile) for 30 min on ice. The BTFA-labeled protein was then purified as described above.

To detect ^{19}F PREs to our BTFA-labeled HqRRM1,2, we purchased the AGGGA*U oligo that contained a specific phosphorothioate between the A5 and U6 and reacted this oligo with 3-(2-iodoacetamidomethyl)-PROXYAL (IAM-PROXYAL). In brief, 0.3 mM AGGGA*U oligonucleotide was dissolved in 300 μL of 50 mM TEAA buffer, and 10 equiv of IAM-PROXYAL in 300 μL of TEAA pH 6.5/DMF (2:1 ratio) was added to the reaction mixture, which was then incubated at 50 °C for 8 h. The reaction mixture was washed with chloroform to remove excess IAM-PROXYAL and further purified with anion exchange chromatography followed by size exclusion chromatography in water. The final samples that contain the spin-labeled AGGGA*U were kept under vacuum centrifuge to remove water and exchanged into 20 mM sodium phosphate, 20 mM NaCl, pH 6.2. All ^{19}F NMR experiments were performed at 305 K on a 500 MHz Bruker spectrometer equipped with a PRODIGY probe. All spectral data were processed with Topspin3.0.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b05366.

HqRRM1,2 crystallographic unit cell information; comparison of NMR and X-ray structures of HqRRM1,2; NMR HSQC spectra of HqRRM1,2; temperature dependence of global ^{15}N T_1 relaxation data; comparison of HqRRM1,2 and HqRRM1,2^{P105A} T_1, T_2 relaxation data; solution NMR structure of HqRRM2; comparison of HSQC spectra of HqRRM1,2 mutants used for PRE measurements; experimental SEC-SAXS data; experimental ^{15}N R_{ex} data; crystallographic and NMR structure calculation statistics; and summary of NMR and SAXS HqRRM1,2 and FqRRM1,2 solution properties, including Figures S1−S11 and Tables S1−S4 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*bst18@case.edu

ORCID

Blanton S. Tolbert: 0000-0003-2456-0443

Author Contributions

†S.R.P. and L.-Y.C. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by grants from the National Institutes of Health (NIH): R01GM101979 (to B.S.T.) and U54GM103297 (to J.A.S.). This research also used resources of the Advanced Photon Source, a United States Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract DE-AC0206CH11357. Additional support was provided by Grant P41 GM103622 from the National Institute of General Medical Sciences of the NIH. Use of the Pilatus 31M detector was provided by NIH Grant 1S10OD018090-1. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation, and the Michigan Technology Tri-Corridor is thanked for the support of this research program (Grant 085P1000817). The authors would also like to acknowledge Dr. Hsuan-Chun Lin for assistance with running ModelFree.

■ REFERENCES

(1) Geuens, T.; Bouhy, D.; Timmerman, V. *Hum. Genet.* **2016**, *135* (8), 851−867.

(2) Dreyfuss, G.; Matunis, M. J.; Pinol-Roma, S.; Burd, C. G. *Annu. Rev. Biochem.* **1993**, *62*, 289−321.

(3) Han, S. P.; Tang, Y. H.; Smith, R. *Biochem. J.* **2010**, *430* (3), 379−92.

(4) Caputi, M.; Zahler, A. M. *J. Biol. Chem.* **2001**, *276* (47), 43850−9.

(5) Dominguez, C.; Fisette, J. F.; Chabot, B.; Allain, F. H. *Nat. Struct. Mol. Biol.* **2010**, *17* (7), 853−61.

(6) Huelga, S. C.; Vu, A. Q.; Arnold, J. D.; Liang, T. Y.; Liu, P. P.; Yan, B. Y.; Donohue, J. P.; Shiue, L.; Hoon, S.; Brenner, S.; Ares, M., Jr.; Yeo, G. W. *Cell Rep.* **2012**, *1* (2), 167−78.

(7) Clery, A.; Blatter, M.; Allain, F. H. *Curr. Opin. Struct. Biol.* **2008**, *18* (3), 290−8.

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673

(8) Dominguez, C.; Allain, F. H. *Nucleic Acids Res.* **2006**, *34* (13), 3634−3645.
(9) Schaub, M. C.; Lopez, S. R.; Caputi, M. *J. Biol. Chem.* **2007**, *282* (18), 13617−13626.
(10) Ravi Chandra, B.; Gowthaman, R.; Akhouri, R. R.; Gupta, D.; Sharma, A. *Protein Eng., Des. Sel.* **2004**, *17* (2), 175−182.
(11) Papaleo, E.; Saladino, G.; Lambrughi, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. *Chem. Rev.* **2016**, *116* (11), 6391−423.
(12) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (4), 486−501.
(13) Chang, S. L.; Szabo, A.; Tjandra, N. *J. Am. Chem. Soc.* **2003**, *125* (37), 11379−84.
(14) Ramelot, T. A.; Yang, Y.; Pederson, K.; Shastry, R.; Kohan, E.; Janjua, H.; Xiao, R.; Acton, T. B.; Everett, J. K.; Prestegard, J. H.; Montelione, G. T.; Kennedy, M. A. *BMRB* **2012**, 18698.
(15) Gobl, C.; Madl, T.; Simon, B.; Sattler, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *80*, 26−63.
(16) Mackereth, C. D.; Madl, T.; Bonnal, S.; Simon, B.; Zanier, K.; Gasch, A.; Rybin, V.; Valcarcel, J.; Sattler, M. *Nature* **2011**, *475* (7356), 408−11.
(17) Battiste, J. L.; Wagner, G. *Biochemistry* **2000**, *39* (18), 5355−65.
(18) Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am. Chem. Soc.* **2007**, *129* (17), 5656−64.
(19) Tria, G.; Mertens, H. D.; Kachala, M.; Svergun, D. I. *IUCrJ* **2015**, *2* (2), 207−217.
(20) Farber, P. J.; Mittermaier, A. *Biophys. Rev.* **2015**, *7* (2), 191−200.
(21) Neudecker, P.; Lundstrom, P.; Kay, L. E. *Biophys. J.* **2009**, *96* (6), 2045−54.
(22) Baldwin, A. J.; Kay, L. E. *Nat. Chem. Biol.* **2009**, *5* (11), 808−14.
(23) Kleckner, I. R.; Foster, M. P. *Biochim. Biophys. Acta, Proteins Proteomics* **2011**, *1814* (8), 942−68.
(24) Vallurupalli, P.; Bouvignies, G.; Kay, L. E. *J. Am. Chem. Soc.* **2012**, *134* (19), 8148−61.
(25) Burnouf, D.; Ennifar, E.; Guedich, S.; Puffer, B.; Hoffmann, G.; Bec, G.; Disdier, F.; Baltzinger, M.; Dumas, P. *J. Am. Chem. Soc.* **2012**, *134* (1), 559−65.
(26) Dumas, P.; Ennifar, E.; Da Veiga, C.; Bec, G.; Palau, W.; Di Primo, C.; Pineiro, A.; Sabin, J.; Munoz, E.; Rial, J. *Methods Enzymol.* **2016**, *567*, 157−80.
(27) Uren, P. J.; Bahrami-Samani, E.; de Araujo, P. R.; Vogel, C.; Qiao, M.; Burns, S. C.; Smith, A. D.; Penalva, L. O. *RNA Biol.* **2016**, *13* (4), 400−11.
(28) Tamayo, J. V.; Teramoto, T.; Chatterjee, S.; Hall, T. M. T.; Gavis, E. R. *Cell Rep.* **2017**, *19* (1), 150−161.
(29) Ruiz, D. M.; Turowski, V. R.; Murakami, M. T. *Sci. Rep.* **2016**, *6*, 28504.
(30) Xiao, Y.; Liddle, J. C.; Pardi, A.; Ahn, N. G. *Acc. Chem. Res.* **2015**, *48* (4), 1106−14.
(31) Briggs, S. D.; Smithgall, T. E. *J. Biol. Chem.* **1999**, *274* (37), 26579−83.
(32) Marquez, J. A.; Smith, C. I.; Petoukhov, M. V.; Lo Surdo, P.; Mattsson, P. T.; Knekt, M.; Westlund, A.; Scheffzek, K.; Saraste, M.; Svergun, D. I. *EMBO J.* **2003**, *22* (18), 4616−4624.
(33) Begley, M. J.; Taylor, G. S.; Kim, S. A.; Veine, D. M.; Dixon, J. E.; Stuckey, J. A. *Mol. Cell* **2003**, *12* (6), 1391−402.
(34) Minor, Z. O. a. W. *Methods Enzymol.* **1997**, *276* (20), 307−326.
(35) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (2), 213−221.
(36) Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60* (12), 2126−2132.
(37) Bricogne, G.; Blanc, E.; Brandl, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P.; Smart, O. S.; Vonrhein, C.; Womack, T. O. *BUSTER*, version 2.11.2; Global Phasing Ltd.: Cambridge, UK, 2017.
(38) Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (1), 12−21.
(39) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6* (3), 277−293.
(40) Lee, W.; Westler, W. M.; Bahrami, A.; Eghbalnia, H. R.; Markley, J. L. *Bioinformatics* **2009**, *25* (16), 2085−7.
(41) Lee, W.; Tonelli, M.; Markley, J. L. *Bioinformatics* **2015**, *31* (8), 1325−7.
(42) Mandel, A. M.; Akke, M.; Palmer, A. G., III. *J. Mol. Biol.* **1995**, *246* (1), 144−163.
(43) Sjodt, M.; Clubb, R. T. *Bio. Protoc.* **2017**, *7* (7), 2207.
(44) Hennig, J.; Warner, L. R.; Simon, B.; Geerlof, A.; Mackereth, C. D.; Sattler, M. *Methods Enzymol.* **2015**, *558*, 333−62.
(45) Rieping, W.; Habeck, M.; Bardiaux, B.; Bernard, A.; Malliavin, T. E.; Nilges, M. *Bioinformatics* **2007**, *23* (3), 381−2.
(46) Brunger, A. T. *Nat. Protoc.* **2007**, *2* (11), 2728−33.
(47) Berjanskii, M. V.; Wishart, D. S. *J. Am. Chem. Soc.* **2005**, *127* (43), 14970−1.
(48) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Marius Clore, G. *J. Magn. Reson.* **2003**, *160* (1), 65−73.
(49) Laskowski, R. A. *Nucleic Acids Res.* **2009**, *37* (Database), D355.
(50) Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; Svergun, D. I. *J. Appl. Crystallogr.* **2017**, *50* (4), 1212−1225.
(51) Nielsen, S. S.; Toft, K. N.; Snakenborg, D.; Jeppesen, M. G.; Jacobsen, J. K.; Vestergaard, B.; Kutter, J. P.; Arleth, L. *J. Appl. Crystallogr.* **2009**, *42* (5), 959−964.
(52) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. *J. Appl. Crystallogr.* **2003**, *36* (5), 1277−1282.
(53) Svergun, D. *J. Appl. Crystallogr.* **1992**, *25* (4), 495−503.
(54) Svergun, D.; Barberato, C.; Koch, M. H. J. *J. Appl. Crystallogr.* **1995**, *28* (6), 768−773.

DOI: 10.1021/jacs.8b05366
J. Am. Chem. Soc. 2018, 140, 11661−11673