Received: 23 December 2024 | Revised: 21 March 2025 | Accepted: 3 April 2025

DOI: 10.1002/pro.70127

RESEARCH ARTICLE

THE PROTEIN SOCIETY WILEY

An evolutionarily conserved tryptophan cage promotes folding of the extended RNA recognition motif in the hnRNPR-like protein family

Ernest S. Atsrim [1] | **Catherine D. Eichhorn** [1,2]

[1]Department of Chemistry, University of Nebraska, Lincoln, Nebraska, USA

[2]Nebraska Center for Integrated Biomolecular Communication, Lincoln, Nebraska, USA

Correspondence
Catherine D. Eichhorn, Department of Chemistry, University of Nebraska, 639 North 12th St, Lincoln, NE 68588, USA.
Email: ceichhor@unl.edu

Funding information
National Institutes of Health, Grant/Award Number: 1R35GM143030; Nebraska Center for Integrated Biomolecular Communication, Grant/Award Number: P20 GM113126

Review Editor: Aitziber L. Cortajarena

Abstract

The heterogeneous nuclear ribonucleoprotein (hnRNP) R-like family is a class of RNA binding proteins in the hnRNP superfamily with diverse functions in RNA processing. Here, we present the 1.90 Å X-ray crystal structure and solution NMR studies of the first RNA recognition motif (RRM) of human hnRNPR. We find that this domain adopts an extended RRM (eRRM1) featuring a canonical RRM with a structured N-terminal extension (N_{ext}) motif that docks against the RRM and extends the β-sheet surface. The adjoining loop is structured and forms a tryptophan cage motif to position the N_{ext} motif for docking to the RRM. Combining mutagenesis, solution NMR spectroscopy, and thermal denaturation studies, we evaluate the importance of residues in the N_{ext}–RRM interface and adjoining loop on eRRM folding and conformational dynamics. We find that these sites are essential for protein solubility, conformational ordering, and thermal stability. Consistent with their importance, mutations in the N_{ext}–RRM interface and loop are associated with several cancers in a survey of somatic mutations in cancer studies. Sequence and structure comparison of the human hnRNPR eRRM1 to experimentally verified and predicted hnRNPR-like proteins reveals conserved features in the eRRM.

KEYWORDS
atypical RRM, biomolecular NMR, protein dynamics, protein–protein interactions, thermal denaturation, Trp-cage

1 | INTRODUCTION

Heterogeneous nuclear ribonucleoproteins (hnRNPs) are RNA-binding proteins with diverse functions in cellular processes (Dreyfuss et al. 1993). Of these, the hnRNPR-like family is broadly involved in RNA processing and includes hnRNPR, hnRNPQ (also known as SYNCRIP and NASP1), dead end protein 1 (DND1), RBM46, RBM47, APOBEC1 complementation factor (A1CF), and GRY-RBP (Lau et al. 2001). hnRNPR plays essential roles in splicing and transcription regulation (Agbo et al. 2021; Briese et al. 2018; Fukuda et al. 2009; Ji et al. 2022; Jiang et al. 2023; Kabat et al. 2009; Wang et al. 2024). In mammals, the major isoform of hnRNPR consists of a predicted N-terminal α-helical bundle (NαB), three tandem RNA recognition motifs (RRMs), and a C-terminal RGG-box domain (Han et al. 2010) (Figure 1a). A second isoform is also expressed in which the NαB is absent (Ghanawi et al. 2021; Huang et al. 2005; Peng et al. 2009). Highly expressed in the nervous system (Rossoll et al. 2002), hnRNPR function is critical to axonal growth and neuronal development (Briese et al. 2018; Glinka et al. 2010; McWhorter et al. 2003; Zare et al. 2024). Deletion of

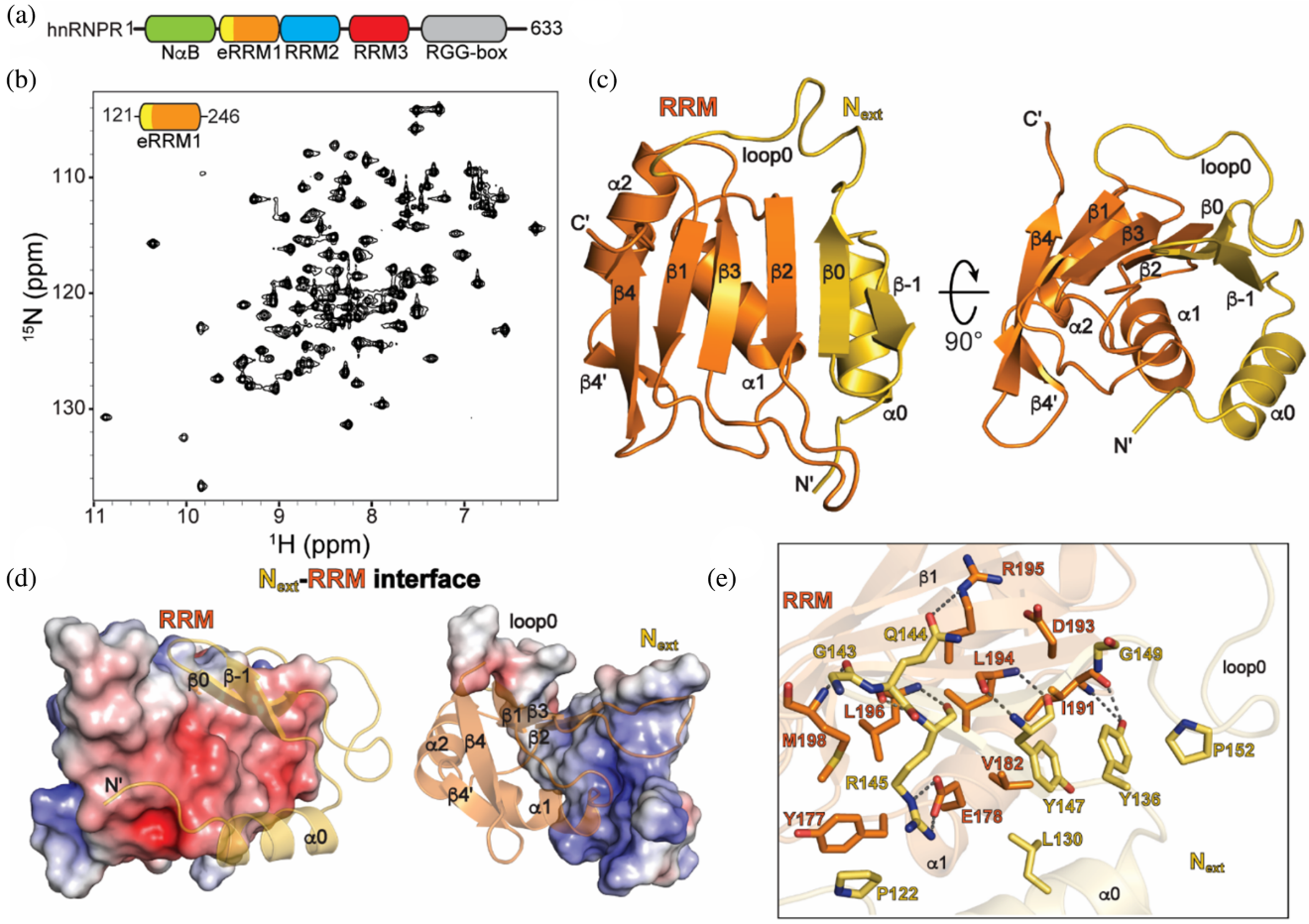

FIGURE 1 The first RNA binding domain of hnRNPR is an eRRM. (a) Domain topology of human hnRNPR. (b) ^{1}H-^{15}N HSQC spectrum of WT eRRM1 construct shows well-folded protein at 25°C. (c) 1.90 Å X-ray crystallographic structure of WT eRRM1 shows N_{ext} (colored gold) and RRM (colored orange) motifs. (d) Electrostatic surface potential map of the buried interface between the RRM and N_{ext} motifs show a charged interface. (e) Key residues participate in hydrogen bonding and hydrophobic interactions at the N_{ext}–RRM interface.

RRM1 and RRM2 impaired β-actin content in growth cones and neurite growth (Rossoll et al. 2003). Overexpression and/or dysfunction of hnRNPR is linked to spinal muscular atrophy (Gascon et al. 2024; Gillentine et al. 2021; Hu et al. 2024; Jiang et al. 2023) and cancer metastasis (Chen et al. 2019; Li et al. 2022; Wang et al. 2024; Yang et al. 2023). Mutations in the C-terminal RGG-box are associated with developmental disorders (Duijkers et al. 2019; Gillentine et al. 2021). hnRNPR interacts with proteins such as Yb1 (Ghanawi et al. 2021), SMN (Rossoll et al. 2002), HMGC (Agbo et al. 2021), and ALS-associated proteins TDP-43 and FUS (Gascon et al. 2024). hnRNPR has several identified RNA substrates including MHC class I mRNA (Reches et al. 2016), β actin mRNA (Glinka et al. 2010), UPF3B mRNA (Wang et al. 2024), ASCL1 mRNA (Hu et al. 2024), and 7SK noncoding RNA (Briese et al. 2018; Briese and Sendtner 2021; Ji et al. 2021; Ji et al. 2022). Prior studies indicate that the first two RRMs are the primary RNA recognition sites (Rossoll et al. 2002; Rossoll et al. 2003; Wang et al. 2024).

Despite its significance, remarkably little is known regarding the structural and biophysical properties of hnRNPR. While an hnRNPR AlphaFold2 (AF2) predicted structural model is available, the only experimentally determined high-resolution structure is a solution NMR structural ensemble of RRM3 (PDB ID 2DK2). Canonical RRMs consist of an antiparallel β-sheet composed of four to five β-strands, with two α-helices that lie underneath the β-sheet (Maris et al. 2005). The β-sheet surface serves as the RNA recognition site, with conserved ribonucleoprotein (RNP) sequence motifs on β1 (RNP2) and β3 (RNP1) that typically contain aromatic and/or basic residues (Maris et al. 2005). RRMs are frequently decorated with secondary structure elements at the N- and/or C-termini to promote RNA and/or protein substrate recognition and specificity (Daubner et al. 2013; Eichhorn et al. 2018; Maris et al. 2005; Singh et al. 2012). In particular, the U1-70K spliceosomal protein contains an RRM with an N-terminal α-helix, which improves RNA binding specificity and affinity (Gopan et al. 2022). The first RRM of

DND1 and hnRNPQ contains an N-terminal α-helix and β-hairpin, which contribute to the RNA binding surface (Duszczyk et al. 2022; Hobor et al. 2018). All three atypical RRMs have been named extended RRMs (eRRMs) to denote their N-terminal secondary structure extension.

The tryptophan cage is one of the smallest protein motifs, requiring approximately 18–20 aa residues to adopt a stable folded structure (Neidigh et al. 2002). Comprised of hydrophobic residues that organize around a central tryptophan residue, the tryptophan cage motif was first identified in the C-terminus of the exendin-4 peptide (Neidigh et al. 2001). This motif was isolated and stabilized through a series of substitutions to generate the Trp-cage peptide (Neidigh et al. 2002). The Trp-cage peptide is a model system for protein folding studies owing to its fast folding kinetics, with numerous computational and NMR studies (Meuzelaar et al. 2013; Zhou 2003). However, this motif has not been identified within a naturally occurring protein, or within a folded domain.

Here, we combine X-ray crystallography, solution NMR spectroscopy, thermal denaturation studies, and mutagenesis to determine the structure and conformational dynamics of the first RRM of the human hnRNPR (Hs-hnRNPR). We find that this domain is an eRRM consisting of an N-terminal extension (N_{ext}) motif docked to a canonical RRM with structural similarity to hnRNPQ and DND1 eRRMs. Features include a structured loop connecting the N_{ext} to the RRM that organizes around a tryptophan residue in the β-sheet, forming a tryptophan cage motif that is conserved among hnRNPR-like proteins. Residues at the N_{ext}–RRM interface, as well as in the connecting loop, are highly conserved among members of the hnRNPR-like family. N-terminal truncations or substitutions to residues in the N_{ext}–RRM interface result in reduced protein solubility, reduced thermal stability, and increased conformational dynamics. Similarly, substitutions within the tryptophan cage result in reduced thermal stability and increased conformational dynamics, showing the importance of this motif in eRRM folding. Importantly, residues identified in this study to be important for protein folding are correlated with cancer-related missense mutations, suggesting a role in human health and disease. Together, this study characterizes features that define the eRRM1, identifies the tryptophan cage as a naturally occurring motif, and demonstrates the importance of stable association of the N_{ext} with the RRM for eRRM folding.

2 | RESULTS

2.1 | Structure of the first RRM domain of hnRNPR is an extended RRM

Constructs of the first RRM domain of hnRNPR were designed with N- and C-terminal boundaries determined from sequence comparison to hnRNPR-like proteins and the AF2 predicted model (Figure S1). Human hnRNPR eRRM1 (WT eRRM1, residues 121–246) containing an N-terminal histidine tag and TEV cleavage site was overexpressed in *Escherichia coli* and purified (see section 4). The solution state NMR ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum showed excellent chemical shift dispersion indicating a stably folded protein (Figures 1b and S2). NMR resonance assignments were performed using standard triple-resonance experiments, with 94.83% completeness for backbone resonances (N, H, Cα, Cβ, CO). Residues D179–L181 (α1) could not be assigned due to line broadening. We performed crystal screening of constructs for X-ray crystallographic structure determination and identified conditions yielding crystals that diffracted to 1.90 Å (Table 1).

WT eRRM1 folds as an eRRM containing a canonical RRM domain with an N-terminal extension (Figure 1c). The canonical RRM contains a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4'\beta_4$ topology and is comprised of a β-sheet with five anti-parallel β-strands and two α-helices that lie underneath the β-sheet. This secondary structure is consistent with predictions made from chemical shift assignments using TALOS+ (Shen et al. 2009; Shen and Bax 2015) (Figure S3). On the β-sheet surface, the RRM domain has conserved RNA recognition sequences RNP2 on β1 (V167–F168–V169) and RNP1 on β3 (Y208–A209–F210) (Burd and Dreyfuss 1994) and an electropositive surface potential (Figure S4). An 18 aa structured loop (loop0, G149–T165) links β0 in the N_{ext} to β1 in the RRM. The N_{ext} contains an α0β-1β0 topology and is comprised of an α-helix that lies underneath a β-hairpin. The N_{ext} β-hairpin docks alongside the β2-α1 side of the canonical RRM, extending the β-sheet surface by two β-strands. The N_{ext}–RRM buried interface is highly charged, with an electronegative surface potential on the β2-α1 side of the RRM and an electropositive surface potential on the α0-β0 side of the N_{ext} (Figure 1d). There are several interactions between N_{ext} and RRM subunits including hydrogen bonds between residues G143 (loopβ-1-β0)–M198 (β2), Q144 (β0)–R195 (β2), R145 (β0)–L196 (β2), R145 (β0)–E178 (α1), Y147 (β0)–L194 (β2), and G149 (loop0)–D193 (β2) as well as hydrophobic contacts between residues P122–Y177, L130–V182, Y136–P152, and Y136–I191 (Figure 1e).

2.2 | A structurally conserved tryptophan cage stabilizes loop0 connecting N_{ext} to RRM

In the X-ray crystal structure, the 18 aa loop0 that links the N_{ext} to the canonical RRM is highly structured, where primarily hydrophobic residues are coordinated around β_2 residue W192 with extensive van der Waals (vdW) contacts (Figure 2a). The W192 indole ring is positioned between P150–P151 and Q160–P161, has

TABLE 1 Crystallography statistics for *H. sapiens* hnRNPR eRRM1.

Statistics	
Data collection	
Wavelength (Å)	0.72929
Space group	P31 2 1
Cell dimensions	
a, *b*, *c* (Å)	77.023, 77.023, 52.893
α, *β*, *γ* (°)	90.0, 90.0, 120.0
Resolution (Å)	1.90
R_{merge}	0.033 (2.174)
$R_{measure}$	0.035 (2.281)
R_{pim}	0.011 (0.686)
I/σ	40.7 (1.5)
Completeness (%)	100 (99.9)
No. of total reflections	304500
No. of unique reflections	14604
Multiplicity	20.9 (21)
Wilson B factor (Å^2)	49.44
$CC_{1/2}$	1.000 (0.626)
Refinement	
Resolution (Å)	31.13–1.90
No. of reflections	14582
R_{work}/R_{free}	0.1945/0.2221
No. of atoms	1038
Protein	972
Water	65
Average B factors (Å^2)	63.49
Protein	62.49
Water	65.51
Clashscore	1.03
RMSD	
Bond lengths (Å)	0.007
Bond angles (°)	0.95
Ramachandran plot	
Favored (%)	99.20
Allowed (%)	0.80
Outliers (%)	0.00

an edge-to-face interaction with Y156, and contacts D193 (β2) and T212 (β3) on the β-sheet (Figure 2a). This organization bears a striking resemblance to the tryptophan cage motif in the Trp-cage peptide (Neidigh et al. 2002) (Figure 2b). The Trp-cage peptide also has a central tryptophan residue (W6) positioned between two proline residues and has a similar edge-to-face interaction with a tyrosine residue (Y3). Rather than residing on a β-sheet as observed in eRRM1, W6 is in an α-helix and interacts with α-helical residues Y3 and L7. However, the positioning of Y3 and L7 side chains are nearly identical to T212 (β3) and D193 (β2) on eRRM1, respectively, despite the significant secondary structure differences. An X-ray crystal structure reported for the *Drosophila melanogaster* hnRNPQ (Dm-hnRNPQ) eRRM1 domain has near-identical sequence and structural similarity to Hs-hnRNPR in this region (Figure 2c,d) (Hobor et al. 2018). Here, residue W155 in Dm-hnRNPQ replaces residue Y156 in Hs-hnRNPR for maintained edge-to-face interactions with the β2 residue W191 in Dm-hnRNPQ. A valine (V159) is positioned above W191 rather than Q160 in Hs-hnRNPR for maintained vdW contacts with W191 (Figure 2c).

To gain insights into the sequence conservation of the tryptophan cage motif across the hnRNPR-like family of proteins, we compared the Trp-cage peptide sequence to human hnRNPR, hnRNPQ, DND1, RBM46, RBM47, and APOBEC1 complementation factor (A1CF) proteins (Figure 2d,e). In all cases, a Gly-Gly-Pro sequence is found at the transition from the secondary structured element to the loop, promoting a sharp turn in loop0. In addition, a proline is conserved at position 161, P19 in the Trp-cage peptide, positioned above the tryptophan residue. In humans, hnRNPR and hnRNPQ contain a two aa insertion in loop0 compared to DND1, RBM46, RBM47, and A1CF. Notably, residues in loop0 are conserved, including an aromatic residue at position 156. While a glutamine is conserved in hnRNPR and hnRNPQ at position 160, a proline is located at this position in the synthetic Trp-cage and other members of the hnRNPR-like family. Interestingly, DND1 and RBM46 both contain a triple proline repeat sequence at residues 160–162, in common with Trp-cage peptide residues 17–19 (Figure 2e).

To identify sequence conservation across hnRNPR proteins, we curated a list of 10 known and 215 predicted/hypothetical hnRNPR proteins identified in a Uniprot BLAST search (see section 4). Among the 10 known hnRNPR sequences, eRRM1 contains near-exact sequence identity (Figure S5). In particular, residues at the N_{ext}–RRM interface and tryptophan cage in loop0 are identical. Sequence similarity is conserved at positions 154, 215, and 228 as polar; position 159 as hydrophobic; and position 217 as negatively charged. The only dissimilarity occurs at position 187, which is conserved as a lysine across all species except in *Rattus norvegicus*, where it is a glutamic acid. We expanded our sequence analysis to include predicted/hypothetical hnRNPR sequences and observed similarly high conservation in the eRRM1, particularly in the N_{ext} and N_{ext}–RRM interface (Figure S6). Within loop0, the N- and C-terminal ends are highly conserved while central residues at positions 154–157 are variable. In addition, the loop between β1 and α1 is highly variable. Taken together, both the sequence and structure of loop0 are highly conserved as a tryptophan cage motif in the eRRM1. In particular, the tryptophan residue in β2 has a dual role where the backbone participates in N_{ext}–RRM interface interactions, and the indole sidechain

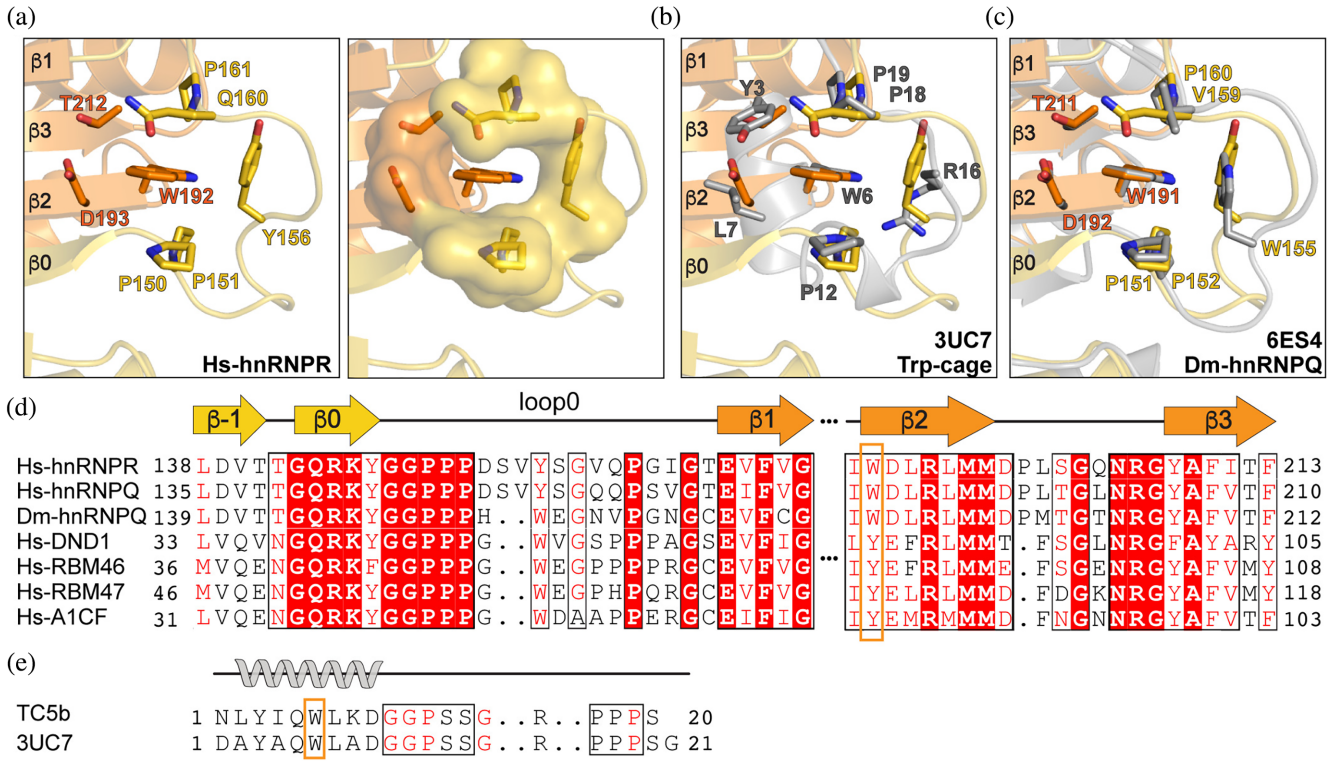

FIGURE 2 A conserved tryptophan cage motif structures the N_{ext} loop0. (a) hnRNPR eRRM1 residues in loop0 and the β-sheet stabilize loop0 Left: sidechains encircling W192 are shown in surface representation. (b) Hs-hnRNPR eRRM1 superimposed with the X-ray crystal structure of the Trp-cage peptide, colored gray (PDB ID 3UC7, chain C) (Scian et al. 2012). (c) Hs-hnRNPR eRRM1 superimposed with the X-ray crystal structure of the Dm-hnRNPQ eRRM1, colored gray (PDB ID 6ES4, chain A) (Hobor et al. 2018). (d) Sequence alignment of eRRM1 domains among human hnRNPR-like proteins (Hs-hnRNPR, Hs-hnRNPQ, Hs-DND1, Hs-RBM46, Hs-RBM47, Hs-A1CF) and experimentally determined eRRM structures (Hs-hnRNPR, Dm-hnRNPQ, Hs-DND1) shows high conservation for residues at the N_{ext}–RRM interface and tryptophan cage. (e) Sequence and secondary structure of representative example Trp-cage peptides. Residues with high sequence similarity to loop0 of hnRNPR-like family proteins are colored red. The central tryptophan residue is indicated in an orange outline.

participates in the tryptophan cage motif. To our knowledge, this study reports the first example of a tryptophan cage motif naturally occurring in a biological protein.

2.3 | Conserved residues at the N_{ext}–RRM interface and loop0 are required for protein folding stability

We next investigated the impact of truncations and point substitutions on eRRM1 chemical environment and thermal stability, evaluated using NMR chemical shift perturbation (CSP) analysis and thermal denaturation studies. The WT eRRM1 construct has a melting temperature (Tm) of 59 ± 1°C and a highly cooperative melting transition (Figures 3a and S2b and Table 2). To determine the importance of the N_{ext} motif in the eRRM1, we generated a construct of the canonical RRM domain lacking N_{ext} (aa 162–246). This construct was primarily observed in the insoluble fraction in *E. coli*, with little soluble expression (Figure S7). We next extended the N-terminus to residue 116 (Ext), which includes the full linker region between the NαB and eRRM1 domains. In the backbone amide ^{1}H-^{15}N HSQC spectrum, the spectra for Ext and WT eRRM1 constructs were nearly identical (Figure S8). Four additional resonances were observed for residues E117–K120 that could be assigned from the ^{15}N-edited NOESY spectrum. CSPs were observed for residues G121–D123 (N-terminus), D175–L176 (loopβ1-α1), and G203–N205 (loopβ2-β3) (Figure S8a–c), consistent with interactions observed in the X-ray crystal structure. This construct showed a modest increase in Tm compared to WT (ΔTm +3.0 ± 1.4°C) (Figures 3a and S8d). Together, these data are consistent with the N-terminal boundary of the eRRM1 beginning at residue 121.

To evaluate the contribution of specific residues at the N_{ext}–RRM interface on protein thermal stability, we next performed alanine substitutions to residues R133 (α0), R145 (β0), Y177 (loopβ1-α1), and R195 (β2) (Figure 3b and Table 2). R133 (α0) is positioned toward helix α1 but does not make direct contacts to the canonical RRM domain in the X-ray crystal structure. Alanine substitution (R133A) shows significant CSPs for N_{ext} α-helical residues E124, R133A, and T134, and

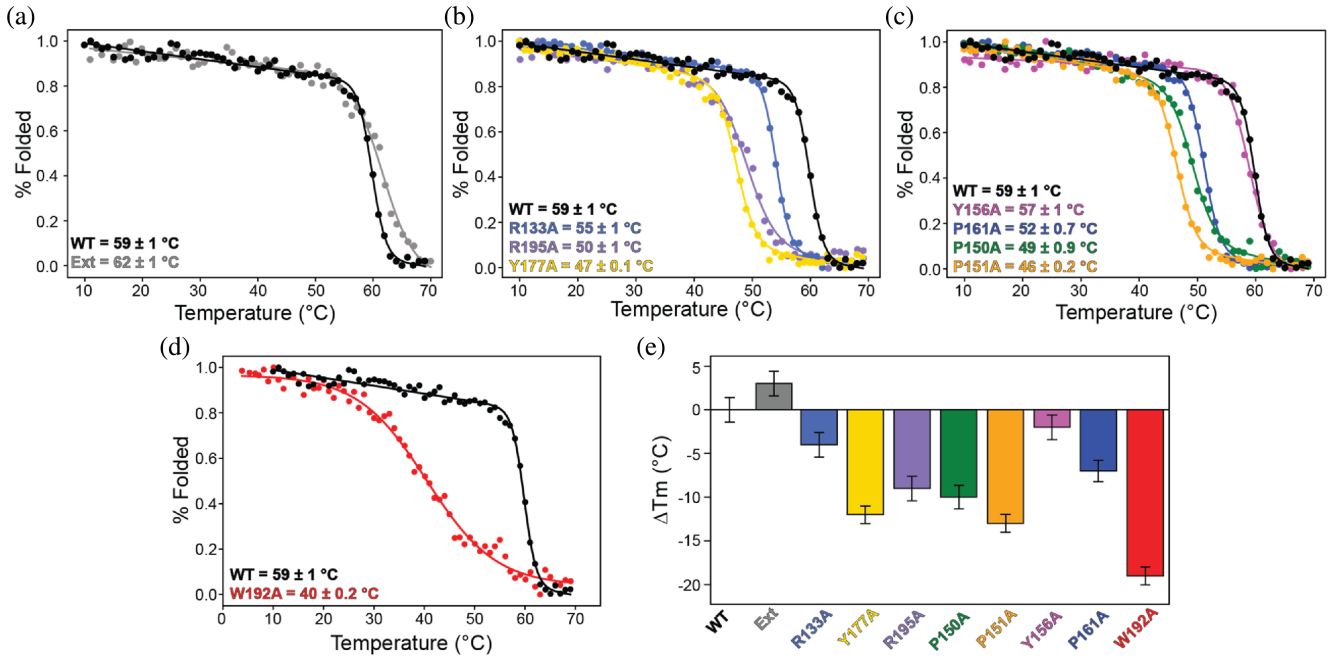

FIGURE 3 Impact of sequence variation on eRRM1 thermal stability using thermal denaturation studies. (a–c) Thermal denaturation plots for (a) WT (colored in black) and Ext (colored in gray) constructs; (b) alanine substitutions to residues at N_{ext}–RRM interface; (c) alanine substitution of loop0 residues in the tryptophan cage motif; (d) W192A construct (colored in red) shows severe reduction in thermal stability; (e) bar plot summarizing melting temperature data for constructs compared to WT eRRM1.

TABLE 2 Thermal melting statistics.

Construct	Protein boundary (aa)	Tm (°C)
WT	121–246	59 ± 1
N-terminal extension and truncation		
Ext	116–246	62 ± 1
RRM1	162–246	ND
Point substitutions at the N_{ext}–RRM1 interface		
R133A	121–246	55 ± 1
R145A	121–246	ND
R195A	121–246	50 ± 1
Y177A	121–246	47 ± 0.1
Point substitutions in loop0		
P150A	121–246	49 ± 0.9
P151A	121–246	46 ± 0.2
Y156A	121–246	57 ± 1
P161A	121–246	52 ± 0.7
W192A	121–246	40 ± 0.2

Note: ND due to insoluble expression.

to a lesser extent RRM residues F185–K187 (α1), G189 and I191 (loopα1-β2), and W192 (β2) (Figure S9a–c). A modest reduction in the Tm (ΔTm −4.0 ± 1.4°C) was observed (Figures 3b and S9d). The R145 (β0) sidechain has extensive interactions that span both N_{ext} and RRM domains including a cation–π interaction with Y177 (loopβ1-α1) and a salt bridge to E178 (α1) (Figure 1e). Consistent with its unique environment, the Hε resonance is significantly downfield shifted (10.8 ppm) relative to other arginine Hε resonances (∼7–8 ppm) in the ^{1}H-^{15}N HSQC spectrum, indicating decreased shielding (Figure S10). R145A does not express in *E. coli*, indicating the importance of these interactions on protein solubility. Alanine substitution of Y177 (Y177A) showed significant CSPs for L176–Y177A, at the substitution site, and to a lesser extent residues E124–Y147 (α0) (Figure S11a–c). Y177A has a significantly reduced Tm (ΔTm −12 ± 1.0°C) (Figures 3b and S11d). The R195 (β2) sidechain is positioned on the β-sheet surface and hydrogen bonds to Q144 (β0), D193 (β2), T212 (β3), E166 (β1), and a coordinated water molecule (Figure S12). Consistent with these interactions, R195A showed CSPs across the β-sheet surface, particularly for residues G143–R145 (β0), Y147 (α0), V167–F168 (β1), and D193–L196 (β2) near the substitution site (Figure S12a–c). R195A has a significantly reduced Tm (ΔTm −9.0 ± 1.4°C) compared to WT (Figures 3b and S12e). Together, these data support the importance of stable association between N_{ext} and RRM subunits for protein solubility and thermal stability.

We next performed alanine substitutions to residues in loop0 to investigate the importance of conserved residues in the tryptophan cage on eRRM1 thermal stability (Figure 3c). A construct substituting Y156 to alanine (Y156A) showed CSPs for residues D153–S154 (loop0), G162–I163 (loop0), and C214 (loopβ3-α2) (Figure S13a–c). A minor reduction in Tm was observed

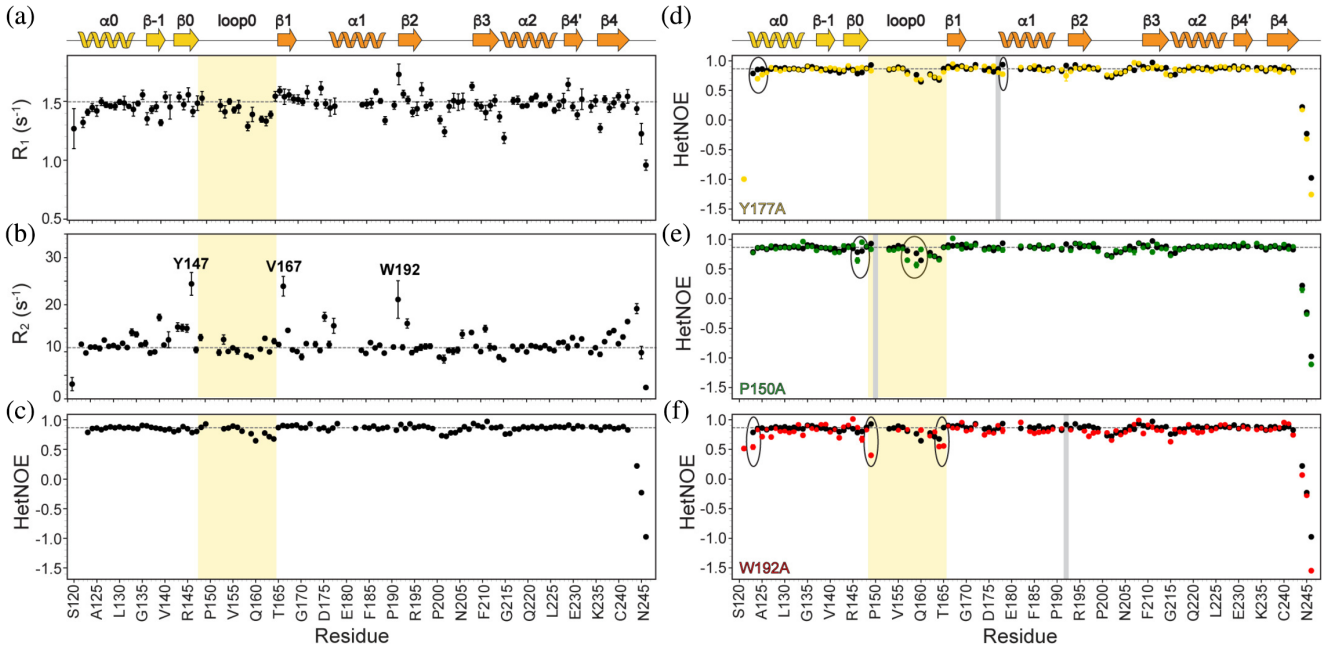

FIGURE 4 hnRNPR eRRM1 conformational dynamics show an ordered protein with flexible ends. (a–c) Plots of WT eRRM1. (a) R_1 relaxation rates; (b) R_2 relaxation rates, with labeled residues indicating sites of chemical exchange; (c) ^{1}H-^{15}N NOE values. (d–f) ^{1}H-^{15}N NOE values for WT eRRM1 (colored in black) and point substitution constructs (d) Y177A (colored in yellow), (e) P150A (colored in green), and (f) W192A (colored in red). The eRRM1 secondary structure topology is depicted above panels (a) and (d). The substitution site is shown as a gray bar, loop0 is highlighted in light yellow, and differences indicated with a circle.

relative to WT (ΔTm −2.0 ± 1.4°C) (Figures 3c and S13d). P161A shows significant CSPs for residues I163–T165 (loop0), T212 (β3), and G215 (β3) (Figure S14a–c), with a modest reduction in Tm (ΔTm −7.0 ± 1.2°C) (Figures 3c and S14d). In the P150A construct, CSPs were observed for residues T137–L138 (β-1), Y147–G149 (loop0), and S157 (loop0) (Figure S15a–c). Similar CSPs were observed for the P151A construct, particularly residues G149, Y156, G158, and Q160 (loop0) (Figure S16a–c). In the P151A construct, residues W192 and L194 also showed significant CSPs. Both P150A and P151A constructs show substantial reductions in Tm (ΔTm −10.0 ± 1.3°C and ΔTm −13 ± 1.0°C, respectively) (Figures 3c, S15d, and S16d).

W192 is located both at the N_{ext}–RRM interface, with backbone interactions to G149 (β0), and in the tryptophan cage as the central residue. Alanine substitution of the central W192 (β2) residue in the tryptophan cage (W192A) dramatically reduced sample stability, resulting in precipitation within 24 h and a severe reduction in thermal stability (ΔTm −19 ± 1.0°C) with reduced melting cooperativity (Figures 3d and S17). Significant CSPs are observed in loop0 and across the β-sheet, particularly in β3 and β1 strands (Figure S17a–c). In summary, all point substitutions that disrupt observed sidechain interactions at the N_{ext}–RRM interface and loop0 result in reduced thermal stability (Figure 3e). Y177A, P151A, and W192A substitutions showed the greatest reductions in thermal stability, indicating their significance in eRRM1 folding through stabilizing the N_{ext}–RRM interface and tryptophan cage motif interactions, respectively. Together, these results show the importance of the stable formation of the N_{ext}–RRM interface and tryptophan cage motif on eRRM1 protein folding.

2.4 | Point substitutions enhance conformational dynamics in loop0 and N_{ext}–RRM interface

We used solution NMR spectroscopy to investigate WT eRRM1 conformational dynamics at fast (ps–ns) timescales. Longitudinal (R_1) and transverse (R_2) relaxation rates were measured for amide nitrogen resonances and ^{1}H-^{15}N heteronuclear NOE values for backbone amide resonances at 25°C (Figure 4a–c). These data together show an overall highly ordered protein with a flexible N-terminal residue G121 and C-terminal residues N245–N246. Similarly, ^{1}H-^{15}N NOE values of the Ext construct gradually reduce approaching the N-terminus, with a minimum value for residue E117, consistent with G121 as the N-terminal boundary of the eRRM1 (Figure S8e). ^{1}H-^{15}N NOE values are slightly elevated for residues K146–G148 (β0) in Ext compared to WT, indicating that the additional N-terminal residues result in increased order in β0 at the N_{ext}–RRM interface. N-terminal residue D123 (α1) has similar values as structured regions, consistent with the hydrophobic contact observed between

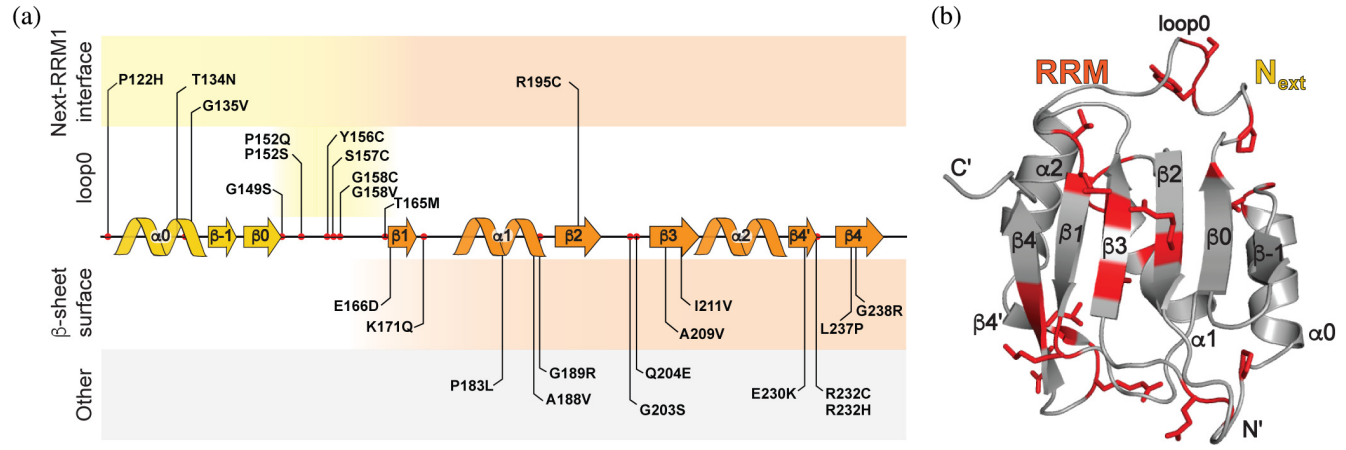

FIGURE 5 Missense mutations in the N_{ext}–RRM interface and loop0 are associated with cancers. (a) COSMIC-annotated missense mutations mapped onto the eRRM1 secondary structure topology. (b) COSMIC-annotated missense mutations (colored in red with sidechains shown in stick representation) mapped onto the X-ray crystal structure of the hnRNPR eRRM1 domain.

P122 and Y177 (α1) in the X-ray crystal structure (Figure 1e). Loop residues V159–G164 (loop0) and L201–Q204 (loopβ2-β3) have reduced values, indicating these loops have fast internal motions. Elevated R_2 values are observed for residues Y147 (β0) and V167 (β1), located at β-strand ends that connect to loop0, as well as residue W192 (β2), located in the tryptophan cage (Figure 4b). These elevated values indicate additional contributions to R_2 from chemical exchange, which may be due to slow motions in the β-strand residues connected to loop0 and at the N_{ext}–RRM interface.

We next compared the conformational dynamics of WT eRRM1 to the point substitution constructs described above to evaluate the impact on eRRM1 conformational dynamics. In the N_{ext}–RRM interface, ^{1}H-^{15}N NOE values of R133A and R195A constructs are nearly identical to WT, indicating no change in fast internal motions (Figures S9e and S12f) despite both constructs having reduced thermal stability compared to WT (Figure 3e). In contrast, in the Y177A construct, ^{1}H-^{15}N NOE values are reduced for N-terminal residues E124–A125 (α0) relative to WT, indicating increased mobility at the N-terminus, particularly in helix α0 (Figures 4d and S18a). These increased dynamics are likely due to destabilizing the hydrophobic contact between P122-Y177A and are consistent with the substantial reduction in thermal stability compared to WT.

We next compared ^{1}H-^{15}N NOE values for loop0 point substitution constructs to evaluate their impact on eRRM1 conformational dynamics. ^{1}H-^{15}N NOE values in the Y156A construct are nearly identical to WT, indicating no change in relative internal motions and consistent with this construct's similar thermal stability to WT (Figures 3e and S13e). In the P161A construct, ^{1}H-^{15}N NOE values are slightly reduced for loop0 residues V159–Q160 and G162 that are adjacent to the substitution site (Figure S14e). Similarly, in both P150A and P151A constructs, ^{1}H-^{15}N NOE values are reduced for loop0 residues S157–Q160 (Figures 4e and S16e), indicating increased dynamics compared to WT. The W192A construct, which has the most deleterious impact on thermal stability (Figure 3e), showed substantially increased dynamics compared to WT (Figure 4f). In particular, ^{1}H-^{15}N NOE values are reduced for residues at the N-terminus, loop0, and β2 indicating increased dynamics at the N_{ext}–RRM interface and loop0 as a result of replacing the central tryptophan indole ring with an alanine methyl group.

Despite the significant reductions observed in the melting temperatures of P150A, P151A, Y177A, and W192A constructs (Figure 3e) the ^{1}H-^{15}N NOE values overall remained similar to WT ^{1}H-^{15}N NOE values (Figures 4d–f and S18) indicating similar order on ps–ns timescales, on which the ^{1}H-^{15}N NOE experiments report. As a control to ensure sufficient equilibration time during the heteronuclear NOE experiment, we repeated these experiments with an incremental delay equal to 10 times the average T_1 value (Figure S18). Nevertheless, the point substitution constructs continued to show similar ^{1}H-^{15}N NOE values when compared to WT, with differences localized to loop0 and the N-terminus. From these data, we conclude that all constructs have overall similar fast timescale motions to WT, with localized increased dynamics at the N-terminus or loop0. Together, these results indicate that the reduced thermal stability observed in point substitution constructs is due to increased dynamics in loop0. These increased dynamics likely reduce the association between N_{ext} and RRM motifs, leading to a reduced melting temperature.

2.5 | Missense mutations in N_{ext} motif are associated with cancer

Prior literature has shown disease-associated mutations in hnRNPR at RRM2 (Wang et al. 2024) and the C-terminal RGG-box repeat (Duijkers et al. 2019;

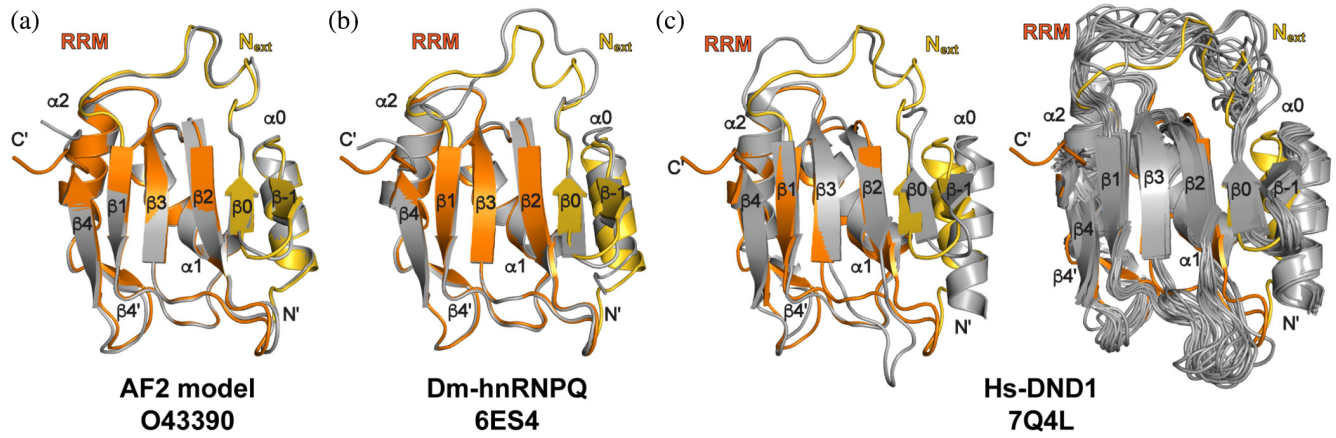

FIGURE 6 Comparison of predicted and high-resolution structures of eRRM1 domains in the hnRNPR-like family. (a) AF2 model (colored in gray) superimposed on X-ray crystal structure of Hs-hnRNPR eRRM1; (b) superimposition of Hs-hnRNPR eRRM1 and Dm-hnRNPQ eRRM1 (colored in gray); (c) superimposition of Hs-hnRNPR eRRM1 and DND1 (colored in gray), with the lowest energy model shown on the left and the structural ensemble shown on the right.

Gillentine et al. 2021). To investigate the biomedical relevance of eRRM1, we surveyed the COSMIC (Catalog of Somatic Mutations in Cancer) database to map sequence variants onto hnRNPR domains (Sondka et al. 2018; Sondka et al. 2024). Of the 513 variant sequences present, there were 172 missense, or single amino acid residue, mutations (Data S2). Missense mutations were present in all domains, with 38.4% in the RGG-box, 16.9% in eRRM1, 16.3% in RRM3, 12.8% in RRM2, 12.2% in NαB, and 3.4% in linker regions (Figure S19a,b). Missense mutations are associated with numerous cancer histology subtypes, most frequently adenocarcinoma (38.3%), not specified (28.8%), and squamous cell carcinoma (11.9%) (Figure S19c). Within the eRRM1 domain, missense mutations are most frequently present in the adenocarcinoma cancer subtype (37.3%) and are found in the large intestine (21.6%) and skin (15.7%) (Figure S19d,e). Of the 27 unique missense mutations in eRRM1, 13 were present in the N_{ext} motif, loop0, and N_{ext}–RRM1 interface residues including at residues P122, T134, G135, G149, P152, Y156–G158, T165–E166, and R195 (Figure 5). Our sequence conservation analysis and mutagenesis studies showed that several of these residues are highly conserved and required for protein thermal stability, suggesting that destabilization of the eRRM1 may have disease implications.

2.6 | Comparison of human hnRNPR eRRM1 to hnRNPR-like protein family

The Hs-hnRNPR eRRM1 X-ray crystal structure is extremely similar to the predicted AF2 model, with an overall backbone RMSD of 0.31 Å (Figure 6a). Comparison to the previously determined X-ray crystal structure of the Dm-hnRNPQ eRRM1 (Hobor et al. 2018) also shows high similarity, with an overall backbone RMSD of 0.50 Å (aa 122–242) (Figure 6b). Differences are observed in loop0, helix α0 orientation, and the positioning of the N_{ext} β-turn (Figure 6b), which may be due to a 2 aa deletion between the triple proline repeat and the aromatic residue in the eRRM1 of Dm-hnRNPQ compared to human hnRNPR and hnRNPQ (Figure 2d).

A solution NMR structural ensemble was reported for the human hnRNPR-like protein DND1 containing tandem eRRM1–RRM2 domains bound to an RNA oligomer (Duszczyk et al. 2022) (Figure S20). The heavy-atom RMSD between the Hs-hnRNPR eRRM1 and the lowest-energy model in the human DND1 (Hs-DND1) eRRM1 structural ensemble is 1.5 Å for all residues (aa 16–134) and 0.99 Å for structured residues in the RRM (Figure 6c). Differences are primarily observed in the N_{ext} motif, particularly helix α0, loop0, and the RRM loopβ2-β3. These differences are likely due to the different method of structure determination and/or the presence of RNA substrate. Inspection of the Hs-DND1 NMR structural ensemble shows higher RMSD among models in loop0 and loopβ2-β3 regions (Figure 6c), suggesting dynamics in these loops. Similarly, our NMR relaxation data support fast dynamics in these loops in the human hnRNPR eRRM1 (Figure 4a–c). From comparison of these examples of eRRMs for representative proteins in the hnRNPR-like family, we identify the required elements for the eRRM fold to be (1) the N_{ext} motif, which has an αββ topology; (2) electrostatic interactions between the N_{ext} and RRM; and (3) the tryptophan cage motif in loop0.

3 | DISCUSSION

hnRNPR and hnRNPR-like proteins perform essential roles in RNA processing. However, biophysical and

structural information on hnRNPR is sparse, limiting a complete understanding of how hnRNPR achieves its function in RNA transcription and splicing regulation. Here, we determined the X-ray crystal structure of the human hnRNPR N-terminal RRM1 and identified an atypical eRRM consisting of a N_{ext} motif that docks alongside a canonical RRM. We evaluated eRRM1 conformational dynamics using NMR spectroscopy, which showed an overall ordered protein with flexible N- and C-termini. However, these dynamics may differ in the full-length hnRNPR, where the N-terminal end is linked to the NαB domain and the C-terminal end is linked to RRM2.

Our work demonstrates that stable association of the N_{ext} motif with the RRM is required for protein solubility and thermal stability. Point substitutions that disrupt N_{ext}–RRM interactions result in increased dynamics at the N-terminus and reduced thermal stability. For example, alanine substitution of Y177, which destabilizes the hydrophobic contact to P122, results in increased dynamics at the N-terminus as well as a substantial reduction in thermal stability compared to WT. Further, truncation of the N_{ext} motif results in a loss of protein solubility. These findings are consistent with previous research that evaluated the impact of N-terminal truncations in the hnRNPR-like protein A1CF. Truncating the first 13 residues at the A1CF N-terminus reduced activity to 33% of full-length A1CF, with a complete loss of activity when truncating two additional residues (Mehta and Driscoll 2002). Our data explain this loss of activity due to truncation of the A1CF eRRM1 domain, which we predict to begin at residue 14 (Figure S1b).

The eRRM1 is further stabilized by a tryptophan cage motif in loop0, which positions the N_{ext} motif for docking to the β2-α1 side of the RRM. The tryptophan cage motif was first identified in the exendin-4 peptide (Neidigh et al. 2001), later modified to the synthetic Trp-cage peptide (Neidigh et al. 2002), and to our knowledge, the eRRM represents the first reported example of a tryptophan cage motif within a naturally occurring protein domain. The central tryptophan residue has a dual role in eRRM1 architecture, also participating in the N_{ext}–RRM interface. Loop0 residues are highly conserved among hnRNPR proteins and hnRNPR-like family proteins. Using combined NMR spectroscopy and thermal denaturation studies, we show that alanine substitutions to loop0 residues lead to increased local dynamics and reduced thermal stability. In particular, P150A, P151A, and W192A substitutions showed the largest reductions in thermal stability compared to WT. While the heteronuclear NOE experiments showed overall similar internal motions compared to WT, reduced values were observed in loop0, indicating increased dynamics at fast timescales in this region. These increased dynamics arising from destabilization of the tryptophan cage motif likely reduce the stable association of the N_{ext} motif to the RRM, leading to reduced thermal stability. Consistent with our results demonstrating the importance of the tryptophan cage motif in eRRM folding, loop0 residues are highly conserved in hnRNPR proteins, as well as among the hnRNPR-like family. A survey of somatic mutations in cancer studies shows cancer-associated mutations in this loop, suggesting that stable formation of the tryptophan cage in loop0 is necessary for cellular function.

RRMs often have additional secondary structure elements that aid in substrate recognition and specificity. Here, the eRRM is a unique example of an extension that is nearly half the size of the RRM and extends the β-sheet surface from five strands to seven strands. The structural basis for hnRNPR eRRM1-RNA recognition remains unknown, and it remains to be seen how the hnRNPR eRRM1 coordinates with the NαB and tandem RRMs to bind RNA and protein substrates. However, a previously determined NMR structure of the human DND1 eRRM1–RRM2 tandem domain bound to RNA provides insights into eRRM1-RNA recognition (Duszczyk et al. 2022). The RNA substrate is sandwiched between eRRM1 and RRM2 domains and binds to the eRRM1 β-sheet surface and loopβ2-β3 (Figure S20). Beyond the canonical interactions with RNP1 and RNP2 sequences, the RNA substrate also interacts with β2 residue R88 (R195 in hnRNPR) and N_{ext} β-turn residues N37 and Q39 (T141 and Q144 in hnRNPR). We anticipate that the hnRNPR eRRM1 may have a similar mode of RNA recognition in which the β2-strand, β2-β3 loop, and N_{ext} β-hairpin contribute to binding. The DND1 eRRM1 does not interact with loop0 or Y85, the equivalent residue to W192 in hnRNPR. The high sequence conservation of the tryptophan cage, along with our mutagenesis studies, suggests a structural role rather than a function in RNA recognition.

A comparison of available examples of eRRMs in the hnRNPR-like family, along with biophysical characterization of the hnRNPR eRRM1, permitted identification of the required elements for the eRRM fold, which were found to be a N_{ext} motif that stably associates with an otherwise canonical RRM and a tryptophan cage motif in the intervening loop0. This study demonstrates the importance of the N_{ext} motif for eRRM folding and solubility, expanding understanding of atypical RRM structure and folding. Our finding that the tryptophan cage motif can be stably inserted within a loop to promote docking of two protein subunits has the potential to be leveraged for protein design applications. In summary, this work provides fundamental biophysical and structural insights into the eRRM1 fold to begin to characterize its role in RNA processing and cellular function.

4 | METHODS

4.1 | Protein construct design and cloning

The amino acid sequence for human hnRNPR protein (Uniprot ID O43390-1) was used to generate the gene sequence optimized for *E. coli* codon bias and purchased as a gBlock (Integrated DNA Technologies, IDT). The gene was cloned into a pET vector containing an N-terminal His-tag, a TEV protease recognition site, and a kanamycin resistance gene. Forward and reverse primers, designed using Benchling and ordered from IDT, were used to clone protein constructs. The plasmids were then transformed into DH5α cells for propagation. Afterward, the plasmids were purified from the DH5α cells using a miniprep plasmid cleanup kit (Zymo Research). Whole Plasmid Sequencing was performed by Plasmidsaurus™ using Oxford Nanopore Technology with custom analysis and annotation. A complete list of protein sequences used in this study is provided in Table S1.

4.2 | Protein expression and purification

Plasmids were transformed into *E. coli* NiCo21(DE3) competent cells (New England Biolabs). Cells were cultured with shaking in LB broth media with 0.05 mg/mL kanamycin at 37°C to an OD600 of 0.6. Culture was transferred to 18°C, expression induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and grown for 18–20 h with shaking. For NMR experiments, cells were cultured in M9 media and supplemented with ^{13}C-labeled glucose (Cambridge Isotope Laboratories) and/or ^{15}N-labeled ammonium chloride (Cambridge Isotope Laboratories). Cells were harvested by centrifugation at 6000*g* for 15 min. Cells were resuspended in resuspension buffer (20 mM Na_2HPO_4/NaH_2PO_4, 1.5M NaCl, 20 mM imidazole, 10% glycerol, 1 mM tris(2-carboxyethyl) phosphine (TCEP), pH 7.0). 0.1 mM PMSF and 1 mg lysozyme were added prior to sonication to lyse cells. After sonication, soluble and insoluble fractions were separated by centrifugation at 36,000*g* for 45 min, and the supernatant was filtered using a 0.45 μm syringe filter. Affinity purification was performed using a nickel nitriloacetic acid (Ni-NTA) affinity column (Qiagen) attached to an AKTA™ start system (Cytiva). After loading the supernatant, the column was washed with 10 column volumes (CV) of resuspension buffer followed by elution in a linear gradient with elution buffer (20 mM Na_2HPO_4/NaH_2PO_4, 300 mM NaCl, 1.0M imidazole, 10% glycerol, 1 mM TCEP, pH 7.0). Fractions were analyzed by SDS-PAGE 4–12% Bis-Tris NuPAGE™ gels (ThermoScientific). To remove the His-tag, protein was dialyzed against storage buffer (20 mM Na_2HPO_4/NaH_2PO_4, 150 mM NaCl, 1 mM TCEP, pH 6.0) in the presence of TEV protease (Addgene #92414, recombinantly expressed and purified in-house) (Raran-Kurussi et al. 2017) for 3–4 h at room temperature or overnight at 4°C. After dialysis, cleaved protein was filtered using a 0.45 μm syringe filter and purified by a second Ni–NTA purification step (Qiagen). As a final purification step, proteins were purified by size exclusion chromatography using a HiLoad 16/600 Superdex 75pg (Cytiva) column attached to an AKTA™ Pure M25 FPLC system (Cytiva) in storage buffer. Fractions were analyzed by 4–12% Bis-Tris NuPAGE™ gel and inspection A260/A280 ratios to identify fractions with pure protein. Protein concentration was determined from the absorbance measured at 280 nm using a NanoDrop One™ spectrophotometer (Thermo Fisher Scientific) using Beer's law. Pure fractions were concentrated to 0.1–0.8 mM using a 3 kDa Amicon™. Expasy ProtParam (Wilkins et al. 1999) was used to compute protein extinction coefficients, molecular weights, and theoretical pI values.

4.3 | X-ray crystallography

Proteins were crystalized using the hanging drop vapor diffusion method. Proteins were prepared in crystallization buffer (20 mM HEPES, 150 mM NaCl, 1 mM TCEP pH = 7.0) at 16 mg/mL. Rod-like crystals appeared within 48 h and were harvested at 200–300 μm in size. Crystals formed under reservoir conditions of 0.04M monobasic potassium phosphate, 20% (v/v) glycerol, and 8% (w/v) polyethylene glycol 8000 with a 1:1 protein-to-reservoir ratio. Data was collected at 100 K at the Stanford Synchrotron Radiation Lightsource at the SLAC National Accelerator Laboratory on beamline 12-2 using 0.72929 Å X-rays with a Dectris Pilatus 6M detector. Data from a single crystal was used for structure determination. Data was indexed and integrated using XDS (Kabsch 2010). Data reduction was performed using Aimless, Pointless, and Ctruncate in the CCP4 suite (Winn et al. 2011). Phases were determined using molecular replacement against the X-ray crystal structure of the drosophila hnRNPQ eRRM (PDB ID 6ES4) (Hobor et al. 2018). Model building and refinement was performed using Coot version 0.9.8.94 (Emsley et al. 2010) and PHENIX version 1.21.1 (Liebschner et al. 2019) with TLS refinement (Painter and Merritt 2006). Final data collection, phasing, and refinement statistics are reported in Table 1.

4.4 | NMR spectroscopy

NMR experiments were performed at 25°C on a 600 MHz Bruker spectrometer equipped with a triple-resonance HCN cryoprobe and Avance Neo console. NMR samples were prepared in storage buffer with added 5% D_2O at 0.3–0.8 mM concentrations in 3 or 5 mm NMR tubes (Norell). Backbone (N, H, Cα, Cβ,

CO) resonance assignments were performed using standard triple resonance experiments (Cavanagh 2007; Reid et al. 1997). Data were processed using NMRPipe (Shen et al. 2009) and analyzed using NMRFAM-Sparky 1.470 (Lee et al. 2015) in the NMRbox virtual machine (Maciejewski et al. 2017).

^{1}H-^{15}N heteronuclear nuclear Overhauser effect (NOE) experiments (hsqcnoef3gpsi) from the Bruker experimental suite were recorded in an interleaved manner with 32 scans and 2 s incremental delay for 0.5 mM protein samples. For WT, P150A, Y177A, and W192A constructs heteronuclear NOE experiments were also performed with a 7 s incremental delay. The heteronuclear NOE is reported as the residue-specific ratio of peak intensity between the saturated and unsaturated experiments (I_{sat}/I_{unsat}). Error was estimated as the standard deviation of noise in the saturated (σI_{sat}) and unsaturated (σI_{unsat}) experiments (Farrow et al. 1994; Metcalfe et al. 2004): $\frac{\sigma_{NOE}}{NOE} = \sqrt{\frac{\sigma I_{sat}}{I_{sat}}^2 + \frac{\sigma I_{unsat}}{I_{unsat}}^2}$. R_1 and R_2 values were obtained from T_1 (hsqct1etf3gpsi3d) and T_2 (hsqct2etf3gpsi3d) relaxation experiments with 32 scans and 2 s incremental delay. Data was processed in NMRPipe (Shen et al. 2009) and relaxation rates were calculated using the Function and Data Analysis (FuDA) software package (Hansen et al. 2007). For T_1 experiments, delays were 20 ms, 60 ms × 2, 200 ms, 400 ms, 800 ms × 2, and 1200 ms. T_2 loops (2 × 2, 4, 6, 8, 12 × 2, 16) had a 0.01696 ms variable delay. Weighted average chemical shift perturbations were calculated using the equation $\sqrt{\Delta H^2 + 0.1\Delta N^2}$ (Cavanagh 2007).

4.5 | Circular dichroism

Thermal denaturation experiments were performed on a Jasco815 spectrometer equipped with a Peltier temperature control device. Protein samples were prepared in storage buffer at concentrations ranging between 18 and 25 μM. CD spectra were collected at 222 nm with the following parameters: temperature range of 10–90°C (except for the W192A construct, where the temperature range used was 4–80°C), 1 s hold time, 1.0°C/s ramp rate, 1°C sampling interval, 5 s wait time, 1 s data integration time, 2 nm band width, and standard sensitivity. Data were analyzed using in-house python code (https://github.com/eichhorn-lab). Due to protein aggregation at temperatures above 70°C, data from 10 to 70°C (except for W192A, which included data from 4 to 70°C), were normalized using the equation (data point − min value)(max value − min value) to compute percent folded protein. The data was fitted with a logistic function of $\frac{1}{1+e^{\frac{T-Tm}{k}}}$, where Tm is the melting temperature and k is the slope factor. Independent experiments were performed in duplicate. A complete list of fitted parameters and replicates for each construct is provided in Table S2.

4.6 | Bioinformatics

The human hnRNPR sequence (O43390-1) was retrieved from the UniProt database (UniProt 2024) and used as a query in UniProt BLAST (Altschul et al. 1990) with default search parameters. Of 250 initial hits, 25 were excluded to remove redundant entries, achieved sequences, and multiple isoforms from the same species. The selected homologous sequences were then aligned using the EMBL Clustal Omega Multiple Sequence Alignment (MSA) tool (Goujon et al. 2010; Sievers et al. 2011). The alignment quality was visually inspected to ensure accurate residue mapping. To assess the evolutionary conservation of individual residues, the MSA of the full-length hnRNPR was analyzed using the ConSurf web server (Ashkenazy et al. 2016; Yariv et al. 2023) with default parameters. Conservation scores were mapped onto the AF2 predicted model (AF-O43390-F1). For cancer-related missense mutation analysis, data was downloaded from the COSMIC database (Sondka et al. 2018; Sondka et al. 2024) (Data S1). An in-house python script was used to extract missense mutations, sort data by cancer histology subtype, primary tissue, and domain location, and generate plots (https://github.com/eichhorn-lab).

AUTHOR CONTRIBUTIONS

Ernest S. Atsrim: Writing – review and editing; formal analysis; investigation; visualization; validation. **Catherine D. Eichhorn:** Conceptualization; investigation; funding acquisition; writing – original draft; writing – review and editing; formal analysis; supervision; visualization.

ACKNOWLEDGMENTS

We acknowledge funding support from the National Institutes of Health (1R35GM143030) and the Nebraska Center for Integrated Biomolecular Communication (P20 GM113126). Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (P30GM133894). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. We acknowledge assistance

from Linh Hua in buffer preparation. We thank all members of the Eichhorn lab for critical reading of the manuscript and providing constructive feedback.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

PDB coordinates are deposited to the PDB with ID 9EJY. NMR resonance assignments, R_1 and R_2 relaxation, and ^{1}H-^{15}N NOE data are deposited to BMRB ID 52734. Python scripts are deposited to github (https://github.com/eichhorn-lab). The data that support the findings of this study are openly available in Human hnRNPR extended eRRM1 domain at https://www.rcsb.org, reference number 9EJY.

ORCID

Catherine D. Eichhorn https://orcid.org/0000-0001-8624-1961

REFERENCES

Agbo J, Akinyemi AR, Li D, Zhang Q, Li X, Li W, et al. RNA-binding protein hnRNPR reduces neuronal cholesterol levels by binding to and suppressing HMGCR. J Integr Neurosci. 2021;20:265–76.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016;44:W344–50.

Briese M, Saal-Bauernschubert L, Ji C, Moradi M, Ghanawi H, Uhl M, et al. hnRNP R and its main interactor, the noncoding RNA 7SK, coregulate the axonal transcriptome of motoneurons. Proc Natl Acad Sci U S A. 2018;115:E2859–68.

Briese M, Sendtner M. Keeping the balance: the noncoding RNA 7SK as a master regulator for neuron development and function. Bioessays. 2021;43:e2100092.

Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. Science. 1994;265:615–21.

Cavanagh J. Protein NMR spectroscopy: principles and practice. Amsterdam, Boston: Academic Press; 2007.

Chen EB, Qin X, Peng K, Li Q, Tang C, Wei YC, et al. HnRNPR-CCNB1/CENPF axis contributes to gastric cancer proliferation and metastasis. Aging (Albany NY). 2019;11:7473–91.

Daubner GM, Clery A, Allain FH. RRM-RNA recognition: NMR or crystallography… and new findings. Curr Opin Struct Biol. 2013;23:100–8.

Dreyfuss G, Matunis MJ, Pinol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem. 1993;62:289–321.

Duijkers FA, McDonald A, Janssens GE, Lezzerini M, Jongejan A, van Koningsbruggen S, et al. HNRNPR variants that impair homeobox gene expression drive developmental disorders in humans. Am J Hum Genet. 2019;104:1040–59.

Duszczyk MM, Wischnewski H, Kazeeva T, Arora R, Loughlin FE, von Schroetter C, et al. The solution structure of dead end bound to AU-rich RNA reveals an unusual mode of tandem RRM-RNA recognition required for mRNA regulation. Nat Commun. 2022;13:5892.

Eichhorn CD, Yang Y, Repeta L, Feigon J. Structural basis for recognition of human 7SK long noncoding RNA by the La-related protein Larp7. Proc Natl Acad Sci U S A. 2018;115:E6457–66.

Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of coot. Acta Crystallogr D Biol Crystallogr. 2010;66:486–501.

Farrow NA, Zhang O, Forman-Kay JD, Kay LE. A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. J Biomol NMR. 1994;4:727–34.

Fukuda A, Nakadai T, Shimada M, Hisatake K. Heterogeneous nuclear ribonucleoprotein R enhances transcription from the naturally configured c-fos promoter in vitro. J Biol Chem. 2009;284:23472–80.

Gascon E, Zaragoza P, Calvo AC, Osta R. Sporadic amyotrophic lateral sclerosis skeletal muscle transcriptome analysis: a comprehensive examination of differentially expressed genes. Biomolecules. 2024;14:377.

Ghanawi H, Hennlein L, Zare A, Bader J, Salehi S, Hornburg D, et al. Loss of full-length hnRNP R isoform impairs DNA damage response in motoneurons by inhibiting Yb1 recruitment to chromatin. Nucleic Acids Res. 2021;49:12284–305.

Gillentine MA, Wang T, Hoekzema K, Rosenfeld J, Liu P, Guo H, et al. Rare deleterious mutations of HNRNP genes result in shared neurodevelopmental disorders. Genome Med. 2021;13:63.

Glinka M, Herrmann T, Funk N, Havlicek S, Rossoll W, Winkler C, et al. The heterogeneous nuclear ribonucleoprotein-R is necessary for axonal beta-actin mRNA translocation in spinal motor neurons. Hum Mol Genet. 2010;19:1951–66.

Gopan G, Ghaemi Z, Davis CM, Gruebele M. Spliceosomal SL1 RNA binding to U1-70K: the role of the extended RRM. Nucleic Acids Res. 2022;50:8193–206.

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 2010;38:W695–9.

Han SP, Tang YH, Smith R. Functional diversity of the hnRNPs: past, present and perspectives. Biochem J. 2010;430:379–92.

Hansen DF, Yang D, Feng H, Zhou Z, Wiesner S, Bai Y, et al. An exchange-free measure of 15N transverse relaxation: an NMR spectroscopy application to the study of a folding intermediate with pervasive chemical exchange. J Am Chem Soc. 2007;129:11468–79.

Hobor F, Dallmann A, Ball NJ, Cicchini C, Battistelli C, Ogrodowicz RW, et al. A cryptic RNA-binding domain mediates Syncrip recognition and exosomal partitioning of miRNA targets. Nat Commun. 2018;9:831.

Hu T, Zeng C, Song Z, Liu Q, Chen S, Huang W, et al. HNRNPA2B1 and HNRNPR stabilize ASCL1 in an m6A-dependent manner to promote neuroblastoma progression. Biochim Biophys Acta Mol Basis Dis. 2024;1870:167050.

Huang J, Chen XH, Wu K, Xu P. Cloning and expression of a novel isoform of heterogeneous nuclear ribonucleoprotein-R. Neuroreport. 2005;16:727–30.

Ji C, Bader J, Ramanathan P, Hennlein L, Meissner F, Jablonka S, et al. Interaction of 7SK with the Smn complex modulates snRNP production. Nat Commun. 2021;12:1278.

Ji C, Deng C, Antor K, Bischler T, Schneider C, Fischer U, et al. hnRNP R negatively regulates transcription by modulating the association of P-TEFb with 7SK and BRD4. EMBO Rep. 2022;23:e55432.

Jiang T, Qu R, Liu X, Hou Y, Wang L, Hua Y. HnRNPR strongly represses splicing of a critical exon associated with spinal muscular atrophy through binding to an exonic AU-rich element. J Med Genet. 2023;60:1105–15.

Kabat JL, Barberan-Soler S, Zahler AM. HRP-2, the Caenorhabditis elegans homolog of mammalian heterogeneous nuclear ribonucleoproteins Q and R, is an alternative splicing factor that binds to UCUAUC splicing regulatory elements. J Biol Chem. 2009;284:28490–7.

Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010;66:125–32.

Lau PP, Chang BH, Chan L. Two-hybrid cloning identifies an RNA-binding protein, GRY-RBP, as a component of apobec-1 editosome. Biochem Biophys Res Commun. 2001;282:977–83.
Lee W, Tonelli M, Markley JL. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics. 2015; 31:1325–7.
Li Y, Wang H, Wan J, Ma Q, Qi Y, Gu Z. The hnRNPK/A1/R/U complex regulates gene transcription and translation and is a favorable prognostic biomarker for human colorectal adenocarcinoma. Front Oncol. 2022;12:845931.
Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Biol Crystallogr. 2019;75:861–77.
Maciejewski MW, Schuyler AD, Gryk MR, Moraru II, Romero PR, Ulrich EL, et al. NMRbox: a resource for biomolecular NMR computation. Biophys J. 2017;112:1529–34.
Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J. 2005;272:2118–31.
McWhorter ML, Monani UR, Burghes AH, Beattie CE. Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding. J Cell Biol. 2003;162:919–31.
Mehta A, Driscoll DM. Identification of domains in apobec-1 complementation factor required for RNA binding and apolipoprotein-B mRNA editing. RNA. 2002;8:69–82.
Metcalfe EE, Zamoon J, Thomas DD, Veglia G. (1)H/(15)N heteronuclear NMR spectroscopy shows four dynamic domains for phospholamban reconstituted in dodecylphosphocholine micelles. Biophys J. 2004;87:1205–14.
Meuzelaar H, Marino KA, Huerta-Viga A, Panman MR, Smeenk LE, Kettelarij AJ, et al. Folding dynamics of the Trp-cage miniprotein: evidence for a native-like intermediate from combined time-resolved vibrational spectroscopy and molecular dynamics simulations. J Phys Chem B. 2013;117:11490–501.
Neidigh JW, Fesinmeyer RM, Andersen NH. Designing a 20-residue protein. Nat Struct Biol. 2002;9:425–30.
Neidigh JW, Fesinmeyer RM, Prickett KS, Andersen NH. Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. Biochemistry. 2001;40: 13188–200.
Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr. 2006;62:439–50.
Peng ZY, Huang J, Lee SC, Shi YL, Chen XH, Xu P. The expression pattern of heterogeneous nuclear ribonucleoprotein R in rat retina. Neurochem Res. 2009;34:1083–8.
Raran-Kurussi S, Cherry S, Zhang D, Waugh DS. Removal of affinity tags with TEV protease. Methods Mol Biol. 2017; 1586:221–30.
Reches A, Nachmani D, Berhani O, Duev-Cohen A, Shreibman D, Ophir Y, et al. HNRNPR regulates the expression of classical and nonclassical MHC class I proteins. J Immunol. 2016;196: 4967–76.
Reid DG, MacLachlan LK, Edwards AJ, Hubbard JA, Sweeney PJ. Introduction to the NMR of proteins. Methods Mol Biol. 1997;60: 1–28.
Rossoll W, Jablonka S, Andreassi C, Kroning AK, Karle K, Monani UR, et al. Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of beta-actin mRNA in growth cones of motoneurons. J Cell Biol. 2003; 163:801–12.
Rossoll W, Kroning AK, Ohndorf UM, Steegborn C, Jablonka S, Sendtner M. Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp/hnRNP-Q: a role for Smn in RNA processing in motor axons? Hum Mol Genet. 2002;11:93–105.
Scian M, Lin JC, Le Trong I, Makhatadze GI, Stenkamp RE, Andersen NH. Crystal and NMR structures of a Trp-cage miniprotein benchmark for computational fold prediction. Proc Natl Acad Sci U S A. 2012;109:12521–5.
Shen Y, Bax A. Protein structural information derived from NMR chemical shift with the neural network program TALOS-N. Methods Mol Biol. 2015;1260:17–32.
Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR. 2009;44:213–23.
Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.
Singh M, Wang Z, Koo BK, Patel A, Cascio D, Collins K, et al. Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. Mol Cell. 2012;47: 16–26.
Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. Nat Rev Cancer. 2018;18: 696–705.
Sondka Z, Dhir NB, Carvalho-Silva D, Jupe S, Madhumita, McLaren K, et al. COSMIC: a curated database of somatic variants and clinical data for cancer. Nucleic Acids Res. 2024;52: D1210–7.
UniProt Consortium. UniProt: the universal protein knowledgebase in 2025. Nucleic Acids Res. 2024;53:D609–17.
Wang H, Qian D, Wang J, Liu Y, Luo W, Zhang H, et al. HnRNPR-mediated UPF3B mRNA splicing drives hepatocellular carcinoma metastasis. J Adv Res. 2024;68:257–70.
Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999;112:531–52.
Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 2011;67:235–42.
Yang Y, Sun JD, Xiang ZL. Pan-cancer analysis of the oncogenic role of HNRNPR in human tumors. Cell Biol Int. 2023;47:1406–26.
Yariv B, Yariv E, Kessel A, Masrati G, Chorin AB, Martz E, et al. Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf. Protein Sci. 2023;32:e4582.
Zare A, Salehi S, Bader J, Schneider C, Fischer U, Veh A, et al. hnRNP R promotes O-GlcNAcylation of eIF4G and facilitates axonal protein synthesis. Nat Commun. 2024;15:7430.
Zhou R. Trp-cage: folding free energy landscape in explicit water. Proc Natl Acad Sci U S A. 2003;100:13280–5.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Atsrim ES, Eichhorn CD. An evolutionarily conserved tryptophan cage promotes folding of the extended RNA recognition motif in the hnRNPR-like protein family. Protein Science. 2025;34(5):e70127. https://doi.org/10.1002/pro.70127