Received: 26 February 2021 | Revised: 12 April 2021 | Accepted: 13 April 2021

DOI: 10.1002/pro.4083

FULL-LENGTH PAPER

THE PROTEIN SOCIETY WILEY

Human DND1-RRM2 forms a non-canonical domain swapped dimer

Pooja Kumari | **Neel Sarovar Bhavesh**

Transcription Regulation Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India

Correspondence
Pooja Kumari and Neel Sarovar Bhavesh, Transcription Regulation group, International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Marg, New Delhi 110067, India.
Email: kashyappooja1902@gmail.com (P. K.) and neelsb@icgeb.res.in (N. S. B.)

Funding information
International Centre for Genetic Engineering and Biotechnology, New Delhi; Department of Biotechnology, Government of India, Grant/Award Numbers: BT/PR13018/BRB/10/731/2009, BT/INF/22/SP22660/2017

Abstract

RNA recognition motif (RRM) being the most abundant RNA binding domain in eukaryotes, is a major player in cellular regulation. Several variations in the canonical βαββαβ topology have been observed. We have determined the 2.3 Å crystal structure of the human DND1-RRM2 domain. The structure revealed an interesting non-canonical RRM fold, which is maintained by the formation of a 3D domain swapped dimer between β_1 and β_4 strands across protomers. We have delineated the structural basis of the stable domain swapped dimer formation using the residue level dynamics of protein explored by NMR spectroscopy and MD simulations. Our structural and dynamics studies substantiate major determinants and molecular basis for domain swapped dimerization observed in the RRM domain.

KEYWORDS

crystal structure, DND1, domain swapped dimerization, NMR spectroscopy, RNA binding protein, RNA recognition motif

1 | INTRODUCTION

RNA recognition motif (RRM) or ribonucleoprotein domain (RNP), is the most abundant nucleic acid binding domain in higher vertebrates. It is present either in single copy, multiple copies, or in conjunction with other domains in the same protein.[1] RRMs are essential in regulating post-transcriptional gene regulation by virtue of their interactions. They modulate several major processes in mRNA metabolism, which are; capping, splicing, polyadenylation, export, translation, and decay.[2] Nuances in interaction affinity is mediated by RNA binding domain structure, RNA structure, auxiliary protein interaction, PTM based regulation, and so forth. RRMs with their varied types of interactions with proteins/nucleic acids happen to be highly functionally versatile.[3] Several variations have been observed in the β1–α1–β2–β3–α2–β4 topology of an RRM domain such as extended secondary structural elements in the N- and C- terminus, formation of β hairpin in the loop region, variation in the α-helix and β-sheet length, presence of additional loop, so forth. Often found to be crucial during sequence specific interactions, RRM domain structural variation have shown pronounced plasticity in order to elicit cellular regulation.[4]

Human DND1 (dead end protein homolog1) also known as DND microRNA-mediated repression inhibitor 1 is an RNA binding protein containing two RRMs (RNA recognition motif) in tandem and a double stranded RNA binding domain at the C-terminus separated by a flexible linker of 40 residues. It is located at position q31.3 on

Abbreviations: *CCR4*, carbon catabolite repressor 4; CSPs, chemical shift perturbation; DELTA-BLAST, domain enhanced lookup time accelerated BLAST; DND1, dead end protein homolog1; dsRBD, double stranded RNA binding domain; HSQC, heteronuclear single quantum coherence; MD simulation, molecular dynamics simulation; NMR, nuclear magnetic resonance; NOESY, nuclear overhauser effect spectroscopy; *NOT*, negative on TATA; RMSD, root mean square deviation; RMSF, root mean square fluctuation; RNP, ribonucleoprotein; RRM, RNA recognition motif; SAD, single wavelength anomalous diffraction; UTR, untranslated region.

human chromosome 5, and is highly conserved in vertebrates. It is essential for migration and viability of primordial germ cells in zebrafish.[5] In mice, complete loss of DND1 protein leads to early embryonic lethality.[6] Introduction of nonsense mutations (R190X) named as Ter in RRM2 of DND1 in mice and (W289X) in dsRBD of rats, results in the formation of a truncated DND1 protein, which is followed by germ cell depletion, found to be oncogenic and result in testicular germ cell tumors.[7–9] In humans, reports of uncontrolled regulation at the protein level or mutation in the *dnd1* gene have been associated with testicular cancer, and tongue squamous cell carcinoma.[10–12] Although the molecular function associated with DND1 has been studied in zebrafish, Xenopus, rodents, and human cells, there is a disparity in the function so far unraveled. On the one hand DND1 is known to positively modulate the stability and translation of AU rich element containing mRNA, but on the other hand it has also been reported to associate with proteins of macromolecular complexes CCR4-NOT, which destabilize mRNAs.[13,14] The specification of transcripts for deadenylation is usually mediated through adapter proteins that interact with the enzyme complexes as well as its cognate RNA sequences.[15] Mouse DND1 is also known to co-localize with NANOS2 protein for targeted mRNA processing.[16,17] DND1 targets are involved in chromatin regulation linking DND1 to epigenetic regulation.[18] DND1 plays a multifaceted role and is pivotal in post transcriptional regulation. Although no information on the three-dimensional structure is available, still structure prediction tools utilizing sequence-based prediction methods have proposed multi-domain architectures for DND1 protein showing three separated domains.[18] We still lack a comprehensive model for the precise role of DND1.

Here, we report the 2.3 Å resolution crystal structure of the DND1-RRM2 domain, which reveals the formation of a domain swapped dimer, first to be reported in an RRM domain. Domain swapping is a process of protein oligomerization in which identical structural motif is exchanged between two protomers.[19] Using a combination of biochemical studies, NMR spectroscopy and MD simulation, we have further deduced determinants of the conformational flexibility in the hinge loop, which allowed domain swapped dimer formation in DND1-RRM2.

2 | RESULTS

2.1 | RRM2 of DND1 contains pseudo RNP motifs and a conserved HRAAAMA motif

DND1 consists of three domains as predicted by DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST), which are RRM1 (58–136), RRM2 (139–217), and dsRBD (254–332) double stranded RNA binding domain. The analysis of DND1 sequence showed significant bias in composition consisting of 50% aliphatic residues with distinctly higher number of leucine, alanine, glycine and proline amino acid residues, 7% aromatic, 20% neutral, 13% basic, and 7% acidic residues (Figure S1). To identify the structural features and their conservation, which are responsible for DND1 function, the sequence alignment of DND1 sequence from various vertebrate organisms is shown in Figure S2. We observed that the ribonucleoprotein (RNP) motifs, RNP1 and RNP2, corresponding with RNP1 and RNP2 consensus sequence which is (R/K)-G-(F/Y)-(G/A)-(F/Y)-V-X-(F/Y) and (L/I)-(F/Y)-(V/I)-X-(N/G)-L, respectively, is well conserved in case of RRM1.[20] However, major differences were found in RRM2 domain and dsRBD domain, where the RNP motifs were neither well defined nor well conserved. Despite this, a highly conserved HRAAAMA motif was found in the RRM2 domain, which is a putative ATPase and promotes nanos1 translation.[21] Studies from Xenopus DND1 recombinant protein showed that RRM1 predominantly binds to nanos1 mRNA.[21] PAR-CLIP studies performed by Yamaji et al. showed that DND1 predominantly binds to UU(A/U) trinucleotide motif in the 3′ untranslated regions of mRNA.[13] We have investigated RNA binding using isothermal titration calorimetry and titrated monomeric RRM2 which is having essentially the same binding area as compared to the domain swap dimer, with UUUUUU and UUAUUU RNA sequences. We found that RRM2 alone shows no interaction with these RNA sequences as shown in Figure S3. As the RNP sites containing the aromatic residues, which normally interact with RNA, are not well conserved, hence, RRM2 does not play a role in canonical RNA recognition.

2.2 | DND1-RRM2 forms a stable domain swapped dimer

The RRM2 construct contained a 21 amino acid long N-terminus hexa-histidine tag followed by the RRM2 domain E139-K217 residues shown in Figure S4. The domain architecture is shown in Figure 1(a). The purified DND1-RRM2 domain crystallized and the native crystal diffracted X-rays to a resolution of 2.4 Å. DND1-RRM2 domain showed low sequence homology to any known protein structure except with RBM47 (PDB id 2DIS), which was determined using solution-state NMR spectroscopy. Therefore, the crystal structure of DND1-RRM2 was determined using selenium based single wavelength anomalous diffraction (SAD) method. The crystal diffracted to 2.3 Å and belonged to $I4_1$ space group.

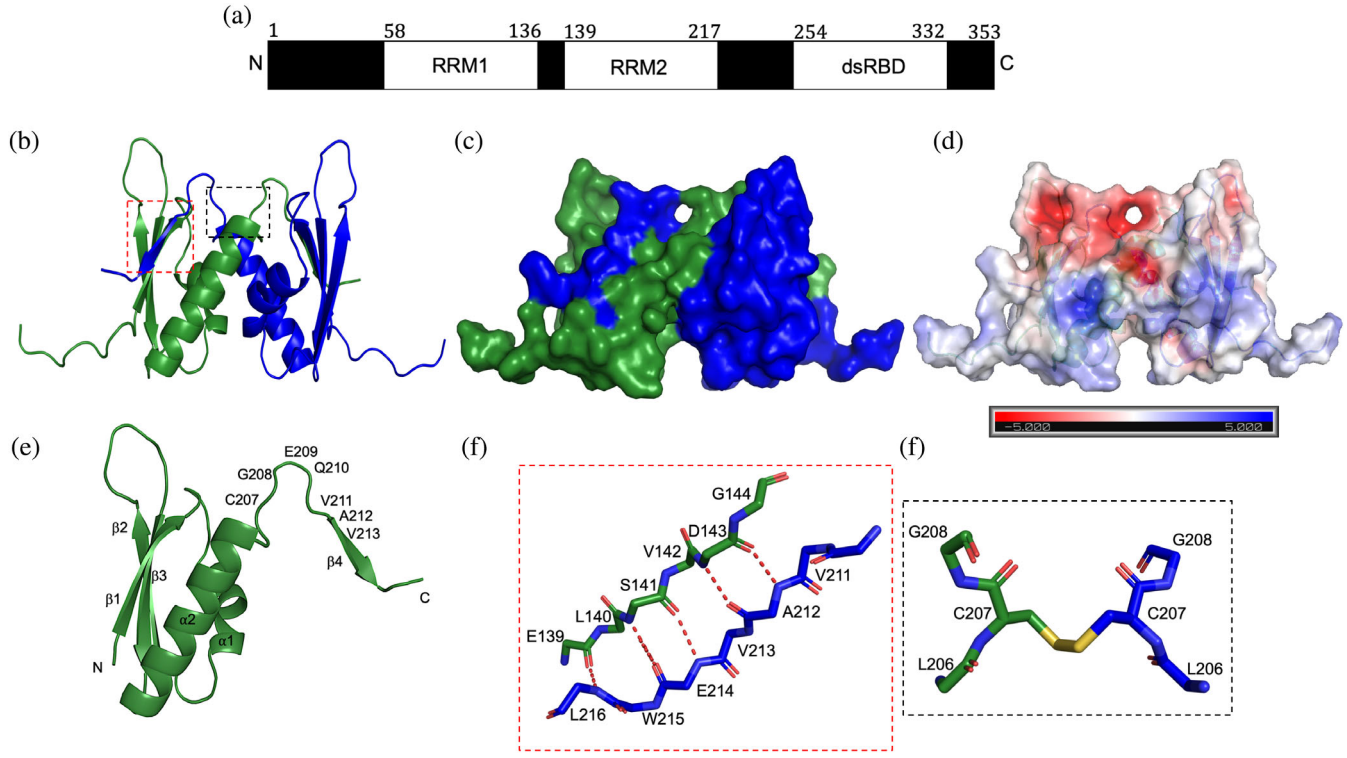

FIGURE 1 (a) Domain organization of DND1 protein showing its domains and their boundaries. (b) Overall structure in ribbon representation showing two chains in blue and green showing domain swapped dimer formation. (c) Space-filling model of the DND1-RRM2 domain-swapped dimer is shown to highlight the quaternary arrangement of the monomers. (d) Electrostatic potential surface generated by the APBS method as implemented in PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC showing charge polarity. (e) Structure of open monomer with labeled residues in hinge loop and secondary structure. (f) The backbone trace of the exchanged region is shown, and red dashed lines show the hydrogen bonds formed between the two chains. (g) Disulphide bond between two cysteine residues from chain A and chain B shown by yellow solid line

DND1-RRM2 domain crystallized with one molecule in the asymmetric unit and four copies in the unit cell. The molecule in the asymmetric unit had continuous electron density from N-terminus to C-terminus except for the initial 12 residues comprising the N-terminus hexa-histidine tag. Data processing and structure refinement statistics are given in Table 1. Structural analysis revealed a striking non-canonical feature where complete RRM motif comprising of four β sheets and two α helices forming a β_1-α_1-β_2-β_3-α_2-β_4 fold was maintained by the formation of a globular dimer structure (Figure 1(b,c)) (PDB id 6LE1). A domain swapped homodimer between β_1-strand of one monomer and β_4-strand of another monomer of DND1-RRM2 resulted in the formation of a non-canonical RRM fold. A hinge loop between α_2-helix and β_4-strand mediates domain swap resulting in formation of an anti-parallel β-sheet across the chains. The solvent accessible area and buried surface area for the protein were 7,255.7 and 2000.6 Å^2 with solvation energy of −55.2 kcal/mol. Domain-swapped dimerization is one way by which surface exposed and strained hydrophobic residues in the loop can be stabilized. The dimer assembly comprising of two RRM2 molecules had a surface area of 10,989.8 Å^2 and buried area of 4,330.9 Å^2. Free energy of solvation gained upon the formation of the domain swapped dimer structures is −16.6 Kcal/mol. The examination of surface electrostatic potential shown in Figure 1(d) revealed a bias in polarity with strong negative charge near the apical loops while a strong positive charge near the basal loop regions shown in Figure 1(d). The domain swapped anti-parallel β-sheet formation is stabilized by hydrogen bonds between E139 to D143 in β_1 and A212 to L216 in β_4 and disulphide bond between C207 across the protomers shown in Figure 1(f,g). Complete list of all the interactions and their distance between the two chains is given in Table S1.

Given the dimeric domain organization of the protein in the crystal, we set out to define whether this dimeric association holds true in solution also. We concentrated the purified protein to 1 mM and ran it on GE Healthcare 16/60 Hiload Superdex 75 column calibrated with five standard proteins and found two peaks corresponding to monomer as well as dimer population (Figure 2(c)). Therefore, the RRM2 dimer association observed in the

TABLE 1 Data processing and structure refinement statistics. Numbers in parentheses denote the highest-resolution shell

Space group	I 4_1
Unit cell parameters (Å, °)	$a = 44.62$, $b = 44.62$, $c = 111.33$ $\alpha = \beta = \gamma = 90$
Matthews coefficient ($Å^3Da^{-1}$)	2.64
Solvent content (%)	53.41
Data-collection temperature (K)	100
Detector	Pilatus 6 M
Wavelength (Å)	0.972
Resolution limits (Å):	31.57–2.3 (2.34–2.30)
Total reflections	29,990 (1616)
Unique reflections	4,476 (237)
Data completeness %	92.2 (100)
R_{merge}[a]	0.092 (0.52)
{I/σ(I)}	15.7 (3.7)
Anomalous completeness	89.4 (95.2)
Multiplicity	6.7 (6.8)
Anomalous multiplicity	3.5 (3.6)
$CC_{½}$	0.997 (0.921)
Wilson B-factor (Å2)	32.0
Mosaicity (°)	0.2
Refinement statistics	
R-value work	0.21
R-value free	0.24
B factor	34.0
Ramachandran statistics	
Preferred region	98.85%
Allowed region	1.15%
Outliers	0
Data deposition	
PDB id	6LE1

[a]$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i^{th} observation of reflection hkl and $\langle I(hkl)\rangle$ is the average intensity of the i^{th} observations.

crystal was also present in solution, instead of being a crystallization artifact. Since the structure showed formation of a disulphide bond between C207 residues in RRM2 domain, we explored its role in dimer formation. We observed that conditions in which 5 mM β-mercaptoethanol was present, showed only monomer population while in its absence, the protein showed two populations comprising of the monomer as well as dimer (Figure 2(a)). This result was further validated by non-reducing SDS-PAGE where the loading dye lacked β-mercaptoethanol, the monomer and dimer could be distinguished as seen in Figure 2(d). Hence, the monomer to dimer switch is being restricted by adding β-mercaptoethanol. This could be possible, when the domain in its monomeric form has exposed Cys bound β-mercaptoethanol (similar to cases in PDB ids 2RDP and 4TN4). However, at a higher protein concentration the β-mercaptoethanol bound protein after attaining open conformation experiences steric hindrance. This would not allow the formation of the domain swapped dimer. To investigate the dependence of this monomer dimer switch on the cysteine residue, we mutated cysteine residue to serine C207S and found both monomer and dimer shown in Figure 2(b). Hence, we concluded the disulphide bond formation is not the only determinant of monomer dimer switch. Although, it increases the propensity of dimer formation by disulphide bond formation between the protomers.

In order to explore the concentration dependence of dimer formation, we performed thermal shift assay where the protein was incrementally heated from 25 to 99 °C. As the protein unfolds, the hydrophobic core is exposed providing nonpolar region for the fluorescent dye to interact. The melting curve of protein is shown in Figure S5(a,b) at different concentrations harboring both populations in different ratios that is, monomer and dimer. The concentrations of recombinant protein used were 0.1 mM (blue), 1 mM (green), 2 mM (red). The protein at 0.1 mM concentration showed one melting peak while two peaks were observed in the case of 1 and 2 mM concentrations. The thermal denaturation profile indicates a well-folded protein with a low initial fluorescence at room temperature and as the temperature increases, it showed cooperative unfolding with high fluorescence giving rise to a sigmoid graph. The dimer unfolds first as it is supported by fewer bonds when compared to monomer as observed in the structure. The observed T_m was ~50 °C, which indicates that the overall domain is stable. We also examined C207S for dimer formation at different protein concentrations and found that the protein with serine residue also forms domain swapped dimer (Figure S5(c,d)).

2.3 | DND1-RRM2 forms closed monomeric structure

Structural analysis of the monomer was also performed to address whether the monomer had a closed structure where β1 and β4 are folded on to itself opposed to the open structure as observed in the crystal structure described in this study. Sequence-specific backbone resonance assignment of the RRM2 domain was achieved using standard triple-resonance 3D NMR spectra

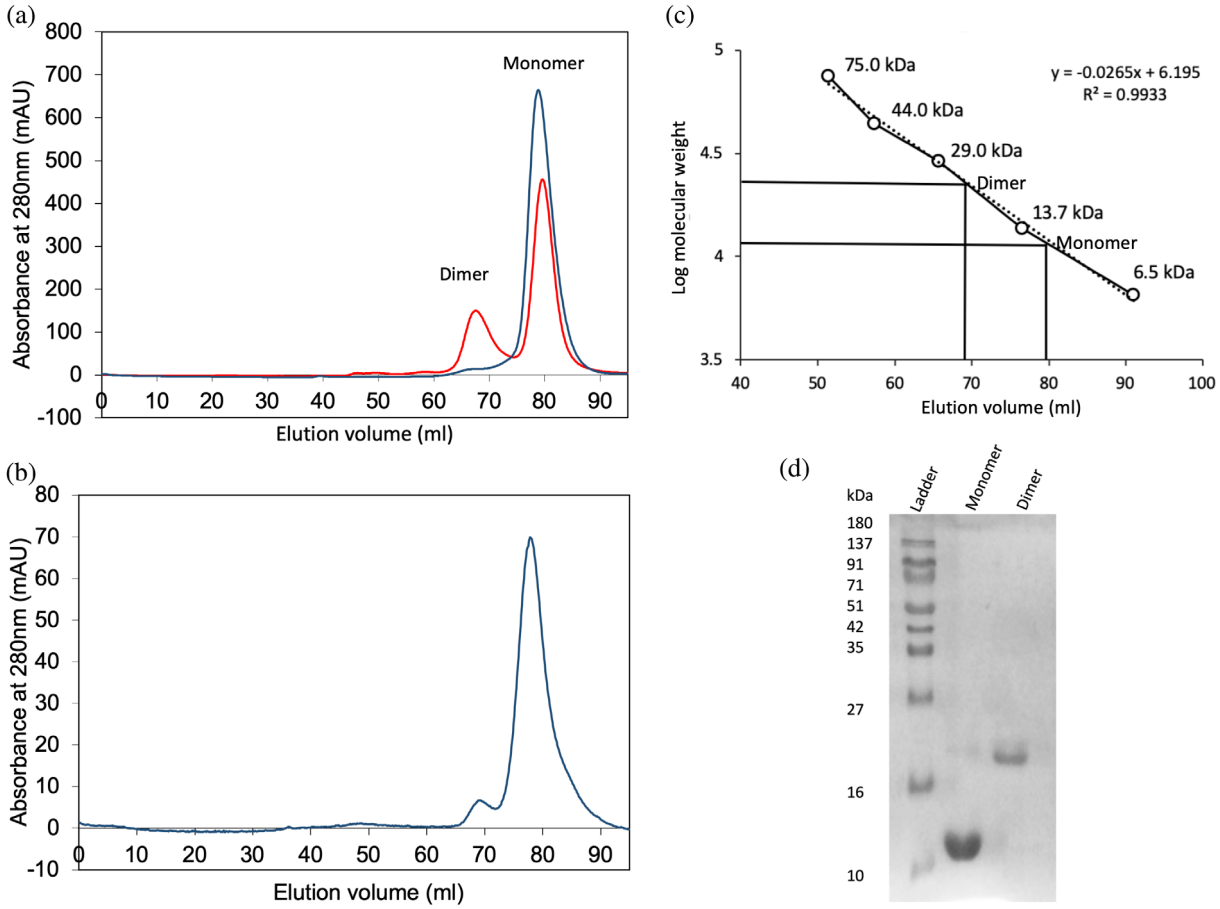

FIGURE 2 (a) SEC profile of DND1-RRM2 in presence of 5 mM β-mercaptoethanol showing only monomer in blue and SEC profile of DND1-RRM2 in absence of β-mercaptoethanol showing both monomer and dimer population in solution in red. (b) SEC profile of DND1-RRM2 C207S mutant showing both dimer and monomer population. (c) Plot of the logarithm of molecular weight versus elution volume generated using the elution profile of the calibration mix and the apparent molecular weights of monomer and dimer (d) Non-reducing SDS-PAGE showing the protein ladder in lane 1, monomer in lane 2, and dimer in lane 3

allowing 93% of backbone assignment excluding prolines and hexa-histidine tag. 2D [^{15}N,^{1}H] HSQC spectrum with labeled assigned resonance peaks is shown in Figure S6(a). 3D ^{15}N-edited [^{1}H,^{1}H]-NOESY (mixing time = 120 ms) spectrum measured on *U*-^{15}N labeled protein was used to examine the monomer structure. We observed across the strands *N,N*(*i,j*) nOe cross peaks between S141 on E214 and D143 on A212 in the 3D ^{15}N-edited [^{1}H,^{1}H]-NOESY spectrum conforming the closed monomeric conformation (Figure S7). CS Rosetta was used for backbone chemical shift-based structure prediction of monomer, which revealed that it was indeed a closed structure (Figure S8(a)). Structure alignment of closed monomer and open monomer chain showed an RMSD of 11.225 Å (Figure S8(b)), which is quite large indicating high conformational flexibility. For a closed monomer to open, it would require many interactions to disrupt and overcome a large kinetic energy barrier to form a dimer. For the monomer to open, conformational flexibility should be observed in the loop region and the exchanged β sheets. For this purpose, *U*-^{15}N labeled protein was purified and both monomer and dimer population were collected separately after size exclusion chromatography. 2D [^{15}N,^{1}H] HSQC spectra were acquired for both the species, which provided another evidence of existence of both the monomer and the dimer in solution state. Most of the resonance peaks in the 2D [^{15}N,^{1}H] HSQC spectra of monomer and dimer overlapped, which could be attributed to the similar structural fold in both. However, a few resonance peaks of dimer (those of the residues belonging to the β1, and β4 strands; α2–β4 and α1–β2 loops) did not overlap with those in the 2D [^{15}N, ^{1}H] HSQC spectrum of monomer indicating plausible structural differences in those regions. We also observed increased line-width of resonances in the 2D [^{15}N,^{1}H] HSQC spectrum for the dimer form as compared to monomer due to an increase in the rotational correlation time of the dimer. Several

resonances in the dimer spectra showed line broadening such as M138, L140, S141, G144, A193, K197, L199, V200, G202, Q203, C207, G208, and W215 (Figure S6(b)). The chemical shift perturbations (CSPs) for these were calculated by comparing spectra of monomer and dimer by increasing the contour level for the dimer to find the shifted resonances (Figure 3(a)). However, shifted resonance peaks corresponding to V200, Q203, C207, G208 were completely beyond detection and hence were excluded from CSP calculation. The chemical shift perturbation (CSPs) calculated after comparison of the backbone amide chemical shifts revealed that the residues involved in domain swapped dimerization. Several resonances belonging to β1 and β4 such as E139-G144 and V211-L216 showed high perturbation due to direct structural changes upon monomer-dimer switch marked in Figure 3(b). The resonances of amino acid residues like S204 at the start of α2–β4 loop showed maximum change while C207, and G208 amino acid residues within the hinge loop became broad beyond detection. The mean value of CSP was 0.034 ppm. Also, the residues belonging to loop between α1 and β2 such as L156, A157, L161, G162 showed perturbation which could occur due to indirect effects in spatial changes shown in Figure 3(c).

2.4 | Both monomer and domain swapped dimer are more stable than open monomer

Molecular dynamics simulations were performed for the closed monomer, the open monomer and the dimer for

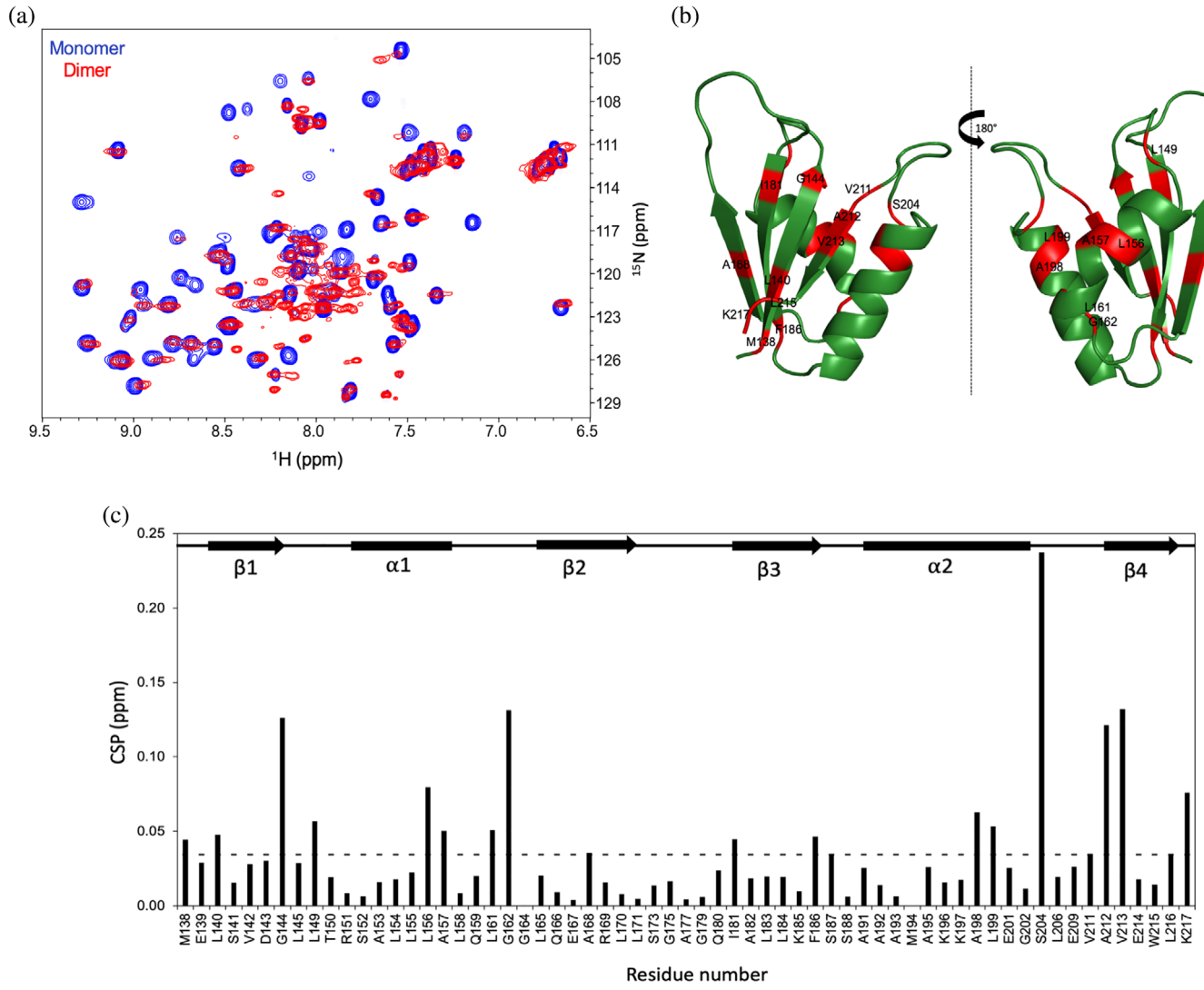

FIGURE 3 (a) Superposition of the 2D [^{15}N,^{1}H] HSQC spectra of the DND2 RRM2 in monomer and dimer states shown in blue and red, respectively; (b) Cartoon representation showing the residues having CSP above the mean value is marked in red; (c) CSP between monomer and dimer is shown for each residue with the mean value marked as dashed line, the secondary structural arrangement of RRM2 is indicated as arrows (strand) and lines (helix) on the top

200 ns with His-tag coordinates deleted from the structure in order to reduce unwanted fluctuations. The goal for these simulations was to understand protein flexibility, mobility, and degree of rotation in the loop between α2 and β4, in case of closed monomer, open monomer and dimer. The RMSD of all three trajectories got stabilized after 125 ns as shown in Figure 4. The RMSD-Time evolution (Figure 4(a–c)) showed that the RMSD value was quite high around 8.0 Å in the case of open monomer while it was around 3.6 Å in the case of closed monomer and domain swapped dimer. This indicated that the closed monomer and dimer conformation was much more stable compared to the intermediate open monomer conformation. The root mean square fluctuation (RMSF) values were computed for each residue in DND1-RRM2 domain to recognize flexible and constrained region of the protein useful for characterizing local changes along the protein chain. Local changes in these residues indicate flexible residues having higher RMSF compared to constrain region with lower RMSF (Figure 4(d–f)). RMSF value of each residue in DND1 fluctuates maximally in the loop region between β1 and α1 (G144-S153), β2 and β3 (P172-G179) in all three simulations. While the fluctuation in the hinge loop between α2 and β4 (S204-V211) and the succeeding residues varies within the simulations having low RMSF in the closed monomer, comparatively high in the dimer and highest in the open monomer. The α helix and β strands showed minimal fluctuations and were more rigid. The MD simulation studies on RRM2 support our claim of relatively higher dynamics in hinge loop attributing to domain swapped dimer formation, making it more stable in comparison to open monomer.

3 | DISCUSSION

Monomeric topology dictates domain swapping and temporary conditions such as high protein concentration or pH change allow a closed monomer to open.[22–24] The conformational flexibility depends upon hinge loop length, hinge loop residues, and solvent exposed surface area. The experimental evidence suggests that the length of the hinge loop if reduced leads to a more favored dimer conformation. On a contrary note, very long loop disfavors domain swapping as it allows higher conformational plasticity.[25–27] In the case of DND1-RRM2, an eight amino acid residue long hinge loop offers desired conformational flexibility. A recent study had shown that introduction of a hydrophobic five residue motif comprising the sequence QVVAG in the hinge loop leads to domain swapping irrespective of the position be it N-terminus, C-

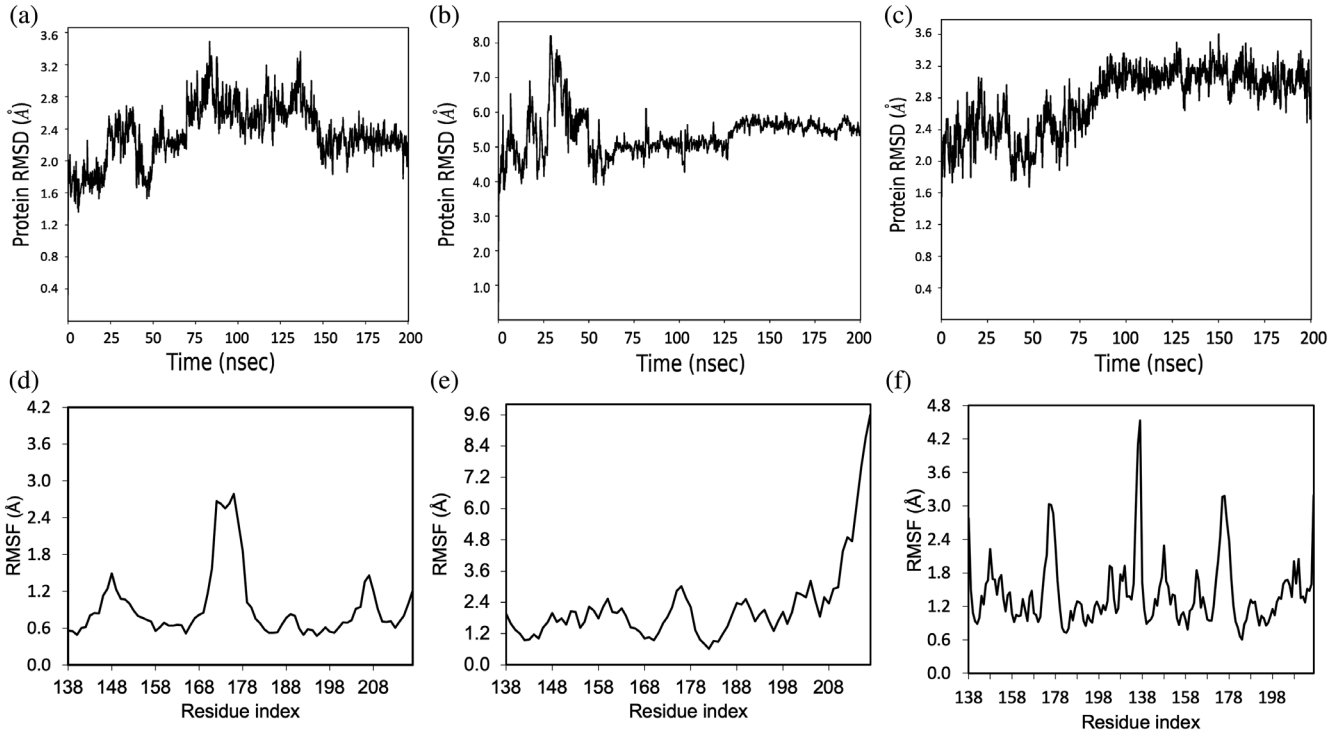

FIGURE 4 Root-mean-square deviation (RMSD) of Cα versus simulation time curves for all MD simulations (a) closed monomer, (b) open monomer, (c) domain swapped dimer. Protein root mean square fluctuation (RMSF) values of Cα along each residue in DND1-RRM2 domain (d) Closed monomer, (e) Open monomer, (f) domain swapped dimer

terminus or the middle of the protein.[27] The hinge loop between α2 and β4 in DND1-RRM2 comprises a similar surface exposed hydrophobic sequence GEQVAV shown in Figure 1(e), which allows the RRM2 domain to open and domain swap. Protein oligomerization imparts a larger binding surface, the possibility of allosteric control, and formation of active sites having diverse biological implications. It also acts as a regulatory mechanism for protein functions, and an evolutionary strategy to create protein interaction network and molecular machines.[19] Our results indicate that monomer to dimer switch is governed by several determinants such as hinge loop residues, its length, hydrophobic residues, concentration of protein, disulphide bond, and so forth.

Variations in the RRM fold as shown in Figure 5 depicts the extraordinary plasticity in the RRM domain of several proteins. N- and C-terminal extension in the form of additional α helix or β sheet allow recognition of additional nucleotides due to increase in binding surface as seen in PTB RRM3 where β5 strand is antiparallel to β2.[4] Loop mediated RRM-RNA interaction is also observed in case of Fox1 with conserved aromatic residues.[28] In case of pseudo RRM SRSF1, the interaction is mediated by conserved α-helix-1.[29] The RRM domain in GW182 protein lacks aromatic residues that preclude its RNA interaction possibility but the domain deletion impairs miRNA mediated silencing suggesting its role in protein interaction.[30] While our study shows the formation of domain swapped dimer and precludes canonical RNA recognition due to absence of conserved RNP sites. These variations in the scaffold as well as binding mechanism challenge the canonical RNP-centric view, where versatile mode of interaction with nucleotides or proteins is achieved by modifications or extensions in the domain to meet specific cellular regulatory requirement. Previous studies have reported 3D domain swapping possibly linked to protein aggregation and misfolding,[31,32] no such information about DND1 aggregation is yet known.

Full length DND1 protein from *Xenopus* has been shown to homodimerize *in vitro* and *in vivo* as well as it is also capable of forming heterodimers with specific subunits of the eIF3 complex. However, the smallest region required for homodimerization is mapped to residue 257–327 and differs from the region of domain swap

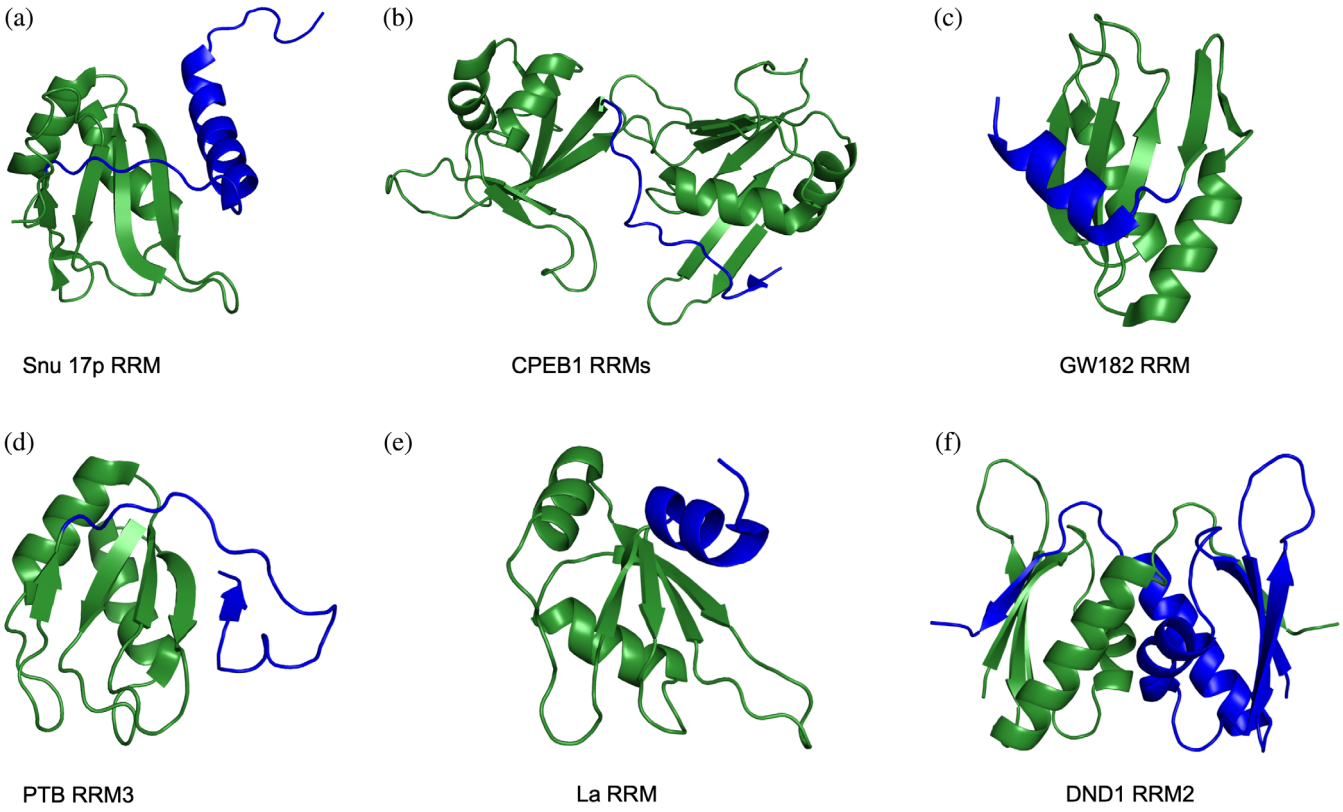

FIGURE 5 Few examples of non-canonical extension of RRM domain, canonical β1–α1–β2–β3–α2–β4 fold is shown in green while the non-canonical extension is shown in blue. (a) Snu 17p RRM with additional C-terminal α-helix (PDB id 2MKC), (b) CPEB1-RRM1with addition β strand (PDB id 2MKH), (c) GW182 RRM with additional C term α helix (PDB id 2WBR), (d) PTB RRM3 with addition β strand PDB ID-2ADC, (e) La RRM with additional C term α helix (PDB id 1S79), (f) DND1-RRM2 with domain swapped dimerization (PDB id 6LE1)

dimerization mentioned in this study.[21] While the biological relevance of dimerization in the functioning of human DND1 remains elusive, our results show the major determinant, which allowed 3D domain swapping in RRM2 domain, increasing the binding surface area, which may allow this protein to exhibit multifaceted role during post transcriptional regulation. Our study supports a model of DND1 architecture that could allow varied roles for its RRMs, where the domain swapped dimer could facilitate non-canonical interactions.

4 | MATERIALS AND METHODS

4.1 | Protein domain prediction and multiple sequence alignment

Domain enhanced lookup time accelerated BLAST (DELTA-BLAST)[33] of Human DND1 was performed and multiple sequence alignment was done using clustal omega[34] from protein sequences from different phyla of vertebrates obtained from uniport consisting of *Homo sapiens* Q8IYX4, *Mus musculus* Q6VY05, *Gallus gallus* A0A1L1RQS8, *Anolis carolinensis* G1KMZ1, *Xenopus laevis* Q6DCB7, *Danio rerio* Q7T1H5. RNP sites were recognized by alignment with well-known consensus sequence.[20]

4.2 | Cloning, overexpression, and purification

Human *DND*1 (UniprotID Q8IYX4) coding sequence corresponding to RRM12 (56–217) was codon optimized for overexpression in *Escherichia coli* and chemically synthesized from GeneArt (Life Technologies). Sub-clones corresponding to the RRM1 (residues 56–136) and RRM2 (139–217) and the combined construct (56–217) were prepared. pET28a expression vector utilizing *Nde*I and *Xho*I restriction sites were used for sub-cloning of all three constructs. The plasmid vector containing the desired construct was transformed in *E. coli* BL21(DE3) *CodonPlus* cells and was used for protein overexpression having a thrombin-cleavable hexa histidine tag. The overexpression was done in *E. coli* BL21 (DE3) codon plus cells at 37/25 °C for 16–20 h using 0.5 mM IPTG at $A_{600} \sim 0.6$–0.8. Purification of the overexpressed protein containing hexa-histidine tag was done by nickel-nitrilotriacetic acid (Ni-NTA) affinity resin, SP Sepharose cation exchange column followed by gel permeation chromatography on GE Healthcare 16/60 Hiload Superdex 75 in buffer consisting of 25 mM Tris–HCl (pH-8.0), 150 mM NaCl, 5 mM β-mercaptoethanol. L-Selenomethionine labeled protein was subsequently overproduced in selenomet base medium plus nutrient mix media supplemented with 40 mg/L L-Selenomethionine. *U*-^{15}N labeled and *U*-^{15}N and ^{13}C labeled-protein was prepared by over-expressing the protein in minimal M9 medium containing 1 g/L $^{15}NH_4Cl$ and 6 g/L glucose (for *U*-^{15}N-labeled protein) or 1 g/L $^{15}NH_4Cl$ and 2.5 g/L ^{13}C-glucose (for *U*-^{15}N- and ^{13}C-labeled protein).

Primers	Sequence
DND1-RRM2-F-NdeI	5′ ATTCCATATGGAACTGAGCGTT GATGGTCTG 3′
DND1-RRM2-R-XhoI	5′ CCGCTCGAGTTTCAGCCATTCA ACTGCAAC 3′

4.3 | Crystallization, data collection, and structure determination

Protein was concentrated to 30 mg/ml which is ~3 mM and crystallized by hanging drop method at 16 °C in Hampton crystal structure screen G2 condition consisting 30% (wt/vol) PEG-MME 5000, 0.1 M MES monohydrate pH -6.5, 0.2 M ammonium sulfate and then with 100 mM of KCl from Hampton additive screen, crystals were flash-frozen after a brief soaking in mother liquor supplemented with 20% glycerol. X-ray diffraction data were collected at 100 K on beamline ID23A at a wavelength of 0.972 Å with a Pilatus_6M_F detector at the European Synchrotron Radiation Facility (ESRF), Grenoble. Data were processed by autoPROC.[35] Structure was determined by the single wavelength anomalous diffraction (SAD) method using autosol and autobuild in Phenix.[36] The model was refined by iterative cycles of manual model building in Coot.[37] Refinement progress was monitored by tracking the R_{work}/R_{free} ratio (with R_{free} representing 5% of total reflections).[38,39] Domain swapped structure, surface area and stabilization energy was calculated using PDBePISA.[40] Crystallographic data collection and structure refinement statistics are presented in Table 1. Structural images were prepared with the PyMOL Molecular Graphics System, version 2.0 (Schrödinger, LLC).

4.4 | NMR spectroscopy

All NMR measurements were performed in 50 mM sodium phosphate pH 7.0 buffer containing 100 mM NaCl and 10% (vol/vol) D_2O measured on a 500 MHz Bruker Avance III spectrometer equipped with 5 mm TCI

cryo-probe at 298 K. Data were processed using Topspin 3.6 pl7 (Bruker). Sequence-specific backbone assignments of the protein were achieved using a set of standard double- and triple-resonance spectra[41] namely 2D [^{15}N,^{1}H] HSQC, 2D [^{13}C, ^{1}H] HSQC, 3D HNCA, 3D HNCO, 3D HNcaCO, 3D CBCAcoNH, 3D HNCACB, 3D hCcconNH, 3D HcccoNH, 2D [^{13}C,^{1}H]-HSQC [aliphatic (−1 to 5 ppm $^{1}H^{ali}$), aromatic (4.7–10 ppm $^{1}H^{aro}$)] and 3D ^{15}N [^{1}H,^{1}H]-NOESY (NOESY mixing time 120 ms). Protein backbone resonances were manually assigned using Computer Aided Resonance Assignment (CARA)[42] software with ^{1}H, ^{13}C, and ^{15}N shifts referenced indirectly to the DSS methyl proton resonance at 0 ppm in all spectra. Chemical shift based monomer structure prediction was done using CS-Rosetta.[43] Changes in chemical shifts (chemical shift perturbations, $\Delta\delta$) of backbone amide protons were tracked on 2D [^{15}N,^{1}H] HSQC spectra recorded for monomer and dimer. CSPs were calculated using the equation.

$$\Delta\delta_{15N^H,H^N} = \sqrt{\left(\frac{\Delta\delta_{15N^H}}{5}\right)^2 + (\Delta\delta_{H^N})^2}$$

where $\Delta\delta_{H}{}^{N}$ and $\Delta\delta^{15}{}_{N}{}^{H}$ are the changes in backbone amide chemical shifts for $^{1}H^{N}$ and ^{15}N resonances, respectively.

4.5 | Isothermal titration calorimetry

The monomeric state was purified in 25 mM sodium phosphate pH 7.0 buffer containing 100 mM NaCl and 3 mM β-mercaptoethanol. ITC experiments were performed with 0.1 mM protein inside the ITC cell and 1 mM RNA in the syringe prepared in the same buffer on a GE MicroCal iTC$_{200}$ Calorimeter at 303 K. Fourty injections of 1 μl each were sequentially given containing RNA at 750 rpm mixing which gave heats that were further analyzed using Origin 7 software.

4.6 | Site directed mutagenesis

C207S mutation in DND1-RRM2 were introduced by PCR amplification of plasmids pET28a DND1-RRM2 with primers containing the desired mutation followed by Dpn1 treatment. The presence of the desired mutation was ascertained by sequencing. The resulting plasmids were introduced into *E. coli* BL21 (DE3) codon plus cells. Protein was purified as described above in buffer consisting 25 mM Tris–HCl pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol.

Primers	Sequence
Forward	5′ CAGAGCCATCTGAGTGGTGAACAGG 3′
Reverse	5′ CCTGTTCACCACTCAGATGGCTCTG 3′

4.7 | Molecular dynamics simulation

For molecular dynamics (MD) simulations the Desmond program[44] and OPLS3e[45] force field was used, with a simulation time set to 200 ns for closed monomer, open monomer and domain swapped dimer. For the simulation temperature was 300 K, pressure was 1.0325 bar, while cut off radius was set to 10 Å. The whole system was considered as isothermal–isobaric (NPT) ensemble class. TIP3P model[46] was used for modeling of the solvent. MD system consisted of one molecule of the protein placed into the cubic box. Input and output files in the case used were prepared on protein preparation wizard,[47] analyzed and visualized with Maestro[48] graphical user interface (GUI).

4.8 | Thermal shift assay

Melting temperature and concentration dependent dimer formation was tested using thermal shift assay for RRM2 protein in 25 mM TRIS pH 8.0 buffer containing 150 mM NaCl. Increasing concentration of protein was taken and mixed with 5X SYPRO Orange dye and real time PCR run was setup initially at 25 °C, ramping up in increments of 1 °C to a final temperature of 95 °C, data was analyzed on Bio-Rad CFX manager software.[49]

5 | ACCESSION CODES

The atomic coordinates and structure factors (PDB id 6LE1) have been deposited in the Protein Data Bank (http://wwpdb.org). Partial sequence-specific backbone resonance assignment of DND1-RRM2 has been deposited in the Biological Magnetic Resonance Bank (BMRB) with accession number 50151 (https://bmrb.io/).

ACKNOWLEDGMENTS

The study was supported by the ICGEB, New Delhi core funds. Pooja Kumari is a recipient of CSIR senior research fellowship. We thank Dr. Alexander Popov (ID23-1 beamline, ESRF) for help with X-ray diffraction data collection. X-ray diffraction data collection at ID23 structural biology beamline was facilitated by the European Synchrotron Radiation Facility (ESRF)

Grenoble, France Access Program, which is supported by Department of Biotechnology (DBT), Government of India (BT/INF/22/SP22660/2017). The authors wish to thank ICGEB, New Delhi and DBT for providing financial support for the high field NMR spectrometers at the ICGEB, New Delhi. The Department of Biotechnology, Government of India funded the ITC instrument through grant number BT/PR13018/BRB/10/731/2009 to Neel Sarovar Bhavesh.

AUTHOR CONTRIBUTIONS

Neel Sarovar Bhavesh: Conceptualization; investigation; methodology; resources; supervision; validation. **Pooja Kumari:** Conceptualization; investigation; methodology.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

ORCID

Pooja Kumari https://orcid.org/0000-0002-9486-6723
Neel Sarovar Bhavesh https://orcid.org/0000-0002-7248-4978

REFERENCES

1. Lunde BM, Moore C, Varani G. RNA-binding proteins: Modular design for efficient function. Nat Rev Mol Cell Biol. 2007;8:479–490.
2. Kinzy TG, De Stefano LA, Esposito AM, et al. A birth-to-death view of mRNA from the RNA recognition motif perspective. Biochem Mol Biol Educ. 2008;36:1–8.
3. Maris C, Dominguez C, Allain FHT. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J. 2005;272:2118–2131.
4. Afroz T, Cienikova Z, Cléry A, Allain FHT. One, two, three, four! How multiple RRMs read the genome sequence. Struct Large RNA Mol Compl. 2015;558:235–278.
5. Weidinger G, Stebler J, Slanchev K, et al. Dead end, a novel vertebrate germ plasm component, is required for zebrafish primordial germ cell migration and survival. Curr Biol. 2003;13:1429–1434.
6. Zechel JL, Doerner SK, Lager A, Tesar PJ, Heaney JD, Nadeau JH. Contrasting effects of Deadend1 (Dnd1) gain and loss of function mutations on allelic inheritance, testicular cancer, and intestinal polyposis. BMC Genet. 2013;14:54.
7. Youngren KK, Coveney D, Peng X, et al. The Ter mutation in the dead end gene causes germ cell loss and testicular germ cell tumours. Nature. 2005;435:360–364.
8. Northrup E, Zschemisch NH, Eisenblätter R, et al. The ter mutation in the rat Dnd1 gene initiates gonadal teratomas and infertility in both genders. PLoS One. 2012;7:e38001.
9. Noguchi T, Noguchi M. A recessive mutation (ter) causing germ cell deficiency and a high incidence of congenital testicular teratomas in 129/Sv-ter mice. J Natl Cancer Inst. 1985;75:385–392.
10. Sijmons RH, Vos YJ, Herkert JC, et al. Screening for germline DND1 mutations in testicular cancer patients. Fam Cancer. 2010;9:439–442.
11. Linger R, Dudakia D, Huddart R, et al. Analysis of the DND1 gene in men with sporadic and familial testicular germ cell tumors. Genes Chromosomes Cancer. 2008;47:247–252.
12. Liu X, Wang A, Heidbreder CE, et al. MicroRNA-24 targeting RNA-binding protein DND1 in tongue squamous cell carcinoma. FEBS Lett. 2010;584:4115–4120.
13. Yamaji M, Jishage M, Meyer C, et al. DND1 maintains germline stem cells via recruitmentof the CCR4 – NOT complex to target mRNAs. Nat Publ Gr. 2017;543:568–572.
14. Kedde M, Strasser MJ, Boldajipour B, et al. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell. 2007;131:1273–1286.
15. Webster MW, Stowell JA, Passmore LA. RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-not. Elife. 2019;8:e40670.
16. Suzuki A, Niimi Y, Shinmyozu K, Zhou Z, Kiso M, Saga Y. Dead end1 is an essential partner of NANOS2 for selective binding of target RNAs in male germ cell development. EMBO Rep. 2016;17:37–46.
17. Niimi Y, Imai A, Nishimura H, et al. Essential role of mouse dead end1 in the maintenance of spermatogonia. Dev Biol. 2019;445:103–112.
18. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46:296–303.
19. Liu Y, Eisenberg D. 3D domain swapping: As domains continue to swap. Protein Sci. 2002;11:1285–1299.
20. Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. Science. 1994;265:615–621.
21. Aguero T, Jin Z, Owens D, et al. Combined functions of two RRMs in dead-end1 mimic helicase activity to promote nanos1 translation in the germline. Mol Reprod Dev. 2018;85:896–908.
22. Yang S, Cho SS, Levy Y, et al. Domain swapping is a consequence of minimal frustration. Proc Natl Acad Sci U S A. 2004;101:13786–13791.
23. Yang S, Levine H, Onuchic JN. Protein oligomerization through domain swapping: Role of inter-molecular interactions and protein concentration. J Mol Biol. 2005;352:202–211.
24. Bennett MJ, Schlunegger MP, Eisenberg D. 3D domain swapping: A mechanism for oligomer assembly. Protein Sci. 1995;4:2455–2468. https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.5560041202
25. Rousseau F, Schymkowitz J, Itzhaki LS. Implications of 3D domain swapping for protein folding, misfolding and function. Adv Exp Med Biol. 2012;747:137–152.
26. Ogihara NL, Ghirlanda G, Bryson JW, Gingery M, DeGrado WF, Eisenberg D. Design of three-dimensional domain-swapped dimers and fibrous oligomers. Proc Natl Acad Sci U S A. 2001;98:1404–1409.
27. Nandwani N, Surana P, Negi H, et al. A five-residue motif for the design of domain swapping in proteins. Nat Commun. 2019;10:452.
28. Clery A, Allain, FHT. From structure to function of RNA binding domains - RNA binding proteins. ZJ Lorkovic RNA Binding Proteins. 2013;137–158. Boca Raton: CRC Press.
29. Clery A, Sinha R, Anczukow O, et al. Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a

noncanonical mode of RNA recognition. Proc Natl Acad Sci U S A. 2013;110:2802–2811.
30. Eulalio A, Tritschler F, Büttner R, Weichenrieder O, Izaurralde E, Truffault V. The RRM domain in GW182 proteins contributes to miRNA-mediated gene silencing. Nucleic Acids Res. 2009;37:2974–2983.
31. Jaskólski M. 3D domain swapping, protein oligomerization, and amyloid formation. Acta Biochim Pol. 2001;48:807–827.
32. Bennett MJ, Sawaya MR, Eisenberg D. Deposition diseases and 3D domain swapping. Structure. 2006;14:811–824.
33. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. Biol Direct. 2012;7:12.
34. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. Mol Syst Biol. 2011;7:539.
35. Vonrhein C, Flensburg C, Keller P, et al. Data processing and analysis with the autoPROC toolbox. Acta Cryst D. 2011;67:293–302.
36. Adams PD, Afonine PV, Bunkóczi G, et al. PHENIX: A comprehensive python-based system for macromolecular structure solution. Acta Cryst D. 2010;66:213–221.
37. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. Acta Cryst D. 2004;60:2126–2132.
38. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. Acta Cryst D. 2011;67:235–242.
39. Kovalevskiy O, Nicholls RA, Long F, Carlon A, Murshudov GN. Overview of refinement procedures within REFMAC 5: Utilizing data from different sources. Acta Cryst D. 2018;74:215–227.
40. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007;372:774–797.
41. Bax A, Grzesiek S. Methodological advances in protein NMR. Acc Chem Res. 1993;26:131–138.
42. Anon Keller RLJ. The computer aided resonance assignment tutorial. Goldau: CANTINA Verlag, 2004 CARA website. http://cara.nmr-software.org/portal/.
43. Shen Y, Lange O, Delaglio F, et al. Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci U S A. 2008;105:4685–4690.
44. Anon Schrödinger Release 2019–4: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2019. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2019.
45. Harder E, Damm W, Maple J, et al. OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. J Chem Theory Comput. 2016;12:281–296.
46. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J Interaction models for water in relation to protein hydration. 1981. pp. 331–342.
47. Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013;27:221–234.
48. Anon Maestro Schrödinger Release 2019–4: Maestro, Schrödinger, LLC, New York, NY, 2019.
49. Huynh K, Partch CL. Current protocols in protein science: Analysis of protein stability and ligand interactions by thermal shift assay. Curr. Protoc. Protein Sci. 2015;79:28.9.1–28.9.14.

SUPPORTING INFORMATION

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

How to cite this article: Kumari P, Bhavesh NS. Human DND1-RRM2 forms a non-canonical domain swapped dimer. *Protein Science*. 2021;30:1184–1195. https://doi.org/10.1002/pro.4083