Title

Rational design yields RNA-binding zinc finger domains with altered sequence specificity.

Journal

RNA, 31(2)

Authors

Liang, Qishan

Xiang, Joy

Yeo, Gene

et al.

Publication Date

2025-01-22

DOI

10.1261/rna.080329.124

Peer reviewed

Rational design yields RNA-binding zinc finger domains with altered sequence specificity

QISHAN LIANG,[1,2,3,4] JOY S. XIANG,[2] GENE W. YEO,[2,3,4] and KEVIN D. CORBETT[2,4,5]

[1]Department of Chemistry and Biochemistry, UC San Diego, La Jolla, California 92093, USA
[2]Department of Cellular and Molecular Medicine, UC San Diego, La Jolla, California 92093, USA
[3]Sanford Stem Cell Institute, UC San Diego, La Jolla, California 92093, USA
[4]Center for RNA Technologies and Therapeutics, UC San Diego, La Jolla, California 92093, USA
[5]Department of Molecular Biology, UC San Diego, La Jolla, California 92093, USA

ABSTRACT

Targeting and manipulating endogenous RNAs in a sequence-specific manner is essential for both understanding RNA biology and developing RNA-targeting therapeutics. RNA-binding zinc fingers (ZnFs) are excellent candidates as designer proteins to expand the RNA-targeting toolbox, due to their compact size and modular sequence recognition. Currently, little is known about how the sequence of RNA-binding ZnF domains governs their binding site specificity. Here, we systematically introduced mutations at the RNA-contacting residues of a well-characterized RNA-binding ZnF protein, ZRANB2, and measured RNA binding of mutant ZnFs using a modified RNA bind-n-seq assay. We identified mutant ZnFs with an altered sequence specificity, preferring to bind a GGG motif instead of the GGU preferred by wild-type ZRANB2. Further, through a series of all-atom molecular dynamics simulations with ZRANB2 and RNA, we characterized changes in the hydrogen-bond network between the protein and RNA that underlie the observed sequence specificity changes. Our analysis of ZRANB2–RNA interactions both in vitro and in silico expands the understanding of ZnF-RNA recognition rules and serves as a foundation for eventual use of RNA-binding ZnFs for programmable RNA targeting.

Keywords: ZRANB2; RNA-binding zinc finger; rational design; RBNS

INTRODUCTION

Proper regulation of RNA processing, localization, translation, and degradation is critical for the health of all cells, and defects in any of these processes can lead to disease (Gerstberger et al. 2014). As such, manipulation of endogenous RNAs in a sequence-specific manner represents an important strategy for both fundamental research into RNA-related pathways, and for the eventual development of RNA-targeting therapeutics. Existing sequence-specific RNA-targeting tools include antisense oligos (ASOs), designer RNA-binding proteins of the PUF (Pumilio and FBF) protein family, and CRISPR–Cas proteins. ASOs target RNA transcripts through complementary base-pairing and, as such, are highly specific and readily programmable (Lauffer et al. 2024). Designer proteins like PUF are versatile in diverse RNA manipulations through fusion with different protein effectors (Zhao et al. 2018; Sugimoto et al. 2020). RNA-targeting CRISPR–Cas proteins combine these advantages, harnessing a guide RNA for specific recognition through complementary base-pairing, and enabling targeting of functional domains for purposes like RNA degradation and modification (Konermann et al. 2018; Lau and Suh 2018). While each of these tools enables RNA targeting in unique ways, other drawbacks limit their use in laboratory and/or clinical settings. For example, ASOs are difficult to deliver specifically to affected tissues and have to be frequently redosed (Lauffer et al. 2024); PUF proteins show limited RNA sequence specificity and cannot efficiently target RNA sequences over 8 nt in length (Zhao et al. 2018; Sugimoto et al. 2020); and CRISPR–Cas proteins are large and therefore challenging to deliver as transgenes. In addition, the bacterial origin of CRISPR–Cas proteins introduces concerns about their potential immunogenicity when used in a clinical setting (Konermann et al. 2018; Lau and Suh 2018).

One class of proteins with the potential to overcome the challenges of existing RNA-targeting tools is RNA-binding

Corresponding authors: geneyeo@ucsd.edu, kcorbett@ucsd.edu
Handling editor: Fatima Gebauer
Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.080329.124. Freely available online through the *RNA* Open Access option.

zinc finger proteins (ZnFs). Modular arrays of DNA-binding ZnFs have been used for programmable, sequence-specific recognition of DNA for over two decades (Isalan and Choo 2001; Pabo et al. 2001; Beerli and Barbas 2002). More recently, a class of RNA-binding ZnF proteins has been identified (Nguyen et al. 2011), whose individual ZnF domains (each ~3 kDa) specifically bind three-base single-stranded RNA (ssRNA) motifs and can be assembled into arrays to target longer sequences. For example, the two ZnF domains of *Homo sapiens* ZRANB2 (ZnF Ran-binding domain-containing protein 2) each bind a core GGU motif with micromolar affinity (Nguyen et al. 2011), and an artificial array of three ZRANB2 ZnFs binds 5′-GGUGGUGGU-3′ with a dissociation constant (K_d) of ~10 nM (O'Connell et al. 2012). Further, a 6-ZnF array was shown to recognize a 20 nt RNA sequence in a modular manner both in vitro and in living cells (De Franco et al. 2019). These data suggest that modular arrays of RNA-binding ZnFs could represent an RNA-targeting tool with distinct advantages: (1) compactness: only four ZnFs (~100 amino acids total) may be required for site specificity in the human transcriptome (12 RNA bases); (2) minimal immunogenicity: ZnF arrays could be assembled from individual domains with human origins.

While RNA-binding ZnFs hold promise as programmable sequence-specific targeting reagents, the number of ZnFs with defined target sequences is extremely limited. Here, we undertook a rational design approach to develop mutants of a single ZRANB2 ZnF domain with altered RNA sequence specificity. Using a modified RNA bind-n-seq (RBNS) approach (Lambert et al. 2014; Dominguez et al. 2018), we systematically tested a total of 38 single and double mutants of a ZRANB2 ZnF domain for RNA-binding affinity and sequence specificity. We identified one single mutant that retained high RNA-binding affinity but showed an altered sequence specificity, preferring GGG rather than the GGU targeted by the wild-type ZnF. Using a series of molecular dynamics (MD) simulations of the ZnF-RNA complex, we find that this mutant adopts a distinctive conformation and hydrogen bonding pattern that accounts for its altered sequence specificity. Overall, our engineering and characterization of ZRANB2 ZnF-RNA binding enhances the understanding of ZnF-RNA recognition principles and serves as a basis for the engineering of compact, designer RNA-targeting proteins.

RESULTS

RBNS characterizes the RNA-binding profile of ZnFs

The *H. sapiens* alternative splicing regulator ZRANB2 is a member of the broad SR protein family, which typically contains at least one RNA recognition motif (RRM) and a serine/arginine (SR) rich domain implicated in protein–protein interactions (Fig. 1A; Loughlin et al. 2009; Shepard and Hertel 2009). ZRANB2 encodes two RanBP2-type ZnFs (here termed N′ and C′) at its N terminus, which are 48% identical to one another at the amino acid level, and both recognize ssRNA with a 5′-AGGUAA-3′ motif, where "GGU" is the core 3-mer motif (Fig. 1B; Loughlin et al. 2009). To establish an experimental framework for rational design and testing of RNA-targeting ZnFs, we designed a modular 2-ZnF array with one copy of the ZRANB2 N′-ZnF (the "anchor" ZnF) followed by a short linker (amino acid sequence: GSGSG) and a second ZnF (the "test" ZnF; Fig. 1C). In combination with a modified RBNS assay (described below), this experimental design enables rapid assembly and testing of diverse test ZnFs for RNA-binding affinity and specificity. We generated an initial panel of 2-ZnF arrays with test ZnFs from ZRANB2 (N′ and C′, both of which recognize GGU) and a second RBP, RBM5, whose ZnF recognizes GGG (Nguyen et al. 2011; Soni et al. 2023). As a negative control, we also generated a mutant of the ZRANB2 C′-ZnF with a key tryptophan residue mutated to alanine (W79A); this residue interacts with the two guanine residues of the core GGU motif through pi-stacking (Loughlin et al. 2009), and its mutation to alanine is expected to abolish RNA binding.

We designed a modified high-throughput RBNS assay (Lambert et al. 2014; Dominguez et al. 2018) to assess the RNA-binding characteristics of the recombinant 2-ZnF arrays (Fig. 1C). In RBNS, a randomized pool of RNAs is affinity-purified by an immobilized RNA-binding protein, then the purified RNA pool is deep-sequenced and compared to the input RNA pool to calculate the fold-enrichment of each sequence motif. By performing RBNS at a range of protein concentrations, an estimate of binding affinity for each sequence motif can be determined, in addition to the identification of preferred binding motifs (Lambert et al. 2014). We designed an RNA pool with the sequence 5′-AAAGGUNNNNNNAAA-3′, where "N" represents any base (A/U/G/C) mixed at equal frequencies in the library. The leading GGU motif serves as a binding site for the anchor ZnF, allowing the test ZnF to bind preferred sequences within the six-base random window. Native ZRANB2 has a 23-residue disordered linker between its N′ and C′ ZnFs, and the protein recognizes tandem GGU motifs with a spacing of two to five bases between motifs (Loughlin et al. 2009); we anticipated that the shorter five-residue linker in our 2-ZnF arrays would enable recognition of more closely spaced motifs. The six-base random window in our RNA pool enables the test ZnF to selectively bind RNAs with preferred motifs spaced between zero and three bases from the anchor ZnF's GGU motif.

We performed RBNS with our initial panel of 2-ZnF arrays at four different protein concentrations (160, 320, 640, and 1280 nM), deep-sequenced the input and purified RNA pools, and performed two analyses: first, we calculated the fold-enrichment for each six-base sequence by

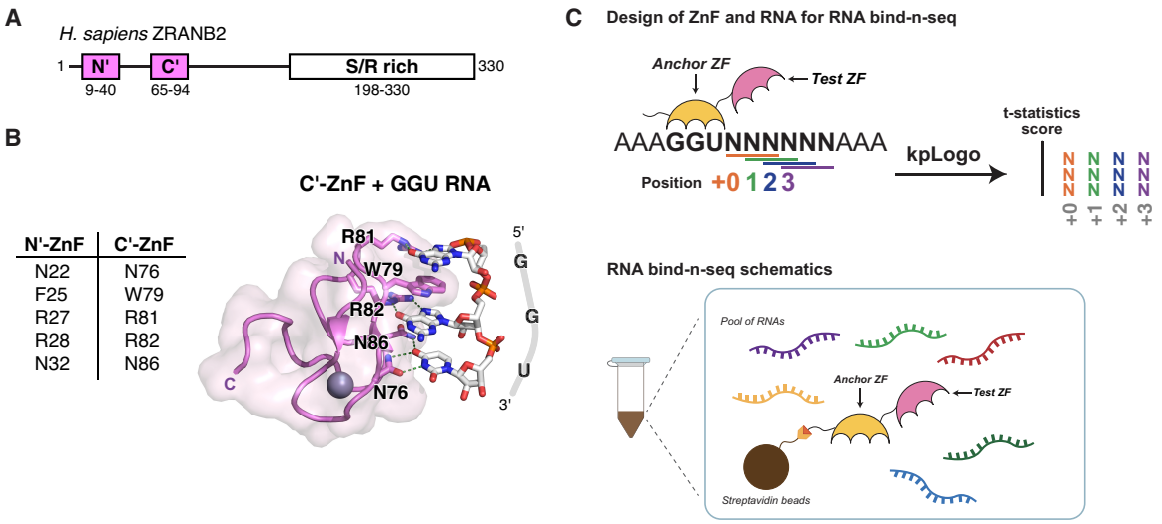

FIGURE 1. Design of ZRANB2 ZnF mutant and RBNS assay. (*A*) Domains of human ZRANB2 protein. (*B*) Structure of human ZRANB2 C′-ZnF in complex with GGU RNA (PDB: 3g9y) and the homology between N′-ZnF and C′-ZnF. (*C*) Schematics of the modified RBNS assay with the ZnF array.

comparing input to purified RNA pools. Second, we used kpLogo (Wu and Bartel 2017) to search for enriched 3-mer motifs at each position (+0, +1, +2, and +3) within the six-base random window. We designed the protein concentrations in the RBNS to be at or slightly above the estimated affinity between wild-type ZRANB2-ZnF and a GGU RNA (K_d ~100 nM) (Loughlin et al. 2009), to better capture the RNA-binding specificity of the mutants due to the potential loss of RNA binding. In our initial panel of 2-ZnF arrays, we found that the ZRANB2 N′-ZnF, ZRANB2 C′-ZnF, and RBM5 ZnF all showed high selectivity (specificity) at the lowest tested protein concentration (160 nM), with the most-enriched sequences being 100 to 200-fold enriched over input levels (Fig. 2A–C). Fold-enrichments for preferred sequences were lower at higher protein concentrations, indicating that at these concentrations the protein nonspecifically binds a wide range of sequences. The *k*-means clustering analysis revealed distinct clusters among sequences, each characterized by similar patterns of the fold-enrichment values at four different protein concentrations (Supplemental Fig. S1A). The most highly enriched sequences for the two ZRANB2 ZnFs showed a preponderance of sequences containing GGU motifs within the six-base random window, while the RBM5 ZnF showed enrichment of sequences containing GGG motifs (Supplemental Fig. S1B). In agreement with these findings, kpLogo analysis with both ZRANB2 ZnFs showed GGU as the most preferred motif at all positions (+0 to +3), while the RBM5 ZnF showed GGG as the preferred motif at most positions (Fig. 2E,F). As expected, the ZRANB2 C′-ZnF W79A mutant showed essentially no specific enrichment, and kpLogo analysis did not identify consistently preferred motifs across different protein concentrations (Fig. 2D,F). Overall, these data show that the 2-ZnF array design coupled to our modified RBNS assay can robustly identify preferred RNA-binding motifs of individual ZnFs.

Rational mutagenesis of ZRANB2 ZnF in search of novel RNA-binding motifs

Next, we attempted to rationally design ZRANB2 ZnF mutants with altered RNA sequence specificity. Based on a crystal structure of the ZRANB2 C′-ZnF bound to a 5′-AGGUAA-3′ RNA (Loughlin et al. 2009), we identified all protein side chains within hydrogen bonding distance of the RNA bases in the core GGU motif (Fig. 1B). Because we had observed higher expression and better resistance to proteolysis during purification with the 2-ZnF array containing the ZRANB2 N′-ZnF compared to the C′-ZnF (Fig. 1A), we chose to engineer the N′-ZnF. We designed a panel of 18 single mutants of the ZRANB2 N′-ZnF with one of three residues—R27, R28, or N32—mutated to a different residue with altered hydrogen-bonding characteristics and/or side-chain length (D, E, N, R, Q, H, or S; Table 1). We chose residues R27, R28, and N32 because these residues are predicted to be within hydrogen bonding distance of the bound RNA bases, based on the corresponding C′-ZnF crystal structure, in which Loughlin et al. (2009) showed that (1) R81 (R27 equivalent) interacts with base G1 through side chain and water-mediated backbone hydrogen bond; (2) G2 forms a bidentate interaction with R82 (R28 equivalent); and (3) N86 (N32 equivalent) side chain connects to G2 through a water-mediated hydrogen bond and interacts with the O4 carbonyl group of U3. All mutants could be expressed and purified equivalently to the unmutated 2-ZnF array (Supplemental Fig. S2), suggesting that the mutations do not affect folding or stability of the protein. We then performed RBNS on each protein

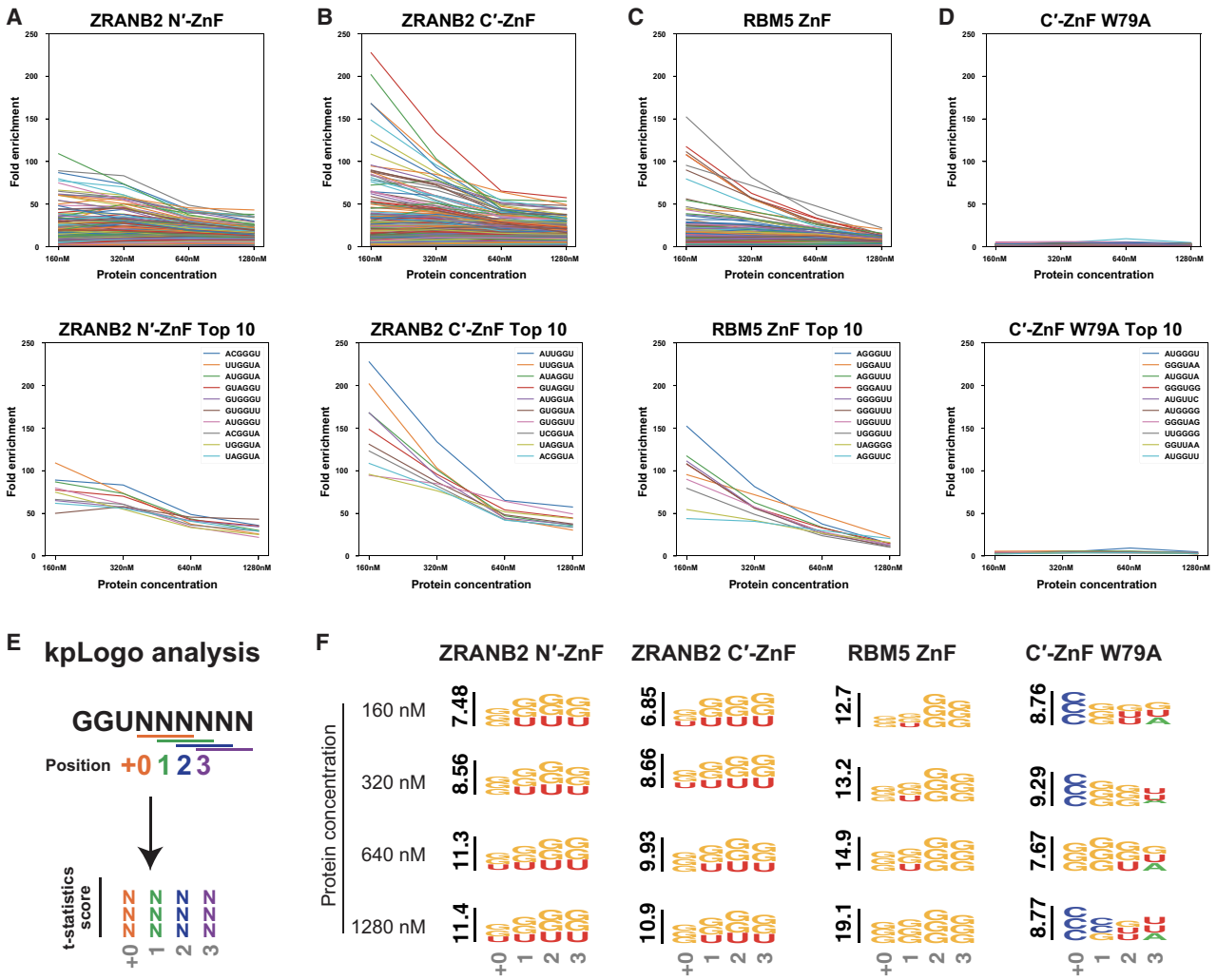

FIGURE 2. Benchmarking of the modified RBNS assay. (*A–D*) Fold enrichment score of all RNA 6-mers (*top* panel) or top 10 RNA 6-mers of ZRANB2 N′-ZnF, ZRANB2 C′-ZnF, RBM5 ZnF, and ZRANB2 C′-ZnF W79A mutant. (*E*) KpLogo analysis schematics. (*F*) KpLogo results of ZRANB2 N′-ZnF, ZRANB2 C′-ZnF, RBM5 ZnF, and ZRANB2 C′-ZnF W79A mutant.

with the RNA pool containing the six-base random window, at four protein concentrations.

A summary of the RBNS RNA-binding profile of all single mutants is shown in Table 2. Of the 18 single mutant ZnFs, six (R28E, R28Q, R28H, R28S, N32D, and N32E) showed almost no enrichment of any RNA sequences at any tested protein concentration, indicating that these mutants completely lost the ability to bind RNA (Supplemental Fig. S3A–F). Five of the 18 (R27D, R27E, R28D, R28N, and N32S) showed limited enrichment only at the highest tested protein concentrations, indicating that these mutants retain only residual RNA-binding activity (Supplemental Fig. S3G–K). A further five mutants (R27N, R27Q, R27H, R27S, and N32Q) showed some level of specificity as demonstrated by significant fold-enrichment of preferred motifs at 320 nM and 640 nM protein concentrations (Fig. 3A–F), with increasing fold-enrichment at higher protein concentrations. R27N and R27Q mutants showed enrichment values for preferred motifs at 1280 nM equivalent to that observed with the wild-type N′-ZnF at 160 nM, indicating that these mutants retain sequence-specific binding but show overall lower RNA-binding affinity than the wild-type ZnF. Another mutant, N32H, showed high fold-enrichment at relatively low protein concentration (320 nM) (Fig. 3G). Finally, N32R mutant (Fig. 3H) showed high fold-enrichment for preferred motifs at the lowest protein concentrations, and KpLogo analysis showed that these mutants showed an altered RNA motif preference of GGG, in contrast to the wild-type ZnF's preferred GGU motif. Altered preference for the third base in the core motif is consistent with the location of this residue within hydrogen bonding distance of the "U" base in the crystal structure of the ZRANB2 C′-ZnF bound to RNA (Supplemental Fig. S4A; Loughlin et al. 2009). Our analysis of single mutants showed that the N32R mutant shows an altered motif preference (GGG rather than GGU) while

TABLE 1. Single mutant design of ZnF

Single mutant no.	R27	R28	N32
1	D		
2	E		
3	N		
4	Q		
5	H		
6	S		
7		D	
8		E	
9		N	
10		Q	
11		H	
12		S	
13			R
14			D
15			E
16			Q
17			H
18			S

retaining overall high RNA-binding affinity (Fig. 3G), as indicated by high fold-enrichment of preferred motifs at low protein concentrations.

We next designed double mutants of the N′-ZnF in search of further alternation of the motif preference, involving N32 and a nearby residue, N22, mutated to different combinations of D, E, R, Q, H, or S. All double mutants could be expressed and purified similarly to the single mutant 2-ZnF arrays (Supplemental Fig. S4B) without disruption of protein folding or solubility. Based on the location of these two residues near the third RNA base in the core motif, we reasoned that double mutants of N32 and N22 might further alter the RNA sequence preference at the third base of the motif. Although N32D, N32E, and N32S appeared to disrupt RNA binding and selectivity in the context of a single mutant, we could not rule out the possibility of a cooperative mutation pair at both N32 and N22 that might recover RNA binding. We designed and purified a panel of 20 double mutants in the 2-ZnF array (Table 3) and assayed each by RBNS. Using our modified RBNS assay, we found that eight mutants, including all tested double mutants bearing N32R, exhibited variable effects on RNA binding, but none of them further changed the preferred RNA-binding motif from GGG (Supplemental Fig. S5; Table 4). Among them, both N22R/N32R and N22H/N32R appear to retain high fold-enrichment for preferred motifs at the lowest protein concentration, indicating high RNA-binding affinity (Supplemental Fig. S5A–C). The other 12 of 20 tested double mutants showed compromised RNA binding (Supplemental Fig. S6).

All-atom MD simulations reveal the ZnF-RNA recognition mechanism

Our RNA-binding data show that the single point mutation N32R alters the preferred RNA-binding motif of the ZRANB2 N′-ZnF from GGU to GGG. To gain a molecular understanding of how this mutation could alter RNA recognition, we performed all-atom MD simulations using AMBER14 (Peters et al. 2010; Case et al. 2023a,b). We built our starting model from the known structure of the ZRANB2 C′-ZnF in complex with GGU RNA (PDB: 3g9y), which is highly homologous to the N′-ZnF (Fig. 1A). To mimic the effects of the N32R mutation in the N′-ZnF, we generated a mutation in the structurally equivalent residue in the C′-ZnF (N86R; termed "NRmut"). To investigate the differences in protein–RNA interaction for both WT and NRmut ZnFs, we designed four models (WT + GGU RNA, WT + GGG RNA, NRmut + GGU RNA, NRmut + GGG RNA) and performed 50 nsec simulations in triplicate for each model. Root mean squared distance (RMSD) analysis indicated that the models remained stable throughout the simulations in all replicates (Supplemental Fig. S7).

We first investigated the effects of the ZnF mutation on the overall structural flexibility of the protein–RNA complex. We examined root mean squared fluctuation (RMSF) of the protein and RNA backbone for all four models (Fig. 4A). The protein backbone dynamics were not affected by the N-to-R mutation and remained similar when the ZnF was in complex with either GGU or GGG RNA. However, the GGU RNA displayed a higher overall flexibility in complex with the NRmut, in contrast to the GGG RNA. This indicates the NRmut-GGU RNA binding is less stable compared to NRmut-GGG RNA binding. In addition, we observed the highest variation of RMSF in both the protein and RNA backbone across replicates in the NRmut-GGU RNA complex (Supplemental Fig. S8A), corroborating our speculation.

We then sought to understand the hydrogen bonding interactions between the ZnF and the bound ssRNA, which play essential roles in specificity to intermolecular interactions (Hubbard and Kamran Haider 2010). We calculated the hydrogen-bond relative intensity score based on the fraction of time a given protein–RNA hydrogen bonding interaction existed (determined and tracked by CPPTRAJ [Roe and Cheatham 2013]) over the course of each 50 nsec simulation. The fraction of time for all hydrogen-bond donor and acceptor pairs is summed for every combination of protein and RNA residues, defining the "hydrogen-bond relative intensity" of each interacting protein + RNA residue pair (Fig. 4B; Supplemental Fig. S8B). In the WT + GGU RNA simulations, we observe long-lived hydrogen bonding interactions between RNA base G1 and

TABLE 2. RNA-binding profile in RBNS of all single mutants

	Mutant RNA-binding profile		
Single mutant	Affinity	Top 10 sequence motifs	Predominant kpLogo
R27D	N/A	N/A	N/A
R27E	N/A	N/A	N/A
R27N	Lower	AUGGGU	GGG
		GUGGGU	
		AUGGUA	
		GGUGGG	
		GUGGUA	
		GUGGGG	
		AUGGGA	
		UUGGGU	
		ACGGGA	
		AUGGGG	
R27Q	Lower	AUGGGU	GGG
		AUGGUA	
		GUGGGU	
		GUGGUA	
		ACGGUA	
		GGUGGG	
		AUGGGA	
		ACGGGA	
		GUAGGU	
		GUGGGG	
R27H	Lower	GUGGGU	GGU
		GUGGUA	
		UGGGUG	
		GUGGUU	
		UGGGUA	
		GUAGGU	
		UUGGUA	
		AUGGUA	
		GGUGGU	
		UUGGGU	
R27S	Lower	UGGGUA	GGU
		GUGGGU	
		ACGGGU	
		AUGGUA	
		AUGGGU	
		ACGGUA	
		UGGGUG	
		GUGGGG	
		GUAGGU	
		UUGGUA	
R28D	N/A	N/A	N/A
R28E	N/A	N/A	N/A
R28N	N/A	N/A	N/A
R28Q	N/A	N/A	N/A
R28H	N/A	N/A	N/A

Continued

TABLE 2. *Continued*

	Mutant RNA-binding profile		
Single mutant	Affinity	Top 10 sequence motifs	Predominant kpLogo
R28S	N/A	N/A	N/A
N32R	Similar	UGGGUA	GGG
		UUGGUA	
		AUGGUA	
		UGGGGU	
		ACGGGU	
		GUGGGU	
		UGGGUU	
		UUGGUG	
		GUGGUA	
		UUGGGU	
N32D	N/A	N/A	N/A
N32E	N/A	N/A	N/A
N32Q	Lower	AUGGGU	GGG
		UUGGUA	
		AUGGUA	
		GUGGGU	
		UGGGUA	
		GGUGGG	
		UUGGGU	
		GUGGUA	
		AGGGGA	
		UGGGUG	
N32H	Lower	UUGGUA	GGU
		ACGGGU	
		UGGGUA	
		AUGGUA	
		AUGGGU	
		UAGGUA	
		UGGGGU	
		GUGGGU	
		GUGGUA	
		UUGGGU	
N32S	N/A	N/A	N/A

protein residues Asp68 and Arg81; between RNA base G2 and protein residues Val77 (backbone) and Arg82; and RNA base U3 and protein residues Asn76, Asn86. We also detected hydrogen bonding between Trp79 and G2; this residue is stably stacked between bases G1 and G2 during the simulations. When the uridine (U3) was mutated to a guanosine (G3), the hydrogen bonds of all nucleobase-amino acid side-chain pairs were weakened by an average of 14.2% (Supplemental Fig. S9), indicating the weaker interaction of WT ZnF with a GGG 3-mer compared to GGU. Intriguingly, the asparagine to arginine mutation in NRmut completely disrupted the interaction of this residue with RNA base U3 in the NRmut + GGU model (Fig. 4B, bottom left panel), indicating the loss of specific recognition of U3 by NRmut. In contrast, the NRmut + GGG RNA simulation showed a restored hydrogen-bond interaction between the newly introduced arginine and RNA base G3, along with an additional interaction with RNA base G2. Besides the restored hydrogen bonding between the protein and the third RNA base, the total hydrogen-bond intensity was also restored to levels equivalent to those in the WT ZnF + GGU RNA simulations (Supplemental Fig. S9), consistent with our experimental results that the NRmut prefers to bind the RNA motif GGG over GGU. Interestingly, each base in the GGG motif is contacted equivalently by an arginine residue in

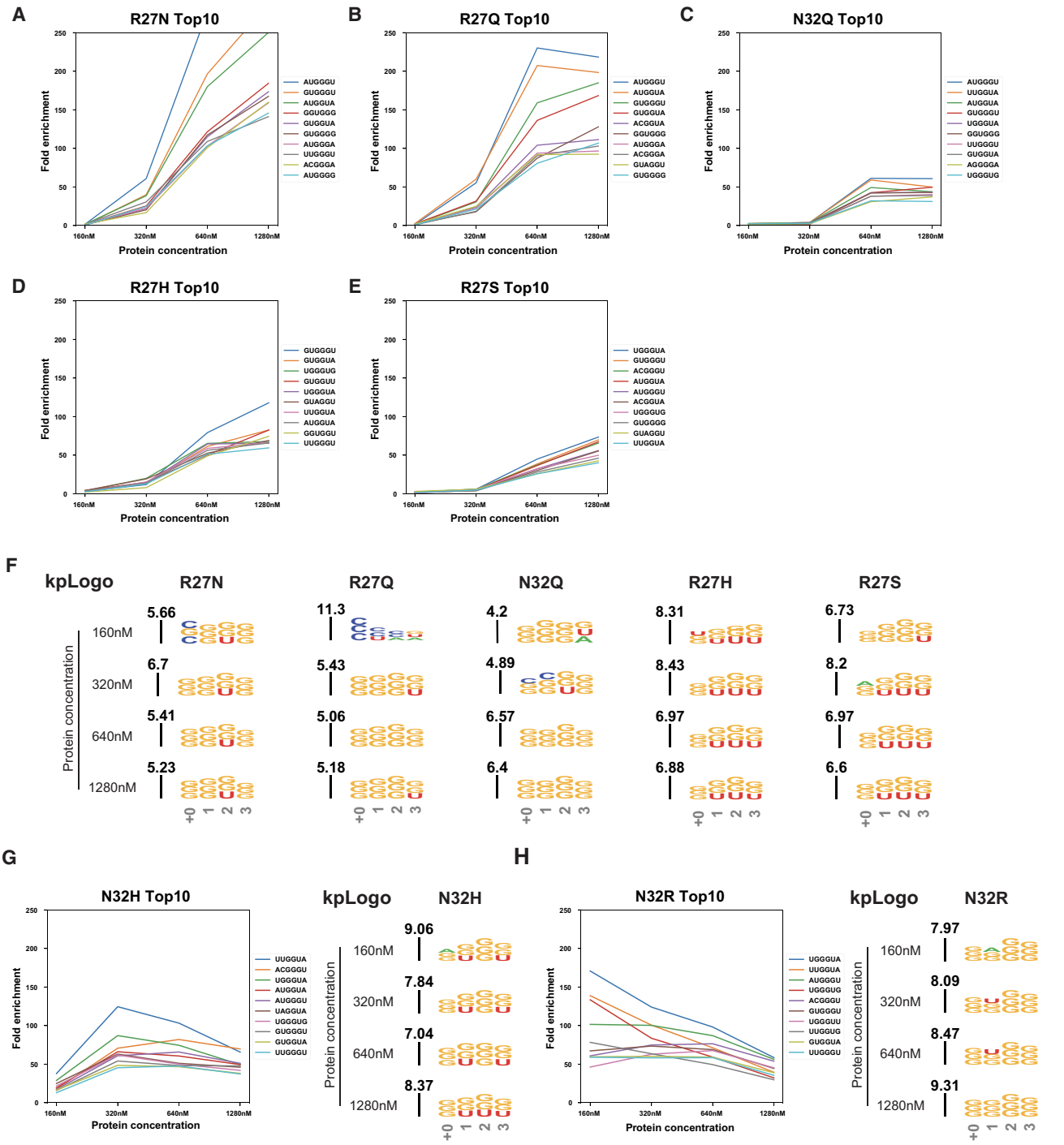

FIGURE 3. RBNS results of selective ZnF single mutants. (*A–E*) Top 10 enriched RNA 6-mers of R27N, R27Q, N32Q, R27H, and R27S. This group of single mutants showed overall lower RNA-binding affinity than the wild-type ZnF, despite retaining some degree of specificity. (*F*) KpLogo results of R27N, R27Q, N32Q, R27H, and R27S. (*G*) Top 10 enriched RNA 6-mers and kpLogo results of N32H. (*H*) Top 10 enriched RNA 6-mers and kpLogo results of N32R, which retained RNA-binding specificity and RNA-binding affinity.

NRmut: R81 contacts G1, R82 contacts G2, and R86 contacts G3.

Finally, we investigated the molecular mechanism underlying the altered RNA-binding preference in NRmut by examining the protein–RNA binding conformations in the MD simulation trajectories. Hierarchical clustering derived five representative conformations from each 50 nsec simulation and the average model of the major conformation (>50% of the simulation time) was visualized (see Materials and Methods). Structural overlays of the average

TABLE 3. Double mutant design of ZnF

Double mutant no.	N22	N32
1	R	R
2	R	H
3	D	R
4	D	H
5	E	R
6	E	D
7	E	E
8	E	Q
9	E	H
10	E	S
11	Q	R
12	Q	H
13	H	R
14	H	H
15	S	R
16	S	D
17	S	E
18	S	Q
19	S	H
20	S	S

model of the major conformation (>50% of the simulation time) in the ZnF WT and NRmut in complex with GGU RNA showed a high degree of overlap of the RNA–protein complex backbone (Fig. 4C). As expected, the closest N–O distance between the newly introduced Arginine residue in NRmut and the RNA base U3 (5.1 Å) is much larger than the median distance between amide C=O acceptors and N–H donors in a typical hydrogen bond (~2.9 Å) (Fig. 4C,D; Bissantz et al. 2010), indicating loss of hydrogen-bond interactions. Strikingly, structural overlays of representative structures of ZnF NRmut + GGU RNA and NRmut + GGG RNA revealed a significant displacement of the protein backbone (Fig. 4E). NRmut seemed to be displaced away from the GGG RNA compared to the simulations with GGU RNA. The closest N–O distance between Arg86 and G3 was measured 2.9 Å, indicating the restored hydrogen bonding for NRmut with GGG RNA in contrast to GGU RNA. The results showed that the NRmut likely adapts to GGG RNA binding by the backbone displacement and rotation, despite the interaction pair changes from short amino acid side chain–small nucleobase (Asn–Uridine) to long amino acid side chain–bulky nucleobase (Arg–Guanine). In summary, the MD simulations rationalize how a single amino acid substitution (from asparagine to arginine) can alter the RNA motif recognition by an RNA-binding ZnF, thereby deepening our understanding of the mechanisms involved in protein–RNA recognition.

DISCUSSION

Here we show that rational design of single mutations in RNA-contacting residues can alter the sequence specificity for an RNA-binding ZnF without compromising RNA-binding affinity, opening the door to further rational design or screening of mutants with further alterations in sequence specificity. Our modified RBNS assay robustly profiles both sequence specificity and overall RNA-binding affinity and represents a promising avenue for high-throughput screening of ZnF mutants. Our MD simulations show that the NRMut ZnF likely adapts to its preferred GGG motif through both main-chain and side-chain movement, revealing a high degree of flexibility in this family of RNA-binding proteins. Our work has defined a set of key interactions that define the specificity for the third RNA base in the ZRANB2 ZnF preferred motif; we anticipate that future work will identify interactions that define specificity for the first and second bases of this motif, enabling

TABLE 4. RNA-binding profile in RBNS of all double mutants

ZnF double mutant		Mutant RNA-binding profile	
N22	N32	Affinity	Predominant kpLogo
R	R	Similar	GGU
R	H	Lower	GGG
D	R	Lower	GGG
D	H	N/A	N/A
E	R	Lower	GGG
E	D	N/A	N/A
E	E	N/A	N/A
E	Q	N/A	N/A
E	H	N/A	N/A
E	S	N/A	N/A
Q	R	Lower	GGG
Q	H	N/A	N/A
H	R	Similar	GGG
H	H	Lower	GGG
S	R	Lower	GGG
S	D	N/A	N/A
S	E	N/A	N/A
S	Q	N/A	N/A
S	H	N/A	N/A
S	S	N/A	N/A

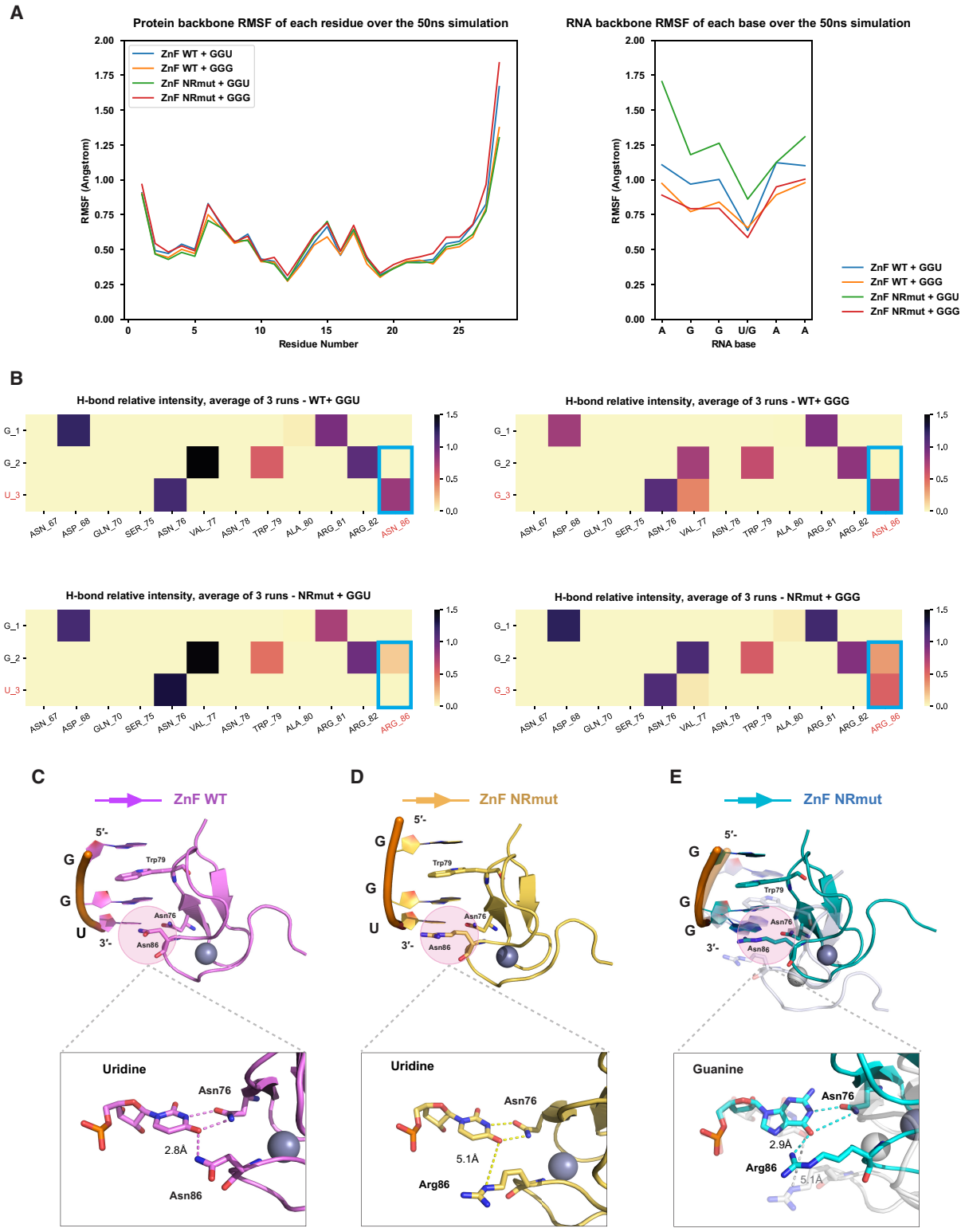

FIGURE 4. MD simulation of ZnF-RNA complexes. (*A*) RMSF of each protein residue and RNA base during the simulation (average of three replicates of 50 nsec simulations) in four models (ZnF WT + GGU, ZnF WT + GGG, ZnF NRmut + GGU, ZnF NRmut + GGG). (*B*) Heat map of H-bond relative intensity (magnitude indicated by color bar) between the three RNA bases and the RNA-contacting residues during the simulation (average of three replicates of 50 nsec simulations). (*C–E*) Representative structures of ZnF-RNA complexes during the simulations. Note that the orientations in the *lower* panels are different from the *top* panels (slightly rotated). (*C*) The representative structure of ZnF WT + GGU (green); (*D*) the representative structure of ZnF NRmut + GGU (yellow); (*E*) the representative structure of ZnF NRmut + GGG (cyan) overlaying with the representative structure of ZnF NRmut + GGU (white). The protein residues highlighted in the *top* panel are Arg86/Asn86 and Trp79.

rational design of individual ZnFs and ZnF arrays for high-specificity RNA targeting.

MATERIALS AND METHODS

Purification of ZRANB2 ZnF and mutants

The DNA sequences of N′ and C′ ZnF in human ZRANB2 (NP_005446.2) were codon-optimized for expression in the *Escherichia coli* strain. A 2-ZnF array of ZnF_WT-ZnF* was cloned into a modified UC Berkeley MacroLab vector 2CT (Addgene 29706) to generate an N-terminal fusion to a His_6-MBP (maltose-binding peptide) tag with a GSGSG-linker in between the ZnFs. A streptavidin-binding peptide (SBP) tag was cloned at the N-terminal of the 2-ZnF array. The His-MBP-SBP-ZnF_WT-ZnF* protein was expressed in *E. coli* strain Rosetta2 pLysS (EMD Millipore) by growing cultures in 2xYT media to mid-log phase at 37°C, followed by induction with 0.25 mM IPTG at 18°C for 16 h, supplemented with 100 µM $ZnCl_2$. For protein purification, cells were harvested by centrifugation, suspended in resuspension buffer (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole, 2 mM β-Mercaptoethanol, and 10% glycerol) and lysed by sonication. Lysates were clarified by centrifugation (16,000 rpm 30 min); then the supernatant was loaded onto a 1 mL Ni^{2+} affinity column (HisTrap HP, Cytiva) pre-equilibrated with the resuspension buffer. The column was washed with a buffer containing 20 mM imidazole and 500 mM NaCl for 10 column volume (CV), followed by loading 0.5 mg MBP-TEV protease (Tobacco Etch Virus nuclear-inclusion-a endopeptidase) onto the HisTrap column for on-column His-MBP tag cleavage. The MBP-TEV protease was expressed and purified from pKR1043 (Addgene 8835) as described (Kapust et al. 2001). After 4 h TEV cleavage at room temperature, the His-MBP tag-free protein was washed by 10 CV with TEV wash buffer (20 mM Tris pH7.5, 150 mM KCl, 2 mM β-ME, 10% glycerol). The protein was then concentrated by ultrafiltration, aliquoted, and frozen at −80°C for future use. For His-MBP tagged protein, no on-column TEV protease cleavage was carried out, and the protein was eluted with a buffer containing 250 mM imidazole and 150 mM KCl.

RNA bind-n-seq (RBNS)

RBNS experiments with an RNA pool containing a randomized window were performed as indicated in the literature (Lambert et al. 2014; Dominguez et al. 2018). In brief, randomized RNA oligonucleotides (6 nt) containing a sequence of "AAAGGUNNNNNNAAA" flanked by adaptor sequences were synthesized through in vitro transcription from synthesized DNA libraries (Integrated DNA Technologies). The pool of RNA was incubated with the SBP-tagged recombinant 2-ZnF-array at five concentrations of protein (160, 320, 640, 1280 nM), while keeping RNA concentrations constant at 1 µM. RNA–protein complexes were immobilized with streptavidin-conjugated affinity magnetic resin and were subjected to multiple washing steps. Then the RNA was eluted and prepared for deep sequencing. A total of 0.2–0.5 million reads per RBP pulldown concentration were targeted to ensure more than 50-fold coverage for each unique RNA sequence (4096 sequences × 50 fold coverage = 204,800 reads minimum per sample). The original pool of RNA was also sequenced with a similar depth as the input library.

RBNS data processing analysis

The fold enrichment of each RNA 6-mer was calculated as the frequency of each 6-mer in the pulldown library reads divided by its frequency in the input library. Fold enrichments were ranked and plotted using Matplotlib in python. To identify the core 3-mer motif from the enriched 6-mers, kpLogo (Wu and Bartel 2017) was used to report the 3-mer enrichment at each position using Student's *t*-test. The command "kpLogo weighted_RBNS_file.txt -weighted -k 3 -plot s -fontsize 50" was used to generate kpLogo output, with 6-mer RNA sequences weighted by their fold enrichments as input. K-means clustering of the RNA 6-mers was performed using the Scikit-learn package in python, with the fold enrichment of each sequence as the features. Number of clusters *k* was set to four, and *k* cluster centroids were initialized using *k*-means++ and a maximum of 500 iterations were performed. *K*-mean clustering code is: KMeans(n_clusters = 4, init="k-means++", _init = 50, _iter = 500, random_state = 42).

All-atom molecular dynamics simulations

The ZnF-RNA complexes were constructed based on crystal structure (PDB: 3g9y). All crystallographic waters were preserved. The NRmut and the RNA mutant (AGGUAA → AGGGAA) were prepared by inducing virtual mutations to the ZRANB2-C′-ZnF-RNA structure using the mutagenesis plugin available in PyMOL. Protein and RNA atoms were represented using the AMBER (Case et al. 2023a,b) ff14SB force field (Maier et al. 2015), and the OL3 parameters (Cheatham and Case 2013), respectively. Zinc chelation was represented using Zinc AMBER Force Field (ZAFF) (Peters et al. 2010). The protein and RNA are solvated in a cubic TIP3P water box with 10 Å distance from the solute to the boundaries of the simulation box. Hydrogen atoms were added through the LeaP program in AMBER. A varying number of Na^+ ions were added to neutralize the system. The simulation cell was then replicated infinitely in three dimensions to impose periodic boundary conditions. All MD simulations were performed under periodic boundary conditions using the CUDA accelerated version of PMEMD implemented in the Amber14 (Case et al. 2023a,b) suite of programs on TSCC (San Diego Supercomputer Center 2022). The structures were first relaxed using a combination of steepest descent (10,000 steps, 2 fsec each) and conjugate gradient minimization (10,000 steps, 2 fsec each). The lengths of bonds involving hydrogen atoms were constrained by SHAKE. The particle mesh Ewald method (PME) was employed to treat the long-range electrostatics. The nonbonded cutoff for the long-range interactions was set to 10 Å. Heating was conducted from 0 to 300 K in 800 psec, using Langevin dynamics with a collision frequency of 2 $psec^{-1}$ as temperature control, followed by pre-equilibrium with 800–1000 psec at the same condition of heating. Another pre-equilibrium was performed in an NPT (constant number of atoms, pressure, and temperature) ensemble at 300 K and 1 atm, using Langevin dynamics for temperature regulation and a Berendsen barostat for pressure control. No restraints were applied after this stage. A following 10 nsec equilibrium step was conducted with similar conditions as pre-equilibrium, except

that the Monte Carlo barostat was used for pressure control. A 5 nsec of postequilibrium run was conducted before the production simulations and excluded from the analysis. Subsequently, three replicates of 50 nsec unbiased MD simulations were carried out for the ZnF WT + GGU RNA, ZnF WT + GGG RNA, ZnF NRmut + GGU RNA, ZnF NRmut + GGG RNA complexes.

MD simulation data analysis

The CPPTRAJ (Roe and Cheatham 2013) module was used to analyze all MD trajectories. The RMSD and RMSF of the ZnF-RNA complexes from each MD trajectory were calculated, with respect to the Cα atoms of the protein backbone and the C3′ and C4′ atoms of the RNA backbone in reference to the final structure of the postequilibrium run. Hydrogen bonds between residues were defined and measured by the CPPTRAJ package. The visualization of the MD trajectories was rendered using visual molecular dynamics (VMD) (Humphrey et al. 1996), and data were plotted using Matplotlib. Hierarchical clustering was performed after stripping water and ions and based on the protein and RNA backbone positions excluding hydrogen atoms (command: cluster c1 hieragglo epsilon 3.0 clusters 5 averagelinkage rms :1-36@C,N,O,CA,P,O3′,O5′,C3′,C4′,C5′&!@H=). The average model of the top #1 representative cluster was used for overlay. PyMOL was used for structure visualization and overlay.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

COMPETING INTEREST STATEMENT

G.W.Y. is an SAB member of Jumpcode Genomics and a cofounder, member of the Board of Directors, on the SAB, equity holder, and paid consultant for Locanabio and Eclipse BioInnovations. G.W.Y. is a distinguished visiting professor at the National University of Singapore. G.W.Y.'s interests have been reviewed and approved by the University of California, San Diego in accordance with its conflict-of-interest policies.

ACKNOWLEDGMENTS

The authors acknowledge support from the National Institutes of Health (R35 GM144121 to K.D.C., R01 HG004569, HG009889, HG011864, NS103172, and MH131907 to G.W.Y.). Q.L. was supported by an individual predoctoral fellowship from the American Heart Association. This publication includes data generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000 that was purchased with funding from a National Institutes of Health SIG grant (#S10 OD026929) and data generated at the Sequencing Core Facility, La Jolla Institute, utilizing an Illumina NovaSeq 6000 that was acquired through the Shared Instrumentation Grant (SIG) Program (#S10OD025052). Computational resources were provided by the Department of Defense High Performance Computing Modernization Program (HPCMP) and the Triton Shared Computing Cluster (TSCC) at the San Diego Supercomputer Center (SDSC). We would also like to thank the Stem Cell Genomics Core at the Sanford Stem Cell Institute for providing sequencing services.

Received November 17, 2024; accepted November 27, 2024.

REFERENCES

Beerli RR, Barbas CF. 2002. Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol* **20:** 135–141. doi:10.1038/nbt0202-135

Bissantz C, Kuhn B, Stahl M. 2010. A medicinal chemist's guide to molecular interactions. *J Med Chem* **53:** 5061–5084. doi:10.1021/jm100112j

Case DA, Aktulga HM, Belfon K, Ben-Shalom IY, Berryman JT, Brozell SR, Cerutti DS, Cheatham TEC III, Cisneros GA, Cruzeiro VWD, et al. 2023a. *Amber 2023*. University of California, San Francisco.

Case DA, Aktulga HM, Belfon K, Cerutti DS, Cisneros GA, Cruzeiro VWD, Forouzesh N, Giese TJ, Götz AW, Gohlke H, et al. 2023b. AmberTools. *J Chem Inf Model* **63:** 6183–6191. doi:10.1021/acs.jcim.3c01153

Cheatham TE III, Case DA. 2013. Twenty-five years of nucleic acid simulations. *Biopolymers* **99:** 969–977. doi:10.1002/bip.22331

De Franco S, Vandenameele J, Brans A, Verlaine O, Bendak K, Damblon C, Matagne A, Segal DJ, Galleni M, Mackay JP, et al. 2019. Exploring the suitability of RanBP2-type zinc fingers for RNA-binding protein design. *Sci Rep* **9:** 1–13. doi:10.1038/s41598-019-38655-y

Dominguez D, Freese P, Alexis MS, Su A, Hochman M, Palden T, Bazile C, Lambert NJ, Van Nostrand EL, Pratt GA, et al. 2018. Sequence, structure, and context preferences of human RNA binding proteins. *Mol Cell* **70:** 854–867.e9. doi:10.1016/j.molcel.2018.05.001

Gerstberger S, Hafner M, Tuschl T. 2014. A census of human RNA-binding proteins. *Nat Rev Genet* **15:** 829–845. doi:10.1038/nrg3813

Hubbard RE, Kamran Haider M. 2010. *Hydrogen bonds in proteins: role and strength*. John Wiley & Sons, Hoboken, NJ.

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph* **14:** 33–38. doi:10.1016/0263-7855(96)00018-5

Isalan M, Choo Y. 2001. Engineering nucleic acid-binding proteins by phage display. In *DNA-protein interactions: principles and protocols* (ed. Moss T), pp. 417–429. Humana Press, Totowa, NJ.

Kapust RB, Tözsér J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. 2001. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **14:** 993–1000. doi:10.1093/protein/14.12.993

Konermann S, Lotfy P, Brideau NJ, Oki J, Shokhirev MN, Hsu PD. 2018. Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell* **173:** 665–676.e14. doi:10.1016/j.cell.2018.02.033

Lambert N, Robertson A, Jangi M, McGeary S, Sharp PA, Burge CB. 2014. RNA bind-n-seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell* **54:** 887–900. doi:10.1016/j.molcel.2014.04.016

Lau C-H, Suh Y. 2018. In vivo epigenome editing and transcriptional modulation using CRISPR technology. *Transgenic Res* **27:** 489–509. doi:10.1007/s11248-018-0096-8

Lauffer MC, van Roon-Mom W, Aartsma-Rus A. 2024. Possibilities and limitations of antisense oligonucleotide therapies for the treatment of monogenic disorders. *Commun Med* **4:** 1–11. doi:10.1038/s43856-023-00419-1

Loughlin FE, Mansfield RE, Vaz PM, McGrath AP, Setiyaputra S, Gamsjaeger R, Chen ES, Morris BJ, Guss JM, Mackay JP.

2009. The zinc fingers of the SR-like protein ZRANB2 are single-stranded RNA-binding domains that recognize 5′ splice site-like sequences. *Proc Natl Acad Sci* **106:** 5581–5586. doi:10.1073/pnas.0802466106

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. 2015. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* **11:** 3696–3713. doi:10.1021/acs.jctc.5b00255

Nguyen CD, Mansfield RE, Leung W, Vaz PM, Loughlin FE, Grant RP, MacKay JP. 2011. Characterization of a family of RanBP2-type zinc fingers that can recognize single-stranded RNA. *J Mol Biol* **407:** 273–283. doi:10.1016/j.jmb.2010.12.041

O'Connell MR, Vandevenne M, Nguyen CD, Matthews JM, Gamsjaeger R, Segal DJ, MacKay JP. 2012. Modular assembly of RanBP2-type zinc finger domains to target single-stranded RNA. *Angew Chem Int Ed* **51:** 5371–5375. doi:10.1002/anie.201200866

Pabo CO, Peisach E, Grant RA. 2001. Design and selection of novel Cys_2His_2 zinc finger proteins. *Annu Rev Biochem* **70:** 313–340. doi:10.1146/annurev.biochem.70.1.313

Peters MB, Yang Y, Wang B, Füsti-Molnár L, Weaver MN, Merz KM Jr. 2010. Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). *J Chem Theory Comput* **6:** 2935–2947. doi:10.1021/ct1002626

Roe DR, Cheatham TEI. 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* **9:** 3084–3095. doi:10.1021/ct400341p

San Diego Supercomputer Center. 2022. *Triton Shared Computing Cluster*. University of California, San Diego. Service. doi:10.57873/T34W2R

Shepard PJ, Hertel KJ. 2009. The SR protein family. *Genome Biol* **10:** 242. doi:10.1186/gb-2009-10-10-242

Soni K, Jagtap PKA, Martínez-Lumbreras S, Bonnal S, Geerlof A, Stehle R, Simon B, Valcárcel J, Sattler M. 2023. Structural basis for specific RNA recognition by the alternative splicing factor RBM5. *Nat Commun* **14:** 4233. doi:10.1038/s41467-023-39961-w

Sugimoto M, Suda A, Futaki S, Imanishi M. 2020. Effective RNA regulation by combination of multiple programmable RNA-binding proteins. *Appl Sci* **10:** 6803. doi:10.3390/app10196803

Wu X, Bartel DP. 2017. kpLogo: positional k-mer analysis reveals hidden specificity in biological sequences. *Nucleic Acids Res* **45:** W534–W538. doi:10.1093/nar/gkx323

Zhao Y-Y, Mao M-W, Zhang W-J, Wang J, Li H-T, Yang Y, Wang Z, Wu J-W. 2018. Expanding RNA binding specificity and affinity of engineered PUF domains. *Nucleic Acids Res* **46:** 4771–4782. doi:10.1093/nar/gky134

MEET THE FIRST AUTHOR

Qishan Liang

***Meet the First Author(s)* is an editorial feature within *RNA*, in which the first author(s) of research-based papers in each issue have the opportunity to introduce themselves and their work to readers of *RNA* and the RNA research community. Qishan (Lisa) Liang is the first author of this paper, "Rational design yields RNA-binding zinc finger domains with altered sequence specificity." Lisa recently completed her PhD in biochemistry and molecular biophysics at University of California San Diego and is currently a postdoctoral associate at Scripps Research.**

What are the major results described in your paper and how do they impact this branch of the field?

We engineered an RNA-binding zinc finger (ZnF) protein for sequence-specific RNA-binding through rational mutagenesis and investigated the zinc finger–RNA interaction both in vitro and in silico. Our findings broaden the knowledge of ZnF-RNA recognition principles and provide a basis for the future application of RNA-binding ZnFs in programmable RNA targeting.

What led you to study RNA or this aspect of RNA science?

RNAs and RNA-binding proteins compose a fascinating concerto in the cells—complex yet melodic. I was drawn to protein structures and dynamics as an undergrad. As I learned more about RNAs, especially their flexible structures, I became very interested in protein–RNA interactions at the atomic level. Co-advised by the Kevin Corbett lab and the Gene Yeo lab in my PhD, I had the opportunity to study RNA-binding proteins with the guidance of structural biology experts and RNA biology experts. Engineered RNA-binding proteins (ZnF in this case) have great potential for next-generation therapeutics for RNA-relevant diseases, which is also why I am passionate about studying RNA-binding proteins and RNAs.

During the course of these experiments, were there any surprising results or particular difficulties that altered your thinking and subsequent focus?

We originally aimed to engineer an array of zinc fingers for sequence-specific RNA targeting. However, using the standard RNA bind-n-seq (RBNS) protocol with a 20 nt randomized RNA window, it was extremely hard to reveal the binding motif of a ZnF array, due to the natural "gap" between individual binding motifs of each ZnF. We then decided to characterize one ZnF mutant at a time, with the WT-mutant ZnF array design and only 6 nt randomized window for the mutant ZnF and subsequently figured out that kpLogo was an effective way to extract the ZnF-binding motif.

Continued

If you were able to give one piece of advice to your younger self, what would that be?

It's normal to feel frustrated as a human being trying to understand the rules governing Mother Nature. Frustration means you are approaching the boundaries of human knowledge—keep the big picture in mind and keep trying.

Are there specific individuals or groups who have influenced your philosophy or approach to science?

The co-mentorship in the Corbett lab and the Yeo lab during my PhD trained my interdisciplinary approach and engineering mindset in science. The zinc finger protein engineering project is an example—it started with the goal of a compact and programmable RNA-targeting tool, with a variety of skill sets involved, including structural biology, biochemistry, computational biology, and bioinformatics. I learned the importance of seeking resources and inspirations from different experts, as well as the importance of communicating efficiently between disciplines. These can really spark novelty and allow me to tackle challenges in unexplored fields.