



Conserved linker length in double dsRBD proteins from plants restricts interdomain motion

Florencia C. Mascali^a, Roberta Crespo^a, Leandro C. Tabares^c, Rodolfo M. Rasia^{a,b,*}

^a Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET-UNR), Rosario 2000, Santa Fe, Argentina

^b Área Biofísica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario 2000, Santa Fe, Argentina

^c Institute for Integrative Biology of the Cell (I2BC), Department of Biochemistry, Biophysics and Structural Biology, Université Paris-Saclay, CEA, CNRS UMR 9198, F-91198 Gif-sur-Yvette, France

ARTICLE INFO

Keywords:

dsRBD
Paramagnetic relaxation enhancement
DEER
Protein linkers

ABSTRACT

Double stranded RNA binding domains (dsRBDs) are ubiquitous in all kingdoms of life. They can participate both in RNA and protein recognition and are usually present in multiple copies in multidomain proteins. We analyzed the linkers between dsRBDs in different proteins and found that sequences corresponding to plant proteins have a highly conserved linker length. In order to assess the importance of linker length in the conformational freedom of double dsRBD plant proteins, we introduced lanthanide binding tags (LBTs) in different positions of the dsRBD containing protein HYL1 from *Arabidopsis thaliana*. These constructs were used to obtain conformational restraints from Double electron–electron resonance (DEER) measurements on doubly labeled proteins and from paramagnetic relaxation enhancement (PRE) in single labeled samples. Fitting the experimental datasets to a computational model of the ensemble created by allowing freedom to the linker region we found that the domains tend to explore a particular region of the allowed conformational space. The high conservation in linker length suggests that this restricted conformational sampling is functional, possibly hindering HYL1-dsRBD2 from contacting the substrate dsRNA and allowing it to participate in protein-protein interactions.

1. Introduction

Multidomain proteins are ubiquitous in nature. Domain shuffling and duplication have helped organisms to achieve a wide range of functions by combining a limited number of individual modules [1]. The number, kind and order of domains, or protein architecture, defines the function of a protein. The domains in multidomain proteins are often connected by sequences of amino acids of varying length and rigidity. Although these linker sequences can be thought of as simple covalent anchors between domains, their structural properties can modulate the behavior of the protein as a whole. In some cases, rigid linkers can function as spacers, keeping a fixed distance between domains. Even if they do not have a defined structure, linker length and sequence can limit the available space allowed for inter-domain conformations [2].

The double stranded RNA binding domain (dsRBD) is widely distributed among eukaryotic, bacterial and viral proteins, where they are usually present in multiple copies, alone or combined with other functional domains [3]. Most dsRBDs bind dsRNA in a non-sequence-specific fashion, and owe their specificity to the

recognition of the three dimensional conformation of the RNA molecule [4]. Nevertheless, recent findings show that some dsRBDs are capable of not only recognizing the imperfect structure of particular dsRNA substrates, but show also a base-specific readout of the substrate [3,5]. Besides dsRNA recognition, dsRBDs have been shown to participate in protein-protein interactions [6–9].

Multiple dsRBDs are usually found arranged in tandem in multidomain proteins. In some cases, this architecture results in a higher affinity for their partner dsRNAs mediated by cooperative binding. In other cases, the presence of multiple dsRBDs in a protein allowed for functional divergence, where some dsRBDs acquired protein binding capabilities acting as inter- or intra-protein recognition modules. For example among the five dsRBD domains of Staufen, domains 1, 3 and 4 recognize RNA whereas domains 2 and 5 participate in the localization of the protein through protein:protein interactions [6].

Small interfering RNAs (siRNAs) and microRNAs (miRNAs) are 21 nt RNA molecules that regulate gene expression at a post transcriptional level. Small RNAs originate from longer, mostly double stranded RNA molecules, and are released by RNaseIII-type enzymes with the

* Corresponding author.

E-mail address: rasia@ibr-conicet.gov.ar (R.M. Rasia).

participation of dsRBD containing helper proteins [10]. In higher eukaryotes, the RNA binding protein DGCR8 interacts with the RNase III enzyme DROSHA to generate pre-miRNAs in the nucleus while the dsRBD containing proteins TRBP and PACT interact with the enzyme DICER in the processing of pre-miRNAs or long dsRNAs into the 21 nt fragments. Similarly, in plants, proteins from the DRB family interact with Dicer-like proteins, forming complexes that process an array of small RNAs [11–13]. The linker regions between dsRBDs in these helper proteins are highly variable. For example, in DGCR8 the linker region is actually structured and the two dsRBD domains form a compact fold [14], while the regions surrounding the first dsRBD of RDE-4 have helical structural elements but are mostly unstructured [15].

HYL1 is a protein that participates in miRNA biogenesis in plants as a part of a processing complex that includes the RNaseIII enzyme Dicer-Like 1 (DCL1) and the RNA binding protein Serrate (SE) [12,13]. Analysis of HYL1 sequence shows a domain architecture consisting of two dsRBDs followed by a long unstructured C-terminal tail that contains six tandem repeats of a 21 residues sequence [16]. The mutant phenotype of *hyl1*⁻ plants can be complemented with a construction containing only the two N-terminal dsRBDs, showing that the C-terminal tail is dispensable for its function in vivo [17]. The two dsRBD domains have different functions. While the first (D1) shows high affinity for dsRNA substrates, the second (D2) has only moderate dsRNA binding capacity [7,18]. High affinity substrate binding by D1 is essential for the protein activity in plants [19]. In turn D2 seems to participate in protein-protein interactions, mediating low affinity dimerization of HYL1 and its interaction with DCL1, the miRNA processing ribonuclease [7,8]. In HYL1 from *Arabidopsis thaliana* D1 and D2 are connected by a 17-residue linker. The NMR spectra of the individual domains and of a construction containing the two domains linked together are identical, showing that the two domains behave as independent units.

In the present work we studied how the linker between D1 and D2 of HYL1 restricts the mobility between both domains. Sequence analysis of linkers between double dsRBD plant proteins show a high conservation in linker length and some conservation in sequence. Inter domain distances measured by double electron–electron resonance (DEER) [20–22] on protein constructs with two Gd(III) loaded lanthanide binding tags (LBTs) show a wide distribution of distances. NMR paramagnetic relaxation enhancement (PRE) on single labeled protein samples shows an asymmetry in transient domain-domain interactions. By means of ensemble simulations and calculation of PREs we found that the domains explore a restricted conformational space. Altogether our results put forward the functional importance of the linkers between dsRBD modules in multidomain proteins.

2. Materials and methods

2.1. Analysis of linker sequences

The DRBs sequences and taxonomy were obtained from the UniProt Database. The sequences alignments were done with the program MEGA version 10.0.4 [23] with the default options of the software MUSCLE [24]. The phylogeny reconstruction was made through the Maximum Likelihood Statistical Method. The analysis was performed using custom Python scripts.

2.2. Construction of LBT containing variants of HYL1

A synthetic gene containing the sequence L-D1-L-D2-L (Supplementary Figure 1) was designed for the construction of plasmids expressing HYL1 with lanthanide binding tags (LBTs) in different positions. The LBT sequence corresponds to the sequence optimized for Gd(III) binding reported by Daughtry and coworkers [25]. The first fourteen residues of D1 do not have secondary structure, so they were excluded from the construct in order to reduce the flexibility in the N-terminal domain. We purchased the synthetic gene cloned in pUC57 into the NcoI/XhoI

restriction sites. The gene was first cloned into the same restriction sites of the pET-TEV vector [26]. The double LBT labeled proteins L-D1-L-D2 and L-D1-D2-L were prepared from the plasmid by digestion and ligation onto the same plasmid using the SalI and SacI restriction sites respectively. The last double LBT labeled protein gene was prepared by using the NdeI/XhoI restriction sites. The obtained NdeI/XhoI fragment was ligated into the same restriction sites of an empty pET-TEV vector to produce the D1-L-D2-L gene. The D1-L-D2 and D1-D2-L genes were prepared from the plasmids D1-L-D2-L and L-D1-D2-L respectively, using the previously described protocol. Finally, the L-D1-D2 gene was prepared from the plasmid L-D1-L-D2 by digestion and ligation onto the same plasmid using the SacI restriction site. All mutations were confirmed by DNA sequencing via the Maine's University service. Residue numbering for each protein used throughout the paper corresponds to the amino acid sequence of wild type HYL1.

2.3. Protein expression and purification

Expression plasmids were transformed in *E. coli* BL21(DE3) cells which were then grown at 37 °C in Erlenmeyer flasks shaken at 180 rpm in either M9 minimal medium supplemented with 1 g/l ¹⁵N-NH₄Cl or 1 g/l ¹⁵N-NH₄Cl and 2 g/l [U-¹³C]-Glucose (Cambridge Isotope Laboratories), or in LB broth. Protein expression was induced with 0.25 mM IPTG at OD₆₀₀ ≈ 0.7 and cells were incubated for an additional 16 h at 20 °C. The cells were collected by centrifugation at 4000 g for 10 min at 4 °C, followed by resuspension into a 20 ml solution containing 100 mM phosphate, 10 mM Tris, 5 mM β-mercaptoethanol, 8 M Urea pH 8. The suspension was disrupted by sonication and the soluble fraction was clarified by centrifugation for 1 h at 20,000 g. The unfolded proteins were purified using a Ni(II) column and refolded by dialysis in 100 volumes of 100 mM Pi, 50 mM NaCl, 50 mM arginine, 50 mM glutamic acid and 1 mM β-mercaptoethanol pH 7. The refolded proteins were dialyzed against 50 mM Tris pH 8.0, 500 mM NaCl, 1 mM β-mercaptoethanol and digested with His-tagged TEV protease. The digested proteins were further purified with a Ni(II) column. The final purification was carried out with a G75 size exclusion chromatography column equilibrated with 20 mM HEPES, 500 mM NaCl, 1 mM β-mercaptoethanol, pH 7.0.

Gd(III) and Lu(III) derivatives were prepared by mixing the LBT containing proteins with one equivalent of GdCl₃ or LuCl₃·6H₂O per LBT respectively.

2.4. RNA synthesis

The RNA construct was produced by *in-vitro* transcription with T7 RNA polymerase, using annealed oligonucleotides. Briefly, a mix was prepared containing 1X transcription buffer (Tris 40 mM pH8; DTT 5 mM; spermidine 1 mM; Triton-X100 0.01%; PEG 8000 80 mg/ml), 4 mM each rNTP (rA; rC; rG; rU), 20 mM MgCl₂, 40 μg/ml BSA, 1 unit pyrophosphatase, and the annealed template at 35 μg/ml. The reaction was started by addition of T7 RNA polymerase and allowed to proceed for 3 h at 37 °C. Then, 50 units of RNase-free DNase were added and the mix incubated further for 30 min at 37 °C. The reaction was then diluted eight-fold in 20 mM Tris, 10 mM EDTA, 8 M Urea, pH 8.0 and loaded on a Q-sepharose column equilibrated with the same buffer. The column was eluted with a gradient from 0 to 1 M NaCl in the same buffer. Fractions containing RNA, as determined by A₂₆₀, were checked on denaturing 5% PAGE. The fractions with the desired transcript were pooled, dialyzed 3 times against 200 volumes of H₂O and lyophilized for storage before use. The secondary structure of the RNA fragment was calculated using the UNAFold web server [27] (Supplementary Figure 6).

2.5. DEER measurements and data processing

LBT:Gd(III) DEER measurements were performed as described before

[22] using a 94 GHz Pulsed-EPR Bruker Elexsys II 680 EPR spectrometer equipped with a Bruker “power upgrade 2” and an Oxford Instruments CF935 flow cryostat. Measurements were carried out at 10 K using the standard four-pulse, dead-time-free DEER sequence [21] with a four-step two-phase cycle [28]. The dipolar evolution times were 3500 ns. For all measurements, a π -pump pulse of 18 ns was applied at a magnetic field of 15 G above the field position of the Gd(III) signal maxima and the spin-echo was detected 70 MHz below the frequency of the pump pulses using $\pi/2 = 14$ and $\pi = 28$ ns pulses. The pulse sequence was repeated 100 times per point with a recovery time of 550 μ s. The data shown in Fig. 2 corresponds to 160 to 200 averages for a total acquisition time of 4 to 6 h. For the experiments with HYL1 bound to a miRNA precursor, the protein samples were mixed with in vitro transcribed miR172a precursor in a 1:1 ratio. The DEER time traces were analyzed using Deer Analysis [29,30]. The data shown in Fig. 2C correspond to the DEER time traces after background correction. The distance distributions were obtained using a Tikhonov regularization factor of 250. The error was estimated using the DeerAnalysis validation tool.

2.6. NMR spectroscopy

HSQC NMR spectra were acquired at 298 K on a 700 MHz Bruker Avance III spectrometer equipped with a TXI probehead using pulse sequences from the standard Bruker library. All spectra were processed with NMRPipe [31] and analyzed with CCPNMR [32]. Assignments for the signals were obtained from the BMRB entries 17,141 (dsRBD2) and 17,143 (dsRBD1). The I/I₀ ratios were calculated from the signal intensities for each ¹⁵N-¹H crosspeaks in equivalent protein samples containing either Gd(III) (I) or Lu(III) (I₀).

2.7. HYL1 ensembles generation and PRE calculations

2.7.1. Generation of conformers for the ensemble

The initial 100 models for the LBT containing constructs were generated using Modeller v 9.19 [33]. We constructed 100 initial models using as templates the crystal structures for each domain of HYL1 (3ADG and 3ADJ for D1 and D2, respectively) and the crystal structure of a Ca (II) loaded LBT (PDB 1TJB). Then we selected the best model based on the DOPE parameter. Starting from this model we generated ensembles by producing new conformers using PyRosetta [34]. Briefly, a residue among those defined as movable is picked randomly. Then the Phi and Psi angles are randomly selected from a library of random coil values provided with the software Flexible Meccano [35]. The energy of the new structure is evaluated by measuring the *fa_{atr}* score. Only movements that do not increase this score or with an incrementation lower than a probability value are accepted. If no acceptable score is achieved after 100 times, the original phi and psi values are conserved for that structure. By doing this analysis clashing between different parts of the protein are avoided. The procedure is repeated until all available angles were randomized, and the modeled conformer is stored as a member of the ensemble.

2.7.2. PRE calculation

Experimentally, the effect of the paramagnetic ion was calculated from the ratio of the amide HN signal intensities obtained from ¹H-¹⁵N-HSQC spectra of samples loaded either with Lu(III) (I₀) or Gd(III) (I). This intensity ratio relates to the PRE:

$$\frac{I}{I_0} = \frac{R_2^0}{R_{2M}^{PRE} + R_2^0}$$

The R_2^0 values for each residue are the ¹⁵N T2 values measured on a diamagnetic sample using the hsqct2etf3gpsi3d pulse sequence from the standard Bruker library. In order to translate the measured values in geometric constraints, we obtained the expected values of I/I₀ for each

conformer in the ensemble. The paramagnetic relaxation enhancement induced by the LBT was calculated as follows [36]

$$R_{2M}^{PRE} = \frac{k'_{Solomon} + k'_{Curie}}{r^6}$$

where

$$k'_{Solomon} = \frac{(\mu_0)^2 \gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{15} \left[4\tau_c + \frac{\tau_c}{1 + (\omega_I - \omega_S)^2 \tau_c^2} + \frac{3\tau_c}{1 + \omega_I^2 \tau_c^2} + \frac{6\tau_c}{1 + (\omega_I + \omega_S)^2 \tau_c^2} + \frac{6\tau_c}{1 + \omega_S^2 \tau_c^2} \right]$$

$$k'_{Curie} = \frac{1}{5} \left(\frac{\mu_0^2}{4\pi} \right) \frac{\omega_I^2 g_e^2 \mu_B^4 S^2 (S+1)^2}{(3k_B T)^2} \left[4\tau_{Curie} + \frac{3\tau_{Curie}}{1 + \omega_I^2 \tau_{Curie}^2} \right]$$

Here ω_I is the nuclear Larmor frequency, ω_S is the electron Larmor frequency, g_e is the electron Landé factor, μ_B is the Bohr magneton, μ_0 the vacuum permeability, S the spin quantum number (7/2 for Gd(III)) and τ the correlation times related to the relaxation mechanisms:

$$\tau_c^{-1} = \tau_e^{-1} + \tau_r^{-1} + \tau_M^{-1}$$

$$\tau_{Curie}^{-1} = \tau_r^{-1} + \tau_M^{-1}$$

where τ_e is the electron relaxation time, τ_r the rotational correlation time and τ_M the exchange time. The value of τ_M was selected based on the expected time range of conformational fluctuation in the flexible linker region [37]. The resulting PRE effect is relatively insensitive to τ_M within the 2 – 50 ns range (Supplementary Figure 2). We used constant values of 5 ns for τ_r and τ_M and 2 ns for τ_e . With these values kept constant, the values of R_{2M}^{PRE} and of the I/I₀ ratio for each amide HN depend only on the distance between the nuclei and the paramagnetic ion.

2.8. Conformer clustering

In order to select a subensemble of conformers that fits the experimental data, we clustered the conformers based on the similitude of calculated I/I₀ profiles. The I/I₀ profiles for the domains directly bound to the LBT are fairly constant (see Section 3.3 in Results) and hence are not influenced by the conformational sampling of the domains with respect to each other. Therefore, the distance between two models was defined as the squared sum of the differences between the I/I₀ values of each residue for D1 (positions 18 – 88) considering the C-terminal Gd (III) and for the D2 (positions 104 – 174) considering the N-terminal Gd (III). The clustering processes with each distance matrix were done with the python library `scipy.cluster.hierarchy.linkage` in the “ward” mode.

Similarly, the agreement of the I/I₀ profile for different conformers with the experimental data was calculated as the squared sum of the differences between the calculated and experimental I/I₀ values for selected positions. The calculated values for D2 (positions 104 – 174) of L-D1-D2-L models considering the N-terminal Gd(III) was compared with the experimental values for D2 obtained with the L-D1-D2 construct. In the same way, the calculated values for D1 (positions 18 – 88) of L-D1-D2-L models with a C-terminal Gd(III) was compared with experimental values for D2 obtained with the D1-D2-L construct.

3. Results

3.1. Bioinformatic analysis of linker sequences

To obtain sequences of dsRBD containing proteins, we searched the database UniProt [38] with the query “dsRBD”. We obtained 50,188 proteins and recorded their sequence, domains position and taxonomy data. We kept sequences corresponding to metazoans and viridiplantae

harboring two or more dsRBDs for further analysis (Supplementary Table 1). We then measured the lengths of linker sequences between consecutive dsRBDs (Supplementary Figure 3). The animal proteins did not show any particular length conservation. In contrast, the plant proteins presented a high conservation of linkers length of 17 residues (Fig. 1A). The high number of sequences with 17-residue linkers could be related to biases in plant species representativity in the database. Counting the number of sequences from plant species shows that the linker length conservation is distributed among eudicotyledoneae (data not shown).

We then focused on proteins containing only two dsRBDs and 17 residues long linkers. A sequence logo of this group of sequences shows a high conservation in the final 4 residues of the linker, ETG(V/L/I) (Fig. 1B), suggesting that these residues are actually a part of the second dsRBD in these proteins, but are not detected by domain prediction algorithms.

All the plant double dsRBD proteins analyzed belong to the DRB family. The members of this family presumably participate in small RNA processing, interacting with different Dicer-Like proteins (DCLs) [39]. HYL1 is a member of this family, alternatively named as DRB1. We produced an alignment of all non-redundant proteins and sorted them in four groups, corresponding to DRB1 (88 sequences), DRB2 (100 sequences), DRB3–5 (126 sequences) and DRB6 (36 sequences), according to the classification by Deragon and coworkers [39]. The DRB4 sequences are absent from this analysis, as DRB4 linkers are 9 residues long, and the dsRBDs from these proteins show restricted mobility [40]. The linker sequence logos for these groups show a high conservation for DRB2, DRB3–5 and DRB6 linkers, but much lower for DRB1 (HYL1) linkers (Fig. 1B). In order to determine whether the sequence conservation would extend to the whole protein, we calculated the sequence identity for each dsRBD and the linker separately (Supplementary Figure 4). Remarkably, while the sequence identity is overall similar in the DRB2, DRB3–5 and DRB6 groups, the identity between linkers in DRB1 (HYL1) is much lower than the identity between domains. This suggests that length conservation is more functionally important than sequence or composition, at least in DRB1 (HYL1).

3.2. Interdomain interactions evidenced by PRE and DEER

In order to determine the space explored by the two dsRBD domains in HYL1 we engineered LBTs at each end of the protein. The use of lanthanides as tags in NMR experiments allows the measurement of long range restrictions due to their effect on the relaxation and chemical shifts of nuclei in a range of up to 20 nm and of the relative orientations of pair of atoms when the lanthanide possess an anisotropic magnetic susceptibility tensor [41]. The introduction of an LBT sequence allows the binding of lanthanides to the protein without the attachment of external chelating tags [42].

In all experiments we used Lu(III) as a diamagnetic reference and Gd

(III) as paramagnetic tag, as this latter ion gives rise to large PRE effects but has an isotropic magnetic susceptibility. The PRE patterns obtained are shown in Fig. 2B. Introduction of Gd(III) in the C-terminus of the protein gives rise to a clear PRE pattern in D2, that is consistent with the known structure. Some signals corresponding to D1 also show a weaker PRE effect, indicating the existence of short-lived long-range interactions between the C-terminal end of the protein and the N-terminal dsRBD. A PRE on the distal domain arises due to the existence of conformations where the lanthanide at the C-terminus is close to the N-terminal domain. These conformations are necessarily transient and not stable as indicated by the dynamic nature of the linker and the conformational independence of the domains within the whole protein [18]. Loading Gd(III) on the N-terminal LBT gives a different result. Whereas the PRE pattern on the proximal D1 is again consistent with the dsRBD structure, a larger global PRE effect is observed on the distal D2 domain.

A possible source for this effect is weak direct Gd(III) binding to D2, but PRE measurements carried on the protein without LBTs showed global broadening of signals with no localized PRE effect on HYL1, ruling out this possibility (Supplementary Figure 5A). Previous works have suggested the possibility of HYL1 dimerization, mediated by D2 [7–9]. The asymmetric behavior of the L-D1-D2 constructs could be due to this reported dimerization. To verify this, we performed PRE measurements on samples with ^{15}N -D1-D2 and ^{14}N -L[Gd(III)]-D1-D2 or ^{14}N -D1-D2-L[Gd(III)] (Supplementary Figure 5B). No selective PRE effect was observed, indicating that these constructs do not dimerize under our experimental conditions. These experiments indicate that the asymmetry in PRE profiles between L-D1-D2 and D1-D2-L is due to a restriction in interdomain motions. The relative location of the N- and C-termini of the dsRBDs also plays a role, as the reach of the Gd(III)-LBT differs depending on where it is located (Fig. 2A).

We then obtained direct distance restrictions by means of DEER experiments on samples with two LBTs (Fig. 2C). We produced constructs with LBTs located at the N- and C-termini or at the C-terminus of domain 1 (Supplementary Figure 1). The distance distribution obtained for the construct with LBTs flanking the D1 domain (L[Gd(III)]-D1-L[Gd(III)]-D2) has a define peak at 4.3 nm, in good agreement with the known structure. The D1-L[Gd(III)]-D2-L[Gd(III)] shows a distinct peak at 2.5 nm and a longer, less defined peak, with a maximum around 4.3 nm. The latter distance corresponds to the length of the domain plus LBTs as seen in the L[Gd(III)]-D1-L[Gd(III)]-D2 construct. The shorter one could arise from the LBTs coming close due to the flexibility conferred by the linker region intervening between the end of the LBT and the N-terminus of D2. For the construct with LBTs in both ends of the protein (L[Gd(III)]-D1-D2-L[Gd(III)]) the distance distributions obtained are broader, highlighting the conformational flexibility conferred by the linker.

The small amount of sample needed for DEER measurements allowed us to investigate the influence of dsRNA binding in the conformational sampling of the protein. Both domains of HYL1 can bind dsRNA, but whereas D1 shows high affinity, D2 shows a moderate affinity and

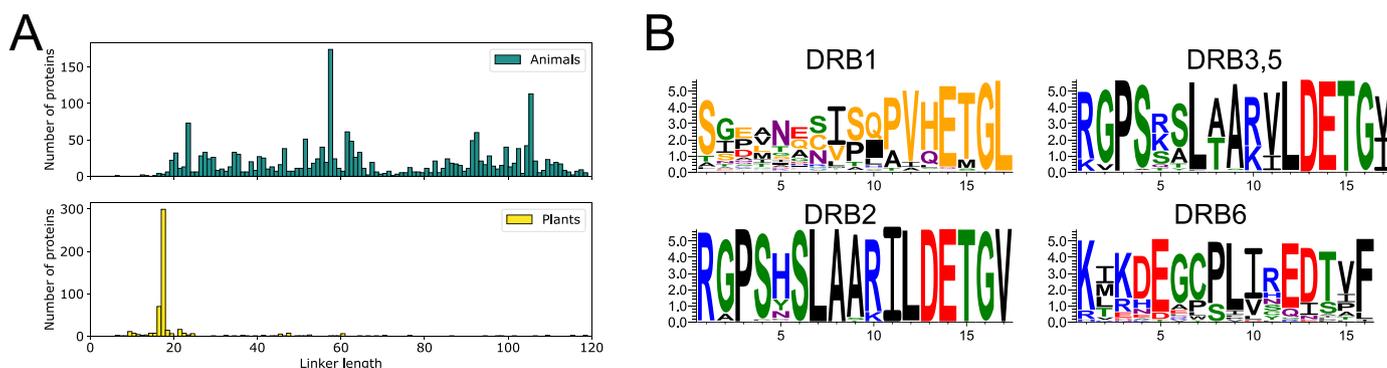


Fig. 1. A. Histogram of linker lengths in double dsRBD containing proteins. B. sequence logos for the linker regions in the DRB1, DRB2, DRB3,5 and DRB6 families. The sequence of *A. thaliana* HYL1 is colored yellow in the DRB1 logo.

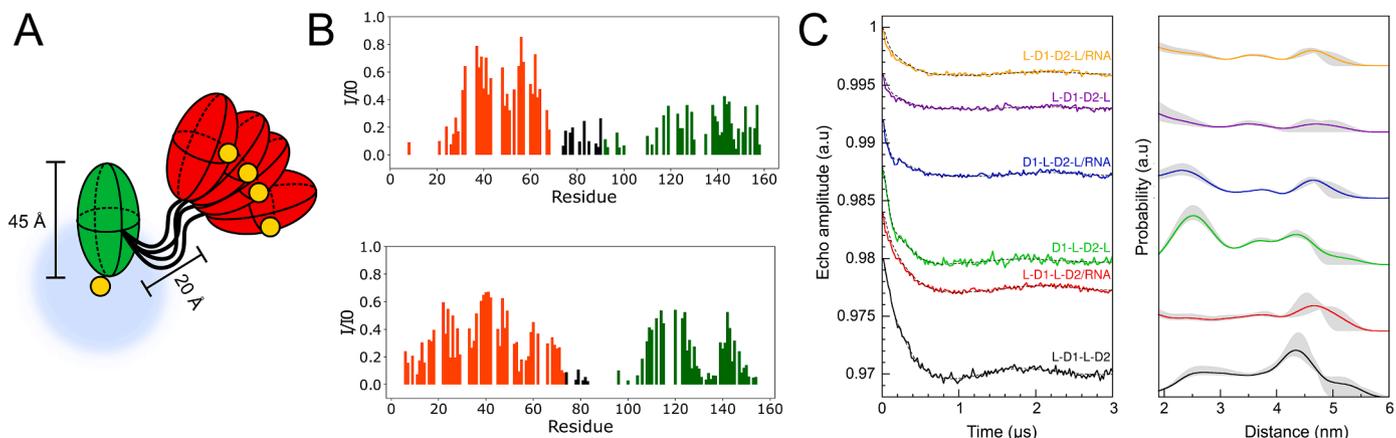


Fig. 2. A. Schematic depiction of HYLI-D1-D2. The red and green ellipsoids represent the first and second dsRBD domain respectively. The yellow spheres represent the location of the LBT. Note that the N-terminus of the polypeptide chain enters the dsRBD at the equator of the ellipsoid whereas the C-terminus exits at a pole. B. PRE patterns of the L-D1-D2 (top) and D1-D2-L (bottom) constructs. Red bars correspond to D1, black bars to the linker and green bars to D2. C. DEER time traces and distance distributions for protein constructs with two Gd(III) loaded LBTs. The gray area in the distance distributions correspond to the error estimated using the DeerAnalysis validation tool.

intermediate exchange in NMR [7,18]. In the presence of a RNA construct corresponding to the sequence of the miR172 precursor (pri-miR172) (Supplementary Figure 6 [27]), the longer distances found in both free L[Gd(III)]-D1-L[Gd(III)]-D2 and D1-L[Gd(III)]-D2-L[Gd(III)] are increased by about 0.5, presumably due to the bound RNA hindering the LBTs from coming closer. The broad distance distributions of the L[Gd(III)]-D1-D2-L[Gd(III)] are not modified, indicating that the protein as a whole remains flexible, with D1 more strongly bound to RNA and D2 more loose and free to participate in interactions with other partners.

3.3. Modelling the conformational propensities of HYLI

The experimental results obtained show that in spite of being structurally independent within the full protein, the two domains of

HYLI do have restrictions in terms of their relative orientations and distances. In order to understand the structural basis of these restrictions we modelled an ensemble of physically plausible protein conformations to then find the ones that best explain the observed interactions. For modeling, we considered the domains as rigid bodies and allowed for conformational flexibility mainly on the linker. We constructed 100 initial models of L-D1-D2-L with the software Modeller using the crystal structures for each domain of HYLI (3ADG and 3ADJ for D1 and D2, respectively [7]) and the crystal structure of a Ca(II) loaded LBT (PDB 1TJB [43]) as templates. The best model was selected and an ensemble of conformers was then generated starting from this model by randomizing the Phi/Psi angles of the residues corresponding to the interdomain linker region (residues 90–104) and those corresponding to the linker between the protein and the LBT (residues 17,18,175 and 176). The protocol for angle randomization was programmed using

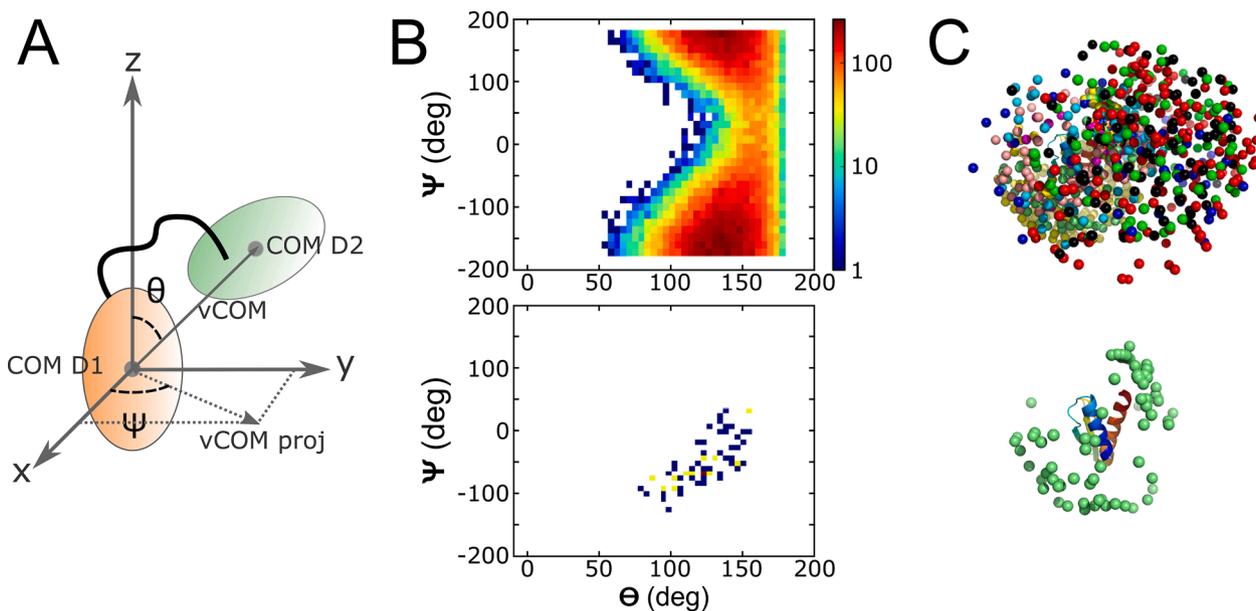


Fig. 3. A. Angle definitions used in the description of relative domain orientations. The reference frame is defined by the moment of inertia tensor of D1. B. Histograms for the distribution of orientations in the whole ensemble (top) and in the subensemble that best fits the PRE data (bottom). The histograms are color coded based on the number of structures in the ensembles with the corresponding Ψ and Θ angles: the color bar indicate the number of structures in the top chart, whereas purple pixels denote a single structure and yellow pixels two structures in the bottom chart. C. Sampling of positions of D2 respect to D1 in the whole ensemble (top) and in the subensemble that best fits the PRE data (bottom). D1 is plotted in cartoon representation and the spheres represent the COMs of D2 on different conformers. Only 1:100 COMs are represented for the whole ensemble, which are color coded according to the subensemble they belong to.

PyRosetta, as described in the Methods section. Using this protocol, we generated an ensemble of 50,000 conformers.

To evaluate whether this protocol allows a good coverage of the available conformational space, we measured the relative orientation of the two domains as follows. We calculated the center of mass of each domain (COM) and the vector that connects the COM of D1 with the COM of D2 (vCOM) (Fig. 3A). In order to define a reference orientation for the protein, we calculated the moment of inertia tensor of D1 in each conformer. We used the eigenvectors of the inertia tensor of D1 as reference frame. We then calculated the projections of the vCOM on this reference frame. This allows the definition of the relative conformation of the domains using a spherical coordinate system, where θ and ψ have the usual meaning (see Fig. 3A). Inspection of the histogram of these angles calculated for each of the conformers shows that the available conformational space as defined in these coordinates is restricted. The length of the linker does not allow D2 to go above D1 (θ values are always greater than ca. 55°). Also, the position where the linker exits the D1 ellipsoid and enters the D2 ellipsoid creates a further restriction, a forbidden region for θ values smaller than ca. 120° around $\psi \approx 0$. In all, our protocol generates a good coverage of the conformational space available for the relative location of the protein domains (Fig. 3B).

We then calculated the distances from each residue to both Gd(III) in each conformer. The R_{2M}^{PRE} and expected I/I_0 values were calculated from those distances as explained in the Materials and Methods section (Supplementary Figure 7). We observed that the calculated PRE profile of the nearby domain was very similar for all conformers and in good agreement with the experimental data. The conformation of the linker between the LBT and the domain was also randomized, so this result is not trivial. Therefore we ignored the effects on the proximal domain. The largest variations were found on the distant domain, and the results of the calculated PRE on both the L-D1-D2 and D1-D2-L constructs were simultaneously used in the conformation clustering step. In this way we included the restraints from both constructs to define the conformational space explored by the protein. This is shown in Supplementary Figure 7, where the data as a function of the protein sequence is displayed duplicated: the left half of each panel corresponds to the L-D1-D2 construct and the right half to the D1-D2-L construct. The experimental data in red corresponds to the distal domain and the data in black corresponds to the proximal domain.

To select the structures from the ensemble that are in better agreement with the experimental data, we decided to cluster them in groups of conformations with similar I/I_0 patterns, dividing the whole set in three clusters. We then analyzed the correlation between each cluster and the experimental data. For this, the distance between the PREs calculated for each model and the experimental data were calculated and plotted. Visual analysis showed that the cluster containing most structures had the models with the largest differences with respect to the experimental PRE data. Those structures were then eliminated and the remaining structures were combined and clustered again. This process resulted in 10 new clusters. We analyzed the filtered dataset using the same pipeline and found a single cluster (cluster I, see Supplementary Figure 8) that shows the best correlation to the experimental PRE values. The angular distribution of this cluster shows a preferential location of domain 2 with respect to domain 1 (Fig. 3C).

4. Discussion

Individual domains in multidomain proteins are often connected by linker sequences that are predicted to be mostly disordered. These linker regions can function as passive spacers, but can also include molecular recognition elements that participate in interactions. The existence of sequence ensemble relationships, where conservation of features denotes an evolutionary pressure on otherwise disordered linker regions that show low conservation, was recently demonstrated on a viral protein [2].

MicroRNA biogenesis is carried out by protein complexes that

include multi domain proteins containing exclusively dsRBDs. The precise function of these helper proteins is currently not well understood, but their importance in the efficiency of the processing reactions is well established [44–48].

We found that the length of the linkers connecting the dsRBDs in double-dsRBD proteins from plant is strikingly conserved (an exception being DRB4, see below), suggesting a functional requirement. In our previous work with HYL1 we determined that the linker region is flexible [18]. Our current analysis on a large set of HYL1 homologs show little sequence conservation along the linker. This led us to think that the linker region acts in this protein as a fixed length tether for both domains and prompted us to study the conformational space that it allows for both dsRBDs.

We resorted to magnetic resonance methods to provide a quantitative description of the influence of the linker connecting the two dsRBDs of HYL1 in the restriction of the conformational space allowed for interdomain movement. Construction of an ensemble of conformers allowing freedom to the linker residues backbone angles shows that the linker length limits the accessible locations of the domains relative to each other. By fitting simulated PRE profiles to our experimental data we found that the actual conformational space is even more restricted with respect to what is physically allowed.

HYL1 has been proposed to dimerize through the dsRBD2 region [7–9]. We have not observed this behavior in our samples, presumably due to differences in the limits of the constructs used. But in a HYL1 dimer, the highly conserved linker length and the observed restriction in conformational freedom could optimize the relative positioning of the two free dsRBD1 to bind the precursor miRNA substrate.

The regions separating the dsRBDs in miRNA processing helper proteins have a large variety of lengths and conformations. Despite being separated by a 43 residues linker, the two domains of DGCR8 are not independent. In this case the linker adopts helical structures that pack together with an extra C-terminal helix in domain 2 to adopt a compact structure [14]. In the case of RDE4, the dsRBDs have additional structural elements within the relatively long (63 residues) linker, but do not interact with each other [15]. The linker region was also shown to be essential for interaction with Dicer [49] and in vivo gene silencing in *C. elegans* [46]. This contrasts with the case of HYL1 and the other plant DRB proteins.

DRB4, the dsRBD helper protein for DCL4, shares the same domain architecture as HYL1. But in this case the linker between dsRBDs is much shorter, only 9 residues. Deshmukh and collaborators demonstrated that while both domains of DRB4 are independent structural units in solution, the short linker restricts their relative orientation, and this orientation is important for dsRNA binding [40]. They further showed that the linker hinders dsRNA recognition by the domain that follows it. In all, the linker appears to play a role in substrate selection and function of DRB4. A work from Sattler and collaborators demonstrated the influence of linker length on the affinity, binding mode and function of *D. melanogaster* Loqs [50]. The authors showed that shortening of the linker leads to lower affinity and more pronounced sliding of the dsRBDs over dsRNA. They further argue that a shorter linker makes difficult for the protein to arrange both dsRBDs conveniently over a rigid dsRNA molecule and suggest that a minimal linker length is necessary for a functional interaction of two tandem dsRBDs. In the case of HYL1, we found by DEER that binding to RNA does not alter the distance distribution of the domains, suggesting that the linker length, despite being larger than that of DRB4, also limits simultaneous binding of substrate RNA by the two dsRBDs. dsRBDs appear to slide along dsRNA [51], and the dynamics of this sliding movement in double dsRBD proteins correlates with the length of the linker connecting the domains. TRBP, with a 61 residue linker shows more sliding than PACT or Staufen, where the RNA binding dsRBDs are separated by 25 residues [52]. The selection of a 15 residue long linker in plant DRBs could also influence the sliding activity of these proteins over substrate RNAs.

The human Dicer-TRBPD3 complex has recently been solved by cryo-

EM [53]. The structure shows the location of TRBPD3, and hints the possible location of the two N-terminal dsRNA binding domains. The distance between the domains is large, requiring the long linker present between the domains. It is therefore unlikely that binding of HYL1 to DCL1 of DRB4 to DCL4 proceeds in a similar fashion, considering the little flexibility conferred by the much shorter linkers.

In summary, the linker region between dsRBD domains in plant DRB proteins has a highly conserved length of 17 residues. This linker restricts the conformational freedom of the domains. When the protein is bound to dsRNA, the distance distribution of the two domains is not altered, suggesting that both domains cannot bind simultaneously to dsRNA. The conformational sampling of the domains as determined from PRE measurements is even more restricted than the space allowed by the linker length.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by grants PICT 2012-1702, PICT 2013-3281 and PICT 2017-2807 from ANPCyT, Argentina, and the mobility grants PICT-CNRS DEERNA and ECOS-Sud A14B02. FCM received a doctoral Fellowship from CONICET, Argentina. We benefited from the biophysics platform of the Institute for Integrative Biology of the Cell supported by French Infrastructure for Integrated Structural Biology (FRISBI, ANR-10-INBS-0005), the Infrastructure for Biology, Health and Agronomy (IBISA) and the Ile-de-France DIM-ELICIT program.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.jmro.2023.100109](https://doi.org/10.1016/j.jmro.2023.100109).

References

- [1] R. Dias, A. Manny, O. Kolaczowski, B. Kolaczowski, Convergence of domain architecture, structure, and ligand affinity in animal and plant RNA-binding proteins, *Mol. Biol. Evol.* 34 (2017) 1429–1444, <https://doi.org/10.1093/molbev/msx090>.
- [2] N.S. González-Foutel, J. Glavina, W.M. Borcherds, M. Safranchik, S. Barrera-Vilarmau, A. Sagar, A. Estaña, A. Barozet, N.A. Garrone, G. Fernandez-Ballester, C. Blanes-Mira, I.E. Sánchez, G. de Prat-Gay, J. Cortés, P. Bernadó, R.V. Pappu, A. S. Holehouse, G.W. Daughdrill, L.B. Chemes, Conformational buffering underlies functional selection in intrinsically disordered protein regions, *Nat. Struct. Mol. Biol.* 29 (2022) 781–790, <https://doi.org/10.1038/s41594-022-00811-w>.
- [3] G. Masliah, P. Barraud, F.H.-T. Allain, RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence, *Cell. Mol. Life Sci.* (2012), <https://doi.org/10.1007/s00018-012-1119-x>.
- [4] K.-Y. Chang, A. Ramos, The double-stranded RNA-binding motif, a versatile macromolecular docking platform, *FEBS J.* 272 (2005) 2109–2117, <https://doi.org/10.1111/j.1742-4658.2005.04652.x>.
- [5] R. Stefl, F.C. Oberstrass, J.L. Hood, M. Jourdan, M. Zimmermann, L. Skrisovska, C. Maris, L. Peng, C. Hofr, R.B. Emeson, F.H.-T. Allain, The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove, *Cell* 143 (2010) 225–237, <https://doi.org/10.1016/j.cell.2010.09.026>.
- [6] M.L. Glegghorn, C. Gong, C.L. Kielkopf, L.E. Maquat, Staufein1 dimerizes through a conserved motif and a degenerate dsRNA-binding domain to promote mRNA decay, *Nat. Struct. Mol. Biol.* 20 (2013) 515–524, <https://doi.org/10.1038/nsmb.2528>.
- [7] S.W. Yang, H.-Y. Chen, J. Yang, S. Machida, N.-H. Chua, Y.A. Yuan, Structure of Arabidopsis hyponastic leaves1 and its molecular implications for miRNA processing, *Structure* 18 (2010) 594–605, <https://doi.org/10.1016/j.str.2010.02.006>.
- [8] X. Yang, W. Ren, Q. Zhao, P. Zhang, F. Wu, Y. He, Homodimerization of HYL1 ensures the correct selection of cleavage sites in primary miRNA, *Nucleic Acids Res.* 42 (2014) 12224–12236, <https://doi.org/10.1093/nar/gku907>.
- [9] N.P. Achkar, S.K. Cho, C. Poulsen, A.L. Arce, D.A. Re, A.J. Giudicatti, E. Karayekov, M.Y. Ryu, S.W. Choi, J. Harholt, J.J. Casal, S.W. Yang, P.A. Manavella, A quick HYL1-dependent reactivation of microRNA production is required for a proper developmental response after extended periods of light deprivation, *Dev. Cell* 46 (2018) 236–247, <https://doi.org/10.1016/j.devcel.2018.06.014>, e6.
- [10] A. Heyam, D. Lagos, M. Plevin, Dissecting the roles of TRBP and PACT in double-stranded RNA recognition and processing of noncoding RNAs: roles of TRBP and PACT in double-stranded RNA recognition, *WIREs RNA* 6 (2015) 271–289, <https://doi.org/10.1002/wrna.1272>.
- [11] D.A. Ré, P.L.M. Lang, C. Yones, A.L. Arce, G. Stegmayer, D. Milone, P.A. Manavella, Alternative usage of miRNA-biogenesis co-factors in plants at low temperatures, *Development* (2019), dev.172932, <https://doi.org/10.1242/dev.172932>.
- [12] M.-H. Han, S. Goud, L. Song, N. Fedoroff, The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 1093–1098, <https://doi.org/10.1073/pnas.0307969100>.
- [13] R. Mencia, L. Gonzalo, I. Tossolini, P.A. Manavella, Keeping up with the miRNAs: current paradigms of the biogenesis pathway, *J. Exp. Bot.* (2022) erac322, <https://doi.org/10.1093/jxb/erac322>.
- [14] S.Y. Sohn, W.J. Bae, J.J. Kim, K.-H. Yeom, V.N. Kim, Y. Cho, Crystal structure of human DGCR8 core, *Nat. Struct. Mol. Biol.* 14 (2007) 847–853, <https://doi.org/10.1038/nsmb1294>.
- [15] S.C. Chiliveri, M.V. Deshmukh, Structure of RDE-4 dsRBDs and mutational studies provide insights into dsRNA recognition in the *Caenorhabditis elegans* RNAi pathway, *Biochem. J.* 458 (2014) 119–130, <https://doi.org/10.1042/BJ20131347>.
- [16] C. Lu, N. Fedoroff, A mutation in the Arabidopsis HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin, *Plant Cell* 12 (2000) 2351–2366.
- [17] F. Wu, L. Yu, W. Cao, Y. Mao, Z. Liu, Y. He, The N-terminal double-stranded RNA binding domains of Arabidopsis hyponastic leaves1 are sufficient for pre-microRNA processing, *Plant Cell* 19 (2007) 914–925, <https://doi.org/10.1105/tpc.106.048637>.
- [18] R.M. Rasia, J.L. Mateos, N.G. Bologna, P. Burdisso, L. Imbert, J.F. Palatnik, J. Boisbouvier, Structure and RNA interactions of the plant microRNA processing-associated protein HYL1, *Biochemistry* (2010).
- [19] P. Burdisso, F. Milia, A.L. Schapiro, N.G. Bologna, J.F. Palatnik, R.M. Rasia, Structural determinants of Arabidopsis thaliana hyponastic leaves 1 function in vivo, *PLoS ONE* 9 (2014), e113243, <https://doi.org/10.1371/journal.pone.0113243>.
- [20] A.D. Milov, A.B. Ponomarev, Y.D. Tsvetkov, Double resonance in electron spin echo: hloled biradical systems and the sensitized photolysis of decalin, (1984).
- [21] R.E. Martin, M. Pannier, F. Diederich, V. Gramlich, M. Hubrich, H.W. Spiess, Determination of end-to-end distances in a series of TEMPO diradicals of up to 2.8 nm length with a new four-pulse double electron resonance experiment, *Angew. Chem. Int. Ed.* 37 (1998) 2833–2837, [https://doi.org/10.1002/\(SICI\)1521-3773\(19981102\)37:20<2833::AID-ANIE2833>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1521-3773(19981102)37:20<2833::AID-ANIE2833>3.0.CO;2-7).
- [22] F.C. Mascali, H.Y.V. Ching, R.M. Rasia, S. Un, L.C. Tabares, Using genetically encodable self-assembling Gd^{III} spin labels to make in-cell nanometric distance measurements, *Angew. Chem. Int. Ed.* 55 (2016) 11041–11043, <https://doi.org/10.1002/anie.201603653>.
- [23] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, X. MEGA, Molecular evolutionary genetics analysis across computing platforms, *Mol. Biol. Evol.* 35 (2018) 1547–1549, <https://doi.org/10.1093/molbev/msy096>.
- [24] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, *BMC Bioinform.* 5 (2004) 113, <https://doi.org/10.1186/1471-2105-5-113>.
- [25] K.D. Daughtry, L.J. Martin, A. Sarraju, B. Imperiali, K.N. Allen, Tailoring encodable lanthanide-binding tags as MRI contrast agents, *ChemBioChem* 13 (2012) 2567–2574, <https://doi.org/10.1002/cbic.201200448>.
- [26] K. Houben, D. Marion, N. Tarbouriech, R.W.H. Ruijgrok, L. Blanchard, Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family, *J. Virol.* 81 (2007) 6807–6816, <https://doi.org/10.1128/JVI.00338-07>.
- [27] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res.* 31 (2003) 3406–3415, <https://doi.org/10.1093/nar/gkg595>.
- [28] E. Matalon, T. Huber, G. Hagelueken, B. Graham, V. Frydman, A. Feintuch, G. Otting, D. Goldfarb, Gadolinium(III) spin labels for high-sensitivity distance measurements in transmembrane helices, *Angew. Chem. Int. Ed.* 52 (2013) 11831–11834, <https://doi.org/10.1002/anie.201305574>.
- [29] G. Jeschke, G. Panek, A. Godt, A. Bender, H. Paulsen, Data analysis procedures for pulse ELDOR measurements of broad distance distributions, *Appl. Magn. Reson.* 26 (2004) 223–244, <https://doi.org/10.1007/BF03166574>.
- [30] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. R. Timmel, D. Hilger, H. Jung, DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data, *Appl. Magn. Reson.* 30 (2006) 473–498, <https://doi.org/10.1007/BF03166213>.
- [31] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR* 6 (1995) 277–293.
- [32] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, *Proteins: Struct., Function Bioinform.* 59 (2005) 687–696, <https://doi.org/10.1002/prot.20449>.

- [33] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.* 234 (1993) 779–815, <https://doi.org/10.1006/jmbi.1993.1626>.
- [34] S. Chaudhury, S. Lyskov, J.J. Gray, PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta, *Bioinformatics* 26 (2010) 689–691, <https://doi.org/10.1093/bioinformatics/btq007>.
- [35] P. Bernadó, L. Blanchard, P. Timmins, D. Marion, R.W.H. Ruigrok, M. Blackledge, A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 17002–17007, <https://doi.org/10.1073/pnas.0506202102>.
- [36] M. Fragai, C. Luchinat, G. Parigi, E. Ravera, Conformational freedom of metalloproteins revealed by paramagnetism-assisted NMR, *Coord. Chem. Rev.* 257 (2013) 2652–2667, <https://doi.org/10.1016/j.ccr.2013.02.009>.
- [37] A. Abyzov, N. Salvi, R. Schneider, D. Maurin, R.W.H. Ruigrok, M.R. Jensen, M. Blackledge, Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent NMR relaxation, *J. Am. Chem. Soc.* 138 (2016) 6240–6251, <https://doi.org/10.1021/jacs.6b02424>.
- [38] The UniProt Consortium, UniProt: a hub for protein information, *Nucleic Acids Res.* 43 (2015) D204–D212, <https://doi.org/10.1093/nar/gku989>.
- [39] M. Clavel, T. Pélissier, T. Montavon, M.-A. Tschopp, M.-N. Pouch-Pélissier, J. Descombin, V. Jean, P. Dunoyer, C. Bousquet-Antonelli, J.-M. Deragon, Evolutionary history of double-stranded RNA binding proteins in plants: identification of new cofactors involved in easiRNA biogenesis, *Plant Mol. Biol.* 91 (2016) 131–147, <https://doi.org/10.1007/s11103-016-0448-9>.
- [40] S.C. Chiliveri, R. Aute, U. Rai, M.V. Deshmukh, DRB4 dsRBD1 drives dsRNA recognition in *Arabidopsis thaliana* tasi/siRNA pathway, *Nucleic Acids Res.* 45 (2017) 8551–8563, <https://doi.org/10.1093/nar/gkx481>.
- [41] G. Otting, Prospects for lanthanides in structural biology by NMR, *J. Biomol. NMR* 42 (2008) 1–9, <https://doi.org/10.1007/s10858-008-9256-0>.
- [42] K. Barthelmes, A.M. Reynolds, E. Peisach, H.R.A. Jonker, N.J. DeNunzio, K. N. Allen, B. Imperiali, H. Schwalbe, Engineering encodable lanthanide-binding tags into loop regions of proteins, *J. Am. Chem. Soc.* (2011).
- [43] M. Nitz, M. Sherawat, K.J. Franz, E. Peisach, K.N. Allen, B. Imperiali, Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide, *Angew. Chem. Int. Ed. Engl.* 43 (2004) 3682–3685, <https://doi.org/10.1002/anie.200460028>.
- [44] R.C. Wilson, A. Tambe, M.A. Kidwell, C.L. Noland, C.P. Schneider, J.A. Doudna, Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis, *Mol. Cell* 57 (2015) 397–407, <https://doi.org/10.1016/j.molcel.2014.11.030>.
- [45] Z. Dong, M.-H. Han, N. Fedoroff, The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 9970–9975, <https://doi.org/10.1073/pnas.0803356105>.
- [46] D. Blanchard, P. Parameswaran, J. Lopez-Molina, J. Gent, J.F. Saynuk, A. Fire, On the nature of in vivo requirements for *rde-4* in RNAi and developmental pathways in *C. elegans*, *RNA Biol.* 8 (2011) 458–467, <https://doi.org/10.4161/rna.8.3.14657>.
- [47] J. Han, Y. Lee, K.-H. Yeom, J.-W. Nam, I. Heo, J.-K. Rhee, S.Y. Sohn, Y. Cho, B.-T. Zhang, V.N. Kim, Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex, *Cell* 125 (2006) 887–901, <https://doi.org/10.1016/j.cell.2006.03.043>.
- [48] A.C. Partin, K. Zhang, B.-C. Jeong, E. Herrell, S. Li, W. Chiu, Y. Nam, Cryo-EM structures of human Drosha and DGCR8 in complex with primary microRNA, *Mol. Cell* 78 (2020) 411–422, <https://doi.org/10.1016/j.molcel.2020.02.016>, e4.
- [49] G.S. Parker, T.S. Maity, B.L. Bass, dsRNA binding properties of RDE-4 and TRBP reflect their distinct roles in RNAi, *J. Mol. Biol.* 384 (2008) 967–979, <https://doi.org/10.1016/j.jmb.2008.10.002>.
- [50] J.-N. Tants, S. Fesser, T. Kern, R. Stehle, A. Geerloff, C. Wunderlich, M. Juen, C. Hartmüller, R. Böttcher, S. Kunzelmann, O. Lange, C. Kreutz, K. Förstemann, M. Sattler, Molecular basis for asymmetry sensing of siRNAs by the *Drosophila* Loqs-PD/Dcr-2 complex in RNA interference, *Nucleic Acids Res.* 45 (2017) 12536–12550, <https://doi.org/10.1093/nar/gkx886>.
- [51] H.R. Koh, M.A. Kidwell, K. Ragunathan, J.A. Doudna, S. Myong, ATP-independent diffusion of double-stranded RNA binding proteins, *Proc. Natl. Acad. Sci. U.S.A.* 110 (2013) 151–156, <https://doi.org/10.1073/pnas.1212917110>.
- [52] X. Wang, L. Vukovic, H.R. Koh, K. Schulten, S. Myong, Dynamic profiling of double-stranded RNA binding proteins, *Nucleic Acids Res.* 43 (2015) 7566–7576, <https://doi.org/10.1093/nar/gkv726>.
- [53] Z. Liu, J. Wang, H. Cheng, X. Ke, L. Sun, Q.C. Zhang, H.-W. Wang, Cryo-EM structure of human dicer and its complexes with a pre-miRNA substrate, *Cell* 173 (2018) 1191–1203, <https://doi.org/10.1016/j.cell.2018.03.080>, e12.